

Κεφάλαιο 2: Βιολογία Συστημάτων, Τεχνολογίες Omics

Γεώργιος Θεοδωρίδης

Σύνοψη

Στο παρόν κεφάλαιο γίνεται σύντομη περιγραφή του πεδίου της Βιολογίας Συστημάτων και των προσεγγίσεων μη-στοχευμένης ανάλυσης δειγμάτων για την εύρεση βιοδεικτών. Περιγράφονται σύντομα τα βασικά υποπεδία: Γονιδιωματική, Πρωτεομική, ενώ δίνεται πιο αναλυτική περιγραφή του πεδίου της Μεταβολομικής και των εφαρμογών αυτής. Περιγράφονται εφαρμογές στις επιστήμες της ζωής: προσεγγίσεις για την εύρεση βιοδεικτών στις κλινικές επιστήμες (διαγνωστική), βιοτεχνολογία, φυτική παραγωγή, ιχθυλασιμότητα, ενώ δίνονται συνοπτικά οι κύριες ανάγκες και εφαρμογές της Βιοανάλυσης.

Προαπαιτούμενη Γνώση

Πριν την ανάγνωση του βιβλίου χρήσιμη γνώση για τον αναγνώστη μπορεί να βρεθεί σε συγγράμματα *Ενόργανης Ανάλυσης*, *Βιοχημείας* και στο σύγγραμμα του μαθήματος *Ειδικές Μέθοδοι Διαχωρισμού και Χημικής Ανάλυσης*

2.1. Ολιστική Ανάλυση

«Παραδοσιακά», στόχος του Αναλυτικού Χημικού είναι να ανιχνεύει και να προσδιορίζει τα συστατικά διαφόρων δειγμάτων ή να ανιχνεύει και να ταυτοποιεί άγνωστες ενώσεις σε άγνωστα δείγματα. Μέχρι πρόσφατα οι τεχνολογικές δυνατότητες περιόριζαν την ανάλυση στη μέτρηση-προσδιορισμό μικρού αριθμού μορίων και ως εκ τούτου η σχετική βιβλιογραφία αναφέρεται κατά μεγάλο ποσοστό σε ειδικές μεθόδους, με ζητούμενα χαρακτηριστικά την ευαισθησία, την αποτελεσματικότητα, την ειδικότητα και την εκλεκτικότητα ως προς τον προσδιορισμό αυτών των ενώσεων. Κατά την ανάπτυξη νέων μεθόδων, στόχος είναι η απομόνωση των προσδιοριζόμενων συστατικών και η μείωση παρεμποδίσεων από όλα τα υπόλοιπα συστατικά του δείγματος. Τα τελευταία χρόνια, η ανάπτυξη των φασματοσκοπικών τεχνικών ανάλυσης και της υπολογιστικής ισχύος έχει κάνει εφικτή την προοπτική η οποία έως πρόσφατα θεωρούνταν ανέφικτη: την ολιστική ανάλυση, δηλαδή, την ανάλυση δειγμάτων με στόχο τη λήψη του ολικού «αποτυπώματος» του δείγματος: μια «ολική χαρτογράφηση» των συστατικών του. Ολιστική ανάλυση σημαίνει στην πράξη ότι επιλέγεται μια μέθοδος η οποία μπορεί να συλλέξει δεδομένα, πλούσια σε πληροφορίες, χωρίς αρχικές προεπιλογές για την ανίχνευση των ορισμένων συστατικών. Στόχος της μέτρησης είναι η ανίχνευση του μέγιστου δυνατού αριθμού από τα συστατικά ενός δείγματος. Η ειδικότητα στη μέτρηση και την απόκριση κάποιων συστατικών θεωρείται μη επιθυμητή. Τα δεδομένα αναλύονται με προηγμένα υπολογιστικά εργαλεία, όπως η στατιστική ανάλυση πολλαπλών μεταβλητών και τεχνικές όπως Principal Component Analysis (PCA), Partial Least Squares-Discriminant Analysis (PLS-DA) καθώς και άλλες. Οι κύριες αναλυτικές τεχνικές οι οποίες εφαρμόζονται στην ολιστική ανάλυση είναι οι τεχνικές συστοιχιών (array), η φασματοσκοπία πυρηνικού συντονισμού (NMR) και η φασματομετρία μαζών σε διάφορες μορφές της είτε μόνη της είτε σε συνδυασμό με υγρή χρωματογραφία (LC-MS), αέρια χρωματογραφία (GC-MS), ή τριχοειδή ηλεκτροφόρηση (CE-MS).

2.2. Βιολογία Συστημάτων

Πεδίο εφαρμογής των ολιστικών αναλυτικών τεχνολογιών είναι οι μελέτες των υπο-ενοτήτων της βιολογίας συστημάτων, τα οποία προσεγγίζονται με τις τεχνολογίες - omics: γονιδιωματική, πρωτεομική, μεταβολομική κλπ. Η Βιολογία Συστημάτων (BS) είναι η μελέτη της βιολογίας ως ολοκληρωμένο σύστημα γενετικών, πρωτεϊνικών, μεταβολικών και βιοσυνθετικών δράσεων, οι οποίες βρίσκονται σε αλληλεξάρτηση και συν-διακύμανση. Αντί να ξεκινούμε από συγκεκριμένη επιστημονική θεωρία (υπόθεση) όπου αναλύουμε περιορισμένα επιμέρους υποσύνολα και ανεξάρτητες συνιστώσες ενός οργανισμού (όπως το μεταβολισμό των σακχάρων, μικρό αριθμό ιντερλευκινών, ή την έκφραση ενός γονιδίου), στη BS προσπαθούμε να μετρήσουμε όλες τις συνιστώσες να τις συσχετίσουμε, ώστε να προσδιορίσουμε τις αλληλεπιδράσεις μεταξύ αυτών σε ένα οργανισμό και να οδηγηθούμε από τα δεδομένα σε νέες υποθέσεις.

Η κατανόηση της πορείας και του τελικού αποτελέσματος διαφόρων παρεμβάσεων σε ζώντες οργανισμούς μπορεί να είναι δύσκολη και αβέβαιη, όταν δεν έχουμε πλήρη γνώση των εμπλεκόμενων βιοσυνθετικών μηχανισμών. Η πορεία μιας συμβατικής μελέτης περιλαμβάνει την εκτέλεση διαφορετικών βιοχημικών/εργαστηριακών μετρήσεων που μας δίνουν περιορισμένη πληροφορία για την κατάσταση του υπό μελέτη συστήματος. Για παράδειγμα γίνονται κλινικές μετρήσεις στις επιστήμες της ζωής ή εκτιμήσεις ωριμότητας φρούτων στον τομέα τροφίμων, με μετρήσεις συγκεκριμένων μορίων στόχων με διαφορετικές τεχνολογίες, μεθόδους και πρωτόκολλα μέτρησης ή βαθμονόμησης. Με την τακτική αυτή, οι περισσότερες λειτουργίες παραμένουν ουσιαστικά, μερικώς, μόνο χαρτογραφημένες, οπότε προσπαθούμε να καταλάβουμε την πορεία των φαινομένων βασιζόμενοι σε διάσπαρτα ευρήματα.

Από τα παραπάνω γίνεται κατανοητό ότι η «παραδοσιακή» προσέγγιση της βιολογίας στοχεύει στη μελέτη της λειτουργίας ανεξάρτητων γονιδίων, πρωτεϊνών ή μεταβολιτών. Για παράδειγμα μια βιοτεχνολογική μελέτη μπορεί να στοχεύει στην αύξηση της παραγωγής ενός τελικού προϊόντος (π.χ. ενός αλκαλοειδούς με φαρμακευτική χρήση) από ένα φυτό ή κυτταρική καλλιέργεια. Η γνώση στην οποία βασιζόταν έως πρόσφατα μια τέτοια μελέτη περιοριζόταν στη μελέτη της επίδρασης σε ορισμένα σημεία της βιοσυνθετικής οδού, για παράδειγμα σε κάποια (συνήθως 2 ή 3) στάδια πριν από τη σύνθεση του επιθυμητού προϊόντος, εξετάζοντας περιορισμένα δεδομένα σχετικά με τη λειτουργία του εξεταζόμενου συστήματος, όπως την έκφραση μιας καταλυτικής πρωτεΐνης και τις συγκεντρώσεις μιας ή δύο πρόδρομων ενώσεων. Αντίστοιχα στις επιστήμες υγείας, οι δυνατότητες θεραπείας πολύπλοκων, πολυ-παραμετρικών ασθενειών όπως ο καρκίνος ή ο διαβήτης περιορίζονται, όταν μελετώνται μερικές μόνον παραμέτροι κάθε φορά (π.χ. τα επίπεδα μερικών γενικών δεικτών, όπως τρασφεράσες, ιντερλευκίνες κλπ.).

Σε αντίθεση με αυτήν την τακτική, στη ΒΣ γίνεται προσπάθεια κατανόησης των οργανισμών στο σύνολό τους, δηλαδή όχι ξεχωριστά ενός ενζύμου ή ενός γονιδίου αλλά συνδυαστική μελέτη της δράσης του ενζύμου ή της έκφρασης του γονιδίου με τη φυσιολογία του οργανισμού. Για παράδειγμα, το ανοσοποιητικό σύστημα ενός οργανισμού δεν είναι αποτέλεσμα απλών μηχανισμών ή γονιδίων και δεν μπορεί να περιγραφεί ικανοποιητικά με τη μέτρηση μόνον των συγκεντρώσεων ή της δραστηριότητας κάποιων ανοσοσφαιρινών. Στη ΒΣ το ανοσοποιητικό σύστημα θεωρείται ως σύνολο αλληλεπιδράσεων διαφορετικών γονιδίων, πρωτεϊνών, μηχανισμών και εξωτερικών παραγόντων με στόχο την επιτυχή αντιμετώπιση ασθενειών, επιθέσεων και μολύνσεων. Επομένως, για να γίνει πρόβλεψη της αντίδρασής του σε μια ορισμένη επέμβαση ή αγωγή (π.χ. μεταμόσχευση), οι πληροφορίες μπορεί να βρίσκονται κρυμμένες στο γονιδίωμα, το σύνολο των πρωτεϊνών ή το σύνολο των μεταβολιτών του οργανισμού ή στο συνδυασμό και τη σύνθεση των πληροφοριών.

Για παράδειγμα, στις εφαρμογές της ΒΣ στη διαγνωστική αναλύουμε βιολογικά υγρά ασθενών και υγιών οργανισμών, μετρώντας μη στοχευμένα, τα χιλιάδες βιομόρια που περιέχονται σε αυτά και προσπαθούμε να εντοπίσουμε τυχόν διαφορές στις συγκεντρώσεις των βιομορίων μεταξύ των πληθυσμών. Εάν βρεθούν διαφορές, προσπαθούμε στη συνέχεια να ταυτοποιήσουμε το/τα μόρια δείκτη (βιοδείκτης, biomarker) και να καταλάβουμε για ποιο λόγο αυτοί διαφέρουν ανάμεσα στις δύο ομάδες.

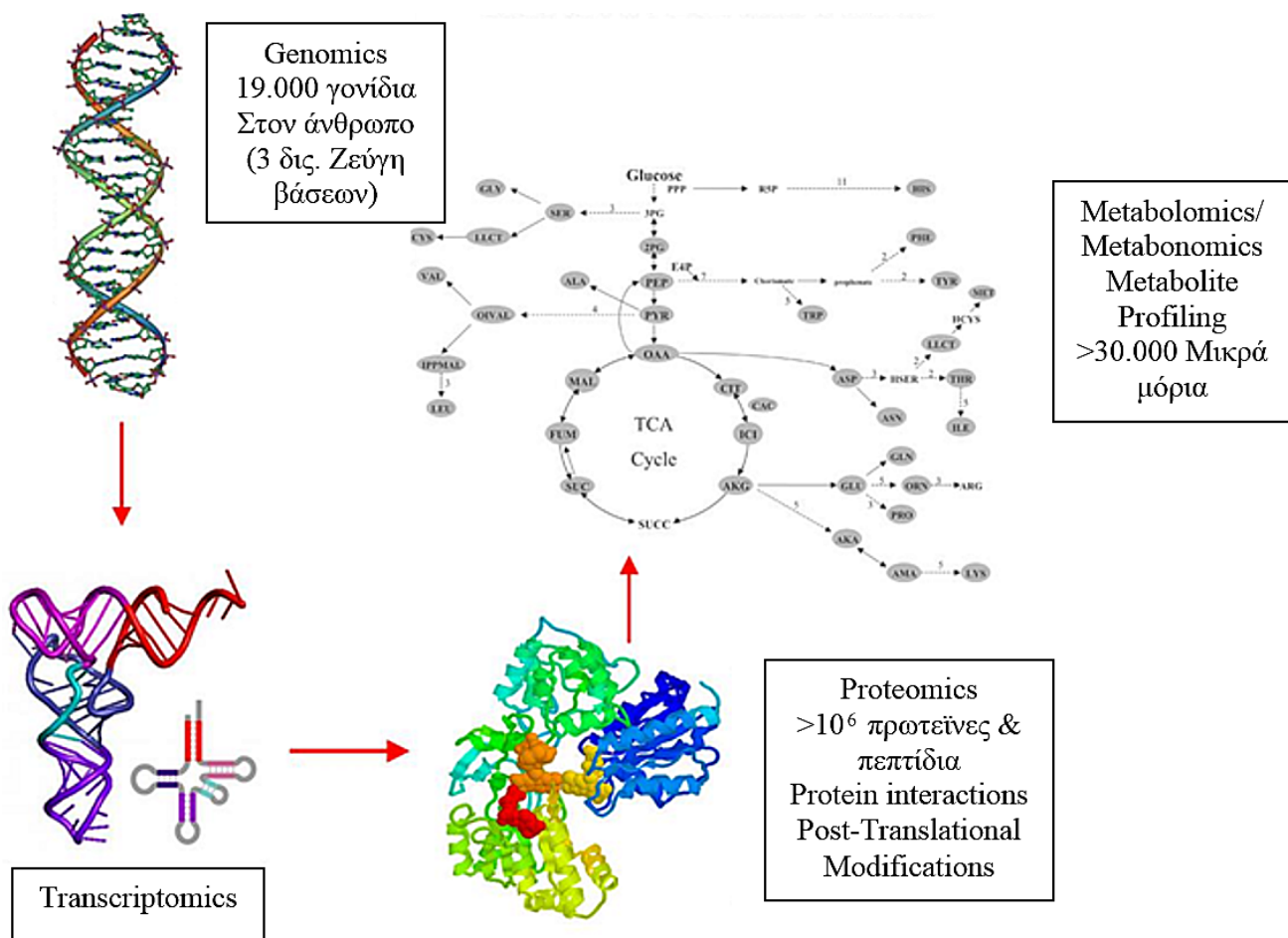
Αποκρυπτογραφώντας τις οδούς βιοσύνθεσης, οι ερευνητές μπορούν να προβλέψουν την επίδραση διαφόρων ερεθισμάτων (εξωγενών ουσιών, ασθένειας, γενετικής τροποποίησης) σε κάποιον οργανισμό. Με τη συσσώρευση αδιαμφισβήτητων δεδομένων, την ολοκλήρωση της περιγραφής των βιοσυνθετικών οδών και την αποκωδικοποίηση των μηχανισμών, τα μοντέλα πρόβλεψης γίνονται πιο ακριβή και αξιόπιστα, γεγονός που οδηγεί την έρευνα και ανάπτυξη πιο κοντά στην επίλυση περίπλοκων βιολογικών προβλημάτων.

Λόγω της φύσης της μη-στοχευμένης ανάλυσης, η βιολογία συστημάτων επεξεργάζεται μεγάλες ποσότητες δεδομένων που περιγράφουν ένα ή περισσότερα συστήματα. Για να γίνει εκμετάλλευση της πληροφορίας από τέτοια περίπλοκα και ογκώδη δεδομένα, πρέπει αυτά να μελετηθούν συστηματικά με προηγμένες μαθηματικές μεθόδους και στατιστικά εργαλεία, τα οποία αναδεικνύουν και προβάλλουν με αποτελεσματικό τρόπο τα αποτελέσματα. Ο συνδυασμός επιστημών, όπως βιολογία, ιατρική, χημεία, μαθηματικά βιοστατιστική/βιοπληροφορική, δίνει τη δυνατότητα αποτελεσματικότερης αντιμετώπισης τέτοιων πολύπλοκων προβλημάτων. Επιπλέον, η ΒΣ μπορεί να συνδυάσει ευρήματα από την ανάλυση γονιδίων, πρωτεϊνών, μεταβολιτών μεταξύ τους αλλά και με τη φυσιολογία (κλινική εικόνα) του υπό μελέτη συστήματος και να επιτρέψει την καλύτερη κατανόηση των πραγματικών μηχανισμών των βιολογικών δράσεων και την πρόβλεψη της πορείας του μελετώμενου οργανισμού.

Η ΒΣ αναδύθηκε μέσα από την πρόοδο της μελέτης του ανθρώπινου γονιδιώματος και την κατανόηση των μηχανισμών από το σύστημα των γονιδίων, πρωτεϊνών ως τη σύνθεση και αλληλεπίδραση βιομορίων. Ο χώρος των διαφόρων πεδίων -omics προϋπήρχε της ΒΣ, αλλά χαρακτηριζόταν από διάσπαση: διάσπαρτες μετρήσεις, μικρή συνέργεια στη μελέτη μεταξύ επιπέδων της γονιδιακής έκφρασης (συνδυασμός γονιδιακής με πρωτεϊνική πληροφορία). Η συνδυασμένη μελέτη εν γένει των διαφόρων πεδίων – omics (**Σχήμα 2.1**)

συσσωρεύει νέα γνώση και επιτρέπει τη μελέτη αυτής με αποτελεσματικό τρόπο, ώστε να αποκαλύψει «κρυμμένη» πληροφορία η οποία μπορεί να οδηγήσει στην εύρεση βιοδεικτών για την ανάπτυξη νέων φαρμάκων και τη διάγνωση ασθενειών (μεταξύ άλλων εφαρμογών). Με τις προσεγγίσεις της ΒΣ δε συσχετίζουμε απλώς τις παρατηρήσεις μεταξύ τους αλλά, ξεκινώντας, συχνά, χωρίς κάποια θεωρία, δομείται το βιολογικό σύστημα με βάση μοριακές πληροφορίες και αδιαμφισβήτητα δεδομένα: τα δεδομένα οδηγούν στη θεωρία κι όχι το αντίθετο.

Οι εφαρμογές της νέας γνώσης η οποία παράγεται είναι πολλές και πολύ σημαντικές. Για παράδειγμα, στην ανάπτυξη φαρμάκων πολλές ενώσεις οι οποίες είναι σε στάδιο υποψήφιας φαρμακευτικής ένωσης (candidate) μπορούν να απορριφθούν σε αρχικό στάδιο μελέτης, μειώνοντας έτσι το χρόνο και το κόστος ανάπτυξης νέων φαρμάκων. Η απόρριψη μπορεί να βασιστεί στην πρόβλεψη τοξικής δράσης της υποψήφιας φαρμακευτικής ένωσης ή στη μικρή αποτελεσματικότητά της δράσης του. Παράδειγμα σχεδιασμού τέτοιας μελέτης δίνεται συνοπτικά στην παράγραφο 2.4.



Σχήμα 2.1. Τα τέσσερα βασικά πεδία της Βιολογίας συστημάτων και η σχετική σύνδεσή τους

2.3. Πεδία της Βιολογίας Συστημάτων

Τα επιμέρους πεδία της βιολογίας συστημάτων δίνονται στο Σχήμα 2.1 και είναι:

Γονιδιωματική: Η μελέτη του συνόλου των γονιδίων ενός οργανισμού. Ο καθορισμός της δομής, της αλληλουχίας των γονιδίων και η διαλεύκανση της λειτουργίας τους και του ρόλου τους. Κατανοώντας τη βιολογική διαφοροποίηση σε επίπεδο γονιδιώματος, μπορούμε να βγάλουμε συμπεράσματα για την προέλευση ατομικών χαρακτηριστικών και τάσεις εκδήλωσης νόσων. Η γονιδιωματική είναι το πιο αναπτυγμένο και ώριμο πεδίο της ΒΣ. Έχει, ήδη, ολοκληρωθεί η χαρτογράφηση του γονιδιώματος σε διάφορους οργανισμούς, ενώ πλέον υπάρχει μεγάλος αριθμός εμπορικών εταιριών που προσφέρουν

υπηρεσίες πλήρους Γονιδιωματικής ανάλυσης. Οι τεχνικές έχουν, ουσιαστικά, φτάσει σε τέτοιο τεχνολογικό επίπεδο, ώστε στόχος της έρευνας και ανάπτυξης στο πεδίο είναι η ταχύτερη και οικονομικότερη ανάλυση.

Παρόλο που τα άτομα ενός είδους, είναι πολύ όμοια σε γενετικό επίπεδο (σε ποσοστό κοντά στο 99.9%), υπάρχουν διαφορές στην αλληλουχία του σε συχνότητα περίπου 1 σε κάθε 300 βάσεις νουκλεοτιδίων. Λαμβάνοντας υπόψη ότι το σύνολο των βάσεων (που είναι κοντά στα 3 δισεκατομμύρια στον άνθρωπο), το ποσοστό μεταφράζεται σε 10.000.000 βάσεις οι οποίες διαφέρουν μεταξύ των ατόμων. Οι διαφοροποιήσεις μεμονωμένων βάσεων ονομάζονται Single Nucleotide Polymorphisms (SNPs) και η μελέτη τους αποτελεί ένα μεγάλο τμήμα της γονιδιωματικής.

Βασικά εργαλεία της έρευνας αποτελούν οι DNA μικροδιατάξεις (μικροσυστοιχίες), οι οποίες εφαρμόζονται 1) στη χαρτογράφηση του γονιδιώματος ή γενοτύπηση (Genome-wide genotyping) που εξετάζει ποια γονίδια υπάρχουν σε ένα άτομο, 2) στη μελέτη της γονιδιακής έκφρασης στους διάφορους ιστούς (Tissue-specific gene expression) η οποία μελετά ποια γονίδια χρησιμοποιούνται για την παραγωγή πρωτεϊνών και 3) στη μελέτη μεταλλάξεων (Mutational analysis).

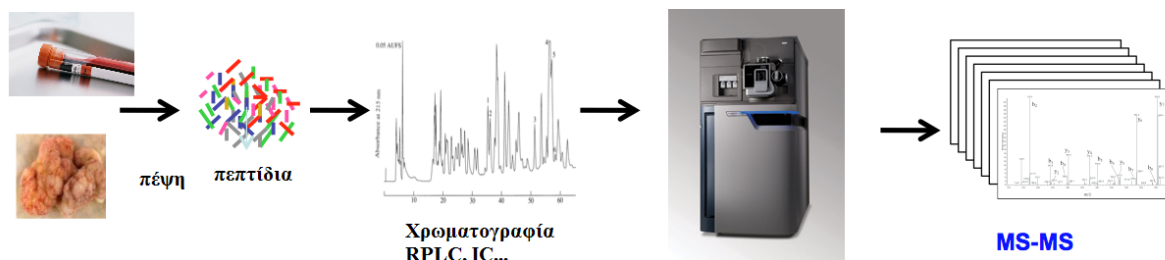
Ένα σύγχρονο πεδίο που αναπτύσσεται τα τελευταία χρόνια είναι η Επιγενετική. Αντικείμενό της είναι η μελέτη των (ενίοτε) κληρονομήσιμων αλλαγών στη λειτουργία των γονιδίων, λόγω κυτταρικών και φαινοτυπικών μεταβολών. Είναι η μελέτη της επίδρασης στη δράση ενός γονιδίου (π.χ. από λειτουργικό σε μη-λειτουργικό), χωρίς να αλλάζει η αλληλουχία των βάσεων στο DNA του γονιδίου. Οι επιγενετικοί παράγοντες οδηγούν κύτταρα στη διαφοροποίηση και στην εξέλιξή τους σε διάφορους ιστούς, παρότι τα κύτταρα έχουν το ίδιο DNA. Οι επιγενετικοί μηχανισμοί (μεθυλίωση βάσεων, ακετυλίωση ιστονών) παραμένουν σημαντικοί σε όλη τη διάρκεια της ζωής. Μπορεί να εκκινούν επιλεκτικά κάποια γονίδια σε διαφορετικούς ιστούς ως απάντηση σε περιβαλλοντικά και άλλα ερεθίσματα.

Μεταγραφωματική (Transcriptomics) είναι η μελέτη του συνόλου των μεταγραφωμάτων (mRNA) τα οποία παράγονται σε ένα κύτταρο ή είδος κυττάρων ή οργανισμό. Η ανάλυση γίνεται με μεθόδους υψηλής απόδοσης, όπως αυτές οι οποίες χρησιμοποιούνται στη γονιδιωματική, για παράδειγμα η ανάλυση μικροσυστοιχιών. Σύγκριση των συνόλων των μεταγραφωμάτων επιτρέπει την ταυτοποίηση των γονιδίων τα οποία εκφράζονται διαφορετικά σε διακριτούς πληθυσμούς κυττάρων, ή για παράδειγμα την απόκριση σε διαφορετικές θεραπείες· αναφέρεται και ως expression profiling.

Η Πρωτεομική (proteomics) πραγματεύεται την ανάλυση του πρωτεώματος, δηλαδή, του συνόλου των πρωτεϊνών και πεπτιδίων τα οποία κωδικοποιούνται και παράγονται από το γονιδίωμα ενός οργανισμού. Στόχοι της είναι 1) ο διαχωρισμός, η ταυτοποίηση, ο χαρακτηρισμός και η αλληλούχιση πρωτεϊνών, 2) η μελέτη αλληλεπίδρασης πρωτεϊνών σε σύμπλοκα και 3) η ποσοτικοποίησή τους και συσχέτιση με την κυτταρική λειτουργία. Η πρωτεομική χρησιμοποιεί μια ποικιλία τεχνικών, για να επιτύχει τους στόχους της όπως η ηλεκτροφόρηση πηκτής (διαχωρισμός), η τριχοειδής ηλεκτροφόρηση (διαχωρισμός) και η υγρή χρωματογραφία σε συνδυασμό με τη φασματομετρία μάζας (διαχωρισμός και ποσοτικοποίηση), η φασματομετρία μάζας σε σειρά (αλληλούχιση), η φασματοσκοπία NMR (αλληλεπιδράσεις πρωτεϊνών), και η κρυσταλλογραφία (μελέτη τρισδιάστατης δομής). Στο **Σχήμα 2.2**, δίνεται η πορεία μιας πρωτεομικής μελέτης. Η πρωτεομική αντιπροσωπεύει το πιο δυναμικό τμήμα των ολιστικών αναλυτικών τεχνολογιών και ως εκ τούτου μεγάλος αριθμός ερευνητών και χρηματοδοτούμενης έρευνας εντοπίζεται στην πρωτεομική. Η πρωτεομική προσφέρει σημαντικές πληροφορίες για τη γονιδιακή έκφραση και την κατανόηση βιοχημικών μηχανισμών οι οποίες έχουν άμεση εφαρμογή και δυνητικό αποτέλεσμα στη μελέτη επιστημών ζωής (γνώση στο μηχανισμό εκδήλωσης ασθένειας) αλλά και στις μοριακές επιστήμες φυτών ή επιστήμες τροφίμων κ.ά.

Μεταβονομική ή μεταβολομική (metabonomics/metabolomics) είναι η ποσοτική μέτρηση της αντίδρασης των ζωντανών συστημάτων σε παθοφυσιολογική διέγερση ή γενετική τροποποίηση μέσω της μέτρησης του προφίλ των μικρών μορίων (μεταβολιτών, όπως οργανικά οξέα, αμινοξέα, σάκχαρα, νουκλεοτίδια κ.ά.). Ουσιαστικά είναι η ολιστική μελέτη της σύστασης, δυναμικής, και των αλληλεπιδράσεων των μεταβολιτών, σε σχέση με παρεμβολές ή αλλαγές στο περιβάλλον τους, σε κύτταρα, ιστούς και βιολογικά υγρά. Οι μεταβολίτες μπορούν να χωριστούν σε δύο κατηγορίες: τους πρωτογενείς και τους δευτερογενείς. Οι πρωτογενείς μεταβολίτες συνδέονται άμεσα με την ανάπτυξη και την αναπαραγωγή των οργανισμών, ενώ οι δευτερογενείς δε συνδέονται άμεσα με αυτήν τη διαδικασία, ωστόσο παίζουν σημαντικό ρόλο στις αλληλεπιδράσεις του οργανισμού με το περιβάλλον του. Αποτελούν τα αντιδρώντα και προϊόντα του μεταβολισμού και γι' αυτό οι μεταβολές των συγκεντρώσεων συγκεκριμένων ομάδων μεταβολιτών μας δίνει πληροφορίες της αντίδρασης των συστημάτων σε διάφορες παρεμβολές (περιβαλλοντικές, γενετικές ή άλλες).

Η μελέτη των παρεμβολών μπορεί να γίνει σημαντικό εργαλείο για το χαρακτηρισμό σύνθετων φαινοτύπων καθώς και την ανακάλυψη βιοδεικτών (biomarkers) για συγκεκριμένες φυσιολογικές αντιδράσεις.



Ποσοτική Ανάλυση

Μέτρηση της "αφθονίας κορυφής"
Μετρήσεις του σήματος κάθε πεπτιδίου
Στατιστική ανάλυση

Protein	Ratio	Quasi.fdr	Counts	Spec1 Rep1	Spec1 Rep2	Spec1 Rep3	Spec2 Rep1	Spec2 Rep2	Spec2 Rep3
MX2	-37.59	0.0000	47	0	0	0	16	15	16
IGF2BP3	-36.52	0.0010	61	0	0	0	18	22	21
ACSS1	-36.40	0.0004	56	0	0	0	20	17	19
HSD17B4	-36.17	0.0012	48	0	0	0	17	14	17
AGPS	-35.50	0.0009	30	0	0	0	10	9	11
MMP1	-35.50	0.0009	30	0	0	0	10	11	9
IFIT2	-35.11	0.0037	23	0	0	0	9	7	7
COL6A3	-7.090	0.0040	549	2	1	1	186	206	153
MVP	-5.864	0.0047	237	1	2	1	81	87	65
OAS3	-5.554	0.0039	96	1	0	1	32	33	29
KRT6B	4.115	0.0042	110	36	36	32	1	2	3
MSH2	4.432	0.0018	113	38	37	33	2	1	2
POGZ	34.19	0.0040	33	12	12	9	0	0	0
WDR70	34.76	0.0047	18	7	6	5	0	0	0
RFC4	34.98	0.0031	21	6	8	7	0	0	0
UGT8	36.73	0.0002	26	9	8	9	0	0	0
LSP1	37.22	0.0004	99	31	36	32	0	0	0

Επεξεργασία δεδομένων
Αναζήτηση σε βάσεις δεδομένων
Ταυτοποίηση πεπτιδίων
Κανονικοποίηση
Σύγκριση ομάδων...

Η ταυτοποίηση βασίζεται στη μέτρηση της μάζας του πεπτιδίου, στη μέτρηση της μάζας των θραυσμάτων και η αλληλούχιση γίνεται με βοήθεια κατάλληλου λογισμικού

Σχήμα 2.2. Η πορεία τυπικής μελέτης πρωτεομικής. Αίμα ή άλλο δείγμα από τον υπό μελέτη οργανισμό υπόκειται σε κατεργασία για παραλαβή των πρωτεϊνών και στη συνέχεια ακολουθεί διαχωρισμός των πρωτεϊνών με ηλεκτροφόρηση ή χρωματογραφία. Επιλέγεται μια πρωτεΐνη η οποία υφίσταται πέψη με πρωτεολυτικό ένζυμο, για διάσπασή της σε πεπτίδια. Τα πεπτίδια αναλύονται σε σύστημα LC-MS(/MS) όπου επιτυγχάνεται, σε διαφορετικά στάδια, διαχωρισμός, ταυτοποίηση, αλληλούχιση και ποσοτικοποίηση των πεπτιδίων.

Κύριες τεχνικές ανάλυσης είναι οι φασματοσκοπικές (φασματοσκοπία NMR, φασματομετρία μάζας συζευγμένη με χρωματογραφία) ενώ τεχνικές λεπτομέρειες με παραδείγματα δίνονται στη συνέχεια του κεφαλαίου. Τα τελευταία χρόνια αναπτύσσεται η απεικόνιση με MS (MS imaging), που περιγράφεται εν συντομία στο κεφάλαιο 12.

Πέραν αυτών των βασικών πεδίων, υπάρχει μια σειρά δραστηριοτήτων, όπως λιπιδιομική, γλυκομική, μεταλομική, φαρμακογενομική. Η μεταλομική αναπτύσσεται στο κεφάλαιο 5. Η λιπιδιομική αναπτύσσεται, ταχέως, και εστιάζει στη χαρτογράφηση των λιπιδίων ενός κυττάρου, οργάνου ή οργανισμού. Πραγματοποιείται, κυρίως, με χρήση LC-MS, GC-MS, στοχεύοντας στην ανάλυση ιστών ή αίματος για τον προσδιορισμό φωσφολιπιδίων (φωσφατιδυλοχολινών), γλυκεριδίων, χολικών οξέων, ελεύθερων λιπαρών οξέων, ακυλοκαρνιτινών και άλλων ομάδων λιπόφιλων ενώσεων. Έως τις μέρες μας αρκετές εκατοντάδες διακριτών λιπιδίων έχουν ανιχνευθεί σε διάφορα δείγματα προερχόμενα από ζώα ή φυτά.

2.4. Βιοαναλυτικές προκλήσεις στα πεδία της Βιολογίας Συστημάτων

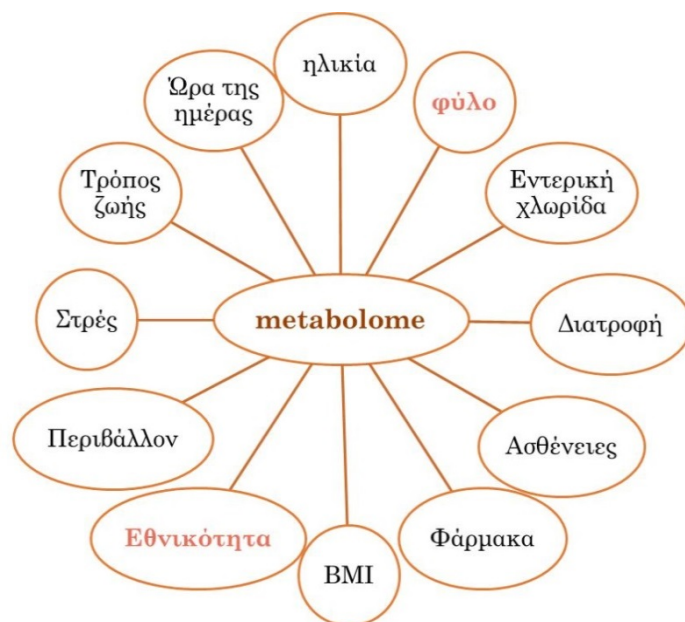
Η ολιστική αντίληψη της αναλυτικής διαδικασίας είναι κοινή ανάμεσα στα διάφορα πεδία omics, όμως, η φύση των πεδίων αυτών είναι διαφορετική και γι' αυτόν το λόγο οι βιοαναλυτικές προκλήσεις και πιθανά προβλήματα που καλείται να λύσει ο ερευνητής είναι διαφορετικά σε κάθε πεδίο.

Τα κύτταρα ενός οργανισμού έχουν κοινό γονιδιακό φορτίο το οποίο μπορεί να προσδιοριστεί σε διάφορα σημεία του οργανισμού (π.χ. διαφορετικοί ιστοί και βιολογικά υγρά του ανθρώπου σίελο, αίμα κ.ά.). Όμως τα κύτταρα διαφοροποιούνται ανάλογα με το υποσύνολο των γονιδίων που είναι ενεργά και κωδικοποιούν διαφορετικές πρωτεΐνες οι οποίες καθορίζουν την κυτταρική λειτουργία. Επομένως, το σύνολο των πρωτεϊνών μπορεί να αλλάζει δραματικά από ιστό σε ιστό. Επίσης, η κατανομή των πρωτεϊνών ή των μεταβολιτών είναι διαφορετική μεταξύ ιστών, υγρών ή άλλων δειγμάτων. Έτσι το πρωτόμα και το μεταβόλωμα του ήπατος, για παράδειγμα, είναι τελείως διαφορετικά από τα αντίστοιχα σύνολα σε εγκέφαλο,

αίμα και ούρα. Άρα διαφορετικά δείγματα από τον ίδιο οργανισμό θα δώσουν διαφορετικά μεταβολικά ή πρωτεϊνικά προφίλ.

Το γονιδίωμα ενός οργανισμού είναι σχετικά σταθερό, με την έννοια ότι δεν παρατηρούνται αλλαγές στη διάρκεια μιας ημέρας, ενώ σε μεγαλύτερα χρονικά διαστήματα αλλαγές και (επι)γενετικές τροποποιήσεις είναι σχετικά περιορισμένες. Το γεγονός αυτό επιτρέπει (ανάμεσα στα άλλα) την πραγματοποίηση ελέγχων πατρότητας ή δικανικών ελέγχων ακόμη σε πολύ παλαιά δείγματα (δεκαετιών): το γονιδιακό φορτίο αίματος ή άλλων δειγμάτων είναι «στατικό» και προσφέρει τη δυνατότητα ταυτοποίησης ατόμων (π.χ. ύποπτοι συμμετοχής ή θύματα εγκληματικής ενέργειας τα οποία ταυτοποιούνται από το DNA τους) και διερεύνησης της γενετικής σύνδεσής τους. Σε αντίθεση, το πρωτέωμα ή το σύνολο των μεταβολιτών ενός δείγματος, αλλάζει δραστικά στη διάρκεια ακόμη και μιας ημέρας ακολουθώντας το βιολογικό ρολόι του οργανισμού. Οι συγκεντρώσεις σε ορμόνες, σάκχαρα, αμινοξέα, οργανικά οξέα και άλλα μόρια μπορεί, επίσης, να αυξομειωθούν δραστικά (σε κάποιες περιπτώσεις πάνω από δύο τάξεις μεγέθους) στα διάφορα βιολογικά δείγματα λόγω ήπιας επίδρασης, όπως φυσική άσκηση, διατροφή, κάπνισμα ή άλλη δραστηριότητα. Οι παράγοντες που επηρεάζουν το μεταβολικό προφίλ δίνονται παραστατικά στο **Σχήμα 2.3**. Η μεταβλητότητα αυτή επιβάλλει ιδιαίτερη επιμέλεια στο σχεδιασμό πειραμάτων και δυσχεραίνει την εξαγωγή ασφαλών συμπερασμάτων.

Πέραν αυτού, το πρωτέωμα και το μεταβόλωμα είναι πιο περίπλοκα από το γονιδίωμα. Ο αριθμός των γονιδίων είναι μικρότερος από τον αριθμό πρωτεϊνών ή μεταβολιτών (19.000-20.000 περίπου γονίδια στον άνθρωπο, >1.000.000 πρωτεΐνες και άγνωστος αριθμός μεταβολιτών, με εκτιμήσεις που κυμαίνονται από 30.000 έως >100.000). Τα γονίδια σχηματίζονται από DNA, το οποίο περιέχει τέσσερις επαναλαμβανόμενες αζωτούχες βάσεις: την αδενίνη, τη θυμίνη, τη γουανίνη και την κυτοσίνη. Η αλληλουχία αυτών των βάσεων καθορίζει την αλληλουχία των 20 αμινοξέων κατά τη μετάφραση. Ενώ, όμως, η αλληλούχιση του DNA προσφέρει γνώση του λειτουργικού γονιδιώματος, η γνώση της αλληλουχίας των αμινοξέων δε συνεπάγεται γνώση της τρισδιάστατης δομής μιας πρωτεΐνης, της συγκέντρωσης ή των τάσεων αυτής, της λειτουργίας της πρωτεΐνης και των αλληλεπιδράσεών της με άλλες πρωτεΐνες. Επιπλέον, οι πρωτεΐνες τροποποιούνται ομοιοπολικά μετά τη μετάφραση με μια μεγάλη ποικιλία τρόπων, δυσχεραίνοντας ακόμη περισσότερο την ολοκλήρωση της χαρτογράφησης του πρωτεώματος.



Σχήμα 2.3. Παράγοντες που επηρεάζουν το μεταβολικό προφίλ. Οι δύο παράγοντες με κόκκινο χρώμα είναι γενετικά καθορισμένοι. Το μεταβόλωμα είναι προϊόν δυναμικής αλληλεπίδρασης μεταξύ γενετικής επίδρασης, περιβαλλοντικής επίδρασης, τρόπου ζωής και άλλων παραγόντων.

Στο επίπεδο των μικρών μορίων, οι διάφορες οργανικές ενώσεις (μεταβολίτες) οι οποίες συνθέτουν το μεταβόλωμα, παρουσιάζουν πολύ διαφορετικές φυσικοχημικές ιδιότητες: από πολύ πολικές (οξέα, βάσεις) έως πολύ άπολες (λιπίδια). Παρουσιάζουν μεγάλες διαφορές στη μοριακή τους δομή, πολλές πιθανές

ισομέρειες (π.χ. πολλά σάκχαρα με μοριακή μάζα 180 amu) και απαντώνται στα διάφορα βιολογικά δείγματα σε πολύ μεγάλο εύρος συγκεντρώσεων το οποίο να φτάνει τις 7 τάξεις μεγέθους.

Χαρακτηριστικό παράδειγμα των βιοαναλυτικών προκλήσεων είναι η διερεύνηση της δομής άγνωστων μορίων στη λιπιδιομική, λόγω των πολλών πιθανών ισομερών. Τα γλυκεροφωσφολιπίδια αποτελούνται από ένα σκελετό γλυκερόλης, όπου συνδέονται αλκυλο-, ακυλο- ή αλκενυλομάδες. Αυτές οι αλυσίδες είναι συχνά διαφορετικού μήκους ανθρακικής αλυσίδας, διαφορετικού βαθμού κορεσμού και διαφορετικής θέσης των διπλών δεσμών. Αυτή η πολυπλοκότητα έχει ως αποτέλεσμα να υπάρχουν 10 ή περισσότερες πιθανές διαφορετικές μοριακές δομές για έναν ευρεθέντα μοριακό τύπο. Τελικά, λόγω των πολλών παραλλαγών, η ταυτοποίηση της πραγματικής δομής μπορεί να μη γίνεται πάντα εφικτή. Περαιτέρω παράγοντες που μπορεί να δημιουργήσουν προβλήματα στη μεταβολομική και μπορεί να προέρχονται από τη συλλογή και κατεργασία δείγματος, αναφέρονται στην αρχή της παραγράφου 2.5.2.

Στο πεδίο της πρωτεομικής, περιορισμοί και προβλήματα που παραμένουν για επίλυση είναι τα:

- 1) Το υψηλό κόστος της ανάλυσης και τα χρονοβόρα πειράματα.
- 2) Η αδυναμία ανάλυσης πρωτεϊνών οι οποίες βρίσκονται σε πολύ χαμηλή συγκέντρωση.
- 3) Ο μέγιστος αριθμός πεπτιδίων ο οποίος μπορεί να ανιχνευθεί σε μια ανάλυση είναι της τάξης των 5.000 πεπτιδίων, που είναι πολύ μικρό ποσοστό του αναμενόμενου συνόλου (περί το 10⁶). Σημειωτέον ότι μια τέτοια ανάλυση απαιτεί χρονοβόρα ανάπτυξη μεθόδου, χρονοβόρα κατεργασία δείγματος από πολύ εξειδικευμένο αναλυτή και τα πιο εξελιγμένα αναλυτικά συστήματα. Μετά την ανάπτυξη, μια ανάλυση απαιτεί 6-7 ώρες τουλάχιστον ανά δείγμα.
- 4) Προβλήματα με τον καθαρισμό και τη σταθερότητα των πρωτεϊνών.
- 5) Προβλήματα στο διαχωρισμό πρωτεϊνών οι οποίες είναι είτε πολύ μικρές είτε πολύ μεγάλες, είτε πολύ υδρόφιλες, είτε πολύ υδρόφοβες.

Για τους παραπάνω λόγους, η τεχνολογική ολοκλήρωση της χαρτογράφησης του πρωτεώματος ή του μεταβολώματος δεν είναι ακόμη εφικτή και υπάρχει μεγάλη ανάγκη για ανάπτυξη νέων μεθόδων. Για παράδειγμα, η χαρτογράφηση του μεταβολώματος με τεχνική RPLC-MS προσφέρει κάλυψη ενός μόνο ποσοστού. Η χαρτογράφηση του μεταβολώματος με GC-MS μπορεί να προσφέρει κάλυψη ενός ακόμη μικρότερου ποσοστού. Τα δύο αυτά υποσύνολα μπορεί να αλληλεπικαλύπτονται κατά 30-50%· αυτό σημαίνει ότι με τη χρήση και των δυο μεθόδων συνδυαστικά, μπορεί ο αναλυτής να ανιχνεύσει ένα άγνωστο ποσοστό του συνόλου των μεταβολιτών οι οποίοι πραγματικά υπάρχουν στο δείγμα. Η αβεβαιότητα αυτή ακόμη και στην εκτίμηση του ποσοστού κάλυψης του συνόλου, καταδεικνύει την ανάγκη ανάπτυξης του πεδίου. Η κάλυψη, δηλαδή η ανίχνευση και ημι-ποσοτική ανάλυση των ενώσεων εξαρτάται και από άλλους παράγοντες, για παράδειγμα τον τρόπο κατεργασίας του δείγματος. Αν για παράδειγμα γίνει εκχύλιση με οργανικό διαλύτη πριν την ανάλυση, υπάρχει σοβαρή πιθανότητα μη παραλαβής πολικών οργανικών μορίων. Αν γίνει ρύθμιση του pH, ανάλογα με την τελική τιμή pH, υπάρχει σοβαρή πιθανότητα να μην παραληφθούν ορισμένες ιονικές ή ιονιζόμενες ενώσεις. Όλα αυτά τα προβλήματα συμβάλλουν στο να παραμένει η τεχνολογία ατελής προς το παρόν και δείχνουν την ανάγκη για ανάπτυξη νέων πιο αποτελεσματικών μεθόδων.

2.5. Εφαρμογές της βιολογίας συστημάτων

Οι εφαρμογές που έχουν τα πεδία της ΒΣ είναι μεγάλης σημασίας και πολλές. Η χαρτογράφηση των επιμέρους συστημάτων προσφέρει νέα γνώση σε μοριακό πεδίο για τους μηχανισμούς της ζωής και τις αιτίες διάφορων φαινομένων (ανάπτυξη, γήρανση, ασθένεια, τοξικότητα κ.ά.). Η παρέμβαση βασίζεται, πλέον, σε δεδομένα και γνώση της αρχικής αιτίας (evidence-based), σε αντίθεση με την έως πρόσφατα αντιμετώπιση, κυρίως, των συμπτωμάτων (φλεγμονή, πόνος ή πυρετός). Για παράδειγμα, εφόσον γίνει γνωστή μια γενετική τροποποίηση η οποία οδηγεί σε λάθος κωδικοποίηση και παραγωγή λανθασμένης πρωτεϊνικής αλληλουχίας, στόχος της θεραπευτικής αντιμετώπισης γίνεται η παρεμπόδιση της μετάφρασης ή η απενεργοποίηση (μπλοκάρισμα) της μη κανονικής πρωτεΐνης, ώστε να σταματήσει ο μηχανισμός της ασθένειας στην αρχή του.

Η ΒΣ έχει χρησιμοποιηθεί σε διαφορετικούς τομείς των βιολογικών και των κλινικών επιστημών και σχετικά παραδείγματα αναφέρονται παρακάτω:

1. Κλινική διάγνωση και πρόγνωση ασθενειών. Η χρήση της ΒΣ στην κλινική διαγνωστική εστιάζεται στην εύρεση βιοδεικτών. Βιοδείκτες ασθένειας αποκαλούνται τα βιομόρια τα οποία παρουσιάζουν διαφοροποίηση στη συγκέντρωσή τους κατά την εκδήλωση και πορεία μιας ασθένειας.

Η εύρεση βιοδεικτών (οι βιοδείκτες αναφέρονται και στο κεφάλαιο 15) μπορεί να βοηθήσει στην κατανόηση των μηχανισμών ασθένειας, να οδηγήσει στην ανακάλυψη νέας γενιάς φαρμάκων, τα οποία και θα καταπολεμούν την αιτία της ασθένειας, και να βοηθήσει στην άμεση και αποτελεσματική παρακολούθηση της πορείας θεραπείας και ανταπόκρισης σε φαρμακευτική αγωγή. Είναι γνωστό ότι, συχνά, η αρχή της ασθένειας προηγείται πολλούς μήνες ή και χρόνια (όπως π.χ. σε νευροεκφυλιστικές παθήσεις) της εκδήλωσης συμπτωμάτων και της τελικής διάγνωσης.

Η έγκαιρη και ασφαλής διάγνωση βασισμένη σε απόλυτα στοιχεία (όπως συγκεντρώσεις επικυρωμένων βιοδεικτών) και όχι σε μόνο σε κλινικές εξετάσεις οι οποίες κάποιες φορές μπορεί να είναι υποκειμενικές, μπορεί να βοηθήσει στην έγκαιρη αντιμετώπιση της νόσου, την προετοιμασία και τον καλύτερο προγραμματισμό της ζωής του ασθενούς και της οικογένειάς του. Έπειτα η συνεχής παρακολούθηση των συγκεντρώσεων μοριακών δεικτών μπορεί να βοηθήσει στη λήψη αποφάσεων σε κλινικό επίπεδο, για παράδειγμα στην εξατομικευμένη θεραπευτική, στην αναγνώριση της απόκρισης ή μη του ασθενούς σε αγωγή, πολύ νωρίτερα από τις όποιες αλλαγές στη φυσιολογία. Για παράδειγμα, η μέτρηση βιοδεικτών απόκρισης σε θεραπεία μπορεί να γίνει μέσα σε λίγα λεπτά. Αντίθετα, η αλλαγή στην κλινική εικόνα λόγω της ανταπόκρισης στη θεραπεία μπορεί να απαιτεί μέρες, γεγονός που μπορεί να έχει μεγάλο κόστος αλλά και να επιδράσει ακόμη και στην υγεία και τη ζωή του ασθενούς ο οποίος δεν ανταποκρίνεται στην αγωγή και παραμένει υπό μη αποτελεσματική αγωγή.

Στην έρευνα για βιοδείκτες, η επιλογή των ομάδων και ο πειραματικός σχεδιασμός πρέπει να ακολουθεί κανόνες και σαφή κριτήρια επιλεξιμότητας ή απόρριψης των ατόμων/πειραματόζωων. Επίσης, είναι σημαντικό να συλλέγονται δεδομένα, όπως διατροφικά, στυλ ζωής, συνήθειες όπως κατανάλωση αλκοόλ ή κάπνισμα, δημογραφικά στοιχεία, δείκτης μάζας/σώματος, στοιχεία φαρμακευτικής αγωγής κ.ά. τα οποία μπορεί στη συνέχεια να χρησιμοποιηθούν για να δοθεί ερμηνεία σε ευρήματα, όπως μη αναμενόμενες ομαδοποιήσεις δειγμάτων. Για να γίνει αυτό κατανοητό, σε μια επιδημιολογική μελέτη όπου συλλέγονται 1000 δείγματα ούρων γυναικών, μπορεί η ολιστική ανάλυση να αναδείξει μια ομαδοποίηση γυναικών οι οποίες βρίσκονται στην εμμηνόπαυση, ή βρίσκονται σε εγκυμοσύνη, ή ορμονική ή άλλη θεραπεία. Εάν δεν έχει γίνει η σχετική συλλογή των δημογραφικών στοιχείων, έρευνα ετών και πολύ υψηλού κόστους μπορεί να σταματήσει στο σημείο αυτό χωρίς να δώσει ουσιαστικό αποτέλεσμα.

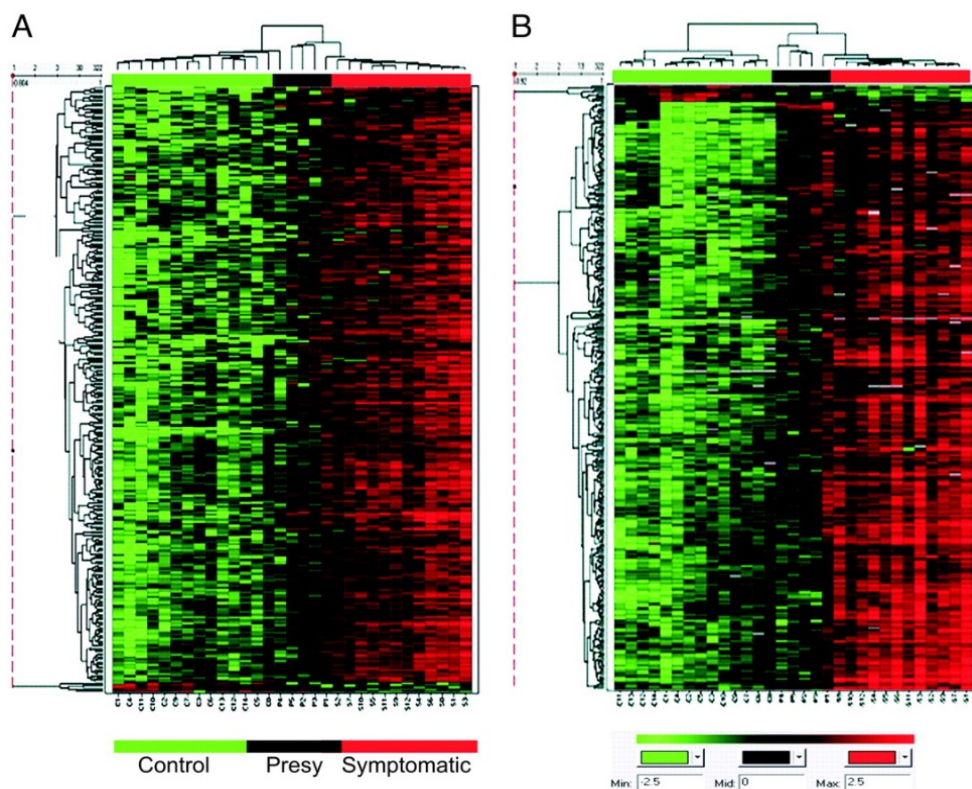
Για να χαρακτηριστεί ένα μόριο ή μια γενετική τροποποίηση ως βιοδείκτης πρέπει να παρουσιάζει χαρακτηριστικές, επαναλήψιμες και αξιόπιστες διαφοροποιήσεις μεταξύ των συγκρινόμενων ομάδων πληθυσμού. Η επικύρωση των διαφορών αυτών πρέπει να είναι ενδεδειγμένη και στατιστικά αποδεκτή, μέσω διαφορετικών μεθόδων στατιστικής ανάλυσης, όπως με Student's *t* test, θηκογράμματα ή πολυπαραμετρική στατιστική ανάλυση.

Στην πράξη όλα τα πεδία των επιστημών της ζωής παρουσιάζουν εντατική δραστηριότητα προς την εύρεση βιοδεικτών. Πολύ μεγάλος αριθμός μελετών πραγματοποιήθηκε ή πραγματοποιείται για την εύρεση βιοδεικτών σε διάφορα δείγματα για ασθένειες όπως καρκίνος, καρδιοαγγειακά νοσήματα, αυτοάνοσα νοσήματα, διαβήτης αλλά και καταστάσεις όπως παχυσαρκία, φυσική άσκηση κ.ά.

Στο **Σχήμα 2.4** δίνονται τα αποτελέσματα της μελέτης γονιδιακής έκφρασης σε δείγματα αίματος ασθενών με νόσο Huntington, με δύο διαφορετικές μικροσυστοιχίες. Τα αποτελέσματα που δίνονται στο σχήμα, προέρχονται από Ανάλυση Ιεραρχικής Ομαδοποίησης (Hierarchical Clustering). Στο **Σχήμα 2.5** δίνεται η πορεία για την εύρεση πρωτεϊνικού καρκινικού βιοδείκτη σε πλάσμα αίματος. Στο **Σχήμα 2.6** δίνεται ένα παράδειγμα από την εφαρμογή μεταβολομικής για την εύρεση βιοδεικτών παχυσαρκίας μέσω της ανάλυσης ούρων από παχύσαρκα και φυσιολογικά πειραματόζωα. Στο σχήμα φαίνεται η διαφοροποίηση των δειγμάτων ούρων με βάση την ανάλυσή τους με LC-MS και πολυπαραμετρική ανάλυση.

2. Εξατομικευμένη ιατρική (Personalized medicine) είναι ένας από τους πιο μεγαλεπίβλους στόχους εφαρμογής της ΒΣ και των τεχνολογιών omics. Αφορά τη συλλογή πληροφοριών και δεδομένων "omics" (genomic, proteomic, metabolomic) για καθένα άνθρωπο καθώς και κλινικών και περιβαλλοντικών δεδομένων, τα οποία στη συνέχεια θα συσχετίζονται για την ιατρική διάγνωση, την απόφαση για ιατρική πράξη, φαρμακευτική αγωγή και δοσολογία. Θεωρείται ότι με τον τρόπο αυτό είναι δυνατή η βελτιστοποίηση της θεραπείας και της απόκρισης του ασθενούς. Το πιο ανεπτυγμένο έως τώρα τμήμα είναι το πεδίο της φαρμακογενομικής, όπου η χαρτογράφηση του γονιδιώματος είναι, απολύτως, εφικτή και δυνατή ακόμη και από διάφορες ιδιωτικές εταιρίες. Επιπλέον, έχουν ήδη προταθεί δεκάδες γονιδιακοί δείκτες για διάφορες

ασθένειες. Η σύγκριση του γενότυπου ενός ασθενούς με γενότυπους αναφοράς μπορεί να αναδείξει διαφοροποιήσεις και να υποδείξει πιθανότητα για εκδήλωση νόσου. Με βάση αυτήν την πληροφορία μπορεί ο ιατρός να επιτύχει ταχύτερη διάγνωση και αποτελεσματικότερη θεραπεία: για παράδειγμα μπορεί ο ιατρός να βοηθηθεί στην επιλογή θεραπείας από γενετική πληροφορία η οποία μπορεί να υποδείξει χαμηλή ανταπόκριση σε κάποιο φάρμακο.

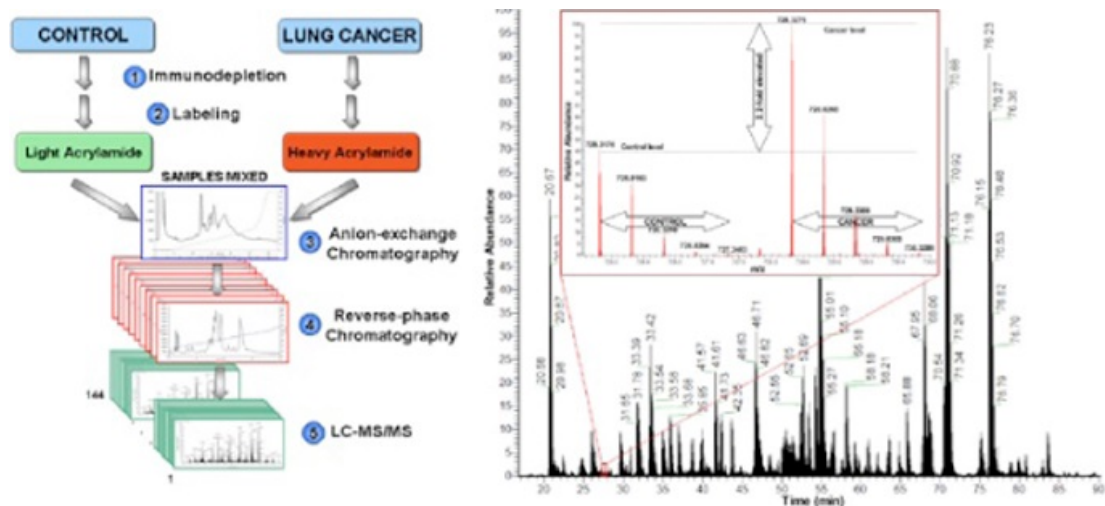


Σχήμα 2.4. Μελέτη γονιδιακής έκφρασης σε ασθενείς με νόσο Huntington. Ανάλυση Ιεραρχικής Ομαδοποίησης (Hierarchical Clustering), σε δείγματα αίματος ασθενών για 322 γονίδια των οποίων η έκφραση διαφοροποιείται σε πλάκα μικροσυστοιχίας Affymetrix (A) και Amersham Biosciences (B). Τα γονίδια αυτά βρέθηκαν να έχουν σημαντική διαφοροποίηση ($P < 0.0005$), λόγος μεταβολής (>1.8 ή <0.6) στον πληθυσμό της μελέτης που νοσούσε. Ο πληθυσμός περιλάμβανε 17 ασθενείς με τη νόσο Huntington και 14 υγιείς μάρτυρες. Κάθε στήλη αντιστοιχεί σε δείγμα και κάθε γραμμή σε ένα γονίδιο. Με κόκκινο φαίνονται τα γονίδια με υψηλά επίπεδα έκφρασης και με πράσινο τα γονίδια με χαμηλά επίπεδα έκφρασης. Παρατηρούνται οι ομαδοποιήσεις των δειγμάτων σε υγιείς (πράσινο χρώμα), προσυμπτωματικούς (μαύρο χρώμα) και συμπτωματικούς ασθενείς (κόκκινο χρώμα, με βάση την έκφραση των εξεταζόμενων γονιδίων [Borovecki et al., 2005]).

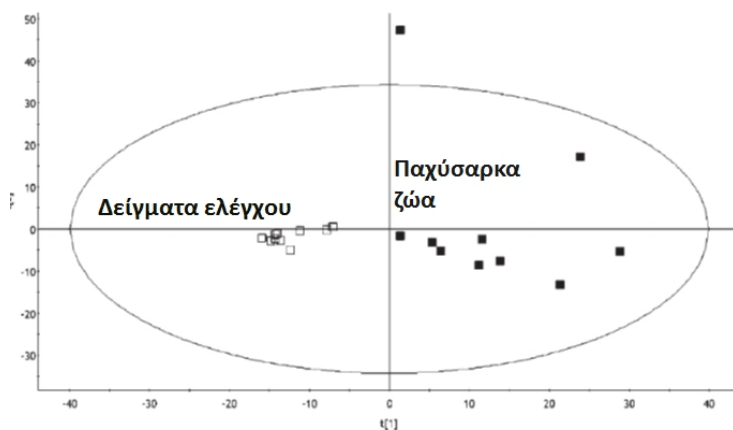
3. Εξατομικευμένη ιατρική (Personalized medicine) είναι ένας από τους πιο μεγαλεπίβουλους στόχους εφαρμογής της ΒΣ και των τεχνολογιών omics. Αφορά τη συλλογή πληροφοριών και δεδομένων "omics" (genomic, proteomic, metabalomic) για καθένα άνθρωπο καθώς και κλινικών και περιβαλλοντικών δεδομένων, τα οποία στην συνέχεια θα συσχετίζονται για την ιατρική διάγνωση, την απόφαση για ιατρική πράξη, φαρμακευτική αγωγή και δοσολογία. Θεωρείται ότι με τον τρόπο αυτό είναι δυνατή η βελτιστοποίηση της θεραπείας και της απόκρισης του ασθενούς. Το πιο ανεπτυγμένο έως τώρα τμήμα είναι το πεδίο της φαρμακογενομικής, όπου η χαρτογράφηση του γονιδιώματος είναι, απολύτως, εφικτή και δυνατή ακόμη και από διάφορες ιδιωτικές εταιρίες. Επιπλέον, έχουν ήδη προταθεί δεκάδες γονιδιακοί δείκτες για διάφορες ασθένειες. Η σύγκριση του γενότυπου ενός ασθενούς με γενότυπους αναφοράς μπορεί να αναδείξει διαφοροποιήσεις και να υποδείξει πιθανότητα για εκδήλωση νόσου. Με βάση αυτήν την πληροφορία μπορεί ο ιατρός να επιτύχει ταχύτερη διάγνωση και αποτελεσματικότερη θεραπεία: για παράδειγμα μπορεί ο ιατρός να βοηθηθεί στην επιλογή θεραπείας από γενετική πληροφορία η οποία μπορεί να υποδείξει χαμηλή ανταπόκριση σε κάποιο φάρμακο.

Ειδικά για τη θεραπεία του καρκίνου, έχει βρεθεί τα τελευταία χρόνια μεγάλη γενετική διαφοροποίηση κάποιων μορφών καρκίνου οι οποίες στη συμβατική ιατρική εμφανίζονται ως ίδιες και

κατηγοροποιούνται μαζί. Γενετικές αναλύσεις υποδεικνύουν μεγάλη ετερογένεια σε όγκους ή ακόμη και διαφοροποιήσεις μέσα στον ίδιο όγκο. Αυτές οι εξελίξεις βοηθούν την ανάπτυξη φαρμάκων τα οποία μπορεί να επιτύχουν πολύ καλά αποτελέσματα σε τέτοιες μορφές καρκίνου, παρόλο που δεν έδωσαν ικανοποιητικά αποτελέσματα σε γενικότερες μελέτες.



Σχήμα 2.5. Πρωτεομική ανάλυση για την ταυτοποίηση καρκινικού βιοδείκτη σε πλάσμα αίματος. Αριστερά: δείγματα πλάσματος από ασθενείς με καρκίνο του πνεύμονα και αντίστοιχα δείγματα από φυσιολογικούς μάρτυρες (controls), στα οποία απομακρύνθηκαν οι πρωτεΐνες υψηλής συγκέντρωσης (βήμα 1), έγινε επισήμανση με ισότοπα ακρυλαμίδιου (ελαφρύ για κανονικό και βαρύ για καρκίνο) (βήμα 2). Έπειτα από ανάμειξη, τα δείγματα κλασματοποιήθηκαν σε επίπεδο πρωτεϊνών σε 144 κλάσματα (βήματα 3 και 4). Καθένα από αυτά τα κλάσματα αναλύθηκε με γρήγη χρωματογραφία-φασματομετρία μάζας σε σειρά (LC-MS / MS) μετά από πέψη της πρωτεΐνης με θρυψίνη (βήμα 5). Δεξιά: Το χρωματογράφημα δείχνει την πολυπλοκότητα των αποτελεσμάτων από αναλύσεις LC -MS/MS. Ένα πεπτίδιο της ινσουλίνης, ταυτοποιήθηκε με έκλουσή του σε περίπου 27,5 λεπτά στο χρωματογράφημα, και η λεπτομερής εικόνα υποδεικνύει ότι το πεπτίδιο που προέρχεται από δείγματα καρκίνου (επισημασμένο βαρύ ισότοπο), όπου έχει περίπου 2 φορές μεγαλύτερης έντασης σήμα σε σχέση με δείγματα από φυσιολογικούς μάρτυρες. [Faca et al., 2010].



Σχήμα 2.6. Αποτελέσματα μεταβολομικής μελέτης σε μοντέλο ζώου με παχυσαρκία. Διαφοροποίηση δειγμάτων ούρων από παχύσαρκα (σκούρα τετράγωνα) και φυσιολογικά πειραματόζωα (ανοικτά τετράγωνα) μετά τη μεταβολομική ανάλυση σε LC-TOF-MS και PCA. Οι δείκτες που βρέθηκαν ως υπεύθυνοι για τη διαφοροποίηση των παχύσαρκων ζώων, περιλάμβαναν λιπίδια όπως: Myristoyl L- α -lysophosphatidylcholine, 1-Linoleoylglycerophosphocholine, 1-Palmitoyllysophosphatidylcholine και άλλες φωσφοχολίνες, που εμφάνισαν υψηλότερες συγκεντρώσεις στα ούρα των παχύσαρκων ζώων [Lofthus et al, 2008].

Έως πρόσφατα η ανάπτυξη των φαρμάκων και η εφαρμογή τους στη θεραπευτική βασιζόταν στην αντίληψη ότι το φάρμακο θα εφαρμοστεί σε μεγάλο τμήμα του πληθυσμού σε μία συγκεκριμένη δόση. Η εξατομικευμένη θεραπεία βαδίζει με τη λογική: το κατάλληλο φάρμακο για τον κατάλληλο ασθενή και στην

κατάλληλη δόση. Η θεραπευτική αγωγή προσαρμόζεται μετά τη λήψη δεδομένων (εργαστηριακών, κλινικών και δεδομένων μελετών ΒΣ) (Σχήμα 2.7). Οι ερευνητές συνεκτιμούν τη γενετική ή πρωτεομική πληροφορία σε συνδυασμό με την αντίδραση του οργανισμού του ασθενούς και αποφασίζουν για την καταλληλότερη φαρμακευτική ένωση: σχήμα, δοσολογία, συχνότητα εφαρμογής. Αυτό μπορεί να βοηθήσει την επίτευξη αποτελεσματικής, μη επεμβατικής, οικονομικής θεραπείας. Για παράδειγμα, το φάρμακο ταμοξιφαίνη (Tamoxifen) χρησιμοποιούνταν συχνά στη θεραπεία καρκίνου με θετικούς υποδοχείς οιστρογόνων για μείωση του κινδύνου εμφάνισης καρκίνου του μαστού. Όμως το 65% των γυναικών ανέπτυξαν ανοχή, μειώνοντας το θεραπευτικό αποτέλεσμα. Μετά από έρευνα βρέθηκε ότι αυτό οφειλόταν στο γεγονός ότι γυναίκες με μετάλλαξη στο γονίδιο CYP2D6, δεν μπορούσαν να μεταβολίσουν το φάρμακο προς τον ενεργό του μεταβολίτη και, επομένως, η θεραπεία αυτή δεν ήταν αποτελεσματική. Έκτοτε γίνεται έλεγχος ως προς τη συγκεκριμένη μετάλλαξη πριν από την εφαρμογή της αγωγής, ώστε οι ασθενείς να έχουν από την αρχή αποτελεσματική θεραπεία. Ασθενείς με τη συγκεκριμένη μετάλλαξη υποβάλλονται σε εναλλακτική αγωγή.

Με την εξέλιξη των υπόλοιπων πεδίων Omics, η εξατομικευμένη θεραπεία έχει μετακινηθεί από την αρχική εστίαση των μοντέλων της στη γονιδιωματική και περιλαμβάνει πλέον και πρωτεϊνικούς, μεταβολικούς ή άλλους δείκτες.

Μελλοντικός στόχος είναι με την εξατομικευμένη ιατρική να είμαστε σε θέση να μπορούμε να προβλέψουμε για κάθε άνθρωπο ποιες νόσους έχει την πιθανότητα να εμφανίσει, την πορεία της νόσου καθώς και την ανταπόκριση στη θεραπεία.

Κάτι τέτοιο, βέβαια, ακόμα δεν είναι απολύτως εφικτό, καθώς η λήψη μεγάλου όγκου πληροφορίας και η διαχείριση όλων αυτών των πληροφοριών είναι τεχνικά δύσκολη και υπόκειται σε περιορισμούς. Προς το παρόν πιο κοντινός στόχος είναι η κατηγοριοποίηση του πληθυσμού με βάση τα 'omic' προφίλ τους (patient stratification) και στη συνέχεια η προσαρμογή της θεραπείας ανάλογα με την κατηγορία του πληθυσμού στην οποία ανήκει, λαμβάνοντας υπόψη και κάποιους άλλους προσωπικούς παράγοντες του ατόμου για το οποίο προορίζεται.



Σχήμα 2.7. Η θεωρία της εξατομικευμένης θεραπείας σε απλοποιημένο σχήμα. Ο πληθυσμός που πάσχει από την ασθένεια είναι ετερογενής και μπορεί να χωριστεί σε 3 υποομάδες των οποίων ο φαινότυπος μπορεί να περιγραφεί ικανοποιητικά από τα ευρήματα μελέτης ΒΣ. Με βάση αυτήν τη γνώση, ο ιατρός μπορεί να επιλέξει τρεις διαφορετικές αγωγές που θα έχουν καλύτερο θεραπευτικό αποτέλεσμα.

4. Έλεγχος Τοξικότητας, για την αναγνώριση και παρακολούθηση τοξικότητας σε όργανα και ιστούς. Οι κλασικές μελέτες φαρμακοκινητικής και μεταβολισμού που μελετούν μόνον τους γνωστούς μεταβολίτες του φαρμάκου και δεν εξετάζουν την υπόλοιπη βιοσύνθεση, θα αγνοήσουν τέτοια φαινόμενα, οπότε υπάρχει ο

κίνδυνος μια ένωση που παρουσιάζει τοξικότητα, να προχωρήσει σε επόμενα στάδια ανάπτυξης και να απορριφθεί μετά από μεγαλύτερο χρόνο και κόστος ανάπτυξης ή ακόμη μετά από την κυκλοφορία της και τις τραγικές επιπτώσεις της στην υγεία των ανθρώπων.

Πρόσφατο παράδειγμα κατά το οποίο τοξικά φάρμακα διέφυγαν του μηχανισμού συμβατικού ελέγχου, είναι η απόσυρση (κατά τα έτη 2004, 2005) δύο φαρμάκων αναστολέων της COX-II, τα οποία χορηγούνταν για θεραπεία αρθρίτιδας. Την εποχή της απόσυρσής τους από την αγορά τα φάρμακα συνταγογραφούνταν σε εκατοντάδες χιλιάδες πασχόντων και είχαν στις ΗΠΑ κύκλο εργασιών άνω των \$ 3 δισεκατομμυρίων. Οι προκλινικοί έλεγχοι απέτυχαν να αναδείξουν την ισχυρή καρδιοτοξικότητα των φαρμάκων οδηγώντας στο θάνατο εκατοντάδων ασθενών.

Μελέτη ΒΣ μπορεί να εφαρμοστεί για παράδειγμα για την προκλινική μελέτη φαρμάκων (πριν από την εμπορική τους κυκλοφορία). Χορηγείται το φάρμακο σε πειραματόζωα (ποντίκια ή/και αρουραίους) και ακολουθεί ολιστική ανάλυση δειγμάτων ούρων και πλάσματος με χρήση NMR, LC-MS ή άλλων τεχνικών. Χαρτογραφείται η σύσταση των βιολογικών υγρών ή των ιστών που έχουν υποστεί βλάβη, προσφέροντας πληροφορία για το μηχανισμό τοξικότητας (βιοσυνθετικών οδών και ροών). Για παράδειγμα, μπορεί να αποκαλυφθεί υψηλή συγκέντρωση πεπτιδίων ή άλλων ενδογενών ενώσεων, οι οποίες λόγω της επίδρασης του φαρμάκου συσσωρεύονται και δρουν τοξικά στον ιστό, χωρίς, όμως, να συνδέονται άμεσα με το μεταβολισμό του φαρμάκου, δηλαδή, δεν είναι μεταβολίτες αυτού. Αυτό μπορεί να συμβεί για πολλούς λόγους π.χ. αδρανοποίηση ενζύμου, μπλοκάρισμα συνενζύμων και άλλοι μηχανισμοί οι οποίοι δεν μπορούν να προβλεφτούν. Το αποτέλεσμα είναι η διαταραχή της βιοσύνθεσης, η συσσώρευση ενώσεων ή η έλλειψη άλλων ενδιάμεσων μεταβολιτών απαραίτητων για τη βιοσύνθεση άλλων ενώσεων.

5. Βιοτεχνολογικές μελέτες σε φυτά, για την ανακάλυψη βιομορίων τα οποία επηρεάζονται από γενετικές τροποποιήσεις, φλεγμονές ή προσβολές από άλλους παράγοντες. Ειδικά σε τρόφιμα και ποτά υψηλής αξίας (π.χ. κρασί, λάδι, μέλι) η χαρτογράφηση του γονιδιώματος ή των μεταβολιτών επιτρέπει την κατηγοριοποίησή τους ως προς τη γεωγραφική προέλευση και τους ελέγχους αυθεντικότητας.

Στα φυτά και τρόφιμα φυτικής προέλευσης μεγάλο ενδιαφέρον παρουσιάζει ο υβριδισμός και η μεταφορά γονιδιακού υλικού μεταξύ φυτών ή καλλιεργειών και στη συνέχεια ο έλεγχος της επιτυχούς παρέμβασης, ο οποίος συχνά περιλαμβάνει τον προσδιορισμό λειτουργικών πρωτεϊνών και εύρεση των τάσεων των μεταβολιτών σε σχέση για παράδειγμα με την ωρίμανση καρπών, την έκθεση στο ηλιακό φως ή άλλους παράγοντες. Οι τεχνικές Omics βοηθούν και επιταχύνουν την ανάπτυξη νέων μεθόδων υβριδισμού για την παραγωγή διαποικιλιακών φυτών με επιθυμητές ιδιότητες ή την ανάπτυξη νέων τεχνικών φυτοπροστασίας. Όπως στην περίπτωση παραγωγής νέων ποικιλιών τομάτας, ο στόχος μπορεί να είναι η αυξημένη συγκέντρωση συστατικών, όπως λυκοπένιο, αργινίνη, κιτρικό οξύ ή άλλα επιθυμητά βιομόρια τα οποία να βελτιώνουν τα ποιοτικά χαρακτηριστικά του καρπού (άρωμα, γεύση, διάρκεια ζωής, χρώμα ή άλλο). Η ΒΣ προσφέρει γνώση για την καλύτερη επιλογή των ποικιλιών προς υβριδισμό (αυτές που περιέχουν ενεργά γονίδια τα οποία εκφράζουν και παράγουν υψηλότερα επίπεδα των επιθυμητών συστατικών). Στη συνέχεια, μετά τη διασταύρωση, επιλέγονται τα υβρίδια που, όντως, ικανοποιούν με τον καλύτερο τρόπο τις αρχικές απαιτήσεις και στόχους. Ο έλεγχος δε γίνεται μόνο στη συγκέντρωση ενός μορίου- στόχου π.χ. αργινίνη, αλλά μπορεί να καλύπτει μεγάλο μέρος ή όλη τη στοχευμένη βιοσυνθετική οδό (arginine pathway).

6. Μελέτες στις επιστήμες διατροφής. Σε συνδυασμό με την άνω ερευνητική δραστηριότητα, τρόφιμα με ορισμένες επιθυμητές ιδιότητες (π.χ. πλούσια σε βιταμίνες ή πολυφαινόλες) χορηγούνται σε πληθυσμούς (ανθρώπους ή πειραματόζωα) και ελέγχεται η διαθεσιμότητα και η κατανομή των διατροφικών συστατικών στον οργανισμό. Μπορεί να πραγματοποιηθεί στοχευμένη ανάλυση για τη μέτρηση των επιπέδων των επιθυμητών συστατικών (όπως αναφέρθηκε παραπάνω) αλλά και η συνολική διαφοροποίηση του προφίλ των βιολογικών δειγμάτων. Για παράδειγμα, η διατροφή με συγκεκριμένο τρόφιμο μπορεί να μειώσει τις συγκεντρώσεις λιπιδίων (ανεπιθύμητων σε σχέση με την υγεία του καταναλωτή), ενώ ταυτόχρονα αυξάνει τις συγκεντρώσεις άλλων επιθυμητών λιπιδίων στο αίμα ή σε άλλο ιστό.

7. Μελέτες της επίδρασης δραστηριοτήτων και συνηθειών του ανθρώπου. Αυτές μπορεί να περιλαμβάνουν το κάπνισμα, τη φυσική άσκηση, την κατανάλωση αλκοόλης και άλλες συνήθειες. Για την περίπτωση της άσκησης έχουν αναπτυχθεί πολλά μοντέλα άσκησης και προπόνησης για διαφορετικά αθλήματα ή ζητούμενα αποτελέσματα. Η παρακολούθηση των αλλαγών σε μοριακό επίπεδο μπορεί να δώσει άμεση και αξιόπιστη πληροφορία για το μηχανισμό ανταπόκρισης του οργανισμού στην κατανάλωση ενέργειας, την εξάντληση των μηχανισμών τροφοδοσίας (ATP, γλυκόλυση), το πιθανό οξειδωτικό στρες.

2.5.1. Μελέτες Βιολογικών Συστημάτων στην πράξη

Για την πραγματοποίηση αναλύσεων στα πεδία Omics απαιτούνται ισχυρές αναλυτικές μέθοδοι, ικανές να παράγουν το πλήρες προφίλ γονιδιώματος, μεταβολιτών, πρωτεϊνών από πολυσύνθετα βιολογικά δείγματα, τα οποία μπορεί να είναι από βακτηριακής μέχρι ανθρώπινης προέλευσης.

Η εισαγωγή διαταραχών και η παρατήρηση των επιδράσεών τους είναι ένας τρόπος για τη διεξοδική μελέτη και την απόκτηση γνώσης για το εξεταζόμενο σύστημα. Αυτή η προσέγγιση απαιτεί τη διεξοδική παρακολούθηση του συστήματος, έτσι ώστε οι επιδράσεις των διαταραχών να μπορούν να αναγνωριστούν και να ερμηνευθούν χωρίς λάθη. Αυτό σημαίνει ότι πρέπει να εντοπιστούν οι διαφορές που οφείλονται μόνο στην επιβαλλόμενη διαταραχή από μεταβολές γενικές οι οποίες μπορεί να οφείλονται στην ανάπτυξη του οργανισμού, τις αλλαγές εποχής κλπ. Αυτή είναι και η βασική απαίτηση για τη βαθύτερη κατανόηση του συστήματος. Σε μελέτες βιολογικών συστημάτων, οι διαταραχές αυτές μπορεί να είναι διαφόρων ειδών, όπως είναι:

1. τοξική προσβολή,
2. περιβαλλοντικές αλλαγές,
3. γενετική τροποποίηση ή
4. συγκεκριμένη ασθένεια.

Τέτοιες παρεμβολές έχουν αντίκτυπο σε μοριακό επίπεδο. Εξαιτίας της πολυπλοκότητας των βιολογικών δειγμάτων και της ευαισθησίας των σημερινών αναλυτικών συστημάτων, τα αποτελέσματα τέτοιου είδους μελετών είναι εξαιρετικά ογκώδη και πλούσια σε πληροφορία. Η πληροφορία αποτελείται από μεγάλο αριθμό συσχετιζόμενων μεταβλητών των οποίων η ερμηνεία είναι εξαιρετικά επίπονη.

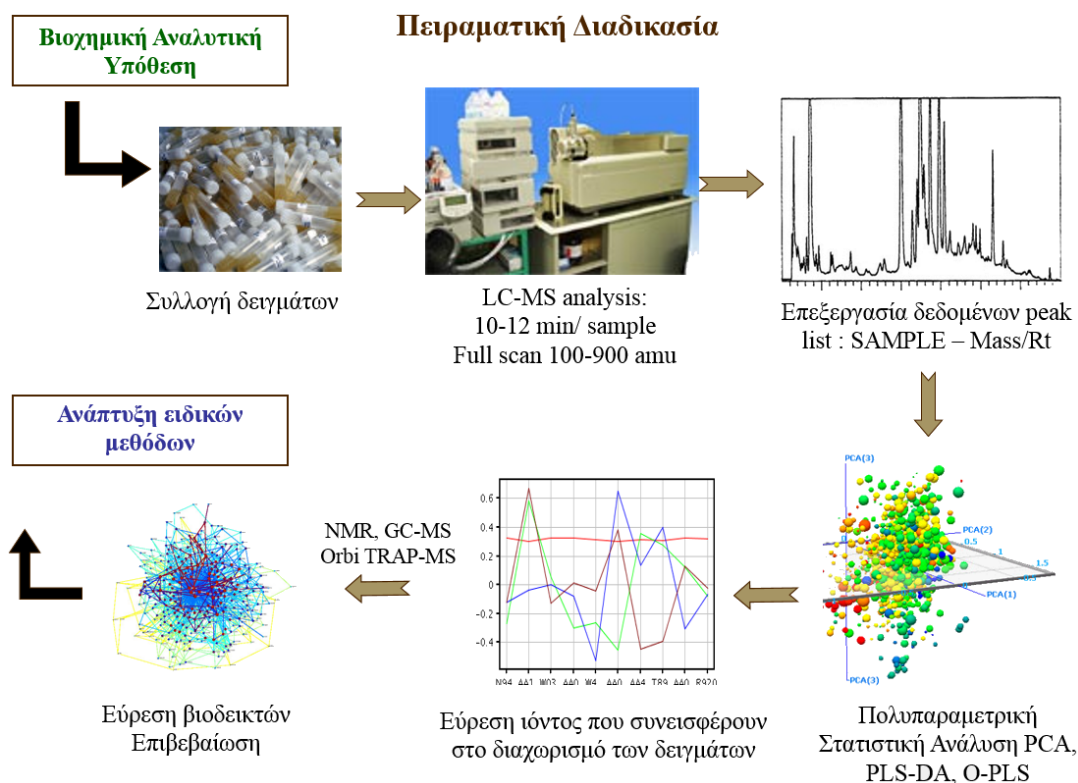
Για τη μελέτη όλων αυτών των μεταβλητών χρησιμοποιούνται, κυρίως, πολυπαραμετρικές μέθοδοι προβολής (παράγραφος 2.7), γιατί οι μέθοδοι αυτές είναι σχεδιασμένες να αντιμετωπίζουν ακριβώς αυτού του είδους τις καταστάσεις και έχει αποδειχθεί ότι είναι κατάλληλες για την επεξεργασία και ανάλυση δεδομένων από πολύπλοκα βιολογικά δείγματα. Για την πλήρη αξιοποίηση αυτών των μεθόδων, συχνά τα δεδομένα πρέπει να υποστούν προκατεργασία, ώστε να είναι κατάλληλα για την περαιτέρω ανάλυση.

Γι' αυτόν το λόγο είναι απαραίτητα σταθερά και ορθολογιστικά πρωτόκολλα επεξεργασίας δεδομένων, ώστε να διατηρούνται όλες οι απαραίτητες πληροφορίες που θα επιτρέψουν την αποτελεσματική πολυπαραμετρική ανάλυση. Με αυτόν τον τρόπο μπορεί να διεξαχθεί αξιόπιστη ανάλυση των δεδομένων, η οποία οδηγεί στη βαθύτερη κατανόηση του συστήματος υπό μελέτη. Η επεξεργασία δεδομένων στη μεταβολομική ανάλυση αναλύεται στο τέλος του κεφαλαίου.

2.5.2. Πορεία της μελέτης

Στο **Σχήμα 2.9** παριστάνεται η ολική πορεία της μεταβολομικής ανάλυσης με χρήση LC-MS. Ξεκινώντας από μια αρχική ιδέα (π.χ. τη μελέτη σταθερότητας δειγμάτων ούρων, τη μελέτη μεθόδων εκχύλισης, την ανακάλυψη βιοδεικτών αρθρίτιδας σε γονιδιακά τροποποιημένα ποντίκια), πρέπει τα δείγματα να συλλεχθούν σύμφωνα με αυστηρά προεπιλεγμένα κριτήρια, ώστε να εξαλειφθεί η πιθανότητα εισαγωγής μη απαραίτητων παραγόντων στατιστικού ή χημικού θορύβου. Για παράδειγμα, δείγματα αίματος τα οποία έχουν συλλεχθεί σε διαφορετικούς σωληνίσκους ή σύριγγες, μπορεί να παρουσιάσουν μεγάλες διαφορές στο προφίλ των μεταβολιτών λόγω διαφορετικών προσθέτων των σωλήνων/συρίγγων (όπως PEG) τα οποία δίνουν πολύ ισχυρό σήμα στο LC-MS.

Αντιστοίχως, δείγματα αίματος τα οποία συλλέγονται σε διαφορετική ώρα ή αφήνονται για παραλαβή ορού για διαφορετικό χρονικό διάστημα (κάποια δείγματα για 20 λεπτά, κάποια άλλα για 40 λεπτά) μπορεί να περιέχουν διαφορετικές συγκεντρώσεις βιομορίων, επειδή οι συγκεντρώσεις αυτών κυμαίνονται στη διάρκεια της ημέρας ή επειδή άλλαξαν στη διάρκεια της διαφορετικής διάρκειας αναμονής στον εργαστηριακό πάγκο. Τέτοια ευρήματα δεν προέρχονται από την πραγματική βιολογική/κλινική αιτία (π.χ. ασθένεια) και μπορεί να ακυρώσουν ακριβές και πολύχρονες μελέτες, λόγω αύξησης της αβεβαιότητας. Σε συμβατική φαρμακοκινητική ανάλυση, η σταθερότητα του φαρμάκου μπορεί να ελεγχθεί με ειδική μελέτη. Αυτό, όμως, δεν είναι δυνατό για το σύνολο των ενδογενών μεταβολιτών ή πρωτεϊνών του δείγματος: γι' αυτόν το λόγο χρειάζεται επιμελής σχεδιασμός και προσοχή κατά την παραλαβή των δειγμάτων μελετών ΒΣ.



Σχήμα 2.9. Πορεία πραγματοποίησης μιας μεταβολομικής μελέτης με χρήση LC-MS

Εφόσον ακολουθηθεί ενδελεχής και αναλυτικός πειραματικός σχεδιασμός, τα δείγματα αναλύονται στο LC-MS με ολική σάρωση στο φασματογράφο μαζών και βαθμωτή έκλουση στο χρωματογράφο, ώστε να συλλεχθεί όσο πιο πλούσιο προφίλ είναι δυνατόν: να εκλουσθούν και να μετρηθούν όσο περισσότερες ενώσεις γίνεται. Το χρωματογράφημα μοιάζει με κύβο που έχει στις τρεις διαστάσεις του: το χρόνο, την κλίμακα m/z και την ένταση σήματος. Τα δεδομένα εξετάζονται ως ιοντικό χρωματογράφημα και μελετάται η σταθερότητα του συστήματος με κριτήρια τη σταθερότητα απόκρισης του ανιχνευτή και τη σταθερότητα των χρόνων συγκράτησης. Μελετώνται δείγματα ελέγχου ποιότητας με ιδιαίτερη προσοχή. Εφόσον οι μετρήσεις είναι ικανοποιητικές, τα δεδομένα αναλύονται με ειδικό λογισμικό το οποίο συρρικνώνει τις δύο διαστάσεις: χρόνο και κλίμακα m/z , σε μία, την ταυτότητα της κορυφής, οπότε παράγεται πίνακας που περιέχει τις εντάσεις σήματος για κάθε κορυφή και δείγμα. Ο πίνακας εισάγεται για στατιστική ανάλυση με στόχο τη μελέτη τάσεων μεταξύ των δειγμάτων. Αν για παράδειγμα στη μελέτη παχυσαρκίας, μια χωριστή ομάδα που αντιπροσωπεύει δείγματα από παχύσαρκους ασθενείς/πειραματόζωα, μελετάται η αξιοπιστία του στατιστικού μοντέλου, η ικανότητα κατηγοριοποίησης και πρόβλεψης των δειγμάτων. Εάν αυτά είναι ικανοποιητικά, μελετώνται οι βιοδείκτες/μεταβλητές που ευθύνονται για την ομαδοποίηση. Ο αναλυτής αναζητεί ποιες κορυφές έχουν διαφορετική συγκέντρωση στις ομάδες και συνεισφέρουν στην αναλυτική τους διαφοροποίηση. Εφόσον οι δείκτες είναι αξιόπιστοι και επαναλήψιμοι, ακολουθεί μελέτη με επιπλέον τεχνικές για την εύρεση της δομής τους. Αυτό είναι συχνά το πιο δύσκολο τμήμα της μελέτης, όπου μπορεί να μην υπάρξει τελικό αποτέλεσμα. Η μελέτη ξεκινά με αναζήτηση σε τοπικές βιβλιοθήκες ή βιβλιοθήκες στο internet (massbank, HMDB, lipidMaps κλπ.) για εύρεση πιθανών μοριακών τύπων και δομών με βάση τη μετρούμενη μοριακή μάζα. Ακολουθούν επιπλέον αναλύσεις MS-MS για λήψη δομικής πληροφορίας (χαρακτηριστικά θραύσματα) ή σε ανεξάρτητες τεχνικές. Για να υπάρξει απόλυτη ταυτοποίηση πρέπει να αναλυθεί το δείγμα επιμολυσμένο με την υποψήφια ένωση και να παρατηρηθεί αύξηση σήματος και ταύτιση φασμάτων. Στο τέλος, οι επιβεβαιωμένοι βιοδείκτες πρέπει να τοποθετηθούν στη βιοσυνθετική οδό και να αποκαλυφθεί ο μηχανισμός αύξησης ή μείωσης της συγκέντρωσής τους στην κάθε περίπτωση.

Οι πορείες ανάλυσης και ανάπτυξης μεθόδου στη πρωτεομική περιέχουν κάποια αντίστοιχα στάδια, αλλά γενικά μπορεί να διαφέρουν πάρα πολύ (κεφάλαιο 12). Σε όλα τα πεδία της ΒΣ, μεγάλο κομμάτι της εργασίας (συνήθως, το περισσότερο χρονοβόρο) αποτελεί η επεξεργασία των δεδομένων για την ανάκτηση της πληροφορίας.

2.6. Επεξεργασία και ανάλυση δεδομένων

Η ανάκτηση της πληροφορίας από δεδομένα ανάλυσης ΒΣ δεν είναι εύκολη. Την τελευταία δεκαετία, η πρόοδος της αναλυτικής τεχνολογίας μάς έχει εφοδιάσει με αξιόπιστα όργανα τα οποία παράγουν τεράστιους όγκους δεδομένων υψηλής αξίας με ταχύτατους ρυθμούς (κάτι το οποίο ήταν και το ζητούμενο των αναλυτικών χημικών). Οι τεράστιοι αυτοί όγκοι δεδομένων και ο χειρισμός τους αποτελούν, πλέον, το μεγαλύτερο πρόβλημα που αντιμετωπίζει σήμερα και θα αντιμετωπίζει στο προσεχές μέλλον η επιστημονική κοινότητα που δραστηριοποιείται στη ΒΣ.

Για να έχει ο αναγνώστης μια εικόνα της διαδικασίας, αναπτύσσεται στη συνέχεια η διαδικασία επεξεργασίας δεδομένων στο πεδίο της μεταβολομικής. Στη μεταβολομική υπάρχει στροφή από τις τεχνικές του NMR προς τεχνικές που βασίζονται στο MS, κυρίως, λόγω του ότι η φασματοσκοπία μαζών είναι πιο ευαίσθητη και προσφέρει επιπλέον διαχωρισμό σε χρωματογραφικές και ηλεκτροφορητικές τεχνικές για την απόκτηση περιεκτικών δεδομένων. Οι τεχνικές MS προσφέρουν υψηλό ρυθμό παραγωγής δεδομένων και πολύ μεγάλο δυναμικό εύρος (μετρούν συγκεντρώσεις από mM έως 10^{-10} M).

Η επεξεργασία των δεδομένων είναι δύσκολη ακόμη και για δεδομένα NMR που είναι πιο απλά διδιάστατα δεδομένα (καταγραφή της έντασης σε σχέση με τη χημική μετατόπιση). Υπάρχει σοβαρή διένεξη σε σχέση με την καλύτερη μέθοδο για την ανάκτηση των δεδομένων (data extraction). Χρησιμοποιούνται τεχνικές χωρισμού του φάσματος σε μικρές περιοχές (binning) τα οποία στη συνέχεια εξετάζονται χωριστά αλλά και τεχνικές αναγνώρισης και ευθυγράμμισης κορυφών σε όλο το φάσμα (peak-picking and alignment). Για την επεξεργασία πολυδιάστατων δεδομένων LC-MS, GC-MS, η κατάσταση είναι, σαφώς, πιο περίπλοκη.

Το βασικό πρόβλημα με τις μεθόδους ανάκτησης δεδομένων είναι η έλλειψη προτυποποίησής τους: όταν γίνεται ανάκτηση από τα ίδια δεδομένα, με n διαφορετικές μεθόδους, μπορεί να ληφθούν n διαφορετικά σετ δεδομένων (datasets) και πιθανότατα n διαφορετικά στατιστικά μοντέλα (εάν ακολουθηθεί ένα και μόνο είδος στατιστικής ανάλυσης). Σε εφαρμογές στις οποίες λαμβάνουν χώρα μεγάλης κλίμακας αλλαγές, όπως για παράδειγμα σε μοντέλα τοξικότητας, οι διαφορές στη μέθοδο ανάκτησης δεδομένων ίσως να μην είναι τόσο επιδραστικές ως προς το τελικό αποτέλεσμα και τα τελικά ευρήματα μπορεί να είναι πρακτικά τα ίδια. Σε άλλες περιπτώσεις, όμως, όπου οι αλλαγές στο σύνολο των μεταβολιτών είναι ελάχιστες (π.χ. επιδημιολογικές μελέτες περιβαλλοντικής έκθεσης ανθρώπων), ακόμη και οι μικρές διαφοροποιήσεις στην επεξεργασία των δεδομένων μπορεί να έχουν δραματικό αντίκτυπο στα αποτελέσματα μιας μελέτης. Γι' αυτόν το λόγο χρειάζεται προσοχή και μεθοδικότητα κατά την επεξεργασία των δεδομένων, η οποία πρέπει να γίνει με τις ίδιες βασικές παραμέτρους για όλα τα δείγματα, εκτός εάν συντρέχει κάποιος σοβαρός και ειδικός λόγος.

Τεχνικές όπως η LC-MS, η GC-MS, και κυρίως η GCxGC-MS, έχουν ως αποτέλεσμα πολύ πιο πολύπλοκα δεδομένα (τριδιάστατα ή ακόμη και τετραδιάστατα) σε σχέση με τα δεδομένα του NMR. Επίσης, οι φασματογράφοι μάζας έχουν ένα πλήθος λειτουργιών και μεθόδων καταγραφής (acquisition modes) καθώς και άλλων παραμέτρων, οι οποίες αυξάνουν την πολυπλοκότητα των δεδομένων. Για παράδειγμα, υπάρχουν οι εξής περιπτώσεις:

1. Ανάκτηση δεδομένων σε centroid ή profile data acquisition mode.
2. Σταθερός χρόνος σάρωσης (όργανα τύπου TOF, Q) ή μεταβλητός (όργανα τύπου IT).
3. Θετική, αρνητική ή εναλλασσόμενη πολικότητα.
4. Σάρωση (full scan MS) ή επιλογή ιόντων (SIM MS).
5. Χαμηλή ή πολύ υψηλή διαχωριστική ικανότητα.
6. Την ικανότητα για συνδυασμό όλων των παραπάνω.
7. Τύποι δεδομένων και μέθοδοι καταγραφής μοναδικοί για το κάθε όργανο.

Τα διάφορα αυτοματοποιημένα εργαλεία εξαγωγής δεδομένων χρησιμοποιούν διαφορετικούς αλγόριθμους. Οι πιο συχνά απαντώμενες προσεγγίσεις περιλαμβάνουν τις εξής διεργασίες:

1. Ανίχνευση κορυφών και ολοκλήρωση, ακολουθούμενη από ευθυγράμμιση (alignment).
2. Στρέβλωση του επιπέδου (warping of the m/z –retention time plane (alignment at raw data level)), αφαίρεση σήματος υποβάθρου, ακολουθούμενη από εντοπισμό κορυφών και ολοκλήρωσή τους.

3. Αφαίρεση των ισοτοπικών κορυφών μετά τα βήματα 1 και 2 (de-isotoping of data).
4. Εξαγωγή μεταβολιτών με τη χρήση φασματοσκοπικών δεδομένων (deconvolution, φασματοσκοπικές βιβλιοθήκες, ακριβής μάζα, κλπ.).

Υπάρχουν περίπου 20 διαφορετικά υπολογιστικά εργαλεία για την αυτόματη εξαγωγή δεδομένων, εμπορικά και δωρεάν διαθέσιμα, τα οποία δίνονται στους **Πίνακες 2.1** και **2.2** αντίστοιχα.

Το τελικό αποτέλεσμα της εφαρμογής αυτών των λογισμικών είναι ο πίνακας δεδομένων που εξάγεται με μορφή xls, csv, txt, και εισάγεται για στατιστική ανάλυση.

ΟΝΟΜΑ	ΚΑΤΑΣΚΕΥΑΣΤΗΣ	ΕΦΑΡΜΟΓΕΣ
BlueFuse	BlueGnome, Cambridge, UK	Metabolomics με δεδομένα MS και NMR
Chenomx NMR Suite	Chenomx, Edmonton, Canada	Metabolomics με δεδομένα NMR
Genedata Expressionist	Genedata, Basel, Switzerland	Πλατφόρμα για γενική χρήση σε τεχνολογίες –omics. Το τμήμα των metabolomics δέχεται δεδομένα MS
LineUp	Infometrix	Χρωματογραφική ευθυγράμμιση
MarkerLynx	Waters, Milford, MA, USA	Metabolomics με δεδομένα LC-MS
MarkerView	Applied Biosystems, Foster City, CA	Metabolomics με δεδομένα LC-MS
MassHunter Profiling Software	Agilent Technologies, Santa Clara, CA	Proteomics με δεδομένα LC-MS
Metabolic Profiler	Bruker Daltonic & Bruker Biospin, Billerica, MA, USA	Metabolomics με δεδομένα MS και NMR
MS Resolver	Pattern Recognition Systems, Bergen, Norway	Δεδομένα GC-MS και LC-MS
Profile	Phenomenone Discoveries, Saskatoon, Canada	Metabolomics με δεδομένα MS και NMR
Rosetta Elucidation	Rosetta Biosoftware, Seattle, WA, USA	Proteomics με δεδομένα LC-MS
Sieve	Thermo Fisher Scientific, Waltham, MA, USA	Proteomics με δεδομένα LC-MS

Πίνακας 2.1. Εμπορικά λογισμικά για την ανάλυση δεδομένων μεταβολομικής

2.7. Πολυπαραμετρική Στατιστική Ανάλυση

Μετά την αρχική επεξεργασία δεδομένων, ακολουθεί στατιστική ανάλυση με διάφορες πολυπαραμετρικές και μονοπαραμετρικές τεχνικές στατιστικής ανάλυσης. Για την ανάλυση δεδομένων μικροσυστοιχιών μπορεί να χρησιμοποιηθεί μια πληθώρα στατιστικών μεθόδων, αλγορίθμων ταξινόμησης και διάκρισης των στοιχείων, τα οποία εντάσσονται στα πλαίσια της πολυπαραμετρικής ανάλυσης. Η πολυπαραμετρική στατιστική ανάλυση αφορά αλγορίθμους ανάλυσης δεδομένων που χρησιμοποιούν περισσότερες από μία μεταβλητές σε κάθε βήμα. Υπάρχουν δύο ευρείες κατηγορίες στατιστικής ανάλυσης που αναφέρονται ως επιβλεπόμενες και μη επιβλεπόμενες τεχνικές. Στις επιβλεπόμενες, η ανάλυση των αποτελεσμάτων προϋποθέτει τη γνώση των κατηγοριών στις οποίες ανήκουν τα δείγματα. Μ' αυτές τις μεθόδους γίνεται δυνατή η διάκριση γονιδίων, μεταβολιτών κ.ά. που διαφέρουν σε συγκεκριμένες καταστάσεις (παράδειγμα τέτοιας τεχνική είναι η SAM). Στις μη επιβλεπόμενες, τα δεδομένα αναλύονται ανεξαρτήτως διαθέσιμης πληροφορίας. Ουσιαστικά είναι μια «τυφλή» ανάλυση, που αποσκοπεί στην κατηγοριοποίηση των δεδομένων των πειραματικών μετρήσεων (παράδειγμα ή PCA και η HCL).

Στην πολυπαραμετρική ανάλυση, αλγόριθμοι ομαδοποίησης ταξινομούν τα δεδομένα και τις ομάδες, με βάση το διαχωρισμό τους σύμφωνα με τα αποτελέσματα της μελέτης. Εφαρμόζονται διάφορες τεχνικές ομαδοποίησης: ιεραρχικές (hierarchical), όπου η ταξινόμηση που προκύπτει έχει έναν αυξανόμενο αριθμό εμφωλευμένων κλάσεων και το αποτέλεσμα μοιάζει με φυλογενετική ταξινόμηση, και μη ιεραρχικοί αλγόριθμοι (non – hierarchical), όπως ο k – means, οι οποίοι, απλώς, διαχωρίζουν τα αντικείμενα σε διαφορετικές ομάδες, χωρίς να προσπαθούν να βρουν τη σχέση μεταξύ των ξεχωριστών αντικειμένων, που αποτελεί και το ζητούμενο στις ιεραρχικές ομαδοποιήσεις.

ΟΝΟΜΑ	ΕΦΑΡΜΟΓΕΣ	ΠΛΑΤΦΟΡΜΑ	ΤΥΠΟΣ ΑΔΕΙΑΣ
Chrompare	Metabolomics με δεδομένα GC-FID	Microsoft Excel Visual Basic	Διαθέσιμο για download
COMSPARI	Metabolomics με δεδομένα LC-MS και GC-MS	Γραμμένο σε γλώσσα C, για κάθε πρόσφατη πλατφόρμα, όπως Linux και Windows	GNU General Public Licence
Continuous profile models	Proteomics με δεδομένα LC-MS	Toolbox για το Matlab	Δωρεάν για εκπαιδευτική και ερευνητική χρήση, διαθέσιμος πηγαίος κώδικας
HiRes	Metabolomics με δεδομένα NMR	Γραμμένο σε γλώσσα C++, για Windows	Δωρεάν για εκπαιδευτικούς και κλινικούς σκοπούς
LCMSWARP	Proteomics με δεδομένα LC-MS και LC-MS/MS	Γραμμένο σε γλώσσα C++	Άγνωστο
MapQuant	Proteomics με δεδομένα LC-MS	Γραμμένο σε γλώσσα C, για Windows και Linux	Harvard University open-source compatible licence
MathDAMP	Metabolomics με δεδομένα LC-MS, GC-MS και CE-MS	Πακέτο για το Mathematica	Δωρεάν
metAlign	Για δεδομένα GC-MS και LC-MS	Windows	Δωρεάν
MET-IDEA	Metabolomics με δεδομένα LC-MS, GC-MS και CE-MS	Windows, πλατφόρμα .NET	Δωρεάν διαθέσιμο για ακαδημαϊκούς σκοπούς κατόπιν αίτησης
MSFACTs	Metabolomics με δεδομένα LC-MS και GC-MS	Γραμμένο σε Java	Δωρεάν διαθέσιμο για ακαδημαϊκούς σκοπούς κατόπιν αίτησης
MSight	Proteomics με δεδομένα LC-MS	Windows	Δωρεάν
msInspect	Proteomics με δεδομένα LC-MS	Γραμμένο σε Java, απαιτεί τη στατιστική γλώσσα R	Δωρεάν λογισμικό διαθέσιμο με άδεια Apache 2.0
MZmine	Metabolomics με δεδομένα LC-MS και GC-MS	Γραμμένο σε Java	GNU General Public Licence
SpecArray	Proteomics με δεδομένα LC-MS	Γραμμένο σε γλώσσα C, για Linux	GNU General Public Licence
Xalign	Proteomics με δεδομένα LC-MS	Γραμμένο σε γλώσσα C++	Διαθέσιμο μετά από αίτηση στον προγραμματιστή
XCMS	Metabolomics με δεδομένα LC-MS και GC-MS	Γραμμένο στη στατιστική γλώσσα R	GNU General Public Licence

Πίνακας 2.2. Δωρεάν διαθέσιμα λογισμικά για την ανάλυση δεδομένων metabolomics και proteomics

Η ιεραρχική ομαδοποίηση έχει το πλεονέκτημα ότι είναι σχετικά απλή και μπορεί να γίνει εύκολα αντιληπτή μέσω γραφημάτων. Είναι η πιο διαδεδομένη μέθοδος ανάλυσης δεδομένων γονιδιακής έκφρασης. Αποτελεί μια συνενωτική μέθοδο κατά την οποία το κάθε δεδομένο εντάσσεται σε ομάδες, οι οποίες ομαδοποιούνται και μεταξύ τους μέχρι να ολοκληρωθεί η διαδικασία η οποία θα δώσει ένα ιεραρχικό δένδρο. Η διαδικασία προχωρά με απλό τρόπο. Αρχικά, υπολογίζεται ένας πίνακας αποστάσεων για όλα τα γονίδια που πρέπει να ομαδοποιηθούν. Στη συνέχεια, ο πίνακας αποστάσεων εξερευνάται για την εύρεση των δύο πιο όμοιων γονιδίων ή ομάδων. Κάθε ομάδα περιέχει αρχικά ένα μόνο γονίδιο, δηλαδή, έχουμε τόσες ομάδες όσα και τα γονίδια. Αν υπάρχουν περισσότερα του ενός ζευγάρια με ίσες αποστάσεις, καλό θα ήταν να έχει καθοριστεί εκ των προτέρων ένας κανόνας απόφασης για τέτοιες περιπτώσεις. Τώρα, τα δύο γονίδια – ομάδες ενώνονται και παράγουν μια ομάδα με δύο μέλη το ελάχιστο. Συνεχίζοντας, επανυπολογίζονται οι αποστάσεις μεταξύ της νέας ομάδας και των υπόλοιπων ομάδων. Δεν είναι απαραίτητο να υπολογιστούν εκ νέου όλες οι αποστάσεις, παρά μόνον αυτές που συμπεριλαμβάνουν την ομάδα που σχηματίστηκε. Τέλος, επαναλαμβάνονται οι διαδικασίες σχηματισμού ομάδων με βάση τις αποστάσεις, μέχρι να ομαδοποιηθούν όλα τα δεδομένα σε μία και μόνο ομάδα.

Αν διαθέτουμε προχωρημένη γνώση για τον αριθμό των ομάδων στις οποίες πρέπει να διαχωριστούν τα δεδομένα μας, ο αλγόριθμος *k – means* είναι μια καλή εναλλακτική πρόταση στις ιεραρχικές μεθόδους. Εδώ τα δεδομένα χωρίζονται σε προκαθορισμένο αριθμό (*k*) ομάδων, έτσι ώστε τα δεδομένα να είναι παρόμοια εντός μιας ομάδας και ανόμοια μεταξύ διαφορετικών ομάδων. Ο αλγόριθμος δεν παράγει δένδρογράμματα.

Οι *SOM (self organizing maps)* είναι αλγόριθμοι οι οποίοι βασίζονται σε νευρωνικά δίκτυα και ανήκουν στις διαχωριστικές προσεγγίσεις. Ένας SOM διακρίνει τα γονίδια σε σειρά διαχωρισμών με βάση την ομοιότητα των διανυσμάτων έκφρασης σε σύγκριση με διανύσματα αναφοράς τα οποία καθορίζονται για κάθε διαχωρισμό.

Οπτικοποίηση είναι η τεχνική κατά την οποία δημιουργούνται γραφήματα τα οποία μας επιτρέπουν να εξάγουμε συμπεράσματα για τα δεδομένα μας. Ο ανθρώπινος νους μπορεί να αντιλαμβάνεται εικόνες σε τρισδιάστατο χώρο μόνο. Όταν τα δεδομένα ξεπερνούν τις τρεις διαστάσεις, δεν μπορούμε να κάνουμε αποτελεσματική διάκριση χαρακτηριστικών και προτύπων και να αντιληφθούμε τις τάσεις διαφοροποίησης μέσα στα δεδομένα. Για το λόγο αυτό έχουν εξελιχθεί αλγόριθμοι που μας επιτρέπουν να αντιληφθούμε πιο εύκολα τα δεδομένα. Γίνεται ειδική μορφοποίηση και τελική παρουσίαση των δεδομένων και, πλέον, γινόμαστε ικανοί να διακρίνουμε τα στοιχεία που μας ενδιαφέρουν. Στη συνέχεια της παραγράφου περιγράφονται εν συντομία μέθοδοι οπτικοποίησης οι οποίες χρησιμοποιούνται, συχνά, κατά την ανάλυση δεδομένων βιολογίας συστημάτων.

PCA (principal component analysis – ανάλυση κυρίων συνιστωσών). Πολλές από τις τεχνικές μέτρησης που χρησιμοποιούνται στις επιστήμες ζωής, συγκεντρώνουν δεδομένα για πολύ περισσότερες μεταβλητές ανά δείγμα, από τον αριθμό των δειγμάτων. Για παράδειγμα, οι μικροσυστοιχίες DNA μετρούν τα επίπεδα χιλιάδων mRNA από εκατοντάδες δείγματα. Αυτή η πολυδιάστατη συλλογή δεδομένων κάνει δύσκολη την οπτικοποίηση των δειγμάτων και περιορίζει την εξερεύνηση των δεδομένων. Η ανάλυση κύριων ή πρωταρχικών συνιστωσών είναι ένας μαθηματικός αλγόριθμος που μειώνει τις διαστάσεις των δεδομένων, ενώ ταυτόχρονα διατηρεί το μεγαλύτερο μέρος της πληροφορίας: επομένως, διατηρεί τη δυνατότητα απόκτησης γνώσης για τις αποκλίσεις μεταξύ δεδομένων άρα και δειγμάτων. Αυτό γίνεται δυνατό με το να αναγνωρίζει κατευθύνσεις οι οποίες ονομάζονται πρωταρχικές συνιστώσες, κατά μήκος της μέγιστης απόκλισης. Με βάση λίγες συνιστώσες, κάθε δείγμα αναπαριστάται από σχετικά λίγα νούμερα σε σχέση με τις τιμές των χιλιάδων μεταβλητών που έχουμε στην πραγματικότητα. Τα δείγματα τώρα, μπορούν να αναπαρασταθούν σε διάγραμμα και μπορούμε να δούμε ομοιότητες και διαφορές μεταξύ τους, αλλά και να τα ομαδοποιήσουμε.

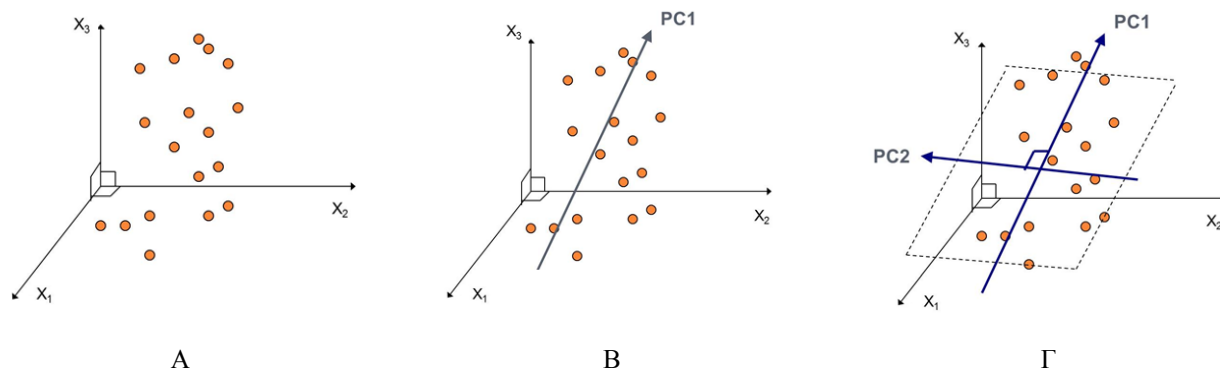
Η PCA είναι, δηλαδή, μια μέθοδος οπτικοποίησης των δεδομένων από τον πολυδιάστατο στο γεωμετρικό τρισδιάστατο (συνήθως) χώρο, έτσι ώστε να γίνονται πιο κατανοητά τα αποτελέσματα και ευρήματα. Η PCA μετατρέπει τον πειραματικό χώρο σε λιγότερες διαστάσεις βάσει του πίνακα συσχέτισης (correlation matrix) των δεδομένων, στηριζόμενη σε γραμμικό μετασχηματισμό. Στόχος είναι να μειωθούν οι διαστάσεις των δεδομένων, ενώ ταυτόχρονα διατηρείται το μεγαλύτερο μέρος της διακύμανσής τους, για να εντοπιστούν οι ομοιότητες και οι διαφορές μεταξύ των στοιχείων, όπως και οι ομαδοποιήσεις που προκύπτουν από το μεγάλο πλήθος των δεδομένων. Γίνεται προσπάθεια μελέτης όσο το δυνατόν λιγότερων μεταβλητών και συγκεκριμένα των πιο αντιπροσωπευτικών μεταβλητών του πειράματος, χωρίς να χαθεί σημαντική πληροφορία για το βαθμό συσχέτισης των αποτελεσμάτων. Αυτό είναι περισσότερο εφικτό όταν τα πειραματικά δεδομένα σχετίζονται πολύ μεταξύ τους και μπορούν να παραλειφθούν κάποια ή να υπολογιστεί ένας μέσος όρος. Στην περίπτωση που η συσχέτιση των αρχικών μεταβλητών δεν είναι μεγάλη, τότε η ανάλυση με PCA καθίσταται ακατάλληλη μέθοδος.

Η PCA ξεκινάει με τη δημιουργία του πίνακα των δεδομένων όπου σειρές του πίνακα είναι τα δείγματα που εξετάστηκαν και οι στήλες του είναι οι εξεταζόμενες μεταβλητές (αναλυτικό σήμα, αναλυόμενες ενώσεις). Ο αλγόριθμος της PCA ανακατασκευάζει τον αρχικό πίνακα μεταβλητών σε έναν πίνακα με μικρότερο αριθμό στηλών οι οποίες ονομάζονται κύριες συνιστώσες και ανταποκρίνονται στο γραμμικό συνδυασμό των αρχικών μεταβλητών. Οι νέοι άξονες είναι κάθετοι μεταξύ τους έτσι ώστε οι συντεταγμένες των παρατηρήσεων να είναι οι τιμές των κύριων συνιστωσών. Οι νέες μεταβλητές είναι γραμμικός συνδυασμός των αρχικών μεταβλητών. Η πρώτη συνιστώσα αντιστοιχεί στο άνυσμα που περιγράφει τη μέγιστη διαφοροποίηση μεταξύ των δειγμάτων. Η δεύτερη συνιστώσα είναι άνυσμα σε γωνία 90 μοιρών με την πρώτη συνιστώσα και εξηγεί τη δεύτερη μεγαλύτερη διακύμανση, δηλαδή τη μέγιστη διαφοροποίηση της διακύμανσης που δεν μπορεί να εξηγήσει η πρώτη συνιστώσα, και ούτω καθεξής με τις υπόλοιπες συνιστώσες.

Για να γίνει πιο αντιληπτός ο τρόπος λειτουργίας της PCA, περιγράφεται συνοπτικά η μελέτη συγκεκριμένης εργασίας. Η ανάλυση είχε ως αποτέλεσμα 704 κορυφές, σε 18 δείγματα. Ο αλγόριθμος της PCA θα δημιουργήσει ένα σύννεφο (Σχήμα 2.10Α), και θα τραβήξει το άνυσμα κατά μήκος του μεγαλύτερου ποσοστού διακύμανσης των δεδομένων. Αυτό το άνυσμα είναι η κύρια συνιστώσα 1 (Σχήμα 2.10Β). Έπειτα θα βρεθεί η αμέσως επόμενη κατεύθυνση με το μεγαλύτερο ποσοστό δεδομένων, αυτή, δηλαδή, που περιγράφει το 2ο μεγαλύτερο ποσοστό διαφοροποίησης. Αυτή θα είναι η δεύτερη συνιστώσα που θα είναι κάθετη στην πρώτη και τα δεδομένα θα μπορούν, πλέον, να προβληθούν σε διδιάστατο χώρο (Σχήμα 2.10Γ) σε μορφή επιπέδου, όπως δίνεται στα Σχήματα 2.6 και 2.8. Όταν προστεθεί η 3η κύρια συνιστώσα λαμβάνουμε τρισδιάστατα γραφήματα, όπως αυτό του Σχήματος 2.9. Από τις κύριες συνιστώσες, συνήθως, εξετάζονται ανά δύο ή ανά τρεις οι 4-5 πρώτες συνιστώσες οι οποίες εκφράζουν το μεγαλύτερο ποσοστό της διακύμανσης στα δεδομένα.

SAM (Significance Analysis of Microarrays). Η SAM είναι μια στατιστική μέθοδος η οποία χρησιμοποιείται, κυρίως, στη γονιδιωμική. Αναγνωρίζει γονίδια με στατιστικά σημαντικές αλλαγές στην έκφραση συνταυριάζοντας ένα σύνολο t tests τα οποία αφορούν συγκεκριμένα γονίδια. Κάθε γονίδιο παίρνει μια συγκεκριμένη τιμή που βασίζεται στην αλλαγή της έκφρασής του σχετικά με την τυπική απόκλιση επαναλαμβανόμενων μετρήσεων για το γονίδιο. Τα γονίδια με τιμές μεγαλύτερες από ένα δεδομένο όριο καθορίζονται ως, δυναμικώς, σημαντικά. Το ποσοστό αυτών των γονιδίων που θα προκύψουν τυχαία, αποτελεί το ρυθμό λανθασμένης ανακάλυψης (FDR), ο οποίος προσδιορίζεται από ανάλυση πολλαπλών συνδυασμών των μετρήσεων.

Support Vector Machine (SVM). Είναι αλγόριθμοι οι οποίοι μπορούν να μαθαίνουν από παραδείγματα και εκπαιδευτικά υποσύνολα δεδομένων, ώστε να αναγνωρίζουν και να κατηγοριοποιούν αντικείμενα και δείγματα. Βασική τους εφαρμογή στη βιοιατρική είναι στην κατηγοριοποίηση δεδομένων γονιδιακής έκφρασης από πειράματα μικροσυστοιχιών. Αποτελούν επιβλεπόμενους αλγόριθμους οι οποίοι χρησιμοποιούν έναν αρχικό αριθμό δεδομένων για εκπαίδευση της συνάρτησής τους και, στη συνέχεια, είναι σε θέση να αναγνωρίσουν και να ξεχωρίσουν τα υπόλοιπα δεδομένα με βάση την έκφρασή τους. Με λίγα λόγια, ένα βασικό τους πλεονέκτημα είναι ότι χρησιμοποιούν βιολογικές πληροφορίες για να καθορίσουν τα στοιχεία έκφρασης τα οποία χρειάζονται για να χαρακτηρίσουν τις ομάδες και, μετά, είναι ικανοί να κατηγοριοποιήσουν οποιοδήποτε άλλο δείγμα/αντικείμενο.



Σχήμα 2.10. Α: Προβολή στο χώρο των δειγμάτων σε διάγραμμα κύριων συνιστωσών. Κάθε κουκκίδα αντιστοιχεί σε ένα δείγμα. Β: Κύρια συνιστώσα 1, περιγράφει ποσοστό διαφοροποίησης των δειγμάτων μεγαλύτερο από οποιαδήποτε άλλη συνιστώσα. Γ: Κύρια συνιστώσα 2, περιγράφει το 2^ο μεγαλύτερο ποσοστό διαφοροποίησης.

Αναφορές

Borovecki F. et al. (2005). *Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease*. PNAS;102, 11023-11028.

Faca V., Krasnoselsky A., Hanash S. (2007). *Innovative proteomic approaches for cancer biomarker discovery*, Fred Hutchinson Cancer Research Center, Seattle, WA, USA BioTechniques, 43, 279–283

- Loftus N., Miseki K., Iida J., Gika H.G., Theodoridis G., Wilson I.D. (2008). *Profiling and Biomarker Identification in plasma from different Zucker rat strains via high mass accuracy multistage mass spectrometric analysis using liquid chromatography/mass spectrometry with a quadrupole ion trap-time of flight mass spectrometer*, Rapid Communications in Mass Spectrometry 22, 2547-2554.
- Pechlivanis A., Kostidis S., Saraslanidis P., Petridou A., Tsalis G., Mougios V., Gika H.G., Mikros E., Theodoridis G.A., (2010). *H-NMR-based metabolomic investigation of the effect of two different exercise sessions on the metabolic fingerprint of human urine*, J. Proteome Res., 9 6405-16.

Βιβλιογραφία

- Bolouri H. (2009). *Personal Genomics and Personalized Medicine*, Imperial College Press
- Brown T.A. (2002). *Genomes, 2nd edition*, Oxford: Wiley-Liss.
- Chadwick B.P. (2015). *Epigenetics: Current Research and Emerging Trends*, Caister Academic Press.
- Fuentes M., LaBaer J., (2014). *Proteomics: Targeted Technology, Innovations and Applications*, Caister Academic Press.
- Ginsburg G.S., Willard H.F. (2012). *Genomic and Personalized Medicine*, Second Edition: V1-2, Academic Press.
- Issaq H.J., Veenstra T.D. (2013). *Proteomic and Metabolomic Approaches to Biomarker Discovery*, Academic Press.
- Lindon J.C., Nicholson J.K. (2007). *The Handbook of Metabonomics and Metabolomics*, Elsevier.
- Liu Y. (2014). *Omics in Clinical Practice: Genomics, Pharmacogenomics, Proteomics, and Transcriptomics in Clinical Research*, Yu Liu, 2014, Apple Academic Press.
- Mishra N. (2010). *Introduction to Proteomics: Principles and Applications*, Willey.
- Twyman R., (2004). *Principles of Proteomics*, Taylor & Francis.

Δικτυογραφία

National Institutes of Health (NIH) Human Microbiome Project (HMP)	http://hmpdacc.org/
The Human Metabolome Database (HMDB)	www.hmdb.ca/
The Human Proteome Project	www.thehpp.org/
KEGG (Kyoto Encyclopedia of Genes and Genomes)	www.kegg.jp
LIPID MAPS Lipidomics Gateway	www.lipidmaps.org/
Proteomics Database	https://www.proteomicsdb.org/
ExpASY: SIB Bioinformatics Resource Portal	www.expasy.org
Gene Database	www.ncbi.nlm.nih.gov/gene