

11. Βιοπληροφορική και Βιολογία Συστημάτων

Η ραγδαία ανάπτυξη των βιολογικών επιστημών τις τελευταίες δεκαετίες οδήγησε στη δημιουργία τεράστιου όγκου δεδομένων, τα οποία δεν μπορούν πλέον να αναλυθούν με τις παραδοσιακές μεθόδους. Η αλληλούχηση ολόκληρων γονιδιωμάτων, η μαζική παραγωγή δεδομένων RNA, οι μελέτες του συνόλου των πρωτεϊνών και των μεταβολιτών, αλλά και η χρήση εικόνων υψηλής ανάλυσης έχουν μετατρέψει τη βιοϊατρική σε μια επιστήμη που εξαρτάται άμεσα από την πληροφορική. Η **βιοπληροφορική** (bioinformatics) γεννήθηκε ως απάντηση σε αυτή την ανάγκη: είναι ο κλάδος που συνδυάζει τη βιολογία και την επιστήμη των υπολογιστών για τη διαχείριση, ανάλυση και ερμηνεία βιολογικών δεδομένων. Ο όρος «βιοπληροφορική» πρωτοχρησιμοποιήθηκε τη δεκαετία του 1970, αλλά έγινε γνωστός με την πρόοδο της γονιδιωματικής τη δεκαετία του 1990, ιδίως με το Πρόγραμμα Ανθρώπινου Γονιδιώματος (Human Genome Project). Το 2001 δημοσιεύτηκε η πρώτη προσχέδια αλληλουχία του ανθρώπινου γονιδιώματος, ένα ορόσημο που κατέστησε σαφές ότι η κατανόηση της ανθρώπινης βιολογίας και παθολογίας απαιτεί προηγμένα υπολογιστικά εργαλεία. Σήμερα, η βιοπληροφορική δεν αποτελεί απλώς ένα βοηθητικό εργαλείο, αλλά έναν κεντρικό πυλώνα της βιοϊατρικής έρευνας και της κλινικής πράξης. Από τη μελέτη των μοριακών μηχανισμών ασθενειών μέχρι την ανάπτυξη νέων φαρμάκων και την **εξατομικευμένη ιατρική**¹ και θεραπεία, η εφαρμογή της βιοπληροφορικής αξιοποιεί τα δεδομένα που παράγονται καθημερινά στην έρευνα και στη διάγνωση.

11.1 Βιολογικά δεδομένα

Το αντικείμενό της βιοπληροφορικής είναι εξαιρετικά ευρύ και αφορά όλους τους τομείς της βιολογίας πέρα από την Ιατρική Βιολογία. Βασική προϋπόθεση για την εφαρμογή της βιοπληροφορικής είναι η ύπαρξη **μεγάλων σειρών δεδομένων** (big data). Όπως ήταν εμφανές σε αρκετά από τα κεφάλαια του βιβλίου η ζωή χρησιμοποιεί κώδικες με τους οποίους οι μηχανισμοί λειτουργούν συντηρητικά και συντονισμένα. Από την κωδικοποίηση των ζευγών βάσεων στο DNA και την κωδικοποίηση του γενετικού κώδικα, μέχρι των δομικό κώδικα που κρύβεται πίσω από τις φυσικοχημικές ιδιότητες των μονομερών των βιομορίων, η ζωή βασίζεται πάνω στην πληροφορία για την διατήρηση αλλά και την εξέλιξη της. Η βιοπληροφορική καλείται να διαχειριστεί νουκλεοτιδικές και αμινοξικές αλληλουχίες αλλά και πληροφορίες δομών των μακρομορίων ή πληροφορίες για την παρουσία μεταβολιτών και άλλων βιομορίων. Η συσσώρευση αυτών των δεδομένων έκανε επιτακτική τη δημιουργία και διαχείριση **βάσεων βιολογικών δεδομένων (Πίνακας 11.1)**. Τα δεδομένα από τις βάσεις δεδομένων σε συνδυασμό με τα πειραματικά ή κλινικά δεδομένα με τη βοήθεια αλγορίθμων, δηλαδή υπολογιστικών προγραμμάτων, οδηγούν σε ολοκληρωμένες αναλύσεις που επιτρέπουν την ερμηνεία της βιολογικής πληροφορίας και τη μετάφρασή της σε κλινικά χρήσιμα συμπεράσματα. Μέσα από αυτή τη διαδικασία αναδεικνύονται πρότυπα γονιδιακής έκφρασης, ταυτοποιούνται παθογόνες μεταλλάξεις, εντοπίζονται νέοι **βιοδείκτες**² και δημιουργούνται μοντέλα που προβλέπουν την πορεία μιας νόσου ή την ανταπόκριση ενός ασθενούς σε μια θεραπεία. Η αλληλεπίδραση ανάμεσα στα δεδομένα και τους αλγορίθμους

¹ Η ιατρική που αξιοποιεί τα προσωπικά βιολογικά δεδομένα του κάθε ασθενούς και προσαρμόζει την αντιμετώπιση και τη θεραπεία εξατομικευμένα. Ονομάζεται και ιατρική ακριβείας.

² Βιολογικά μόρια που έχουν συσχετισθεί με την ανάπτυξη, την εξέλιξη και θεραπεία μιας ασθένειας.

μετατρέπει τη βιοπληροφορική από μια απλή αποθήκη πληροφοριών σε ένα δυναμικό εργαλείο ιατρικής ακριβείας.

Πίνακας 11.1 Τύποι Βιολογικών Δεδομένων στη Βιοπληροφορική.

Τύπος δεδομένων	Περιγραφή	Βάσεις Δεδομένων (με αριθμό)	Ιατρική Χρησιμότητα
Γενετικά δεδομένα (DNA)	Αλληλουχίες DNA από γονιδιώματα οργανισμών	1. GenBank, 2. Ensembl, 3. UCSC Genome Browser	Εντοπισμός γονιδίων, μεταλλάξεων, διάγνωση κληρονομικών νοσημάτων
Μεταγραφικά δεδομένα (RNA)	Προφίλ γονιδιακής έκφρασης, αλληλουχίες RNA	4. GEO, 5. ArrayExpress, 6. GTEx	Κατανόηση μηχανισμών νόσου, ανάλυση καρκινικών προφίλ
Πρωτεϊνικά δεδομένα	Αλληλουχίες και ιδιότητες πρωτεϊνών	7. UniProt, 8. PDB, 9. Pfam	Συσχέτιση μεταλλάξεων με λειτουργία πρωτεϊνών, σχεδιασμός φαρμάκων
Δομικά δεδομένα	Τρισδιάστατες δομές πρωτεϊνών και μακρομορίων	8. PDB	Πρόβλεψη δράσης φαρμάκων, ερμηνεία παθογόνων αλλαγών
Γονιδιωματικές παραλλαγές	πολυμορφισμοί, μικροελλείψεις, επαναλήψεις, δομικές αλλαγές DNA	10. ClinVar, 11. dbSNP, 12. OMIM, 13. HGMD	Ταυτοποίηση παθογόνων μεταλλάξεων, εξατομικευμένη θεραπεία
Μεταβολομικά δεδομένα	Προφίλ μεταβολιτών σε κύτταρα/ιστούς	14. HMDB, 15. MetaboLights	Ανίχνευση βιοδεικτών, παρακολούθηση ασθενειών
Μικροβίωμα	Γονιδιωματικά δεδομένα μικροοργανισμών που ζουν στον άνθρωπο	16. HMP (Human Microbiome Project)	Συσχέτιση μικροβιώματος με υγεία, λοιμώξεις, ανοσολογικές διαταραχές
Κλινικά δεδομένα	Ιατρικοί φάκελοι, εργαστηριακές εξετάσεις, απεικονιστικά δεδομένα	17. TCGA, κλινικές βάσεις	Συσχέτιση μοριακών ευρημάτων με κλινικά χαρακτηριστικά, πρόγνωση

1. GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>

2. Ensembl: <https://www.ensembl.org/>

3. UCSC Genome Browser: <https://genome.ucsc.edu/>

4. GEO: <https://www.ncbi.nlm.nih.gov/geo/>

5. ArrayExpress: <https://www.ebi.ac.uk/arrayexpress/>

6. GTEx: <https://gtexportal.org/>

7. UniProt: <https://www.uniprot.org/>

8. PDB: <https://www.rcsb.org/>

9. Pfam: <http://pfam.xfam.org/>
10. ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>
11. dbSNP: <https://www.ncbi.nlm.nih.gov/snp/>
12. OMIM: <https://www.omim.org/>
13. HGMD: <http://www.hgmd.cf.ac.uk/>
14. HMDB: <https://hmdb.ca/>
15. MetaboLights: <https://www.ebi.ac.uk/metabolights/>
16. HMP: <https://www.hmpdacc.org/>
17. TCGA: <https://www.cancer.gov/tcga>

Οι πρώτες και πιο γνωστές είναι οι **βάσεις γενετικών αλληλουχιών**, όπως η **GenBank** στις Ηνωμένες Πολιτείες, η **EMBL-Bank** στην Ευρώπη και η **DDBJ** στην Ιαπωνία. Σε αυτές αποθηκεύονται δισεκατομμύρια αλληλουχίες DNA και RNA από διαφορετικούς οργανισμούς, μεταξύ αυτών και ο άνθρωπος, οι οποίες είναι διαθέσιμες για ελεύθερη αναζήτηση. Εξίσου σημαντικοί είναι οι **περιηγητές γονιδιωμάτων**, όπως ο **UCSC Genome Browser (Εικόνα 11.1)** και ο **Ensembl**, οι οποίοι επιτρέπουν την οπτική αναπαράσταση και την εξερεύνηση του ανθρώπινου γονιδιώματος αλλά και άλλων οργανισμών-μοντέλων. Μέσα από αυτά τα εργαλεία μπορεί κανείς να δει τη θέση ενός γονιδίου, τις ρυθμιστικές περιοχές του, τις παραλλαγές που έχουν καταγραφεί στον πληθυσμό και τη συσχέτισή του με νοσήματα. Για τον κλινικό γιατρό, η ανίχνευση μιας σπάνιας παραλλαγής μπορεί να οδηγήσει σε διάγνωση μιας κληρονομικής πάθησης, ενώ η γνώση συγκεκριμένων μεταλλάξεων σε ογκογονίδια καθορίζει την επιλογή στοχευμένων θεραπειών. Πέρα από τις βάσεις γενετικών δεδομένων, υπάρχουν εξειδικευμένες βάσεις που αφορούν τις πρωτεΐνες. Η **UniProt** παρέχει πληροφορίες για την αλληλουχία, τη δομή και τη λειτουργία των πρωτεϊνών, ενώ η **PDB** (Protein Data Bank) συγκεντρώνει τρισδιάστατες δομές που έχουν προσδιοριστεί πειραματικά. Οι πληροφορίες αυτές είναι πολύτιμες για την κατανόηση του μηχανισμού δράσης ενζύμων, για την ερμηνεία παθογόνων μεταλλάξεων που αλλοιώνουν τη δομή μιας πρωτεΐνης και για την ανάπτυξη νέων φαρμάκων που στοχεύουν συγκεκριμένες πρωτεϊνικές περιοχές (**Κεφάλαιο X πρωτεΐνες**).

Ιδιαίτερη θέση κατέχουν οι βάσεις που συνδέουν άμεσα τις γενετικές παραλλαγές με την κλινική εικόνα. Η **ClinVar** καταγράφει παραλλαγές DNA που έχουν ενοχοποιηθεί ή συσχετιστεί με ασθένειες, ενώ η **OMIM** (Online Mendelian Inheritance in Man) αποτελεί τη μεγαλύτερη ηλεκτρονική εγκυκλοπαίδεια γενετικών νοσημάτων, καταγράφοντας τη σχέση μεταξύ γονιδίων και φαινοτύπων. Για τον γιατρό που αντιμετωπίζει έναν ασθενή με ασαφή κλινική εικόνα, αυτές οι βάσεις παρέχουν τα απαραίτητα εργαλεία ώστε να ερμηνεύσει εργαστηριακά ευρήματα και να συνδέσει μια παραλλαγή με την πιθανή παθολογική της σημασία. Η πρόσβαση σε αυτές τις πληροφορίες έχει πλέον απλοποιηθεί σε τέτοιο βαθμό ώστε ο γιατρός αποτελεί συμμετόχο στη διαδικασία ερμηνείας των μοριακών δεδομένων και όχι απλώς αποδέκτης έτοιμων αναφορών από ένα εργαστήριο γενετικής.

Παρακάτω ακολουθεί ένα απλοποιημένο παράδειγμα **καταχώρησης** στην πιο διαδεδομένη βάση Genbank. Με τον όρο καταχώρηση αναφερόμαστε σε μία **ολοκληρωμένη μονάδα**

πληροφορίας που αφορά ένα συγκεκριμένο τύπο βιολογικού δεδομένου. Συχνά η καταχώρηση συνοδεύεται από στοιχεία που σχετίζονται αυτή την καταχώρηση και ονομάζονται **μεταδεδομένα**. Τα μεταδεδομένα διευκολύνουν την αναζήτηση, επιτρέπουν τη σωστή ερμηνεία και καθιστούν δυνατή τη διαχείριση και συσχέτιση της καταχώρησης με άλλα βιολογικά δεδομένα. Η κάθε σειρά μεταδεδομένων έχει αριθμηθεί για την υποβοήθηση της επεξήγησης. Η καταχώρηση είναι διαθέσιμη στην GenBank στο σύνδεσμο https://www.ncbi.nlm.nih.gov/nuccore/NM_001101

1.	LOCUS	NM_001101	1812 bp	mRNA	linear	PRI 28-MAY-2025
2.	DEFINITION	Homo sapiens actin beta (ACTB), mRNA.				
3.	ACCESSION	NM_001101				
4.	VERSION	NM_001101.5				
5.	KEYWORDS	RefSeq; MANE Select.				
6.	SOURCE	Homo sapiens (human)				
7.	ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
8.	PRIMARY	REFSEQ_SPAN	PRIMARY_IDENTIFIER	PRIMARY_SPAN	COMP	
		1-1336	AK130157.1	1-1336		
		1337-1812	BC009636.1	71-546		
9.	FEATURES	Location/Qualifiers				
10.	source	1..1812 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon: 9606 " /chromosome="7" /map="7p22.1" 1..1812 /gene="ACTB" /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1; THC8" /note="actin beta" /db_xref="GeneID: 60 " /db_xref="HGNC: HGNC:132 " /db_xref="MIM: 102630 "				
11.	gene					
12.	exon	1..78 /gene="ACTB" /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1; THC8" /inference="alignment:Splign:2.1.0"				
13.	exon	79..207 /gene="ACTB" /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1; THC8" /inference="alignment:Splign:2.1.0"				
14.	CDS	85..1212 /gene="ACTB" /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1; THC8" /note="beta cytoskeletal actin; PS1TP5-binding protein 1; I(2)-actin" /codon_start=1 /product="actin, cytoplasmic 1" /protein_id="NP_001092.1" /db_xref="CCDS: CCDS5341.1 " /db_xref="GeneID: 60 " /db_xref="HGNC: HGNC:132 " /db_xref="MIM: 102630 " /translation="MDDDI AALVVDNGSGMCKAGFAGDDAPRAVFP SIVGRPRHQGV M VGMGQKDSYVGDEA QSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHP VLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGR TTGIVMDSGDGVT HTVPIYEGYALPHAILRLDLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCY VALDFEQEMATAASSSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHE TTFNSIMKCDVDIRKLDYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPE RKYSVWIGGSILASLSTFQQMWISKQ EYDESGPSIVHRKCF"				
15.	exon	208..447 /gene="ACTB" /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1; THC8" /inference="alignment:Splign:2.1.0"				

```

16.      exon                448..886
                        /gene="ACTB"
                        /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1;
                        THC8"
                        /inference="alignment:Splign:2.1.0"
17.      exon                887..1068
                        /gene="ACTB"
                        /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1;
                        THC8"
                        /inference="alignment:Splign:2.1.0"
18.      exon                1069..1812
                        /gene="ACTB"
                        /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1;
                        THC8"
                        /inference="alignment:Splign:2.1.0"
19.      regulatory         1787..1792
                        /regulatory_class="polyA_signal_sequence"
                        /gene="ACTB"
                        /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1;
                        THC8"
                        /note="hexamer: AATAAA"
20.      polyA site         1812
                        /gene="ACTB"
                        /gene_synonym="BKRNS; BNS; BRWS1; CSMH; DDS1; PS1TP5BP1;
                        THC8"
                        /note="major polyA site"
21. ORIGIN
    1 accgccgaga ccgcgtccgc cccgcgagca cagagcctcg cctttgccga tccgccgccc
    61 gtccacaccc gccgccagct caccatggat gatgatatcg ccgcgctcgt cgtcgacaac
    121 ggctccggca tgtgcaaggc cggtcttcgcg gccgacgatg ccccccgggc cgtcttcccc
    181 tccatcgttg gccgccccag gcaccagggc gtgatgggtg gcatgggtca gaaggattcc
    241 tatgtgggag acgaggccca gagcaagaga gccatcctca ccctgaagta ccccatcgag
    301 cacggcatcg tcaccaactg ggacgacatg gagaaaatct gccaccacac cttctacaat
    361 gagctgcgtg tggctcccga ggagcaccgc gtgctgctga ccgaggcccc cctgaacccc
    421 aaggccaacc gcgagaagat gacccagatc atgtttgaga ctttcaacac cccagccatg
    481 tacgttgcta tccaggctgt gctatccctg tacgcctctg gccgtaccac tggcatcggt
    541 atggactccg gtgacggggg caccacact gtgccatctc acgaggggta tgcctctccc
    601 catgccatcc tgcgtctgga cctggctggc cgggacctga ctgactacct catgaagatc
    661 ctacccgagc gcggctacag cttcaccacc acggccgagc gggaaatcgt gcgtgacatt
    721 aaggagaagc tgtgctacgt cgccctggac ttcgagcaag agatggccac ggctgcttcc
    781 agtcctctcc tggagaagag ctacgagctg cctgacggcc aggtcatcac cattggcaat
    841 gagcggttcc gctgccctga ggcactcttc cagccttctc tcctgggcat ggagtcctgt
    901 ggcattccac aaactacctt caactccatc atgaagtgtg acgtggacat ccgcaaaagc
    961 ctgtacgcca acacagtgtc gtctggcggc accaccatgt accctggcat tgcgacagc
    1021 atgcagaagg agatcactgc cctggcacc cgcacaatga agatcaagat cattgtctct
    1081 cctgagcgca agtactccgt gtggatcggc ggctccatcc tggcctcgct gtccaccttc
    1141 cagcagatgt ggatcagcaa gcaggagtat gacgagtcgg gccccctccat cgtccaccgc
    1201 aaatgcttct agcgggacta tgacttagtt gcgttacacc ctttcttgac aaaacctaac
    1261 ttgcgcagaa aacaagatga gattggcatg gctttatttg ttttttttgt tttgttttgg
    1321 tttttttttt ttttttggtc tgactcagga tttaaaaact ggaacggtga aggtgacagc
    1381 agtcgggttg agcagagcat ccccaaagt cacaatgtgg ccgaggactt tgattgcaca
    1441 ttgttgtttt tttaatagtc attocaaata tgagatgcgt tgttacagga agtcccttgc
    1501 catcctaaaa gccacccccc ttctctctaa ggagaatggc ccagtcctct cccaagtcca
    1561 cacaggggag gtgatagcat tgctttcgtg taaattatgt aatgcaaaat ttttttaatc
    1621 ttgcgcctaa tactttttta ttttgtttta ttttgatga tgagccttcg tgccccccct
    1681 tccccctttt ttgtcccca acttgagatg tatgaaggct tttgtctctc ctgggagtg
    1741 gtggaggcag ccagggttta cctgtacact gacttgagac cagttgaata aaagtgcaca
    1801 ccttaaaaat ga
    //

```

Η κάθε σειρά περιέχει έναν ή περισσότερους τύπου μεταδεδομένων που αναλύονται ως εξής:

1. LOCUS → Δίνει βασικές πληροφορίες για την αλληλουχία:

- όνομα (*NM_001101*),
- μήκος (*1812 bp*),
- τύπος μορίου (*mRNA*),
- μορφή (*linear*),
- οργανισμική ομάδα (*PR1 = primate*),
- ημερομηνία τελευταίας ενημέρωσης (*28-MAY-2025*).

2. DEFINITION → Σύντομη περιγραφή του τι αντιπροσωπεύει η αλληλουχία. Εδώ: το mRNA του ανθρώπινου *ACTB* (β-ακτίνη).

3. ACCESSION → Σταθερός μοναδικός κωδικός αναγνώρισης της καταχώρησης (όχι εκδόσεων).
Π.χ. NM_001101.

4. VERSION → Ο αριθμός accession μαζί με τον δείκτη έκδοσης της αλληλουχίας. Π.χ. NM_001101.5 σημαίνει 5η αναθεώρηση.

5. KEYWORDS → Λέξεις-κλειδιά για ευκολότερη αναζήτηση.

6. SOURCE → Οργανισμός προέλευσης του δείγματος.

7. ORGANISM → Πλήρης ταξινομική ιεραρχία του οργανισμού.

8. PRIMARY → Δείχνει τα αρχικά (πρωτογενή) δεδομένα που χρησιμοποιήθηκαν για να δημιουργηθεί η συναρμολογημένη αλληλουχία RefSeq.

9. FEATURES → Ενότητα σχολιασμένων χαρακτηριστικών (annotation) για τα βιολογικά στοιχεία της αλληλουχίας.

10. source → Γενική πληροφορία για το πλήρες εύρος της αλληλουχίας:

- οργανισμός,
- είδος μορίου (mRNA),
- χρωμόσωμα (7),
- χάρτης θέσης (7p22.1),
- ταξινομικό αναγνωριστικό.

11. gene → Περιγράφει το γονίδιο *ACTB* και ονόματα, συνώνυμα και αριθμό καταχώρησης σε άλλες βάσεις δεδομένων

12–13. exon (1..78, 79..207)
→ Σηματοδοτούν τα εξώνια που εντοπίζονται σε αυτά τα τμήματα της αλληλουχίας, μαζί με συνώνυμα και πηγή ευθυγράμμισης.

14. CDS (85..1212) → *Coding DNA Sequence* (κωδική περιοχή):

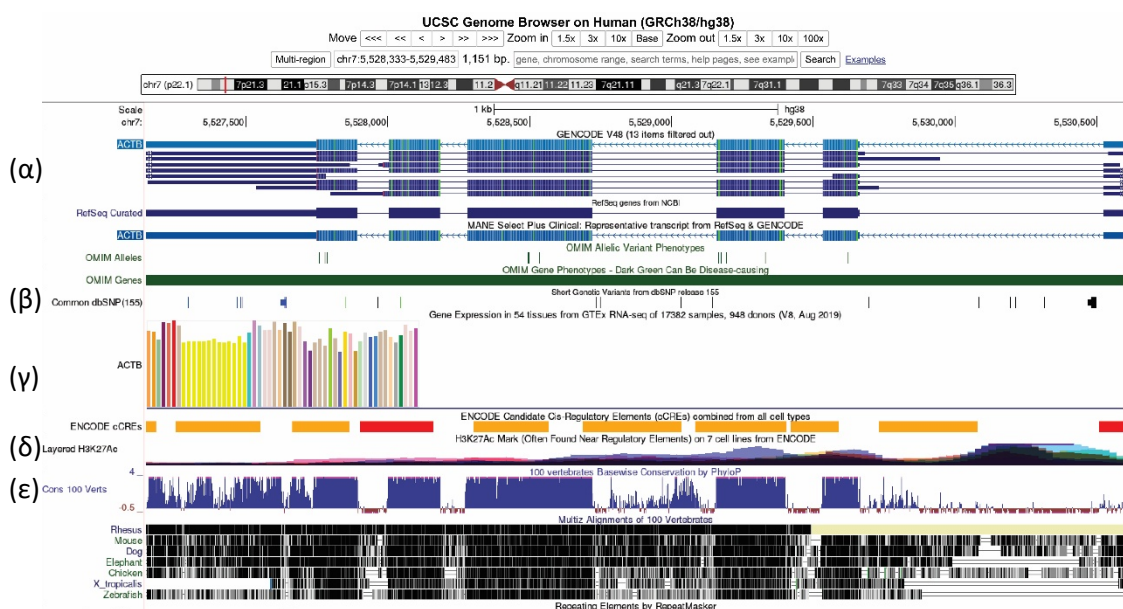
- Πού ξεκινά και τελειώνει η μετάφραση.
- Περιγραφή προϊόντος πρωτεΐνης (actin, cytoplasmic 1).
- Αναγνωριστικά πρωτεΐνης (NP_001092.1).
- Σειρά αμινοξέων σε μορφή μετάφρασης.

15–18. exon (208..1812) → Πρόσθετα εξώνια του γονιδίου *ACTB* σε αντίστοιχες θέσεις.

19. regulatory → Ρυθμιστική αλληλουχία (1787..1792): μοτίβο AATAAA πολυαδενυλίωσης.

20. polyA_site → Θέση πολυαδενυλίωσης (1812) — σημείο στο οποίο ολοκληρώνεται η ωρίμανση του mRNA με polyA tail.

21. ORIGIN → Η ίδια η αλληλουχία βάσεων (DNA/RNA), σε νουκλεοτιδική μορφή. Παρατηρείστε ότι η καταχώρηση αυτή αφορά mRNA αλλά η αναπαράσταση γίνεται με τις αζωτούχες βάσεις του DNA. Η χρήση των βάσεων DNA είναι καθολική στη GenBank καθώς ένας αλγόριθμος «μεταγραφής» μπορεί να συνάγει τις βάσεις του RNA με τη χρήση του μεταδεδομένου του τύπου του νουκλεϊκού οξέος από το LOCUS.



Εικόνα 11.1 Άποψη του UCSC Genome Browser στην χρωμοσωμική περιοχή που είναι το γονίδιο της ακτίνης B στο γονιδίωμα του ανθρώπου. Η εικόνα που προέρχεται από το UCSC Genome Browser απεικονίζει ένα τμήμα του ανθρώπινου γονιδιώματος hg38 στην περιοχή μεταξύ των νουκλεοτιδίων 5.528.333-5.529.483 του χρωμοσώματος 7 (Chr7), όπου βρίσκεται το γονίδιο ACTB. (α) Η μπλε μπάρα (track) αναπαριστά την περιοχή των εξονίων του γονιδίου και οι σκουρόχρωμες μπλε μπάρες, από κάτω, αναπαριστούν τα διαφορετικά μετάγραφα που προκύπτουν μετά από εναλλακτική ωρίμανση. Τα βελάκια πάνω στη μπλε μπάρα αποτελούν ένδειξη ότι η μεταγραφή βρίσκεται στο αρνητικό (αντίρροπο) κλώνο. (β) Παράλληλα εμφανίζονται επιπλέον επίπεδα όπως το OMIM που συνδέει την περιοχή με φαινότυπους και ασθένειες και το dbSNP που καταγράφει γνωστούς πολυμορφισμούς. (γ) Δεδομένα έκφρασης του γονιδίου σε διάφορους ιστούς μετά από RNA-seq αναδुकείουν τα μοτίβα έκφρασης του κάθε γονιδίου στον οργανισμό. (δ) Το «layered H3K27Ac» που εμφανίζεται ως επιλογή στον UCSC Genome Browser είναι ένα track που σχετίζεται με επιγενετικές τροποποιήσεις και συγκεκριμένα με την ακετυλίωση της λυσίνης 27 της ιστόνης H3 (H3K27Ac) σε διάφορους ιστούς (ιστογράμματα διαφόρων χρωμάτων). Το πάχος του κάθε ιστογράμματος αντιστοιχεί με την τοποθέτηση της H3K27Ac στην κάθε βάση της μεταγραφόμενης περιοχής. Η συγκεκριμένη τροποποίηση λειτουργεί ως δείκτης ενεργών ρυθμιστικών περιοχών του γονιδιώματος, προέρχεται από δεδομένα του προγράμματος ENCODE και είναι βασισμένη σε πειράματα ChIP-seq. (ε) Τέλος, περιέχονται δεδομένα συγκριτικής γονιδιωματικής, όπως η συντηρητικότητα σε διαφορετικά είδη μέσω πολλαπλής στοίχισης (π.χ. έναντι 100 άλλων σπονδυλωτών). Η εικόνα αποτελεί αποτύπωση από το σύνδεσμο του UCSC Genome Browser (<https://genome.ucsc.edu>).

11.2 Ανάλυση Αλληλουχιών DNA και RNA

Η ανάλυση των βιολογικών αλληλουχιών αποτελεί την καρδιά της βιοπληροφορικής και ίσως το πιο θεμελιώδες εργαλείο για τη σύγχρονη ιατρική έρευνα και κλινική πράξη. Οι αλληλουχίες DNA, RNA και πρωτεϊνών είναι οι «γλώσσες» με τις οποίες εκφράζεται η βιολογική πληροφορία, και η κατανόησή τους επιτρέπει να συνδυαστούν τα μοριακά δεδομένα με τον φαινότυπο του ασθενούς. Δεν είναι καθόλου τυχαίοι οι όροι αντιγραφή, μεταγραφή και μετάφραση που χρησιμοποιούνται στη βιολογία αφού θέλουν να εκφράσουν ακριβώς την αναφορά σε ένα σύστημα γραφής και κωδικοποίησης της γλώσσας της ζωής.

Η επεξεργασία των αλληλουχιών απαιτεί τη χρήση προγραμματιστικών εντολών χειρισμού κειμένου (string), καθώς κάθε αλληλουχία DNA, RNA ή πρωτεΐνης μπορεί να θεωρηθεί ως μια «**αλυσίδα χαρακτήρων**» που αποτελείται από τα γράμματα που περιλαμβάνει ο γενετικός κώδικας (νουκλεοτίδια και αμινοξέα). Έτσι, οι ίδιες μέθοδοι που εφαρμόζονται στην πληροφορική για την αναζήτηση, την αντικατάσταση ή την ταξινόμηση χαρακτήρων σε ένα κείμενο, βρίσκουν άμεση εφαρμογή και στη βιοπληροφορική. Με αυτόν τον τρόπο μπορούν να εντοπιστούν συγκεκριμένα μοτίβα βάσεων, να μετρηθεί η συχνότητα εμφάνισης ενός νουκλεοτιδίου, να ανιχνευθούν μεταλλάξεις ή να συγκριθούν δύο αλληλουχίες.

Η στοίχιση αλληλουχιών (sequence alignment) είναι η διαδικασία με την οποία δύο ή περισσότερες αλληλουχίες συγκρίνονται για να εντοπιστούν περιοχές ομοιότητας αλλά και διαφοροποίησης (**Εικόνα 11.2**). Η ομοιότητα αυτή υποδηλώνει συχνά κοινή εξελικτική καταγωγή, άρα και πιθανή κοινή λειτουργία. Η στοίχιση μπορεί να είναι **ολική**, όταν συγκρίνεται ολόκληρο το μήκος των αλληλουχιών, ή **τοπική**, όταν αναζητούνται μόνο οι περιοχές με υψηλή ομοιότητα. Η τοπική στοίχιση είναι ιδιαίτερα χρήσιμη στη βιοϊατρική, διότι οι κρίσιμες μεταλλάξεις συχνά περιορίζονται σε συγκεκριμένα τμήματα ενός γονιδίου ή μιας πρωτεΐνης.

Αλληλουχία1	CCATGTCAAGTTCCGGCTCCTGAATTCTTTACGGAGGTG	39
Αλληλουχία2	CCATGCCAGGTTCGGCCCC-GAATTCTTCACAGAAGTG	38
	***** ** ***** ** ***** ** ** *	

Αλληλουχία1	YIAKQRQISFVKSHFSRSDILFLIYQDIKELGME	34
Αλληλουχία2	YIAKQRQ--IVKSHFSRQDILDLIYQDLKELGME	32
	***** : ***** .*** ***** :*****	

Εικόνα 11.2 Στοίχιση αλληλουχιών. Η πρώτη στοίχιση της εικόνας αποτελεί σύγκριση δύο αλληλουχιών νουκλεοτιδίων. Οι ομόλογες περιοχές στοιχίζονται η μία κάτω από την άλλη και στις θέσεις που υπάρχει συμφωνία μεταξύ των νουκλεοτιδίων των δύο αλληλουχιών ο αλγόριθμος τοποθετεί έναν αστερίσκο (*). Τα κενά ανάμεσα στους αστερίσκους υποδεικνύουν την διαφοροποίηση στη συγκεκριμένη θέση μεταξύ των δύο αλληλουχιών. Στο τέλος της κάθε γραμμής αναφέρονται τα νουκλεοτίδια της κάθε αλληλουχίας μέχρι το τέλος της δεδομένης γραμμής. Η δεύτερη στοίχιση αποτελεί σύγκριση μεταξύ δύο αλληλουχιών αμινοξέων. Οι αστερίσκοι ακολουθούν τη λογική που προαναφέρθηκε, ωστόσο εμφανίζονται και άλλα δύο σύμβολα της στοίχισης. Η άνω και κάτω τελεία (:) υποδεικνύει μεγάλη ομοιότητα στις φυσικοχημικές ιδιότητες των δύο αμινοξέων, ενώ η τελεία (.) υποδεικνύει μικρότερη ομοιότητα. Το κενό στην περίπτωση της στοίχισης αμινοξέων σημαίνει απόλυτη απόκλιση στις ιδιότητες, π.χ. ένα υδρόφοβο αμινοξύ έχει αντικατασταθεί από ένα υδρόφιλο. Για να είναι ευκρινής η στοίχιση απαιτεί την χρήση της γραμματοσειράς Courier, καθώς σε αυτήν το κάθε σύμβολο καταλαμβάνει διάστημα ίσου πλάτους.

Η **πολλαπλή στοίχιση αλληλουχιών (Multiple Sequence Alignment, MSA)** επιτρέπει τη σύγκριση πολλών αλληλουχιών ταυτόχρονα. Μέσω της MSA εντοπίζονται συντηρημένα

μοτίβα, τα οποία μπορεί να υποδεικνύουν περιοχές κρίσιμες για τη λειτουργία ενός μακρομορίου. Για παράδειγμα, η στοίχιση πολλών ειδών αιμοσφαιρινών ανέδειξε την ύπαρξη συγκεκριμένων **συντηρημένων** θέσεων αμινοξέων που είναι απαραίτητες για τη δέσμευση της αίμης. Η ανάλυση αυτή όχι μόνο επιβεβαίωσε τη λειτουργική σημασία αυτών των θέσεων, αλλά επέτρεψε και την κατανόηση της εξελικτικής σταθερότητας του μηχανισμού δέσμευσης του οξυγόνου σε διαφορετικούς οργανισμούς. Η MSA χρησιμοποιείται επίσης για την κατασκευή φυλογενετικών δέντρων, καθώς οι ομοιότητες και οι διαφορές ανάμεσα σε πολλές αλληλουχίες δίνουν πληροφορίες για τις εξελικτικές σχέσεις μεταξύ ειδών ή γονιδιακών οικογενειών. Επιπλέον, επιτρέπει την πρόβλεψη δομικών και λειτουργικών περιοχών σε πρωτεΐνες που δεν έχουν μελετηθεί πειραματικά, με βάση τη σύγκρισή τους με άλλες πρωτεΐνες γνωστής δομής.

Το **FASTA** υπήρξε ένα από τα πρώτα προγράμματα στοίχισης και εξακολουθεί να χρησιμοποιείται, αν και έχει σε μεγάλο βαθμό αντικατασταθεί από το BLAST. Το πιο ευρέως χρησιμοποιούμενο εργαλείο απλής στοίχισης κατά ζεύγη είναι το **BLAST (Basic Local Alignment Search Tool)**, που δημιουργήθηκε στις αρχές της δεκαετίας του 1990 και από τότε καθιερώθηκε ως το βασικό πρόγραμμα για τη σύγκριση αλληλουχιών. Με το BLAST, ο ερευνητής εισάγει μια αλληλουχία DNA, RNA ή πρωτεΐνης και το πρόγραμμα επιστρέφει τις πιο όμοιες αλληλουχίες από τις μεγάλες βάσεις δεδομένων. Τα πιο διαδεδομένα λογισμικά για πολλαπλή στοίχιση είναι τα **Clustal Omega**, το **MUSCLE** και το **MAFFT** τα οποία έχουν και τη δυνατότητα δημιουργίας φυλογενετικών δένδρων που αποτυπώνουν γραφικά τη συγγένεια μεταξύ των συγκρινόμενων αλληλουχιών (**Εικόνα 11.3**).

Αλληλουχία1	CCATGTCAAGTTCCGCATCCTGAATTCTTTACGGAGGTA	39
Αλληλουχία4	CCATGTTCAGGTTCCAAAACCTGAATTCTTTACAGAGGTA	39
Αλληλουχία5	CCATGTCAAGTTCCAAAACCTGAATTCTTTACAGAGGTA	39
Αλληλουχία2	CGATGCGGGGTTCCG-GCCCCGAATTCTTCACAGAAGTG	38
Αλληλουχία3	CCATGGCAGGTTCCG-GCTCCGAATTCTTCAGAGAACTG	38
	* * * * * * * * * * * * * * * * * * * * * * * *	



Εικόνα 11.3. Πολλαπλή στοίχιση και φυλογενετικά δέντρα. Στην ανάλυση διακρίνονται οι θέσεις συντήρησης (αστερίσκοι) και οι διαφοροποιήσεις μεταξύ όλων των αλληλουχιών. Δηλαδή, η έλλειψη αστερίσκου σημαίνει την απουσία ταυτότητας έστω και μία από τις συγκρινόμενες αλληλουχίες. Το φυλογενετικό δέντρο που προκύπτει απεικονίζει τις εξελικτικές σχέσεις μεταξύ τους, με τις Αλληλουχίες 1, 4 και 5 να εμφανίζουν μεγαλύτερη συγγένεια σε σχέση με τις Αλληλουχίες 2 και 3. Οι κόμβοι αποτελούν τους κοινούς προγόνους των αντίστοιχων κλάδων που ξεκινούν από αυτούς τους κόμβους, ενώ το μήκος των κλάδων εξαρτάται από το βαθμό διαφοροποίησης σε επίπεδο αλληλουχίας. Παρατηρήστε τις

διαφορές και τις ομοιότητες στις αλληλουχίες και πώς αυτές απεικονίζονται εξελικτικά στο δέντρο.

Η κατασκευή **φυλογενετικών δέντρων** βασίζεται στη στοίχιση πολλών αλληλουχιών (Multiple Sequence Alignment – MSA). Από τη στιγμή που στοιχίζονται οι αλληλουχίες, υπολογίζεται η γενετική απόσταση μεταξύ τους, δηλαδή το πόσο διαφορετικές είναι. Με μαθηματικούς αλγορίθμους, αυτές οι αποστάσεις μετατρέπονται σε ένα δενδροειδές σχήμα που αναπαριστά τις πιθανές πορείες της εξέλιξης της δεδομένης γονιδιακής περιοχής. Στο δέντρο, κάθε κόμβος αντιπροσωπεύει ένα κοινό πρόγονο και κάθε κλαδί δείχνει την εξελικτική απόκλιση. Όσο πιο κοντά βρίσκονται δύο αλληλουχίες στο δέντρο, τόσο πιο πρόσφατο είναι το κοινό τους εξελικτικό σημείο ή αλλιώς **κοινός πρόγονος (Εικόνα 11.3)**.

11.3 Τεχνολογίες αλληλούχησης

Η εξέλιξη των τεχνολογιών αλληλούχησης υπήρξε καθοριστική για την πρόοδο της βιοϊατρικής επιστήμης. Από την πρώτη μέθοδο Sanger, που επέτρεψε την ανάγνωση του DNA με ακρίβεια αλλά σε περιορισμένη κλίμακα, έως τις σύγχρονες πλατφόρμες αλληλούχησης επόμενης γενιάς (Next Generation Sequencing, NGS) και τις νεότερες τεχνολογίες τρίτης γενιάς με δυνατότητα ανάγνωσης πολύ μεγάλων μορίων DNA, η πορεία αυτή αντικατοπτρίζει τη διαρκή ανάγκη της επιστήμης για πιο γρήγορη, φθηνή και ολοκληρωμένη πρόσβαση στο γονιδίωμα.

Η **μέθοδος Sanger**, που αναπτύχθηκε στα τέλη της δεκαετίας του 1970, αποτέλεσε το θεμέλιο λίθο της γονιδιωματικής. Βασίζεται στη χρήση τροποποιημένων νουκλεοτιδίων DNA (dideoxynucleotides), τα οποία σταματούν την επιμήκυνση της αλυσίδας DNA κατά την αντιγραφή. Με τη βοήθεια φθορίζοντων χρωστικών, κάθε τύπος νουκλεοτιδίου αναγνωρίζεται ξεχωριστά, επιτρέποντας την ανάγνωση της αλληλουχίας. Δηλαδή γνωρίζοντας το πιο είναι το νουκλεοτίδιο που τερμάτισε την αλυσίδα γνωρίζουμε στη δεδομένη θέση ποιο είναι το νουκλεοτίδιο της αλληλουχίας. Για πάνω από δύο δεκαετίες, η μέθοδος Sanger αποτέλεσε τη βασική τεχνολογία για τη μελέτη γονιδίων, για την ταυτοποίηση μεταλλάξεων και για την ολοκλήρωση του **Προγράμματος Ανθρώπινου Γονιδιώματος**, το οποίο διήρκεσε σχεδόν 15 χρόνια και κόστισε δισεκατομμύρια δολάρια. Το πλεονέκτημά της ήταν η ακρίβεια· μπορούσε να δώσει σχεδόν τέλεια ανάγνωση μικρών τμημάτων DNA. Το μειονέκτημα, ωστόσο, ήταν η χαμηλή της απόδοση, καθώς λίγες εκατοντάδες βάσεις μπορούσαν να αναλυθούν ανά αντίδραση (μέχρι ~800 βάσεις), καθιστώντας την ακατάλληλη για την εποχή της «γονιδιωματικής». Η μέθοδος Sanger εξακολουθεί να χρησιμοποιείται μέχρι σήμερα, κυρίως για μικρής κλίμακας επιβεβαιώσεις γενετικών πολυμορφισμών και για ειδικές διαγνωστικές εξετάσεις, αλλά η κυριαρχία της έληξε με την εμφάνιση των τεχνολογιών δεύτερης γενιάς.

Η **δεύτερη γενιά** τεχνολογιών, γνωστή ως **Next Generation Sequencing (NGS)**, εισήχθη στις αρχές της δεκαετίας του 2000 και αποτέλεσε πραγματική επανάσταση. Σε αντίθεση με τη μέθοδο Sanger, οι πλατφόρμες NGS έχουν τη δυνατότητα **μαζικής παράλληλης αλληλούχησης**, δηλαδή εκατομμύρια ή και δισεκατομμύρια τμήματα DNA μπορούν να αλληλουχηθούν ταυτόχρονα. Η ανάγνωση του κάθε τμήματος οδηγεί στην παραγωγή μιας **ανάγνωσης**, δηλαδή μίας αλληλουχίας DNA που έχει προέλθει από αλληλούχηση.

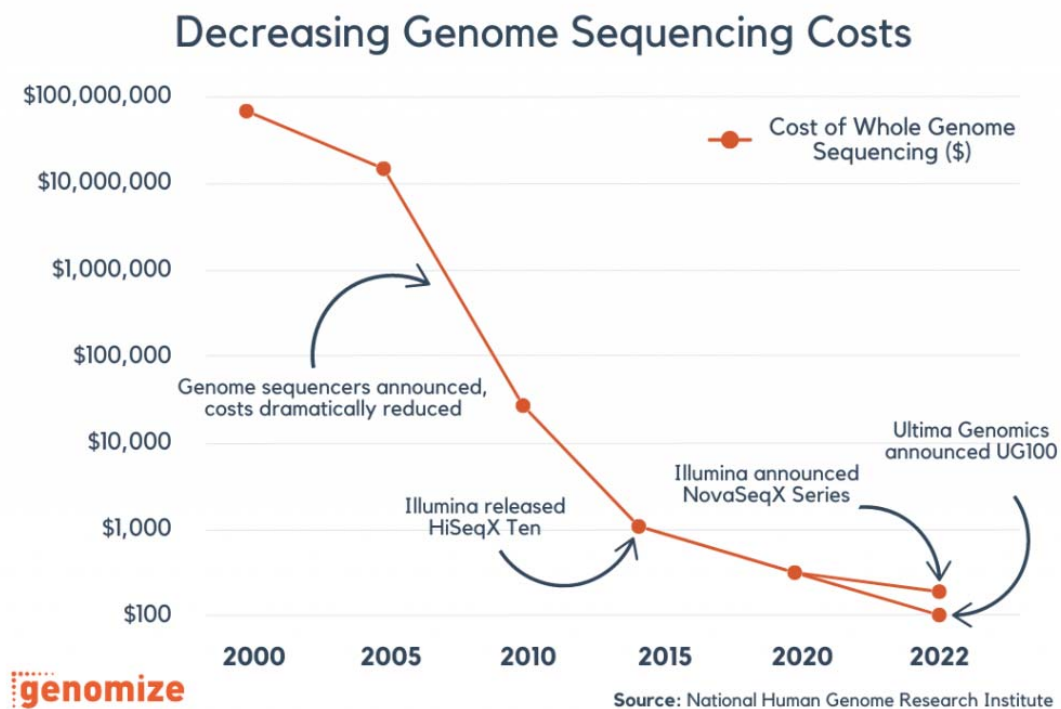
Οι κυριότερες τεχνικές που αναπτύχθηκαν ήταν:

- **Illumina sequencing (αλληλούχηση κατά τη σύνθεση):** Η πιο διαδεδομένη μέθοδος NGS, που χρησιμοποιεί φθορίζοντα νουκλεοτίδια τερματιστές της σύνθεσης του

DNA. Κάθε κύκλος αλληλούχησης καταγράφει την ενσωμάτωση ενός νουκλεοτιδίου, και η διαδικασία επαναλαμβάνεται μέχρι να διαβαστεί όλο το θραύσμα. Παρέχει εξαιρετικά υψηλή ακρίβεια (>99,9%) και είναι η βασική τεχνολογία για κλινικές εφαρμογές.

- **Ion Torrent:** Χρησιμοποιεί την ανίχνευση αλλαγών pH κατά την ενσωμάτωση νουκλεοτιδίων. Δηλαδή όταν ένα νουκλεοτίδιο ενσωματωθεί κατά τον πολυμερισμό του DNA η μικρή αλλαγή στο pH που προκαλεί η αντίδραση καταγράφεται μαζί με τον τύπο του νουκλεοτιδίου που των προκάλεσε.

Με την εμφάνιση της τεχνολογίας NGS, η **ολική γονιδιωματική αλληλούχηση (WGS)** έγινε τεχνικά εφικτή σε ευρεία κλίμακα και χρησιμοποιείται σήμερα στην κλινική πράξη. Διάφορες άλλες κλινικές εφαρμογές ήταν επίσης άμεσες. Στη γενετική, χιλιάδες σπάνια σύνδρομα διαγνώστηκαν χάρη στην πιο στοχευμένη αλλά και πιο οικονομική μέθοδο της αλληλούχησης των εξονίων, δηλαδή του τμήματος εκείνου του γονιδιώματος που καταλήγει στα ώριμα RNA. Η μεθοδολογία αυτή ονομάστηκε **εξωμική αλληλούχηση** από την προσθήκη της κατάληξης «ωμική» που υποδηλώνει μια ολιστική προσέγγιση στη λέξη εξόνιο. Η στοχευμένη ανάλυση μεταλλάξεων σε καρκινικού όγκους ή σε υγρές βιοψίες (συνήθως αίμα) έδωσε τη δυνατότητα ανίχνευσης μεταλλάξεων που σχετίζονται με την εμφάνιση, την εξέλιξη αλλά και τη θεραπεία των καρκίνων.



Εικόνα 11.4. Η μείωση του κόστους αλληλούχησης ενός πλήρους γονιδιώματος μεταξύ 2000 και 2022. Τα βέλη δείχνουν το πώς η ανάπτυξη της τεχνολογίας αλληλούχησης βοήθησε στη μείωση του κόστους. (Πηγή: NHGRI)

Παρά τα πλεονεκτήματά της, η NGS έχει αρκετούς περιορισμούς. Ο βασικότερος περιορισμός είναι ότι παράγει **σύντομες αναγνώσεις (short reads)**, μήκους 50–300 βάσεων, γεγονός που καθιστά δύσκολη την ανάλυση περιοχών με πολλά επαναλαμβανόμενα στοιχεία ή μεγάλες δομικές παραλλαγές. Για να ξεπεραστεί αυτό, αναπτύχθηκαν οι τεχνολογίες τρίτης γενιάς.

Οι τεχνολογίες τρίτης γενιάς χαρακτηρίζονται από τη δυνατότητα παραγωγής **μεγάλων αναγνώσεων (long reads)**, δηλαδή αλληλουχιών χιλιάδων έως και εκατοντάδων χιλιάδων βάσεων χωρίς κατακερματισμό. Οι δύο κυριότερες πλατφόρμες είναι:

- **Pacific Biosciences (PacBio):** Βασίζεται στην ανίχνευση φθοριζόντων νουκλεοτιδίων σε πραγματικό χρόνο καθώς το ένζυμο DNA πολυμεράση αντιγράφει το μόριο. Παράγει αναγνώσεις άνω των 10.000 βάσεων, με δυνατότητα υψηλής ακρίβειας μέσω κυκλικής αλληλούχησης³ (circular consensus sequencing, CCS).
- **Oxford Nanopore Technologies (ONT):** Χρησιμοποιεί νανοπόρους από πρωτεΐνες μέσα από τους οποίους διέρχεται το DNA ή το RNA. Οι αλλαγές στο ηλεκτρικό ρεύμα καταγράφουν τη σειρά των βάσεων. Μπορεί να παράγει αναγνώσεις άνω των 100.000 βάσεων και προσφέρει τη μοναδική δυνατότητα φορητών συσκευών (π.χ. MinION), που χρησιμοποιούνται ακόμα και επί τόπου στο πεδίο ή σε λιγότερο εξοπλισμένα εργαστήρια (**Εικόνα 11.6**)

Οι τεχνολογίες αυτές επιτρέπουν την πλήρη ανάλυση σύνθετων περιοχών του γονιδιώματος, την ανίχνευση μεγάλων δομικών παραλλαγών και την ακριβή χαρτογράφηση επαναλαμβανόμενων στοιχείων.

Μεθοδολογία δημιουργίας βιβλιοθήκης για NGS δεύτερης γενιάς

Η δημιουργία βιβλιοθήκης αλληλούχησης αποτελεί το κρίσιμο βήμα προετοιμασίας δειγμάτων για όλες τις πλατφόρμες NGS. Παρότι κάθε τεχνολογία έχει τις δικές της ιδιαιτερότητες, τα βασικά στάδια της διαδικασίας είναι κοινά και περιλαμβάνουν: απομόνωση και ποιοτικό έλεγχο DNA/RNA, κατακερματισμό, επιδιόρθωση άκρων και προσθήκη προσαρμογέων (adapters), ενίσχυση μέσω PCR και τελικό έλεγχο ποιότητας.

Απομόνωση και ποιοτικός έλεγχος: Η διαδικασία ξεκινά με την προσεκτική απομόνωση του DNA ή RNA από το βιολογικό δείγμα. Η καθαρότητα και η ακεραιότητα του νουκλεϊκού οξέος είναι καθοριστικές, καθώς επηρεάζουν άμεσα την ποιότητα της βιβλιοθήκης. Ο ποσοτικός και ποιοτικός έλεγχος πραγματοποιείται με φασματοφωτομετρία, φθορισμομετρία ή ηλεκτροφόρηση σε πηκτή.

Κατακερματισμός: Για να είναι συμβατά τα μόρια με τις πλατφόρμες NGS, το DNA πρέπει να κοπεί σε μικρότερα τμήματα, συνήθως μήκους 200–500 βάσεων. Αυτό μπορεί να γίνει μηχανικά, με υπερήχους, ενζυμικά, με ειδικές νουκλεάσες ή χημικά. Σε αυτή τη φάση γίνεται και η μετατροπή του RNA σε cDNA. Το κάθε δείγμα σε αυτή τη φάση αντιπροσωπεύεται από μια τεράστια ποικιλία τμημάτων που στο σύνολό τους ονομάζονται **βιβλιοθήκη**.

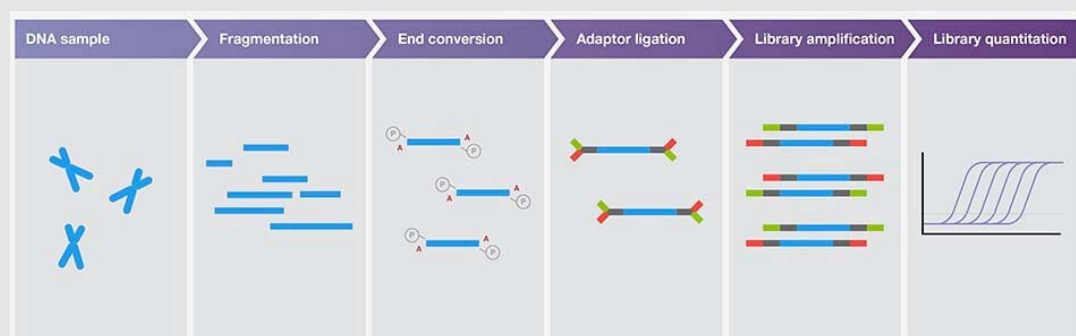
Επιδιόρθωση άκρων: Μετά τον κατακερματισμό, τα άκρα των θραυσμάτων συχνά είναι άνισα και μερικώς αποδομημένα από τη διαδικασία κατακερματισμού. Γι' αυτό χρησιμοποιούνται ειδικά μείγματα ενζύμων που επιδιορθώνουν τα άκρα ώστε να βελτιστοποιεί η αποτελεσματικότητα σύνδεσης με τους προσαρμογείς.

³ Το DNA κόβεται σε τμήματα τα οποία με ειδικούς προσαρμογείς σχηματίζουν κυκλικές δομές. Ένα μόριο πολυμεράσης ξεκινάει την αντιγραφή του κυκλικού μορίου, και καθώς το DNA είναι κυκλικό, η πολυμεράση διαβάζει πολλές φορές το ίδιο μόριο σε διαδοχικούς κύκλους. Αυτό μειώνει την καταχώρηση λανθασμένων βάσεων.

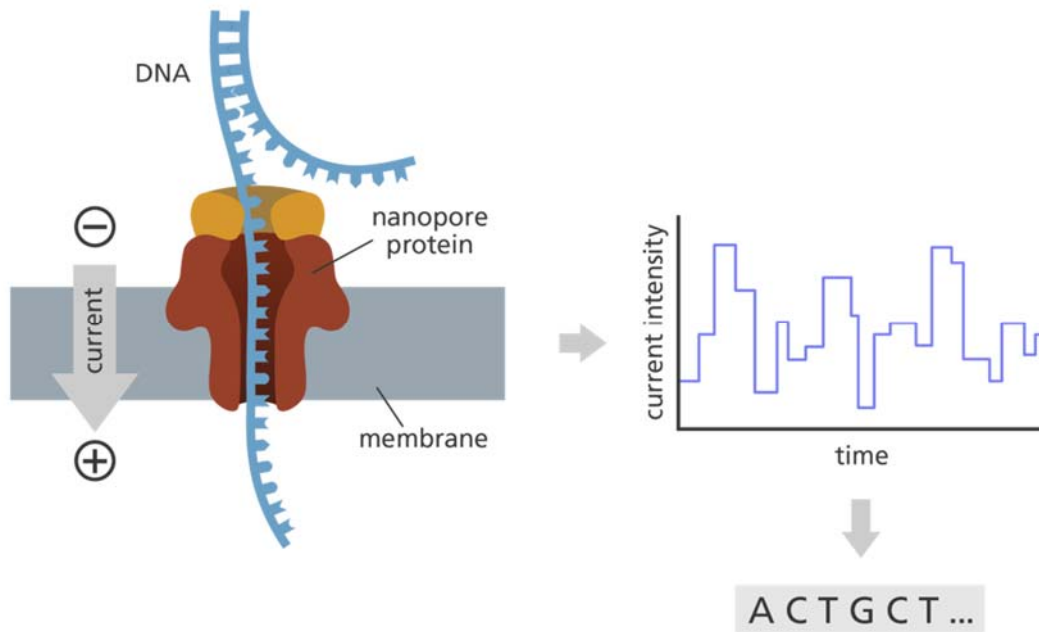
Προσθήκη προσαρμογών και αναγνωριστικών αλληλουχιών: Ειδικές συνθετικές αλληλουχίες DNA, οι προσαρμογείς adaptors και αναγνωριστικές αλληλουχίες (index barcodes), συνδέονται στα άκρα των θραυσμάτων με τη βοήθεια ενζύμων. Οι προσαρμογείς επιτρέπουν την πρόσδεση του κάθε μορίου DNA στη σταθερή φάση του μέσου πάνω στο οποίο θα γίνει η αλληλούχηση. Οι αναγνωριστικές αλληλουχίες αποτελούν ένα ολιγονουκλεοτίδιο με δεδομένη αλληλουχία χαρακτηριστική για το κάθε δείγμα και επιτρέπουν τον συνδυασμό πολλών δειγμάτων στην ίδια ανάλυση (multiplexing). Η βιοπληροφορική ανάλυση μπορεί στη συνέχεια να αναγνωρίσει τις αναγνώσεις με την ίδια αναγνωριστική αλληλουχία και να την αντιστοιχήσει σε ένα δεδομένο δείγμα (demultiplexing).

Ενίσχυση μέσω PCR: Μετά την προσθήκη προσαρμογών, η βιβλιοθήκη συχνά υποβάλλεται σε ενίσχυση μέσω PCR για να αυξηθεί η ποσότητα DNA και να δημιουργηθούν επαρκή αντίγραφα. Ως εκκινητές χρησιμοποιούνται συμπληρωματικά ολιγονουκλεοτίδια προς τους σταθερούς προσαρμογείς που είναι κοινοί σε όλες τις βιβλιοθήκες.

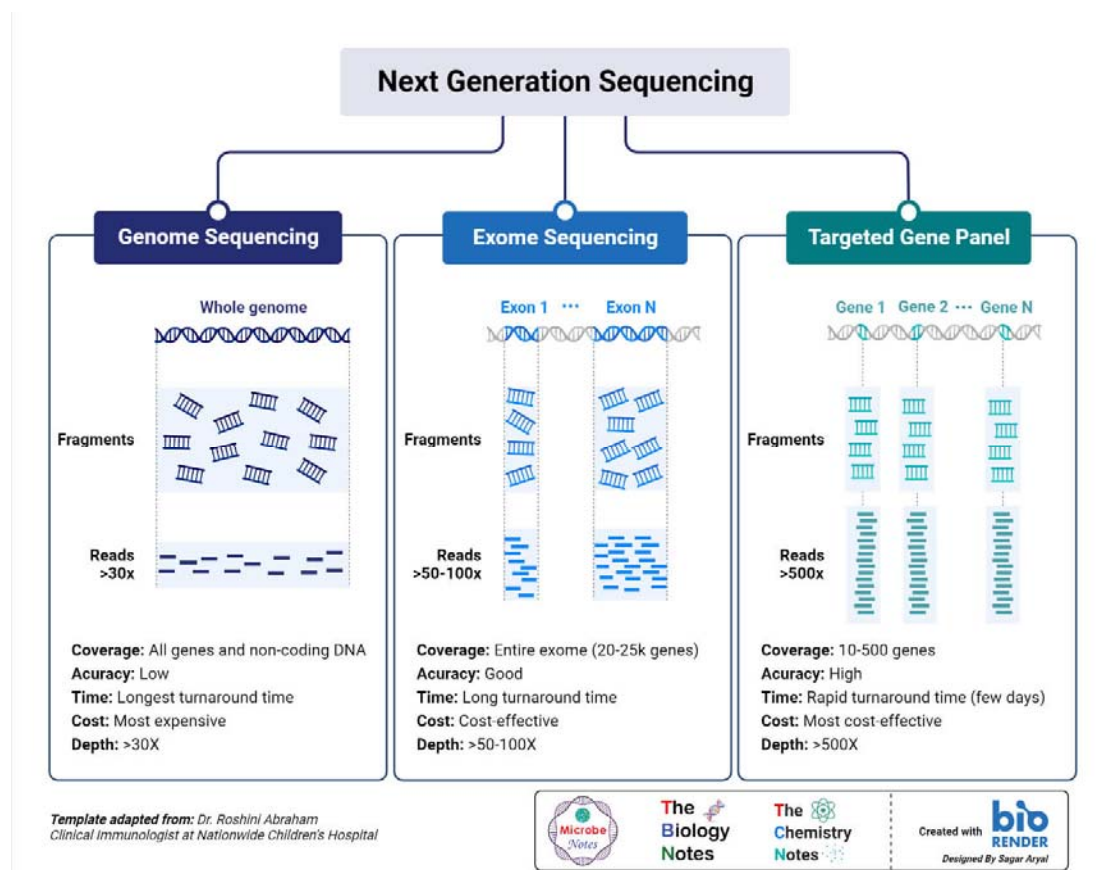
Έλεγχος ποιότητας και ποσοτικοποίηση: Πριν την αλληλούχηση, η βιβλιοθήκη αξιολογείται εκ νέου για το μέγεθος, την καθαρότητα και τη συγκέντρωση. Ο ακριβής έλεγχος είναι ουσιώδης, διότι υπερβολικά μεγάλες ή μικρές βιβλιοθήκες, καθώς και λάθος συγκεντρώσεις, οδηγούν σε αποτυχία αλληλούχησης ή σε χαμηλή ποιότητα δεδομένων.



Εικόνα 11.5 Η μεθοδολογία της δημιουργίας βιβλιοθήκης για ένα NGS δεύτερης γενιάς.



Εικόνα 11.6 Αρχή λειτουργίας της τεχνολογίας Αλληλούχησης Νανοπόρου. Ένα μόριο DNA διέρχεται μέσα από μια πρωτεΐνη-νανοπόρο που είναι ενσωματωμένη σε μεμβράνη. Καθώς κάθε νουκλεοτίδιο περνά, προκαλεί χαρακτηριστικές μεταβολές στο δυναμικό της μεμβράνης. Οι αλλαγές αυτές καταγράφονται ως σήμα έντασης ρεύματος σε σχέση με τον χρόνο και αποκωδικοποιούνται υπολογιστικά στην αντίστοιχη αλληλουχία βάσεων (π.χ. ACTGCT...).



Εικόνα 11.7 Συγκριτική απεικόνιση των βασικών προσεγγίσεων αλληλούχησης επόμενης γενιάς (NGS). Η ολική γονιδιωματική αλληλούχηση (Genome Sequencing) καλύπτει ολόκληρο το γονιδίωμα, συμπεριλαμβανομένων των μη κωδικών περιοχών, αλλά απαιτεί τον περισσότερο χρόνο και το υψηλότερο κόστος. Η εξονική αλληλούχηση (Exome Sequencing) επικεντρώνεται αποκλειστικά στις κωδικές περιοχές (εξόνια) και αποτελεί μια πιο οικονομική και αποδοτική λύση για την ανίχνευση κλινικά σημαντικών μεταλλάξεων. Τα στοχευμένα γονιδιακά πάνελ (Targeted Gene Panels) εξετάζουν επιλεγμένα γονίδια ή ομάδες γονιδίων με πολύ υψηλή ακρίβεια και ταχύτητα, αποτελώντας την πλέον πρακτική προσέγγιση για κλινική διάγνωση σε εξειδικευμένες ενδείξεις.

Ο ρόλος της αλληλούχησης στην αντιμετώπιση των αιματολογικών κακοηθειών

Η ταχεία πρόοδος της τεχνολογίας αλληλούχησης νέας γενιάς έχει μεταμορφώσει την κατανόηση και την κλινική αντιμετώπιση των αιματολογικών κακοηθειών. Μέσα σε λιγότερο από δύο δεκαετίες, η τεχνολογία του NGS εξελίχθηκε από ερευνητικό εργαλείο σε βασική πρακτική των αιματολογικών κλινικών.

Η αναγνώριση των σωματικών «οδηγών» μεταλλάξεων (driver mutations), δηλαδή των μεταλλάξεων που εμφανίζονται τυχαία και οδηγούν στην ανάπτυξη και εξέλιξη μιας κακοήθειας, είναι καίρια για την κατανόηση της βιολογίας της νόσου και την επιλογή θεραπευτικών στρατηγικών. Οι πρόσφατες ταξινομήσεις του Παγκόσμιου Οργανισμού Υγείας (2022) και οι κατευθυντήριες οδηγίες του European LeukemiaNet (ELN, 2022) δίνουν πρωτεύοντα ρόλο στις γενετικές-γονιδιωματικές προσεγγίσεις. Όχι μόνο για τη διάγνωση, αλλά και για την πρόγνωση, τον εντοπισμό της υπολειπόμενης νόσου και τη λήψη αποφάσεων θεραπείας. Έτσι, η NGS εντάσσεται πλέον ως αναπόσπαστο στοιχείο της αιματολογικής διαγνωστικής.

Τα πάνελ DNA είναι η πιο διαδεδομένη εφαρμογή. Ο Παγκόσμιος Οργανισμός Υγείας εισήγαγε το 2022 νέες οντότητες και πρόδρομες καταστάσεις, όπως η κλωνική αιμοποίηση απροσδιόριστου δυναμικού (CHIP) και η κλωνική κυτταροπενία απροσδιόριστης σημασίας (CCUS). Και οι δύο σχετίζονται με την παρουσία σωματικών μεταλλαγών (συχνά σε γονίδια όπως *Dnmt3a*, *Tet2*, *Asx1*), χωρίς όμως να πληρούν τα κριτήρια κακοήθειας. Η μέθοδος NGS επιτρέπει την ευρεία ανίχνευση αυτών των καταστάσεων, παρέχοντας προγνωστικές πληροφορίες για την μελλοντική εξέλιξη τους. Στη μυελοδυσπλαστική νόσο (MDS), αναδείχθηκε η σημασία μεταλλαγών όπως του γονιδίου *Sf3b1* (με θετική πρόγνωση) και του *Trp53* (με δυσμενή πρόγνωση, ειδικά όταν εμπλέκονται παραπάνω από μία μεταλλαγές «multi-hit»). Αντίστοιχα, στην οξεία μυελογενή λευχαιμία (AML), η μέθοδος NGS καθόρισε νέες διαγνωστικές κατηγορίες. Η οντότητα «AML με μυελοδυσπλαστικά χαρακτηριστικά» εγκαταλείφθηκε και αντικαταστάθηκε από την «AML myelodysplasia-related (MR)», η οποία ορίζεται από μεταλλάξεις σε οκτώ γονίδια (π.χ. *Asx1*, *Ezh2*, *Srsf2*). Σημαντικό είναι και το πεδίο της κληρονομικής προδιάθεσης. Γονιδιακές μεταλλάξεις (π.χ. στα γονίδια *Ddx41*, *Runx1*, *Trp53*) προδιαθέτουν σε αιματολογικές κακοήθειες και έχουν άμεση κλινική σημασία, τόσο για τον ασθενή όσο και για τους συγγενείς, κυρίως όταν εξετάζεται η μεταμόσχευση μυελού.

Πίνακας 11.2 Ο ρόλος του NGS στη διάγνωση και θεραπεία των αιματολογικών κακοηθειών.

Νόσος / Κατηγορία	Μετάλλαξη ή σύντηξη που	Κλινική σημασία	Επιλογή/Καθοδήγηση θεραπειάς
-------------------	-------------------------	-----------------	------------------------------

ανιχνεύεται με NGS			
οξεία μυελογενής λευχαιμία	FLT3-ITD / FLT3-TKD	Καθορίζει προγνωστική ομάδα	Midostaurin, Gilteritinib
οξεία μυελογενής λευχαιμία	IDH1/IDH2	Παρουσία οδηγεί σε στοχευμένη θεραπεία	Ivosidenib (IDH1), Enasidenib (IDH2)
οξεία μυελογενής λευχαιμία	CEBPA	Ευνοϊκή πρόγνωση → λιγότερο επιθετική θεραπεία	Αποφυγή πρώιμης μεταμόσχευσης· κλασική χημειοθεραπεία
οξεία μυελογενής λευχαιμία με προδιάθεση στο DDX41	DDX41 κληρονομικής DDEX41 + δευτερογενούς εμφάνισης	Κληρονομική προδιάθεση	Επηρεάζει επιλογή δότη για μεταμόσχευση, καλή ανταπόκριση σε χημειοθεραπεία
μυελοδυσπλαστική νόσος	SF3B1	Ευνοϊκή πρόγνωση	Καθοδήγηση για πιο ήπια θεραπευτικά σχήματα
μυελοδυσπλαστική νόσος / οξεία μυελογενή λευχαιμία	multi-hit TP53	Πολύ δυσμενής πρόγνωση	Κλινικές δοκιμές ή πρώιμη μεταμόσχευση
οξεία λεμφοβλαστική λευχαιμία	BCR::ABL1 σύντηξη	Υποκατηγορία με κακή πρόγνωση	Αναστολείς κινάσης τυροσίνης (Imatinib, Dasatinib)
οξεία λεμφοβλαστική λευχαιμία	BCR::ABL1-like (CRLF2, JAK2, ABL1, PDGFRB κ.λπ.)	Μοιάζει βιολογικά με Ph+ ALL	Αναστολείς κινάσης τυροσίνης ή αναστολείς JAK
οξεία λεμφοβλαστική λευχαιμία	ETV6::RUNX1-like ή TCF3::HLF	Ειδικές υποκατηγορίες	Εντατικοποίηση θεραπείας· κλινικές δοκιμές
λεμφοπλασματοκυτταρικό ό λ έμφωμα	MYD88 L265P	Βοηθά στη διάγνωση	Αναστολέας της BTK Ibrutinib
πολλαπλούν μυέλωμα	KRAS / NRAS / BRAF / DIS3	Συχνές μεταλλάξεις με προγνωστική σημασία	Αναστολέας του BRAF, νέες κλινικές δοκιμές
οξεία λεμφοβλαστική λευχαιμία/ πολλαπλούν μυέλωμα – μετρήσιμη υπολειπόμενη νόσος	Αναδιατάξεις IGH/TCR	Εκτίμηση ανταπόκρισης	Απόφαση για συνέχιση, αλλαγή ή ενίσχυση θεραπείας
μεταμόσχευση μυελού	Πάνελ πολυμορφισμών	Έγκαιρη ανίχνευση υποτροπής	Πρόληψη/τροποποίηση ανοσοκαταστολής ή δόσης δότη

Αρχικά, το NGS αναπτύχθηκε και χρησιμοποιήθηκε ευρέως για την ανάλυση του DNA, ωστόσο, με την εξέλιξη των βιολογικών ερωτημάτων και την ανάγκη κατανόησης όχι μόνο του στατικού γονιδιώματος αλλά και της δυναμικής του έκφρασης, η ίδια πλατφόρμα αλληλούχησης προσαρμόστηκε ώστε να επιτρέψει και την ανάλυση του RNA μέσω της μεθόδου **RNA-seq (RNA sequencing, αλληλούχηση RNA)**. Η βασική αρχή πίσω από αυτήν την εφαρμογή είναι ότι το RNA αντιπροσωπεύει την «ενεργή» πληροφορία του κυττάρου σε δεδομένη χρονική στιγμή, καθώς αποτυπώνει ποια γονίδια εκφράζονται και σε ποιο βαθμό. Το σύνολο των μεθόδων του RNA-seq περιγράφονται με τον γενικό όρο «**μεταγραφωμική**». Έτσι, ενώ το NGS του DNA προσφέρει τη χαρτογράφηση του γενετικού δυναμικού ενός οργανισμού, το RNA-seq με τη βοήθεια των ίδιων τεχνολογιών δίνει πρόσβαση στη λειτουργική του κατάσταση. Στην πράξη, η διαδικασία ξεκινά με την απομόνωση RNA από ένα βιολογικό δείγμα, στη συνέχεια ακολουθεί η μετατροπή του σε συμπληρωματικό DNA (cDNA) μέσω της δράσης της αντίστροφης μεταγραφάσης, και τελικά το cDNA εισάγεται σε πλατφόρμες NGS. Η μέτρηση του αριθμού των αναγνώσεων που αντιστοιχούν σε κάθε RNA μπορεί να προσφέρει στη σχετική αύξηση ή μείωση της έκφρασής του μεταξύ των συγκρινόμενων δειγμάτων/καταστάσεων. Δηλαδή το RNA-seq αποτελεί μια μεθοδολογία μέτρησης της γονιδιακής έκφρασης σε επίπεδο RNA. Η χρήση του NGS στο RNA-seq δεν περιορίζεται μόνο στη μέτρηση της γονιδιακής έκφρασης, αλλά προσφέρει επίσης τη δυνατότητα ανακάλυψης νέων μεταγραφικών παραλλαγών, όπως η χαρτογράφηση του εναλλακτικού ματίσματος, η μεταγραφή μη κωδικών RNA, καθώς και χιμαιρικών μεταγράφων που σχετίζονται με ασθένειες όπως ο καρκίνος. Η συνεχής εξέλιξη των πλατφορμών NGS ανοίγει τον δρόμο για ακόμα πιο λεπτομερείς και ολοκληρωμένες RNA-seq μελέτες, όπως η ανάλυση **RNA-seq μεμονωμένων κυττάρων (single-cell RNA-seq)**, όπου εξετάζεται το μεταγραφικό προφίλ μεμονωμένων κυττάρων και όχι μόνο ο μέσος όρος ενός ιστού, που τις περισσότερες φορές περιέχει πολλαπλούς κυτταρικούς τύπους.

11.4 Βιοπληροφορικές μέθοδοι ανάλυσης του NGS

Η αλληλούχηση επόμενης γενιάς έχει οδηγήσει σε μια πρωτοφανή πληθώρα δεδομένων, τα οποία καθιστούν αναγκαία τη χρήση εξελιγμένων βιοπληροφορικών μεθόδων για την ανάλυση και την ερμηνεία τους. Τα δεδομένα αυτά, είτε αφορούν γονιδιωματική, είτε μεταγραφωμική, είτε άλλες «-ωμικές» αναλύσεις, δεν μπορούν να αξιοποιηθούν χωρίς την ύπαρξη ενός οργανωμένου υπολογιστικού πλαισίου. Η επεξεργασία των δεδομένων NGS συνήθως διακρίνεται σε τρία διαδοχικά στάδια.

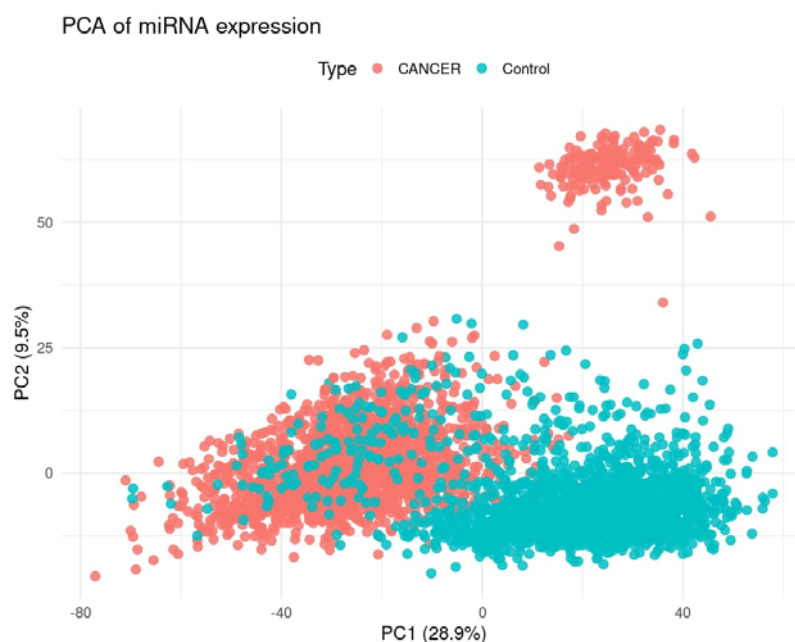
Το πρώτο είναι ο **έλεγχος ποιότητας των ακατέργαστων δεδομένων (raw data)**, δηλαδή των αναγνώσεων που προκύπτουν απευθείας από το μηχάνημα αλληλούχησης. Σε αυτό το στάδιο εντοπίζονται και απομακρύνονται χαμηλής ποιότητας αναγνώσεις ή τεχνικά εσφαλμένες αναγνώσεις ώστε τα εξαγόμενα δεδομένα να είναι αξιόπιστα.

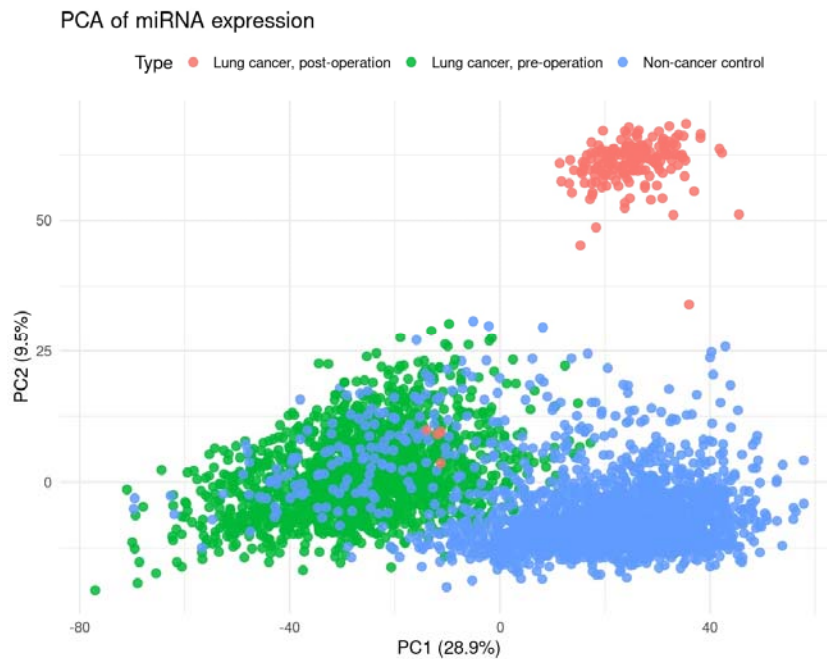
⁴ Η ανάλυση χιμειρισμού μετά από μεταμόσχευση μυελού με τη χρήση πάνελ πολυμορφισμών (SNP panel) μέσω τεχνολογίας NGS επιτρέπει την ακριβή διάκριση και ποσοτικοποίηση των κυττάρων του δότη και του λήπτη. Με αυτόν τον τρόπο παρακολουθείται η επιτυχία της εγκατάστασης του μοσχεύματος και ανιχνεύεται έγκαιρα τυχόν υποτροπή της νόσου ή απώλεια του μοσχεύματος.

Στο δεύτερο στάδιο ακολουθεί η **στοίχιση** ή η **εκ νέου συναρμολόγηση (de novo assembly)**. Στην περίπτωση της στοίχισης, οι αναγνώσεις συγκρίνονται με ένα γνωστό γονιδίωμα αναφοράς (reference genome), ώστε να τοποθετηθούν στη σωστή θέση. Έτσι μπορούν να εντοπιστούν γενετικές αποκλίσεις, όπως η απώλεια χρωμοσωμικών τμημάτων και γονιδιακών περιοχών, ή για να μετρηθεί η έκφραση και η εν γένει παρουσία ενός γονιδίου σε ένα δείγμα. Στην εκ νέου συναρμολόγηση, αντίθετα, οι αναγνώσεις συνδυάζονται μεταξύ τους χωρίς γονιδίωμα αναφοράς, με σκοπό την ανακατασκευή ενός νέου γονιδιώματος. Δηλαδή η μέθοδος λειτουργεί ως ένα ιδιότυπο παζλ αλληλουχιών με σκοπό την παραγωγή του πλήρους γονιδιώματος. Η προσέγγιση αυτή είναι σημαντική όταν μελετώνται νέοι οργανισμοί ή παθογόνα χωρίς γνωστή αλληλουχία.

Το τρίτο στάδιο είναι η **ανίχνευση γενετικών παραλλαγών (variant calling)**, όπου οι στοιχισμένες αλληλουχίες συγκρίνονται με το γονιδίωμα αναφοράς ώστε να εντοπιστούν νουκλεοτιδικόι πολυμορφισμοί (SNPs), μικρές ή μεγάλες απαλοιφές, επαναλήψεις ή δομικές αναδιατάξεις. Τα αποτελέσματα αυτά συνδέονται με βάσεις δεδομένων όπως η ClinVar ή η dbSNP για την αξιολόγηση της πιθανής κλινικής σημασίας τους. Επιπλέον, σε αναλύσεις RNA-seq πραγματοποιείται **ποσοτικοποίηση της γονιδιακής έκφρασης**, ώστε να συγκριθούν τα επίπεδα έκφρασης μεταξύ διαφορετικών δειγμάτων ή συνθηκών.

Πολλά προγράμματα έχουν αναπτυχθεί για κάθε στάδιο της ανάλυσης. Για τον έλεγχο ποιότητας χρησιμοποιούνται εργαλεία όπως το **FastQC**, για τη στοίχιση το **BWA**, το **Bowtie** και το **STAR**, ενώ για την εκ νέου συναρμολόγηση χρησιμοποιούνται τα **Trinity**, **SPAdes** και **Megahit**. Στην ανίχνευση παραλλαγών, δημοφιλή είναι τα **GATK (Genome Analysis Toolkit)** και **FreeBayes**, τα οποία παρέχουν εξειδικευμένες διαδικασίες φιλτραρίσματος και βαθμονόμησης. Για τη μέθοδο RNA-seq, χρησιμοποιούνται προγράμματα όπως το **Cufflinks** και το **DESeq2**. Η πληθώρα αυτών των εργαλείων καθιστά αναγκαία τη δημιουργία ολοκληρωμένων καθετοποιημένων αναλύσεων «**pipeline**», δηλαδή αυτοματοποιημένων ροών εργασίας που ενσωματώνουν όλα τα στάδια με συνεπή και επαναλήψιμο τρόπο. Πλατφόρμες όπως το **Galaxy** παρέχουν ένα φιλικό προς τον χρήστη περιβάλλον, δίνοντας τη δυνατότητα σε μη ειδικούς να εκτελούν περίπλοκες αναλύσεις χωρίς βαθιά γνώση προγραμματισμού. Τέλος σημαντικό κομμάτι της βιοπληροφορικής ανάλυσης αποτελεί η στατιστική επεξεργασία των αποτελεσμάτων για την εξαγωγή συμπερασμάτων.





Εικόνα 11.XX PCA (Principal Component Analysis) βασισμένο σε προφίλ έκφρασης miRNA.

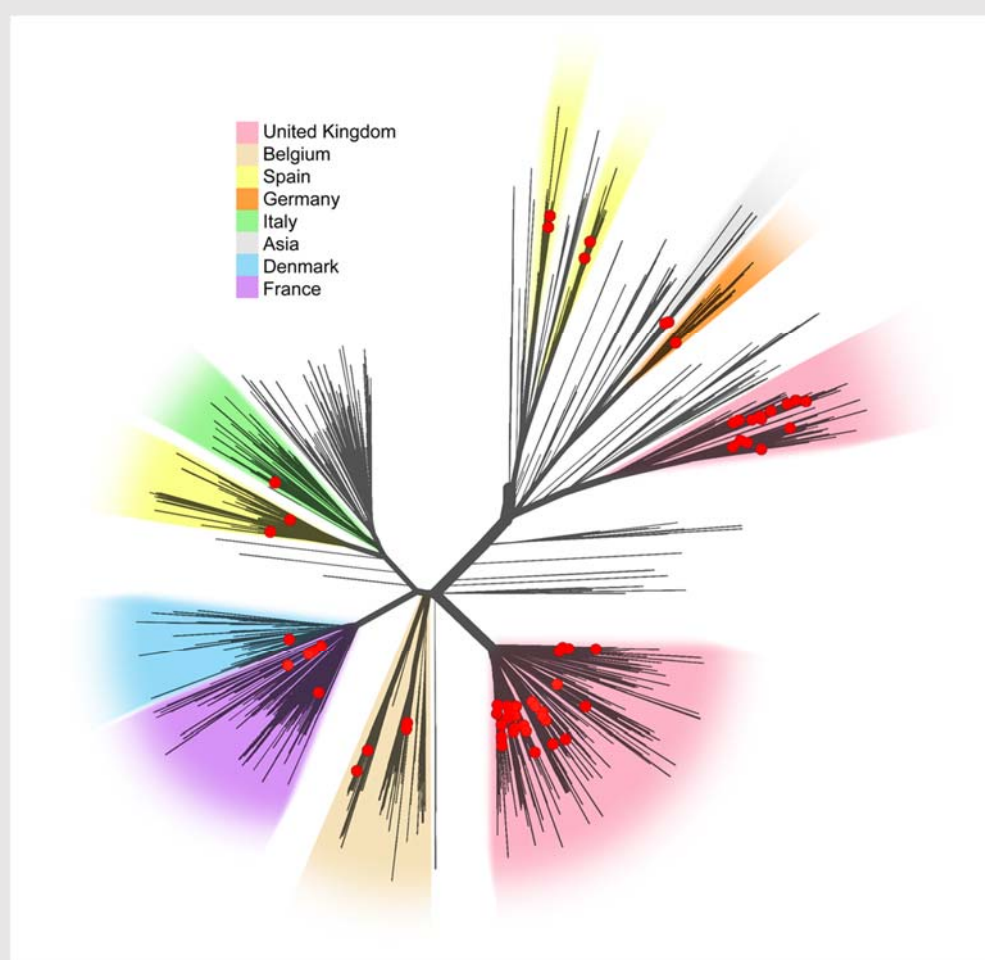
Ένα διάγραμμα PCA τοποθετεί τα δείγματα με βάση την απόκλιση τους σε δύο παραμέτρους/μετρήσεις PC1 και PC2. Οι δύο άξονες αποτελούν μια ποσοτικά σχετική έκφραση των παραμέτρων που διαφοροποιούν τα δείγματα, αντιπροσωπεύοντας ουσιαστικά την διαφορετικότητά τους στα προφίλ έκφρασης των miRNA (ή άλλου μακρομορίου σε άλλες «ωμικές» αναλύσεις). Στο πρώτο διάγραμμα έχουν χρωματιστεί διαφορετικά τα δείγματα που προέρχονται από ασθενείς με καρκίνο (κόκκινο) και υγιείς μάρτυρες (κυανό). Παρατηρήστε ότι τα προφίλ έκφρασης των δειγμάτων από ασθενείς με καρκίνο ομαδοποιούνται διαφορετικά από την ομάδα των υγιών μαρτύρων. Παρατηρήστε επίσης ότι μια υποομάδα ασθενών έχει διαφοροποιηθεί αρκετά από τη βασική ομάδα. Όταν στη μελέτη προστεθούν και άλλα μεταδεδομένα, δηλαδή στην περίπτωση αυτή η αντιστοιχία των δειγμάτων σε προεγχειρητικούς (πράσινο) και μετεγχειρητικούς (κόκκινο) ασθενείς η διαφοροποίηση των υποομάδων μεταξύ τους αλλά και από τους υγιείς μάρτυρες (μπλε) αποκτά διαφορετική ερμηνεία.

ασθενών ξεχωρίζει από ξεχωρίζουν καθαρά από τους υγιείς μάρτυρες, ενώ στη δεύτερη φαίνεται ότι μεταξύ των ασθενών μπορούμε να διακρίνουμε αυτούς που έχουν υποβληθεί σε εκτομή όγκου από εκείνους που δεν έχουν.»

Η αλληλούχηση του πρώτου γονιδιώματος του SARS-CoV-2

Την πρώτη εβδομάδα του Ιανουαρίου 2020, εν μέσω μιας μυστηριώδους επιδημίας πνευμονίας στη Γουχάν, ο Κινέζος ιολόγος Zhang Yongzhen και η ομάδα του στο Κλινικό Κέντρο Δημόσιας Υγείας της Σαγκάης είχαν να αντιμετωπίσουν μια πρόκληση που για πρώτη φορά θα μπορούσε να αντιμετωπιστεί με τις νέες μεθοδολογίες NGS. Στις 3 Ιανουαρίου, έλαβαν δείγματα από ασθενείς που εμφάνιζαν συμπτώματα της νέας ασθένειας. Χρησιμοποιώντας προηγμένες μεθόδους αλληλούχησης RNA-seq, η ομάδα εργάστηκε αδιάκοπα για δύο νύχτες. Περίπου στις 2 π.μ. στις 5 Ιανουαρίου, είχαν ανακατασκευάσει με επιτυχία το πλήρες γονιδίωμα του νέου κορονοϊού - που αργότερα

ονομάστηκε SARS-CoV-2 - το οποίο αποκάλυψε τη γενετική του ομοιότητα με τον ιό SARS. Αναγνωρίζοντας την παγκόσμια σημασία της αλληλουχίας, ο Zhang υπέβαλε το γονιδίωμα σε διεθνείς βάσεις δεδομένων όπως η GenBank και ενημέρωσε τις τοπικές υγειονομικές αρχές για τις πιθανότητες αναπνευστικής εξάπλωσης και την ομοιότητά του με το SARS. Παρόλο που η κινεζική κυβέρνηση είχε περιορισμούς κατά της δημοσίευσης πληροφοριών σχετικά με την επιδημία εκείνη την εποχή, ο Zhang αργότερα έδωσε άδεια στον Edward C. Holmes του Πανεπιστημίου του Σίδνεϋ να δημοσιεύσει τα γονιδιωματικά δεδομένα στο Virological.org στις 11 Ιανουαρίου. Αυτή η έγκαιρη δημοσίευση επέτρεψε στην παγκόσμια επιστημονική κοινότητα να ξεκινήσει την ανάπτυξη διαγνωστικών δοκιμασιών και εμβολίων σχεδόν αμέσως. Η ομάδα του Zhang προχώρησε στην επεξεργασία και την αλληλούχηση δεκάδων χιλιάδων επιπλέον ιικών δειγμάτων, βοηθώντας περαιτέρω την ανταπόκριση της δημόσιας υγείας αλλά και στην παρακολούθηση των στελεχών από το σύνολο της παγκόσμιας επιστημονικής κοινότητας.



Εικόνα 11.6 Η φυλογενετική ανάλυση των γονιδιωμάτων από στελέχη SARS-CoV-2 στην Ελλάδα κατά την πρώτης φάσης της πανδημίας. Ακτινωτό φυλογενετικό δέντρο (radial phylogram) βασισμένο σε χρονολογικά δεδομένα 4668 πλήρων γονιδιωμάτων του SARS-CoV-2 (Nextstrain, 18/05/2020). Τα χρώματα απεικονίζουν τη χώρα προβλεπόμενης προέλευσης των στελεχών, ενώ τα κόκκινα σημεία αντιστοιχούν σε πλήρη γονιδιώματα από στελέχη που εμφανίστηκαν κατά τον πρώτο μήνα της πανδημίας στην Ελλάδα. Η διασπορά των ελληνικών στελεχών σε διαφορετικούς κλάδους υποδηλώνει πολλαπλές ανεξάρτητες εισαγωγές του ιού στην Ελλάδα.

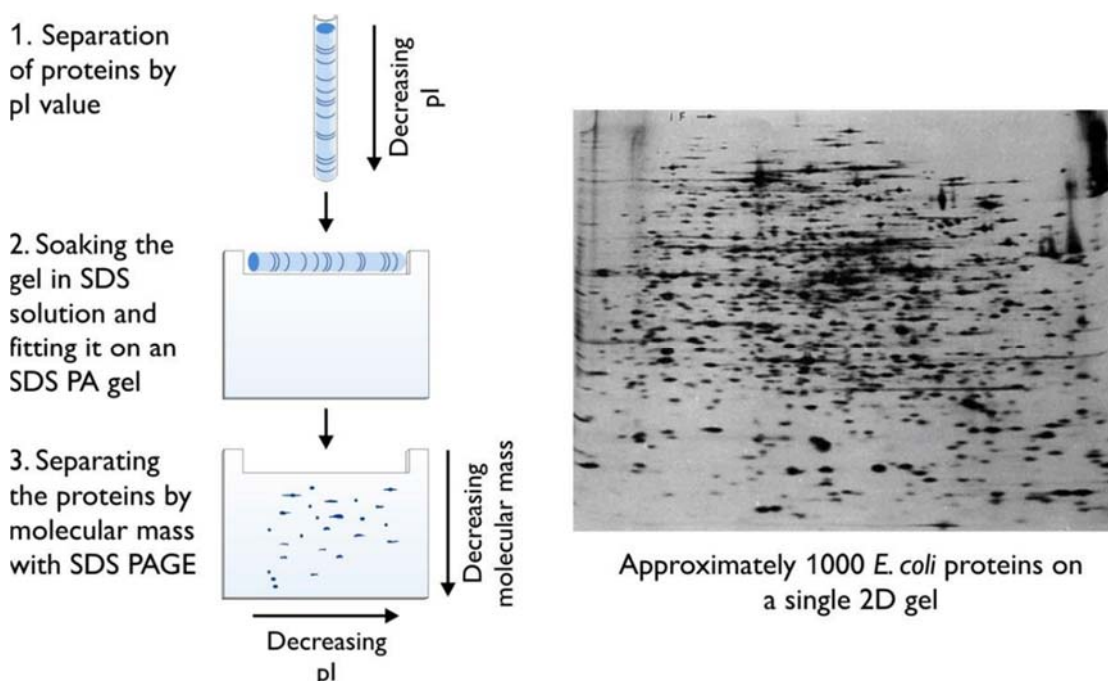
Η **τέταρτη γενιά αλληλούχησης** (Fourth-Generation Sequencing, 4GS) διευρύνει τα όρια της γονιδιωματικής με την εισαγωγή της ***in situ* αλληλούχησης**, δηλαδή της άμεσης ανάλυσης μορίων RNA και DNA μέσα σε κύτταρα και ιστούς. Η καινοτομία αυτή επιτρέπει τη διατήρηση του χωρικού πλαισίου, δίνοντας στους ερευνητές τη δυνατότητα να χαρτογραφήσουν με ακρίβεια τη θέση γενετικών και μεταγραφικών χαρακτηριστικών μέσα σε άθικτα βιολογικά δείγματα. Η προσέγγιση αυτή έχει ιδιαίτερη σημασία για την κατανόηση πολύπλοκων ιστών, όπως ο εγκέφαλος ή το μικροπεριβάλλον των όγκων, όπου οι χωρικές αλληλεπιδράσεις καθορίζουν τη λειτουργία. Τεχνολογίες όπως οι λύσεις χωρικής αλληλούχησης της Illumina και η πλατφόρμα G4X της Singular Genomics υποστηρίζουν αυτοματοποιημένες ροές εργασίας για χαρτογράφηση ολόκληρου του μεταγραφώματος σε ανάλυση κυτταρικού επιπέδου, ακόμα και σε δύσκολα δείγματα, όπως δείγματα εγκλεισμένα σε παραφίνη. Με τις τεχνολογίες 4GS, η γονιδιωματική δεν περιορίζεται πλέον στην απλή καταγραφή αλληλουχιών, αλλά αποκτά μια χωροχρονική διάσταση, καθιστώντας δυνατή την ολοκληρωμένη κατανόηση των βιολογικών συστημάτων στην φυσική τους θέση. Οι εφαρμογές της 4GS στην κλινική έρευνα είναι ήδη εμφανείς. Στη νευρολογία, η *in situ* αλληλούχηση προσφέρει τη δυνατότητα να χαρτογραφηθούν μοριακά προφίλ σε επίπεδο μεμονωμένων νευρώνων, διατηρώντας την τοπογραφία τους μέσα στον εγκέφαλο. Έτσι, μπορούν να αποκαλυφθούν μοτίβα γονιδιακής έκφρασης που συνδέονται με νευροεκφυλιστικές νόσους, όπως η νόσος Alzheimer ή η νόσος Parkinson, και να διερευνηθεί πώς διαφορετικοί πληθυσμοί νευρικών κυττάρων συμβάλλουν στην παθολογία. Στην ογκολογία, η 4GS δίνει τη δυνατότητα μελέτης του μικροπεριβάλλοντος των όγκων με πρωτοφανή λεπτομέρεια. Η ακριβής καταγραφή του πώς τα καρκινικά κύτταρα, τα ανοσοκύτταρα και τα στηρικτικά κύτταρα αλληλεπιδρούν στον χώρο, μέσω ταυτόχρονης ανάλυσης DNA, RNA και πρωτεϊνών, προσφέρει πληροφορίες για τους μηχανισμούς ανοσοδιαφυγής, για πιθανούς νέους θεραπευτικούς στόχους και για την πρόβλεψη της ανταπόκρισης σε ανοσοθεραπεία. Στην παθολογοανατομία, η χωρικά προσδιορισμένη αλληλούχηση μπορεί να μεταμορφώσει τη διάγνωση. Παραδοσιακά, οι ιστολογικές τομές παρέχουν μορφολογικές πληροφορίες, αλλά περιορίζονται στην παρατήρηση της δομής και των χρώσεων. Με την 4GS, οι ίδιες τομές μπορούν να αναλυθούν μοριακά χωρίς να καταστραφεί η δομή τους, προσφέροντας ένα «διπλό» επίπεδο διάγνωσης: μορφολογικό και μοριακό. Αυτό αναμένεται να βελτιώσει την ακρίβεια στην ταξινόμηση όγκων και άλλων παθήσεων.

11.4 Ανάλυση Πρωτεϊνικών Αλληλουχιών

Από το γονίδιο μέχρι την λειτουργική πρωτεΐνη, όπως είδαμε και στο **Κεφάλαιο XX**, μεσολαβούν πολλά στάδια ελέγχου και τροποποιήσεων τόσο του mRNA όσο και των πρωτεϊνών που παράγονται. Η ποικιλία στην λειτουργικότητα δεν απορρέει, λοιπόν, μόνο από τον αριθμό των γονιδίων, αλλά από τον τρόπο με τον οποίο αυτά μεταγράφονται, μεταφράζονται και τροποποιούνται. Η μεταγραφωμική γεφύρωσε τη σχέση γονιδιώματος και γονιδιακής έκφρασης όπως είδαμε στην προηγούμενη παράγραφο. Ωστόσο, ο τελικός τελεστής της λειτουργίας παραμένει να είναι η πρωτεΐνη, για την οποία χρειάστηκε μια επιπλέον δεκαετία ώστε να μπορεί να αναλυθεί ολιστικά αλλά και ποσοτικά με μεθοδολογίες υψηλής απόδοσης. Ο όρος «**πρωτεωμική**» (proteomics) καθιερώθηκε στα μέσα της δεκαετίας του 1990, με αναφορά στο ανάλογο «γονιδιωματική» (genomics). Η πρωτεωμική ασχολείται με τη συστηματική μελέτη του **πρωτεώματος**, δηλαδή του συνόλου των

πρωτεϊνών που περιέχονται σε ένα δείγμα, κύτταρο, ιστό ή οργανισμό υπό συγκεκριμένες συνθήκες. Σε αντίθεση με το γονιδίωμα που αποτελεί μια σχετικά σταθερή οντότητα για κάθε οργανισμό, το πρωτεώμα είναι **δυναμικό και μεταβαλλόμενο**. Η κατανόηση του πρωτεώματος μπορεί να δώσει μια πιο άμεση εικόνα της φυσιολογικής ή παθολογικής κατάστασης ενός οργανισμού σε σχέση με τη γονιδιωματική, ενώ προσεγγίζεται μόνο κατά ένα ποσοστό από τη μεταγραφωμική.

Η μελέτη των πρωτεϊνών έχει μακρά ιστορία. Από τις αρχές του 20ού αιώνα, οι βιοχημικοί ανέπτυξαν μεθόδους απομόνωσης και χαρακτηρισμού πρωτεϊνών με βάση τις φυσικοχημικές τους ιδιότητες, όπως η ηλεκτροφόρηση και η χρωματογραφία. Η πραγματική όμως «γέννηση» της πρωτεωμικής ως ξεχωριστού πεδίου συνδέεται με την ανάπτυξη της **δισδιάστατης ηλεκτροφόρησης (2D-GE)** στη δεκαετία του 1970, η οποία επέτρεψε τον διαχωρισμό εκατοντάδων πρωτεϊνών σε ένα μόνο πήκτωμα. Παρά την καινοτομία, οι περιορισμοί της μεθόδου (χαμηλή ευαισθησία, δυσκολία ανάλυσης μεμβρανικών πρωτεϊνών, περιορισμένη ταυτοποίηση) καθιστούσαν την προσέγγιση ανεπαρκή για την πλήρη χαρτογράφηση του πρωτεώματος. Η καθοριστική πρόοδος προήλθε από την ανάπτυξη της **φασματομετρίας μάζας (MS)** ως εργαλείου για τη βιολογική ανάλυση. Αν και η MS είχε ήδη χρησιμοποιηθεί στη χημεία, η εφαρμογή της στη βιολογία έγινε δυνατή τη δεκαετία του 1980 με την εισαγωγή των ήπιων τεχνικών ιοντισμού ESI (Electrospray Ionization) και MALDI (Matrix-Assisted Laser Desorption/Ionization). Οι τεχνολογίες αυτές επέτρεψαν τον ιονισμό και μέτρηση μεγάλων βιομορίων χωρίς καταστροφή τους, εγκαινιάζοντας μια νέα εποχή στη μαζική ανάλυση πρωτεϊνών και πεπτιδίων. Η συμβολή των μεθόδων αυτών στην πρόοδο της επιστήμης αναγνωρίστηκε διεθνώς με το Βραβείο Νόμπελ Χημείας το 2002 στους John Fenn και Koichi Tanaka. Από εκείνη τη στιγμή, η φασματομετρία μάζας εδραιώθηκε ως η κύρια τεχνολογική πλατφόρμα της πρωτεωμικής, ξεπερνώντας κατά πολύ τις κλασικές βιοχημικές μεθόδους.



Εικόνα 11.7 Αρχή της δισδιάστατης ηλεκτροφόρησης πρωτεϊνών (2D-GE). Στο πρώτο βήμα οι πρωτεΐνες διαχωρίζονται με βάση το ισοηλεκτρικό τους σημείο (pI), δηλαδή στο pH εκείνο

που η πρωτεΐνη έχει μηδενικό φορτίο, ενώ στη συνέχεια εμποτίζονται με SDS και τοποθετούνται σε πήκτωμα πολυακρυλαμίδης. Στο δεύτερο βήμα πραγματοποιείται διαχωρισμός με βάση το μοριακό βάρος μέσω SDS-PAGE. Δεξιά, παράδειγμα εφαρμογής όπου περισσότερες από 1000 πρωτεΐνες του *E. coli* διαχωρίζονται σε ένα μόνο 2D πήκτωμα.

Η πρωτεωμική δεν θα μπορούσε να εξελιχθεί στον βαθμό που βρίσκεται σήμερα χωρίς την πρόοδο της φασματομετρίας μάζας (Mass Spectrometry, MS). Η φασματομετρία μάζας αποτελεί την τεχνολογική «καρδιά» της πρωτεωμικής, καθώς σε συνδυασμό με την υγρή χρωματογραφία υψηλής απόδοσης προσφέρει τη δυνατότητα ποιοτικής και ποσοτικής ανάλυσης χιλιάδων πρωτεϊνών από πολύπλοκα βιολογικά δείγματα, με ακρίβεια και υψηλή ευαισθησία. Ο συνδυασμός της **υγρής χρωματογραφίας (LC)** και δύο διαδοχικών MS ονομάζεται **LC-MS/MS**. Η διαδικασία της LC-MS/MS περιλαμβάνει μια αλληλουχία βημάτων. Αρχικά, οι πρωτεΐνες υφίστανται πρωτεολυτική πέψη, συνήθως με το ένζυμο τρυψίνη, που διασπά τις πολυπεπτιδικές αλυσίδες σε μικρότερα πεπτίδια με προβλέψιμα σημεία θραύσης. Η **τρυψίνη** είναι ένα πρωτεολυτικό ένζυμο (πρωτεάση σερίνης) που χρησιμοποιείται σχεδόν σε όλα τα πειράματα πρωτεωμικής, διότι παρουσιάζει υψηλή ειδικότητα στον κατακερματισμό των πρωτεϊνών. Το χαρακτηριστικό της είναι ότι διασπά τις πολυπεπτιδικές αλυσίδες μετά από τα βασικά αμινοξέα λυσίνη (K) και αργινίνη (R), εκτός εάν ακολουθείται από προλίνη (P). Αυτό δημιουργεί μια σειρά από πεπτίδια με σχετικά ομοιόμορφο μήκος (συνήθως 6–20 αμινοξέα), κατάλληλο για ανάλυση με φασματομετρία μάζας. Η μεγάλη σημασία αυτού του προβλέψιμου μοτίβου είναι ότι καθιστά την τρυψίνη ιδανική για βιοπληροφορική ανάλυση. Από τη στιγμή που γνωρίζουμε την αλληλουχία DNA ενός γονιδίου, μπορούμε να προβλέψουμε την αντίστοιχη πρωτεΐνη και, μέσω υπολογιστικών εργαλείων, να «προσομοιώσουμε» την πέψη με τρυψίνη. Έτσι, μπορεί να δημιουργηθεί ένας *in silico* κατάλογος πεπτιδίων, ο οποίος στη συνέχεια συγκρίνεται με τα πειραματικά δεδομένα φασματομετρίας μάζας. Αν τα φάσματα που προκύπτουν από το δείγμα ταιριάζουν με τα θεωρητικά πεπτίδια της βάσης δεδομένων, η πρωτεΐνη ταυτοποιείται με μεγάλη ακρίβεια. Τα πεπτίδια που προκύπτουν από την πέψη διαχωρίζονται με υγρή χρωματογραφία με βάση την υδροφοβικότητά τους, ώστε να μειωθεί η πολυπλοκότητα και να εισαχθούν διαδοχικά στον φασματογράφο μάζας.

Στον φασματογράφο μάζας τα πεπτίδια μπαίνουν ιονιζόμενα με τις τεχνικές ESI ή MALDI και διαχωρίζονται με τη βοήθεια ηλεκτρικών ή/και μαγνητικών πεδίων στον αναλυτή της συσκευής. Εκεί μετράται ο **λόγος μάζας προς φορτίο (m/z)** των ιόντων όταν προσπίπτουν στον αισθητήρα, δημιουργώντας μια χαρακτηριστική καταγραφή στο φάσμα για το κάθε πεπτίδιο. Κατά το δεύτερο MS, τα πεπτίδια που επιλέγονται με βάση το πόσο έντονη είναι η παρουσία τους, δηλαδή τη σχετική τους αφθονία στο σύνολο των πεπτιδίων, οδηγούνται σε ελεγχόμενη διάσπαση υψηλότερης ενέργειας⁵, με αποτέλεσμα να προκύψει ένα φάσμα από ιοντικά θραύσματα. Το φάσμα αυτό είναι χαρακτηριστικό για το κάθε πεπτίδιο και ενδεικτικό της σειράς με την οποία συνδέονται τα αμινοξέα. Το δεύτερο αυτό MS είναι εκείνο που επιτρέπει την απόδοση αλληλουχίας στα πεπτίδια. Η αλληλουχία των πεπτιδίων μπορεί να οδηγήσει στην αντιστοίχισή τους σε μία πρωτεΐνη του οργανισμού, το δε ύψος της καταγραφής στο φάσμα μπορεί να δώσει ποσοτικό χαρακτήρα στη μέτρηση⁶. Το μεγάλο της πλεονέκτημα της τεχνολογίας LC-MS/MS είναι η δυνατότητα ταυτόχρονης ανάλυσης

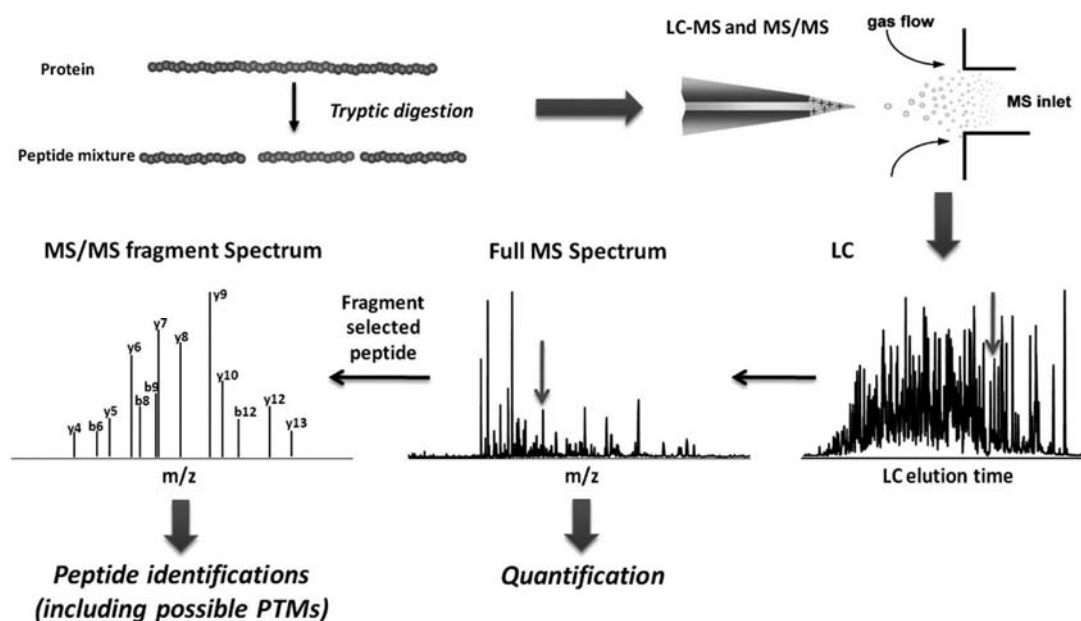
⁵ Με τις τεχνικές Collision-Induced Dissociation (CID) και Higher-Energy Collisional Dissociation (HCD).

⁶ Καθώς τα πεπτίδια κατανέμονται στο σημείο της καταγραφής η πραγματική ποσοτική μέτρηση αφορά στο εμβαδόν της καταγεγραμμένης κορυφής στο φάσμα.

χιλιάδων πρωτεϊνών σε πολύπλοκα δείγματα, όπως ιστοί, καλλιέργειες κυττάρων ή βιολογικά υγρά, ακόμα και με σχετικά μικρή ποσότητα αρχικού υλικού.

Η πιο άμεση εφαρμογή της πρωτεωμικής είναι η ανακάλυψη βιοδεικτών. Η χρήση τεχνικών LC-MS/MS επιτρέπει τη σύγκριση των πρωτεϊνικών προφίλ υγιών και παθολογικών ιστών, καθώς και τη μελέτη βιολογικών υγρών (ορός, ούρα, εγκεφαλονωτιαίο υγρό, κ.α.). Στον καρκίνο, για παράδειγμα, οι βιοδείκτες διακρίνονται σε τρεις κατηγορίες: διαγνωστικοί, που ανιχνεύουν την παρουσία νόσου, προγνωστικοί, που προβλέπουν την πιθανή πορεία της νόσου, και προβλεπτικοί. Ένα χαρακτηριστικό παράδειγμα είναι η υπερέκφραση της πρωτεΐνης HER2 στον καρκίνο μαστού, η οποία είναι προγνωστικός δείκτης για την ανταπόκριση στη θεραπεία με trastuzumab.

Μία από τις πιο εντυπωσιακές και άμεσες κλινικές εφαρμογές της πρωτεωμικής είναι η χρήση της **MALDI-TOF**⁷ φασματομετρίας μάζας στην ταχεία αναγνώριση μικροοργανισμών. Η μέθοδος βασίζεται στη δημιουργία χαρακτηριστικών πρωτεϊνικών «δακτυλικών αποτυπωμάτων» που προέρχονται κυρίως από ριβοσωμικές πρωτεΐνες βακτηρίων και μυκήτων. Με τη σύγκριση των φασμάτων με βάσεις δεδομένων, οι κλινικοί μικροβιολόγοι μπορούν να αναγνωρίσουν μικροοργανισμούς σε λίγα λεπτά, σε αντίθεση με τις παραδοσιακές μεθόδους καλλιέργειας που απαιτούν ημέρες. Η MALDI-TOF έχει πλέον ενσωματωθεί σε πολλά μικροβιολογικά εργαστήρια παγκοσμίως, αλλάζοντας τον τρόπο με τον οποίο γίνεται η διάγνωση λοιμώξεων. Εκτός από την αναγνώριση είδους, η τεχνική χρησιμοποιείται και για τον εντοπισμό αντοχών, π.χ. στην ανίχνευση παραγωγής καρβαπενεμασών σε Gram-αρνητικά βακτήρια.



Εικόνα 11.8. Σχηματική αναπαράσταση της ροής ανάλυσης πρωτεϊνών με LC-MS/MS. Οι πρωτεΐνες υφίστανται πέψη με τρυψίνη σε πεπτίδια, τα οποία διαχωρίζονται με υγρή χρωματογραφία (LC) και εισάγονται στον φασματογράφο μάζας μέσω ιοντισμού. Στο πρώτο

⁷ Η MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization – Time of Flight) είναι τεχνική φασματομετρίας μάζας που συνδυάζει ήπιο ιοντισμό με λέιζερ και καταγραφή με βάση τον χρόνο πτήσης των ιόντων. ο χρόνος πτήσης είναι ο χρόνος που μεσολαβεί από τον ιονισμό και την επιτάχυνση μέχρι την καταγραφή τους στον ανιχνευτή.

στάδιο (Full MS Spectrum) καταγράφεται το συνολικό φάσμα μάζας/φορτίου (m/z) όλων των πεπτιδίων, ενώ στο δεύτερο στάδιο (MS/MS Spectrum) επιλέγονται συγκεκριμένα πεπτίδια που διασπώνται περαιτέρω για τον προσδιορισμό της αλληλουχίας τους. Η ολοκληρωμένη ανάλυση οδηγεί τόσο σε **ταυτοποίηση** όσο και σε **ποσοτικοποίηση** των πρωτεϊνών.

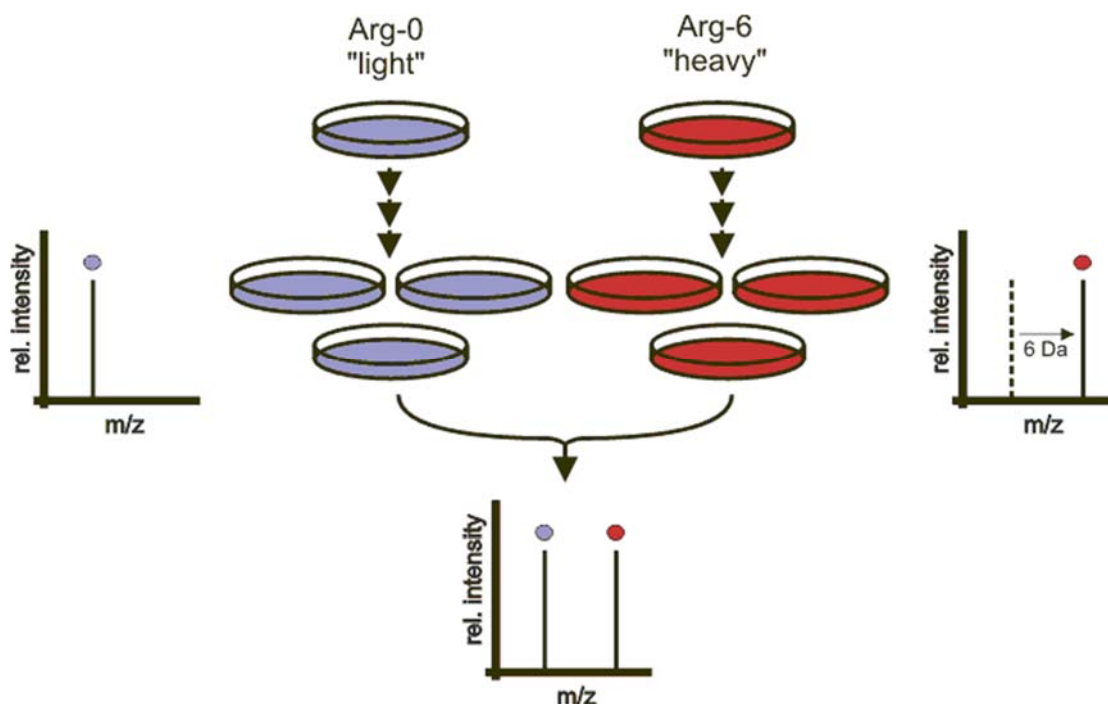
11.6 Βιοπληροφορικές μέθοδοι ανάλυσης της πρωτεωμικής

Η επεξεργασία των δεδομένων φασματομετρίας μάζας (MS data processing) αποτελεί το κρίσιμο πρώτο βήμα σε κάθε ανάλυση πρωτεωμικής, καθώς από την ποιότητα αυτής της διαδικασίας εξαρτάται η αξιοπιστία όλων των επόμενων σταδίων, όπως η ταυτοποίηση, η ποσοτικοποίηση και η ανίχνευση μετα-μεταφραστικών τροποποιήσεων. Τα αρχικά δεδομένα που παράγονται από ένα φασματομέτρο μάζας είναι εξαιρετικά περίπλοκα και συνίστανται σε χιλιάδες έως εκατομμύρια ατομικές μετρήσεις μάζας/φορτίου (m/z), που αντιπροσωπεύουν τα ιόντα των πεπτιδίων ή των πρωτεϊνικών θραυσμάτων. Το πρώτο βήμα είναι η εξομάλυνση του σήματος και η απομάκρυνση του θορύβου. Μέσα από στατιστικές και μαθηματικές τεχνικές φιλτραρίσματος, το σήμα καθαρίζεται ώστε να διατηρηθούν οι κορυφές που αντιστοιχούν σε πραγματικά ιόντα και να εξαλειφθούν τυχαίες διακυμάνσεις. Στη συνέχεια, εφαρμόζεται η διαδικασία της **αναγνώρισης κορυφών**, η οποία αφορά τον ακριβή εντοπισμό των κορυφών που αντιπροσωπεύουν πεπτίδια ή πρωτεΐνες. Η σωστή αναγνώριση των κορυφών είναι ζωτικής σημασίας, διότι οποιαδήποτε απώλεια ή παρανόηση μπορεί να οδηγήσει σε ελλιπή ή λανθασμένη ταυτοποίηση. Για τον σκοπό αυτό χρησιμοποιούνται εξειδικευμένα λογισμικά, όπως το **MaxQuant** ή το **OpenMS**, τα οποία μπορούν να διακρίνουν ανάμεσα σε αληθινά και τεχνητά σήματα. Ένα ακόμη απαραίτητο στάδιο είναι η ευθυγράμμιση των δεδομένων μεταξύ διαφορετικών δειγμάτων ή πειραματικών επαναλήψεων. Το αποτέλεσμα όλων αυτών των βημάτων είναι ένα σύνολο δεδομένων το οποίο μπορεί πλέον να αξιοποιηθεί για πιο σύνθετες αναλύσεις ταυτοποίησης και ποσοτικοποίησης.

Η πρωτεϊνική ταυτοποίηση αποτελεί τον πυρήνα κάθε πρωτεωμικής ανάλυσης, καθώς είναι το βήμα που μετατρέπει τα φάσματα φασματομετρίας μάζας σε βιολογικά χρήσιμη πληροφορία, δηλαδή στη γνώση για το ποια πρωτεΐνη βρίσκεται σε ένα δείγμα. Η διαδικασία βασίζεται στη λεπτομερή ανάλυση των φασμάτων MS/MS, στα οποία επιλεγμένα φάσματα πεπτιδίων λειτουργούν σαν «δακτυλικό αποτύπωμα» που αποκαλύπτει την αλληλουχία του πεπτιδίου και, κατ' επέκταση, την πρωτεΐνη από την οποία προήλθε. Η πλέον διαδεδομένη στρατηγική για την ταυτοποίηση της πρωτεΐνης προέλευσης του πεπτιδίου είναι η αναζήτηση σε βάσεις δεδομένων. Σε αυτήν την προσέγγιση, οι αλληλουχίες πρωτεϊνών από γνωστά γονιδιώματα υποβάλλονται σε υπολογιστική πέψη (*in silico digestion*). Έτσι παράγεται ένας θεωρητικός κατάλογος πεπτιδίων και των αντίστοιχων φασμάτων θραυσμάτων. Τα πειραματικά φάσματα MS/MS συγκρίνονται με τα θεωρητικά και υπολογίζεται ένα σκορ ομοιότητας, το οποίο αντικατοπτρίζει την πιθανότητα σωστής αντιστοίχισης. Λογισμικά όπως το **Mascot**, το **Sequest**, το **Andromeda** (που ενσωματώνεται στο MaxQuant) και το **X!Tandem** έχουν καθιερωθεί διεθνώς για τέτοιες αναζητήσεις. Η προσέγγιση αυτή είναι εξαιρετικά αποτελεσματική όταν υπάρχουν πλήρη και αξιόπιστα γονιδιωματικά δεδομένα, καθώς επιτρέπει την ταυτοποίηση χιλιάδων πρωτεϊνών σε σύνθετα δείγματα με σχετική ακρίβεια. Ωστόσο, η μέθοδος αυτή έχει περιορισμούς όταν μελετώνται οργανισμοί με ατελώς καταγεγραμμένα γονιδιώματα ή όταν υπάρχουν πολλές άγνωστες ισομορφές. Σε αυτές τις περιπτώσεις χρησιμοποιείται μια δεύτερη στρατηγική, γνωστή ως **εκ νέου αποκωδικοποίηση (de novo αποκωδικοποίηση)**. Εδώ ειδικοί αλγόριθμοι αναλύουν το

μοτίβο των θραυσμάτων και συνθέτουν απευθείας την πιθανή αλληλουχία του πεπτιδίου. Αν και η μέθοδος είναι υπολογιστικά πιο απαιτητική και συνήθως λιγότερο ακριβής, προσφέρει ανεκτίμητη αξία για την ανακάλυψη νέων πρωτεϊνικών μορφών, αλλά και πρωτεϊνών λιγότερο μελετημένων οργανισμών.

Η μεγάλη αξία της πρωτεωμικής στην ανάλυση βιολογικών συστημάτων ήρθε από τη δυνατότητα εκτίμησης της σχετικής ή και απόλυτης αφθονίας τους μέσα σε ένα βιολογικό δείγμα. Η ακριβής μέτρηση των επιπέδων έκφρασης πρωτεϊνών είναι ζωτικής σημασίας, διότι οι αλλαγές σε αυτά αντανakλούν τη δυναμική προσαρμογή των κυττάρων σε φυσιολογικές ή παθολογικές συνθήκες. Με άλλα λόγια, ενώ η ταυτοποίηση μάς απαντά στο ερώτημα «ποιες πρωτεΐνες υπάρχουν», η ποσοτικοποίηση μας δείχνει «πόσο εκφράζονται» και «πώς αλλάζουν» σε διαφορετικές καταστάσεις. Στον χώρο της ποσοτικής πρωτεωμικής έχουν αναπτυχθεί δύο βασικές προσεγγίσεις: η **χωρίς σήμανση** (label-free quantification) και η με **ισοτοπική σήμανση** (isotopic labeling). Και οι δύο στηρίζονται σε δεδομένα φασματομετρίας μάζας, αλλά διαφοροποιούνται ως προς τον τρόπο με τον οποίο εκμεταλλεύονται τα φασματικά σήματα. Στην πρώτη περίπτωση, η ποσοτικοποίηση γίνεται άμεσα από την ένταση του σήματος που αποδίδεται σε κάθε ένα πεπτίδιο, ενώ στη δεύτερη, η εισαγωγή διαφορετικών ισοτόπων των χημικών στοιχείων επιτρέπει την ακριβέστερη σύγκριση πολλαπλών δειγμάτων εντός της ίδιας εφαρμογής, εξαλείφοντας έτσι την απαίτηση για ευθυγράμμιση των φασμάτων. Τα λογισμικά εργαλεία που υποστηρίζουν label-free αναλύσεις, όπως το **MaxQuant**, το **Skyline** και το **Progenesis Q1**, παρέχουν εξελιγμένους αλγόριθμους για την αναγνώριση κορυφών, την κανονικοποίηση των εντάσεων και τη στατιστική ανάλυση των διαφορών.



Εικόνα 11.9. Αρχή της μεθόδου SILAC (Stable Isotope Labeling by Amino acids in Cell culture). Στο SILAC μπορούμε να έχουμε διάφορα επίπεδα σήμανσης με άνθρακα 13 στους 6 άνθρακες μίας αργινίνης ή μίας λυσίνης. Έτσι έχουμε 2 καταστάσεις την ελαφριά με τα φυσιολογικά αμινοξέα αργινίνης και λυσίνης (Arg-0, Lys-0) και τη βαριά με άνθρακα 13 ($^{13}\text{C}_6\text{-Arg}$ και $^{13}\text{C}_6\text{-Lys}$). Στην εικόνα τα κύτταρα καλλιεργούνται σε «ελαφρύ» θρεπτικό μέσο με φυσιολογική αργινίνη (Arg-0, μπλε) ή σε «βαρύ» μέσο με ισοτοπικά σημασμένη

αργινίνη (Arg-6, κόκκινο). Η μεταβολική ενσωμάτωση των αμινοξέων στις πρωτεΐνες προκαλεί μετατόπιση μάζας στα αντίστοιχα πεπτίδια, η οποία ανιχνεύεται με φασματομετρία μάζας ως διαφορά 6 Dalton (μετατόπιση της κορυφής στο φάσμα). Όταν τα δύο δείγματα αναμιχθούν και αναλυθούν μαζί, η αναλογία των κορυφών στο φάσμα (μπλε vs κόκκινου) αντανακλά τη σχετική αφθονία της πρωτεΐνης στα δύο δείγματα.

Πώς η πρωτεωμική άλλαξε τη διάγνωση στα κλινικά μικροβιολογικά εργαστήρια

Η εισαγωγή του MALDI-TOF MS στη μικροβιολογική διάγνωση θεωρείται μία από τις σημαντικότερες επαναστάσεις των τελευταίων δεκαετιών. Σε αντίθεση με τις κλασικές μεθόδους που βασίζονται σε καλλιέργειες και βιοχημικά τεστ — διαδικασίες που συχνά απαιτούν 24 έως 48 ώρες για την ταυτοποίηση ενός μικροοργανισμού — η τεχνολογία αυτή προσφέρει τη δυνατότητα διάγνωσης μέσα σε λίγα λεπτά. Η βασική αρχή στηρίζεται στην ανάλυση του «αποτυπώματος πρωτεϊνών» κάθε μικροοργανισμού. Οι πρωτεΐνες, κυρίως ριβοσωμικές, ιονίζονται και αναλύονται βάσει της σχέσης μάζας/φορτίου. Το φάσμα που προκύπτει συγκρίνεται με βάση δεδομένων και οδηγεί σε ασφαλή ταυτοποίηση σε επίπεδο γένους και, συχνά, είδους. Η επίδραση στη ρουτίνα των μικροβιολογικών εργαστηρίων ήταν τεράστια. Ο χρόνος διάγνωσης μειώθηκε κατά περίπου 24 ώρες, επιτρέποντας πιο γρήγορη λήψη θεραπευτικών αποφάσεων, ιδιαίτερα σε περιπτώσεις βαριών λοιμώξεων όπως η σηψαιμία. Παράλληλα, το MALDI-TOF MS αντικατέστησε σε μεγάλο βαθμό τις παλαιότερες μεθόδους αναγνώρισης, προσφέροντας καλύτερη ακρίβεια και μικρότερο κόστος αναλωσίμων. Έτσι, η τεχνολογία δεν αποτελεί απλώς ένα εργαλείο, αλλά μια πραγματική αλλαγή παραδείγματος στη μικροβιολογική διάγνωση.

Η χρήση του MALDI-TOF MS επεκτείνεται σε πολλαπλά επίπεδα της διαγνωστικής μικροβιολογίας, μετατρέποντάς το σε εργαλείο υψηλής αξίας. Η πιο διαδεδομένη εφαρμογή αφορά την ταυτοποίηση μικροοργανισμών από καθαρές καλλιέργειες. Με ένα απλό βήμα προετοιμασίας, ακόμη και μικρές ποσότητες αποικιών μπορούν να οδηγήσουν σε αξιόπιστη ταυτοποίηση. Αυτό είναι ιδιαίτερα σημαντικό σε μικροοργανισμούς που αναπτύσσονται αργά ή απαιτούν πολύπλοκα μέσα καλλιέργειας. Η χρήση του MALDI-TOF MS έχει αποδειχθεί πιο αξιόπιστη από βιοχημικές μεθόδους για ορισμένα είδη, μειώνοντας σημαντικά τον χρόνο αναμονής για τον κλινικό ιατρό. Μία από τις πιο κρίσιμες εφαρμογές είναι η ταχεία ανάλυση θετικών αιμοκαλλιεργείων. Παραδοσιακά, η διαδικασία απαιτεί καλλιέργεια και περαιτέρω βιοχημική ταυτοποίηση, κάτι που μπορεί να καθυστερήσει τη θεραπεία έως και δύο ημέρες. Με τη χρήση MALDI-TOF MS και ειδικών κιτ (όπως το Sepsityper®), η ταυτοποίηση γίνεται μέσα σε 15–20 λεπτά μετά το σήμα θετικότητας. Αυτό επιτρέπει άμεση στοχευμένη θεραπεία, βελτιώνοντας την πρόγνωση και μειώνοντας την αλόγιστη χρήση ευρέως φάσματος αντιβιοτικών. Πέρα από την ταυτοποίηση, το MALDI-TOF MS μπορεί να ανιχνεύσει βιοδείκτες που σχετίζονται με μηχανισμούς αντοχής. Για παράδειγμα, η παρουσία συγκεκριμένων ενζύμων (όπως β-λακταμασών) μπορεί να ανιχνευθεί με ανάλυση των προϊόντων αποδόμησης αντιβιοτικών. Αυτό προσφέρει κρίσιμη πληροφορία 24 ώρες νωρίτερα από το κλασικό αντιβιογράφημα, συμβάλλοντας σε πιο ορθολογική χρήση των αντιμικροβιακών φαρμάκων. Παρόλο που δεν έχει ακόμη ενσωματωθεί στη ρουτίνα, μελέτες έχουν δείξει ότι το MALDI-TOF MS μπορεί να χρησιμοποιηθεί για την ταυτοποίηση παρασίτων όπως πρωτόζωα (π.χ. *Plasmodium falciparum*), εντερικά πρωτόζωα ή εκτοπαράσιτα. Παράλληλα, έχουν γίνει

προσπάθειες για εφαρμογές σε ιούς, αν και απαιτούνται περαιτέρω έρευνες πριν μετατραπεί σε κλινικό εργαλείο.

ε. Δείγματα ούρων και ENY

Εξετάζεται η δυνατότητα ταυτοποίησης παθογόνων απευθείας από βιολογικά υγρά, όπως ούρα ή εγκεφαλονωτιαίο υγρό. Τα πρώτα αποτελέσματα είναι ενθαρρυντικά, όμως χρειάζεται περαιτέρω βελτίωση ώστε να ξεπεραστούν εμπόδια όπως η χαμηλή συγκέντρωση μικροοργανισμών ή οι παρεμβολές από πρωτεΐνες του ξενιστή.

Το MALDI-TOF MS διαθέτει σημαντικά πλεονεκτήματα που εξηγούν την εκρηκτική του διάδοση. Πρώτα απ' όλα, η ταχύτητα: η διαδικασία ταυτοποίησης μπορεί να ολοκληρωθεί σε λιγότερο από 1 λεπτό ανά δείγμα, ενώ με πλάκες 96 θέσεων μπορούν να αναλυθούν δεκάδες δείγματα σε μία ώρα. Δεύτερον, το χαμηλό κόστος αναλωσίμων: απαιτούνται ελάχιστες ποσότητες χημικών, ενώ η συσκευή μπορεί να επαναχρησιμοποιεί πλάκες μετά από καθαρισμό. Αυτό μειώνει δραστικά το λειτουργικό κόστος σε σχέση με βιοχημικά κιτ. Η ακρίβεια αποτελεί άλλο ένα βασικό πλεονέκτημα. Το φάσμα πρωτεϊνών λειτουργεί ως μοναδικό «δακτυλικό αποτύπωμα» για κάθε είδος, οδηγώντας σε σωστή ταυτοποίηση στο μεγαλύτερο ποσοστό των περιπτώσεων. Η μέθοδος είναι επίσης υψηλής ευαισθησίας, απαιτώντας μόλις 10^2 – 10^4 κύτταρα. Παράλληλα, η διαδικασία είναι απλή και δεν χρειάζεται πολύπλοκη ανάλυση από τον χειριστή, καθώς το λογισμικό παρέχει αυτόματα σκορ ταυτοποίησης. Τέλος, το MALDI-TOF MS δίνει τη δυνατότητα ευελιξίας. Οι βάσεις δεδομένων μπορούν να εμπλουτιστούν με νέα είδη, ενώ νέες εφαρμογές (ανίχνευση αντοχής, τυποποίηση στελεχών) μπορούν να προστεθούν στην υπάρχουσα υποδομή, επεκτείνοντας συνεχώς τις δυνατότητές του.

Παρά τα πολλά πλεονεκτήματα, η τεχνολογία δεν είναι χωρίς περιορισμούς. Ένας από τους βασικότερους είναι η ανάγκη για επαρκή ποσότητα. Σε μικροοργανισμούς που αναπτύσσονται αργά, απαιτείται μικρής κλίμακας καλλιέργεια για να συγκεντρωθεί αρκετό υλικό, γεγονός που καθυστερεί τη διαδικασία. Ένα άλλο πρόβλημα είναι η απαιτητική προετοιμασία σε ορισμένες κατηγορίες μικροοργανισμών. Μύκητες, μυκοβακτήρια και βακτήρια με σκληρά κυτταρικά τοιχώματα χρειάζονται ειδικά βήματα εκχύλισης πρωτεϊνών, κάτι που αυξάνει τον χρόνο και το κόστος. Επιπλέον, το MALDI-TOF MS δυσκολεύεται να διαχωρίσει πολύ συγγενικά είδη. Κλασικά παραδείγματα είναι το *Escherichia coli* και το *Shigella*, ή τα είδη του συμπλέγματος *Streptococcus mitis/oralis* που συγχέονται με τον *Streptococcus pneumoniae*. Σε αυτές τις περιπτώσεις, απαιτούνται συμπληρωματικές μοριακές μέθοδοι. Αξιοσημείωτος περιορισμός είναι και η αδυναμία αναγνώρισης μικτών καλλιεργειών. Εάν στο δείγμα υπάρχουν περισσότερα του ενός είδη, το φάσμα που προκύπτει δεν μπορεί να διαχωριστεί με ακρίβεια.

Τα τελευταία χρόνια σημειώνεται πρόοδος στην αυτοματοποίηση της διαδικασίας. Συστήματα όπως το Colibri™ αναλαμβάνουν αυτόματα τη μεταφορά αποικιών από καλλιέργειες και την προετοιμασία για ανάλυση. Αυτό μειώνει περιορίζει τα ανθρώπινα λάθη και εξασφαλίζει ομοιομορφία στα αποτελέσματα. Παράλληλα, έχουν κυκλοφορήσει αναβαθμισμένα εμπορικά συστήματα όπως το Bruker Biotyper Sirius και το VITEK MS

Prime, τα οποία προσφέρουν ευρύτερες βάσεις δεδομένων και δυνατότητα ανάλυσης περισσότερων ειδών. Οι μελέτες δείχνουν ότι το ποσοστό επιτυχούς ταυτοποίησης φτάνει το 97–99% σε ρουτίνα. Επιπλέον, η βελτίωση του λογισμικού επιτρέπει ταχύτερη σύγκριση φασμάτων και πιο εύκολη ενσωμάτωση νέων ειδών. Η τάση προς τον πλήρη αυτοματισμό είναι εμφανής: το MALDI-TOF MS εντάσσεται πλέον σε ολοκληρωμένα συστήματα εργαστηριακού αυτοματισμού, μαζί με ρομπότ που διαχειρίζονται δείγματα και καλλιέργειες. Έτσι, η τεχνολογία όχι μόνο μειώνει τον χρόνο διάγνωσης, αλλά συμβάλλει και στη γενικότερη αναδιοργάνωση των μικροβιολογικών εργαστηρίων προς πιο αποδοτικά και τυποποιημένα μοντέλα.

Πίνακας 11.3 Η εξέλιξη του MALDI-TOF από τη χημεία στην κλινική μικροβιολογία
Έτος **Εξέλιξη**

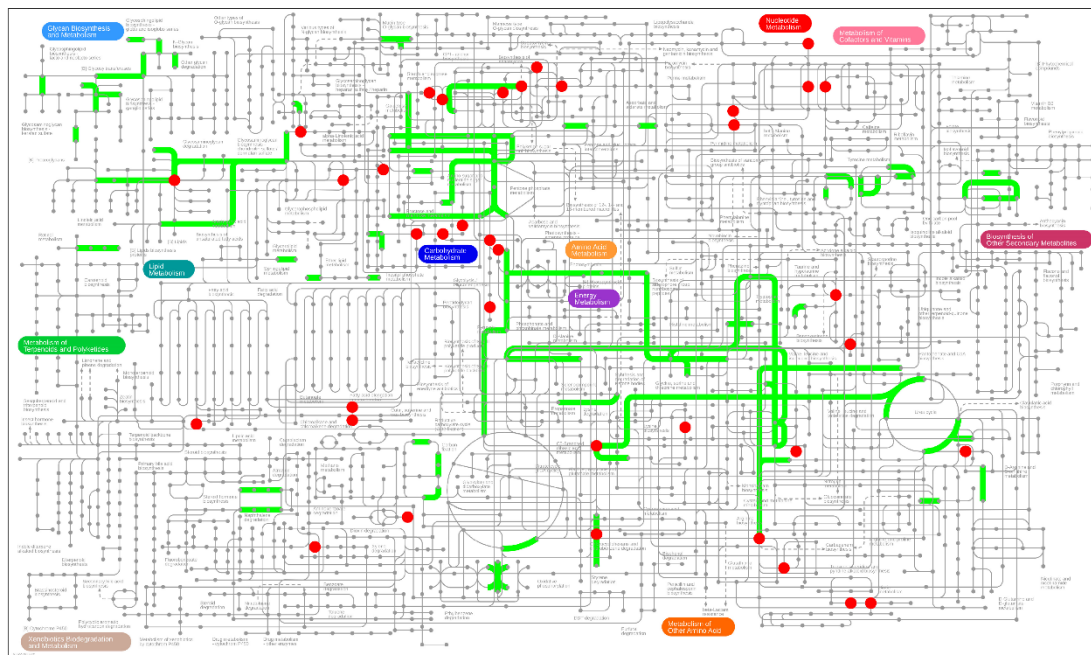
δεκαετία 1910	Η φασματομετρία μάζας (MS) χρησιμοποιείται στις χημικές επιστήμες.
δεκαετία 1980	Το MALDI-TOF MS εφαρμόζεται για την αναγνώριση ολιγονουκλεοτιδίων και πρωτεϊνών.
1994	Το MALDI-TOF MS χρησιμοποιείται για ταχεία αναγνώριση μικροοργανισμών (βακτήρια, μύκητες).
2009	Εισάγεται ευρέως στα κλινικά μικροβιολογικά εργαστήρια με άμεση χρήση αποικιών.
2015 και μετά	Εφαρμόζεται για την ανίχνευση αντοχής στα αντιμικροβιακά.

11.7 Από τις ωμικές τεχνολογίες στην πολύ-ωμική ανάλυση

Η ραγδαία ανάπτυξη των ωμικών τεχνολογιών τα τελευταία τριάντα χρόνια έχει σηματοδοτήσει μια πραγματική επανάσταση στις βιοϊατρικές επιστήμες. Ενώ οι πρώτες προσπάθειες χαρτογράφησης του ανθρώπινου γονιδιώματος αποτέλεσαν την απαρχή ενός νέου παραδείγματος, η εξέλιξη της τεχνολογίας και η πτώση του κόστους στις μεθόδους ανάλυσης υψηλής απόδοσης κατέστησαν δυνατή την ταυτόχρονη μελέτη χιλιάδων βιολογικών μορίων, ανοίγοντας τον δρόμο σε μια ολοκληρωμένη συστημική κατανόηση των ζωντανών οργανισμών. Η άνθιση αυτή δεν περιορίστηκε μόνο στο επίπεδο του DNA και του RNA αλλά επεκτάθηκε σε πολλαπλά άλλα επίπεδα, από τις επιγενετικές τροποποιήσεις μέχρι τα μικρά μόρια που ρυθμίζουν τον κυτταρικό μεταβολισμό. Η ανάγκη κατανόησης της πολυπλοκότητας των βιολογικών συστημάτων οδήγησε στη γέννηση διακριτών, εξειδικευμένων αλλά αλληλοσυμπληρούμενων «-ωμικών» πεδίων, τα οποία σταδιακά άρχισαν να συνδυάζονται μεταξύ τους στο πλαίσιο των λεγόμενων πολυ-ωμικών προσεγγίσεων.

Η επιγενωμική, που μελετά τις χημικές τροποποιήσεις του DNA και των ιστονών, ήρθε να συμπληρώσει την κλασική γονιδιωματική προσέγγιση, προσφέροντας εξηγήσεις για το πώς γονίδια που παραμένουν αμετάβλητα σε επίπεδο αλληλουχίας μπορούν να εκφράζονται διαφορετικά υπό την επίδραση περιβαλλοντικών, διατροφικών ή παθολογικών παραγόντων. Οι τεχνολογίες υψηλής απόδοσης κατέστησαν δυνατή τη χαρτογράφηση του **μεθυλώματος** σε επίπεδο γονιδιώματος, καθώς και την ταυτόχρονη καταγραφή πολλαπλών επιγενετικών τροποποιήσεων. Η αναγνώριση των τροποποιήσεων των ιστονών, όπως π.χ. η ακετυλίωση σε συγκεκριμένα αμινοξέα, με τη χρήση αντισωμάτων έδωσε τη δυνατότητα επιλογής των περιοχών του γονιδιώματος όπου αυτές οι τροποποιημένες ιστόνες εδράζονται. Η **ανοσοκατακρήμνιση** (Immunoprecipitation), δηλαδή η δέσμευση συγκεκριμένων τμημάτων της χρωματίνης μέσω ειδικών αντισωμάτων που προσδένονται σε σφαιρίδια, επιτρέπει την επιλεκτική αλληλούχησή τους (**Chromatin Immunoprecipitation sequencing, ChIP-seq**) και τον εντοπισμό των περιοχών που μεταβάλλουν τη δραστηριότητά τους σε διαφορετικές συνθήκες. Έτσι, σήμερα είναι εφικτό να μελετηθούν μηχανισμοί που εξηγούν την πλαστικότητα της κυτταρικής λειτουργίας, τη διαφοροποίηση των κυττάρων και την παθοφυσιολογία πολύπλοκων ασθενειών, όπως ο καρκίνος.

Εξίσου ραγδαία ήταν η ανάπτυξη της **μεταβολωμικής**, η οποία μελετά το σύνολο των μικρών μορίων που συμμετέχουν στον κυτταρικό μεταβολισμό. Η δυνατότητα καταγραφής εκατοντάδων ή και χιλιάδων **μεταβολιτών** μέσω τεχνικών όπως η φασματοσκοπία πυρηνικού μαγνητικού συντονισμού και η φασματομετρία μάζας επέτρεψε μια δυναμική αποτύπωση της φυσιολογικής κατάστασης ενός οργανισμού. Στην κλινική πράξη, η μεταβολωμική έφερε στο προσκήνιο βιοδείκτες για τον διαβήτη, τις καρδιαγγειακές νόσους, αλλά και την ογκολογία, ενώ στο πεδίο της φαρμακολογίας χρησιμοποιείται για την κατανόηση της φαρμακοκινητικής και της φαρμακοδυναμικής. Η ευαισθησία της μεταβολωμικής στις μεταβολές του περιβάλλοντος την καθιστά πολύτιμο εργαλείο για την παρακολούθηση της επίδρασης της διατροφής, της άσκησης και των τοξικών παραγόντων στην ανθρώπινη υγεία.



Εικόνα 11.10 Ολοκληρωμένη ανάλυση μεταγραφώματος και μεταβολώματος. Η ανάλυση παρουσιάζεται μέσω της πλατφόρμας iPath 3.0, η οποία επιτρέπει τη χαρτογράφηση των βιοχημικών μονοπατιών και την οπτική απεικόνιση της σύγκλισης διαφορετικών τύπων

δεδομένων. Στο γράφημα, τα μονοπάτια που προέρχονται από το μεταβολικό σύνολο αποδίδονται με κόκκινο χρώμα, ενώ εκείνα που προκύπτουν από το γονιδιακό σύνολο με πράσινο. Ο συνδυασμός αυτών των πληροφοριών αναδεικνύει κρίσιμες μεταβολικές και γονιδιακές διεργασίες, προσφέροντας μια ολοκληρωμένη εικόνα της βιολογικής κατάστασης και διευκολύνοντας τον εντοπισμό πιθανών σημείων αλληλεπίδρασης ή ρύθμισης.

Παράλληλα, η λιπιδωμική, ένας εξειδικευμένος κλάδος της μεταβολομικής, επικεντρώνεται στο σύνολο των λιπιδίων ενός κυττάρου ή ιστού. Τα λιπίδια, πέρα από αποθήκες ενέργειας, διαδραματίζουν κρίσιμο ρόλο ως δομικά συστατικά των μεμβρανών και ως σηματοδοτικά μόρια. Η μελέτη τους σε συστημικό επίπεδο αποκάλυψε πολύπλοκα δίκτυα ρύθμισης που σχετίζονται με τη φλεγμονή, την αθηροσκλήρωση, την παχυσαρκία και τον καρκίνο. Η πρόοδος στις αναλυτικές τεχνικές επέτρεψε την ποσοτικοποίηση χιλιάδων λιπιδικών μορίων, καθιστώντας δυνατή τη δημιουργία «λιπιδικών αποτυπωμάτων» με διαγνωστική και προγνωστική αξία.

Η άνθιση όλων αυτών των εξειδικευμένων τεχνολογιών οδήγησε αναπόφευκτα στην αναγνώριση ότι καμία μεμονωμένη «-ωμική» προσέγγιση δεν μπορεί από μόνη της να περιγράψει πλήρως την πολυπλοκότητα της βιολογίας. Έτσι, γεννήθηκαν οι πολυ-ωμικές προσεγγίσεις, οι οποίες συνδυάζουν δεδομένα από διαφορετικά επίπεδα, δημιουργώντας ένα ολοκληρωμένο πλέγμα πληροφοριών. Η πολυ-ωμική ανάλυση απαιτεί εξελιγμένα βιοπληροφορικά εργαλεία για την ενοποίηση, την ερμηνεία και την απεικόνιση των δεδομένων. Η πρόκληση είναι τεράστια, δεδομένου ότι κάθε «-ωμική» τεχνολογία παράγει τεράστιους όγκους δεδομένων, με διαφορετικά επίπεδα ευαισθησίας, θορύβου και τεχνικών περιορισμών. Ωστόσο, η δυνατότητα ενοποίησης αυτών των πληροφοριών οδηγεί σε μια πραγματικά συστημική κατανόηση. Στην ιατρική ακριβείας, οι πολυ-ωμικές προσεγγίσεις έχουν ήδη αρχίσει να μεταμορφώνουν τον τρόπο διάγνωσης και θεραπείας. Στην ογκολογία, για παράδειγμα, η πολυ-ωμική ανάλυση επιτρέπει την ταξινόμηση των όγκων όχι μόνο με βάση τις γενετικές μεταλλάξεις, αλλά και με βάση τις επιγενετικές τροποποιήσεις, το μεταβολικό προφίλ, προσφέροντας εξατομικευμένα θεραπευτικά μονοπάτια. Στις καρδιαγγειακές παθήσεις, ο συνδυασμός λιπιδωμικών και μεταβολομικών δεδομένων αποκαλύπτει μηχανισμούς που οδηγούν στην αθηροσκλήρωση, ανοίγοντας τον δρόμο για νέους στόχους φαρμακευτικής παρέμβασης. Η άνθιση των πολυ-ωμικών προσεγγίσεων συνδέεται άρρηκτα με την πρόοδο της **τεχνητής νοημοσύνης** στη βιοπληροφορική. Τα σύγχρονα εργαλεία **μηχανικής μάθησης** επιτρέπουν την αναγνώριση πολύπλοκων μοτίβων στα πολυδιάστατα δεδομένα, εντοπίζοντας συσχετίσεις που δεν θα ήταν δυνατόν να ανιχνευτούν με κλασικές μεθόδους. Η ανάπτυξη βάσεων δεδομένων, προτύπων αποθήκευσης και εργαλείων ανάλυσης έχει καταστήσει εφικτή την κοινοχρησία και την αξιοποίηση δεδομένων σε παγκόσμιο επίπεδο. Ωστόσο, η ενοποίηση ετερογενών δεδομένων παραμένει πρόκληση, τόσο σε τεχνικό επίπεδο όσο και σε επίπεδο βιολογικής ερμηνείας.

<https://link.springer.com/article/10.1007/s00292-024-01390-x>

<https://www.mdpi.com/2076-2607/12/2/322>

<https://www.jbc.org/article/S0021-9258%2819%2948534-7/fulltext>