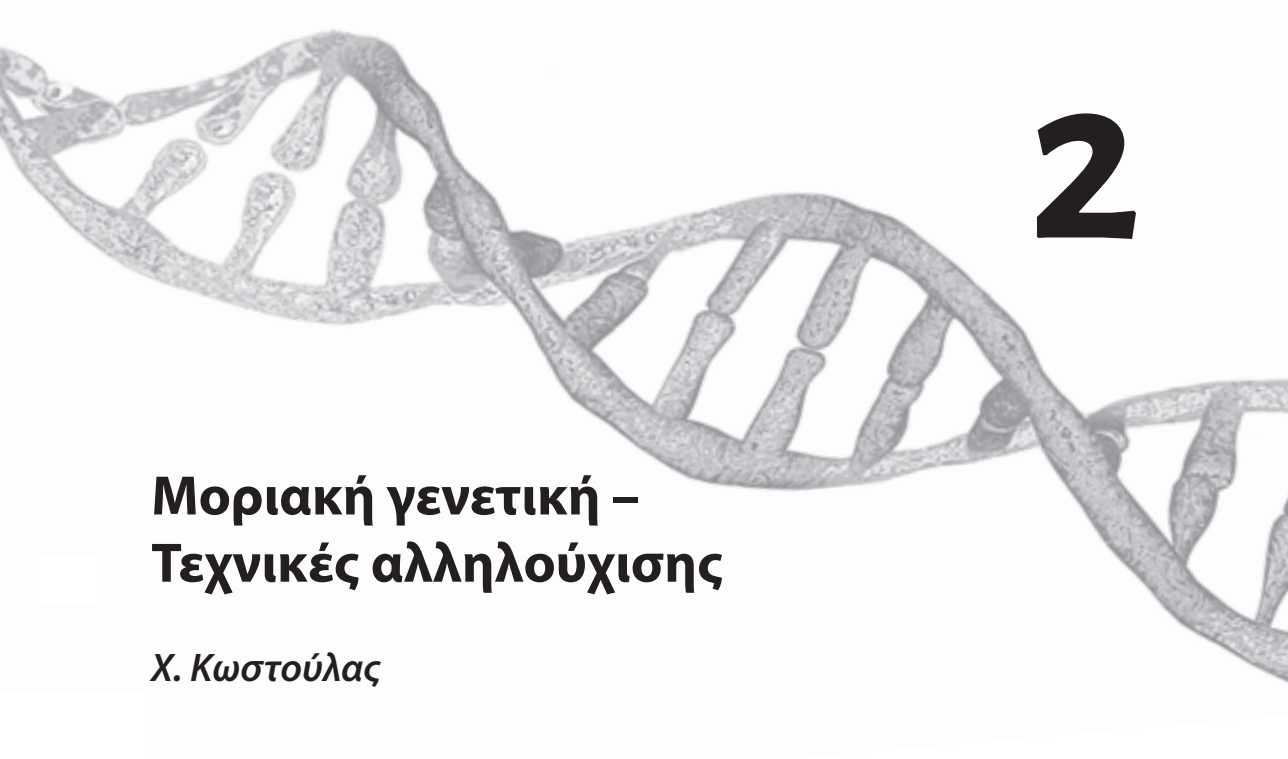


2

Μοριακή γενετική – Τεχνικές αλληλούχισης

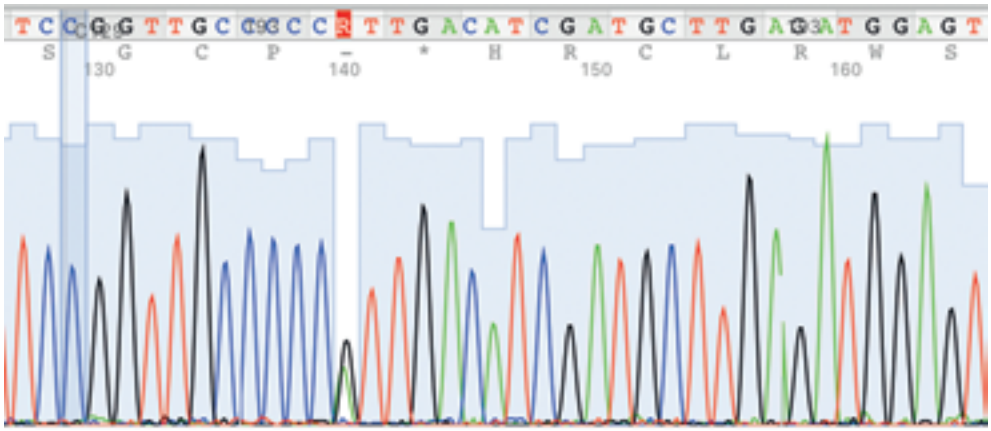
Χ. Κωστούλας

Εισαγωγή

Η μοριακή γενετική είναι ο κλάδος της βιολογίας —και ειδικότερα της γενετικής— που μελετά τη δομή και τη λειτουργία των γονιδίων σε μοριακό επίπεδο, καθώς και την κληρονομικότητα από τη μία γενιά στην άλλη. Το 99,9% του γονιδιώματος (DNA) δεν διαφέρει από άνθρωπο σε άνθρωπο, ενώ το υπόλοιπο 0,1% διαφέρει λόγω των παραλλαγών που εμφανίζονται στην αλληλουχία του DNA. Οι πληροφορίες που αντλούμε από τη μοριακή γενετική μπορούν να χρησιμοποιηθούν τόσο για τον καθορισμό των γενετικών παραλλαγών ή των μεταλλάξεων του DNA που μπορούν να προκαλέσουν κάποιο νόσημα όσο και για τον τρόπο κληρονόμησης του νοσήματος.

Από το 2003 και έπειτα, όταν και αποκωδικοποιήθηκε το ανθρώπινο γονιδίωμα, η αποκτηθείσα γνώση αναφορικά με την αλληλουχία του γονιδιώματος σε συνδυασμό με τις τεχνολογικές εξελίξεις στον τομέα των μοριακών τεχνικών αποτέλεσαν τα κυριότερα μέσα των γενετιστών για την ανίχνευση και την ταυτοποίηση όλων των παραλλαγών είτε αυτές βρίσκονται σε κωδικοποιούσες είτε σε μη κωδικές περιοχές.

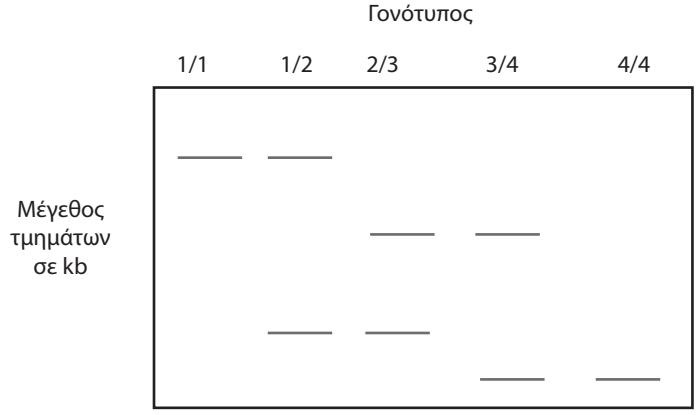
Η μοριακή γενετική ανάλυση βασίζεται σε μια ευρεία γκάμα τεχνικών όπως η αλληλούχιση κατά Sanger **(Εικ. 1)**, η πολλαπλή ενίσχυση ανιχνευτών εξαρτώμενη από την αντίδραση λιγάσης (MLPA – Multiplex Ligation-dependent Probe Amplification) (βλ. Κεφάλαιο 4), Southern Blot **(Εικ. 2)**, η πολλαπλή παράλληλη αλληλούχιση ή αλληλούχιση νέας γενιάς (NGS – Next Generation Sequencing) κ.ά. Οι τεχνικές αυτές μας επιτρέπουν να ανιχνεύσουμε παραλλαγές ενός νουκλεοτιδίου (SNVs – Single Nucleotide Variations), μικρές ενθέσεις ή διαγραφές βάσεων (indels), υπο-μικροσκοπικές παραλλαγές στον αριθμό αντιγράφων ενός ή πολλαπλών τμημάτων DNA



Εικόνα 1 *Χρωματογράφημα αλληλούχισης με τη μέθοδο Sanger.*

(CNVs – Copy Number Variations), καθώς και ελλείμματα ή διπλασιασμούς μεγαλύτερων τμημάτων DNA.

Επιπροσθέτως, η μοριακή γενετική εφαρμόζεται στη διάγνωση μονογονιδιακών νοσημάτων, πολυπαραγοντικών νοσημάτων, πολυγονιδιακών νοσημάτων, ογκολογικών/αιματολογικών νοσημάτων, σε νοσήματα που οφείλονται σε μη φυσιολογική γονεϊκή αποτύπωση (imprinting), σε μονογονεϊκή δισωμία (Uniparental Disomy – UPD), σε τρινουκλεοτιδικές επαναλήψεις (trinucleotide repeat disorders) και σε μιτοχονδριακά νοσήματα. Επίσης, η μοριακή γενετική εφαρμόζεται τόσο σε προγεννητικό όσο και σε μεταγεννητικό επίπεδο σε δείγματα αμνιακού υγρού, τροφοβλάστης, περιφερικού εμβρυϊκού αίματος, περιφερικού ολικού αίματος, βιοψίες συμπαγών όγκων ή μυελού των οστών και υγρές βιοψίες. Επίσης, τα τελευταία χρόνια βρίσκει ευρεία εφαρμογή και στη φαρμακογονιδιωματική.



Εικόνα 2 *Απεικόνιση Southern Blot.*

Αλληλούχιση κατά Sanger

Η αλληλούχιση του DNA είναι η διαδικασία προσδιορισμού των νουκλεοτιδικών βάσεων, δηλαδή της πρωτοταγούς δομής, ενός τμήματος DNA. Η αλληλούχιση κατά Sanger αποτελεί μία από τις σημαντικότερες μοριακές τεχνικές του 20ού αιώνα. Αποτελεί την τεχνική με την οποία επετεύχθη η αποκωδικοποίηση του ανθρώπινου γονιδιώματος (Human Genome Project), η οποία ξεκίνησε το 1990 και ολοκληρώθηκε το 2003.

Η αλληλούχιση Sanger ή αλληλούχιση τερματισμού ή αλληλούχιση διδεοξυνουκλεοτιδίων (ddNTPs) εφευρέθηκε τη δεκαετία του 1970 από τον χημικό Frederick Sanger. Η αρχή της τεχνικής βασίζεται στην επιμήκυνση της αλληλουχίας DNA με την ενσωμάτωση ενός νουκλεοτιδίου μέσω της αντίδρασης του πολυμερισμού.

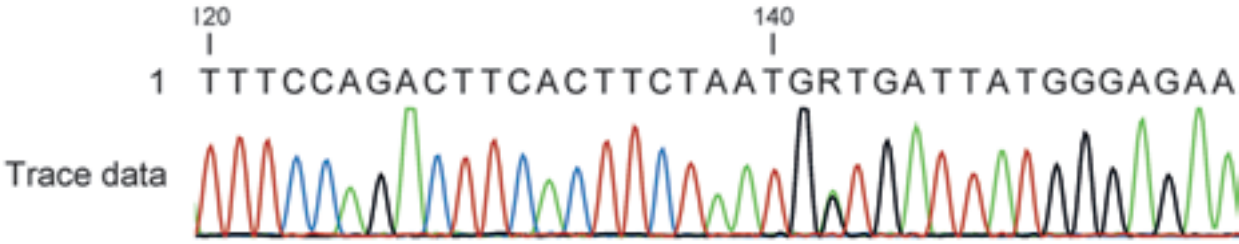
Η αλληλούχιση κατά Sanger απαιτεί την ύπαρξη

- ενός εκκινητή (primer) συμπληρωματικού με την αλληλουχία-στόχο,
- DNA πολυμεράσης για τον πολυμερισμό του εκκινητή,
- 4 δεοξυνουκλεοτιδίων (dATP, dGTP, dCTP, dTTP) και
- 4 διδεοξυνουκλεοτιδίων (ddATP, ddGTP, ddCTP, ddTTP). Το καθένα από αυτά είναι σημασμένο με διαφορετική φθορίζουσα ουσία, η οποία εκπέμπει ακτινοβολία σε διαφορετικό μήκος κύματος. Το νουκλεοτίδιο **A** σημαίνεται με χρώμα **πράσινο**, το **G** με **μαύρο**, το **C** με **μπλε** και το **T** με **κόκκινο** **(Εικ. 3)**.

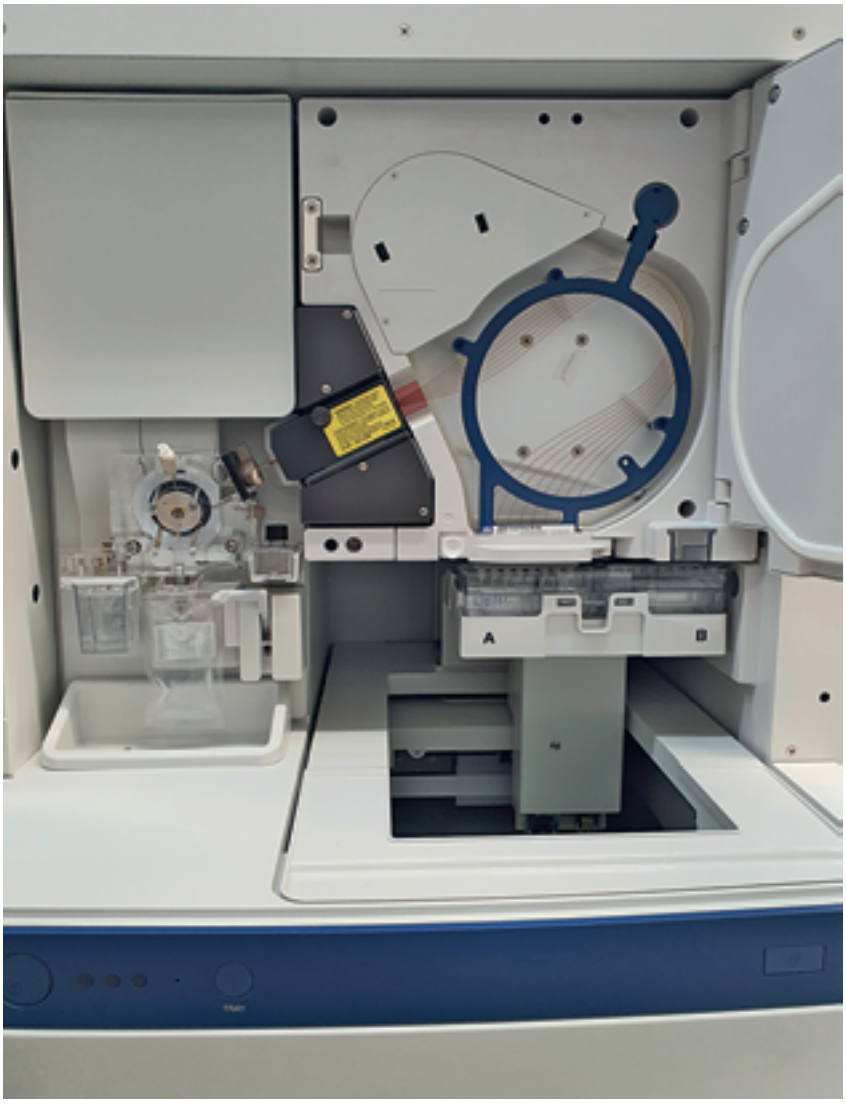
Τα ddNTPs είναι παρόμοια με τα dNTPs, με τη διαφορά, όμως, ως προς το ότι τα ddNTPs στον 3΄ άνθρακα του σακχάρου δεν φέρουν υδροξυλομάδα, όπως τα dNTPs. Έτσι, όταν προστεθεί ένα ddNTP στη νεοσυντιθέμενη αλυσίδα, η πρόσδεση άλλου νουκλεοτιδίου, είτε αυτό είναι δεοξυνουκλεοτίδιο είτε διδεοξυνουκλεοτίδιο, δεν καθίσταται εφικτή, με αποτέλεσμα να πραγματοποιείται τερματισμός σύνθεσης του DNA. Αυτό σημαίνει ότι έχουμε διαφορετικού μεγέθους τμήματα, στο 3΄ άκρο των οποίων η τελευταία βάση φέρει μία φθορίζουσα ουσία. Όταν ολοκληρωθεί η αντίδραση πολυμερισμού, τα τμήματα με διαφορετικό μέγεθος ηλεκτροφορούνται σε ειδικά τριχοειδή σωληνάρια (capillaries), τα οποία περιέχουν ειδικό πήκτωμα, μία διαδικασία γνωστή ως ηλεκτροφόρηση σε τριχοειδή **(Εικ. 4)**.

Τα μικρά σε μέγεθος τμήματα κινούνται πιο γρήγορα στο πήκτωμα, ενώ τα μεγαλύτερα πιο αργά (πρώτο θα περάσει το τμήμα που φέρει ddNTP στην πρώτη βάση μετά τον εκκινητή, δεύτερο θα περάσει το τμήμα που φέρει ddNTP στη δεύτερη βάση μετά τον εκκινητή κ.ο.κ.). Στο τέλος της ηλεκτροφόρησης, τα τμήματα διαφορετικού μεγέθους ανιχνεύονται με laser με βάση τη φθορίζουσα ουσία του ddNTP ανιχνευτή και μέσω του κατάλληλου λογισμικού η εκπομπή του φωτός και το σήμα μετατρέπονται στις τέσσερις χρωστικές **(Εικ. 5)**.

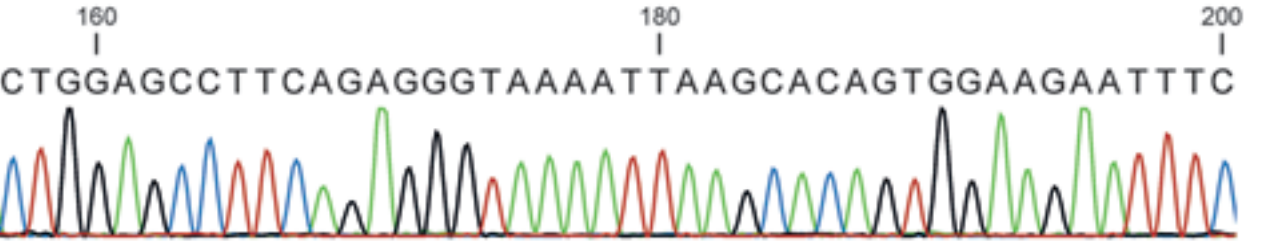
Με την αλληλούχιση κατά Sanger μπορούν να ανιχνευθούν παραλλαγές ενός νου-



Εικόνα 3 *Χρωματογράφημα, στο οποίο η κάθε κορυφή αποτελεί και μια νουκλεοτιδική βάση.*

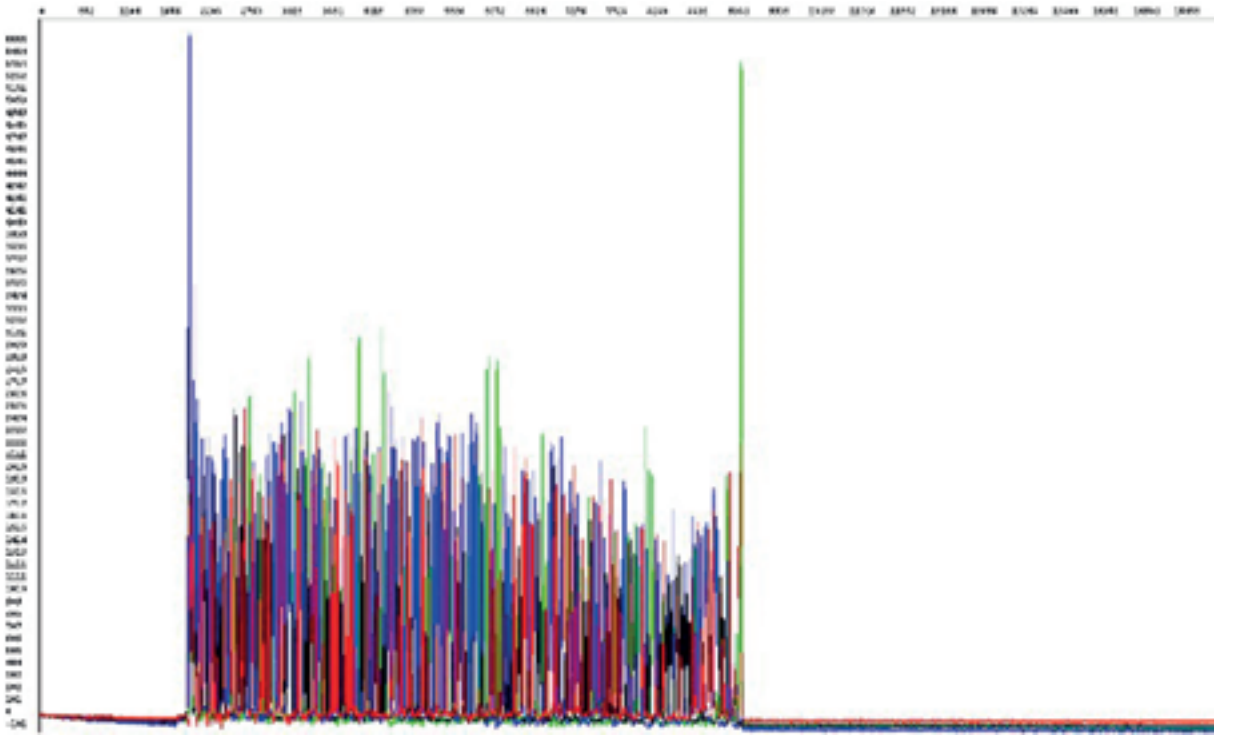


Εικόνα 4 *Αυτόματος αλληλουχητής Sanger της εταιρείας Applied Biosystems.*

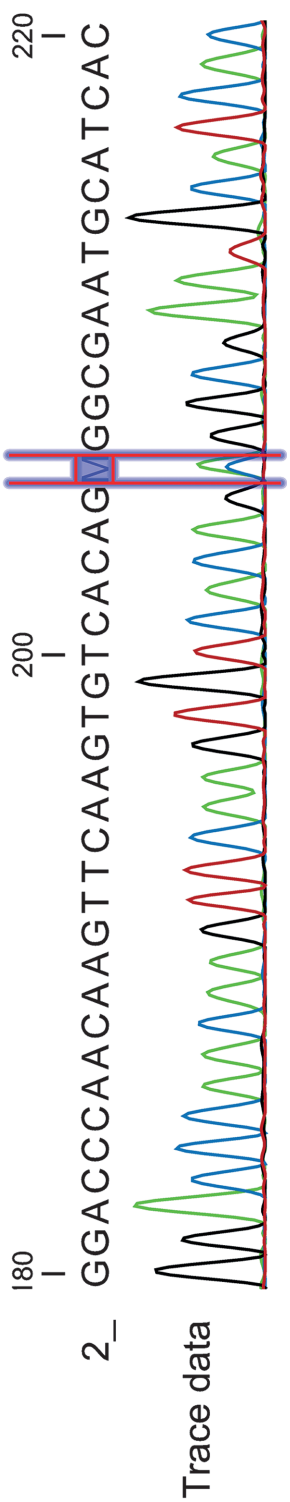


κλεοτιδίου (SNVs – Single Nucleotide Variations) **(Εικ. 6)** και μικρές ενθέσεις ή διαγραφές —γνωστές ή καινούργιες— κάποιων βάσεων (INDELs) **(Εικ. 7)**. Επίσης, αποτελεί την κυριότερη τεχνική για τον έλεγχο και την επιβεβαίωση παραλλαγών στην πρωτοταγή δομή των γονιδίων, οι οποίες σχετίζονται με νοσήματα.

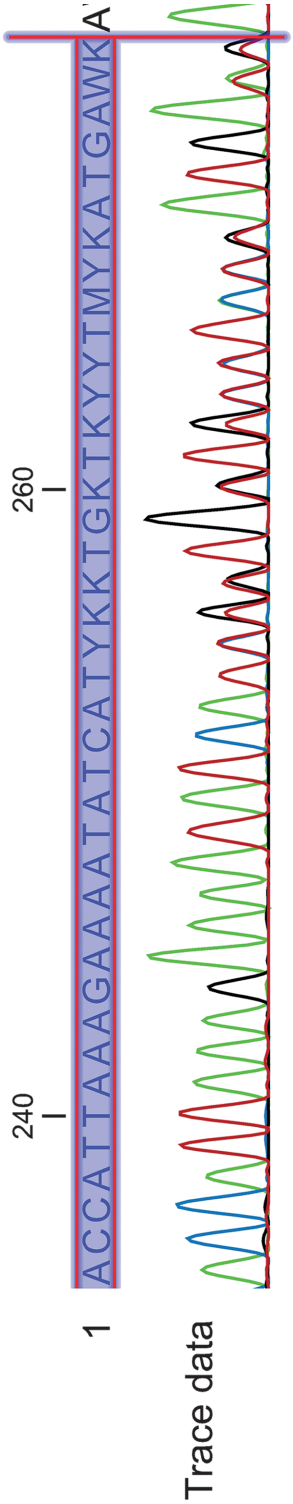
Το πλεονέκτημα της αλληλούχισης κατά Sanger είναι η πολύ μεγάλη ακρίβεια με την οποία προσδιορίζεται η πρωτοταγής δομή σχετικά μεγάλων τμημάτων DNA (έως και 1.200 ζεύγη βάσεων). Αυτό την καθιστά ιδανική λύση για την ταυτοποίηση παραλλαγών τόσο σε συγκεκριμένα εξόνια ενός γονιδίου όσο και σε όλα τα εξόνια



Εικόνα 5 *Μη επεξεργασμένο χρωματογράφημα, με έντονες ισοϋψείς κορυφές.*



Εικόνα 6 *Ετεροζυγωτία σημειακής παραλλαγής με τη χαρακτηριστική εικόνα 2 κορυφών στην ίδια θέση. Ετεροζυγωτία για την παθογονική παραλλαγή c.858C>A (p.Ser286Arg) του γονιδίου LDLR, η οποία προκαλεί οικογενή υπερχοληστερολαιμία.*



Εικόνα 7 *Εικόνα ετεροζυγης διαγραφής ή ένθεσης μιας ή περισσότερων βάσεων. Το χρωματογράφημα μας δείχνει την ετεροζυγη εικόνα ενός φορέα της πιο συχνής παθογονικής παραλλαγής του CFTR γονιδίου, την c.1521_1523del (p.Phe508del), γνωστή ως ΔF508.*

των γονιδίων. Αποτελεί την Gold Standard τεχνική ακόμα και σήμερα για την επιβεβαίωση νέων παραλλαγών που βρέθηκαν με άλλες μοριακές τεχνικές.

Τα σημαντικότερα μειονεκτήματα είναι:

- Έχει σχετικά υψηλό κόστος και είναι αρκετά κοστοβόρα και χρονοβόρα για έλεγχο ταυτόχρονα αρκετών γονιδίων ή μεγάλων γονιδίων για αρκετά μεγάλο αριθμό δειγμάτων.
- Αδυναμία ανίχνευσης ενθέσεων/ελλειμμάτων >20 βάσεων.
- Πιθανότητα αποτυχίας πολλαπλασιασμού (PCR) του ενός αλληλομόρφου (Allele Drop Out – ADO).
- Αδυναμία ανίχνευσης μωσαϊκισμού σε χαμηλά επίπεδα (<20%).
- Αδυναμία προσδιορισμού παραλλαγών με συχνότητα εμφάνισης <10%. Οι παραλλαγές με χαμηλή/μικρή συχνότητα εμφάνισης – Variant Allele Frequency (VAF<10%) σε δείγματα όγκων δεν μπορούν να προσδιοριστούν με αλληλούχιση κατά Sanger.
- Δεν μπορεί να προσδιορίσει εάν οι παραλλαγές βρίσκονται στο ίδιο αλληλόμορφο ή σε διαφορετικό αλληλόμορφο (cis/trans), μια πληροφορία σημαντική για τα υπολειπόμενα νοσήματα.

Η τεχνική MLPA που αναλύεται διεξοδικά στο Κεφάλαιο 4 αποτελεί μία από τις σημαντικότερες τεχνικές για τον έλεγχο των γονιδίων για γενωμικές αναδιατάξεις και παραλλαγές στον αριθμό αντιγράφων τμημάτων DNA. Λόγω αυτής της ικανότητας, η MLPA μπορεί να χρησιμοποιηθεί στη μοριακή διάγνωση γενετικών ασθενειών, η παθογένεια των οποίων σχετίζεται με ελλείμματα ή επαναλήψεις συγκεκριμένων περιοχών του DNA, με νοσήματα τα οποία σχετίζονται με παρουσία αλλαγών στη μεθυλίωση του DNA και με περιπτώσεις μονογονεϊκής δισωμίας (Σύνδρομα Russel-Silver, Kagami-Ogata, Prader-Willi/Angelman κ.ά.).

Αλληλούχιση νέας γενιάς (Next Generation Sequencing – NGS) ή μαζική παράλληλη αλληλούχιση (Massive parallel sequencing)

Όπως αναφέραμε ήδη, το 2003 αποτελεί μία χρονιά ορόσημο για την εξέλιξη της βιολογίας, καθώς ολοκληρώθηκε η ανάλυση του ανθρώπινου γονιδιώματος. Το Human Genome Project, στο οποίο συμμετείχαν εκατοντάδες εργαστήρια και ερευνητές από όλο τον κόσμο, διήρκεσε δώδεκα χρόνια και κόστισε σχεδόν τρία δισεκατομμύρια δολάρια. Η τεχνική που χρησιμοποιήθηκε για την αποκωδικοποίηση, η τεχνολογία αλληλούχισης κατά Sanger, είχε δύο μειονεκτήματα: το ένα σχετίζεται με το υψηλό κόστος και το δεύτερο με το ότι αποτελεί μία σχετικά χρονοβόρα τεχνική, που την καθιστούν μη «αποτελεσματική» για την αλληλούχιση μεγάλων γονιδιωμάτων ή ακόμα και πολλών γονιδίων. Αυτά οδήγησαν στην ανάπτυξη της αλληλούχισης νέας γενιάς (Next Generation Sequencing – NGS). Με την τεχνολογία

NGS, μας δίνεται η δυνατότητα αλληλούχισης ενός πάνελ γονιδίων (στοχευμένη αλληλούχιση, Targeted Sequencing – TS), ολόκληρου του γονιδιώματος (Whole Genome Sequencing – WGS) ή όλων των εξονίων του γονιδιώματος (Whole Exome Sequencing – WES, **Εικ. 6**) σε λίγες ώρες έως ημέρες, ανάλογα με την τεχνολογία και το πρωτόκολλο που χρησιμοποιείται/χρησιμοποιούνται. Είναι αξιοσημείωτο ότι σήμερα με την τεχνολογία NGS μπορεί να αναλυθεί το ανθρώπινο γονιδίωμα μέσα σε μία μέρα και με κόστος 1.000$.

Η τεχνολογία NGS περιλαμβάνει τέσσερα σημαντικά βήματα:

α) τον κατακερματισμό του DNA,
β) τη δημιουργία της βιβλιοθήκης,
γ) τη μαζική παράλληλη αλληλούχιση,
δ) τη βιοπληροφορική ανάλυση και την κλινική ερμηνεία των παραλλαγών που εξορύχθηκαν από τη βιοπληροφορική ανάλυση.

Ο κατακερματισμός του DNA πραγματοποιείται για την περιοχή ενδιαφέροντος σε πολλά μικρότερα θραύσματα (fragments), μεγέθους 75-600 bp. Το DNA μπορεί να κατακερματιστεί είτε με μηχανικές μεθόδους (υπέρηχους – sonication), είτε με μοριακές μεθόδους (πέψη με ένζυμα) είτε και με άλλες μεθόδους. Τα θραύσματα, από τον κατακερματισμό της περιοχής ενδιαφέροντός μας, απομονώνονται με τη χρήση ειδικών ανιχνευτών οι οποίοι είναι συμπληρωματικοί με τις περιοχές-στόχους. Αυτή η μέθοδος αναφέρεται συνήθως ως επιλογή περιοχών ενδιαφέροντος με υβριδισμό (Hybrid capture-based). Μία άλλη μέθοδος επιλογής των περιοχών ενδιαφέροντος είναι η ενίσχυση των περιοχών αυτών με τη χρήση PCR. Με αυτήν τη μέθοδο, πολλά ζεύγη εκκινητών χρησιμοποιούνται για την ενίσχυση των περιοχών-στόχου του DNA χρησιμοποιώντας PCR. Στη συνέχεια, τα τμήματα DNA χρησιμοποιούνται για τη δημιουργία της βιβλιοθήκης.

Η NGS τεχνολογία μας δίνει τη δυνατότητα να αλληλουχίσουμε ταυτόχρονα πολλά διαφορετικά δείγματα σε μια μόνο αντίδραση. Η δημιουργία της βιβλιοθήκης είναι μία διαδικασία κατά την οποία στο 5΄ και 3΄ άκρο των τμημάτων DNA προσδένονται μικρές αλληλουχίες (adaptors) σε κάθε διαφορετικό δείγμα, οι οποίες έχουν συγκεκριμένη αλληλουχία αλλά είναι διαφορετική για το κάθε δείγμα, αποτελώντας έτσι και την «ταυτότητα» του δείγματος. Έπειτα, οι βιβλιοθήκες φορτώνονται σε κατάλληλο πλακάκι ανάλογα με την εταιρεία (flow cell-Illumina, chip-Ion Torrent, **Εικ. 8**) όπου πραγματοποιείται η αλληλούχιση. Τέλος, γίνεται η βιοπληροφορική ανάλυση των δεδομένων με τη χρήση κατάλληλων εργαλείων ανάλυσης και βάσεων δεδομένων για την εύρεση των παραλλαγών στα υπό μελέτη γονίδια αλλά και για την ερμηνεία των παραλλαγών και τη συσχέτιση γονοτύπου – φαινοτύπου.

Τα τελευταία 10 χρόνια, οι τεχνολογίες αλληλούχισης νέας γενιάς (NGS) έχουν αποκτήσει έναν όλο και πιο σημαντικό ρόλο στη διερεύνηση και τη διάγνωση ασθενειών. Με την NGS τεχνολογία μπορούν να ανιχνευθούν SNVs **(Εικ. 9)**, INDELs