



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

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΦΑΡΜΑΚΕΥΤΙΚΗΣ

ΤΟΜΕΑΣ ΦΑΡΜΑΚΕΥΤΙΚΗΣ ΧΗΜΕΙΑΣ

Μεταβολομική μελέτη κυτταρικών σειρών καρκίνου ουροδόχου κύστης με τη χρήση
Φασματοσκοπίας Πυρηνικού Μαγνητικού Συντονισμού και Φασματομετρίας Μάζας

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ

ΠΑΝΑΓΙΩΤΑΚΗΣ ΑΡΙΣΤΕΙΔΗΣ

ΧΗΜΙΚΟΣ

Επιβλέποντες καθηγητές: Εμμανουήλ Μικρός, Ευάγγελος Γκίκας

ΑΘΗΝΑ, 2019

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΦΑΡΜΑΚΕΥΤΙΚΗΣ

ΤΟΜΕΑΣ ΦΑΡΜΑΚΕΥΤΙΚΗΣ ΧΗΜΕΙΑΣ

Μεταβολομική μελέτη κυτταρικών σειρών καρκίνου ουροδόχου κύστης με τη χρήση Φασματοσκοπίας Πυρηνικού Μαγνητικού Συντονισμού και Φασματομετρίας Μάζας

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ

Τίτλος Προγράμματος-Ειδίκευση: «Φαρμακευτική Ανάλυση – Έλεγχος Ποιότητας»

Παναγιωτάκης Αριστείδης, Χημικός

Αριθμός Μητρώου: 161403

Επιβλέποντες Καθηγητές:

Εμμανουήλ Μικρός, Καθηγητής του Τομέα Φαρμακευτικής Χημείας ΕΚΠΑ

Ευάγγελος Γκίκας, Επίκουρος Καθηγητής του Τομέα Φαρμακευτικής Χημείας ΕΚΠΑ

Τριμελής συμβουλευτική Επιτροπή:

1. Εμμανουήλ Μικρός, Καθηγητής του Τομέα Φαρμακευτικής Χημείας ΕΚΠΑ

2. Ευάγγελος Γκίκας, Επίκουρος Καθηγητής του Τομέα Φαρμακευτικής Χημείας ΕΚΠΑ

3. Ιωάννης Ντότσικας, Επίκουρος Καθηγητής του Τομέα Φαρμακευτικής Χημείας ΕΚΠΑ

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή κ. Εμμανουήλ Μικρό, ο οποίος έχει την ευθύνη ανάθεσης του θέματος και επίβλεψη της ερευνητικής εργασίας, για την άριστη συνεργασία και καθοδήγηση. Η μεγάλη θεωρητική και εργαστηριακή εμπειρία του υπήρξε καταλυτική στην πραγματοποίηση της παρούσας ερευνητικής εργασίας.

Στη συνέχεια, θέλω να εκφράσω τις θερμές ευχαριστίες μου στον Επίκουρο Καθηγητή κ. Ευάγγελο Γκίκα, για την καθοδήγηση, τις επισημάνσεις και για την συμπαράστασή του όλον αυτόν τον καιρό. Οι πολλές θεωρητικές του γνώσεις και η μεγάλη εργαστηριακή του εμπειρία αποτέλεσαν αρωγό στην περάτωση της παρούσας ερευνητικής εργασίας.

Επίσης, ευχαριστώ θερμά και τον Επίκουρο Καθηγητή κ. Ιωάννη Ντότσικα, μέλος της τριμελούς επιτροπής, για την βοήθειά του.

Επιπλέον, ευχαριστώ θερμά τον Επίκουρο Καθηγητή κ. Δημήτριο Στραβοπόδη και το εργαστήριο του, για την παροχή των απαραίτητων δειγμάτων για την πραγματοποίηση της εν λόγω εργασίας.

Ιδιαίτερως ευχαριστώ την Δρ. Δήμητρα Μπενάκη για την πολύτιμη βοήθειά της κατά την πραγματοποίηση των πειραμάτων και τη λήψη των φασμάτων NMR και κυρίως για τις συμβουλές και το χρόνο που μου αφιέρωσε.

Ακόμη, ευχαριστώ πολύ τον Δρ. Κατσίκη Σωτήριο για την σημαντική συμβολή του στην αποπεράτωση αυτής της εργασίας, τις γνώσεις του και τη στήριξή του.

Ιδιαίτερες ευχαριστίες αποδίδω στην υποψήφια διδάκτωρ Ηλίου Αικατερίνη για την πολύτιμη και έμπρακτη υποστήριξή της στην εκπόνηση της εργασίας.

Τέλος, δεν θα μπορούσα να μην ευχαριστήσω την οικογένειά μου και τους φίλους μου για την υποστήριξη και την υπομονή που έδειξαν και πάνω από όλα για την ενθάρρυνσή τους όλο αυτό το διάστημα.

ΠΕΡΙΕΧΟΜΕΝΑ

| | |
|--|----|
| ΕΥΧΑΡΙΣΤΙΕΣ..... | 4 |
| ΠΕΡΙΕΧΟΜΕΝΑ | 5 |
| ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ | 7 |
| ABSTRACT | 8 |
| ΠΕΡΙΛΗΨΗ..... | 9 |
| ΚΕΦΑΛΑΙΟ 1..... | 11 |
| ΜΕΤΑΒΟΛΟΜΙΚΗ ΚΑΙ ΑΝΑΛΥΤΙΚΕΣ ΤΕΧΝΙΚΕΣ..... | 11 |
| 1.1 Μεταβολομική | 11 |
| 1.2 Μεταβολομική και καρκίνος | 11 |
| 1.3 Μεταβολομική και καρκίνος της ουροδόχου κύστης | 12 |
| 1.4 Αναλυτικές τεχνικές | 13 |
| 1.4.1 Φασματοσκοπία πυρηνικού μαγνητικού συντονισμού (NMR) | 14 |
| 1.4.2 Φασματομετρία μάζας (MS) | 15 |
| ΚΕΦΑΛΑΙΟ 2..... | 18 |
| ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ | 18 |
| 2.1 Οργανολογία και αντιδραστήρια..... | 18 |
| 2.2 Δείγματα..... | 18 |
| 2.3 Επιλογή πρωτοκόλλου εκχύλισης..... | 19 |
| 2.4 Ανάλυση με NMR..... | 21 |
| 2.5 Ανάλυση με φασματομετρία μάζας..... | 23 |
| ΚΕΦΑΛΑΙΟ 3..... | 26 |
| ΕΠΕΞΕΡΓΑΣΙΑ ΦΑΣΜΑΤΩΝ..... | 26 |
| 3.1 Επεξεργασία φασμάτων NMR..... | 26 |
| 3.2 Επεξεργασία φασμάτων MS..... | 28 |
| ΚΕΦΑΛΑΙΟ 4..... | 30 |
| ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ..... | 30 |
| 4.1 Ανάλυση κύριων συνιστωσών (PCA) | 30 |
| 4.2 Διακρίνουσα ανάλυση μερικών ελαχίστων τετραγώνων (PLS-DA)..... | 35 |
| 4.3 Ορθογώνια διακρίνουσα ανάλυση μερικών ελαχίστων τετραγώνων (OPLS-DA) .. | 35 |

| | | |
|-----|---|----|
| 4.4 | Επικύρωση μοντέλων PLS-DA..... | 36 |
| 4.5 | Επικύρωση υποψήφιων βιοδεικτών | 37 |
| 4.6 | Επιλογή των χαρακτηριστικών που συμβάλλουν στην ομαδοποίηση | 38 |
| | ΚΕΦΑΛΑΙΟ 5..... | 40 |
| | ΑΠΟΤΕΛΕΣΜΑΤΑ | 40 |
| 5.1 | Φασματοσκοπία NMR..... | 40 |
| 5.2 | Χημειομετρική ανάλυση δεδομένων NMR..... | 42 |
| 5.3 | Φασματομετρία μάζας..... | 59 |
| 5.4 | Χημειομετρική ανάλυση δεδομένων MS..... | 62 |
| | ΚΕΦΑΛΑΙΟ 6..... | 74 |
| | ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ..... | 74 |
| | ΠΑΡΑΡΤΗΜΑ | 76 |
| | Παράρτημα 1..... | 76 |
| | Παράρτημα 2..... | 83 |
| | ΒΙΒΛΙΟΓΡΑΦΙΑ..... | 87 |

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

| | |
|---------|--|
| NMR | Μαγνητικός Πυρηνικός Συντονισμός |
| LC | Υγροχρωματογραφία |
| MS | Φασματομετρία Μάζας |
| m/z | Μάζα/Φορτίο |
| HPLC | Υγροχρωματογραφία Υψηλής Απόδοσης |
| UHPLC | Υγροχρωματογραφία Υπερηψηλής Απόδοσης |
| PCA | Ανάλυση Κύριων Συνιστωσών |
| PLS-DA | Διακρίνουσα Ανάλυση Μερικών Ελαχίστων Τετραγώνων |
| OPLS-DA | Ορθογώνια Διακρίνουσα Ανάλυση Μερικών Ελαχίστων Τετραγώνων |
| ROC | Λειτουργικός Χαρακτηριστικός Δέκτης |
| AUC | Εμβαδό Κάτω από τη Καμπύλη |
| VIP | Σημαντικές Μεταβλητές για την Προβολή |
| QC | Δείγμα Ελέγχου Ποιότητας |
| HRMS | Φασματομετρία Μάζας Υψηλής Διακριτικής Ικανότητας |
| RT4 | Grade 1 |
| RT112 | Grade 2 |
| T24 | Grade 3 |
| TCCSUP | Grade 4 |

ABSTRACT

Bladder cancer is the most common urological malignancy. Early diagnosis of bladder cancer is crucial to improve patient outcomes. Currently, metabolomics is a potential technique that can be used to detect bladder cancer.

Plasticity and reprogramming of cellular metabolism have proved to act as strong forces in tumorigenesis. Malignant hallmarks, such as cell survival under stress conditions, as well as the ability to utilize nutrients and successfully encounter high-energy demands, are tightly correlated with metabolic alterations, thus indicating the important roles of metabolic landscapes in Cancer Biology and therapy. Metabolic activities may significantly differ between different malignancy stages (grades) of the same type of cancer, leading to distinct metabolic networks and metabolomes. In this study, we present a comparative NMR- and LC-MS/MS-mediated untargeted metabolic profiling of human urothelial bladder cancer (hUBC) to landscape the metabolic alterations tumor cells are subjected during cancer progression, while we also discuss methodological issues on the metabolomic analysis herein performed.

The metabolic landscapes of grade 1 to grade 4 human urinary bladder cancer cell lines were extensively investigated, using four different cell lines; the RT4 (grade 1), RT112 (grade 2), T24 (grade 3) and TCCSUP (grade 4) ones (10 individual samples from each cell line were examined). The obtained results unveiled diverse and malignancy-grade -specific metabolite collections, critically implicated in amino acid metabolism, tricarboxylic acid (TCA) cycle and energy metabolism, as well as purine and pyrimidine metabolism. A small, but significant, increase was observed in Grade 2 (RT112) compared to grade 1 (RT4) cells, while grade 3 (T24) cells proved to carry surprisingly upregulated levels of most metabolites, with the amino acid metabolic network being emerged as a powerful and promising novel platform for bladder cancer targeted therapy. However, in TCCSUP (grade 4) cells, the levels of most identified metabolites, including those involved in amino acid metabolism, were notably reduced compared to RT112 (grade 2) and T24 (grade 3) cell-line respective ones, likely indicating a “metabolic inversion” of the cellular machinery during late metastasis.

KEYWORDS; metabolomics, “metabolic inversion”, NMR, MS

ΠΕΡΙΛΗΨΗ

Ο καρκίνος της ουροδόχου κύστης είναι η πιο διαδεδομένη κακοήθεια από τους ουρολογικούς τύπους καρκίνου. Η πρόωρη διάγνωσή του είναι μέγιστης σημασίας για την βελτίωση της ποιότητας ζωής των ασθενών. Προς το παρόν η μεταβολομική αποτελεί δυναμική τεχνική που μπορεί να χρησιμοποιηθεί για την ανίχνευση του καρκίνου της ουροδόχου κύστης.

Η πλαστικότητα και η επανατροποποίηση του κυτταρικού μεταβολισμού φαίνεται ότι αποτελούν ισχυρούς παράγοντες στην ογκογένεση. Τα κακοήθη χαρακτηριστικά, όπως η επιβίωση των κυττάρων κάτω από στρεσογόνες συνθήκες, καθώς και η αύξηση κατανάλωσης θρεπτικών συστατικών για την ικανοποίηση των υψηλών ενεργειακών απαιτήσεων, συσχετίζονται στενά με μεταβολικές μετατροπές, υποδηλώνοντας έτσι τον σημαντικό ρόλο της μεταβολομικής στη βιολογία του καρκίνου και στη θεραπεία αυτού. Όσο αφορά τη μεταβολομική, μεταξύ των διαφόρων βαθμίδων κακοήθειας του ίδιου τύπου καρκίνου, παρατηρείται ενίσχυση ή αναστολή της μεταβολικής δραστηριότητας σε μεταβολικά μονοπάτια. Σε αυτήν τη μελέτη, παρουσιάζεται το συγκριτικό μη-στοχευμένο μεταβολομικό προφίλ ανθρώπινου καρκίνου ουροδόχου κύστης(που αποτυπώνεται NMR και LC-MS)ώστε να προσδιορισθούν και να χαρτογραφηθούν οι μεταβολικές αλλαγές των καρκινικών κυττάρων κατά την εξέλιξη του καρκίνου. Ταυτόχρονα συζητούνται θέματα μεθοδολογίας που προέκυψαν κατά την μεταβολομική ανάλυση.

Τα μεταβολικά «τοπία» (landscapes) της βαθμίδας 1 (grade1) έως την βαθμίδα 4 (grade 4) των καρκινικών κυτταρικών σειρών της ανθρώπινης ουροδόχου κύστης διερευνήθηκαν εκτενώς, χρησιμοποιώντας τέσσερις διαφορετικές κυτταρικές σειρές: την RT4 (grade1), RT112 (grade2), T24 (grade3) και TCCSUP (grade4) (μελετήθηκαν δέκα δείγματα από κάθε κυτταρική σειρά). Τα δεδομένα αποκάλυψαν στοιχεία συσχετιζόμενα με κακοήθεια, που εμπλέκονται στον μεταβολισμό των αμινοξέων, του κύκλου των τρικαρβοξυλικών οξέων (TCA) (κύκλος κιτρικού οξέος) και του ενεργειακού μεταβολισμού, καθώς και στον μεταβολισμό πουρινών και της πυριμιδινών. Παρατηρήθηκε πως υπάρχει μικρή αλλά σημαντική διαφορά συγκρίνοντας την κυτταρική σειρά RT112 (grade 2) με την κυτταρική σειρά RT4 (grade 1). Στην κυτταρική σειρά T24 (grade 3) σχεδόν όλοι οι μεταβολίτες βρίσκονταν σε εξαιρετικά αυξημένα επίπεδα συγκέντρωσης, με την μεταβολική πορεία των αμινοξέων να εμφανίζεται ως μία νέα ισχυρά υποσχόμενη πλατφόρμα για την στοχευμένη θεραπεία του καρκίνου της ουροδόχου κύστης. Ωστόσο, στην κυτταρική σειρά TCCSUP(grade 4)τα επίπεδα των μεταβολιτών που ανιχνεύθηκαν, συμπεριλαμβανομένων και εκείνων που συμμετέχουν

στα βιοχημικά μονοπάτια των αμινοξέων, ήταν αισθητά μειωμένα συγκρινόμενα και με την κυτταρική σειρά RT112 (grade 2) και με την κυτταρική σειρά T24 (grade 3) που υποδεικνύει «αναστροφή» στη μεταβολική πορεία των κυττάρων κατά τη διάρκεια της τελευταίας μετάστασης.

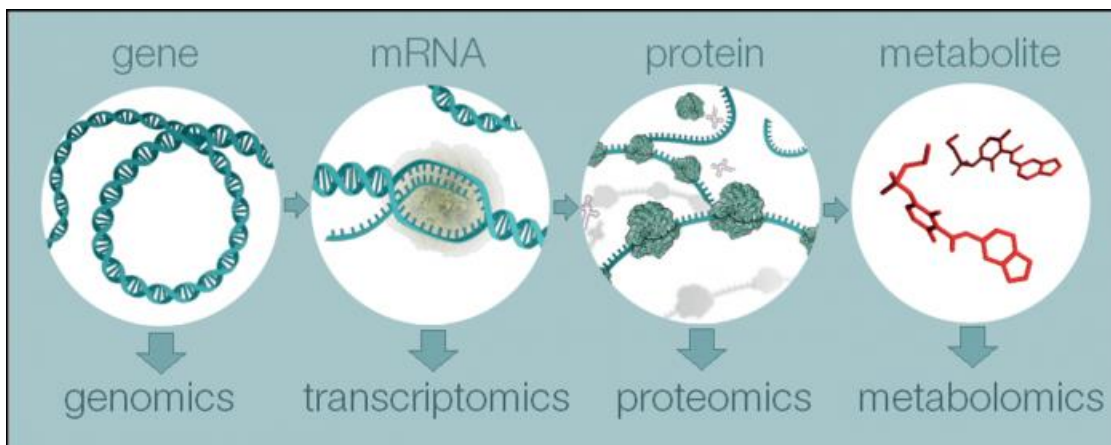
ΛΕΞΕΙΣ-ΚΛΕΙΔΑ: μεταβολομική, «μεταβολομική αναστροφή», φασματοσκοπία πυρηνικού μαγνητικού συντονισμού, φασματομετρία μάζας

ΚΕΦΑΛΑΙΟ 1

ΜΕΤΑΒΟΛΟΜΙΚΗ ΚΑΙ ΑΝΑΛΥΤΙΚΕΣ ΤΕΧΝΙΚΕΣ

1.1 Μεταβολομική

Με τον όρο μεταβολομική αναφερόμαστε στην εκτενή μελέτη που περιλαμβάνει την αναγνώριση και ποσοτικοποίηση μικρών μορίων (μεταβολιτών) που προέρχονται από ένα βιολογικό δείγμα σε μία δεδομένη χρονική στιγμή. Το σύνολο των ανιχνεύσιμων μεταβολιτών ονομάζεται μεταβόλωμα και το βιολογικό δείγμα μπορεί να προέρχεται από κάποιο κύτταρο, ιστό, όργανο καθώς και από βιολογικά υγρά.¹ Η ανάλυση του μεταβόλωματος αποτελεί πάντα μία ιδιαίτερη πρόκληση εξαιτίας της ποικιλίας που υπάρχει στη χημική δομή των μεταβολιτών.² Παρά το έντονο επιστημονικό ενδιαφέρον στο πεδίο της μεταβολομικής μόνο τα τελευταία χρόνια οι τεχνολογίες έχουν επιτρέψει την ανάλυση των μεταβολιτών σε επίπεδο αντίστοιχο με αυτό της πρωτεομικής (proteomics), της γονιδιωματικής (genomics) και των μεταγραφομικής (transcriptomics).



Σχήμα 1. Μια επισκόπηση των τεσσάρων πεδίων "omics", από τη γονιδιωματική έως τη μεταβολική

1.2 Μεταβολομική και καρκίνος

Ο καρκίνος είναι μία ασθένεια για το ανθρώπινο είδος που προκαλεί μεγάλο αριθμό θανάτων κάθε χρόνο³. Η μεταβολομική μπορεί να αξιολογήσει ποσοτικά τους μεταβολίτες που προέρχονται από κάποια καρκινική πηγή (κύτταρο, ιστός ή βιολογικό υγρό) και στην ιδανική περίπτωση να ανακαλύψει κάποιον βιοδείκτη⁴. Η προσέγγιση

μέσω μίας μεταβολομικής μελέτης μπορεί να παρέχει σημαντικές πληροφορίες σχετικά με την ογκογένεση, και να οδηγήσει σε μία πιο αυτοματοποιημένη και έγκαιρη διάγνωση⁵.

1.3 Μεταβολομική και καρκίνος της ουροδόχου κύστης

Ο καρκίνος της ουροδόχου κύστης αποτελεί τον έβδομο σε σειρά εμφάνισης καρκίνο παγκοσμίως για τους άντρες και τον δέκατο έβδομο για τις γυναίκες⁶. Η κλασική διάγνωση του καρκίνου της ουροδόχου κύστης γίνεται με κλινική εξέταση και στη συνέχεια με κυστεοσκόπηση⁷. Παρά τις εκτενείς προσπάθειες και τις επενδύσεις που έχουν γίνει για εναλλακτικούς τρόπους διάγνωσης η κυστεοσκόπηση και η κυτολογία εξακολουθούν να παραμένουν οι κύριες επιλογές.^{8,9}

Η μεταβολομική έχει χρησιμοποιηθεί για τη διάγνωση διαφόρων ειδών καρκίνων όπως ο καρκίνος των ωοθηκών,¹⁰ ο καρκίνος του παγκρέατος^{11,12} και της λευχαιμίας¹⁵. Επίσης έχει αποδειχθεί ότι είναι χρήσιμη στον τομέα της ανακάλυψης βιοδεικτών για τύπους καρκίνου που απαιτείται έγκαιρη διάγνωση και πρόγνωση όπως συμβαίνει στην περίπτωση του καρκίνου της ουροδόχου κύστης¹³. Η μεταβολομική αποτελεί μια δυναμική τεχνική για την ανίχνευση του καρκίνου της ουροδόχου κύστης.

Από την βιβλιογραφία¹⁴⁻¹⁷ έχουν παρατηρηθεί δεκατέσσερις μεταβολίτες που χαρακτηρίζονται ως βιοδείκτες για τον καρκίνο της ουροδόχου κύστης. Οι μεταβολίτες αυτοί είναι: γαλακτικό οξύ (2-υδροξυπροπανικό οξύ, lactate), λευκίνη (leucine), βαλίνη (valine), φαινυλαλανίνη (phenylalanine), γλουταμινικό οξύ (glutamate), ιστοιδίνη (histidine), ασπαραγινικό οξύ (aspartic acid), τυροσίνη (tyrosine), σερίνη (serine), ουρακίλη (uracil), υποξανθίνη (hypoxanthine), καρνιτίνη (carnitine), πυροσταφλικό οξύ (pyruvic acid) και κιτρικό οξύ (citric acid).

Ιδιαίτερη αναφορά¹⁸⁻²¹ γίνεται για τα αυξημένα επίπεδα συγκέντρωσης του γαλακτικού οξέος σε καρκινικά δείγματα ουροδόχου κύστης. Αυτό μπορεί να αποδοθεί στο φαινόμενο Warburg σύμφωνα με το οποίο τα καρκινικά κύτταρα παράγουν ATP κατά τη γλυκόλυση και οδηγούν στην αυξημένη παραγωγή γαλακτικού οξέος²². Το πυροσταφλικό οξύ μπορεί να μετατραπεί σε γαλακτικό οξύ μέσω του ενζύμου γαλακτική αφυδρογονάση.

Όσο αφορά το πυροσταφλικό οξύ εμφανίζεται με χαμηλά επίπεδα συγκέντρωσης σε τέτοιου είδους δείγματα.¹⁹ Γενικά όσοι μεταβολίτες είναι «πρόδρομοι» του γαλακτικού οξέος και σχετίζονται με τον μεταβολισμό της γλυκόλυσης εμφανίζουν χαμηλά επίπεδα

συγκέντρωσης. Επιπρόσθετα, τα χαμηλά επίπεδα του πυροσταφυλικού οξέος συσχετίζονται με τα αυξημένα επίπεδα γαλακτικού οξέος.

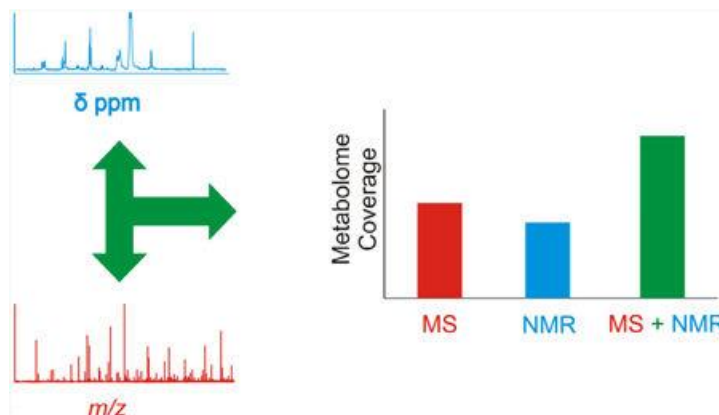
Μειωμένα επίπεδα κιτρικού οξέος και φουμαρικού οξέος που αποτελούν και τα μεταβολικά ενδιάμεσα στην κυτταρική αναπνοή (ιδιαίτερα στον κύκλο του κιτρικού οξέος) παρατηρούνται σε καρκινικά κύτταρα ουροδόχου κύστης²³. Τα χαμηλά επίπεδα των ενδιάμεσων μεταβολιτών στον κύκλο του Krebs υποδεικνύουν την προτίμηση που έχουν τα καρκινικά κύτταρα για τον μεταβολισμό της γλυκόλυσης παρά για αυτόν της οξειδωτικής φωσφορυλίωσης²⁴⁻²⁶.

1.4 Αναλυτικές τεχνικές

Οι κύριες αναλυτικές τεχνικές που χρησιμοποιούνται στη μεταβολομική είναι η φασματοσκοπία πυρηνικού μαγνητικού συντονισμού (nuclear magnetic resonance, NMR) και η φασματομετρία μάζας (mass spectrometry, MS)^{27,28}. Η κάθε μέθοδος έχει τα δικά της πλεονεκτήματα και μειονεκτήματα, είναι όμως σημαντικό ότι μπορούν να εφαρμοστούν συμπληρωματικά²⁹. Ο συνδυασμός των δύο τεχνικών αυτών οδηγεί σε ταχεία αναγνώριση και ποσοτικοποίηση μεταβολιτών που υπάρχουν σε πολύπλοκα δείγματα.

Η φασματοσκοπία πυρηνικού μαγνητικού συντονισμού και η φασματομετρία μάζας αποτελούν τα αναλυτικά εργαλεία που χρησιμοποιούνται κατά κόρον, ξεχωριστά το κάθε ένα εξαιτίας της ευελιξίας, της εφαρμοστικότητας και των ιδιαίτερων δυνατοτήτων τους. Ένα μεγάλο πλεονέκτημα που έχει το NMR είναι ότι απαιτεί ελάχιστη προετοιμασία δείγματος. Είναι εγγενώς ποσοτική μέθοδος, που χαρακτηρίζεται από υψηλή επαναληψιμότητα έχει όμως τον περιορισμό ότι ο μεταβολίτης πρέπει να βρίσκεται σε σχετικά μεγάλη συγκέντρωση ($\geq 1\mu\text{M}$). Από την άλλη πλευρά η φασματομετρία μάζας έχει πολύ υψηλότερη ευαισθησία και μπορεί να πετύχει μεγάλο διαχωρισμό. Η ποσοτικοποίηση όμως στην περίπτωση του MS αποτελεί πρόκληση καθώς ανάλογα με την πολυπλοκότητα του δείγματος παρατηρείται μεταβλητή απόκριση του ανιχνευτή με αποτέλεσμα μειωμένη επαναληψιμότητα και ακρίβεια. Καθώς συχνότατα χρησιμοποιείται η υγροχρωματογραφία (liquid chromatography, LC) ή η αεριοχρωματογραφία (gas chromatography, GC) σε συνδυασμό με τη φασματομετρία μάζας³⁰, η συνολική ακρίβεια παρουσιάζεται συνολικά περισσότερο μειωμένη. Επιπρόσθετα, το NMR παρέχει μεγάλο πλούτο δομικής πληροφορίας ενώ η αντίστοιχη πληροφορία για τη φασματομετρία μάζας είναι πολύ δυσκολότερα ερμηνεύσιμη με το παρόν επίπεδο τεχνολογίας. Συνολικά το NMR

και η φασματομετρία μάζας είναι δύο εξαιρετικά συμπληρωματικές τεχνικές και ο συνδυασμός τους βελτιώνει τη συνολική ποιότητα της μελέτης.^{31–35}



Σχήμα 2.Ο συνδυασμός των δύο τεχνικών καλύπτει καλύτερα ολόκληρο το μεταβόλωμα

1.4.1 Φασματοσκοπία πυρηνικού μαγνητικού συντονισμού (NMR)

Η φασματοσκοπία πυρηνικού μαγνητικού συντονισμού (NMR) είναι μια μορφή φασματομετρίας απορρόφησης, όπου το δείγμα που βρίσκεται σε ισχυρό μαγνητικό πεδίο απορροφά ηλεκτρομαγνητική ακτινοβολία στην περιοχή των ραδιοκυμάτων (4 έως 900 MHz) και οι προσανατολισμένοι πυρήνες των ατόμων από την κατάσταση χαμηλότερης ενέργειας μεταβαίνουν στην κατάσταση υψηλότερης ενέργειας με αντιστροφή του spin. Όταν πραγματοποιηθεί αυτή η αναστροφή, οι πυρήνες έχουν συντονισθεί με την εφαρμοζόμενη ακτινοβολία, απ' όπου προκύπτει και ο όρος «συντονισμός».

Η συχνότητα απορρόφησης της προσπίπτουσας ακτινοβολίας που παρατηρείται από την αλληλοεπίδρασή της με τους πυρήνες του αναλύτη, είναι χαρακτηριστική για κάθε πυρήνα, με αποτέλεσμα η μέθοδος NMR να αποτελεί μια αποτελεσματική μέθοδο για την ταυτοποίηση της δομής οργανικών ενώσεων, για τον ποσοτικό προσδιορισμό τους και για το χημικό «περιβάλλον» των πυρήνων των ατόμων στα μόρια των χημικών ενώσεων.

Τα πρώτα πειράματα σε φασματοσκοπία NMR διεξήχθησαν από τους Bloch, του Πανεπιστημίου Stanford και Purcell, από το Harvard το 1946, οι οποίοι ανεξάρτητα απέδειξαν ότι η απορρόφηση της ηλεκτρομαγνητικής ακτινοβολίας από πυρήνες ατόμων που βρίσκονται σε ισχυρό μαγνητικό πεδίο έχει ως αποτέλεσμα το διαχωρισμό των ενεργειακών τους επιπέδων και τον προσανατολισμό των πυρήνων στις αντίστοιχες στάθμες, μετρώντας την ελάχιστη αυτή ενέργεια που απορροφάται κατά τη

μετάβαση των πυρήνων. Για την ανακάλυψή τους αυτή μοιράσθηκαν το 1952 το βραβείο Νόμπελ Φυσικής. Το 1953 παρουσιάσθηκε το πρώτο εμπορικό όργανο NMR για τη μελέτη χημικών δομών και η μέθοδος αυτή αναπτύχθηκε ταχύτατα με ποικίλες εφαρμογές κυρίως στην οργανική χημεία και βιοχημεία.

Η φασματοσκοπία NMR είναι συμπληρωματική των άλλων φασματοσκοπικών μεθόδων, όπως της φασματομετρίας μάζας και της φασματοσκοπίας υπέρυθρου για την ταυτοποίηση της δομής ακόμα και εξαιρετικά περίπλοκων μορίων³⁶. Επιπλέον συμβάλλει στη στερεοχημική ταυτοποίηση των οργανικών ενώσεων παρέχοντας τη συγκεκριμένη στερεοχημική διάταξη του ανθρακικού σκελετού με τα υδρογόνα σε ένα οργανικό μόριο, δεδομένου ότι οι πυρήνες που εξετάζονται ως επί το πλείστον στο NMR για την οργανική ανάλυση είναι το υδρογόνο (^1H) και ο άνθρακας (^{13}C).

Οι καινούριες εξελίξεις στη φασματομετρία NMR, όπως ο συνδυασμός με την τεχνική μετασχηματισμού Fourier (FT-NMR)³⁷ και η ανάπτυξη δισδιάστατων (2D)³⁸ και πολυδιάστατων (multidimensional) τεχνικών NMR, έδωσαν άλλες διαστάσεις στη χαρτογράφηση της δομής οργανικών ενώσεων.

1.4.2 Φασματομετρία μάζας (MS)

Η φασματομετρία μάζας, είναι ένα αναλυτικό εργαλείο απαραίτητο για τη διερεύνηση της δομής τόσο οργανικών όσο και ανόργανων ουσιών. Βασίζεται στη παραγωγή ιόντων κατά τον ιονισμό του μορίου και τον διαχωρισμό τους βάσει του λόγου m/z , όπου m η μάζα του ιόντος και z το φορτίο του ιόντος. Μπορεί να προσδιορίσει μοριακά βάρη ουσιών έως και για μεγάλα μόρια της τάξης των 10^6 Da. Μπορεί να χρησιμοποιηθεί για την ποιοτική ανάλυση αερίων, υγρών και υπό ειδικές συνθήκες στερεών ενώ συνηθισμένη εφαρμογή είναι η μελέτη της κατανομής ισοτόπων.

Η φασματομετρία ανακαλύφθηκε από τους Thomson (1910) και Aston (1919) και η πρώτη της εφαρμογή ήταν στην ανακάλυψη νέων ισοτόπων. Τα πρώτα εμπορικά όργανα εμφανίστηκαν γύρω στο 1940 και χρησιμοποιήθηκαν στη βιομηχανία πετρελαιοειδών. Η εφαρμογή της φασματομετρίας μάζας σε αναλύσεις σειράς οργανικών μορίων άρχισε κυρίως από το 1960 και μετά.

Αν και η αρχή λειτουργίας του είναι απλή, το φασματόμετρο μάζας είναι ένα πολύπλοκο όργανο με υψηλό κόστος. Λόγω όμως των δυνατοτήτων που έχει στην

εύρεση δομής των ενώσεων, της ταχύτητας και της αξιοπιστίας του, η χρήση του στα αναλυτικά εργαστήρια συνεχώς αυξάνεται.

Η φασματομετρία μάζας συνίσταται στη δημιουργία ιοντικών θραυσμάτων των ενώσεων, το διαχωρισμό τους και την καταγραφή τους ανάλογα με τη μάζα και το φορτίο που φέρουν. Η μελέτη των θραυσμάτων της ένωσης δίνει σπουδαιότερες πληροφορίες για το μοριακό της βάρος, το μοριακό τύπο και τη δομή της.

Τα φασματόμετρα μάζας αποτελούνται από τα εξής τμήματα: α) σύστημα εισαγωγής του δείγματος, β) το θάλαμο ιονισμού, γ) τον αναλυτή μαζών και δ) τον ανιχνευτή. Εκτός από τα τμήματα αυτά περιλαμβάνουν συστήματα υψηλού κενού, καταγραφικά και υπολογιστικά συστήματα για τον έλεγχο λειτουργίας και την επεξεργασία των αποτελεσμάτων.

Η δημιουργία των ιόντων επιτυγχάνεται με διάφορους τρόπους εκ των οποίων ο γνωστότερος είναι ο βομβαρδισμός με δέσμη ηλεκτρονίων μεγάλης ενέργειας EI (Electron Ionization). Κατά τη σύγκρουση των ηλεκτρονίων με μόριο της ένωσης αποβάλλεται αρχικά ένα ηλεκτρόνιο και προκύπτει μια κατιοντική ρίζα:



Το ιόν που προκύπτει έχει μάζα ίση με το μοριακό βάρος της ένωσης και ονομάζεται μοριακό ιόν. Εξ αιτίας της μεγάλης ενέργειας που απέκτησε από τη σύγκρουση, το μοριακό ιόν είναι ασταθές και διασπάται σε διάφορα θραύσματα τα οποία μπορεί να είναι ουδέτερα μόρια, ελεύθερες ρίζες, ανιόντα ή κατιόντα. Τα θυγατρικά ιόντα μπορεί επίσης να διασπαθούν και να δώσουν νέα θραύσματα.

Τα κατιοντικά θραύσματα που δημιουργούνται στο θάλαμο ιοντισμού επιταχύνονται από ένα ηλεκτρικό πεδίο δυναμικού V και στη συνέχεια κατευθύνονται προς τον αναλυτή, ενώ τα ουδέτερα μόρια εξέρχονται από το φασματογράφο από τις αντλίες κενού.

Η βασική λειτουργία του αναλυτή είναι να διαχωρίζει τα ιόντα ανάλογα με την τιμή του m/z του ιόντος (όπου m η μάζα και z το φορτίο του). Οι κυριότεροι αναλυτές μαζών είναι α) τετραπολικός αναλυτής (Quadrupole), β) παγίδα ιόντων (Quadrupole Ion-Trap), γ) χρόνου πτήσης (Time Of Flight, TOF), δ) Orbitrap, ε) αναλυτής κυκλοτρονικού συντονισμού ιόντων με μετασχηματισμό Fourier (Fourier-transform Ion Cyclotron Resonance, FT-ICR), ε) αναλυτής μαγνητικής εκτροπής (Ion magnetic resonance).

Η σύζευξη της φασματομετρίας μάζας με την υγροχρωματογραφία μπορεί να μας οδηγήσει σε μεγάλη ακρίβεια αποτελεσμάτων^{39,40} για την ανίχνευση μεταβολιτών σε πολύπλοκα βιολογικά δείγματα⁴¹. Ακόμα αξίζει να σημειωθεί πως η σύζευξη της

φασματομετρίας μάζας με την αεριοχρωματογραφία (gas chromatography, GC) είναι ιδανική για την ανίχνευση μικρής μοριακής μάζας μεταβολιτών (<650 Da).⁴²

ΚΕΦΑΛΑΙΟ 2

ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

2.1 Οργανολογία και αντιδραστήρια

Η ανάλυση NMR πραγματοποιήθηκε στο φασματόμετρο Avance III 600 (Bruker, Γερμανία), εφοδιασμένο με αυτόματο δειγματολήπτη 60 θέσεων (B-ACS 60) και ανιχνευτή 5 mm PABBI 1H/D-BB σε θερμοκρασία 27°C. Για τον έλεγχο και τη σταθεροποίηση της θερμοκρασίας χρησιμοποιήθηκε ψηφιακός ρυθμιστής BVT-3000 με ρυθμό ροής N₂ 535 L/h.

Για την υδροχρωματογραφία υπέρ-υψηλής απόδοσης χρησιμοποιήθηκε το σύστημα Accela (Thermo Scientific, Γερμανία) που αποτελείται από δυαδική αντλία, αυτόματο δειγματολήπτη, απαερωτή κενού σε σειρά και κλίβανο θερμοστάτησης στηλών. Για τον διαχωρισμό χρησιμοποιήθηκε στήλη αντίστροφης φάσης Hypersil GOLD™ C18 (2.1 x 100 mm, 1.9μm) (Thermo Fisher Scientific). Η φασματομετρία μάζας υψηλής ανάλυσης πραγματοποιήθηκε σε υβριδικό φασματόμετρο μάζας γραμμικής και τροχιακής παγίδας σε σειρά LTQ Orbitrap Discovery XL (Thermo Scientific, Γερμανία), πραγματοποιώντας ιοντισμό με ηλεκτροψεκασμό (electrospray ionization, ESI).

Οι διαλύτες που χρησιμοποιήθηκαν (ακετονιτρίλιο (ACN), μεθανόλη (MeOH), νερό καθώς μυρμηκικό οξύ και διχλωρομεθάνιο) ήταν βαθμού καθαρότητας LCMS, (Sigma-Aldrich, Ελβετία). Ως εσωτερικά πρότυπα για την ανάλυση MS χρησιμοποιήθηκαν η σοχμβίνη, η ρεζερπίνη και η 2-αμινοφαινόλη (Sigma-Aldrich, Steinheim, Γερμανία).

Η φυγοκέντρηση των δειγμάτων έγινε με μικροφυγόκεντρο Mikro 200R (Hettich Lab Technology, Γερμανία), ενώ η εξάτμισή τους έγινε με εξατμιστή GeneVac HT-4XEZ-2 Lyospeed ENABLED (Genevac Ltd, Ηνωμένο Βασίλειο).

2.2 Δείγματα

Χρησιμοποιήθηκαν κυτταρικές σειρές καρκίνου ουροδόχου κύστης σε τέσσερα στάδια. Για το πρώτο στάδιο χρησιμοποιήθηκε η σειρά RT4 (grade 1, A1), για το δεύτερο η σειρά RT112 (grade 2, A2), για το τρίτο η σειρά T24 (grade 3, A3), και για το τέταρτο η σειρά TCCSUP (grade 4, A4)⁴³⁻⁴⁵. Για την πραγματοποίηση των πειραμάτων

NMR και MS χρησιμοποιήθηκαν 10 δείγματα από την κάθε κυτταρική σειρά. Τα δείγματα παρελήφθησαν μέσα σε erpendorfs από το ερευνητικό εργαστήριο του επίκουρου καθηγητή κ.Στραβοπόδη.

2.3 Επιλογή πρωτοκόλλου εκχύλισης

Συνολικά έγιναν τέσσερις δοκιμές για την επιλογή του κατάλληλου πρωτοκόλλου εκχύλισης:

Πείραμα #1 (Πρωτόκολλο Wu)

Σύμφωνα με το πρωτόκολλο αυτό⁴⁶ ακολουθήθηκε η εξής διαδικασία:

- Προσθήκη 200 μL μεθανόλης και 200 μL χλωροφορμίου
- Υπέρηχοι για 5 min (με πάγο στο υδατόλουτρο)
- Προσθήκη 180 μL απιονισμένου νερού (milli Q)
- Vortex για 1 min
- Αναμονή στον πάγο για 15 min
- Φυγοκέντρηση στα 16000 rpm για 20 min στους 4°C

Στη συνέχεια έγινε συλλογή της ανώτερης (υδατικής) φάσης και της κατώτερης (λιπιδικής) φάσης και διατήρηση στους 4°C μέχρι να επαναληφθεί η διαδικασία. Μετά την επανάληψη της διαδικασίας συλλέχθηκε η τελική υδατική (ένωση των δύο υδατικών φάσεων από τις δύο εκχυλίσσεις) και λιπιδική φάση (ένωση των δύο λιπιδικών φάσεων από τις δύο εκχυλίσσεις). Τέλος, οι υδατικές φάσεις μεταφέρθηκαν στο GeneVac όπου έγινε η εξάτμιση των διαλυτών overnight. Οι διαλύτες χρησιμοποιήθηκαν σε θερμοκρασία 4°C.

Πείραμα #2 (Πρωτόκολλο Keun/Athrsuch)

Σύμφωνα με το πρωτόκολλο αυτό⁴⁷ ακολουθήθηκε η εξής διαδικασία:

- Προσθήκη 200 μL χλωροφορμίου και 100 μL μεθανόλης
- Προσθήκη 300μL απιονισμένου νερού (milli Q)
- Vortex για 1 min και διατήρηση στον πάγο
- Φυγοκέντρηση στα 13500 rpm για 10min στους 4°C

Έπειτα έγινε συλλογή της ανώτερης (υδατικής) φάσης και της κατώτερης (λιπιδικής) φάσης και διατήρηση στους 4°C μέχρι να επαναληφθεί η διαδικασία. Μετά την επανάληψη της διαδικασίας συλλέχθηκε η τελική υδατική (ένωση των δύο υδατικών φάσεων από τις δύο εκχυλίσσεις) και λιπιδική φάση (ένωση των δύο λιπιδικών φάσεων

από τις δύο εκχυλίσεις). Τέλος, οι υδατικές φάσεις μεταφέρθηκαν στο GeneVac όπου έγινε η εξάτμισή των διαλυτών overnight. Οι διαλύτες χρησιμοποιήθηκαν σε θερμοκρασία 4°C.

Πείραμα #3 (Πρωτόκολλο Le Belle)

Σύμφωνα με το πρωτόκολλο αυτό⁴⁸ ακολουθήθηκε η εξής διαδικασία:

- Προσθήκη 168 μL μεθανόλης και 84 μL χλωροφορμίου (2:1)
- Υπέρηχοι για 5 min (με πάγο στο υδατόλουτρο)
- Προσθήκη 125 μL χλωροφορμίου και 125 μL απιονισμένου νερού (milli Q)
- Vortex για 1 min (δεν το αναφέρει στο πρωτόκολλο)
- Φυγοκέντρηση στα 13000 rpm για 20min στους 4°C

Έπειτα έγινε συλλογή της ανώτερης (υδατικής) φάσης και της κατώτερης (λιπιδικής) φάσης και διατήρηση στους 4°C μέχρι να επαναληφθεί η διαδικασία. Μετά την επανάληψη της διαδικασίας συλλέχθηκε η τελική υδατική (ένωση των δύο υδατικών φάσεων από τις δύο εκχυλίσεις) και λιπιδική φάση (ένωση των δύο λιπιδικών φάσεων από τις δύο εκχυλίσεις). Τέλος, οι υδατικές φάσεις μεταφέρθηκαν στο GeneVac όπου έγινε η εξάτμισή των διαλυτών overnight. Οι διαλύτες χρησιμοποιήθηκαν σε θερμοκρασία 4°C.

Πείραμα #4 (Πρωτόκολλο παραλλαγή πειράματος 1)

Σύμφωνα με την παραλλαγή του πειράματος 1, είχαμε τα εξής:

- Προσθήκη 200 μL μεθανόλης και 200 μL απιονισμένου νερού (milli Q)
- Φυγοκέντρηση στα 13500 rpm για 15 min στους 4°C
- Συλλογή υπερκείμενης φάσης και εκχύλιση του εναπομείναντος ιζήματος με επανάληψη της διαδικασίας
- Ένωση των δύο υπερκείμενων φάσεων που έχουν προκύψει
- Προσθήκη 200 μL χλωροφορμίου στην υπερκείμενη φάση
- Vortex για 30 sec
- Φυγοκέντρηση στα 13500 rpm για 3 min στους 4°C
- Συλλογή της νέας υπερκείμενης φάσης και προσθήκη σε αυτήν 200 μL χλωροφορμίου
- Vortex για 30 sec
- Φυγοκέντρηση στα 13500 rpm για 3 min στους 4°C

Έπειτα έγινε συλλογή της ανώτερης (υδατικής) φάσης και της κατώτερης (λιπιδικής) φάσης και διατήρηση στους 4°C μέχρι να επαναληφθεί η διαδικασία. Μετά την

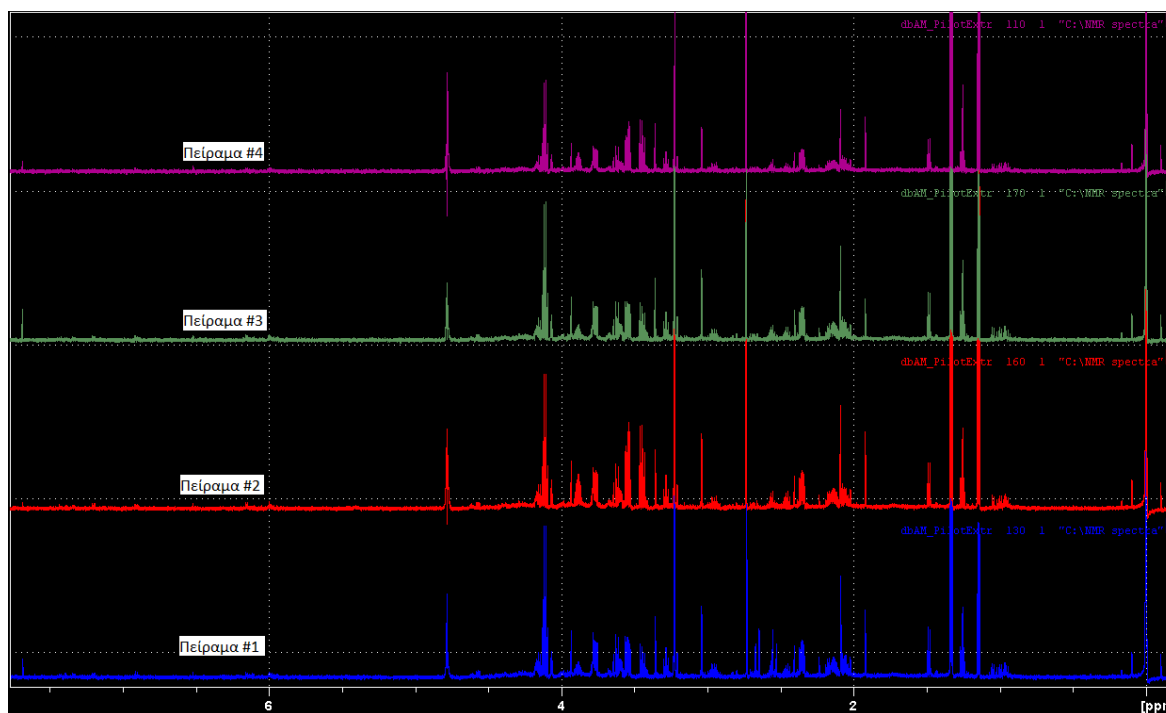
επανάληψη της διαδικασίας συλλέχθηκε η τελική υδατική (ένωση των δύο υδατικών φάσεων από τις δύο εκχυλίσεις) και λιπιδική φάση (ένωση των δύο λιπιδικών φάσεων από τις δύο εκχυλίσεις). Τέλος, οι υδατικές φάσεις μεταφέρθηκαν στο GeneVac όπου έγινε η εξάτμισή των διαλυτών overnight. Οι διαλύτες χρησιμοποιήθηκαν σε θερμοκρασία 4°C.

Όλα τα δείγματα που προετοιμάστηκαν για την επιλογή του καταλληλότερου πρωτοκόλλου εκχύλισης κλείστηκαν με parafilm και αποθηκεύτηκαν στους -80°C. Επίσης να σημειωθεί ότι για τα πιλοτικά αυτά πειράματα χρησιμοποιήθηκαν δείγματα από την κυτταρική σειρά T24 (grade 3, A3).

2.4 Ανάλυση με NMR

Αφού τα εξατμισμένα υδατικά εκχυλίσματα ξεπάγωσαν σε θερμοκρασία δωματίου η ανασύστασή τους πραγματοποιήθηκε με διάλυμα D₂O-Buffer σε αναλογία 90:10. Έγινε προσθήκη 700 μL από το διάλυμα αυτό σε κάθε erpendorf και φυγοκέντρωση στα 20000 rpm για 5 min. Τέλος μεταφέρθηκαν 550 μL από το erpendorf σε σωληνάριο NMR διαμέτρου 5 mm. Η σύσταση του Buffer (ρυθμιστικού διαλύματος) ήταν η εξής: 1.5 M KH₂PO₄/D₂O, 2 mM NaN₃ ως αντιμικροβιακό παράγοντα και 0.1% 1,1,2,2-tetradeutero-3-trimethylsilylpropionic acid, TSP, ως εσωτερικό πρότυπο. Το διάλυμα αυτό είχε pH=7.4.

Παρακάτω φαίνονται τα φάσματα ¹H NMR από τα πειράματα για την επιλογή του καταλληλότερου πρωτοκόλλου εκχύλισης:



Σχήμα 3. Σύγκριση των τεσσάρων μεθόδων εκχύλισης, μετά από ανάλυση με NMR

Επιλέχθηκε το πείραμα #1 (πρωτόκολλο Wu) καθώς με αυτό φαίνεται να έχουμε περισσότερες και σε πολλά σημεία πιο πυκνές κορυφές.

Τα φάσματα NMR ελήφθησαν στο φασματοσκόπιο Bruker Avance III 600 MHz σε θερμοκρασία 300 K (27°C). Η προετοιμασία για τη λήψη του πειράματος βελτιστοποιήθηκε κατά τα πιλοτικά πειράματα για την επιλογή πρωτοκόλλου εκχύλισης και η διαδικασία πραγματοποιήθηκε ομοίως για όλες τις υπόλοιπες μετρήσεις, μέσω του προγράμματος ICON NMR suite.

Τα φάσματα 1D ^1H NMR ελήφθησαν με την εφαρμογή παλμικής αλληλουχίας 1DNOESY (noesygrrr1D), με κορεσμό της κορυφής του νερού (water presaturation) (25 Hz) κατά τη διάρκεια της αποδιέγερσης και του χρόνου μίξης με στόχο την εξάλειψη της κορυφής συντονισμού του νερού. Στο φάσμα 1DNOESY ελήφθησαν 256 σαρώσεις (NS=256), με ανάλυση 64K σημείων (TD=65536) και για φασματικό εύρος SW=12335.526Hz, με χρόνο αποδιέγερσης (relaxation delay) D1=8.0 s, χρόνο λήψης του φάσματος (acquisition time) AQ=2.66s και χρόνο μίξης (mixing time) D8=0.01s.

Το ομοπυρηνικό 2D πείραμα J-resolved (Jres) έχει μικρή χρονική διάρκεια και έτσι ήταν εφικτή η λήψη του για κάθε δείγμα. Για το γρήγορο αυτό πείραμα χρησιμοποιήθηκε η παλμική αλληλουχία jresgrrrqr. Τα φάσματα J-resolved ελήφθησαν με 4 σαρώσεις (NS=4) και ανάλυση 8K σημείων (TD=8192 για την F2, TD=40 για την F1)

σε φασματικό εύρος 10000 Hz (SWH=10000.000 για την F2, SWH=78.125 για την F1), με χρόνο αποδιέγερσης $D1=1$ s και χρόνο λήψης (acquisition time) 0.4 s ($AQ=0.4096000$ για την F2, $AQ=0.2560000$ για την F1). Το φάσμα TOCSY διεξήχθη με 128 σαρώσεις ($NS=128$), με ανάλυση 2K σημείων ($TD=2048$ για την F2, $TD=200$ για την F1) και φασματικό εύρος 9615 Hz ($SWH=9615.385$ τόσο για την F2 αλλά και για την F1), με χρόνο αποδιέγερσης $D1=2$ s και χρόνο λήψης $AQ=0.1$ s ($AQ=0.1064960$ για την F2, $AQ=0.0104000$ για την F1). Το 1H - ^{13}C DEPT 135-HSQC πείραμα πραγματοποιήθηκε με 234 σαρώσεις ($NS=234$) και φασματικό εύρος $SWH=9615$ Hz για την F2 και $SWH=27165$ Hz για την F1, με χρόνο λήψης του φάσματος 0.1 s ($AQ=0.1064960$ για την F2, $AQ=0.0033132$ για την F1).

2.5 Ανάλυση με φασματομετρία μάζας

Μετά την εκχύλιση των ιστών και πριν από την ανάλυση με υγροχρωματογραφία συζευγμένη με την φασματομετρία μάζας, σε κάθε δείγμα προστέθηκαν τα εσωτερικά πρότυπα, ώστε να κανονικοποιηθούν τα αποκτηθέντα δεδομένα και έτσι, να ληφθούν πιο ακριβή αποτελέσματα. Ως κατάλληλα εσωτερικά πρότυπα επιλέχθηκαν η 2-αμινοφαινόλη (M.B.: 109.0522), η υοχιμβίνη (M.B.: 354.1938) και η ρεζερπίνη (M.B.: 608.2728), επειδή δεν περιλαμβάνονται στους γνωστούς μεταβολίτες που υπάρχουν σε καρκινικά κύτταρα ουροδόχου κύστης, είναι εύκολα διαθέσιμες ουσίες, εκκλύονται σε διαφορετικούς και προσεγγιστικά ισαπέχοντες χρόνους ανάλυσης και τα μοριακά βάρη τους καλύπτουν όλο το εύρος του φάσματος μάζας 100-1000 Da.

Τα ξηρά εκχυλίσματα εξατμίστηκαν μέχρι ξηρού δύο φορές. Στη συνέχεια έγινε η επανασύστασή τους με 100 μ L διαλύματος που περιείχε απιονισμένο νερό/ακετονιτρίλιο σε αναλογία 95/5. Ως τελικό διάλυμα ανασύστασης χρησιμοποιήθηκε ίδιου όγκου και αναλογίας διαλυτών διάλυμα συγκέντρωσης εσωτερικών προτύπων 10 μ g/mL. Έπειτα, τα δείγματα περιδινήθηκαν για 2 min, φυγοκεντρήθηκαν στα 13000 rpm για 3 min και στη συνέχεια, μεταφέρθηκαν σε ένθετα φιαλίδια (inserts) των 200 μ L μέσα σε φιαλίδια αυτόματου δειγματολήπτη (vials). Από κάθε δείγμα λήφθηκαν 16 μ L και αναμίχθηκαν πλήρως, δημιουργώντας ένα συνενωμένο δείγμα (pooled sample), το οποίο χρησιμοποιήθηκε ως δείγμα ελέγχου ποιότητας (quality control, QC). Για τον έλεγχο απόδοσης του οργάνου, τρία δείγματα ελέγχου ποιότητας εγχύθηκαν στην αρχή και στο τέλος της ανάλυσης καθώς και ένα μετά από κάθε πέντε εγχύσεις δείγματος.

Για τον χρωματογραφικό διαχωρισμό των δειγμάτων, η κινητή φάση αποτελούνταν από τους διαλύτες **A**: νερό- 0.1% (v/v) μυρμηκικό οξύ, και **B**: ακετονιτρίλιο. Το

πρόγραμμα βαθμιδωτής έκλουσης, για τον θετικό ιοντισμό φαίνεται στον παρακάτω πίνακα (Πίνακας 1):

Πίνακας 1. Πρόγραμμα βαθμιδωτής έκλουσης (gradient) κινητής φάσης που χρησιμοποιήθηκε κατά τον θετικό ιοντισμό

| Χρόνος (min) | Ταχύτητα ροής (mL/min) | Κινητή φάση A (%) | Κινητή φάση B (%) | Curve |
|--------------|------------------------|-------------------|-------------------|-------|
| 0.0 | 0.4 | 100 (98) | 0 (2) | 5 |
| 1.0 | 0.4 | 100 (98) | 0 (2) | 5 |
| 16.0 | 0.4 | 0 | 100 | 5 |
| 20.0 | 0.4 | 0 | 100 | 5 |
| 22.0 | 0.4 | 100 | 0 | 1 |

Το πρόγραμμα βαθμιδωτής έκλουσης, για τον αρνητικό ιοντισμό φαίνεται στον παρακάτω πίνακα (Πίνακας 2):

Πίνακας 2. Πρόγραμμα βαθμιδωτής έκλουσης (gradient) κινητής φάσης που χρησιμοποιήθηκε κατά τον αρνητικό ιοντισμό

| Χρόνος (min) | Ταχύτητα ροής (mL/min) | Κινητή φάση A (%) | Κινητή φάση B (%) | Curve |
|--------------|------------------------|-------------------|-------------------|-------|
| 0.0 | 0.36 | 100 | 0 | 4 |
| 2.0 | 0.36 | 100 | 0 | 4 |
| 17.0 | 0.36 | 0 | 100 | 4 |
| 22.0 | 0.36 | 0 | 100 | 4 |
| 24.0 | 0.36 | 100 | 0 | 1 |

Η θερμοκρασία της στήλης διατηρήθηκε στους 40°C σε όλη τη διάρκεια του πειράματος, ενώ η θερμοκρασία του αυτόματου δειγματολήπτη ορίστηκε στους 8°C. Ο όγκος της έγχυσης ήταν 5 µL.

Χρησιμοποιήθηκε ο ιοντισμός με ηλεκτροψεκασμό (electrospray) και για την παγίδευση των ιόντων έγινε χρήση της τροχιακής παγίδας ιόντων orbitrap (FTMS) για την πλήρη σάρωση των ιόντων. Για να ληφθούν δεδομένα δίδυμης φασματομετρίας μάζας χρησιμοποιήθηκε η γραμμική παγίδα ιόντων σε λειτουργία εξαρτώμενης από τα δεδομένα θραυσματοποίησης (data dependent scanning). Το εύρος σάρωσης μάζας ήταν 100 έως 1000 Da στον θετικό και αρνητικό ιοντισμό. Οι συνθήκες λειτουργίας του φασματομέτρου μάζας υψηλής ανάλυσης, όπως ορίστηκαν φαίνονται στον πίνακα 3. Η ανάλυση πραγματοποιήθηκε με τη χρήση της φασματομετρίας μάζας υψηλής ανάλυσης με μετασχηματισμό Fourier (Fourier transform mass spectrometry, FTMS) πλήρους σάρωσης ιόντος, σε διακριτική ικανότητα 30000, ενώ η καταγραφή του φάσματος μάζας πραγματοποιήθηκε σε κάθε περίπτωση στην κεντροειδή μορφή.

Πίνακας 3. Συνθήκες λειτουργίας του φασματομέτρου μάζας Orbitrap στον θετικό και αρνητικό ιοντισμό

| Παράμετροι | Συνθήκες λειτουργίας φασματομέτρου μάζας | |
|-------------------------------|--|------------|
| | (+) ESI | (-) ESI |
| Θερμοκρασία τριχοειδούς | 356° C | 356° C |
| Τάση τριχοειδούς | 20 V | -60 V |
| Σωληνοειδής φακός (tube lens) | -49 V | 110V |
| Τάση πηγής | 3.10 kV | 3.50 kV |
| Ροή αερίου εκνέφωσης | 30 μονάδες | 30 μονάδες |
| Ροή βοηθητικού αερίου | 10 μονάδες | 10 μονάδες |

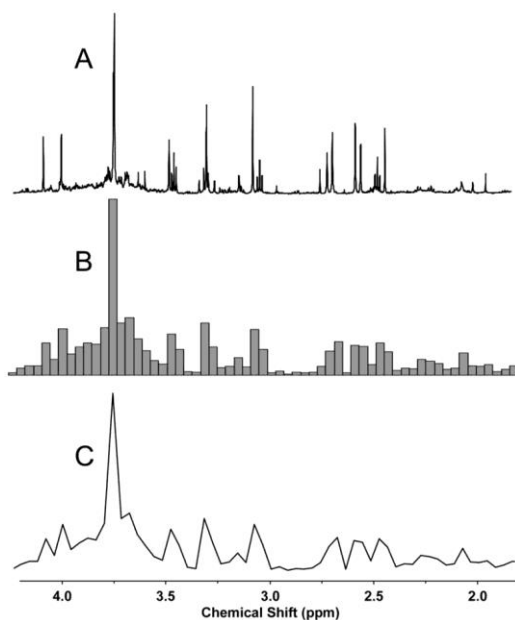
ΚΕΦΑΛΑΙΟ 3

ΕΠΕΞΕΡΓΑΣΙΑ ΦΑΣΜΑΤΩΝ

3.1 Επεξεργασία φασμάτων NMR

Για την ταυτοποίηση των μεταβολιτών στα φάσματα NMR των υδατικών εκχυλισμάτων των δειγμάτων χρησιμοποιήθηκαν συνδυαστικά ειδικά λογισμικά προγράμματα όπως το Chenomx NMR Suite αλλά και το TopSpin (v. 3.5, Bruker BioSpin GmbH). Επίσης, σημαντικό εργαλείο ήταν και οι διαδικτυακές βάσεις δεδομένων όπως η The Human Metabolome Database, HMDB⁴⁹.

Το πρώτο βήμα της στατιστικής επεξεργασίας των δεδομένων είναι η προετοιμασία των φασμάτων NMR, ώστε να ακολουθήσει η στατιστική τους ανάλυση. Αυτό έγινε μέσω του ειδικού λογισμικού AMIX με το οποίο πραγματοποιήθηκε η αναγωγή (data reduction ή bucketing) των δεδομένων σε ισαπέχοντα τμήματα (buckets). Στα υδατικά εκχυλίσματα το bucketing έγινε σε όλη την έκταση του φάσματος (0.70 – 9.40 ppm) χωρίζοντας το σε ισαπέχοντα τμήματα εύρους 0.01 ppm και 0.005 ppm. Από τα φάσματα αφαιρέθηκαν οι περιοχές του υπολειμματικού νερού (4.68 – 5.2 ppm).



Σχήμα 4. Παράδειγμα χρήσης bucketing σε φάσμα ^1H NMR ενός δείγματος ούρων, με ισαπέχοντα τμήματα εύρους 0,04 ppm. Σε αυτό το παράδειγμα, το bucketing μειώνει τον αριθμό των μεταβλητών δεδομένων από 65.536 σε 312, διευκολύνοντας την πολυπαραγοντική ανάλυση με PCA και PLS-DA.

Σε κάθε bucket έγινε ολοκλήρωση των κορυφών που περιέχονται στο φάσμα για κάθε δείγμα. Η κανονικοποίηση των δεδομένων έγινε με την επιλογή «χωρίς κλιμάκωση» (no scaling) με τη βοήθεια του AMIX και ακολούθησε κανονικοποίηση των φασμάτων με το εσωτερικό πρότυπο TSP που περιείχαν τα δείγματα. Επίσης, η κανονικοποίηση των δεδομένων έγινε και με την επιλογή «συνολικής έντασης» (total intensity) καθώς και με την επιλογή «χωρίς κλιμάκωση» (no scaling) με τη βοήθεια του AMIX. Από τα αποτελέσματα των τριών διαδικασιών κρίθηκαν πιο αξιόπιστα αυτά που προέκυψαν από την κανονικοποίηση των δεδομένων με την επιλογή «χωρίς κλιμάκωση» (no scaling) με τη βοήθεια του AMIX και ακόλουθη περαιτέρω κανονικοποίηση των φασμάτων με το εσωτερικό πρότυπο TSP.

Με τον τρόπο αυτό δημιουργήθηκε, ένας πίνακας δεδομένων μέσω του excel και ο οποίος χρησιμοποιήθηκε στην στατιστική ανάλυση. Επίσης, στα φάσματα πραγματοποιήθηκε βαθμονόμηση του φάσματος βάση εσωτερικού προτύπου στα σημεία όπου δεν είχε γίνει σωστά από τον αυτοματισμό έτσι ώστε να επιτευχθούν πιο αξιόπιστα και ερμηνεύσιμα μοντέλα.

Στην απόδοση των κορυφών των NMR φασμάτων αποτέλεσε αρωγός και η χρήση μιας ακόμη δισδιάστατης τεχνικής, που παρουσιάζει την συμμεταβολή των σημάτων. Η τεχνική αυτή, με το όνομα STOCSY^{18,50,51}(statistical total correlation spectroscopy) βασίζεται στην ανάλυση των εντάσεων κορυφών σε έναν μεγάλο αριθμό φασμάτων. Έτσι αποτελεί μέθοδο που προορίζεται για πολλά δείγματα, ενώ η αντίστοιχη, καθαρά φασματοσκοπική τεχνική TOCSY αποτελεί μέθοδο που προορίζεται για ανάλυση ενός και μοναδικού δείγματος. Στην παρούσα εργασία λοιπόν επιστρατεύεται και η χρήση της μονοδιάστατης εφαρμογής STOCSY, που υπολογίζει την συσχέτιση μεταξύ ενός επιλεγμένου σημείου δεδομένων και όλων των υπόλοιπων σημείων δεδομένων στα φάσματα. Με αυτόν τον τρόπο δημιουργείται ένα διάγραμμα δύο διαστάσεων με χημικές μετατοπίσεις στον άξονα των X και συντελεστή συσχέτισης (συνδιακύμανση) στον άξονα των Y, σύμφωνα με την τιμή του οποίου χρωματίζεται το διάγραμμα. Οι υψηλότερες συσχετίσεις που παρατηρούνται, αντιστοιχούν σε σήματα που οφείλονται στον ίδιο μεταβολίτη, ενώ οι λίγο χαμηλότερες αλλά εξίσου υψηλές αντιστοιχούν σε μεταβολίτες, η συγκέντρωση των οποίων μεταβάλλεται με αυτήν του υπό εξέταση μεταβολίτη. Παράλληλα, παρατηρούνται και πολύ χαμηλές συσχετίσεις, οι οποίες οφείλονται σε μεταβολίτες των οποίων η μεταβολή τους δεν συσχετίζεται με τον υπό έλεγχο μεταβολίτη. Τέλος, με αρνητικό πρόσημο εμφανίζονται οι κορυφές που μεταβάλλονται αντιστρόφως ανάλογα με τις υπό εξέταση κορυφές. (Παράρτημα 1)

3.2 Επεξεργασία φασμάτων MS

Η προ-επεξεργασία των δεδομένων από τη φασματομετρία μάζας πραγματοποιήθηκε χρησιμοποιώντας το πρόγραμμα Proteowizard (<http://proteowizard.sourceforge.net/>), τη στατιστική γλώσσα προγραμματισμού R (<https://www.r-project.org/>) και το λογισμικό Microsoft Excel. Αρχικά, τα αρχεία του φασματομέτρου μάζας της μορφής (*raw), μετατράπηκαν στην μορφή mzXML (*mzXML) χρησιμοποιώντας το πρόγραμμα Proteowizard^{52,53}. Στη συνέχεια, χρησιμοποιήθηκε το πακέτο λογισμικού IPO⁵⁴ προκειμένου να βελτιστοποιηθούν οι παράμετροι του προγράμματος επεξεργασίας δεδομένων μάζας XCMS⁵⁵, το οποίο χρησιμοποιείται, στη συνέχεια, για την ανίχνευση των κορυφών, το φιλτράρισμα του θορύβου, τη διόρθωση του χρόνου κατακράτησης και την ευθυγράμμιση των κορυφών. Στο τέλος της διαδικασίας, η παραγόμενη λίστα των κορυφών μαζί με τους χρόνους κατακράτησης και τις αντίστοιχες εντάσεις για όλες τις κορυφές εξήχθη ως αρχείο .csv και ακολούθησε η διαδικασία κανονικοποίησης των δεδομένων (normalization) μέσω του NOREVA (<http://server.idrb.cqu.edu.cn/noreva/>)⁵⁶. Ουσιαστικά ο NOREVA είναι ικανός για την αναγνώριση της κατάλληλης μεθόδου κανονικοποίησης λαμβάνοντας υπόψιν πέντε κριτήρια που είναι τα εξής: α) ικανότητα ελάττωσης της διακύμανσης μεταξύ των δειγμάτων της ίδιας ομάδας⁵⁷, β) επίδραση της μεθόδου για διαφορετικές μεταβολομικές μελέτες⁵⁸ γ) η επαναληψιμότητα της μεθόδου για τον προσδιοριζόμενο μεταβολίτη μεταξύ διαφορετικών δεδομένων⁵⁹, δ) επίδραση της κλιμάκωσης στην ακρίβεια της ομαδοποίησης των δεδομένων⁶⁰⁻⁶², ε) την συσχέτιση που υπάρχει ανάμεσα στα κανονικοποιημένα δεδομένα και στα αρχικά δεδομένα⁵⁸.

Έπειτα, ο τελικός πίνακας εισήχθη στο Microsoft Excel και υπέστη κατάλληλη επεξεργασία με τη χρήση των εντολών «concatenate», «round» και «transpose».

Παράλληλα με την πολυμεταβλητή στατιστική ανάλυση, μελετήθηκε ένας νέος αλγόριθμος για την εξαγωγή χαρακτηριστικών από πολύπλοκα και ογκώδη δεδομένα που δουλεύει μέσω της R, ο αλγόριθμος KODAMA^{63,64}. Ο αλγόριθμος KODAMA (Knowledge Discovery by Accuracy Maximization) αποτελεί μία μη επιβλεπόμενη μέθοδο που έχει σχεδιαστεί ειδικά για την ανάλυση δεδομένων που εμπεριέχουν θόρυβο και είναι ογκώδη. Η βασική ιδέα του αλγορίθμου είναι η χρήση μιας διαδικασίας (που γίνεται κατά επανάληψη) για να προκύψει μία «υποθετική» ομαδοποίηση. Κάθε δείγμα ανατίθεται σε μία ομάδα και έτσι προκύπτουν τόσα δείγματα όσες και οι ομάδες. Ο αλγόριθμος αποφασίζει για τον βέλτιστο αριθμό των ομάδων χρησιμοποιώντας επαναλαμβανόμενη διαδικασία διασταυρούμενης επικύρωσης σε συνδυασμό με κάποια μέθοδο προβολής όπως για παράδειγμα PLS-DA.

Χρησιμοποιείται η ακρίβεια αντιπροσώπευσης που καθορίζεται από διασταυρούμενη επικύρωση. Τέλος, ανακαλύπτεται ο βέλτιστος αριθμός των ομάδων.

ΚΕΦΑΛΑΙΟ 4

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Μετά την προεπεξεργασία των φασματοσκοπικών χρωματογραφικών δεδομένων ακολουθεί η στατιστική ανάλυση. Για να αναλυθούν τέτοια δεδομένα χρησιμοποιούνται στατιστικές μέθοδοι οι οποίες ανήκουν στην κατηγορία της πολυπαραμετρικής ανάλυσης. Η πολυπαραμετρική ανάλυση δεδομένων (MVDA, MultiVariate Data Analysis) αφορά αλγορίθμους ανάλυσης δεδομένων που χρησιμοποιούν περισσότερες από μια μεταβλητές και πιο συγκεκριμένα χρησιμοποιείται για να περιγράψει τη σχέση μεταξύ ομάδων^{65,66}.

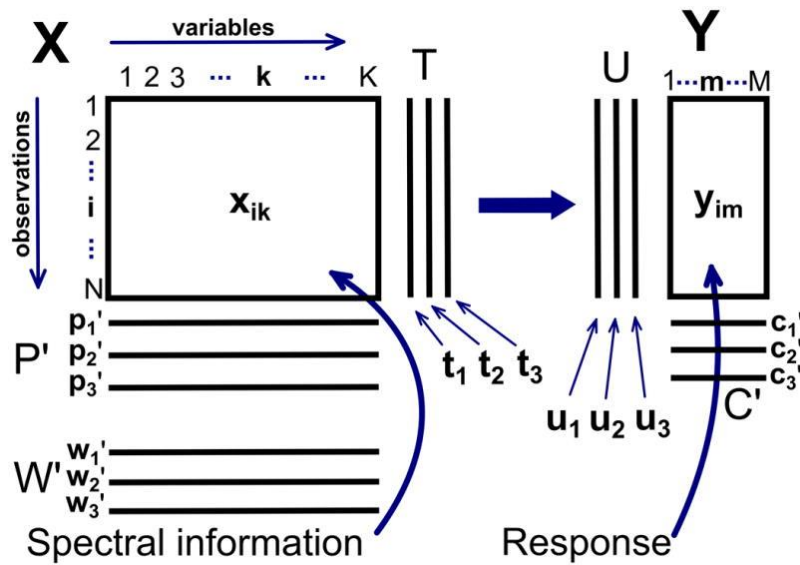
Η MVDA διακρίνεται σε δύο μεγάλες κατηγορίες στις επιβλεπόμενες και στις μη επιβλεπόμενες τεχνικές. Για τις επιβλεπόμενες τεχνικές απαιτείται η πρότερη γνώση των κατηγοριών στις οποίες ανήκουν τα δείγματα ενώ στις μη επιβλεπόμενες, τα δεδομένα αναλύονται χωρίς να υπάρχει τέτοιου είδους πληροφορία για τα δείγματα. Οι πλέον χρησιμοποιούμενες μέθοδοι όσον αφορά τη MVDA είναι η Ανάλυση Κύριων Συνιστωσών (PCA, Principal Component Analysis) και η Διακρίνουσα Ανάλυση Μερικών Ελαχίστων Τετραγώνων (PLS-DA, Partial Least Squares projection to latent structures - Discriminant Analysis). Παραλλαγή της τελευταίας είναι η Ορθογώνια Διακρινούσα Ανάλυση Μερικών Ελαχίστων Τετραγώνων^{67,68} (OPLS-DA, Orthogonal Projection to Latent Structures-Discriminant Analysis).

4.1 Ανάλυση κύριων συνιστωσών (PCA)

Η PCA είναι μη επιβλεπόμενη τεχνική που χρησιμοποιεί ορθογώνιο μετασχηματισμό για να μετατρέψει ένα σύνολο παρατηρήσεων πιθανώς συσχετισμένων μεταβλητών, σε ένα σύνολο τιμών γραμμικά μη συσχετισμένων μεταβλητών που ονομάζονται κύριες συνιστώσες⁶⁹. Ο αριθμός των κύριων συνιστωσών είναι μικρότερος είτε από τον αριθμό των αρχικών μεταβλητών ή τον αριθμό των παρατηρήσεων. Η πρώτη κύρια συνιστώσα παρουσιάζει τη μεγαλύτερη δυνατή διακύμανση (δηλαδή, αντιπροσωπεύει όσο το δυνατόν μεγαλύτερη μεταβλητότητα στα δεδομένα). Κάθε επακόλουθη κύρια συνιστώσα έχει τη μεγαλύτερη υπολειπόμενη δυνατή διακύμανση υπό τον περιορισμό ότι είναι ορθογώνια σε σχέση με τις προηγούμενες κύριες συνιστώσες. Το τελικό αποτέλεσμα είναι μη συσχετισμένα διανύσματα κύριων συνιστωσών που αποτελούν ένα ορθογώνιο σύνολο βάσης. Πρόκειται ουσιαστικά για ένα μαθηματικό αλγόριθμο

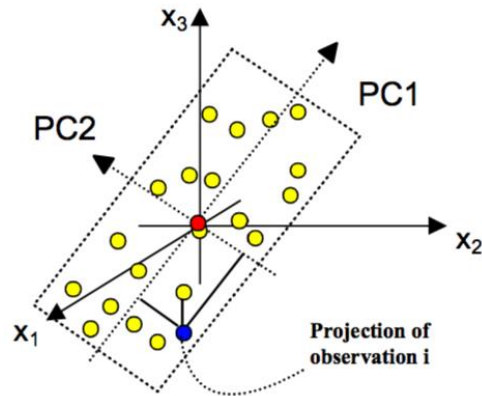
που μειώνει τις διαστάσεις των δεδομένων και παράλληλα διατηρεί το μεγαλύτερο μέρος της πληροφορίας.

Η PCA αποτελεί την βάση της πολυπαραμετρικής ανάλυσης δεδομένων. Πρόκειται για ένα χώρο όπου τα δεδομένα αποτελούνται από ένα πίνακα με N σειρές (παρατηρήσεις) και με K στήλες (μεταβλητές). Οι παρατηρήσεις μπορεί να είναι δείγματα προς ανάλυση, χημικά μείγματα αλλά και βιολογικά δείγματα. Για να χαρακτηριστούν οι ιδιότητες των παρατηρήσεων θα πρέπει να μετρηθούν οι μεταβλητές. Οι μεταβλητές μπορεί να προέρχονται από κάποιο φάσμα NMR, IR, UV αλλά και από χρωματογραφήματα (HPLC, GC).



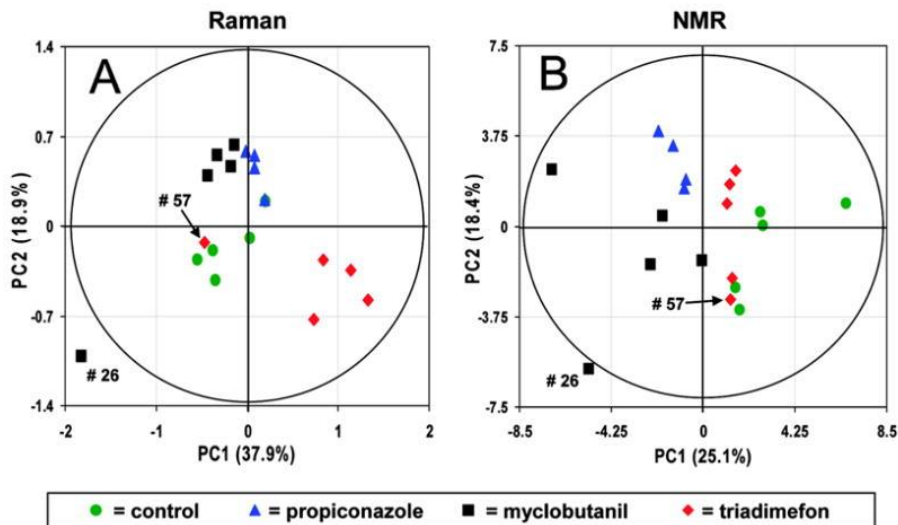
Σχήμα 5. Σε μία μεταβολομική μελέτη ο πίνακας δεδομένων θα περιέχει φασματικά δεδομένα στις σειρές έτσι ώστε η κάθε στήλη να αναπαριστά μία συγκεκριμένη συχνότητα ή bucket

Ουσιαστικά με τη μεθοδολογία PCA βρίσκονται οι κύριες συνιστώσες που είναι οι γραμμικοί συνδυασμοί των αρχικών δεδομένων που περιγράφουν κάθε δείγμα. Ωστόσο, οι κύριες συνιστώσες επιλέγονται έτσι ώστε η πρώτη κύρια συνιστώσα (PC1), να αντιπροσωπεύει το μεγαλύτερο μέρος της διακύμανσης (variation) των δεδομένων, η δεύτερη την αμέσως επόμενη μεγαλύτερη διακύμανση και ούτω καθεξής.



Σχήμα 6: Απεικονίζονται οι κύριες συνιστώσες. Η πρώτη κύρια συνιστώσα η οποία ταυτίζεται με τον άξονα της μεγαλύτερης μεταβολής των δεδομένων (PC1). Γ:Το σύστημα αξόνων των δύο πρώτων κύριων συνιστωσών στον αρχικό χώρο περιγραφής του συστήματος. Οι δύο κύριες συνιστώσες είναι ορθογώνιες μεταξύ τους.

Η οπτικοποίηση των αποτελεσμάτων πραγματοποιείται με τα διαγράμματα αντικειμένων (scores plot) και φορτίων (loadings plot). Τα αντικείμενα αποτελούν τις συντεταγμένες των δειγμάτων και είναι πλέον νέες μεταβλητές. Στο διάγραμμα αυτό κάθε σημείο αντιπροσωπεύει και ένα δείγμα και μπορούν να εξαχθούν συμπεράσματα για πιθανή ομαδοποίηση /διαφοροποίηση μεταξύ των δειγμάτων. Επίσης, μπορεί να διαπιστωθεί και η ύπαρξη έκτροπων τιμών από τα δεδομένα (outliers) σύμφωνα με το κριτήριο T^2 .



Σχήμα 7. Διαγράμματα αντικειμένων από δεδομένα Raman (A) και NMR (B).

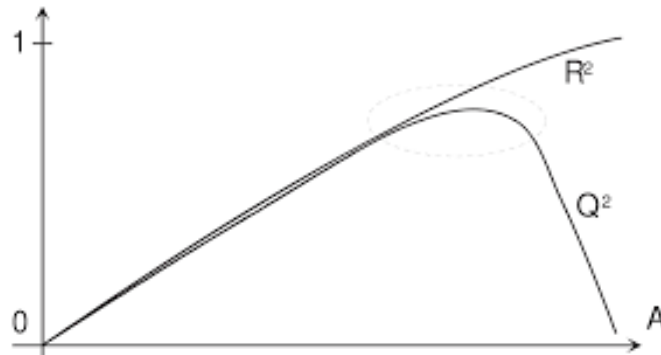
Εκτός από το διάγραμμα αντικειμένων υπάρχει και το διάγραμμα φορτίων (loadings plot). Τα φορτία δείχνουν τις μεταβλητές εκείνες που είναι κυρίως υπεύθυνες για την μετατροπή της θέσης των αρχικών δειγμάτων από τον πίνακα δεδομένων στη νέα τους θέση στο διάγραμμα αντικειμένων.

Συχνά στις μεταβολομικές μελέτες πρέπει να πραγματοποιείται εξομάλυνση των δεδομένων, δηλαδή εξάλειψη του συστηματικού σφάλματος που ενυπάρχει στα δεδομένα και επηρεάζει τα αποτελέσματα της ανάλυσης. Κυρίως στα στατιστικά μοντέλα εφαρμόζεται ως μέθοδος εξομάλυνσης η κλιμάκωση των δεδομένων (scaling) με αποτέλεσμα κάθε μεταβλητή να έχει την ίδια συνεισφορά στο μοντέλο.

Η διαδικασία αυτή μπορεί να γίνει με πολλούς τρόπους αλλά κυρίως χρησιμοποιούνται οι μεθοδολογίες Unit Variance (UV) και το Pareto^{65,70}. Στη μεθοδολογία UV κάθε στήλη (μεταβλητή) πολλαπλασιάζεται με $\frac{1}{sd}$ όπου sd είναι η τυπική απόκλιση της μεταβλητής. Η εφαρμογή του UV έχει σαν αποτέλεσμα να μειώνονται οι «μεγάλες» αριθμητικά μεταβλητές και να μεγαλώνουν οι «μικρές» ώστε κάθε μεταβλητή να έχει παρόμοιας κλίμακας διακύμανση και δεν διαφέρει από τις υπόλοιπες. Στη μεθοδολογία Pareto αντίστοιχα κάθε στήλη (μεταβλητή) πολλαπλασιάζεται με το $\frac{1}{\sqrt{sd}}$ όπου sd είναι και πάλι η τυπική απόκλιση της μεταβλητής.

Ουσιαστικής σημασίας για τη στατιστική ανάλυση είναι η ανακάλυψη πιθανών έκτροπων τιμών. Η ανακάλυψή τους πραγματοποιείται με τη στατιστική δοκιμασία Hotelling's T^2 που δηλώνει ποιές τιμές δεδομένων είναι έκτροπες από αυτές που εμφανίζονται στο διάγραμμα αντικειμένων και έχουν ισχυρή επίδραση σε αυτό. Το Hotelling's T^2 αποτελεί στην ουσία μια γενίκευση του Student's t-test ενώ όταν συνδυαστεί με τα scores καθορίζει την κανονική περιοχή που ανταποκρίνεται σε όρια εμπιστοσύνης 95% ή 99%.

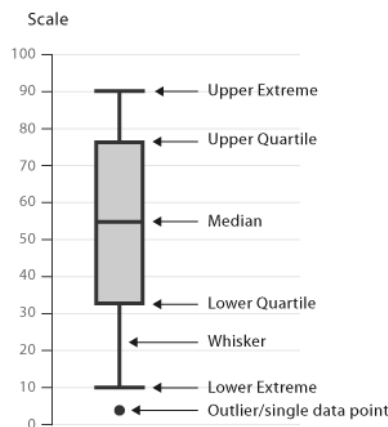
Επιπρόσθετα, οι συντελεστές R^2 (συντελεστής συσχέτισης) και Q^2 (διασταυρούμενος συντελεστής συσχέτισης) δείχνουν την ικανότητα περιγραφής (ή προσαρμογής) και την ικανότητα πρόβλεψης του μοντέλου αντίστοιχα. Ο συντελεστής R^2 δείχνει επίσης το πόσο καλά μπορούν να αναπαραχθούν μαθηματικά τα δεδομένα του συνόλου εκμάθησης.



Σχήμα 8: Γραφική απεικόνιση των R^2 και Q^2 .

Σύμφωνα με το διάγραμμα αυτό τα R^2 και Q^2 τείνουν να έχουν διαφορετική συμπεριφορά όσο αυξάνεται η πολυπλοκότητα του μοντέλου (αυξάνεται όσο συμπεριλαμβάνεται μεγαλύτερος αριθμός κυρίων συνιστωσών). Το R^2 τείνει στο 1 όσο αυξάνεται η πολυπλοκότητα ενώ το Q^2 φαίνεται να εμφανίζει ένα μέγιστο και από εκεί και ύστερα αρχίζει να μειώνεται⁷¹.

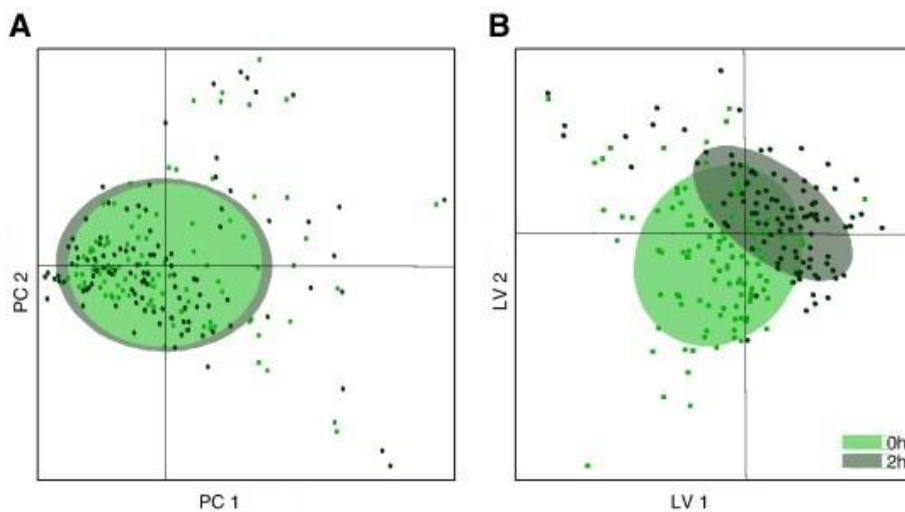
Τέλος ένα ακόμα στατιστικό εργαλείο που χρησιμοποιείται στις μελέτες μεταβολομικής είναι τα θηκογράμματα (boxplots) τα οποία περιγράφουν τα δεδομένα και βοηθούν στην οπτικοποίηση των αποτελεσμάτων. Πρόκειται για ένα διάγραμμα το οποίο παρέχει πληροφορίες σχετικά με την διάμεση τιμή, την ύπαρξη ή όχι κάποιων έκτροπων τιμών καθώς και για το ενδοτεταρτημοριακό εύρος (interquartile range) το χαμηλότερο και υψηλότερο όριο⁷².



Σχήμα 9. Γραφική απεικόνιση θηκογράμματος (box plot)

4.2 Διακρίνουσα ανάλυση μερικών ελαχίστων τετραγώνων (PLS-DA)

Η Διακρίνουσα Ανάλυση Μερικών Ελαχίστων Τετραγώνων (Partial Least Squares - Discriminant Analysis, PLS-DA) αποτελεί επιβλεπόμενη τεχνική πολυπαραμετρικής στατιστικής ανάλυσης⁷³⁻⁷⁵. Πρόκειται για μέθοδο παλινδρόμησης που χρησιμοποιείται για να συσχετίσει τα δεδομένα δύο πινάκων X και Y μέσω ενός πολυπαραγοντικού μοντέλου. Αποτελεί στην ουσία προέκταση της μεθόδου PCA και στηρίζεται στην μέθοδο των προβολών και στην εξαγωγή κυρίων συνιστωσών και για τους δύο πίνακες με στόχο την επίτευξη της καλύτερης δυνατής συσχέτισης μεταξύ των μεταβλητών X και Y. Και σε αυτή την περίπτωση, όπως και στην PCA τα δεδομένα αποτελούνται από ένα πίνακα ο οποίος περιλαμβάνει τις παρατηρήσεις και τις μεταβλητές, με την μόνη διαφορά ότι οι παρατηρήσεις ταξινομούνται σε ομάδες⁷⁶⁻⁷⁸. Τεχνικά προστίθεται μια επιπλέον στήλη όπου καταγράφονται οι ομάδες των παρατηρήσεων. Με τη βοήθεια των κυρίων συνιστωσών που εξάγονται από τις μεταβλητές κάθε ομάδας τα δείγματα θα ταξινομηθούν στην ομάδα όπου ανήκουν.

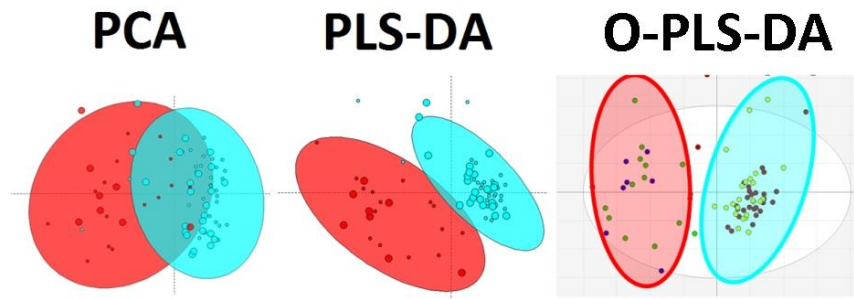


Σχήμα 10. Συγκριτική απεικόνιση των μεθόδων PCA και PLS-DA.

4.3 Ορθογώνια διακρίνουσα ανάλυση μερικών ελαχίστων τετραγώνων (OPLS-DA)

Η Ορθογώνια Διακρίνουσα Ανάλυση Μερικών Ελαχίστων Τετραγώνων (Orthogonal Partial Least Squares – Discriminant Analysis, OPLS-DA) είναι επίσης επιβλεπόμενη τεχνική και αποτελεί επέκταση της προηγούμενης μεθόδου. Χρησιμοποιείται για τις περιπτώσεις εκείνες όπου χρειάζεται πιο σαφής διαχωρισμός μεταξύ των ομάδων των

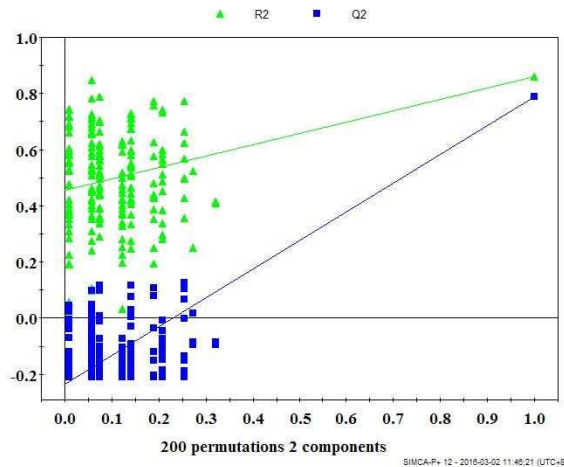
παρατηρήσεων ώστε να γίνουν πιο κατανοητά τα αποτελέσματα^{79,80}. Η OPLS αφαιρεί την διακύμανση των X η οποία είναι κάθετη προς τα Y. Είναι ένα χρήσιμο εργαλείο για την αφαίρεση ανεπιθύμητων συνιστωσών οι οποίες είναι ορθογώνιες ως προς τον πίνακα Y. Η τεχνική αυτή πλεονεκτεί στο γεγονός ότι τα δεδομένα που έχουν αναλυθεί με αυτή την τεχνική είναι πιο εύκολο στη συνέχεια να ερμηνευτούν με λιγότερες και πιο περιεκτικές συνιστώσες.



Σχήμα 11. Συγκριτική απεικόνιση των μεθόδων PCA, PLS-DA και OPLS-DA

4.4 Επικύρωση μοντέλων PLS-DA

Για τον έλεγχο της αξιοπιστίας και της υπερπροσαρμογής των μοντέλων PLS χρησιμοποιούνται τα διαγράμματα μεταθέσεων (permutations plots) τα οποία παράγονται μεταθέτοντας τυχαία τις τιμές Y, παράγοντας με αυτόν τον τρόπο αριθμό συνιστωσών ίσο με εκείνες στο αρχικό μοντέλο^{81,82}. Οι μεταθετικοί έλεγχοι (permutation test) πραγματοποιούνται ώστε να διευκρινιστεί αν πιθανή καλή διάκριση στον διαχωρισμό των ομάδων μπορεί να είναι στατιστικά σημαντική ή να οφείλεται σε τυχαία συσχέτιση. Το διάγραμμα που προκύπτει παρουσιάζει στον άξονα X το συντελεστή διακύμανσης μεταξύ της αρχικής τιμής Y και των τιμών που έχουν προκύψει από τις τυχαίες μεταθέσεις, ενώ στον άξονα Y τις τιμές R^2 και Q^2 για το κάθε μοντέλο⁸³. Ένα μοντέλο κρίνεται αξιόπιστο και ισχυρό, εάν οι τιμές R^2 και Q^2 που προέρχονται από τυχαίες μεταθέσεις των Y είναι χαμηλότερες από τις πραγματικές τιμές R^2 και Q^2 , ενώ αν συμβαίνει το αντίθετο, υπάρχει μεγάλη πιθανότητα τα αποτελέσματα τα μοντέλου να οφείλονται στην τύχη^{84,85}.



Σχήμα 12: Διάγραμμα επικύρωσης μοντέλου PLS-DA.

4.5 Επικύρωση υποψήφιων βιοδεικτών

Η χρήση των βιοδεικτών για την μεταβολομική είναι πολύ σημαντική για τον προσδιορισμό των χαρακτηριστικών των διαφόρων ασθενειών. Ένα αρκετά δημοφιλές στατιστικό εργαλείο που περιγράφει την διαγνωστική ακρίβεια είναι η καμπύλη λειτουργικού χαρακτηριστικού δέκτη ή ROC curve (Receiver Operating Characteristic curve)^{86–89}. Αρχικά εφαρμόστηκε στην ιατρική και σε άλλες επιστήμες ενώ από την δεκαετία του 1990 άρχισε να γίνεται αντικείμενο μελέτης για την στατιστική ανάλυση. Η καμπύλη ROC είναι η γραφική παράσταση που υπάρχει μεταξύ της ευαισθησίας και της ειδικότητας και η τελευταία εκφράζει το ποσοστό των αληθώς θετικών δια του ποσοστού των ψευδών θετικών αποτελεσμάτων.

Η σχέση των αληθώς θετικών ποσοστών και ψευδώς θετικών ποσοστών μπορεί να παρασταθεί γραφικά με μια καμπύλη ROC. Η καμπύλη ROC ορίζεται ως το μοναδιαίο τετράγωνο $[0,1] \times [0,1]$ ξεκινώντας από το σημείο $(0,0)$ και καταλήγοντας στο $(1,1)$. Για να διαπιστώσουμε αν η μέθοδος που χρησιμοποιείται είναι αξιόπιστη αναζητούμε την ύπαρξη του σημείου $(0,1)$ το οποίο αντιστοιχεί σε 100% ευαισθησία (δεν υπάρχουν ψευδώς αρνητικά) και 100% ειδικότητα (δεν υπάρχουν ψευδώς θετικά). Η διαγώνιος $y=x$ αναπαριστά τη διαδικασία τυχαίας πρόβλεψης σε ένα έλεγχο. Το εμβαδόν κάτω από την καμπύλη ROC (Area Under the Curve, AUC) αποτελεί έναν δείκτη διαχωρισμού των δύο ομάδων. Όταν οι δύο κατανομές συμπίπτουν απόλυτα η τιμή του AUC είναι 0.5 και είναι η ελάχιστη που θα μπορούσε να πάρει. Αντίθετα όταν οι δύο κατανομές είναι εντελώς διαφορετικές η τιμή του εμβαδού ανέρχεται στο 1 και είναι η μέγιστη δυνατή.



Σχήμα 13: Καμπύλη roc curve

4.6 Επιλογή των χαρακτηριστικών που συμβάλλουν στην ομαδοποίηση

Αφού πραγματοποιηθεί η στατιστική ανάλυση των δεδομένων το επόμενο στάδιο είναι να βρεθούν εκείνοι οι μεταβολίτες στους οποίους οφείλεται ο παρατηρούμενος διαχωρισμός. Για τον λόγο αυτό χρησιμοποιούνται τα διαγράμματα των Σημαντικών Μεταβλητών για την Προβολή (Variable Importance in the Projection VIP) τα οποία λαμβάνονται από την μέθοδο PLS-DA. Το διάγραμμα αυτό περιγράφει την σχέση των μεταβλητών Y και X. Απαιτείται η τιμή VIP να είναι όσο το δυνατόν υψηλότερη και σε κάθε περίπτωση >1 καθώς όσο πιο υψηλή η τιμή αυτή τόσο πιο σημαντική είναι η συμμετοχή του δεδομένου αυτού στην διάκριση των ομάδων.⁹⁰⁻⁹³

Για την εύρεση των στατιστικά σημαντικών χαρακτηριστικών κατασκευάζεται το διάγραμμα S-plot το οποίο χρησιμοποιείται για την παρουσίαση της συνδιακύμανσης και της συσχέτισης των μεταβλητών με το άνωσμα της πρόβλεψης t[1]. Οι μεταβλητές που απέχουν περισσότερο στον οριζόντιο άξονα του διαγράμματος (X) είναι οι κυρίως υπεύθυνες για τον διαχωρισμό, ενώ αυτές που απέχουν περισσότερο από τον κατακόρυφο άξονα (Y) είναι αυτές που συμβάλλουν περισσότερο στην ομαδοποίηση των δεδομένων.⁹⁴⁻⁹⁶

Παράλληλα συχνά χρησιμοποιούνται και μονοπαραμετρικές (univariate) μέθοδοι ανάλυσης δεδομένων όπως η δοκιμασία κατά Student (t-test) και η. Ο κύριος σκοπός της δοκιμασίας είναι να δείξει αν οι μέσοι όροι των ομάδων ελέγχου και παρέμβασης διαφέρουν στατιστικά σημαντικά μεταξύ τους. Για την δοκιμασία t ισχύει ότι η μηδενική υπόθεση είναι $H_0: \mu_{\text{group1}} = \mu_{\text{group2}}$ και η εναλλακτική υπόθεση είναι $H_a: \mu_{\text{group1}} \neq \mu_{\text{group2}}$. Ο εντοπισμός των μεταβολιτών που ευθύνονται για τον διαχωρισμό των ομάδων εξαρτάται από την τιμή p η οποία αν είναι μικρότερη από κάποια τιμή, ο καθορισμός της οποίας εξαρτάται από την οριζόμενη στάθμη εμπιστοσύνης και συνήθως είναι 0.05, τότε η μηδενική υπόθεση απορρίπτεται. Αντίθετα, αν η τιμή P είναι

μεγαλύτερη από την ορισμένη τιμή τότε ισχύει η μηδενική υπόθεση και θεωρείται ότι δεν υπάρχει σημαντική διαφορά μεταξύ των μέσων τιμών για τις δύο ομάδες. Με άλλα λόγια αυτό σημαίνει ότι ο μεταβολίτης έχει μικρή ικανότητα να συμβάλλει στον διαχωρισμό των ομάδων.⁹⁷⁻⁹⁹

Τέλος κατά την στατιστική ανάλυση λαμβάνει χώρα και μία επιπλέον δοκιμασία η ανάλυση της διακύμανσης (Analysis Of Variance ANOVA)¹⁰⁰⁻¹⁰². Η ANOVA λοιπόν είναι μία στατιστική μέθοδος με την οποία η μεταβλητότητα που υπάρχει σ' ένα σύνολο δεδομένων διασπάται στις επιμέρους συνιστώσες της με στόχο την κατανόηση της σημαντικότητας των διαφορετικών πηγών προέλευσής της. Ελέγχει τη διαφορά στις μέσες τιμές προκειμένου να αποφασιστεί, αν η διακύμανση μεταξύ δύο ή περισσότερων ομάδων είναι μεγαλύτερη από τη διακύμανση εντός των ομάδων. Υπάρχουν δύο είδη αναλύσεων, ο μονοδιάστατος σχεδιασμός (one-way design) και ο παραγοντικός σχεδιασμός (factorial design), που είναι συνήθως διδιάστατος (two-way design).

Για την πραγματοποίηση του συσχετισμένου ελέγχου t, απαιτείται για τα δεδομένα μας να ισχύουν οι ακόλουθες παραδοχές:

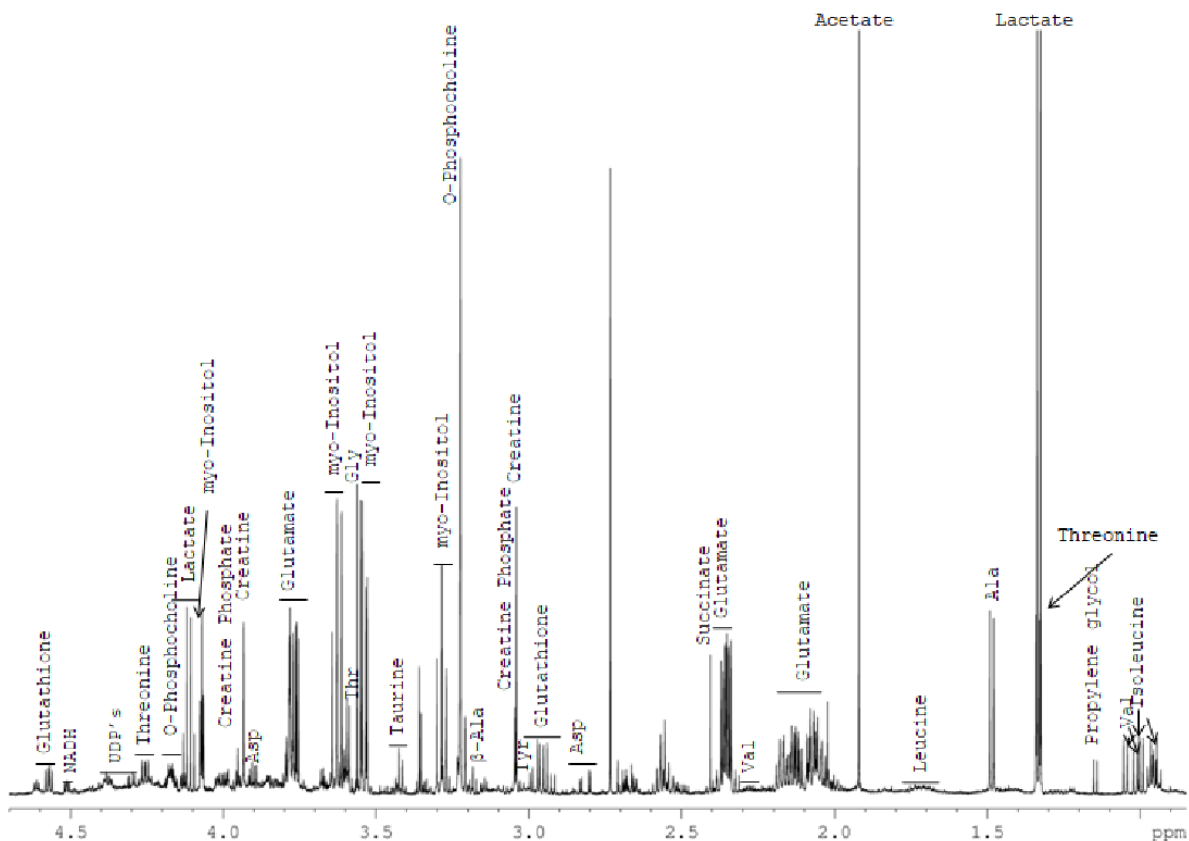
- Η εξαρτημένη μεταβλητή ελέγχου θα πρέπει να προσεγγίζει την κανονική κατανομή. Ο έλεγχος κανονικότητας μπορεί να γίνει ποσοτικά με χρήση του Kolmogorov-Smirnov test^{103,104} ή του Shapiro-Wilk test (Παράρτημα 1).
- Θα πρέπει στις τιμές να μην υπάρχουν σημαντικά ακραίες τιμές (outliers). Το πρόβλημα με τις ακραίες τιμές είναι ότι μπορεί να έχουν αρνητική επίδραση στο t-test, μειώνοντας την ακρίβεια των αποτελεσμάτων που προκύπτουν. Ο έλεγχος για ακραίες τιμές μπορεί να πραγματοποιηθεί με τη δημιουργία και παρατήρηση του θηκογράμματος.
- Τα στοιχεία του δείγματος θα πρέπει να είναι ανεξάρτητα (μη συσχετιζόμενα), το οποίο σημαίνει ότι δεν υπάρχει σχέση μεταξύ των παρατηρήσεων. Η παραβίαση αυτής της παραδοχής συνήθως έχει να κάνει με λανθασμένη σχεδίαση της μελέτης.
- Η εξαρτημένη μεταβλητή ελέγχου πρέπει να είναι ποσοτική μεταβλητή, είτε διαστήματος (interval) είτε αναλογίας (ratio).
- Οι παρατηρήσεις πρέπει να είναι ανεξάρτητες, δηλαδή θα πρέπει να έχει εξασφαλιστεί πως μια παρατήρηση από το ένα δείγμα δεν πρόκειται να ανήκει και στο άλλο.
- Οι τιμές των μεταβλητών θα πρέπει να χαρακτηρίζονται από ομοιογένεια, δηλαδή οι διακυμάνσεις των δύο ομάδων θα πρέπει να είναι ίσες.

ΚΕΦΑΛΑΙΟ 5

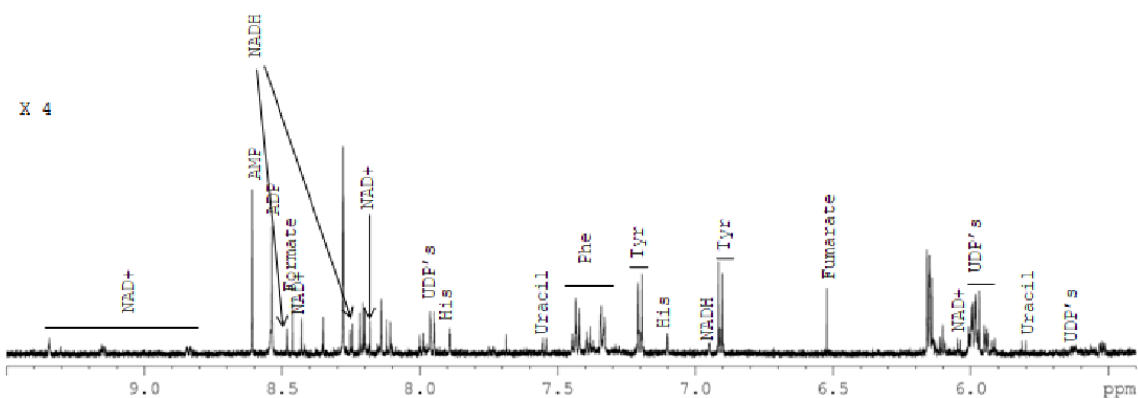
ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Φασματοσκοπία NMR

Με την φασματοσκοπία NMR μελετήθηκαν συνολικά 40 δείγματα, 10 από κάθε κυτταρική σειρά όπως αναφέρθηκε παραπάνω (2.2 Δείγματα). Στις παρακάτω εικόνες παρουσιάζονται χαρακτηριστικές περιοχές αντιπροσωπευτικού φάσματος $1D^1H-NMR$ των υδατικών εκχυλισμάτων με την απόδοση των κορυφών. Επιπλέον οι πίνακες που παρουσιάζονται στη συνέχεια αναφέρονται στο σύνολο των μεταβολιτών που έχουν ανιχνευθεί και ταυτοποιηθεί, οι χημικές μετατοπίσεις καθώς και η πολλαπλότητα των κορυφών συντονισμού.



Εικόνα 1. Αλειφατική περιοχή (4.7-0.7 ppm) αντιπροσωπευτικού φάσματος $1D^1H-NMR$ των υδατικών εκχυλισμάτων της κυτταρικής σειράς T24 (grade 3, A3) σε D_2O , pH 7.4, στους 300 K. Στο φάσμα φαίνεται η απόδοση των μεταβολιτών.



Εικόνα 2. Αρωματική περιοχή (9.5-5 ppm) αντιπροσωπευτικού φάσματος $1D^1H-NMR$ των υδατικών εκχυλισμάτων της κυτταρικής σειράς T24 (grade 3, A3) σε D_2O , pH 7.4, στους 300 K. Στο φάσμα φαίνεται η απόδοση των μεταβολιτών.

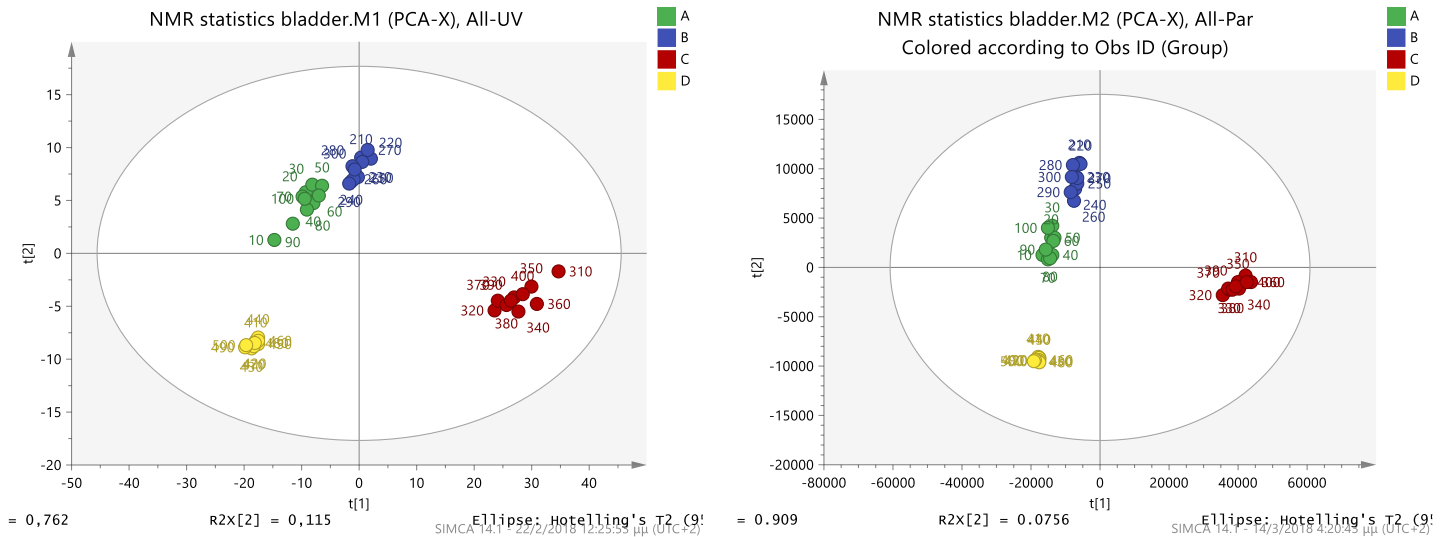
Πίνακας 4. Χημική μετατόπιση μεταβολιτών και πολλαπλότητα κορυφής.
s: singlet (απλή)/ d: doublet (διπλή)/t: triplet (τριπλή)/ dd: doublet of doublets (διπλή διπλών)

| | Metabolites | $^1H \delta$ (ppm) - Multiplicity |
|----|----------------------|---|
| 1 | Leucine | 3.74 q; 1.75 m; 1.71 m; 1.69 m; 0.97 d; 0.96 d |
| 2 | Isoleucine | 0.94 t; 1.02 d, 1.27 m; 3.68 d |
| 3 | Valine | 3.61 d; 2.28 m; 1.05 d; 1.00 d |
| 4 | Threonine | 4.25 m; 3.59 d; 1.33 d |
| 5 | N-Isovaleroylglycine | 0.96 d |
| 6 | Asparagine | 2.94 dd ; 2.87 dd |
| 7 | Methionine | 2.65 t; 2.14 s |
| 8 | GlutamicAcid | 3.76 q; 2.36 m; 2.34 m; 2.13 m, 2.06 m |
| 9 | AsparticAcid | 3.90 q; 2.81 dd; 2.69 dd |
| 10 | Histidine | 7.89 d; 7.10 s |
| 11 | Tyrosine | 7.20 d; 6.91 d; 3.95 dd |
| 12 | Phenylalanine | 7.43 t; 7.38 t; 7.34 d; 4.00 d |
| 13 | Tryptophan | 7.74 d; 7.55 ; 7.33 s; 7.20 t |
| 14 | Glycine | 3.56 s |
| 15 | β -Alanine | 3.18 t; 2.56 t |
| 16 | Alanine | 3.79 q; 1.48 d |
| 17 | Proline | 4.14; 3.43 m; 3.34 m; 2.35; 2.06; 1.99 m |

| | | |
|----|--|--|
| 18 | Taurine | 3.43 t; 3.26 t |
| 19 | Glutathione | 4.57 q; 2.98; 2.93; 2.58 m; 2.54 m; 2.17 m |
| 20 | Malate | 2.67 dd; 2.51 dd ; 2.02 s |
| 21 | Creatine | 3.93 s; 3.04 s; |
| 22 | Creatine Phosphate | 3.95 s; 3.05 s; |
| 23 | Myo-Inositol | 4.07 t; 3.63 t; 3.54 dd; 3.29 t |
| 24 | O-Phosphocholine | 4.17 m; 3.60 m; 3.22 s |
| 25 | sn-Glycero-3-phosphocholine | 3.23 s |
| 26 | O-Acetylcholine | 4.55 m; 3.21 s; 2.14 s |
| 27 | Adenosine Triphosphate (ATP) | 8.54 s; 8.28 s; 6.15 d; 4.61 t; 4.40 m |
| 28 | Adenosine Diphosphate (ADP) | 8.54 s; 8.28 s; 6.15 d; 4.61 t; 4.37 m |
| 29 | Adenosine Monophosphate (AMP) | 8.61 s; 8.28 s; 6.15 d; 4.51 dd; 4.36 m; 4.01 m |
| 30 | Gouanosine Triphosphate (GTP) | 8.14 s; 5.92 d |
| 31 | Nicotinamide adenine dinucleotide (NAD+) | 9.34; 9.15 d; 8.84 d; 8.43 s; 8.20 m; 8.18 s; 6.10 d; 6.04 d |
| 32 | NADH | 8.48 s; 8.24 s; 6.95 m |
| 33 | Adenine | 8.20 s; 8.14 s |
| 34 | Inosine | 8.35 s; 8.22 s; 6.11 s |
| 35 | IMP | 8.25 s |
| 36 | NADPH | 8.48 s; 8.24 s; 6.95 m |
| 37 | UDP-glucose | 7.95 d; 5.56 dd |
| 38 | UDP-N-Acetylglucosamine | 7.96 d; 5.99 d; 5.98 d; 5.52 dd; 2.08 s |
| 39 | UDP-galactose | 8.00 d; 5.63 m |
| 40 | UDP-glucuronate | 7.94 d |
| 41 | UMP | 8.11 d; 5.99 d; 5.97 d; 4.40 t |
| 42 | Uracil | 7.55 d; 5.81 d |
| 43 | Oxypurinol | 8.21 s |
| 44 | Acetate | 1.92 s |
| 45 | Lactic Acid | 4.11 q; 1.33 d |
| 46 | Succinic Acid | 2.41 s |
| 47 | Fumaric Acid | 6.52 s |
| 48 | Malic Acid | 4.30 dd; 2.677 dd; 2.378 dd |
| 49 | Formic Acid | 8.46 s |

5.2 Χημειομετρική ανάλυση δεδομένων NMR

Μετά την τμηματοποίηση (bucketing) των NMR φασμάτων με την χρήση του λογισμικού AMIX¹⁰⁵, ακολουθεί η εισαγωγή του πίνακα που δημιουργείται στο λογισμικό στατιστικής ανάλυσης SIMCAP-14.1¹⁰⁶. Τελικά έγινε επιλογή των καταλληλότερων μοντέλων (βάσει των αποτελεσμάτων διασταυρούμενης επικύρωσης^{107,108} (cross validation, CV), και από αυτά παρουσιάζονται παρακάτω τα διαγράμματα αντικειμένων (scores plots) και φορτίων (loadings plots), καθώς και τα διαγράμματα S-plots¹⁰⁹⁻¹¹¹ και VIP plots^{112,113}.

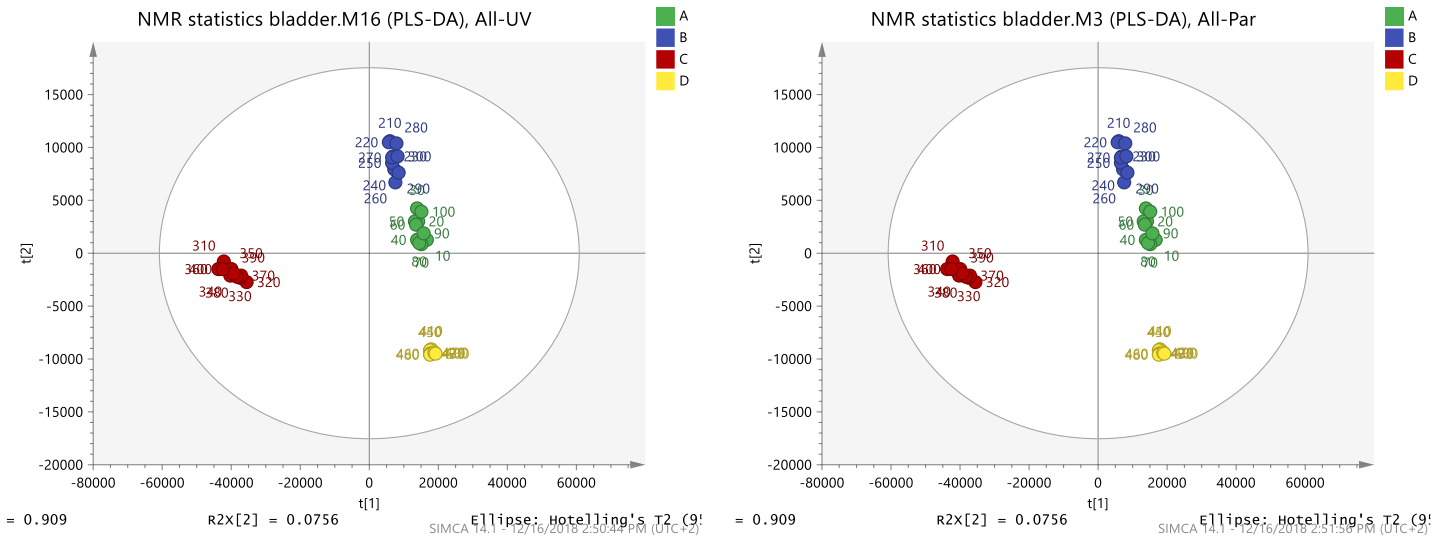


Διάγραμμα 1. Διαγράμματα αντικειμένων PCA για τις τέσσερις κυτταρικές σειρές με UV scaling αριστερά και Pareto scaling δεξιά.

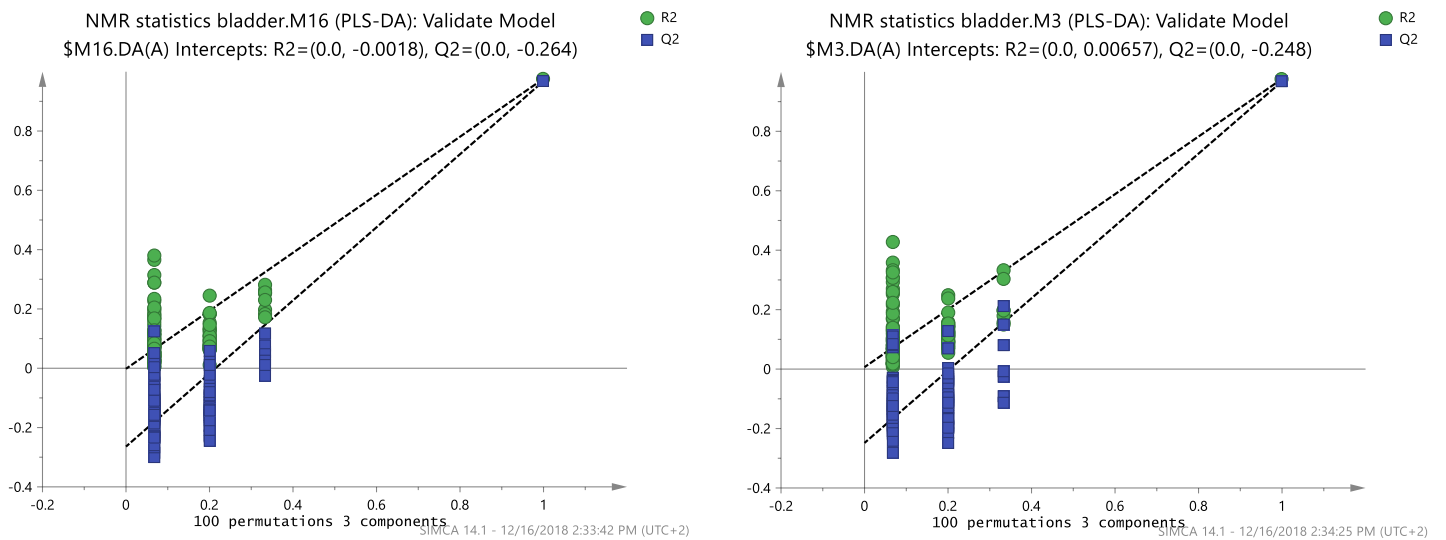
Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, A), με μπλε χρώμα η κυτταρική σειρά RT112 (Grade2, B), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, C), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP(Grade4, D).

Παραπάνω φαίνεται το διάγραμμα αντικειμένων PCA για όλες τις κυτταρικές σειρές με UV scaling αριστερά και Pareto scaling δεξιά. Η περιγραφική ικανότητα του μοντέλου με UV scaling (αριστερό διάγραμμα) είναι $R^2X(\text{cum})=0.897$ και η προβλεπτική του ικανότητα είναι $Q^2(\text{cum})=0.878$. Με Pareto scaling (δεξιό διάγραμμα), που αποτελεί και το πλέον χρησιμοποιούμενο τρόπο κλιμάκωσης για δεδομένα μεταβολομικής^{114,115}, η περιγραφική ικανότητα είναι $R^2X(\text{cum})=0.991$ και η προβλεπτική του ικανότητα είναι $Q^2(\text{cum})=0.985$.

Όπως φαίνεται από τα διαγράμματα PCA έχει ήδη επιτευχθεί ένας καλός διαχωρισμός μεταξύ των τεσσάρων ομάδων. Η πολυπαραμετρική στατιστική ανάλυση συνεχίστηκε με την εφαρμογή της επιβλεπόμενης μεθόδου PLS-DA. Όπως και προηγουμένως, έτσι και για την μέθοδο PLS-DA έγινε χρήση των δύο τρόπων κλιμάκωσης UV και Pareto scaling (διάγραμμα 2). Όπως φαίνεται υπάρχει επίσης καλός διαχωρισμός ενώ οι δύο κλιμακώσεις φαίνεται να μην έχουν μεγάλες διαφορές στο διάγραμμα αντικειμένων PLS-DA. Η επικύρωση του μοντέλου PLS-DA έγινε με τον μεταθετικό έλεγχο (permutation test), επιτρέποντας 100 τυχαίες μεταθέσεις, προκειμένου να αξιολογηθεί η προβλεψιμότητα του μοντέλου και για τους δύο τρόπους κλιμάκωσης (διάγραμμα 3). Στην κλιμάκωση UV (αριστερά) οι τομές ήταν $R^2=(0.0, -0.0018)$, $Q^2=(0.0, -0.264)$, ενώ στην κλιμάκωση Pareto (δεξιά) οι τομές ήταν $R^2=(0.0, 0.00657)$, $Q^2=(0.0, -0.248)$. Και τα δύο μοντέλα θεωρούνται έγκυρα.

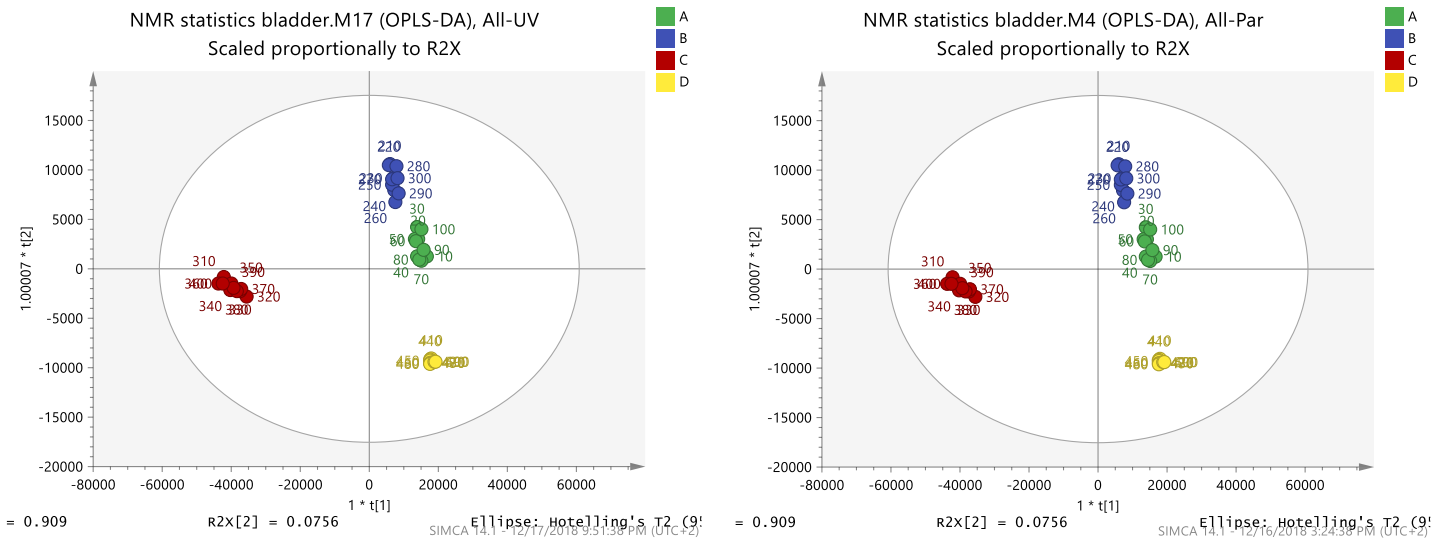


Διάγραμμα 2. Διαγράμματα αντικειμένων PLS-DA για τις τέσσερις κυτταρικές σειρές με UV scaling αριστερά και Pareto scaling δεξιά. Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, A), με μπλε χρώμα η κυτταρική σειρά RT112 (Grade2, B), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, C), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP(Grade4, D).

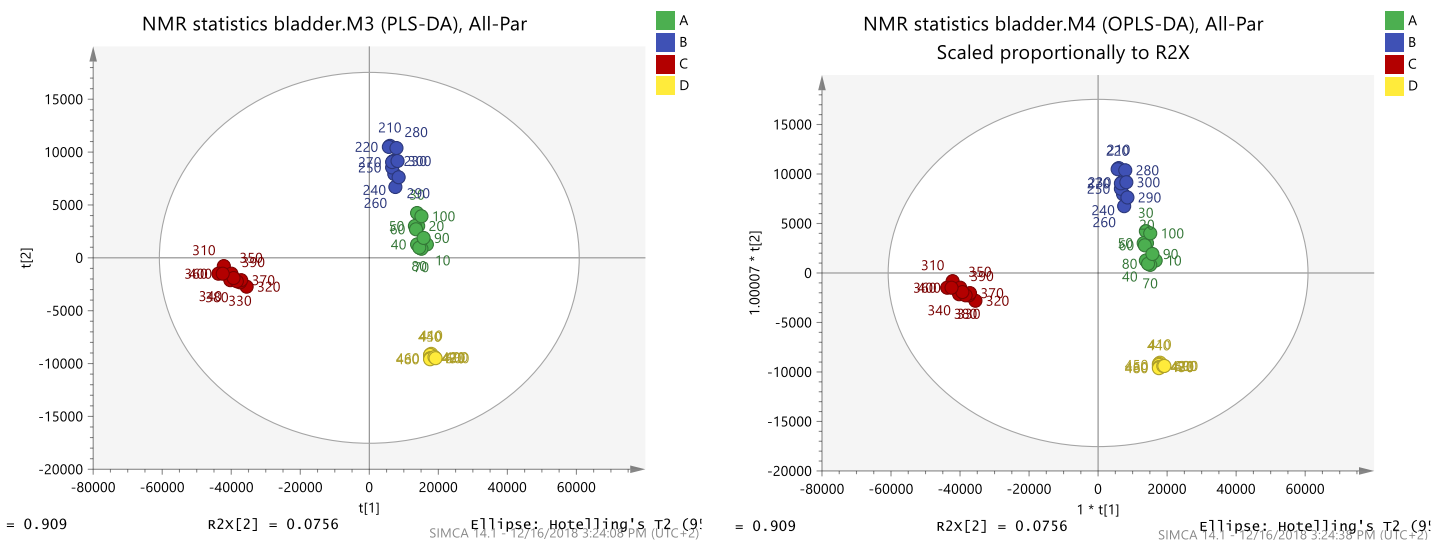


Διάγραμμα 3. Μεταθετικός έλεγχος με 100 τυχαίες μεταθέσεις επιτρεπόμενες για το μοντέλο PLS-DA με UV Scaling αριστερά και Pareto scaling δεξιά.

Τέλος, σε αυτά τα δείγματα πραγματοποιήθηκε στατιστική ανάλυση με την επίσης επιβλεπόμενη μέθοδο OPLS-DA και με UV scaling καθώς και με Pareto scaling όπου και πάλι οι διαφορές ήταν ελάχιστες (Διάγραμμα 4). Επίσης όπως φαίνεται από το διάγραμμα 5 η επιβλεπόμενη μέθοδος PLS-DA και η OPLS-DA παρουσιάζουν την ίδια εικόνα καθώς ο διαχωρισμός μεταξύ των ομάδων ήταν ήδη πολύ καλός από την μέθοδο PLS-DA.



Διάγραμμα 4. Διαγράμματα αντικειμένων OPLS-DA για τις τέσσερις κυτταρικές σειρές με UV scaling αριστερά και Pareto scaling δεξιά. Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, A), με μπλε χρώμα η κυτταρική σειρά RT112 (Grade 2, B), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, C), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP (Grade 4, D).



Διάγραμμα 5. Συγκριτικά διαγράμματα αντικειμένων PLS-DA αριστερά (Pareto scaling) και OPLS-DA δεξιά (Pareto scaling) για τις τέσσερις κυτταρικές σειρές

Στη συνέχεια έγιναν επιμέρους συγκριτικές στατιστικές αναλύσεις με την επιβλεπόμενη μέθοδο OPLS-DA με Pareto scaling για κάθε κυτταρική σειρά ανά δύο. Πραγματοποιήθηκαν δηλαδή έξι επιμέρους αναλύσεις ανά δύο για τις ομάδες μας (grade1-grade2, grade1-grade3, grade1-grade4, grade2-grade3, grade2-grade4, grade3-grade4).

Εν συνεχεία, προκειμένου να εντοπιστούν οι μεταβολίτες εκείνοι που συμβάλουν στον παρατηρούμενο διαχωρισμό των ομάδων μας, υπολογίστηκαν οι τιμές VIP από το μοντέλο PLS-DA και πραγματοποιήθηκε το διάγραμμα S από το μοντέλο OPLS-DA που έγινε για κάθε μία από τις έξι επιμέρους αναλύσεις. Από το διάγραμμα S επιλέχθηκαν τα χαρακτηριστικά εκείνα που βρίσκονται απομακρυσμένα από τους άξονες x και y.

Από το διάγραμμα 7 προκύπτει ότι οι μεταβολίτες που συμβάλλουν στον διαχωρισμό μεταξύ των κυτταρικών σειρών RT4 (Grade 1) και RT112 (Grade 2) είναι οι lactate, myo-inositol, glutamate, o-phosphocholine, και choline.

Από το διάγραμμα 8 προκύπτει ότι οι μεταβολίτες που συμβάλλουν στον διαχωρισμό μεταξύ των κυτταρικών σειρών RT4 (Grade 1) και T24 (Grade 3) είναι οι o-phosphocholine, lactate, myo-inositol, creatine, glutamate, threonine, glycine και glytathione.

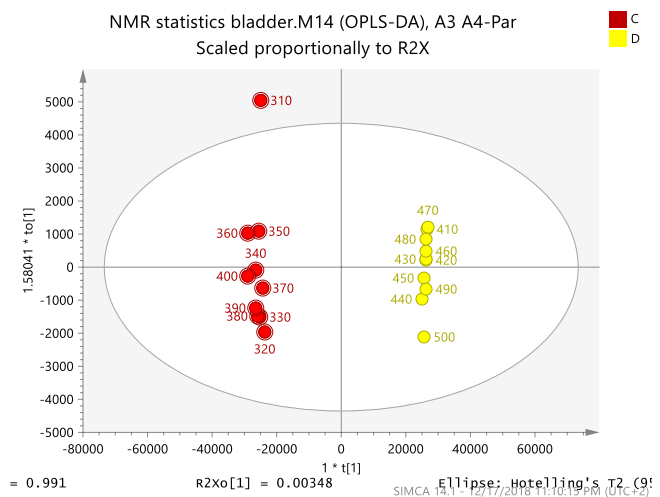
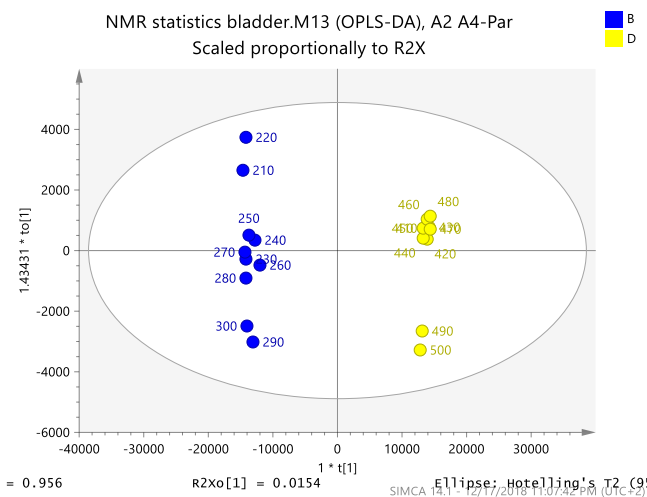
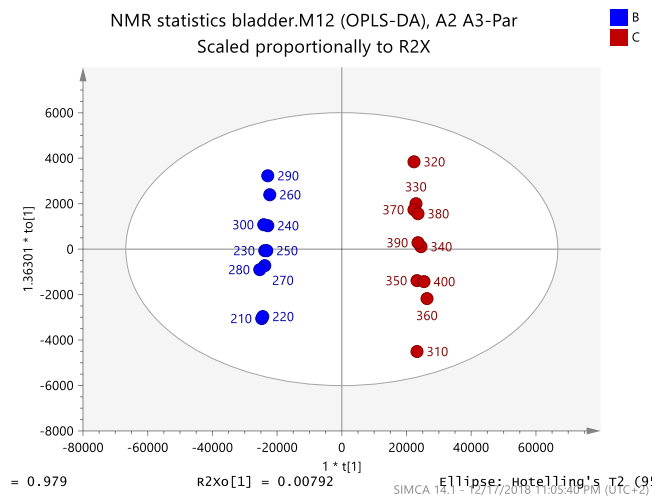
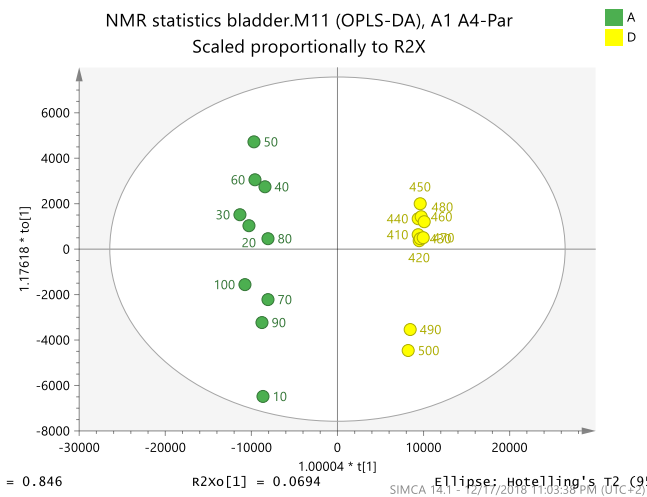
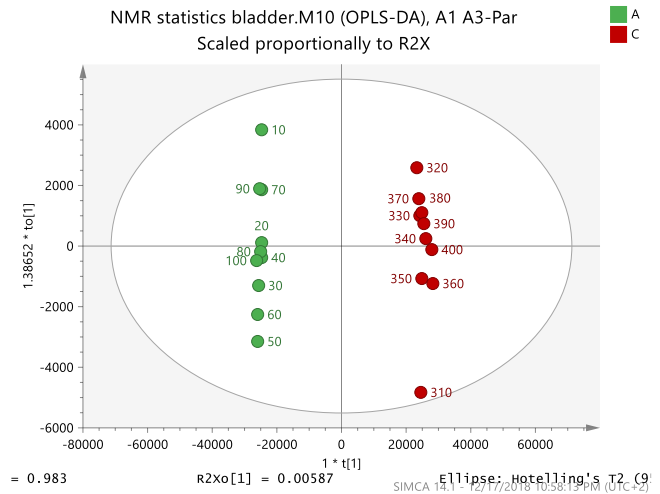
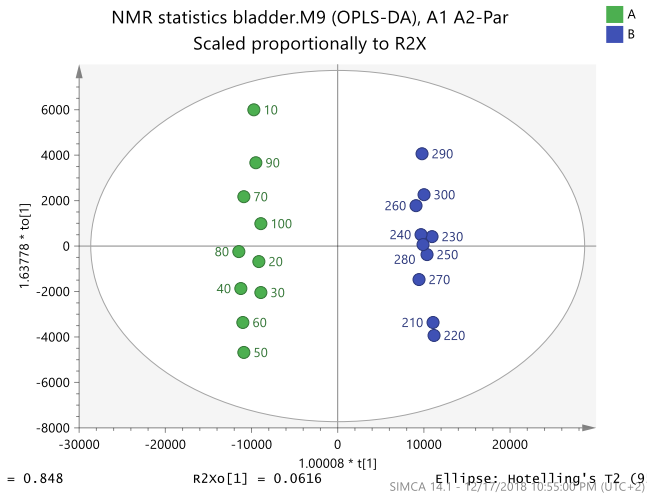
Από το διάγραμμα 9 προκύπτει ότι οι μεταβολίτες που συμβάλλουν στον διαχωρισμό μεταξύ των κυτταρικών σειρών RT4 (Grade 1) και TCCSUP (Grade 4) είναι οι lactate, glutamate, o-phosphocholine, aspartate, propylene glycol και choline.

Από το διάγραμμα 10 προκύπτει ότι οι μεταβολίτες που συμβάλλουν στον διαχωρισμό μεταξύ των κυτταρικών σειρών RT112 (Grade 2) και T24 (Grade 3) είναι οι lactate, glutamate, o-phosphocholine, myo-inositol, propylene glycol, threonine, glycine, histidine, uracil, creatine phosphate και creatine.

Από το διάγραμμα 11 προκύπτει ότι οι μεταβολίτες που συμβάλλουν στον διαχωρισμό μεταξύ των κυτταρικών σειρών RT112 (Grade 2) και TCCSUP (Grade 4) είναι οι lactate, glutamate, o-phosphocholine, myo-inositol, propylene glycol, , glycine, aspartate, choline, UMP, και UDP's.

Από το διάγραμμα 12 προκύπτει ότι οι μεταβολίτες που συμβάλλουν στον διαχωρισμό μεταξύ των κυτταρικών σειρών T24 (Grade 3) και TCCSUP (Grade 4) είναι οι lactate, glutamate, o-phosphocholine, myo-inositol, glycine, threonine, creatine phosphate και creatine.

Στο παράρτημα της παρούσας εργασίας βρίσκονται επιπλέον οι δοκιμασίες normality test, one-way ANOVA, ROC area, t-test, Fold change και υπολογισμός του Z-score.

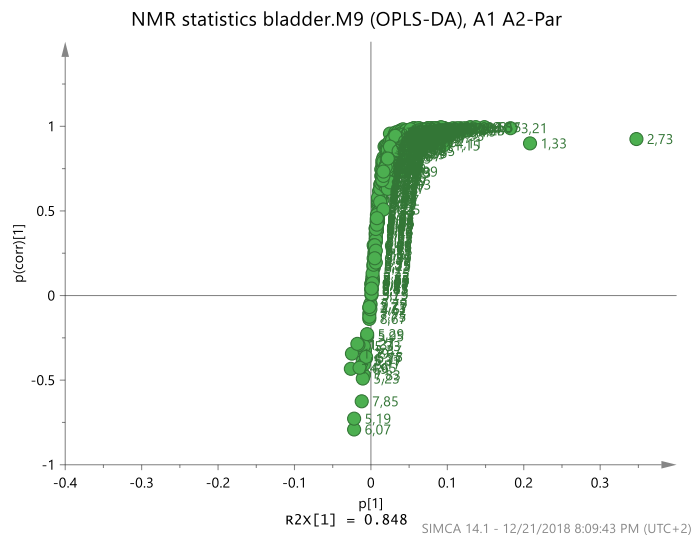


Διάγραμμα 6. Διαγράμματα αντικειμένων OPLS-DA για τις τέσσερις κυτταρικές σειρές (Pareto scaling).

Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, A), με μπλε χρώμα η κυτταρική σειρά RT12 (Grade 2, B), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, C), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP (Grade 4, D).

Πίνακας 5. Οι επιλεγμένοι μεταβολίτες που συμβάλλουν στο διαχωρισμό των τεσσάρων ομάδων.

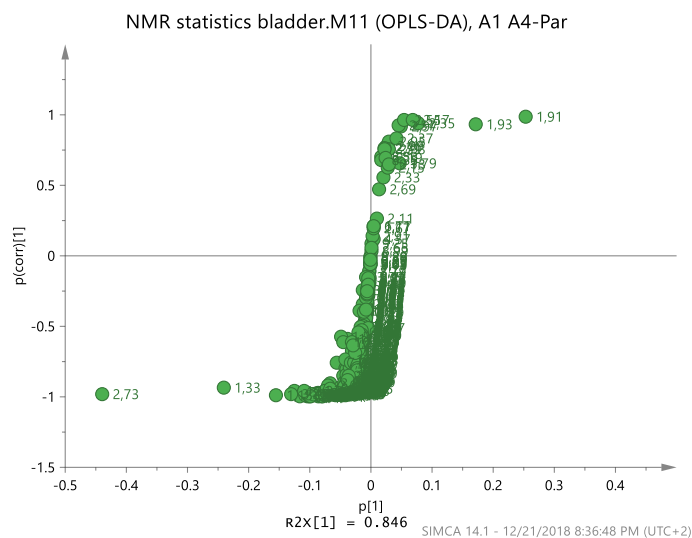
| Τιμή VIP | Metabolite |
|-----------------|-------------------------------|
| 4.75735 | Lactate |
| 4.51112 | O-Phosphocholine |
| 3.57687 | Glutamate |
| 3.12907 | myo-Inositol |
| 3.05812 | Glutamate |
| 2.97982 | myo-Inositol |
| 2.97636 | Glutamate |
| 2.94281 | myo-Inositol |
| 2.74401 | myo-Inositol |
| 2.71671 | Glutamate |
| 2.64766 | myo-Inositol |
| 2.44198 | myo-Inositol |
| 2.40978 | Glutamate |
| 2.37539 | Glutamate |
| 2.32232 | Glutamate |
| 2.24183 | Glutamate |
| 2.18385 | Glutamate |
| 2.16279 | Creatine |
| 2.16182 | Threonine |
| 2.15265 | Glutamate |
| 2.10913 | Creatine + Creatine Phosphate |
| 2.08561 | myo-Inositol |
| 2.0672 | Lactate |
| 2.04073 | Glutamate |
| 1.99441 | O-Phosphocholine |
| 1.96525 | Alanine |
| 1.96459 | Glycine |
| 1.91703 | Glutamate |
| 1.88671 | Glutamate |
| 1.87761 | Creatine |
| 1.87366 | Glutamate |
| 1.87122 | Glutathione |
| 1.86555 | myo-Inositol |
| 1.86456 | Acetate |
| 1.80196 | Alanine |
| 1.75592 | Threonine |
| 1.67976 | Proline |
| 1.66426 | Leucine |
| 1.63134 | Glutathione |
| 1.62002 | Creatine Phosphate |
| 1.57736 | UDP's, UMP |
| 1.57411 | UDP's, UMP |
| 1.57007 | Lactate |
| 1.54544 | Glutathione |
| 1.54071 | Threonine |
| 1.52388 | Threonine |
| 1.50215 | UDP's |



Διάγραμμα 7. Διάγραμμα S plot από το αντίστοιχο μοντέλο OPLS-DA για τις κυτταρικές σειρές RT4 (Grade 1) και RT112 (Grade 2).

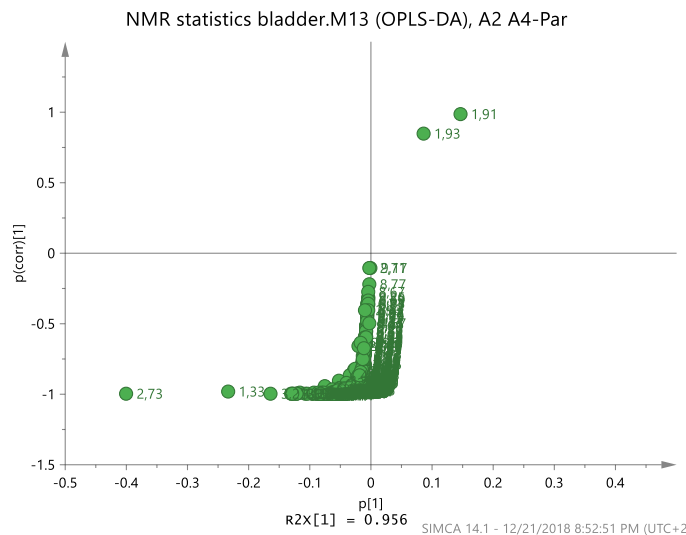
Πίνακας 6. Οι επιλεγμένοι μεταβολίτες που συμβάλλουν στον διαχωρισμό των δύο κυτταρικών σειρών.

| Διάγραμμα S | Τιμή VIP | Metabolite |
|-------------|----------|------------------|
| 4.11 | 2.0672 | Lactate |
| 4.07 | 2.94281 | myo-Inositol |
| 3.79 | 3.05812 | Glutamate |
| 3.77 | 2.97636 | Glutamate |
| 3.63 | 2.97982 | myo-Inositol |
| 3.61 | 2.74401 | myo-Inositol |
| 3.55 | 3.12907 | myo-Inositol |
| 3.53 | 2.64766 | myo-Inositol |
| 3.29 | 2.44198 | myo-Inositol |
| 3.23 | 4.51112 | O-Phosphocholine |
| 3.21 | 1.30353 | Choline |
| 2.35 | 3.57687 | Glutamate |
| 2.09 | 2.04073 | Glutamate |
| 2.07 | 2.71671 | Glutamate |
| 1.33 | 4.75735 | Lactate |



Πίνακας 9. Οι επιλεγμένοι μεταβολίτες που συμβάλλουν στον διαχωρισμό των δύο κутταρικών σειρών.

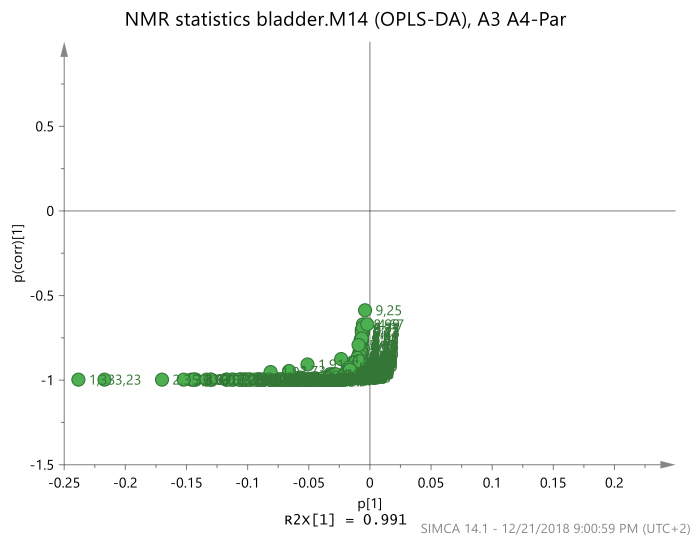
| Διάγραμμα S | Τιμή VIP | Metabolite |
|--------------------|-----------------|-------------------------------|
| 4.17 | 1.99441 | O-Phosphocholine |
| 4.11 | 2.0672 | Lactate |
| 4.07 | 2.94281 | myo-Inositol |
| 3.93 | 2.16279 | Creatine |
| 3.79 | 3.05812 | Glutamate |
| 3.77 | 2.97636 | Glutamate |
| 3.75 | 2.32232 | Glutamate |
| 3.65 | 2.08561 | myo-Inositol |
| 3.63 | 2.97982 | myo-Inositol |
| 3.61 | 2.74401 | myo-Inositol |
| 3.59 | 2.16182 | Threonine |
| 3.57 | 1.96459 | Glycine |
| 3.55 | 3.12907 | myo-Inositol |
| 3.53 | 2.64766 | myo-Inositol |
| 3.29 | 2.44198 | myo-Inositol |
| 3.23 | 4.51112 | O-Phosphocholine |
| 3.05 | 2.10913 | Creatine + Creatine Phosphate |
| 3.03 | 1.87761 | Creatine |
| 2.37 | 2.37539 | Glutamate |
| 2.35 | 3.57687 | Glutamate |
| 2.33 | 1.88671 | Glutamate |
| 2.17 | 2.18385 | Glutamate |
| 2.15 | 2.15265 | Glutamate |
| 2.13 | 2.24183 | Glutamate |
| 2.11 | 1.91703 | Glutamate |
| 2.09 | 2.04073 | Glutamate |
| 2.07 | 2.71671 | Glutamate |
| 2.05 | 2.40978 | Glutamate |
| 2.03 | 1.87366 | Glutamate |
| 1.49 | 1.96525 | Alanine |



Διάγραμμα 11. Διάγραμμα S plot από το αντίστοιχο μοντέλο OPLS-DA για τις κυτταρικές σειρές RT112 (Grade2) και TCCSUP (Grade4).

Πίνακας 10. Οι επιλεγμένοι μεταβολίτες που συμβάλλουν στον διαχωρισμό των δύο κυτταρικών σειρών.

| Διάγραμμα S | Τιμή VIP | Metabolite |
|-------------|----------|------------------|
| 4.13 | 1.57007 | Lactate |
| 4.11 | 2.0672 | Lactate |
| 4.07 | 2.94281 | myo-Inositol |
| 3.99 | 1.57736 | UDP's, UMP |
| 3.63 | 2.97982 | myo-Inositol |
| 3.61 | 2.74401 | myo-Inositol |
| 3.57 | 1.96459 | Glycine |
| 3.55 | 3.12907 | myo-Inositol |
| 3.53 | 2.64766 | myo-Inositol |
| 3.29 | 2.44198 | myo-Inositol |
| 3.23 | 4.51112 | O-Phosphocholine |
| 2.09 | 2.04073 | Glutamate |
| 2.05 | 2.40978 | Glutamate |
| 1.33 | 4.75735 | Lactate |
| 1.93 | 1.86456 | Lactate |

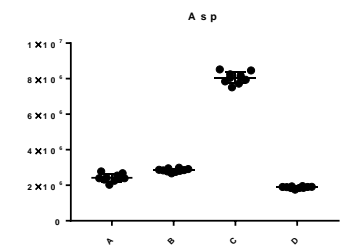
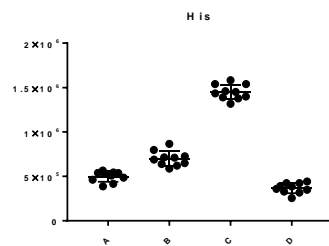
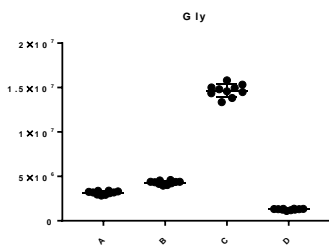
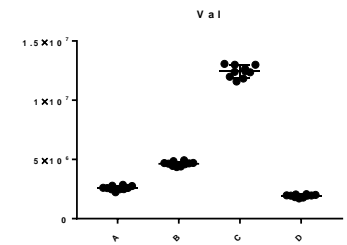
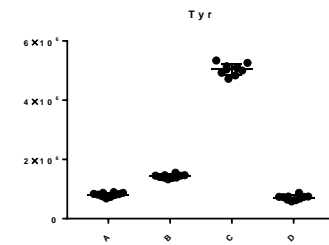
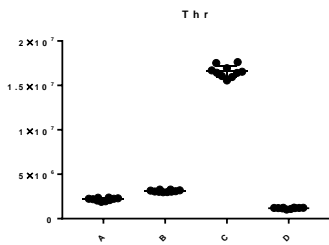
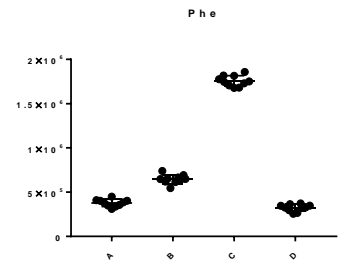
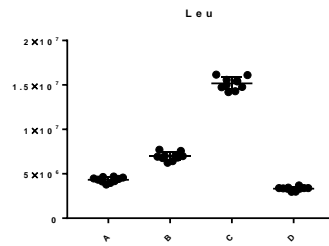
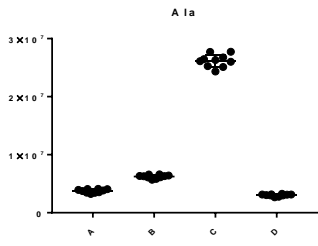


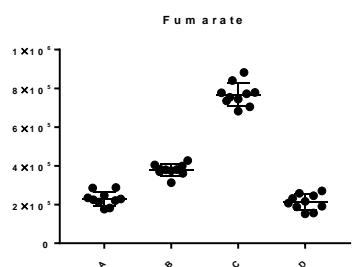
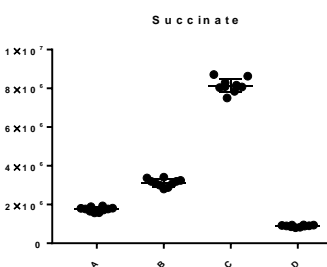
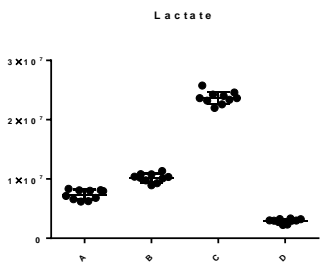
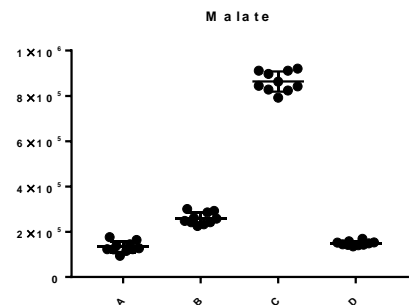
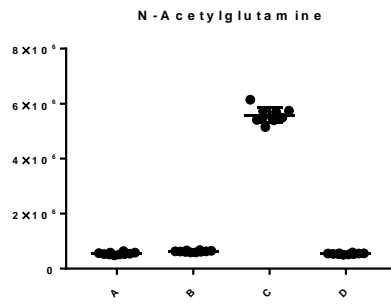
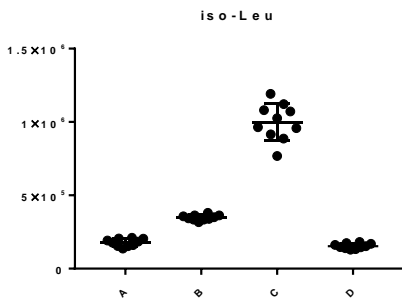
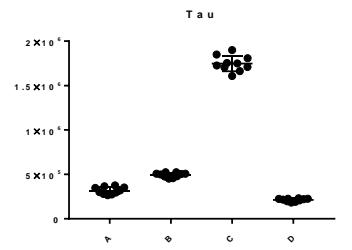
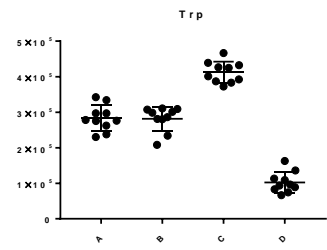
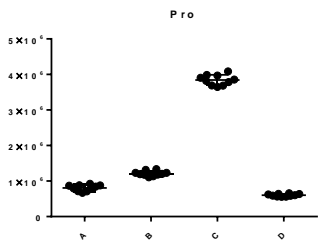
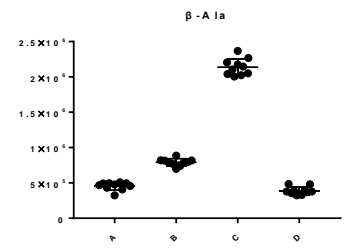
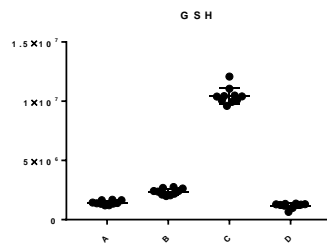
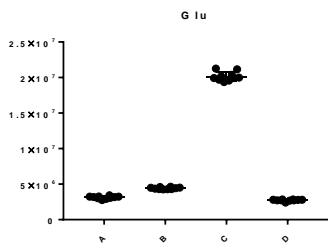
Διάγραμμα 12. Διάγραμμα S plot από το αντίστοιχο μοντέλο OPLS-DA για τις κυτταρικές σειρές T24 (Grade3) και TCCSUP (Grade4).

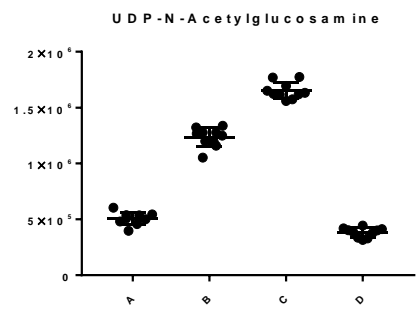
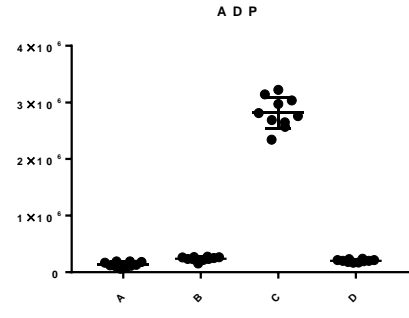
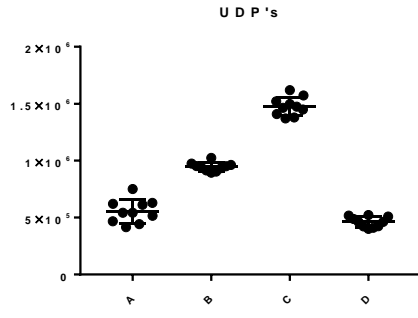
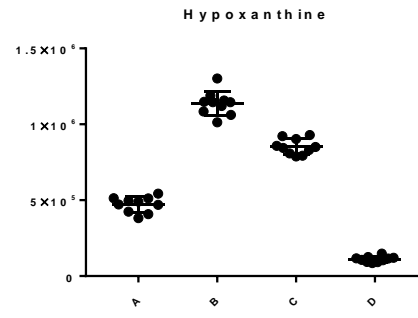
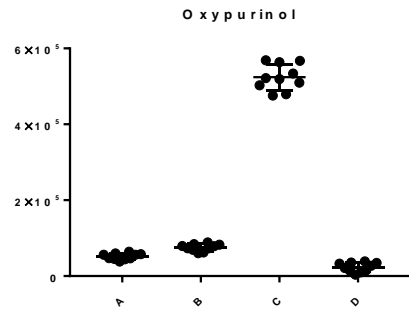
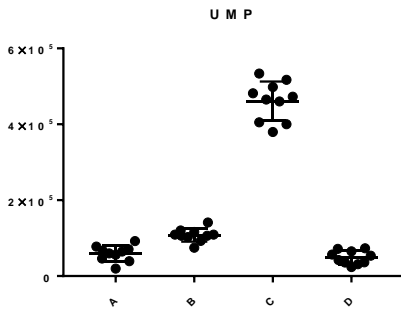
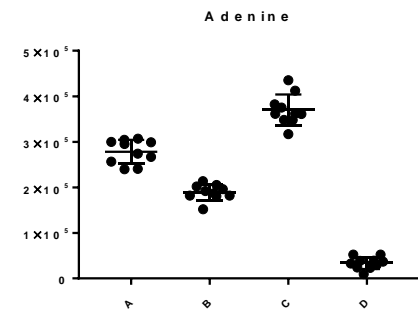
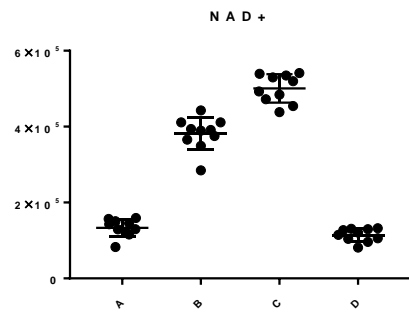
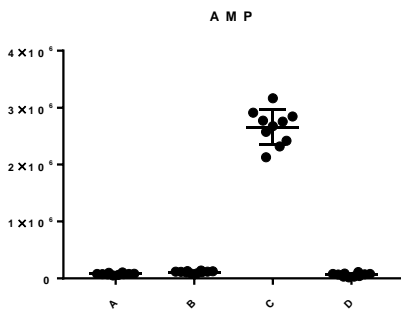
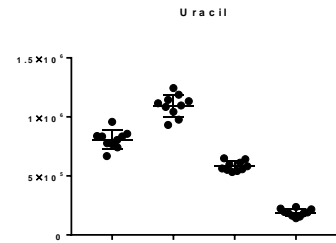
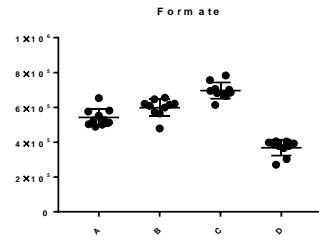
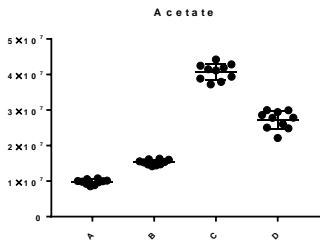
Πίνακας 11. Οι επιλεγμένοι μεταβολίτες που συμβάλλουν στον διαχωρισμό των δύο κυτταρικών σειρών.

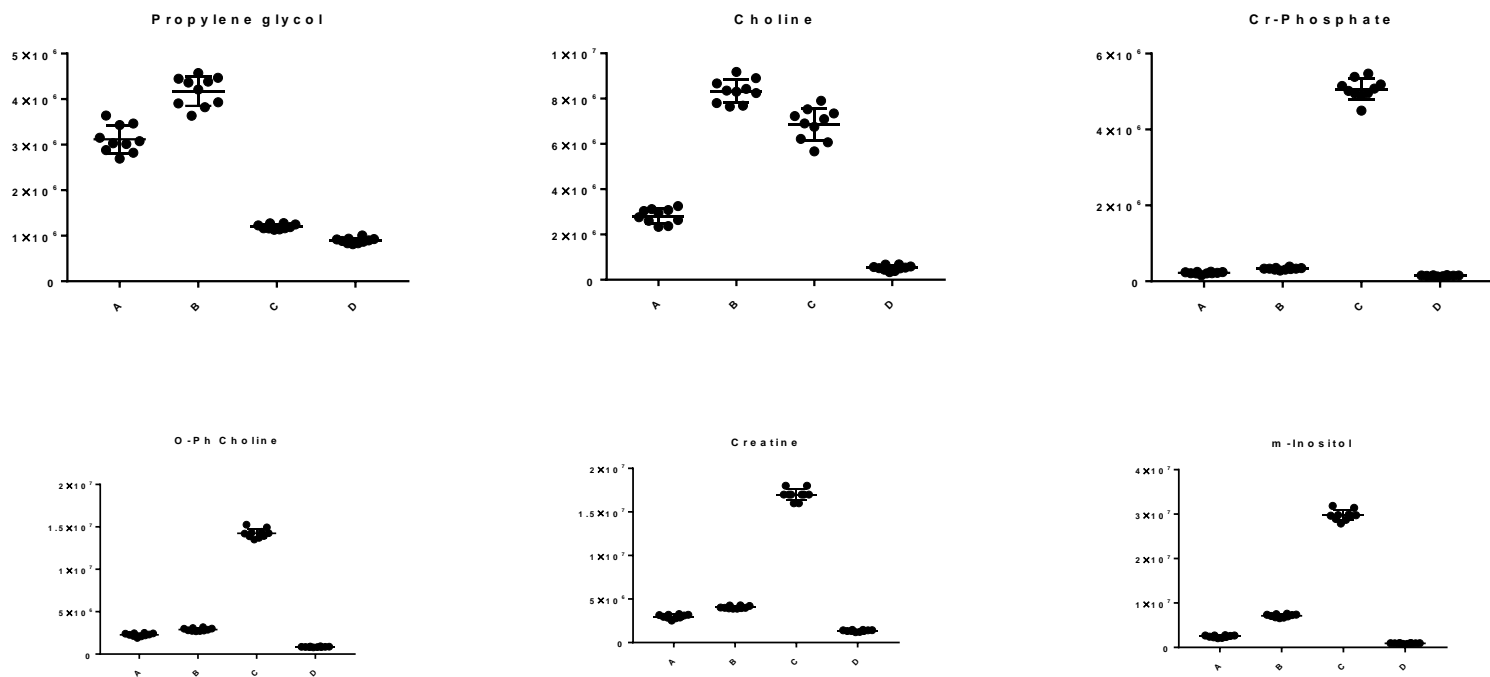
| Διάγραμμα S | Τιμή VIP | Metabolite |
|-------------|----------|-------------------------------|
| 4.17 | 1.99441 | O-Phosphocholine |
| 4.11 | 2.0672 | Lactate |
| 4.07 | 2.94281 | myo-Inositol |
| 3.93 | 2.16279 | Creatine |
| 3.79 | 3.05812 | Glutamate |
| 3.77 | 2.97636 | Glutamate |
| 3.75 | 2.32232 | Glutamate |
| 3.65 | 2.08561 | myo-Inositol |
| 3.63 | 2.97982 | myo-Inositol |
| 3.61 | 2.74401 | myo-Inositol |
| 3.59 | 2.16182 | Threonine |
| 3.57 | 1.96459 | Glycine |
| 3.55 | 3.12907 | myo-Inositol |
| 3.53 | 2.64766 | myo-Inositol |
| 3.29 | 2.44198 | myo-Inositol |
| 3.23 | 4.51112 | O-Phosphocholine |
| 3.05 | 2.10913 | Creatine + Creatine Phosphate |
| 2.37 | 2.37539 | Glutamate |
| 2.35 | 3.57687 | Glutamate |
| 2.17 | 2.18385 | Glutamate |
| 2.15 | 2.15265 | Glutamate |
| 2.13 | 2.24183 | Glutamate |
| 2.09 | 2.04073 | Glutamate |
| 2.07 | 2.71671 | Glutamate |
| 2.05 | 2.40978 | Glutamate |
| 1.33 | 4.75735 | Lactate |
| 4.17 | 1.99441 | O-Phosphocholine |

Συνέχεια έχουν τα θηκογράμματα (box plots) τα οποία χρησιμοποιούνται για να εξακριβώσουμε αν είναι στατιστικά σημαντικές οι διαφορές των μεταβολιτών μεταξύ των ομάδων¹¹⁶. Παρουσιάζονται λοιπόν τα θηκογράμματα για τους μεταβολίτες που έχουν εντοπιστεί στα δείγματα μας. Η όλη διαδικασία έγινε με το λογισμικό πρόγραμμα GraphPad Prisme 7.04.



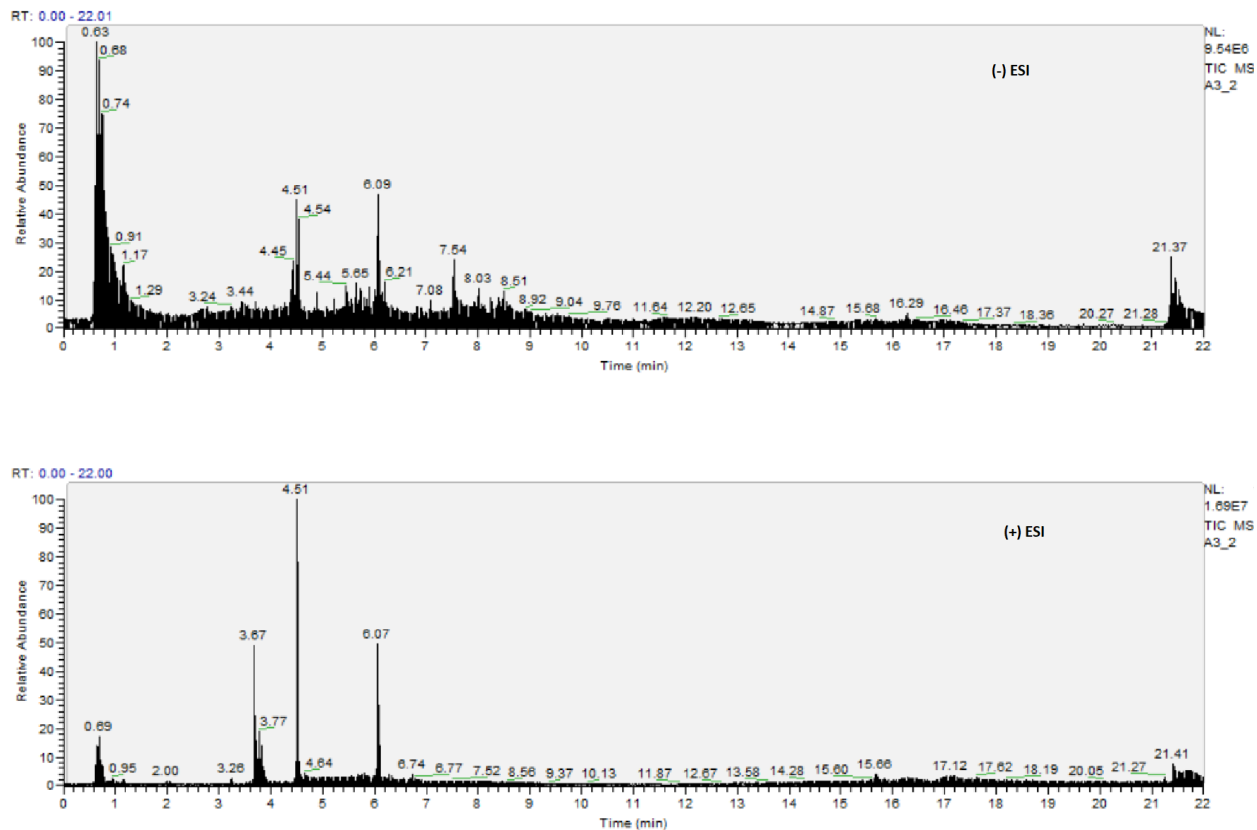






5.3 Φασματομετρία μάζας

Επιλέγοντας τις συνθήκες, που περιγράφηκαν παραπάνω (Ανάλυση με φασματομετρία μάζας), για τη χρωματογραφία /φασματομετρία μάζας, στην εικόνα 3 παρουσιάζονται χαρακτηριστικά χρωματογραφήματα της βασικής κορυφής (base peak intensity) από την κυτταρική σειρά T24 (grade 3, A3) σε αρνητικό και θετικό ιοντισμό. Οι μεταβολίτες διαχωρίστηκαν καλά λόγω της υψηλής διαχωριστικότητας που διαθέτουν οι στήλες με σωματίδια μικρής διαμέτρου ($1.9 \mu\text{m}$) –στήλες UHPLC.



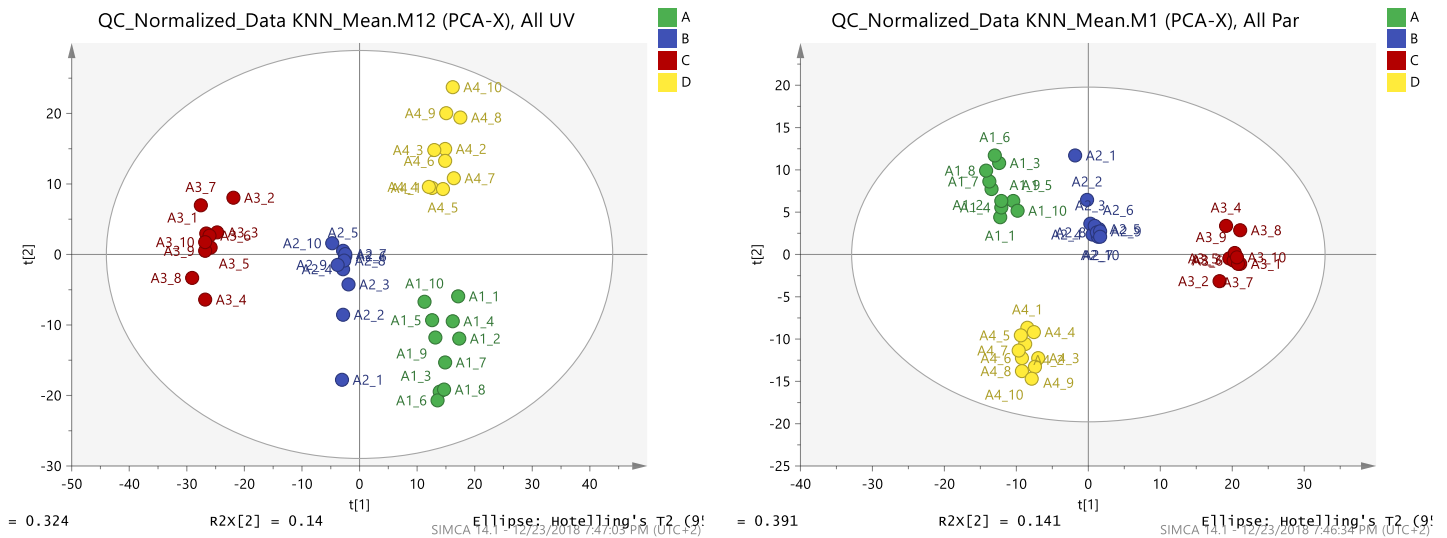
Εικόνα 3. Χρωματογραφήματα βασικής κορυφής ενός δείγματος (A3_2) από την κυτταρική σειρά T24 σε αρνητικό (επάνω) και θετικό ιοντισμό (κάτω).

Μετά την λήψη των φασμάτων, τα δεδομένα της μορφής *raw που προέκυψαν από το Orbitrap μετατράπηκαν στη μορφή *mzXML, η οποία είναι κατάλληλη για την μετέπειτα επεξεργασία με τη στατιστική γλώσσα $R^{117-119}$. Έτσι, προκειμένου να δημιουργηθεί ένας πίνακας που να περιέχει τον λόγο της μάζας προς το φορτίο (m/z), τους χρόνους κατακράτησης των μεταβολιτών και τις αντίστοιχες εντάσεις τους, χρησιμοποιήθηκε το πακέτο XCMS¹²⁰⁻¹²² της στατιστικής γλώσσας R, το οποίο εκτελεί την εύρεση των κορυφών (peak picking), τη διόρθωση του χρόνου κατακράτησης (retention time correction), την ομαδοποίηση των ίδιων κορυφών μεταξύ των δειγμάτων (peak grouping) και την ευθυγράμμιση των χρωματογραφικών κορυφών που παράγονται από δεδομένα μάζας (peak alignment). Όπως είναι γνωστό, η μη στοχευμένη μεταβολομική μελέτη δημιουργεί μεγάλο όγκο δεδομένων και το αποτέλεσμα της επεξεργασίας των δεδομένων εξαρτάται σε μεγάλο βαθμό από τις παραμέτρους του αλγόριθμου. Μη προσεκτική επιλογή παραμέτρων μπορεί να οδηγήσει σε μεροληπτικά αποτελέσματα. Για να βελτιστοποιηθεί η επιλογή παραμέτρων, χρησιμοποιήθηκε το πακέτο λογισμικού IPO^{123,124} προκειμένου να βελτιστοποιήσει τις παραμέτρους του XCMS. Στο πακέτο λογισμικού IPO, η επιλογή των παραμέτρων επιτυγχάνεται με πειραματικό σχεδιασμό (DoE - Design of Experiments).

Για την βελτιστοποίηση των παραμέτρων εξετάστηκε επίσης και το πακέτο arLCMS¹²⁵ μέσω της στατιστικής γλώσσας R. Η επεξεργασία των δεδομένων με το πρόγραμμα XCMS (χρησιμοποιώντας τις βελτιστοποιημένες παραμέτρους) έδωσε πίνακα με τις στήλες mzmin, mzmed, mzmax, rtmin, rtmed, rtmax, npeaks, fold, tstat, rvalue και τα εμβαδά ολοκλήρωσης των κορυφών για κάθε δείγμα. Στη συνέχεια ακολούθησε διαδικασία κανονικοποίησης (normalization) των δεδομένων μέσω της διαδικτυακής πλατφόρμας NOREVA και προέκυψε ο τελικός πίνακας. Για την κανονικοποίηση των δειγμάτων εξετάστηκε και το πακέτο statTarget¹²⁶ μέσω της στατιστικής γλώσσας R. Για την αξιολόγηση των αποτελεσμάτων που προκύπτουν, εξετάστηκε αν τα εσωτερικά πρότυπα εμφανίζονταν σε όλα τα δείγματα και αν οι περιοχές ολοκλήρωσής τους ήταν παρόμοιες με κριτήριο τη % τυπική απόκλιση. Παρατηρήθηκε ότι στον αρνητικό ιοντισμό ανιχνεύθηκαν και τα τρία εσωτερικά πρότυπα. Χρησιμοποιώντας το Microsoft Excel, δημιουργήθηκε πίνακας που περιείχε τις μεταβλητές «mzmed-rtmed» ως στήλες, την κωδικοποίηση του εκάστοτε δείγματος ως σειρές και τις αντίστοιχες περιοχές ολοκλήρωσης ως περιεχόμενο.

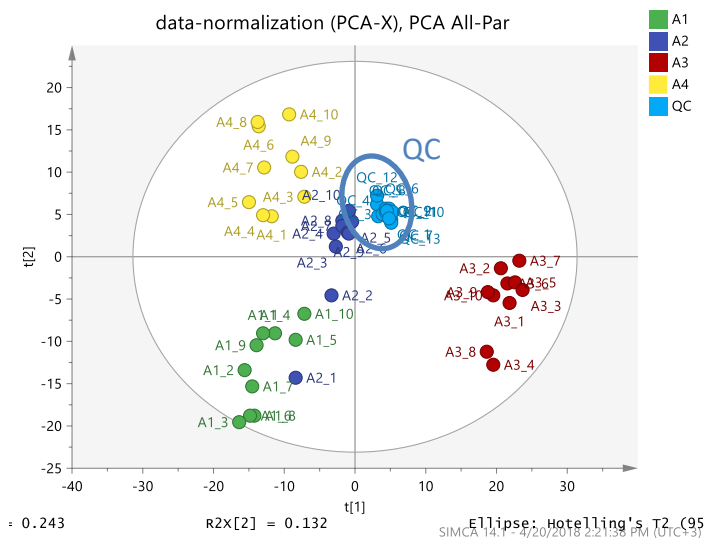
5.4 Χημειομετρική ανάλυση δεδομένων MS

Αφού πραγματοποιήθηκε η προ-επεξεργασία των δεδομένων, στη συνέχεια εφαρμόστηκε πολυμεταβλητή ανάλυση προκειμένου να ανακαλυφθούν οι διαφορές μεταξύ των κυτταρικών σειρών του καρκίνου (PCA, PLS-DA και OPLS-DA). Η μέθοδος PCA χρησιμοποιήθηκε αρχικά για να διερευνήσει πιθανή ομαδοποίηση των δειγμάτων και την ανίχνευση τυχόν ακραίων τιμών, αλλά και για να εξεταστεί η ομαδοποίηση των δειγμάτων ελέγχου. Για τη μέθοδο PCA, δοκιμάστηκαν διάφορες παράμετροι, όπως διαφορετικές μέθοδοι κλιμάκωσης π.χ. UV ή Pareto (που υπολογίζουν διάφορους συνδυασμούς των Q^2 και R^2), προκειμένου να εκτιμηθεί ο αριθμός των παραγόντων που περιγράφουν όσο το δυνατόν καλύτερα τη διακύμανση, χωρίς όμως να αυξάνεται υπερβολικά η πολυπλοκότητα του μοντέλου. Τα αποτελέσματα της κλιμάκωσης UV και Pareto στον αρνητικό ιοντισμό φαίνονται στο Διάγραμμα 13. Παρατηρούμε πως και στις δύο κλιμακώσεις έχουμε καλό διαχωρισμό των ομάδων. Η κλιμάκωση Pareto στον αρνητικό ιοντισμό δείχνει καλύτερη ομαδοποίηση των δειγμάτων σε απεικόνιση PCA σε σύγκριση με εκείνη της κλιμάκωσης UV. Η περιγραφική ικανότητα του μοντέλου με UV scaling (αριστερό διάγραμμα) είναι $R^2_X(\text{cum})=0.536$ και η προβλεπτική του ικανότητα είναι $Q^2(\text{cum})=0.448$. Με Pareto scaling (δεξιό διάγραμμα), η περιγραφική ικανότητα είναι $R^2_X(\text{cum})=0.615$ και η προβλεπτική του ικανότητα είναι $Q^2(\text{cum})=0.553$.



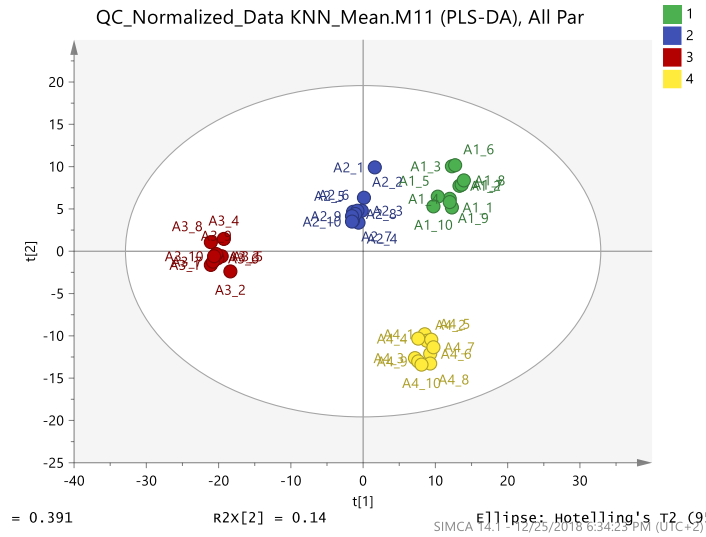
Διάγραμμα 13. Διαγράμματα αντικειμένων PCA για τις τέσσερις κυτταρικές σειρές με UV scaling αριστερά και Pareto scaling δεξιά σε αρνητικό ιοντισμό. Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, A), με μπλε χρώμα η κυτταρική σειρά RT112 (Grade2, B), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, C), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP (Grade4, D).

Όπως φαίνεται από το Διάγραμμα 14 με τη βοήθεια της προσέγγισης PCA, η συνολική ανάλυση UHPLC-ESI-HRMS θεωρείται αξιόπιστη για την στατιστική αξιολόγηση, καθώς τα δείγματα ελέγχου ποιότητας παρουσιάζουν ιδιαίτερα στενή ομαδοποίηση.

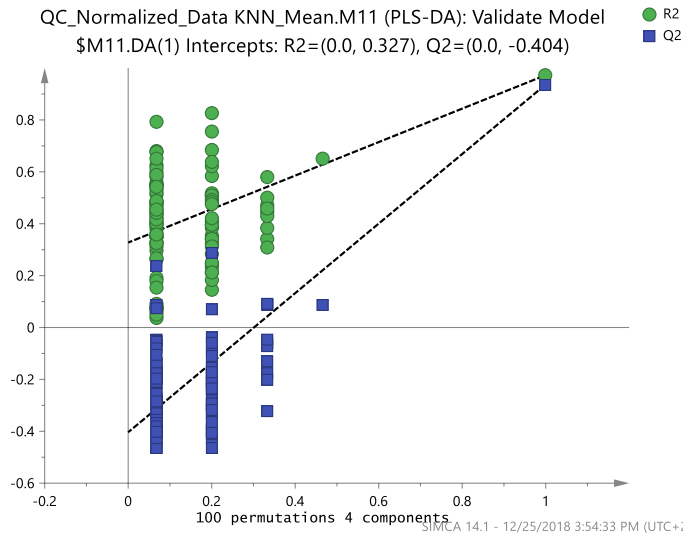


Διάγραμμα 14. Διάγραμμα αντικειμένων PCA για τις τέσσερις κутταρικές σειρές και τα δείγματα ελέγχου ποιότητας με Pareto scaling σε αρνητικό ιοντισμό. Με πράσινο χρώμα η κутταρική σειρά RT4 (Grade 1, A1), με μπλε χρώμα η κутταρική σειρά RT112 (Grade2, A2), με κόκκινο χρώμα η κутταρική σειρά T24 (Grade 3, A3), με κίτρινο χρώμα η κутταρική σειρά TCCSUP (Grade4, A4), με γαλάζιο χρώμα τα δείγματα ελέγχου ποιότητας (QC).

Στη συνέχεια εφαρμόστηκαν μέθοδοι επιβλεπόμενων τεχνικών πολυπαραμετρικής ανάλυσης, όπως PLS-DA και OPLS-DA. Στα μοντέλα αυτά, τα δείγματα ταξινομούνται σε προκαθορισμένες ομάδες, γεγονός που δημιουργεί τον κίνδυνο να δημιουργηθούν είτε υπέρ-προσαρμοσμένα μοντέλα, είτε μοντέλα προκατειλημμένης ομαδοποίησης. Στην προκειμένη περίπτωση, εφαρμόζοντας PLS-DA παρατηρήθηκε σαφής ομαδοποίηση μεταξύ των τεσσάρων ομάδων, όπως μπορεί να φανεί και στο αντίστοιχο διάγραμμα στον αρνητικό ιοντισμό (Διάγραμμα 15). Οι τιμές σε αυτήν την περίπτωση είναι $R^2(\text{cum})=0.661$ και $Q^2(\text{cum})=0.962$. Μετά από την πολυπαραγοντική ανάλυση, ακολουθούν διαδικασίες επικύρωσης προκειμένου να διασφαλιστεί η εγκυρότητα του μοντέλου και να επαληθευτεί η ομαδοποίηση των δειγμάτων. Επιτρέποντας 100 τυχαίες μεταθέσεις οι τομές ήταν $R^2=(0.0, 0.327)$, $Q^2=(0.0, -0.404)$ όπως φαίνεται και από το διάγραμμα 16. Το μοντέλο μπορεί να θεωρηθεί έγκυρο.



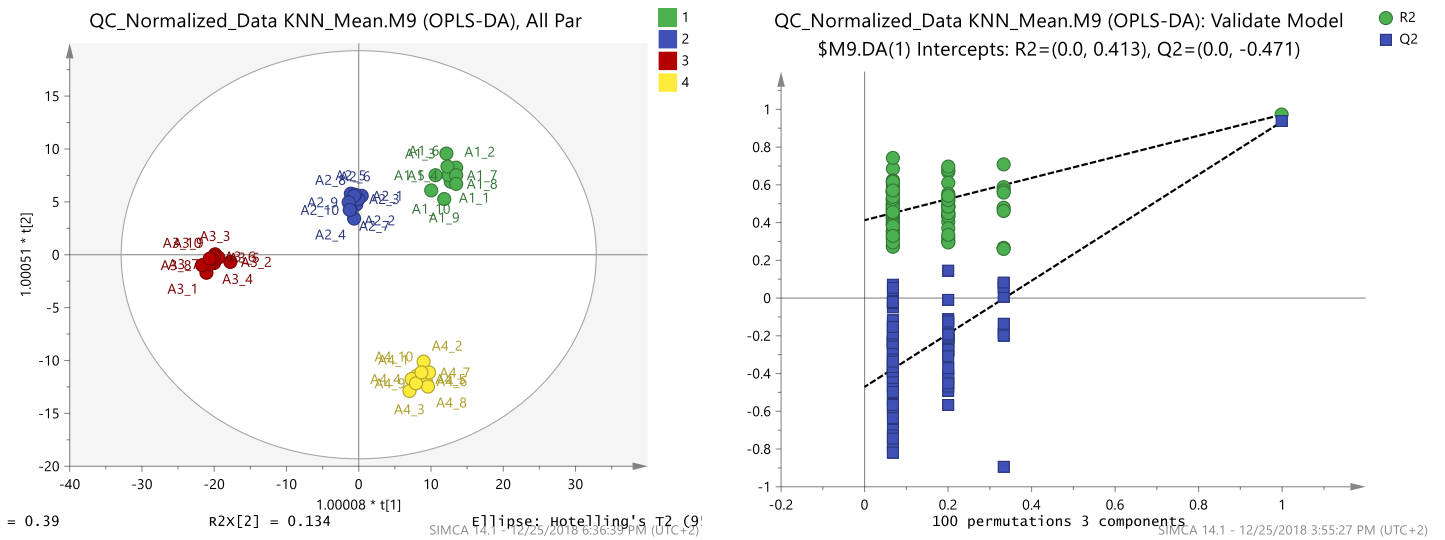
Διάγραμμα 15. Διάγραμμα αντικειμένων PLS-DA για τις τέσσερις κυτταρικές σειρές με Pareto scaling σε αρνητικό ιοντισμό. Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, 1), με μπλε χρώμα η κυτταρική σειρά RT112 (Grade2, 2), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, 3), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP (Grade4, 4).



Διάγραμμα 16. Μεταθετικός έλεγχος με 100 τυχαίες μεταθέσεις επιτρεπόμενες για το μοντέλο PLS-DA με Pareto scaling.

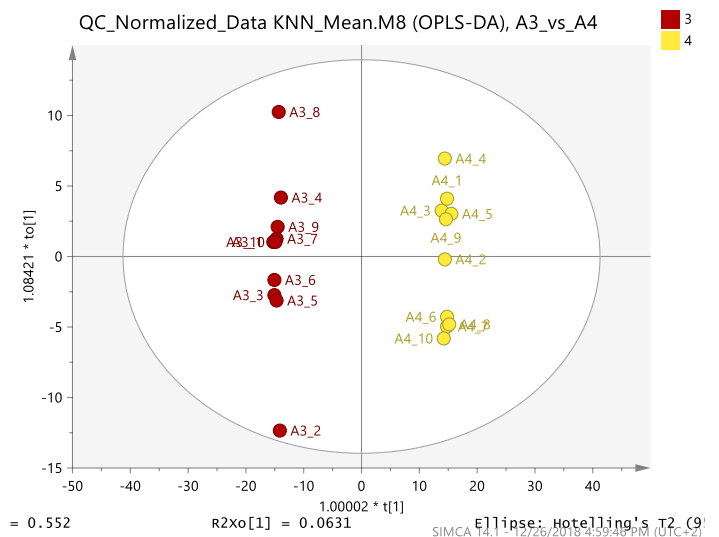
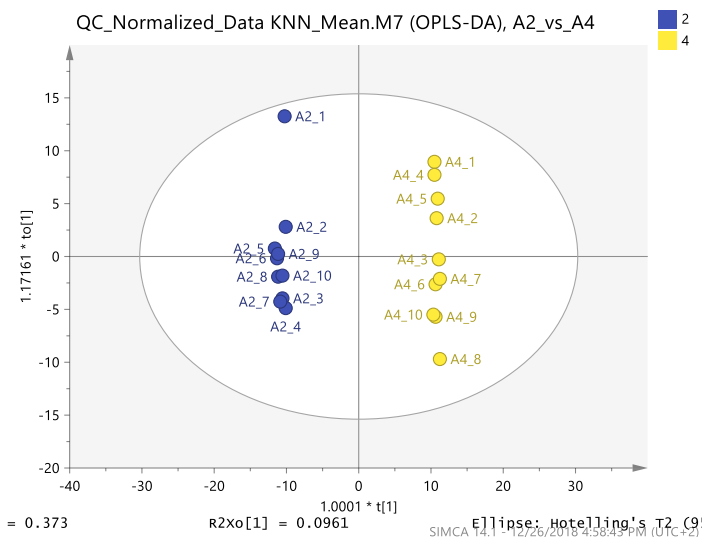
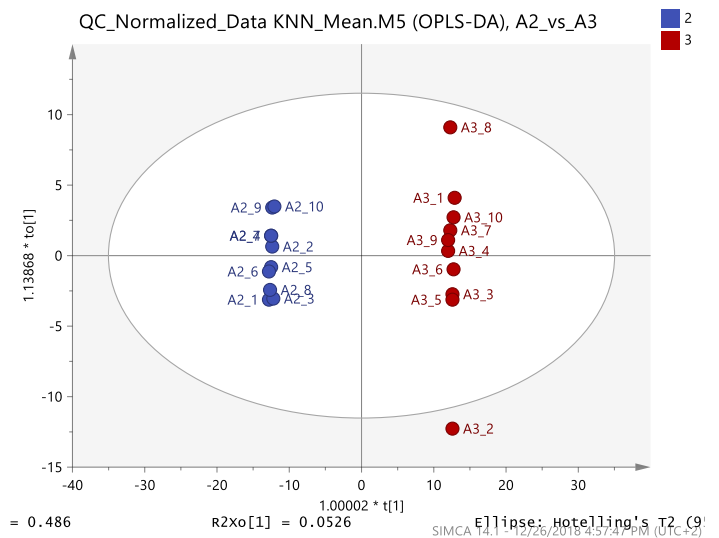
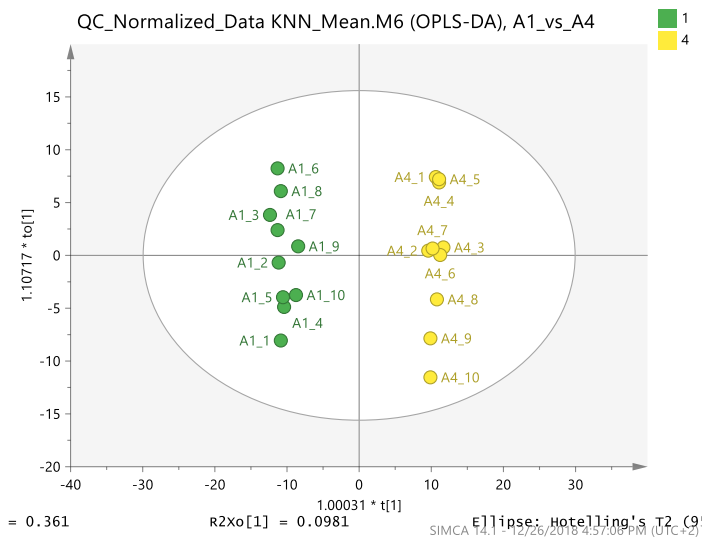
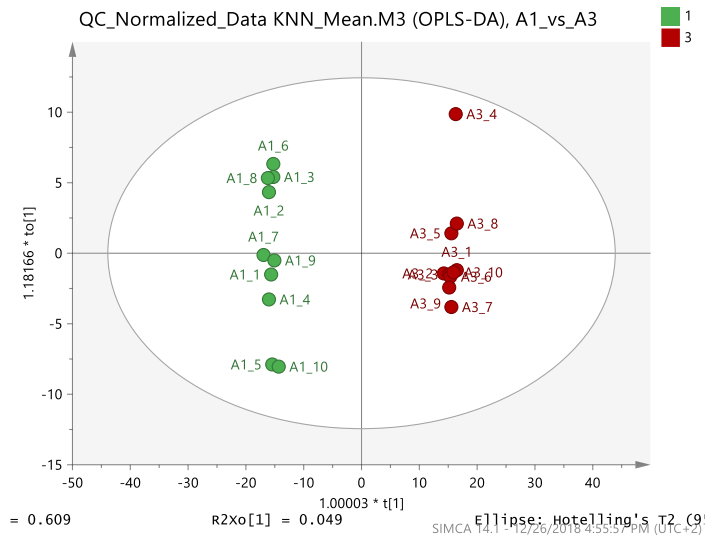
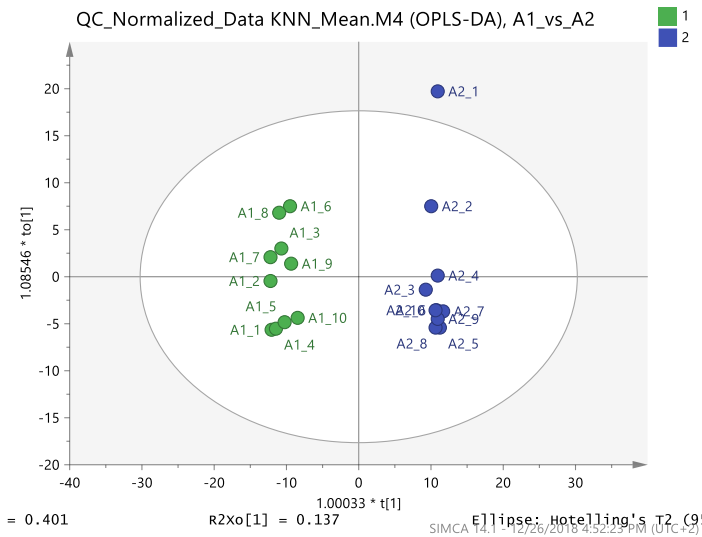
Τέλος, σε αυτά τα δείγματα πραγματοποιήθηκε στατιστική ανάλυση με την επίσης επιβλεπόμενη μέθοδο OPLS-DA με Pareto scaling στον αρνητικό ιοντισμό (διάγραμμα 17, αριστερά). Οι τιμές σε αυτήν την περίπτωση είναι $R^2(\text{cum})=0.661$ και $Q^2(\text{cum})=0.959$. Όπως μπορούμε να διαπιστώσουμε η επιβλεπόμενη μέθοδος PLS-DA και η OPLS-DA παρουσιάζουν την ίδια εικόνα καθώς ο διαχωρισμός μεταξύ των ομάδων

ήταν ήδη πολύ καλός από την μέθοδο PLS-DA. Το μοντέλο επίσης μπορεί να θεωρηθεί έγκυρο (διάγραμμα 17, δεξιά).

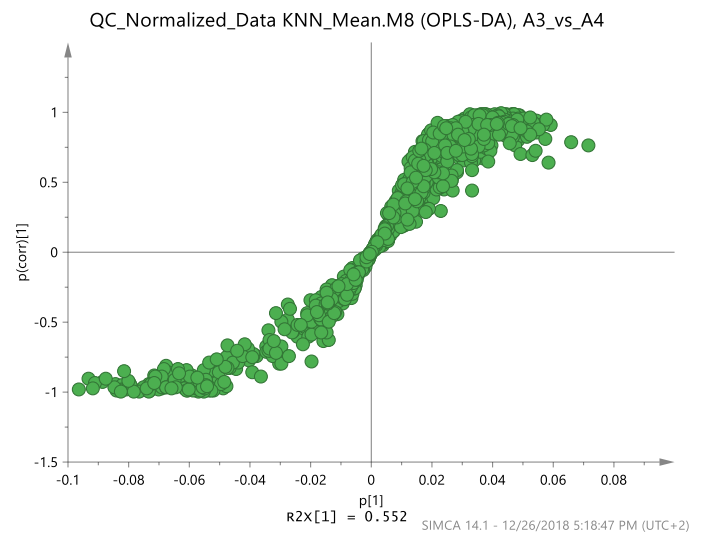
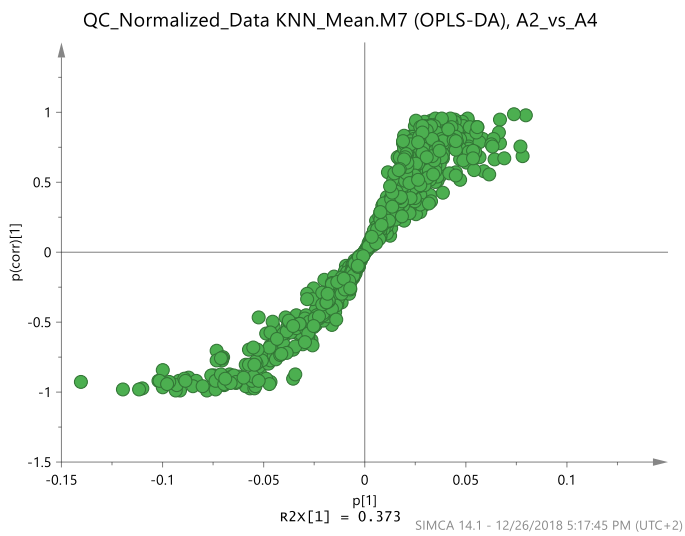
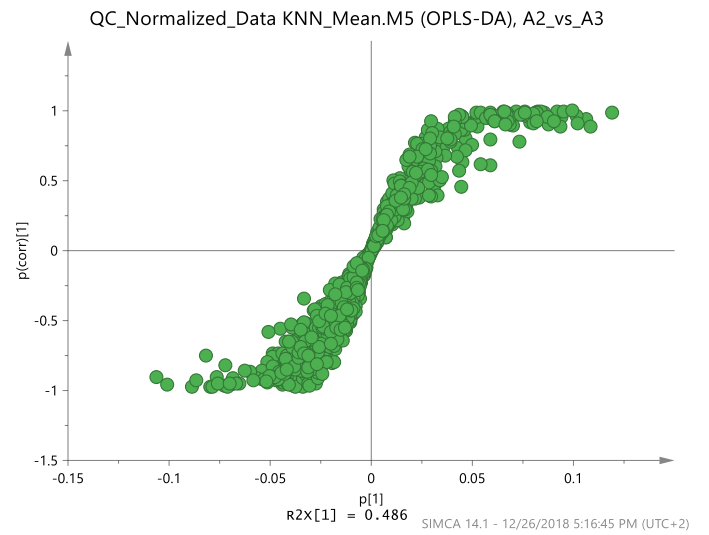
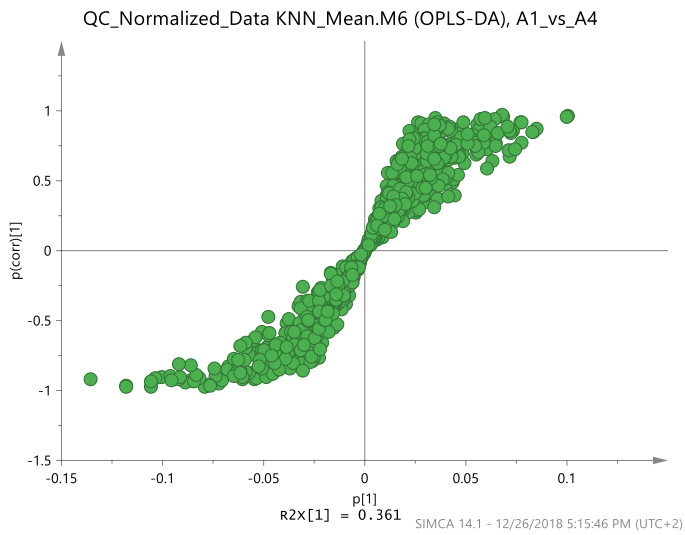
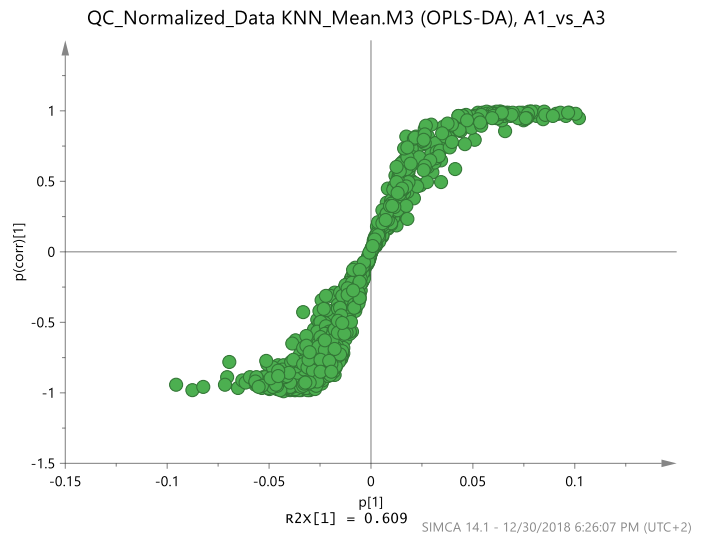
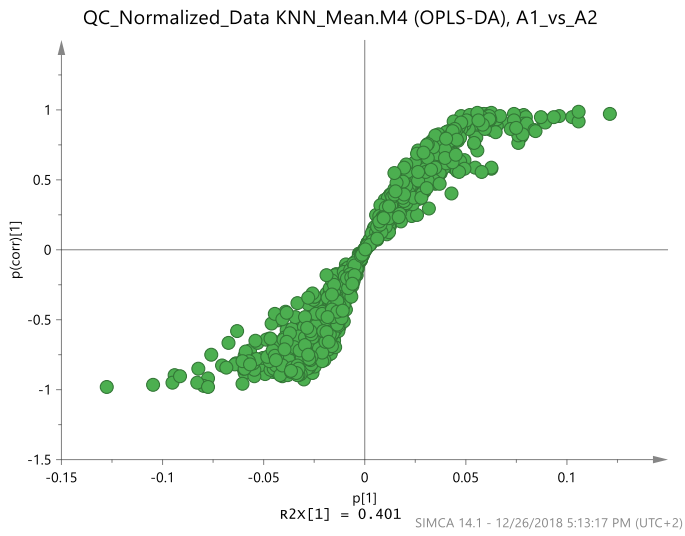


Διάγραμμα 17. Αριστερά: Διάγραμμα αντικειμένων OPLS-DA για τις τέσσερις κυτταρικές σειρές με Pareto scaling σε αρνητικό ιοντισμό. Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, 1), με μπλε χρώμα η κυτταρική σειρά RT112 (Grade 2, 2), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, 3), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP (Grade 4, 4). Δεξιά: Μεταθετικός έλεγχος με 100 τυχαίες μεταθέσεις επιτρεπόμενες για το μοντέλο OPLS-DA με Pareto scaling.

Στη συνέχεια έγινε προσπάθεια να εντοπισθούν τα χαρακτηριστικά που συμβάλουν στον διαχωρισμό των ομάδων. Για αυτόν τον λόγο υπολογίστηκαν οι τιμές VIP από το μοντέλο PLS-DA (Παράρτημα 2) και κατασκευάστηκαν τα διαγράμματα S (S-plot διάγραμμα 19) από τα επιμέρους μοντέλα OPLS-DA (διάγραμμα 18) και για τις έξι συγκρίσεις (ανά δύο οι ομάδες) για τον αρνητικό ιοντισμό.

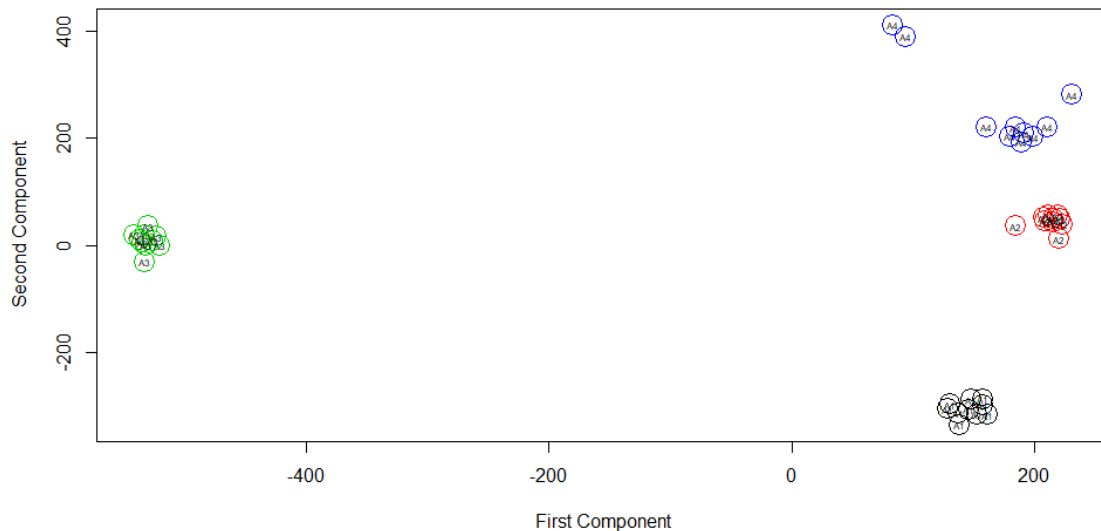


Διάγραμμα 18. Διαγράμματα αντικειμένων OPLS-DA για τις τέσσερις κυτταρικές σειρές (Pareto scaling) σε αρνητικό ιοντισμό. Με πράσινο χρώμα η κυτταρική σειρά RT4 (Grade 1, 1), με μπλε χρώμα η κυτταρική σειρά RT12 (Grade 2, 2), με κόκκινο χρώμα η κυτταρική σειρά T24 (Grade 3, 3), με κίτρινο χρώμα η κυτταρική σειρά TCCSUP (Grade 4, 4).



Διάγραμμα 19. Διαγράμματα Σατό τα αντίστοιχα μοντέλα OPLS-DAγια τον αρνητικό ιοντισμό.

Παράλληλα πραγματοποιήθηκε η δοκιμασία PLS-DA μέσω του στατιστικού πακέτου KODAMA της γλώσσας προγραμματισμού R, για τον αρνητικό ιοντισμό. Μέσω του στατιστικού πακέτου αυτού εφαρμόζεται η δοκιμασία PLS-DA χωρίς όμως να γίνει αρχική ομαδοποίηση των δεδομένων. Αποτελεί στην ουσία μία μη επιβλεπόμενη μέθοδο με την δυνατότητα όμως εναλλακτικά να ορισθούν οι ομάδες και να μετατραπεί σε επιβλεπόμενη μέθοδο. Το διάγραμμα PLS-DA από την δοκιμασία αυτή παρουσιάζεται παρακάτω όπου παρατηρείται καλός διαχωρισμός των ομάδων.



Διάγραμμα 20. Διάγραμμα PLS-DA μέσω του KODAMA για τον αρνητικό ιοντισμό. Με μαύρο χρώμα η κυτταρική σειρά RT4 (Grade 1, A1), με κόκκινο χρώμα η κυτταρική σειρά RT112 (Grade 2, A2), με πράσινο χρώμα η κυτταρική σειρά T24 (Grade 3, A3), με μπλε χρώμα η κυτταρική σειρά TCCSUP (Grade 4, A4)

Με βάση την ανάλυση KODAMA προέκυψε πίνακας με τα φορτία (loadings) όπου η κάθε μεταβλητή (μεταβολίτης) είχε και τη δική του βαθμολογία. Στη συνέχεια προέκυψε και ένας επιπλέον πίνακας που εμφανίζει πάλι τα φορτία (loadings) ύστερα από τον υπολογισμό του λογαρίθμου της τιμής P από τη δοκιμασία Kruskal-Wallis Rank Sums¹²⁷⁻¹³⁰. Οι πίνακες αυτοί εμφανίζονται στο Παράρτημα 2.

Στη συνέχεια παρουσιάζονται πίνακες με τη σχετική θέση που έχει ο κάθε μεταβολίτης όπως προέκυψε από την κάθε δοκιμασία. Στην αριστερή και πρώτη στήλη έχουμε τον μεταβολίτη σε μορφή m/z_rt (rt: retention time, χρόνος κατακράτησης) από τα διαγράμματα S (Διάγραμμα 19) στην επόμενη την τιμή VIP όπως προκύπτει από το μοντέλο PLS-DA. Στις δύο επόμενες έχουμε τη σχετική θέση που βρίσκεται ο κάθε μεταβολίτης μέσω του KODAMA με τη δοκιμασία Kruskal-Wallis Rank Sums και χωρίς αυτή τη δοκιμασία.

Πίνακας 12. Συγκριτικός πίνακας με τις σχετικές θέσεις του κάθε μεταβολίτη από τη σύγκριση των κυτταρικών σειρών RT4 (Grade 1, A1) και RT112 (Grade 2, A2)

| Διάγραμμα S | Τιμή VIP | Kodama (Kruskal) | Kodama |
|--------------------|-----------------|-------------------------|---------------|
| 290.1608_223.58 | 2.34477 | 17 | 51 |
| 555.2096_98.58 | 2.2756 | 48 | 139 |
| 611.145_57.53 | 2.09075 | 61 | 125 |
| 664.1175_61.66 | 2.03038 | 33 | 34 |
| 662.1021_47.6 | 1.72181 | 27 | 39 |
| 633.1312_70.59 | 1.66953 | 8 | 2 |
| 191.0203_53.79 | 2.20072 | 171 | 183 |
| 535.1547_70.34 | 1.66963 | 30 | 4 |
| 283.0685_88.02 | 2.99403 | 86 | 470 |
| 283.0688_72.4 | 2.8322 | 125 | 296 |
| 337.0447_59.32 | 2.19644 | 277 | 8 |
| 248.0405_51.23 | 1.72379 | 313 | 211 |
| 134.0479_52.08 | 1.80036 | 311 | 67 |
| 321.0882_343.78 | 2.97446 | 294 | 1 |
| 800.1436_221.72 | 2.44252 | 100 | 146 |
| 169.9923_212.81 | 1.89053 | 170 | 39 |
| 550.7642_46.49 | 1.63437 | 328 | 7 |
| 165.0563_252.37 | 2.28034 | 125 | 74 |
| 448.8312_45.85 | 1.99098 | 451 | 12 |
| 323.0287_46.17 | 2.7603 | 509 | 157 |
| 323.0289_51.47 | 2.91739 | 75 | 107 |

Πίνακας 13. Συγκριτικός πίνακας με τις σχετικές θέσεις του κάθε μεταβολίτη από τη σύγκριση των κυτταρικών σειρών RT4 (Grade 1, A1) και T24 (Grade 3, A3)

| Διάγραμμα S | Τιμή VIP | Kodama (Kruskal) | Kodama |
|--------------------|-----------------|-------------------------|---------------|
| 664.1175_61.66 | 2.03038 | 33 | 32 |
| 426.0226_50.35 | 2.04466 | 67 | 61 |
| 382.5499_175.7 | 1.93497 | 215 | 10 |
| 346.0562_46.89 | 2.01075 | 81 | 66 |
| 219.1068_161.3 | 1.90319 | 161 | 14 |
| 320.0627_61.08 | 2.10791 | 305 | 63 |
| 490.7923_40.99 | 1.34841 | 244 | 26 |
| 321.0882_343.78 | 2.97446 | 294 | 1 |
| 169.9923_212.81 | 1.89053 | 170 | 39 |
| 599.3193_551.36 | 2.30231 | 104 | 11 |
| 143.0469_53.53 | 1.66695 | 398 | 205 |
| 287.0885_124.66 | 2.5797 | 159 | 75 |

Πίνακας 14. Συγκριτικός πίνακας με τις σχετικές θέσεις του κάθε μεταβολίτη από τη σύγκριση των κυτταρικών σειρών RT4 (Grade 1, A1) και TCCSUP (Grade 4, A4)

| Διάγραμμα S | Τιμή VIP | Kodama (Kruskal) | Kodama |
|--------------------|-----------------|-------------------------|---------------|
| 611.145_57.53 | 2.09075 | 61 | 123 |
| 612.1482_57.92 | 1.78474 | 57 | 120 |
| 709.1215_57.87 | 1.84253 | 90 | 140 |
| 426.0226_50.35 | 2.04466 | 67 | 61 |
| 346.0562_46.89 | 2.01075 | 81 | 66 |
| 321.0882_343.78 | 2.97446 | 294 | 1 |
| 165.0563_252.37 | 2.28034 | 125 | 74 |
| 455.0975_214.26 | 2.43398 | 88 | 769 |
| 151.0269_54.85 | 3.16576 | 4 | 167 |
| 323.0289_51.47 | 2.91739 | 75 | 107 |
| 287.0885_124.66 | 2.5797 | 159 | 75 |
| 181.051_189.69 | 2.66495 | 5 | 115 |

Πίνακας 15. Συγκριτικός πίνακας με τις σχετικές θέσεις του κάθε μεταβολίτη από τη σύγκριση των κυτταρικών σειρών RT112 (Grade 2, A2) και T24 (Grade 3, A3)

| Διάγραμμα S | Τιμή VIP | Kodama (Kruskal) | Kodama |
|--------------------|-----------------|-------------------------|---------------|
| 382.5499_175.7 | 1.93497 | 215 | 4 |
| 346.0562_46.89 | 2.01075 | 81 | 66 |
| 218.1035_160.78 | 2.01585 | 320 | 22 |
| 320.0627_61.08 | 2.10791 | 305 | 63 |
| 316.08_160.85 | 2.01068 | 312 | 6 |
| 174.0414_48.61 | 2.44922 | 258 | 124 |
| 283.0685_88.02 | 2.99403 | 86 | 470 |
| 283.0688_72.4 | 2.8322 | 125 | 296 |
| 599.3193_551.36 | 2.30231 | 104 | 11 |
| 143.0469_53.53 | 1.66695 | 398 | 205 |
| 287.0885_124.66 | 2.5797 | 159 | 75 |

Πίνακας 16. Συγκριτικός πίνακας με τις σχετικές θέσεις του κάθε μεταβολίτη από τη σύγκριση των κυτταρικών σειρών RT112 (Grade 2, A2) και TCCSUP (Grade 4, A4)

| Διάγραμμα S | Τιμή VIP | Kodama (Kruskal) | Kodama |
|--------------------|-----------------|-------------------------|---------------|
| 364.0668_69.66 | 1.66643 | 319 | 649 |
| 312.0954_69.67 | 1.29876 | 431 | 394 |
| 958.5633_1028.81 | 2.12413 | 501 | 605 |
| 957.5613_977.94 | 1.64707 | 387 | 414 |
| 283.0685_88.02 | 2.99403 | 86 | 470 |
| 283.0688_72.4 | 2.8322 | 125 | 296 |
| 455.0975_214.26 | 2.43398 | 88 | 769 |
| 151.0269_54.85 | 3.16576 | 4 | 167 |

Πίνακας 17. Συγκριτικός πίνακας με τις σχετικές θέσεις του κάθε μεταβολίτη από τη σύγκριση των κυτταρικών σειρών T24 (Grade 3, A3) και TCCSUP (Grade 4, A4)

| Διάγραμμα S | Τιμή VIP | Kodama (Kruskal) | Kodama |
|-----------------|----------|------------------|--------|
| 312.0954_69.67 | 1.29876 | 431 | 394 |
| 490.7923_40.99 | 1.34841 | 244 | 26 |
| 539.7413_41.2 | 1.15541 | 139 | 60 |
| 599.3193_551.36 | 2.30231 | 104 | 11 |
| 613.7772_47.2 | 1.31551 | 227 | 376 |
| 143.0469_53.53 | 1.66695 | 398 | 205 |
| 535.5855_186.63 | 1.9396 | 9 | 42 |
| 664.1175_61.66 | 2.03038 | 33 | 32 |
| 117.0201_67.15 | 1.75797 | 1 | 97 |
| 188.0931_221.01 | 1.75797 | 39 | 21 |
| 320.0924_66.58 | 1.80195 | 188 | 104 |
| 535.1547_70.34 | 1.66963 | 30 | 4 |
| 146.0464_44.1 | 1.75632 | 42 | 59 |
| 202.1087_251.11 | 1.62484 | 38 | 24 |
| 259.13_202.08 | 1.72705 | 65 | 13 |

Προκειμένου να ελεγχθεί η αξιοπιστία του κάθε μοντέλου ακολουθήθηκε η εξής διαδικασία: Κατασκευάστηκαν καμπύλες ROC και υπολογίστηκε το AUC για τους πρώτους δέκα μεταβολίτες και για τις έξι συγκρίσεις μεταξύ των ομάδων. Η διαδικασία αυτή έγινε και για τα τρία μοντέλα. Στη συνέχεια υπολογίστηκε το άθροισμα των AUC για τον κάθε μεταβολίτη από τις έξι συγκρίσεις. Δεδομένου ότι το μέγιστο AUC είναι 1, στην περίπτωση των έξι συγκρίσεων θα είναι 6. Τέλος το μέγιστο συνολικό άθροισμα των AUC από το κάθε μοντέλο εφόσον ελήφθησαν υπόψη οι δέκα πρώτοι μεταβολίτες θα είναι 60. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα (πίνακας 18).

Πίνακας 18. Άθροισμα AUC από το κάθε μοντέλο για τους δέκα πρώτους μεταβολίτες σε αρνητικό ιοντισμό

| Metabolite | SUM AUC SIMCA | SUM AUC KODAMA (Kruskal) | SUM AUC KODAMA |
|------------------|---------------|--------------------------|----------------|
| 1 ^{ος} | 5.98 | 5.23 | 6.00 |
| 2 ^{ος} | 5.71 | 5.91 | 5.85 |
| 3 ^{ος} | 5.23 | 5.64 | 5.82 |
| 4 ^{ος} | 5.54 | 5.71 | 5.98 |
| 5 ^{ος} | 5.61 | 4.16 | 6.00 |
| 6 ^{ος} | 5.09 | 4.89 | 5.85 |
| 7 ^{ος} | 6.00 | 5.32 | 5.93 |
| 8 ^{ος} | 5.59 | 5.02 | 5.91 |
| 9 ^{ος} | 4.92 | 5.10 | 5.90 |
| 10 ^{ος} | 5.50 | 5.22 | 5.94 |
| Total SUM | 55.17 | 52.2 | 59.18 |

Στον παρακάτω πίνακα παρουσιάζονται οι δέκα πρώτοι μεταβολίτες όπως αυτοί προέκυψαν από το κάθε μοντέλο με την κωδική μορφή m/z_rt (πίνακας 19).

Πίνακας 19. Οι δέκα πρώτοι μεταβολίτες όπως αυτοί προέκυψαν από το κάθε μοντέλο σε αρνητικό ιοντισμό

| Metabolite | SIMCA | KODAMA (Kruskal) | KODAMA |
|------------------|-----------------|------------------|-----------------|
| 1 ^{ος} | 151.0269_54.85 | 117.0201_67.15 | 321.0882_343.78 |
| 2 ^{ος} | 283.0685_88.02 | 353.049_264.55 | 633.1312_70.59 |
| 3 ^{ος} | 321.0882_343.78 | 164.0721_122.85 | 278.044_59.88 |
| 4 ^{ος} | 323.0289_51.47 | 151.0269_54.85 | 535.1547_70.34 |
| 5 ^{ος} | 283.0688_72.4 | 181.051_189.69 | 957.561_1028.09 |
| 6 ^{ος} | 323.0287_46.17 | 354.0523_264.96 | 316.08_160.85 |
| 7 ^{ος} | 181.051_189.69 | 308.1158_267.22 | 550.7642_46.49 |
| 8 ^{ος} | 287.0885_124.66 | 633.1312_70.59 | 337.0447_59.32 |
| 9 ^{ος} | 174.0414_48.61 | 535.5855_186.63 | 146.0826_161.39 |
| 10 ^{ος} | 800.1436_221.72 | 307.1123_267.5 | 382.5499_175.7 |

Τέλος πραγματοποιήθηκε τόσο για τον αρνητικό όσο και για τον θετικό ιοντισμό μια στοχευμένη ανάλυση για συγκεκριμένους μεταβολίτες η λίστα των οποίων βρίσκεται στο τέλος της εργασίας στο Παράρτημα 2. Στον πίνακα που παρουσιάζεται παρακάτω φαίνονται τα αποτελέσματα τόσο στον αρνητικό όσο και στον θετικό ιοντισμό (πίνακες 20 και 21). Για να προκύψει ο παρακάτω πίνακας χρησιμοποιήθηκαν τα προγράμματα Xcalibur¹³¹ και MZmine¹³²⁻¹³⁴. Στην τελευταία στήλη του πίνακα αναγράφεται το επίπεδο εμπιστοσύνης το οποίο θα πρέπει να είναι >4. Ως αποτέλεσμα, μόνο η ινοσίνη μπορεί να θεωρηθεί ότι βρέθηκε με αξιοπιστία στον αρνητικό ιοντισμό. Να σημειωθεί ότι αν βρεθεί το m/z είναι 2 βαθμοί, αν βρεθεί κάποιο προϊόν προσθήκης (adduct) είναι επίσης 2 βαθμοί, για κάθε προϊόν προσθήκης (adduct). Επίσης, αν βρεθεί κάποιο θραύσμα MS/MS δίνει 1.5 βαθμοί, ενώ αν βρεθεί ισοτοπική αναλογία δίνει 0.5 βαθμοί¹³⁵.

Πίνακας 20. Στοχευμένη ανάλυση σε αρνητικό ιοντισμό

| Substance name | m/z | Rt | Formula | Adducts | | | | | | | MS/MS | | isotope pattern | trust level |
|-------------------------|----------|------|---|---------|---|----|-----------------|------------------|----|-------|-------|--------------|-----------------|-------------|
| | | | | H | K | Na | NH ₄ | FA (Formic acid) | Cl | dimer | dda | deconvoluted | | |
| inosine | 267.0739 | 1.18 | C ₁₀ H ₁₂ N ₄ O ₅ | 2 | | | | 2 | 2 | 2 | 1.5 | | 0.5 | 10 |
| glyceric acid | 142.9757 | 1.51 | C ₃ H ₆ O ₄ | | 2 | | | | | | | | | 2 |
| uridine | 243.0648 | 0.92 | C ₉ H ₁₂ N ₂ O ₆ | | | | | | | | 1.5 | | | 1.5 |
| S-adenosyl-homocysteine | 383.1149 | 8.02 | C ₁₄ H ₂₀ N ₆ O ₅ S | 2 | | | | | | | | | | 2 |
| allantoic acid | 351.1023 | 4.46 | C ₄ H ₈ N ₄ O ₄ | | | | | | | 2 | | | | 2 |

Πίνακας 21. Στοχευμένη ανάλυση σε θετικό ιοντισμό

| Substance name | m/z | Rt | Formula | Adducts | | | | | | MS/MS | | | isotope pattern | trust level |
|----------------|--------|-------|---|---------|---|----|-----------------|------------------|----|-------|-----|--------------|-----------------|-------------|
| | | | | H | K | Na | NH ₄ | FA (Formic acid) | Cl | dimer | dda | deconvoluted | | |
| Homocysteine | 211.95 | 0.6 | C ₄ H ₉ NO ₂ S | | 2 | | | | | | | | | 2 |
| Succinate | 156.98 | 10.02 | C ₄ H ₆ O ₄ | | 2 | | | | | | | | | 2 |

ΚΕΦΑΛΑΙΟ 6

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΣΥΖΗΤΗΣΗ

Ο καρκίνος της ουροδόχου κύστης είναι η πιο διαδεδομένη κακοήθεια από τους ουρολογικούς τύπους καρκίνου¹³⁶. Αποτελεί επίσης έναν κοινά εμφανιζόμενο τύπο καρκίνου παγκοσμίως με πάνω από 330.000 κρούσματα και 130.000 θανάτους ανά χρόνο¹³⁷. Η έγκαιρη διάγνωση είναι επομένως μεγάλης σημασίας καθώς θα μειωθεί η θνησιμότητα και θα βελτιωθεί το επίπεδο ζωής των ασθενών. Όπως αναφέραμε η μεταβολομική είναι μια δυναμική τεχνική που μπορεί να διαγνώσει έγκαιρα τον καρκίνο της ουροδόχου κύστης. Έχει τη δυνατότητα να περιγράψει ποσοτικά όλες τις μικρής μοριακής μάζας ενώσεις που βρίσκονται σε βιολογικά δείγματα¹³⁸.

Σκοπός της παρούσας εργασίας ήταν η μη στοχευμένη μεταβολομική μελέτη με χρήση φασματοσκοπίας NMR και φασματομετρίας μάζας σε τέσσερις κυτταρικές σειρές με καρκίνο ουροδόχου κύστης με σκοπό τη μελέτη του μεταβολώματος κατά την έναρξη και κατά τη διάρκεια εξέλιξης της νόσου. Επίσης, βελτιστοποιήθηκαν και αναπτύχθηκαν διαφορετικές μεθοδολογίες για την προετοιμασία του δείγματος. Για τον σκοπό αυτό μελετήθηκαν τέσσερα διαφορετικά πρωτόκολλα εκχύλισης. Τέλος, έγινε εκτίμηση νέων χημειομετρικών τεχνικών για τη στατιστική ανάλυση των δειγμάτων: μελετήθηκε ο αλγόριθμος KODAMA.

Από τη φασματοσκοπία NMR φάνηκε πως υπάρχει μικρή αλλά σημαντική διαφορά συγκρίνοντας την κυτταρική σειρά RT112 (grade 2) με την κυτταρική σειρά RT4 (grade 1). Στην κυτταρική σειρά T24 (grade 3) σχεδόν όλοι οι μεταβολίτες βρίσκονταν σε εξαιρετικά αυξημένα επίπεδα συγκέντρωσης με την μεταβολική πορεία των αμινοξέων να εμφανίζεται ως μία νέα ισχυρά υποσχόμενη πλατφόρμα για την στοχευόμενη θεραπεία του καρκίνου της ουροδόχου κύστης. Ωστόσο, στην κυτταρική σειρά TCCSUP (grade 4) τα επίπεδα των μεταβολιτών που ανιχνεύθηκαν, συμπεριλαμβανομένων και εκείνων που συμμετέχουν στην μεταβολική πορεία των αμινοξέων, ήταν αισθητά μειωμένα συγκρινόμενα και με την κυτταρική σειρά RT112 (grade 2) και με την κυτταρική σειρά T24 (grade 3). Αυτό πιθανώς μας δείχνει μία «αναστροφή» στη μεταβολική πορεία των κυττάρων κατά τη διάρκεια της τελευταίας μετάστασης. Η μεταβολική αναστροφή πέρα από τη φασματοσκοπία NMR, επιβεβαιώνεται και από τη φασματομετρία μάζας αλλά και από την εφαρμογή του αλγορίθμου KODAMA στα δεδομένα της φασματομετρίας μάζας.

Ο μεταβολισμός της γλυκόζης περιλαμβάνει την γλυκόλυση και τον κύκλο του κιτρικού οξέος. Οι μεταβολίτες εκείνοι που έχουν βρεθεί να παρουσιάζουν σημαντικές διαφορές σε καρκινικά δείγματα είναι η γλυκόζη, η φρουκτόζη, το πυροσταφλικό οξύ, το 2-υδροξυπροπανικό οξύ (lactate), το κιτρικό οξύ και το φουμαρικό οξύ (fumarate). Το 2-υδροξυπροπανικό οξύ (lactate) και το κιτρικό οξύ είναι οι πλέον αντιπροσωπευτικοί μεταβολίτες με τον πρώτο να βρίσκεται υπερεκφρασμένο και τον δεύτερο υποεκφρασμένο. Στην παρούσα εργασία ανιχνεύθηκαν το 2-υδροξυπροπανικό οξύ (lactate) και το φουμαρικό οξύ (fumarate) που αποτελεί και ενδιάμεσο στον κύκλο του κιτρικού οξέος με τα επίπεδα συγκέντρωσής τους να βρίσκονται ιδιαίτερα αυξημένα στην κυτταρική σειρά T24 (grade 3)¹²⁴. Τα αυξημένα επίπεδα 2-υδροξυπροπανικού οξέος (lactate) μπορούν να αποδοθούν στο φαινόμενο Warburg σύμφωνα με το οποίο τα καρκινικά κύτταρα παράγουν ATP κατά τη γλυκόλυση και οδηγούν στην αυξημένη παραγωγή 2-υδροξυπροπανικού οξέος (lactate).

Επίσης παρατηρήθηκαν αυξημένα επίπεδα συγκέντρωσης στην κυτταρική σειρά T24 (grade 3) για τα αμινοξέα αλανίνη, λευκίνη, φαινυλαλανίνη, θρεονίνη, τυροσίνη, βαλίνη, γλυκίνη, ιστιδίνη, ασπαραγινικό οξύ, γλουταμινικό οξύ, γλουταθειόνη, β-αλανίνη, προλίνη, ταυρίνη και ισολευκίνη^{19,139}.

Τέλος, οι μόνοι μεταβολίτες που εμφανίζονται αρχικά με αυξημένα επίπεδα συγκέντρωσης και σταδιακά πέφτουν είναι η τρυπτοφάνη, το φορμικό οξύ και η ουρακίλη.

ΠΑΡΑΡΤΗΜΑ

Παράρτημα 1

Στους παρακάτω πίνακες φαίνονται τα αποτελέσματα από τις στατιστικές δοκιμές one-way ANOVA και t-test από τα δεδομένα του NMR (πίνακας Π1.1). Στον επόμενο πίνακα (πίνακας Π1.2) παρουσιάζονται τα αποτελέσματα για το αν οι μεταβλητές μας ακολουθούν ή όχι κανονική κατανομή. Στον πίνακα Π1.3 φαίνεται το εμβαδό κάτω από τη καμπύλη ROC (AUC)όπως προκύπτει για τον κάθε μεταβολίτη για όλες τις συγκρίσεις μεταξύ των ομάδων. Για την στατιστική ανάλυση χρησιμοποιήθηκαν και οι λόγοι Fold Change¹⁴⁰ και Z-score

Πίνακας Π1.1 Αποτελέσματα της δοκιμασίας one-way ANOVA και t-test μεταξύ των ομάδων. Όταν σημειώνεται Yes τότε έχουμε στατιστικά σημαντική διαφορά ενώ με No δεν έχουμε. Με A η κυτταρική σειρά RT4 (Grade 1), με B η κυτταρική σειρά RT112 (Grade 2), με C η κυτταρική σειρά T24 (Grade 3), με D η κυτταρική σειρά TCCSUP (Grade 4).

| Metabolite | one-way ANOVA | | | | | | t-test | | | | | |
|-------------------------|---------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | A vs B | A vs C | A vs D | B vs C | B vs D | C vs D | A vs B | A vs C | A vs D | B vs C | B vs D | C vs D |
| Alanine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Isoleucine | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Leucine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N-Acetylglutamine | No | Yes | No | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Phenylalanine | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Threonine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Tyrosine | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Valine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Glycine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Proline | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Histidine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Tryptophan | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes |
| Aspartate | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | Yes | Yes |
| Glutamate | Yes | Yes | No | Yes | Yes | Yes | No | Yes | No | Yes | Yes | Yes |
| Glutathione | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
| Taurine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| β-Alanine | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Formate | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Acetate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Lactate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Succinate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fumarate | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Malate | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| ADP | No | Yes | No | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| AMP | No | Yes | No | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| ATP | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| GTP | No | Yes | No | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes |
| NAD+ | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| UDP-N-Acetylglucosamine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| UDP's | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| UMP | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Oxypurinol | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Hypoxanthine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adenine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Uracil | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Creatine | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes |
| CreatinePhosphate | No | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| myo-Inositol | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| O-Phosphocholine | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes |
| Choline | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Propylene glycol | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Πίνακας Π1.2 Αποτελέσματα της δοκιμασίας normality test. Όταν σημειώνεται Yes τότε ακολουθείται κανονική κατανομή ενώ με No όχι. Με A η κυτταρική σειρά RT4 (Grade 1), με B η κυτταρική σειρά RT112 (Grade 2), με C η κυτταρική σειρά T24 (Grade 3), με D η κυτταρική σειρά TCCSUP (Grade 4).

| Metabolite | D' Agostino & Pearson | | | | Shapiro-Wilk | | | | Kolmogorov-Smirnov | | | |
|-------------------------|-----------------------|-----|-----|-----|--------------|-----|-----|-----|--------------------|-----|-----|-----|
| | A | B | C | D | A | B | C | D | A | B | C | D |
| Alanine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Isoleucine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Leucine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| N-Acetylglutamine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Phenylalanine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Threonine | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Tyrosine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Valine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Glycine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Proline | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Histidine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Tryptophan | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Aspartate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Glutamate | Yes | Yes | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes |
| Glutathione | Yes | Yes | No | No | Yes | Yes | Yes | No | Yes | Yes | No | No |
| Taurine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| β-Alanine | No | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Formate | Yes | No | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes | No |
| Acetate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Lactate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| Succinate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Fumarate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Malate | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ADP | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes |
| AMP | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ATP | Yes | Yes | Yes | Yes | No | Yes | Yes | No | Yes | Yes | Yes | No |
| GTP | Yes | Yes | No | Yes | No | No | No | Yes | Yes | No | Yes | No |
| NAD+ | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| UDP-N-Acetylglucosamine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| UDP's | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| UMP | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Oxypurinol | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Hypoxanthine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adenine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Uracil | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Creatine | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | No | Yes |
| CreatinePhosphate | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| myo-Inositol | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| O-Phosphocholine | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Choline | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Propylene glycol | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

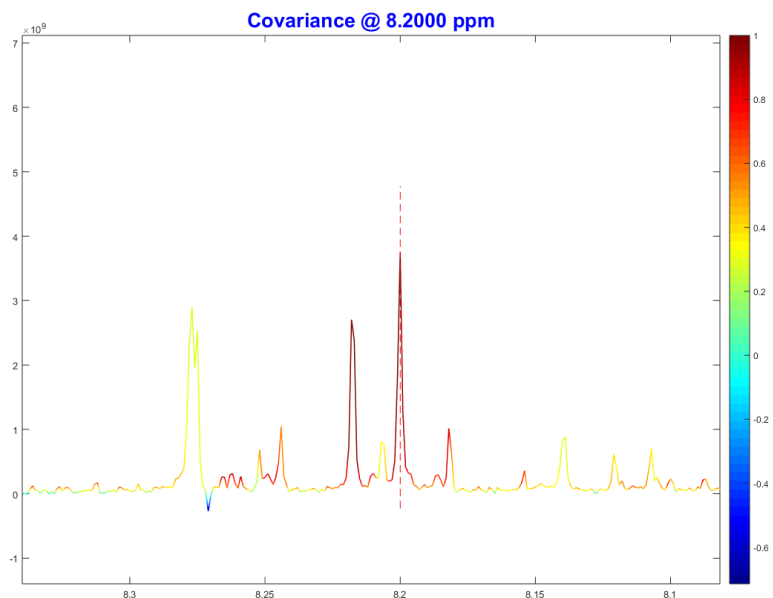
Πίνακας Π1.3 Αποτελέσματα AUC για την κάθε σύγκριση. Με Α η κυτταρική σειρά RT4 (Grade 1), με Β η κυτταρική σειρά RT112 (Grade 2), με C η κυτταρική σειρά T24 (Grade 3), με D η κυτταρική σειρά TCCSUP(Grade 4).

| Metabolite | AUC | | | | | |
|-------------------------|--------|--------|--------|--------|--------|--------|
| | A vs B | A vs C | A vs D | B vs C | B vs D | C vs D |
| Alanine | 1 | 1 | 0.99 | 1 | 1 | 1 |
| Isoleucine | 1 | 1 | 0.78 | 1 | 1 | 1 |
| Leucine | 1 | 1 | 1 | 1 | 1 | 1 |
| N-Acetylglutamine | 0.93 | 1 | 0.51 | 1 | 0.98 | 1 |
| Phenylalanine | 1 | 1 | 0.83 | 1 | 1 | 1 |
| Threonine | 1 | 1 | 0.94 | 1 | 1 | 1 |
| Tyrosine | 1 | 1 | 0.83 | 1 | 1 | 1 |
| Valine | 1 | 1 | 1 | 1 | 1 | 1 |
| Glycine | 1 | 1 | 1 | 1 | 1 | 1 |
| Proline | 1 | 1 | 1 | 1 | 1 | 1 |
| Histidine | 1 | 1 | 0.93 | 1 | 1 | 1 |
| Tryptophan | 0.52 | 1 | 1 | 1 | 1 | 1 |
| Aspartate | 0.96 | 1 | 1 | 1 | 1 | 1 |
| Glutamate | 1 | 1 | 0.95 | 1 | 1 | 1 |
| Glutathione | 1 | 1 | 0.85 | 1 | 1 | 1 |
| Taurine | 1 | 1 | 1 | 1 | 1 | 1 |
| β-Alanine | 1 | 1 | 0.8 | 1 | 1 | 1 |
| Formate | 0.78 | 0.99 | 1 | 0.95 | 1 | 1 |
| Acetate | 1 | 1 | 1 | 1 | 1 | 1 |
| Lactate | 1 | 1 | 1 | 1 | 1 | 1 |
| Succinate | 1 | 1 | 1 | 1 | 1 | 1 |
| Fumarate | 1 | 1 | 0.62 | 1 | 1 | 1 |
| Malate | 1 | 1 | 0.76 | 1 | 1 | 1 |
| ADP | 0.96 | 1 | 0.91 | 1 | 0.84 | 1 |
| AMP | 0.94 | 1 | 0.64 | 1 | 0.96 | 1 |
| ATP | 0.96 | 1 | 0.56 | 1 | 1 | 1 |
| GTP | 0.95 | 1 | 0.72 | 1 | 1 | 1 |
| NAD+ | 1 | 1 | 0.76 | 0.99 | 1 | 1 |
| UDP-N-Acetylglucosamine | 1 | 1 | 0.95 | 1 | 1 | 1 |
| UDP's | 1 | 1 | 0.8 | 1 | 1 | 1 |
| UMP | 0.98 | 1 | 0.65 | 1 | 1 | 1 |
| Oxypurinol | 0.98 | 1 | 0.99 | 1 | 1 | 1 |
| Hypoxanthine | 1 | 1 | 1 | 1 | 1 | 1 |
| Adenine | 1 | 1 | 1 | 1 | 1 | 1 |
| Uracil | 0.99 | 1 | 1 | 1 | 1 | 1 |
| Creatine | 1 | 1 | 1 | 1 | 1 | 1 |
| CreatinePhosphate | 1 | 1 | 0.98 | 1 | 1 | 1 |
| myo-Inositol | 1 | 1 | 0.98 | 1 | 1 | 1 |
| O-Phosphocholine | 1 | 1 | 1 | 1 | 1 | 1 |
| Choline | 1 | 1 | 1 | 1 | 1 | 1 |
| Propylene glycol | 1 | 1 | 1 | 0.97 | 1 | 1 |

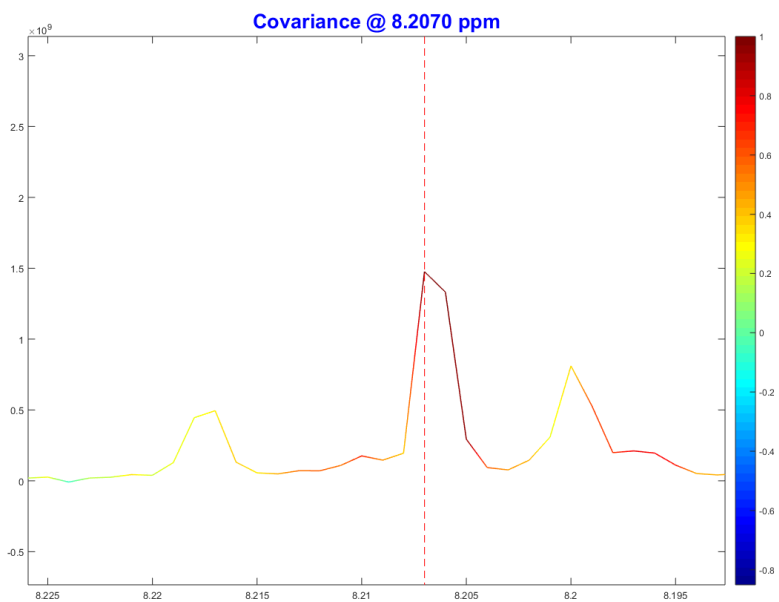
Πίνακας Π1.4 Αποτελέσματα Fold Change και Z-score (ως ομάδα control λήφθηκε η ομάδα A). Με A η κυτταρική σειρά RT4 (Grade 1), με B η κυτταρική σειρά RT112 (Grade 2), με C η κυτταρική σειρά T24 (Grade 3), με D η κυτταρική σειρά TCCSUP (Grade 4).

| Metabolite | Fold Change | | | Z-score | | |
|-------------------------|-------------|----------|----------|----------|----------|----------|
| | B/A | C/A | D/A | B/A | C/A | D/A |
| Alanine | 1.663484 | 6.992112 | 0.806743 | 7.990745 | 72.16669 | -2.32751 |
| Isoleucine | 1.960095 | 5.617831 | 0.864746 | 6.639798 | 31.93587 | -0.93539 |
| Leucine | 1.609273 | 3.510099 | 0.770314 | 8.976212 | 36.98042 | -3.38389 |
| N-Acetylglutamine | 1.143023 | 10.16698 | 0.988247 | 1.774851 | 113.7584 | -0.14585 |
| Phenylalanine | 1.817684 | 5.91547 | 0.855085 | 6.282918 | 37.76948 | -1.1135 |
| Threonine | 1.416474 | 5.384462 | 0.738239 | 2.395137 | 25.21501 | -1.50538 |
| Tyrosine | 1.767084 | 6.245068 | 0.877977 | 9.103421 | 62.24619 | -1.44812 |
| Valine | 1.790674 | 4.806452 | 0.744162 | 11.29696 | 54.38561 | -3.65535 |
| Glycine | 1.332103 | 5.714195 | 0.318241 | 5.391235 | 76.52857 | -11.0674 |
| Proline | 1.502362 | 4.766334 | 0.745129 | 4.661254 | 34.94657 | -2.36486 |
| Histidine | 1.406721 | 2.911951 | 0.739522 | 3.432718 | 16.13684 | -2.19843 |
| Tryptophan | 1.01421 | 1.598774 | 0.374507 | 0.128293 | 5.406002 | -5.64724 |
| Aspartate | 1.176048 | 3.331846 | 0.778045 | 1.998609 | 26.47266 | -2.51978 |
| Glutamate | 1.410954 | 6.435714 | 0.883756 | 6.754748 | 89.34544 | -1.91067 |
| Glutathione | 1.651081 | 7.183366 | 0.87063 | 5.464615 | 51.89788 | -1.08582 |
| Taurine | 1.551963 | 5.499395 | 0.668281 | 4.360801 | 35.54764 | -2.62076 |
| β-Alanine | 1.683763 | 4.544303 | 0.823466 | 9.626952 | 49.90151 | -2.48549 |
| Formate | 1.128682 | 1.284981 | 0.679651 | 1.387493 | 3.072765 | -3.45412 |
| Acetate | 1.562778 | 4.169233 | 2.779624 | 7.858329 | 44.2535 | 24.84972 |
| Lactate | 1.386763 | 3.222043 | 0.393889 | 3.385555 | 19.45078 | -5.30563 |
| Succinate | 1.786852 | 4.652547 | 0.509788 | 11.52747 | 53.51027 | -7.18167 |
| Fumarate | 1.648457 | 3.335524 | 0.920131 | 4.027222 | 14.50469 | -0.49602 |
| Malate | 1.947242 | 6.49596 | 1.118931 | 5.260063 | 30.51924 | 0.660425 |
| ADP | 1.763538 | 20.76109 | 1.492275 | 2.25951 | 58.47824 | 1.456771 |
| AMP | 1.452255 | 34.33477 | 0.794791 | 2.177025 | 160.4639 | -0.98782 |
| ATP | 1.599324 | 6.097853 | 0.883964 | 2.754068 | 23.42611 | -0.53322 |
| GTP | 1.288508 | 11.31109 | 0.879724 | 1.994353 | 71.27696 | -0.83143 |
| NAD+ | 2.86177 | 3.755624 | 0.857567 | 10.70988 | 15.8518 | -0.81935 |
| UDP-N-Acetylglucosamine | 2.43074 | 3.252981 | 0.752958 | 13.09981 | 20.62822 | -2.26191 |
| UDP's | 1.705962 | 2.662977 | 0.833703 | 3.848919 | 9.066589 | -0.90666 |
| UMP | 1.816383 | 7.765539 | 0.824995 | 2.365702 | 19.60506 | -0.50713 |
| Oxypurinol | 1.458259 | 10.16202 | 0.450419 | 2.848729 | 56.95496 | -3.41643 |
| Hypoxanthine | 2.411673 | 1.809644 | 0.235456 | 12.97848 | 7.443614 | -7.02898 |
| Adenine | 0.679887 | 1.330563 | 0.121509 | -3.40047 | 3.511475 | -9.33197 |
| Uracil | 1.355897 | 0.72098 | 0.234043 | 3.720979 | -2.91722 | -8.00826 |
| Creatine | 1.35136 | 5.72559 | 0.453277 | 4.491789 | 60.41198 | -6.98931 |
| CreatinePhosphate | 1.506501 | 22.98698 | 0.657259 | 4.179263 | 181.4201 | -2.82804 |
| myo-Inositol | 2.88783 | 12.05921 | 0.370313 | 19.45699 | 113.9822 | -6.48989 |
| O-Phosphocholine | 1.269326 | 6.257752 | 0.374494 | 3.177276 | 62.0265 | -7.37919 |
| Choline | 2.957327 | 2.442675 | 0.185122 | 17.19426 | 12.67327 | -7.15834 |
| Propylene glycol | 1.337688 | 0.383677 | 0.285835 | 3.465659 | -6.32526 | -7.32941 |

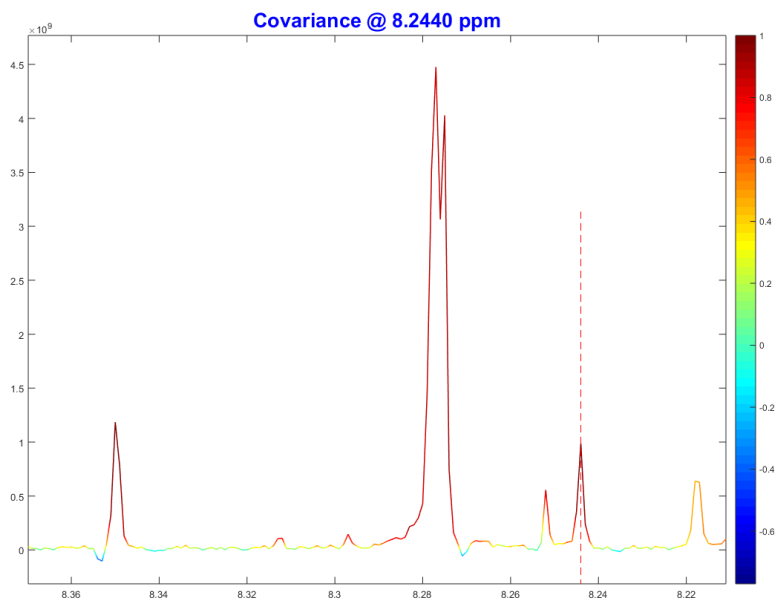
Στη συνέχεια φαίνονται τα αποτελέσματα από το STOCSY.



Εικόνα Π1.1 Πείραμα STOCSY όπου φαίνεται η συνδιακύμανση των δύο κορυφών της υποξανθίνης στα 8.20 και 8.22 ppm



Εικόνα Π1.2 Πείραμα STOCSY όπου επαληθεύεται ότι δεν υπάρχει κάποια συνδιακύμανση με την κορυφή της οξυπουρινόλης στα 8.21 ppm



Εικόνα Π1.3 Πείραμα STOCSY όπου φαίνεται η συνδιακύμανση των δύο κορυφών της ινοσίνης στα 8.24 και 8.35 ppm

Παράρτημα 2

Στη συνέχεια παρουσιάζεται πίνακας με τις τιμές VIPόπως αυτές προέκυψαν από το μοντέλο PLS-DA με όλες τις ομάδες στον αρνητικό ιοντισμό (Διάγραμμα 15, 6.4).

Πίνακας Π2.1 Τιμές VIP από το μοντέλο PLS-DA με Pareto scaling σε αρνητικό ιοντισμό.

| Metabolite | Τιμή VIP | Metabolite | Τιμή VIP |
|------------------|----------|-----------------|----------|
| 151.0269_54.85 | 3.16576 | 554.2035_100.07 | 1.77719 |
| 283.0685_88.02 | 2.99403 | 302.052_61.6 | 1.77521 |
| 321.0882_343.78 | 2.97446 | 535.084_186.63 | 1.7648 |
| 323.0289_51.47 | 2.91739 | 289.0683_56.64 | 1.76017 |
| 283.0688_72.4 | 2.8322 | 117.0201_67.15 | 1.75797 |
| 323.0287_46.17 | 2.7603 | 146.0464_44.1 | 1.75632 |
| 181.051_189.69 | 2.66495 | 789.2104_362.15 | 1.74598 |
| 287.0885_124.66 | 2.5797 | 784.1481_209.94 | 1.7443 |
| 174.0414_48.61 | 2.44922 | 685.1997_357.54 | 1.74145 |
| 800.1436_221.72 | 2.44252 | 243.0628_56.61 | 1.74022 |
| 455.0975_214.26 | 2.43398 | 535.0843_192.42 | 1.73609 |
| 249.0036_54.41 | 2.38501 | 548.7661_45.94 | 1.73059 |
| 290.1608_223.58 | 2.34477 | 259.13_202.08 | 1.72705 |
| 599.3193_551.36 | 2.30231 | 248.0405_51.23 | 1.72379 |
| 165.0563_252.37 | 2.28034 | 662.1021_47.6 | 1.72181 |
| 555.2096_98.58 | 2.2756 | 341.0397_56.54 | 1.71767 |
| 286.9595_54.67 | 2.26572 | 786.1543_208.52 | 1.70299 |
| 152.0301_54.39 | 2.2165 | 130.0627_42.92 | 1.69275 |
| 191.0203_53.79 | 2.20072 | 785.1515_209.38 | 1.68423 |
| 337.0447_59.32 | 2.19644 | 347.9896_60.23 | 1.67345 |
| 958.5633_1028.81 | 2.12413 | 535.1547_70.34 | 1.66963 |
| 822.104_208.16 | 2.11341 | 633.1312_70.59 | 1.66953 |
| 662.1565_182.3 | 2.10923 | 143.0469_53.53 | 1.66695 |
| 320.0627_61.08 | 2.10791 | 364.0668_69.66 | 1.66643 |
| 180.0671_60.09 | 2.09693 | 957.5613_977.94 | 1.64707 |
| 611.145_57.53 | 2.09075 | 550.7642_46.49 | 1.63437 |
| 317.003_59.85 | 2.05265 | 333.0945_56.11 | 1.63427 |
| 426.0226_50.35 | 2.04466 | 642.2136_372.65 | 1.63048 |
| 664.1175_61.66 | 2.03038 | 202.1087_251.11 | 1.62484 |
| 218.1035_160.78 | 2.01585 | 270.9645_52.08 | 1.61511 |
| 346.0562_46.89 | 2.01075 | 685.1998_330.13 | 1.60069 |
| 316.08_160.85 | 2.01068 | 297.8857_44.88 | 1.5961 |
| 448.8312_45.85 | 1.99098 | 354.9258_52.29 | 1.58882 |
| 278.044_59.88 | 1.98705 | 188.0569_61.39 | 1.5857 |
| 535.5855_186.63 | 1.9396 | 188.0931_221.01 | 1.58565 |
| 382.5499_175.7 | 1.93497 | 226.979_52.66 | 1.57946 |
| 244.0661_56.28 | 1.92854 | 219.0776_82.72 | 1.57067 |
| 806.13_208.2 | 1.90802 | 368.9414_51.98 | 1.56611 |
| 219.1068_161.3 | 1.90319 | 383.1149_58.11 | 1.56586 |
| 169.9923_212.81 | 1.89053 | 135.0314_54.8 | 1.55793 |
| 709.1215_57.87 | 1.84253 | 364.8698_44.85 | 1.55391 |
| 425.0012_56.63 | 1.84021 | 283.1372_408.32 | 1.55032 |

| | | | |
|-----------------|---------|------------------|---------|
| 612.905_56.12 | 1.83754 | 357.1055_70.39 | 1.54497 |
| 384.9364_54.54 | 1.83562 | 439.0167_56.63 | 1.54477 |
| 479.1967_249.74 | 1.82585 | 481.0919_58.1 | 1.54235 |
| 476.9726_56.41 | 1.82093 | 464.8053_45.36 | 1.53883 |
| 277.0598_124.54 | 1.81582 | 957.561_1028.09 | 1.53402 |
| 958.5636_977.93 | 1.81051 | 279.0394_56.55 | 1.52809 |
| 320.0924_66.58 | 1.80195 | 303.0837_64.12 | 1.52518 |
| 134.0479_52.08 | 1.80036 | 305.0475_70.64 | 1.52366 |
| 226.0127_58.5 | 1.79482 | 641.2104_372.37 | 1.51944 |
| 373.0296_57.47 | 1.78632 | 302.8879_40.82 | 1.51679 |
| 574.9494_56.13 | 1.78574 | 536.0849_186.74 | 1.5161 |
| 612.1482_57.92 | 1.78474 | 304.9143_1289.08 | 1.51575 |
| 423.1231_312.07 | 1.78288 | 146.0826_161.39 | 1.50628 |
| 290.0715_56.61 | 1.77972 | 911.5557_1029.1 | 1.50063 |

Παρακάτω παρουσιάζεται ο πίνακας όπως αυτός προέκυψε από τη διαδικασία του KODAMA με και χωρίς τη δοκιμασία Kruskal-Wallis Rank Sums.

Πίνακας Π2.2 Τιμές loadings όπως αυτές προκύπτουν από τη διαδικασία του KODAMA και KODAMA με τη διαδικασία του Kruskal-Wallis Rank Sums.

| Metabolites | loadings KODAMA | Metabolites | loadings KODAMA (kruskal) |
|------------------|-----------------|------------------|---------------------------|
| X117.0201_67.15 | 12.08942 | X321.0882_343.78 | 0.467701185 |
| X353.049_264.55 | 12.08942 | X633.1312_70.59 | 0.187744021 |
| X164.0721_122.85 | 12.08942 | X278.044_59.88 | 0.173553393 |
| X151.0269_54.85 | 12.08942 | X535.1547_70.34 | 0.16527022 |
| X181.051_189.69 | 12.08942 | X957.561_1028.09 | 0.145107762 |
| X354.0523_264.96 | 12.01038 | X316.08_160.85 | 0.143615689 |
| X308.1158_267.22 | 12.01038 | X550.7642_46.49 | 0.14346296 |
| X633.1312_70.59 | 11.99772 | X337.0447_59.32 | 0.137391644 |
| X535.5855_186.63 | 11.98501 | X146.0826_161.39 | 0.133747682 |
| X307.1123_267.5 | 11.9352 | X382.5499_175.7 | 0.132332389 |
| X371.0593_258.63 | 11.82476 | X599.3193_551.36 | 0.122179765 |
| X351.1023_266.46 | 11.79645 | X448.8312_45.85 | 0.118797589 |
| X535.084_186.63 | 11.72942 | X259.13_202.08 | 0.112678196 |
| X286.9595_54.67 | 11.69898 | X219.1068_161.3 | 0.109472903 |
| X313.0794_70.66 | 11.67317 | X641.2104_372.37 | 0.107864186 |
| X479.1967_249.74 | 11.66571 | X479.1967_249.74 | 0.102913682 |
| X290.1608_223.58 | 11.61732 | X262.1297_87.84 | 0.09874564 |
| X423.1231_312.07 | 11.59168 | X789.2104_362.15 | 0.095450427 |
| X229.1349_267.04 | 11.56532 | X464.8053_45.36 | 0.094846755 |
| X309.1081_267.33 | 11.51719 | X665.1214_61.76 | 0.093350473 |
| X260.0596_122.74 | 11.49368 | X188.0931_221.01 | 0.092130966 |
| X535.0843_192.42 | 11.44299 | X218.1035_160.78 | 0.091506525 |
| X268.077_70.52 | 11.43249 | X481.0919_58.1 | 0.091159264 |
| X384.9364_54.54 | 11.43249 | X202.1087_251.11 | 0.087663233 |
| X658.6746_42.39 | 11.39513 | X297.8857_44.88 | 0.087621336 |
| X685.1998_330.13 | 11.39395 | X490.7923_40.99 | 0.086169817 |

| | | | |
|------------------|----------|------------------|-------------|
| X662.1021_47.6 | 11.38926 | X383.1149_58.11 | 0.085018942 |
| X387.0325_70.59 | 11.36867 | X536.0849_186.74 | 0.083253329 |
| X133.0149_47.4 | 11.36321 | X638.6932_44.85 | 0.082908881 |
| X535.1547_70.34 | 11.33434 | X549.9098_59.15 | 0.080924146 |
| X372.0626_258.58 | 11.32882 | X357.1055_70.39 | 0.079537858 |
| X315.9998_59.85 | 11.30948 | X664.1175_61.66 | 0.078935421 |
| X664.1175_61.66 | 11.30933 | X382.8418_44.71 | 0.077484181 |
| X305.0475_70.64 | 11.26606 | X305.0475_70.64 | 0.074890503 |
| X135.0314_54.8 | 11.25638 | X164.0721_122.85 | 0.074326247 |
| X574.9494_56.13 | 11.24898 | X511.954_59.18 | 0.073949266 |
| X267.0739_70.67 | 11.24776 | X662.1021_47.6 | 0.073695858 |
| X202.1087_251.11 | 11.24776 | X188.0569_61.39 | 0.073223 |
| X188.0931_221.01 | 11.24199 | X169.9923_212.81 | 0.071938342 |
| X351.0352_70.55 | 11.23014 | X685.1997_357.54 | 0.071673392 |
| X794.6066_42.19 | 11.22646 | X642.2136_372.65 | 0.070632357 |
| X146.0464_44.1 | 11.21832 | X535.5855_186.63 | 0.070192235 |
| X366.0541_70.46 | 11.21258 | X317.003_59.85 | 0.069483274 |
| X347.9896_60.23 | 11.20732 | X387.0325_70.59 | 0.069160266 |
| X357.1055_70.39 | 11.19805 | X303.0837_64.12 | 0.068762072 |
| X662.1565_182.3 | 11.19113 | X372.0626_258.58 | 0.067922506 |
| X685.1997_357.54 | 11.18557 | X248.0331_122.79 | 0.067714256 |
| X555.2096_98.58 | 11.18398 | X303.0505_70.67 | 0.067579939 |
| X536.0849_186.74 | 11.15321 | X451.9327_59.58 | 0.066036294 |
| X365.0507_70.66 | 11.15321 | X485.0098_70.63 | 0.064865325 |

Τέλος, παρουσιάζεται ο πίνακας βάσει του οποίου έγινε η στοχευμένη ανάλυση σε αρνητικό και θετικό ιοντισμό.

Πίνακας Π2.3 Λίστα μεταβολιτών βάσει της οποίας έγινε η στοχευμένη ανάλυση σε αρνητικό και θετικό ιοντισμό.

| Metabolite |
|---|
| 3-ΥΔΡΟΞΥ-ΒΟΥΤΥΡΙΚΟ ΟΞΥ |
| ΒΟΥΤΥΡΙΚΟ ΟΞΥ |
| ΠΡΟΠΙΟΝΙΚΟ ΟΞΥ |
| ΦΟΡΜΙΚΟ (ΜΥΡΜΗΚΙΚΟ) ΟΞΥ |
| ΙΣΟΒΑΛΕΡΙΑΝΙΚΟ ΟΞΥ |
| ΓΛΥΚΟΛΙΚΟ ΟΞΥ |
| ΓΛΥΚΕΡΙΝΙΚΟ ΟΞΥ |
| ΤΡΥΓΙΚΟ ΟΞΥ |
| ΓΛΥΚΟΞΥΛΙΚΟ ΟΞΥ |
| ΥΔΡΟΞΥ-ΠΥΡΟΣΤΑΦΥΛΙΚΟ ΟΞΥ |
| 2-ΟΞΟΓΛΟΥΤΑΡΙΚΟ ΟΞΥ |
| ΑΛΛΥΣΙΝΗ |
| 5-ΥΔΡΟΞΥ-ΛΥΣΙΝΗ |
| 4-ΚΑΡΒΟΞΥ-ΓΛΟΥΤΑΜΙΝΙΚΟ ΟΞΥ |
| ΠΥΡΟΓΛΟΥΤΑΜΙΝΙΚΟ ΟΞΥ (5-ΠΥΡΟΛΙΔΙΝΟ-2-ΚΑΡΒΟΞΥΛΙΚΟ ΟΞΥ) |
| ΤΡΑΝΣ-4-ΥΔΡΟΞΥ-ΠΡΟΛΙΝΗ (L-ΥΔΡΟΞΥ-ΠΡΟΛΙΝΗ) |

| |
|---------------------------|
| S-ΑΔΕΝΟΣΥΛΟ-ΜΕΘΕΙΟΝΙΝΗ |
| ΚΡΕΑΤΙΝΙΝΗ |
| ΦΩΣΦΟΡΙΚΗ ΚΡΕΑΤΙΝΙΝΗ |
| S-ΑΔΕΝΟΣΥΛΟ-ΟΜΟΚΥΣΤΕΪΝΗ |
| ΟΜΟΚΥΣΤΕΪΝΗ |
| ΤΕΤΡΑΪΔΡΟΦΥΛΛΙΚΟ ΟΞΥ |
| N-ΦΟΡΜΥΛΟ-ΜΕΘΕΙΟΝΙΝΗ |
| ΙΝΟΣΙΝΗ |
| ΨΕΥΔΟ-ΟΥΡΙΔΙΝΗ |
| N2-ΔΙ-ΜΕΘΥΛΟ-ΓΟΥΑΝΟΣΙΝΗ |
| N6-ΙΣΟ-ΠΕΝΤΕΝΥΛ-ΑΔΕΝΟΣΙΝΗ |
| ΟΡΟΤΙΚΟ ΟΞΥ |
| ΔΙ-ΥΔΡΟ-ΟΡΟΤΙΚΟ ΟΞΥ |
| ΟΡΟΤΙΔΙΝΗ-5' -P |
| ΚΑΡΒΑΜΥΛΟ-ΦΩΣΦΟΡΙΚΟ |
| N-ΚΑΡΒΑΜΥΛΟ-β-ΑΛΑΝΙΝΗ |
| ΞΑΝΘΟΣΙΝΗ |
| 5-ΦΩΣΦΟΡΙΚΗ-ΞΑΝΘΟΣΙΝΗ |
| ΥΠΟΞΑΝΘΙΝΗ |
| ΞΑΝΘΙΝΗ |
| ΟΥΡΙΚΟ ΟΞΥ |
| ΑΛΛΑΝΤΟΪΝΗ |
| ΑΛΛΑΝΤΟΪΝΙΚΟ ΟΞΥ |
| ΟΡΝΙΘΙΝΗ |
| ΚΙΤΡΟΥΛΛΙΝΗ |
| ΟΥΡΙΑ |
| ΑΡΓΙΝΙΝΟ-ΗΛΕΚΤΡΙΚΟ ΟΞΥ |
| ΑΚΕΤΟΞΙΚΟ ΟΞΥ |
| ΚΥΝΟΥΡΕΝΙΚΟ ΟΞΥ |
| ΞΑΝΘΟΥΡΕΝΙΚΟ ΟΞΥ |
| 2-ΟΞΑΔΙΠΙΝΙΚΟ ΟΞΥ |
| ΔΙ-ΥΔΡΟΞΥ-ΑΚΕΤΟΝΗ |
| ΒΕΤΑΪΝΗ |
| ΧΟΛΙΝΗ |
| ΣΑΡΚΟΣΙΝΗ |
| ΜΕΘΥΛΟ-ΓΛΥΟΞΑΛΗ |
| ΑΣΚΟΡΒΙΚΟ ΟΞΥ |
| 4-ΑΜΙΝΟ-ΒΟΥΤΥΡΙΚΟ |
| ΑΚΕΤΟΝΗ |
| ΚΥΣΤΙΝΗ |
| ΑΜΜΩΝΙΑ |
| ΑΙΘΑΝΟΛΑΜΙΝΗ |
| N-ΥΔΡΟΞΥ-ΑΡΓΙΝΙΝΗ |
| 2,3-ΔΙ-ΦΩΣΦΟ-ΓΛΥΚΕΡΙΝΙΚΟ |
| ΙΠΠΟΥΡΙΚΟ ΟΞΥ |
| ΓΟΥΑΝΙΔΙΝΟΞΙΚΟ ΟΞΥ |
| D-ΞΥΛΟΥΛΟΖΗ |
| 5-ΦΩΣΦΟΡΙΚΗ D-ΞΥΛΟΥΛΟΖΗ |
| L-ΞΥΛΟΥΛΟΖΗ |
| ΔΙ-ΜΕΘΥΛ-ΑΡΓΙΝΙΝΗ |

BIBΛΙΟΓΡΑΦΙΑ

1. Rochfort, S. Metabolomics Reviewed: A New “Omics” Platform Technology for Systems Biology and Implications for Natural Products Research. *J. Nat. Prod.* **68**, 1813–1820 (2005).
2. Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G. & Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **22**, 245–252 (2004).
3. Armitage, E. G. & Barbas, C. Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87**, 1–11 (2014).
4. Spratlin, J. L., Serkova, N. J. & Gail Eckhardt, S. Clinical Applications of Metabolomics in Oncology: A Review. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **15**, 431–440 (2009).
5. Griffin, J. L. & Shockcor, J. P. Metabolic profiles of cancer cells. *Nat. Rev. Cancer* **4**, 551–561 (2004).
6. Burger, M. *et al.* Epidemiology and risk factors of urothelial bladder cancer. *Eur. Urol.* **63**, 234–241 (2013).
7. The Value of Cystoscopy. *Hospital (Rio J.)* **47**, 542 (1910).
8. Al-Abbadi, M. A. Basics of cytology. *Avicenna J. Med.* **1**, 18–28 (2011).
9. Alimi, Q. *et al.* Reliability of urinary cytology and cystoscopy for the screening and diagnosis of bladder cancer in patients with neurogenic bladder: A systematic review. *Neurourol. Urodyn.* **37**, 916–925 (2018).

10. Garcia, E. *et al.* Diagnosis of early stage ovarian cancer by ¹H NMR metabonomics of serum explored by use of a microflow NMR probe. *J. Proteome Res.* **10**, 1765–1771 (2011).
11. OuYang, D., Xu, J., Huang, H. & Chen, Z. Metabolomic profiling of serum from human pancreatic cancer patients using ¹H NMR spectroscopy and principal component analysis. *Appl. Biochem. Biotechnol.* **165**, 148–154 (2011).
12. Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T. & Tomita, M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics Off. J. Metabolomic Soc.* **6**, 78–95 (2010).
13. Rodrigues, D. *et al.* Biomarkers in bladder cancer: A metabolomic approach using in vitro and ex vivo model systems. *Int. J. Cancer* **139**, 256–268 (2016).
14. Hyndman, M. E., Mullins, J. K. & Bivalacqua, T. J. Metabolomics and bladder cancer. *Urol. Oncol. Semin. Orig. Investig.* **29**, 558–561 (2011).
15. Rosser, C. J., Urquidi, V. & Goodison, S. Urinary biomarkers of bladder cancer: an update and future perspectives. *Biomark. Med.* **7**, 779–790 (2013).
16. Vrooman, O. P. J. & Witjes, J. A. Urinary markers in bladder cancer. *Eur. Urol.* **53**, 909–916 (2008).
17. van Rhijn, B. W. G., van der Poel, H. G. & van der Kwast, T. H. Urine markers for bladder cancer surveillance: a systematic review. *Eur. Urol.* **47**, 736–748 (2005).

18. Shariff, M. I. F. *et al.* Urinary Metabotyping of Hepatocellular Carcinoma in a UK Cohort Using Proton Nuclear Magnetic Resonance Spectroscopy. *J. Clin. Exp. Hepatol.* **6**, 186–194 (2016).
19. Dettmer, K. *et al.* Distinct metabolic differences between various human cancer and primary cells. *Electrophoresis* **34**, 2836–2847 (2013).
20. Bansal, N. *et al.* Low- and high-grade bladder cancer determination via human serum-based metabolomics approach. *J. Proteome Res.* **12**, 5839–5850 (2013).
21. Tripathi, P. *et al.* HR-MAS NMR tissue metabolomic signatures cross-validated by mass spectrometry distinguish bladder cancer from benign disease. *J. Proteome Res.* **12**, 3519–3528 (2013).
22. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
23. Pasikanti, K. K. *et al.* Noninvasive urinary metabonomic diagnosis of human bladder cancer. *J. Proteome Res.* **9**, 2988–2995 (2010).
24. Anghileri, L. J., Crone-Escanye, M. C., Thouvenot, P., Brunotte, F. & Robert, J. Mechanisms of gallium-67 accumulation by tumors: role of cell membrane permeability. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **29**, 663–668 (1988).
25. Bucay, A. H. Clinical report: a patient with primary peritoneal mesothelioma that has improved after taking citric acid orally. *Clin. Res. Hepatol. Gastroenterol.* **35**, 241 (2011).
26. Halabe Bucay, A. Hypothesis proved...citric acid (citrate) does improve cancer: a case of a patient suffering from medullary thyroid cancer. *Med. Hypotheses* **73**, 271 (2009).

27. Jin, X. *et al.* Diagnosis of bladder cancer and prediction of survival by urinary metabolomics. *Oncotarget* **5**, 1635–1645 (2014).
28. Pan, Z. & Raftery, D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal. Bioanal. Chem.* **387**, 525–527 (2007).
29. Bingol, K. *et al.* Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis* **8**, 557–573 (2016).
30. Issaq, H. J. *et al.* Detection of Bladder Cancer in Human Urine by Metabolomic Profiling Using High Performance Liquid Chromatography/Mass Spectrometry. *J. Urol.* **179**, 2422–2426 (2008).
31. Marshall, D. D. & Powers, R. Beyond the Paradigm: Combining Mass Spectrometry and Nuclear Magnetic Resonance for Metabolomics. *Prog. Nucl. Magn. Reson. Spectrosc.* **100**, 1–16 (2017).
32. Dunn, W. B., Bailey, N. J. C. & Johnson, H. E. Measuring the metabolome: current analytical technologies. *The Analyst* **130**, 606–625 (2005).
33. Bijland, L. R., Bomers, M. K. & Smulders, Y. M. Smelling the diagnosis: a review on the use of scent in diagnosing disease. *Neth. J. Med.* **71**, 300–307 (2013).
34. Wolfender, J.-L., Marti, G., Thomas, A. & Bertrand, S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromatogr. A* **1382**, 136–164 (2015).
35. Bingol, K. & Brüsweiler, R. Multidimensional approaches to NMR-based metabolomics. *Anal. Chem.* **86**, 47–57 (2014).

36. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **43**, 34–40 (2017).
37. Jankowski, W., Kurek, J., Barczyński, P. & Hoffmann, M. Quantum-chemical, NMR, FT IR, and ESI MS studies of complexes of colchicine with Zn(II). *J. Mol. Model.* **23**, 127 (2017).
38. Bogner, W., Hangel, G., Esmaeili, M. & Andronesi, O. C. 1D-spectral editing and 2D multispectral in vivo ¹H-MRS and ¹H-MRSI - Methods and applications. *Anal. Biochem.* **529**, 48–64 (2017).
39. Karlsson, O. & Hanrieder, J. Imaging mass spectrometry in drug development and toxicology. *Arch. Toxicol.* **91**, 2283–2294 (2017).
40. Lazova, R. & Seeley, E. H. Proteomic Mass Spectrometry Imaging for Skin Cancer Diagnosis. *Dermatol. Clin.* **35**, 513–519 (2017).
41. Zhou, J. & Yin, Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *The Analyst* **141**, 6362–6373 (2016).
42. Fiehn, O. Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr. Protoc. Mol. Biol.* **114**, 30.4.1-30.4.32 (2016).
43. Verma, N., Pink, M., Boland, S., Rettenmeier, A. W. & Schmitz-Spanke, S. Benzo[a]pyrene-induced metabolic shift from glycolysis to pentose phosphate pathway in the human bladder cancer cell line RT4. *Sci. Rep.* **7**, 9773 (2017).
44. Konstantakou, E. G. *et al.* Human bladder cancer cells undergo cisplatin-induced apoptosis that is associated with p53-dependent and p53-independent responses. *Int. J. Oncol.* **35**, 401–416 (2009).

45. Konstantakou, E. G. *et al.* 3-BrPA eliminates human bladder cancer cells with highly oncogenic signatures via engagement of specific death programs and perturbation of multiple signaling and metabolic determinants. *Mol. Cancer* **14**, (2015).
46. Wu, H., Southam, A. D., Hines, A. & Viant, M. R. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal. Biochem.* **372**, 204–212 (2008).
47. Athersuch, T. J. & Keun, H. C. Metabolic profiling in human exposome studies. *Mutagenesis* **30**, 755–762 (2015).
48. Le Belle, J. E., Harris, N. G., Williams, S. R. & Bhakoo, K. K. A comparison of cell and tissue extraction techniques using high-resolution ¹H-NMR spectroscopy. *NMR Biomed.* **15**, 37–44 (2002).
49. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
50. Blaise, B. J., Navratil, V., Emsley, L. & Toulhoat, P. Orthogonal filtered recoupled-STOCSY to extract metabolic networks associated with minor perturbations from NMR spectroscopy. *J. Proteome Res.* **10**, 4342–4348 (2011).
51. Hoijemberg, P. A. & Pelczer, I. Fast Metabolite Identification in Nuclear Magnetic Resonance Metabolomic Studies: Statistical Peak Sorting and Peak Overlap Detection for More Reliable Database Queries. *J. Proteome Res.* **17**, 392–401 (2018).

52. Broeckling, C. D., Hoyes, E., Richardson, K., Brown, J. M. & Prenni, J. E. Comprehensive Tandem-Mass-Spectrometry Coverage of Complex Samples Enabled by Data-Set-Dependent Acquisition. *Anal. Chem.* **90**, 8020–8027 (2018).
53. Adusumilli, R. & Mallick, P. Data Conversion with ProteoWizard msConvert. *Methods Mol. Biol. Clifton NJ* **1550**, 339–368 (2017).
54. Libiseller, G. *et al.* IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics* **16**, 118 (2015).
55. Myers, O. D., Sumner, S. J., Li, S., Barnes, S. & Du, X. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. *Anal. Chem.* **89**, 8689–8695 (2017).
56. Li, B. *et al.* NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* **45**, W162–W170 (2017).
57. Chawade, A., Alexandersson, E. & Levander, F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* **13**, 3114–3120 (2014).
58. A systematic evaluation of normalization methods in quantitative label-free proteomics. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/27694351>. (Accessed: 22nd January 2019)
59. Wang, X., Gardiner, E. J. & Cairns, M. J. Optimal consistency in microRNA expression analysis using reference-gene-based normalization. *Mol. Biosyst.* **11**, 1235–1240 (2015).

60. Normalizing and integrating metabolomics data. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/23150939>. (Accessed: 22nd January 2019)
61. Statistical analysis of metabolomics data. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/23963918>. (Accessed: 22nd January 2019)
62. Gromski, P. S., Xu, Y., Hollywood, K. A., Turner, M. L. & Goodacre, R. The influence of scaling metabolomics data on model classification accuracy. *Metabolomics* **11**, 684–695 (2015).
63. Cacciatore, S., Tenori, L., Luchinat, C., Bennett, P. R. & MacIntyre, D. A. KODAMA: an R package for knowledge discovery and data mining. *Bioinformatics* **33**, 621–623 (2017).
64. Cacciatore, S., Luchinat, C. & Tenori, L. Knowledge discovery by accuracy maximization. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5117–5122 (2014).
65. Liland, K. H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trends Anal. Chem.* **30**, 827–841 (2011).
66. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. (2013). Available at: <https://www.ingentaconnect.com/content/ben/cmb/2013/00000001/00000001/art00008>. (Accessed: 26th November 2018)
67. N. Triba, M. *et al.* PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* **11**, 13–19 (2015).

68. Worley, B. & Powers, R. PCA as a Practical Indicator of OPLS-DA Model Reliability. (2016). Available at: <https://www.ingentaconnect.com/content/ben/cmb/2016/00000004/00000002/art0004>. (Accessed: 26th November 2018)
69. Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **20**, 397–405 (1978).
70. Lan, K., Zhang, Y., Yang, J. & Xu, L. Simple quality assessment approach for herbal extracts using high performance liquid chromatography-UV based metabolomics platform. *J. Chromatogr. A* **1217**, 1414–1418 (2010).
71. Zhang, R. *et al.* Metabolomic Profiling of Post-Mortem Brain Reveals Changes in Amino Acid and Glucose Metabolism in Mental Illness Compared with Controls. *Comput. Struct. Biotechnol. J.* **14**, 106–116 (2016).
72. Martínez-Arranz, I. *et al.* Data in support of enhancing metabolomics research through data mining. *Data Brief* **3**, 155–164 (2015).
73. Gromski, P. S. *et al.* A tutorial review: Metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **879**, 10–23 (2015).
74. Westerhuis, J. A. *et al.* Assessment of PLS-DA cross validation. *Metabolomics* **4**, 81–89 (2008).
75. Westerhuis, J. A., van Velzen, E. J. J., Hoefsloot, H. C. J. & Smilde, A. K. Multivariate paired data analysis: multilevel PLS-DA versus OPLS-DA. *Metabolomics* **6**, 119–128 (2010).

76. Bylesjö, M. *et al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **20**, 341–351 (2006).
77. Uarrota, V. G. *et al.* Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. *Food Chem.* **161**, 67–78 (2014).
78. Gavaghan, C. L., Wilson, I. D. & Nicholson, J. K. Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. *FEBS Lett.* **530**, 191–196 (2002).
79. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion - ScienceDirect. Available at: <https://www.sciencedirect.com/science/article/pii/S0003267013001700>. (Accessed: 3rd December 2018)
80. Tapp, H. S. & Kemsley, E. K. Notes on the practical utility of OPLS. *TrAC Trends Anal. Chem.* **28**, 1322–1327 (2009).
81. Bijlsma, S. *et al.* Large-Scale Human Metabolomics Studies: A Strategy for Data (Pre-) Processing and Validation. *Anal. Chem.* **78**, 567–574 (2006).
82. Rubingh, C. M. *et al.* Assessing the performance of statistical validation tools for megavariable metabolomics data. *Metabolomics* **2**, 53–61 (2006).
83. Mielke, P. W. & Berry, K. J. *Permutation Methods: A Distance Function Approach*. (Springer Science & Business Media, 2007).
84. Kim, H.-J., Fay, M. P., Feuer, E. J. & Midthune, D. N. Permutation tests for joinpoint regression with applications to cancer rates. *Stat. Med.* **19**, 335–351 (2000).

85. Multi-response permutation procedures for a priori classifications: Communications in Statistics - Theory and Methods: Vol 5, No 14. Available at: <https://www.tandfonline.com/doi/abs/10.1080/03610927608827451>. (Accessed: 13th December 2018)
86. Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **9**, 280–299 (2013).
87. Zou, K. H., Hall, W. J. & Shapiro, D. E. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat. Med.* **16**, 2143–2156 (1997).
88. Seli, E. *et al.* Receiver operating characteristic (ROC) analysis of day 5 morphology grading and metabolomic Viability Score on predicting implantation outcome. *J. Assist. Reprod. Genet.* **28**, 137–144 (2011).
89. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond - Pencina - 2008 - Statistics in Medicine - Wiley Online Library. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2929>. (Accessed: 13th December 2018)
90. Galindo-Prieto, B., Eriksson, L. & Trygg, J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *J. Chemom.* **28**, 623–632 (2014).

91. Folch-Fortuny, A., Prats-Montalbán, J. M., Cubero, S., Blasco, J. & Ferrer, A. VIS/NIR hyperspectral imaging and N-way PLS-DA models for detection of decay lesions in citrus fruits. *Chemom. Intell. Lab. Syst.* **156**, 241–248 (2016).
92. Musumarra, G., Barresi, V., Condorelli, D. F., Fortuna, C. G. & Scirè, S. Genome-based identification of diagnostic molecular markers for human lung carcinomas by PLS-DA. *Comput. Biol. Chem.* **29**, 183–195 (2005).
93. Musumarra, G., Barresi, V., Condorelli, D. F., Fortuna, C. G. & Scirè, S. Potentialities of multivariate approaches in genome-based cancer research: identification of candidate genes for new diagnostics by PLS discriminant analysis. *J. Chemom.* **18**, 125–132 (2004).
94. Laiakis, E. C., Morris, G. A. J., Jr, A. J. F. & Howie, S. R. C. Metabolomic Analysis in Severe Childhood Pneumonia in The Gambia, West Africa: Findings from a Pilot Study. *PLOS ONE* **5**, e12655 (2010).
95. Schievano, E., Stocchero, M., Morelato, E., Facchin, C. & Mammi, S. An NMR-based metabolomic approach to identify the botanical origin of honey. *Metabolomics* **8**, 679–690 (2012).
96. Kang, H. J. *et al.* Metabolomic analysis of meju during fermentation by ultra performance liquid chromatography-quadrupole-time of flight mass spectrometry (UPLC-Q-TOF MS). *Food Chem.* **127**, 1056–1064 (2011).
97. Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A. & Hendriks, M. M. W. B. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**, 361–374 (2014).

98. Vinaixa, M. *et al.* A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* **2**, 775–795 (2012).
99. Bartel, J., Krumsiek, J. & Theis, F. J. STATISTICAL METHODS FOR THE ANALYSIS OF HIGH-THROUGHPUT METABOLOMICS DATA. *Comput. Struct. Biotechnol. J.* **4**, e201301009 (2013).
100. Smilde, A. K. *et al.* ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* **21**, 3043–3048 (2005).
101. Xu, Y. & Goodacre, R. Multiblock principal component analysis: an efficient tool for analyzing metabolomics data which contain two influential factors. *Metabolomics* **8**, 37–51 (2012).
102. Arani'bar, N., Ott, K.-H., Roongta, V. & Mueller, L. Metabolomic analysis using optimized NMR and statistical methods. *Anal. Biochem.* **355**, 62–70 (2006).
103. Antoneli, F., Passos, F. M., Lopes, L. R. & Briones, M. R. S. A Kolmogorov-Smirnov test for the molecular clock based on Bayesian ensembles of phylogenies. *PLoS One* **13**, e0190826 (2018).
104. Wallot, S. & Leonardi, G. Deriving inferential statistics from recurrence plots: A recurrence-based test of differences between sample distributions and its comparison to the two-sample Kolmogorov-Smirnov test. *Chaos Woodbury N* **28**, 085712 (2018).
105. Beale*, M. H., Ward, J. L. & Baker, J. M. Establishing Substantial Equivalence: Metabolomics. in *Transgenic Wheat, Barley and Oats: Production and*

- Characterization Protocols* (eds. Jones, H. D. & Shewry, P. R.) 289–303 (Humana Press, 2009). doi:10.1007/978-1-59745-379-0_17
106. Tsugawa, H., Tsujimoto, Y., Arita, M., Bamba, T. & Fukusaki, E. GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA). *BMC Bioinformatics* **12**, 131 (2011).
107. Büscher, J. M., Czernik, D., Ewald, J. C., Sauer, U. & Zamboni, N. Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Anal. Chem.* **81**, 2135–2143 (2009).
108. Redestig, H. *et al.* Compensation for Systematic Cross-Contribution Improves Normalization of Mass Spectrometry Based Metabolomics Data. *Anal. Chem.* **81**, 7974–7980 (2009).
109. Madala, N. E., Piater, L. A., Steenkamp, P. A. & Dubery, I. A. Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited *Nicotiana tabacum* and *Sorghum bicolor* cells. *SpringerPlus* **3**, (2014).
110. Xi, B., Gu, H., Baniyadi, H. & Raftery, D. Statistical Analysis and Modeling of Mass Spectrometry-Based Metabolomics Data. *Methods Mol. Biol. Clifton NJ* **1198**, 333–353 (2014).
111. Lv, H., Hung, C. S., Chaturvedi, K. S., Hooton, T. M. & Henderson, J. P. Development of an Integrated Metabolomic Profiling Approach for Infectious Diseases Research. *The Analyst* **136**, 4752–4763 (2011).

112. Xu, X., Yang, J., Ning, Z. & Zhang, X. ¹H NMR-Based Metabolic Profiling of Urine from Mice Fed Lentinula edodes-Derived Polysaccharides. *Pol. J. Food Nutr. Sci.* **68**, (2018).
113. Putri, S. P. & Fukusaki, E. *Mass Spectrometry-Based Metabolomics: A Practical Guide*. (CRC Press, 2016).
114. Weljie, A. M., Newton, J., Mercier, P., Carlson, E. & Slupsky, C. M. Targeted Profiling: Quantitative Analysis of ¹H NMR Metabolomics Data. *Anal. Chem.* **78**, 4430–4442 (2006).
115. Parsons, H. M., Ludwig, C., Günther, U. L. & Viant, M. R. Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics* **8**, 234 (2007).
116. Statistical methods for handling unwanted variation in metabolomics data. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4544854/>. (Accessed: 1st January 2019)
117. Aggio, R., Villas-Bôas, S. G. & Ruggiero, K. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* **27**, 2316–2318 (2011).
118. Kalhan, S. C. *et al.* Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism* **60**, 404–413 (2011).
119. Grace, S. C. & Hudson, D. A. Processing and Visualization of Metabolomics Data Using R. *Metabolomics - Fundam. Appl.* (2016). doi:10.5772/65405

120. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
121. Patti, G. J., Tautenhahn, R. & Siuzdak, G. Meta-analysis of untargeted metabolomic data from multiple profiling experiments. *Nat. Protoc.* **7**, 508–516 (2012).
122. Benton, H. P., Wong, D. M., Trauger, S. A. & Siuzdak, G. XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. *Anal. Chem.* **80**, 6382–6389 (2008).
123. Narath, S. H. *et al.* An Untargeted Metabolomics Approach to Characterize Short-Term and Long-Term Metabolic Changes after Bariatric Surgery. *PLOS ONE* **11**, e0161425 (2016).
124. Brunius, C., Shi, L. & Landberg, R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. *Metabolomics* **12**, 173 (2016).
125. Yu, T., Park, Y., Johnson, J. M. & Jones, D. P. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* **25**, 1930–1936 (2009).
126. statTarget: A streamlined tool for signal drift correction and interpretations of quantitative mass spectrometry-based omics data - ScienceDirect. Available at: <https://www.sciencedirect.com/science/article/pii/S0003267018309395>. (Accessed: 23rd December 2018)

127. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).
128. Overgaard, A. J. *et al.* Lipidomic and metabolomic characterization of a genetically modified mouse model of the early stages of human type 1 diabetes pathogenesis. *Metabolomics* **12**, (2016).
129. Yin, P. *et al.* Preanalytical Aspects and Sample Quality Assessment in Metabolomics Studies of Human Blood. *Clin. Chem.* clinchem.2012.199257 (2013). doi:10.1373/clinchem.2012.199257
130. Bernini, P. *et al.* Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J. Biomol. NMR* **49**, 231–243 (2011).
131. Vereyken, L., Dillen, L., Vreeken, R. J. & Cuyckens, F. High-Resolution Mass Spectrometry Quantification: Impact of Differences in Data Processing of Centroid and Continuum Data. *J. Am. Soc. Mass Spectrom.* (2018). doi:10.1007/s13361-018-2101-0
132. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
133. Kind, T., Tolstikov, V., Fiehn, O. & Weiss, R. H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal. Biochem.* **363**, 185–195 (2007).

134. Orešič, M. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutr. Metab. Cardiovasc. Dis.* **19**, 816–824 (2009).
135. Sources and scope of European Union law | Fact Sheets on the European Union | European Parliament. Available at: <http://www.europarl.europa.eu/factsheets/en/sheet/6/sources-and-scope-of-european-union-law>. (Accessed: 27th January 2019)
136. Cheng, Y. *et al.* Metabolomics in bladder cancer: a systematic review. *Int. J. Clin. Exp. Med.* **8**, 11052–11063 (2015).
137. Ploeg, M., Aben, K. K. H. & Kiemeny, L. A. The present and future burden of urinary bladder cancer in the world. *World J. Urol.* **27**, 289–293 (2009).
138. Holmes, E., Wilson, I. D. & Nicholson, J. K. Metabolic phenotyping in health and disease. *Cell* **134**, 714–717 (2008).
139. Putluri, N. *et al.* Metabolomic Profiling Reveals Potential Markers and Bioprocesses Altered in Bladder Cancer Progression. *Cancer Res.* **71**, 7376–7386 (2011).
140. van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142 (2006).