



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ**

**Η χρήση της Βιοπληροφορικής
στη μελέτη της νόσου του Καρκίνου**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ευτυχία Σ. Νανά

Επιβλέπων : Δημήτριος Κουτσούρης

Καθηγητής Ε.Μ.Π

Αθήνα, Οκτώβριος 2009



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ**

**Η χρήση της Βιοπληροφορικής
στη μελέτη της νόσου του Καρκίνου**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ευτυχία Σ. Νανά

Επιβλέπων : Δημήτριος Κουτσούρης
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 13 Οκτωβρίου 2009

.....
Δ. Κουτσούρης
Καθηγητής Ε.Μ.Π

.....
Κ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

.....
Π. Τσανάκας
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2009

.....
Ευτυχία Σ. Νανά

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Ε.Μ.Π.

Copyright © Ευτυχία Σ. Νανά, 2009.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Ο καρκίνος είναι πρωταρχικά η ασθένεια του γονιδίου. Ο τρόπος με τον οποίο επιχειρεί την κυριαρχία του επάνω στη ζωή εξαρτάται σε μεγάλο βαθμό από τη μοριακή υπογραφή του καθενός από εμάς. Παραδόξως, στην τεράστια μέχρι τώρα άγνοια προστίθεται το σημερινό χάος της γνώσης. Η Βιοπληροφορική, με τα εργαλεία και το περιβάλλον της, δοκιμάζει να δαμάσει αυτό το χάος. Για πρώτη φορά ο επιστήμονας έχει τη δυνατότητα να ανοίξει έναν απευθείας διάλογο με τη γονιδιακή του ταυτότητα ως ανθρώπου, μεταπηδώντας από την εποχή της παροχής των ιατρικών υπηρεσιών που βασίζονται στη στατιστική σε μια νέα εποχή, αυτή της εξατομικευμένης θεραπείας.

Λέξεις κλειδιά: Βιοπληροφορική, Καρκίνος, Ακολουθιακές μέθοδοι, Μικροδιατάξεις.

Abstract

Cancer is primarily a disease of the gene. The way it tries to prevail upon life greatly depends on the molecular signature of each one of us. Oddly enough, the up-to-know enormous amount of ignorance is added to the current chaos of knowledge. Bioinformatics, with its tool and environment, tries to tame this chaos. For the first time, the scientist has the ability to open a direct dialogue with his/her genetic identity as a human being. We pass from the era of providing medical services that are based on statistics to a new era, that of a personalized treatment.

Key words: Bioinformatics, Cancer, Sequencing Methods, Microarrays.

Ευχαριστίες

Είναι μεγάλη χαρά και αληθινή ευλογία να έχεις δίπλα σου ανθρώπους που, ο καθένας με τον δικό του τρόπο, συμπαρίστανται, σε βοηθούν και ενισχύουν. Είχα αυτή τη χαρά και ευλογία και κατά τη συγγραφή της παρούσας διπλωματικής εργασίας και θα ήθελα να εκφράσω την ευγνωμοσύνη μου σε όλους εκείνους τους ανθρώπους.

Θα ήθελα καταρχάς να ευχαριστήσω τον καθηγητή κ. Δημήτριο Κουτσούρη που μου εμπιστεύτηκε αυτή την διπλωματική και είχε την επίβλεψη της.

Επίσης, θερμές ευχαριστίες θα ήθελα να εκφράσω στον μεταδιδασκτορικό συνεργάτη του εργαστηρίου Βιοιατρικής Τεχνολογίας, τον κ. Ιωάννη Μακρή ο οποίος, καθ' όλη τη διάρκεια της συγγραφής, μου προσέφερε συνεχή βοήθεια και καθοδήγηση. Τον ευχαριστώ και για την εμπιστοσύνη που μου έδειξε και την ελευθερία που μου έδωσε σχετικά με την διαχείριση του χρόνου συγγραφής.

Ευχαριστώ επίσης θερμά, την αγαπητή μου φίλη και συμφοιτήτρια Ιωάννα Μαρκοπούλου, για την καθοριστική συμβολή της και τον τόσο πολύτιμο για εκείνη χρόνο που αφιέρωσε για να με βοηθήσει.

Επίσης θα ήθελα να εκφράσω βαθιά και ειλικρινή ευγνωμοσύνη στον σύζυγό μου, Κωστή Μιχμίζο, για την αδιάκοπη συμπαράστασή του, για την υπομονή και ανοχή του προκειμένου να έχω την δυνατότητα να αφιερώνω χρόνο στην συγγραφή της διπλωματικής μου, αλλά και για την καθοριστική βοήθεια του, χωρίς την οποία ίσως να μην ήταν δυνατόν να ολοκληρωθεί η παρούσα διπλωματική.

Τέλος, ολόψυχα ευχαριστώ τους γονείς μου, Σεραφείμ και Μαρίνα Νανά, για την απύθμενη αγάπη και τις θυσίες τους που από την γέννηση μου με στηρίζουν, με διδάσκουν και με συνοδεύουν μέχρι εδώ αλλά και από εδώ και πέρα. Χωρίς τη δική τους παρουσία στη ζωή μου πολλά, ακόμα και αυτή η διπλωματική, θα ήταν διαφορετικά.

Πίνακας Περιεχομένων

Περίληψη.....	7
Abstract.....	9
Ευχαριστίες.....	11
Πίνακας Περιεχομένων	13
1 Η Επιστήμη της Βιοπληροφορικής	17
1.1 Εισαγωγή στη Βιοπληροφορική	17
1.2 Ιστορική Αναδρομή.....	18
1.3 Μελλοντικοί στόχοι και προκλήσεις.....	21
1.4 Προβληματισμοί.....	23
Βιβλιογραφία.....	24
2 Στοιχεία Βιολογίας.....	27
2.1 Εισαγωγή στο γονιδίωμα	27
2.2 Η Δομή του DNA	30
2.3 Η ροή της γενετικής πληροφορίας	32
2.4 Το ένζυμο - ελικάση.....	34
2.5 Μεταγραφή του DNA.....	35
2.6 Ο γενετικός κώδικας.....	38
2.7 Μετάφραση	39
2.8 Η γονιδιακή ρύθμιση στους προκαρυωτικούς οργανισμούς	42
2.9 Η γονιδιακή ρύθμιση στους ευκαρυωτικούς οργανισμούς	44

2.10	Ποσοστό Έκφρασης Γονιδίων.....	44
2.11	Η υβριδοποίηση των νουκλεικών οξέων χρησιμοποιείται για την ανίχνευση κλώνων γονιδιωματικής ή cDNA βιβλιοθήκης.....	45
	Βιβλιογραφία.....	47
3	Καρκίνος: Μια πολυγονιδιακή νόσος	49
3.1	Τι είναι ο Καρκίνος.....	49
3.2	Καρκινογένεση	49
3.3	Μορφές Καρκίνου	52
3.4	Εξατομικευμένη Θεραπεία του καρκίνου.....	53
3.5	Η χρήση της Βιοπληροφορικής στην μελέτη του καρκίνου.....	53
	Βιβλιογραφία.....	54
4	Ακολουθιακές Μέθοδοι	55
4.1	Εισαγωγή	55
4.2	Ακολουθιακές μέθοδοι.....	55
4.3	Αλγόριθμοι σύγκρισης ακολουθιών	55
4.3.1	Αλγόριθμοι Ταιριάσματος Προτύπου.....	55
4.3.2	Αλγόριθμοι Ευθυγράμμισης Ακολουθιών – Sequence Alignment Algorithms.....	56
4.3.3	Αλγόριθμοι Ευθυγράμμισης Πολλαπλών Ακολουθιών	57
4.3.3.1	Ο αλγόριθμος FASTA.....	57
4.3.3.2	BLAST - Basic Local Alignment Search Tool.....	58
4.3.3.3	CLUSTALw	60
4.4	Βιολογικές Βάσεις Δεδομένων.....	61
4.4.1	Βάσεις Δεδομένων Νουκλεοτιδικών Ακολουθιών	61
4.4.2	Βάσεις Δεδομένων Πρωτεϊνικών Ακολουθιών	62
4.5	Η χρήση των ακολουθιακών μεθόδων στην μελέτη του Καρκίνου	72
4.5.1	Expressed Sequence Tags (ESTs)	73
4.5.2	Serial Analysis of Gene Expression SAGE.....	73
4.5.3	Massively Parallel Signature Sequencing MPSS.....	74
4.5.4	Single Nucleotide Polymorphisms (SNPs)	74
	Βιβλιογραφία.....	75
5	Τεχνικές Μικροδιατάξεων.....	79

5.1	Εισαγωγή στην τεχνολογία των Μικροδιατάξεων	79
5.2	Η Τεχνολογία των Μικροδιατάξεων.....	81
5.2.1	Μικροδιατάξεις cDNA.....	81
5.2.2	Τα Βήματα ενός Πειράματος Υβριδισμού Μικροδιατάξεων.....	83
5.2.3	Κατασκευή μιας μικροδιάταξης.....	84
5.3	Η χρήση των Μικροδιατάξεων στη μελέτη του Καρκίνου	85
5.3.1	Μικροδιατάξεις cDNA.....	86
5.3.2	Μικροδιατάξεις Ολιγονουκλεοτιδίων.....	87
5.3.3	Μικροδιατάξεις SNP	88
5.3.4	Παράδειγμα Χρήσης Μικροδιατάξεων στην Ανίχνευση του Καρκίνου του Μαστού.....	90
	Βιβλιογραφία.....	91
6	Μέθοδοι Πρωτεϊνικής Ανάλυσης.....	97
6.1	Εισαγωγή	97
6.2	Πρωτεωμική	97
6.3	Κλασικές Μέθοδοι Πρωτεϊνικής Ανάλυσης	98
6.4	Η συνεισφορά της Πρωτεϊνικής Ανάλυσης στη μελέτη του Καρκίνου ...	99
	Βιβλιογραφία.....	100
7	Λοιπές Μέθοδοι Μελέτης της Νόσου του Καρκίνου	101
7.1	Εισαγωγή	101
7.2	Η συμβολή των πειραμάτων σε ζώα στη μελέτη του Καρκίνου.....	101
7.3	Εισαγωγή της Βιοπληροφορικής στα πειράματα με GEM πειραματόζωα. 102	
7.3.1	Μέθοδοι πρόβλεψης κλάσης.....	103
7.3.2	Μελέτη των Μεταστάσεων	104
	Βιβλιογραφία.....	105
8	Συμπεράσματα για το Μέλλον της Βιοπληροφορικής στην Αντιμετώπιση του Καρκίνου	107

1

Η Επιστήμη της Βιοπληροφορικής

1.1 Εισαγωγή στη Βιοπληροφορική

Στον επιστημονικό χώρο είναι δύσκολο κανείς να μιλήσει για μια αυτοτελή επιστήμη, τα πράγματα γίνονται ακόμη πιο δύσκολα όταν μιλάμε για την Βιοπληροφορική η οποία ουσιαστικά αποτελεί συνένωση των επιστημών της Βιολογίας και της Πληροφορικής. Μόλις την δεκαετία του 1990 εμφανίζεται ως ανεξάρτητη επιστήμη η οποία «πατά σε ώμους γιγάντων» που η παράλληλη ανάπτυξη τους οδήγησε στην εκπληκτική συνάντησή τους. Για πρώτη φορά η πληροφορική εφαρμόζεται ως εργαλείο για την μοριακή βιολογία παρέχοντας της την υπολογιστική ισχύ και τις μηχανικές μεθόδους ακριβώς τη χρονική στιγμή που εκείνη τις χρειάζεται και έτσι γεννιέται μία νέα επιστήμη, που αναπτύσσεται γοργά φέρνοντας επανάσταση στην επιστημονική κοινότητα αλλά και στην ίδια την κοινωνία.

Τι τελικά είναι η Βιοπληροφορική πρωτοδιατυπώθηκε από τον Δρ. Temple Smith, καθηγητή του Πανεπιστημίου της Βοστώνης και αν θέλαμε να δώσουμε έναν σαφή ορισμό θα λέγαμε ότι

« Η επιστήμη της Βιοπληροφορικής είναι συνένωση των επιστημών της Μοριακής Βιολογίας και της Πληροφορικής. Έχει ως αντικείμενο της την βέλτιστη συνεργασία τους για το καλό (ελπίζουμε) της ανθρωπότητας, πρακτικά στην αποθήκευση, επεξεργασία και ανάλυση των βιολογικών δεδομένων »

Αυτή η καταπληκτική συνένωση μπορεί να γίνει κατανοητή με ένα απλό παράδειγμα. Η αλληλουχία των βάσεων σε έναν κλώνο της αλυσίδας του DNA μπορεί να απεικονιστεί ως συμβολοσειρά (string). Αν απεικονίσουμε την αδενίνη με A την θυμίνη με T την γουανίνη με G και την κυτοσίνη με C, μια αλληλουχία αδενίνη – γουανίνη – κυτοσίνη - θυμίνη αναπαρίσταται από την συμβολοσειρά AGCT. Έτσι η πληροφορία που φέρει η αλληλουχία είναι πλέον εύκολα προσβάσιμη και είναι δυνατόν να αποθηκευτεί, να επεξεργαστεί (πχ με αλγορίθμους σύγκρισης συμβολοσειρών) κτλ.

1.2 Ιστορική Αναδρομή

Η ραγδαία ανάπτυξη της Βιοπληροφορικής και τελικά η αναγωγή της από κράμα επιστημών σε ξεχωριστή επιστήμη δεν ήταν τυχαία. Υποκινήθηκε από την έρευνα αλλά και τις επιχειρηματικές δυνατότητες στη Βιολογία, την Ιατρική και την Μηχανική με την ευρύτερη έννοια και αποτέλεσε ένα είδος επιστημονικής επανάστασης καθώς η ύπαρξη των δεδομένων ,η δυνατότητα διαχείρισής τους και η ανάπτυξη της τεχνολογίας σε τομείς όπως η μελέτη του DNA με ακολουθιακές μεθόδους (DNA Sequencing) , έρχεται να λύσει τους περιορισμούς της έρευνας που βασιζόταν καθαρά στην παρατήρηση πειραμάτων.

Η στροφή αυτή των Βιοεπιστημών προς την Πληροφορική και την τεχνολογία, κατά την δεκαετία του '90, επέφερε κοινωνική αλλαγή. Οι βιομηχανίες πιέζουν για περισσότερα προγράμματα που ασχολούνται με τη Βιοπληροφορική, η ίδια η επιστημονική κοινότητα ενισχύει την ταυτόχρονη λειτουργία δραστηριοτήτων σε αυτόν τον τομέα, στην αγορά εργασίας παρατηρείται ζήτηση και οι χρηματοδότες ερευνητικών προγραμμάτων ζητούν επίμονα να τοποθετηθεί το μεγαλύτερο μέρος των διαθέσιμων χρημάτων σε προγράμματα βιοπληροφορικής με σκοπό την συντόμευση της ολοκλήρωσης τους.

Σήμερα, μας είναι σχεδόν αδύνατον να φανταστούμε πως 20 χρόνια πριν αποτελούσε πρόκληση η δημοσίευση εργασίας με υπολογιστική ανάλυση και επεξεργασία των βιολογικών πληροφοριών. Ειδικά οι Μοριακοί Βιολόγοι, οι οποίοι δεν ήταν συνηθισμένοι να εξάγουν αποτελέσματα με υπολογιστικές μεθόδους θεωρούσαν την πειραματική παρατήρηση ως τον μοναδικό τρόπο έρευνας. Τώρα πλέον , είναι σχεδόν αδιανόητο μία αξιοπρεπής δημοσίευση έρευνας να μην περιλαμβάνει έστω και ένα υπολογιστικό κομμάτι σε οποιοδήποτε επίπεδο, από τον σχεδιασμό ενός πειράματος μέχρι τον σχολιασμό των αποτελεσμάτων του. Είναι λοιπόν εμφανές πως επετεύχθη ένα μεγάλο

άλμα, καθώς η Βιολογία ανήχθη από μία καθαρά πειραματική σε υπολογιστική και πειραματική επιστήμη.

Ας δούμε όμως πως έχει η ιστορία.

Περίπου 60 χρόνια πριν, κάποια από τα βασικά αντικείμενα έρευνας της Μοριακής Βιολογίας παρουσίασαν αλγοριθμικά ζητήματα τα οποία περιελάμβαναν και την μελέτη της δομής του DNA όπως και την κωδικοποίηση της γενετικής πληροφορίας. Στη δεκαετία του '60 δημοσιεύεται ένα πλήθος εξαιρετικής σημασίας papers τα οποία αφορούν την εξέλιξη των γονιδίων και των πρωτεϊνών, τις δομικές ιδιότητες των πολυπεπτιδικών αλυσίδων, την πληροφορία που περιέχουν οι ακολουθίες DNA, τις βασικές αρχές του γενετικού κώδικα, την κατασκευή των φυλογενετικών δένδρων και μια πρωταρχική θεωρία περί της ταξινόμησης ακολουθιών νουκλεοτιδίων ή αμινοξέων.

Εν συνεχεία, κατά την δεκαετία του '70, η διαθεσιμότητα των δεδομένων σχετικά με τις πρωτεϊνικές ακολουθίες και δομές έδωσε την δυνατότητα, για παράδειγμα, να αναπτυχθούν μελέτες που αφορούσαν το ποσοστό των μεταλλάξεων και τον ρόλο των καταλυτών αντίστοιχα. Το πρόβλημα της ταξινόμησης ακολουθιών έρχεται δυναμικά πλέον στο προσκήνιο ενώ ταυτόχρονα στην επιστήμη των υπολογιστών μελετάται ο τρόπος σύγκρισης συμβολοσειρών. Η χρονική αυτή συγκυρία δεν είναι τίποτα άλλο από ένας άψογος συγχρονισμός δύο διαφορετικών δραστηριοτήτων που ουσιαστικά μελετούν τις 2 διαφορετικές όψεις του ίδιου νομίσματος. Οι θεωρητικά σκεφτόμενοι βιολόγοι βλέπουν τις μακρομοριακές ακολουθίες ως αλληλουχίες βάσεων ή αμινοξέων ενώ οι μηχανικοί υπολογιστών ως συμβολοσειρές. Οι δύο επιστήμες ενώνουν τις δυνάμεις τους σε ένα άλλο επίπεδο έρευνας. Δημιουργούνται αλγόριθμοι σύγκρισης μακρομοριακών ακολουθιών ή συμβολοσειρών και προς το τέλος της δεκαετίας παράγονται και τα πρώτα αποθέματα δεδομένων που περιλαμβάνουν πρωταρχικές δομές πρωτεϊνικών μορίων και συμβάλλουν στην δημιουργία των πρώτων βάσεων δεδομένων που ακολουθεί τα επόμενα χρόνια. Ακολουθεί ανάπτυξη στον τομέα των υπολογιστών τόσο στο hardware όσο και στο software η οποία καθιστά δυνατή την αποθήκευση, την διαχείριση και την ανάλυση δεδομένων που πηγάζουν από βιολογικά πειράματα και έχουν υποστεί υπολογιστική ανάλυση, αυτό αποτελεί και τον ορισμό της Βιοπληροφορικής με την στενή έννοια.

Αυτό ήταν. Είχαν πλέον τεθεί τα θεμέλια πάνω στα οποία κτίζεται γοργά το οικοδόμημα της Βιοπληροφορικής. Και πράγματι παρατηρείται εξαιρετική δραστηριότητα γύρω από τα πεδία της ανάλυσης ακολουθιών και αναζήτησης

ομοιοτήτων ([1], [2]) , πρόβλεψης δομών ([3], [4]), μοριακής εξέλιξης ([5], [6]) και σχετικά με τις βάσεις δεδομένων μοριακής Βιολογίας ([7], [8]).

Έτσι, μέχρι τις αρχές της δεκαετίας του 80, τα πρωταρχικά αντικείμενα έρευνας ήταν μακρομοριακές ακολουθίες (DNA, RNA ή πρωτεϊνικά μόρια) που αναλύονταν ως συμβολοσειρές ή παρόμοιες αναπαραστάσεις, και μακρομοριακές δομές που συχνά αναλύονταν με τη βοήθεια καρτεσιανών συντεταγμένων ή/και πιο περίπλοκων συστημάτων. Γενικότερα λοιπόν αντικείμενα μελέτης αποτελούν απλά μόρια, σύνολα μορίων και οι διασυνδέσεις τους προς αναζήτηση ομοιοτήτων σε διάφορα επίπεδα. Παραδείγματα τέτοιων ερευνών αποτελούν η αναγνώριση της περιοχής που αποτελεί την ειδοποιό διαφορά σε μια ομάδα ακολουθιών διαφορετικών ειδών ή η αναζήτηση του προτύπου αναδίπλωσης μέσα σ' ένα σύνολο παρομοίως αναδιπλωμένων πρωτεϊνικών δομών. Μπορεί να γίνει αντιληπτό πως οι δυσκολίες στην ερευνητική διαδικασία αυτή την εποχή είναι τεχνικής και όχι τόσο νοητικής φύσεως και προκύπτουν από την ανεπάρκεια τεχνικών μεθόδων. Ωστόσο μέσα στην επόμενη δεκαετία οι τεχνικές υψηλής απόδοσης που αναπτύσσονται στην βιολογία «σπάνε» αυτούς τους περιορισμούς και παρέχουν την δυνατότητα στους βιολόγους να αγγίξουν το ονειρό τους: Την προέλευση των αρχών που διέπουν την εξέλιξη και ανάπτυξη των έμβιων οργανισμών δια μέσω πολλαπλών κυττάρων, ιστών, οργάνων, ατόμων, πληθυσμών και ειδών.

Στη δεκαετία του '90 λοιπόν, ένα πλήθος από επιτεύγματα «κλειδιά» καθιέρωσαν την ταχεία αναζήτηση βάσεων δεδομένων με την χρήση του BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) [9] . Επίσης αναπτύσσονται τεχνικές για ακολουθιακές μεθόδους μεγάλης κλίμακας αλλά και για άλλου είδους πειραματικές μετρήσεις. Η υπολογιστική ανάλυση παύει πλέον να θεωρείται μπελάς για την έρευνα στην Βιολογία και κρίνεται πλέον απαραίτητη, καθώς το πλήθος των ακολουθιακών δεδομένων απαιτούν γρήγορες, ακριβείς και προσιτές μεθόδους για την ανάλυση αλλά και των σχολιασμό [10].

Ακολουθούν μεγάλες εξελίξεις με την χαρτογράφηση ολοκλήρου του ανθρώπινου γονιδιώματος. Η Βιολογία έχοντας πλέον τα εργαλεία στρέφεται και προς την μελέτη των πρωτεϊνών, της δομής και της λειτουργίας τους αλλά και σε όλους τους ερευνητικούς τομείς της παρατηρείται έντονη δραστηριότητα με τα «φτερά» που της έδωσε η Πληροφορική.

Μετά από όλα αυτά είναι πλέον ολοφάνερο πως η επίδραση στην ιατρική θα ήταν μνημειώδης. Εδώ η επεξεργασία πληροφοριών παίζει κεντρικό ρόλο στην προσομοίωση μοριακών διαδικασιών στα κύτταρα και η εξαγωγή συμπερασμάτων για την επίδραση των φαρμάκων στους ανθρώπους βασίζεται

στο γενετικό τους υπόβαθρο [11] . Πιο συγκεκριμένα, στην κλινική έρευνα (και για τον καρκίνο), εφαρμογές υπολογιστικών προσεγγίσεων περιελάμβαναν πρόγνωση βασισμένη στο γονιδίωμα [12] ,πρώιμη διάγνωση και ανακάλυψη βιοδεικτών για τον καρκίνο [13], ανακάλυψη φαρμακευτικής στόχευσης [14], θεραπευτική στόχευση [15], ανάπτυξη φαρμάκων [16], ανοχή στις κλινικές δοσολογίες φαρμάκων [17] καθώς και διαχείριση κλινικών δεδομένων [18]. Τέλος, αναπτύχθηκαν εξειδικευμένα μέσα όπως το πρόγραμμα ανατομίας γονιδιώματος καρκίνου (Cancer Genome Anatomy Project - CGAP) για την υποστήριξη της έρευνας κατά του καρκίνου [19]. Και γενικότερα, ένα πλήθος εργασιών στο πεδίο της Βιοπληροφορικής ειδικεύτηκαν στην έρευνα κατά της επάρατης νόσου. Παραδείγματα τέτοιων εργασιών είναι η υπολογιστική ανάλυση των επιγενετικών φαινομένων της βιολογίας του καρκίνου [20], όπως τα πρότυπα μεθυλίωσης του DNA και οι ιστονικές τροποποιήσεις [21], οι προσομοιώσεις του κυτταρικού κύκλου κατά την ανάπτυξη του νεοπλάσματος [22], η ανάπτυξη φορμαλισμών τέτοιων προσομοιώσεων [23], η ανάλυση πολυμορφισμών μονού νουκλεοτιδίου (SNPs) [24], η ανακάλυψη της στοχοποίησης του καρκίνου [25], η ανακάλυψη των βιοδεικτών για τον καρκίνο [26], καθώς και η εκμάθηση και η εκπαίδευση της βιοπληροφορικής [27]. Επιπλέον, αναφορές όπως η βιολογική βάση δεδομένων όγκων στα ποντίκια Mouse Tumor Biology Database [28] αποκτούν ολόένα και μεγαλύτερη σημασία, μαζί με τα πειράματα έκφρασης γονιδίων [29], την έρευνα για καινούριες θεραπευτικές μεθόδους [30], την τεχνολογία στοχευμένης πρωτεομικής [31] καθώς και με συνδέσεις με άλλες περιοχές έρευνας όπως είναι η ανοσολογία [32].

Τέλος, η σύνδεση της έρευνας του καρκίνου με τη συστημική βιολογία[33] είναι πολύ νωρίς ακόμη για να αξιολογηθεί επακριβώς. Ωστόσο, εμφανίζεται ως ιδιαίτερα ανοιχτή πηγή έρευνας στο άμεσο μέλλον.

Συμπερασματικά, η βιοπληροφορική αναμένεται να συνεχίσει αυτό το συναρπαστικό διαδραστικό παιχνίδι με το χώρο της γενομικής όσο αφορά στην έρευνα του καρκίνου και να δώσει ιδιαίτερες δυνατότητες στους τομείς της βιοπληροφορικής του καρκίνου και της ογκογενετικής όπως έχουν ονομασθεί [34].

1.3 Μελλοντικοί στόχοι και προκλήσεις

Παρακολουθώντας την μέχρι τώρα πορεία της Βιοπληροφορικής, την ανάπτυξη της, αλλά και την επανάσταση που έφερε στην έρευνα, γνωρίζοντας επίσης την σημερινή πραγματικότητα μπορούμε να συμπεράνουμε τους μελλοντικούς στόχους αλλά και τις προκλήσεις που θα αντιμετωπίσει η

Βιοπληροφορική. Ωστόσο το να μιλά κανείς για το μέλλον πάντοτε περιλαμβάνει ρίσκο και ίσως κάποια από τα συμπεράσματα μας να αποδειχθούν λανθασμένα.

Οι μελλοντικές επιδιώξεις εστιάζονται στο υπάρχον έλλειμμα υπολογιστικής ισχύος που δεν καλύπτει τη βιολογική πολυπλοκότητα. Είναι προφανές ότι η βιοπληροφορική αναπτύχθηκε σε στενή συνεργασία με άλλες τεχνολογικές προσπάθειες στις βιολογικές επιστήμες, κυρίως αυτές της παραγωγής και ανάλυσης δεδομένων. Στις πρώτες ημέρες, το κύριο επίκεντρο ήταν η ανάλυση μοριακών ακολουθιών και δομών, που ακολουθούνταν από την ανάλυση ολόκληρων γονιδίων και των μεταγραφικών προφίλ. Όσο όλο και περισσότερα είδη δεδομένων ανακαλύπτονται σε μεγάλες κλίμακες, κυτταρικά μέρη, παραλλαγές με βάση τον ιστό, ανατομικές ιδιότητες και πληθυσμιακή ποικιλομορφία, οι προκλήσεις για υπολογιστική ανάλυση σε πολλαπλά επίπεδα και σύνθετη ολοκλήρωση αυτών των αναλύσεων είναι μεγαλύτερες από ποτέ. Παρά την αύξηση του αριθμού των δεδομένων και των ταχύτερων υπολογιστών, η κυρίαρχη δύναμη προόδου, όπως και σε κάθε πνευματική δραστηριότητα εξάλλου, θα είναι η ύπαρξη νέων ιδεών και τρόπων σκέψης που θα επιτρέψουν την περαιτέρω ανάπτυξη του τομέα μέχρι και τη μελλοντική εγκαθίδρυσή του ως το βασιλιά των βιολογικών επιστημών.

Σε επιστημονικές κοινότητες «ακούγεται» πως το μέλλον θα προχωρήσει με τη δημιουργία λογισμικών που θα απεικονίζουν τη λειτουργία των πρωτεϊνών και την αποκαλούμενη «έκφραση των γονιδίων», κατασκευή υπολογιστών στους οποίους θα έχει αντικατασταθεί η μνήμη με ζωντανά νευρικά κύτταρα, κάτι που θα πλαισιωθεί με τη δημιουργία νέων ειδικοτήτων, όπως αυτή του βιολόγου-λογισμικού, του γιατρού-προγραμματιστή, του δικτυακού-οικονομολόγου ([35], [36], [37], [38]).

Στις προκλήσεις που θα έχει να αντιμετωπίσει η Βιοπληροφορική ανήκει η ανάλυση σε πραγματικό χρόνο και χωρίς να απαιτούνται chips σιλικόνης. Αυτό μπορεί να αποτελέσει πραγματικότητα με την ενίσχυση της Βιοπληροφορικής από την νανοτεχνολογία. Αυτή η τεχνολογία του μικρόκοσμου κατά πάσα πιθανότητα θα αποτελέσει τη βάση της βιοπληροφορικής στο μέλλον. Αντί λοιπόν να διαχωρίζουμε τις επιστημονικές μεθόδους μετρήσεων (π.χ ακολουθιακές) από τη ανάλυση (π.χ. υπολογισμοί με H/Y), πιθανολογείται ότι κάποιες από τις μετρητικές συσκευές θα έχουν ουσιώδεις υπολογιστικές δυνατότητες με τέτοιον τρόπο ώστε η ανάλυση του όγκου των βιολογικών δεδομένων να μπορεί να γίνει επιτόπου. Αυτού του είδους τα μηχανήματα θα απαιτούν δυνατότητες πολύπλοκης επικοινωνίας, ελέγχου και αισθητήρων, δυνατότητα προγραμματισμού σε πραγματικό χρόνο, υπερσύγχρονες διεπαφές και καταγιστικές λειτουργίες. Δεν χρειάζεται φοβερή φαντασία για να

συνειδητοποιήσει κανείς την τεράστια προοπτική εκμετάλλευσης ενός τέτοιου τύπου εφαρμογών στον τομέα της βιολογικής ([39], [40]) και της βιοϊατρικής έρευνας ([41], [42]). Αυτές οι καινοτομίες θα αλλάξουν ριζικά τις αντιλήψεις μας γύρω από τη διαγνωστική και προληπτική ιατρική [44].

1.4 Προβληματισμοί

Υστερα από ολ' αυτά στεκόμαστε με θαυμασμό μπροστά στην Βιοπληροφορική που αποτελεί όπλο στα χέρια του ανθρώπου. Όπως όμως το μαχαίρι μπορεί να χρησιμοποιηθεί ως εργαλείο αλλά και ως φονικό όπλο ανάλογα θέματα και προβληματισμοί προβάλλουν και γύρω από την χρήση της βιοπληροφορικής. Για παράδειγμα αναφέρουμε πως ίσως δε είναι μακριά η δυνατότητα σύνδεσης ενός υπολογιστή με το άρρωστο μυαλό για να διορθώνεται η σκέψη. Τι όμως θα μπορούσε να προκαλέσει ο έλεγχος της σκέψης σε λάθος χέρια; [35]

Δυστυχώς η έκρηξη της τεχνολογίας συμπίπτει με μια βαθιά κρίση αξιών, κοινωνιών, αυτού του ίδιου του ανθρώπου. Αποδεικνυόμαστε ανέτοιμοι να σηκώσουμε το βάρος των επιτευγμάτων της χωρίς κανένα κόστος. Αν οι πνευματικές αξίες λειτουργούν ως ισχυρή παράμετρος στον αριθμητή, η τεχνολογία μπορεί ελεύθερα να κάνει την δουλειά της στον παρονομαστή του ανθρώπινου κλάσματος [45]. Όπως λέει και ο περίφημος ηθικολόγος Willard Gaylin, « η απώλεια των ανθρώπινων αξιών θα μας κάνει όλους μικρότερους με τίμημα απρόβλεπτο και τρόπους άγνωστους σε καιρούς απροσδιόριστους» [46].

Καθώς προσπαθούμε να μελετήσουμε την επίδραση της σύγχρονης βιοϊατρικής τεχνολογίας στο ανθρώπινο σώμα και την αναφορά της σε αυτό που ονομάζεται άνθρωπος, ψυχή και σώμα, φθαρτή και άφθαρτη υπόσταση μαζί, είναι πολύ φυσικό να ψάχνουμε για κάποια όρια· μέχρι ποιου δηλαδή σημείου μπορεί να παρέμβει η τεχνολογία στον άνθρωπο. Τι είναι λογικά και ηθικά θεμιτό και τι απαγορεύεται. Πριν όμως από τα όρια θα έπρεπε να μελετήσουμε τις αρχές που πρέπει να διέπουν τη σύγχρονη έρευνα. Και τούτο, διότι τα όρια γενικά στο επίπεδο των αξιών δεν είναι πάντοτε σαφή· ούτε πάλι είναι σωστό να σκέπτεται κανείς τη λειτουργία των θεσμών και των επιστημών με βάση τους φραγμούς και τα όριά τους, αλλά με γνώμονα τις αρχές και την ελευθερία τους. Το μυστικό της βιοηθικής προβληματικής δεν βρίσκεται στο τι επιτρέπεται και τι απαγορεύεται, αλλά στο γιατί και πως ενεργούμε. Οι αρχές προσδιορίζουν και τις ορθές κατευθύνσεις και τα αναγκαία όρια.

Ενώ πολλά ακούγονται περί απειλών, κινδύνων, αποκαλυπτικών συνεπειών και καταστροφών, αίσθησή μας είναι ότι η εμφάνιση και πρόοδος της γενετικής, της βιοτεχνολογίας και γενικά της ιατρικής τεχνολογίας μπορεί να

αποδειχθεί περισσότερο ευλογία παρά εφιάλτης. Η βιοϊατρική πρόκληση δεν οδηγεί μόνον σε εμφάνιση καινοφανών κοινωνικών προβλημάτων και γέννηση φόβων διλημάτων ή αδιεξόδων.

Παράλληλα, αφ' ενός μεν το μέγεθος των βιοϊατρικών δυνατοτήτων μας, αφ' ετέρου δε η φτώχεια των αρχών, η κρίση των αξιών και ο γενικότερος αποπροσανατολισμός των συγχρόνων κοινωνιών δικαιολογούν την αναγκαιότητα προσοχής, συνέσεως και σαφών δεοντολογικών διατυπώσεων, οι οποίες όμως βασίζονται σε καλλιεργημένες αξίες και όχι σε νοσηρούς φόβους [47].

Βιβλιογραφία

- [1] Gingeras, T. R. and Roberts, R. J. Steps toward computer analysis of nucleotide sequences. *Science* 209 ,1322–8 (1980).
- [2] Wilbur, W. J. and Lipman, D. J. Rapid similarity searches of nucleic acid and protein data banks. *Proc Nat Acad Sci U S A* 80 , 726–30 (1983)
- [3] Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv Protein Chem* 34 , 167–339 (1981).
- [4] Kabsch, W. and Sander, C. On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc Natl Acad Sci U S A* 81 , 1075–8 (1984)
- [5] Doolittle, R. F. *Similar amino acid sequences: chance or common ancestry?* *Science* 214 , 149–59 (1981).
- [6] Bajaj, M. and Blundell, T. Evolution and the tertiary structure of proteins. *Annu Rev Biophys Bioeng* 13 , 453–92 (1984).
- [7] Philipson, L. The DNA data libraries. *Nature* 332 , 676 (1988)
- [8] Bernstein, F. C. et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 112 , 535–42 (1977)
- [9] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. Basic local alignment search tool. *J Mol Biol* 215 , 403–10 (1990)
- [10] Andrade, M. A. and Sander, C. Bioinformatics: from genome data to biological knowledge. *Curr Opin Biotechnol* 8 , 675–83 (1997)
- [11] Sander, C. Genomic medicine and the future of health care. *Science* 287 , 1977–8 (2000)
- [12] Kallioniemi, O. P. Biochip technologies in cancer research. *Ann Med* 33 , 142–7 (2001)
- [13] Hondermarck, H. et al. Proteomics of breast cancer for marker discovery and signal pathway profiling. *Proteomics* 1 , 1216–32 (2001).
- [14] Pavelic, K. and Gall-Troselj, K. Recent advances in molecular genetics of breast cancer. *J Mol Med* 79, 566–73 (2001).
- [15] Schultze, J. L. and Vonderheide, R. H. From cancer genomics to cancer immunotherapy: toward secondgeneration tumor antigens. *Trends Immunol* 22 , 516–23 (2001).
- [16] Basik, M., Mousses, S. and Trent, J. Integration of genomic technologies for accelerated cancer drug development. *Biotechniques* 35 , 580–2 , 584, 586 passim (2003)
- [17] Damaraju, S., Sawyer, M. and Zanke, B. Genomic approaches to clinical drug resistance. *Cancer Treat Res* 112 , 347–72 (2002).

- [18] Haque, S., Mital, D. and Srinivasan, S. Advances in biomedical informatics for the management of cancer. *Ann N Y Acad Sci* 980 , 287–97 (2002).
- [19] Strausberg, R. L., Greenhut, S. F., Grouse, L. H., Schaefer, C. F. and Buetow, K. H. In silico analysis of cancer through the Cancer Genome Anatomy Project. *Trends Cell Biol* 11, S66–71 (2001).
- [20] Yang, H. H. and Lee, M. P. Application of bioinformatics in cancer epigenetics. *Ann N Y Acad Sci* 1020 , 67–76 (2004)
- [21] Lee, M. P. Genome-wide analysis of epigenetics in cancer. *Ann N Y Acad Sci* 983 , 101–9 (2003)
- [22] Alberghina, L., Chiaradonna, F. and Vanoni, M. Systems biology and the molecular circuits of cancer. *Chembiochem* 5 , 1322–33 (2004).
- [23] Christopher, R. et al. Data-driven computer simulation of human cancer cell. *Ann N Y Acad Sci* 1020, 132–53 (2004).
- [24] Clifford, R. J. et al. Bioinformatics tools for single nucleotide polymorphism discovery and analysis. *Ann N Y Acad Sci* 1020 , 101–9 (2004)
- [25] Desany, B. and Zhang, Z. Bioinformatics and cancer target discovery. *Drug Discov Today* 9 , 795–802 (2004).
- [26] Rhodes, D. R. and Chinnaiyan, A. M. Bioinformatics strategies for translating genomewide expression analyses into clinically useful cancer markers. *Ann N Y Acad Sci* 1020 , 32–40 (2004)
- [27] Umar, A. Applications of bioinformatics in cancer detection: a lexicon of bioinformatics terms. *Ann N Y Acad Sci* 1020 , 263–76 (2004).
- [28] Krupke, D. et al. The Mouse Tumor Biology Database: integrated access to mouse cancer biology data. *Exp Lung Res* 31 , 259–70 (2005)
- [29] Rhodes, D. R. and Chinnaiyan, A. M. Integrative analysis of the cancer transcriptome. *Nat Genet* 37 Suppl,S31–7 (2005)
- [30] Mount, D. W. and Pandey, R. Using bioinformatics and genome analysis for new therapeutic interventions. *Mol Cancer Ther* 4 , 1636–43 (2005).
- [31] Posadas, E. M., Simpkins, F., Liotta, L. A., MacDonald, C. and Kohn, E. C. Proteomic analysis for the early detection and rational treatment of cancer-realistic hope? *Ann Oncol* 16 , 16–22 (2005).
- [32] Strausberg, R. L. Tumor microenvironments, the immune system and cancer survival. *Genome Biol* 6 , 211(2005).
- [33] Khalil, I. G. and Hill, C. Systems biology for cancer. *Curr Opin Oncol* 17 , 44–8 (2005).
- [34] Strausberg, R. L., Simpson, A. J., Old, L. J. and Riggins, G. J. Oncogenomics and the development of new cancer therapies. *Nature* 429 , 469–74 (2004).
- [35] Αλαχιώτης Σ.Ν. Σημειώσεις για τη Βιοηθική και σχετικό με τη Βιοηθική πληροφοριακό υλικό ενημερωτικού χαρακτήρα (2003)
- [36] Krane D. and Raymer M.L., *Fundamental Concepts of Bioinformatics, Pearson Education.*(2003)
- [37] de Silva E, Stumpf MP Complex networks and simple models in biology. *J R Soc Interface*. Dec 22; 2(5):419-30. Review(2006)
- [38] Perez-Iratxeta C, Andrade-Navarro MA, Wren JD. Evolving research trends in bioinformatics. *Brief Bioinform*. Oct 31; (2006)
- [39] O'Brien, T. P. et al. Genome function and nuclear architecture: from gene expression to nanoscience. *Genome Res* 13 , 1029–41 (2003)
- [40] Montemagno, C. D. Integrative technology for the twenty-first century. *Ann N Y Acad Sci* 1013 , 38–49 (2004)

- [41] Cavalcanti, A. Assembly automation with evolutionary nanorobots and sensor-based control applied to nanomedicine. *IEEE Trans Nanotech* 2 , 82–87 (2003)
- [42] Bogunia-Kubik, K. and Sugisaka, M. From molecular biology to nanotechnology and nanomedicine *Biosystems* 65 , 123–38 (2002)
- [43] Segal, E., Friedman, N., Kaminski, N., Regev, A. and Koller, D. From signatures to models: understanding cancer using microarrays. *Nat Genet* 37 Suppl, S38–45 (2005)
- [44] Hood, L., Heath, J. R., Phelps, M. E. and Lin, B. Systems biology and new technologies enable predictive and preventative medicine. *Science* 306 , 640–3 (2004).
- [45] Νικολάου Χατζηνικολάου, Μητροπολίτου Μεσογαίας & Λαυρεωτικής Εναρκτήρια Ομιλία στο Ε΄ Πανελλήνιο Συνέδριο Αγγειολογίας- Αγγειοχειρουργικής. Αθήνα, 23 Ιανουαρίου (1998)
- [46] Willard Gaylin, “Harvesting the Dead: The Potential for Recycling Human Bodies,” *HARPERS* (September 1974), σ. 30.
- [47] Νικολάου Χατζηνικολάου, Μητροπολίτου Μεσογαίας & Λαυρεωτικής, προέδρου της Επιτροπής Βιοηθικής της Ιεράς Συνόδου της Εκκλησίας της Ελλάδος. «Αρχές και αξίες για την οικοδόμηση της Ευρώπης», Μάιος 2003.

2

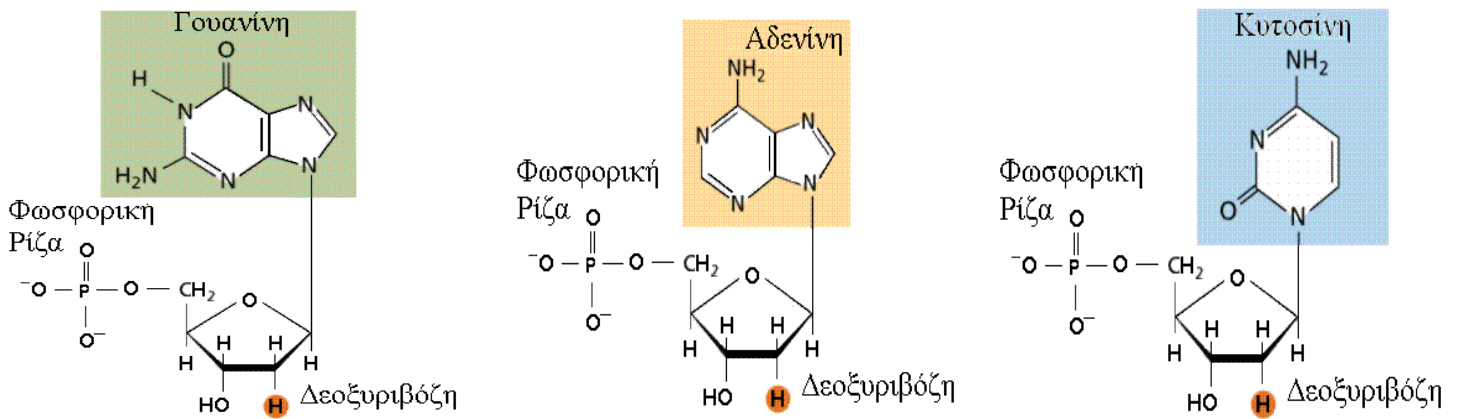
Στοιχεία Βιολογίας

2.1 Εισαγωγή στο γονιδίωμα

Παρ' όλο που το DNA εντοπίστηκε στον πυρήνα του κυττάρου σπέρματος ψαριών το 1869 από τον Friedrich Meischer, έως και το 1944 δεν ήταν γνωστό ότι αποτελεί το γενετικό υλικό των οργανισμών. Οι επιστήμονες από τις αρχές του 1900 έως και τη λήξη του Δευτέρου Παγκοσμίου Πολέμου – τη «Χρυσή Εποχή» της Γενετικής όπως είχε τότε χαρακτηριστεί- πίστευαν ότι τα μόρια που μεταφέρουν τη γενετική πληροφορία είναι οι **πρωτεΐνες** που παρουσιάζουν μεγαλύτερη ποικιλομορφία, επειδή προκύπτουν ως το αποτέλεσμα του συνδυασμού είκοσι διαφορετικών αμινοξέων, ενώ τα νουκλεϊκά οξέα είναι συνδυασμός τεσσάρων μόνο νουκλεοτιδίων.

Κατά τη διάρκεια της δεκαετίας του 1920, ο βιοχημικός P.A. Levene ανέλυσε τα συστατικά του μορίου του DNA. Οι αναλύσεις του έδειξαν ότι το μόριο αυτό αποτελείται από τέσσερις νιτρώδεις βάσεις: τη κυτοσίνη, τη θυμίνη, την αδενίνη και τη γουανίνη και μια φωσφορική ομάδα. Συμπέρανε ότι η βασική μονάδα (το νουκλεοτίδιο) αποτελείται από μια αζωτούχα βάση που συνδέεται με τη δεοξυριβόζη στην οποία συνδέεται και η φωσφορική ρίζα.

Ωστόσο, συμπέρανε λανθασμένα ότι οι αναλογίες των βάσεων ήταν ίδιες και για τις τέσσερις και ότι υπήρχε ένα τετρανουκλεοτίδιο το οποίο ήταν το επαναλαμβανόμενο δομικό στοιχείο του μορίου. Παρ' όλα αυτά, το νουκλεοτίδιο και σήμερα θεωρείται ως το μονομερές δομικό στοιχείο του πολυμερούς νουκλεϊκού οξέος αλλά η απόκλιση από τις τότε πεποιθήσεις γίνεται στον αριθμό των (τεσσάρων) νουκλεοτιδίων, αυτών που δημιουργούνται με κυτοσίνη, εκείνων με γουανίνη, των τρίτων με αδενίνη και των τέταρτων με θυμίνη (εικόνα 2.1).



Εικόνα 2.1: Τα 3 από τα 4 διαφορετικά νουκλεοτίδια του DNA.

Στις άγουρες δύο πρώτες δεκαετίες του 20^{ου} αιώνα, η μελέτη της γενετικής ξεκίνησε με τη διάθεση διασύνδεσης της εργασίας του Mendel και εκείνης των κυτταρικών βιολόγων, κάτι που οδήγησε στη χρωματοσωματική θεωρία της κληρονομικότητας. Ο Garrod πρότεινε τη σύνδεση μεταξύ γονιδίων και «εσωγενών λαθών του μεταβολισμού». Και η αναπόφευκτη ερώτηση τελικά ετέθη: **Τί είναι γονίδιο;** Η απάντηση ήρθε μέσα από τις προσπάθειες των επιστημόνων να φράξουν τις θανατηφόρες συνέπειες μιας μεταδοτικής ασθένειας: της πνευμονίας.

Το 1928, ο Griffith χρησιμοποίησε 2 στελέχη του βακτηρίου πνευμονιόκοκος (*Diplococcus pneumoniae*) τα οποία ξεχωρίζουν μορφολογικά όταν καλλιεργηθούν σε θρεπτικό υλικό λόγω της παρουσίας ή μη ενός προστατευτικού καλύμματος. Το στέλεχος που είχε κάλυμμα σχημάτιζε λείες αποικίες¹ και ήταν παθογόνο, δηλαδή σκότωνε τα ποντίκια που μόλυνε, ενώ εκείνο που δεν είχε κάλυμμα σχημάτιζε αδρές αποικίες και δεν ήταν παθογόνο.

Ο Griffith χρησιμοποίησε υψηλή θερμοκρασία, για να σκοτώσει τα λεία βακτήρια και με αυτά μόλυνε ποντικούς οι οποίοι παρέμειναν ζωντανοί. Όταν όμως ανέμειξε νεκρά λεία βακτήρια με ζωντανά αδρά και με το μείγμα μόλυνε ποντικούς, τότε αυτοί πέθαναν. Στο αίμα των νεκρών ποντικών βρέθηκαν ζωντανά λεία βακτήρια. Ο Griffith συμπέρανε ότι μερικά αδρά βακτήρια «μετασηματίστηκαν» σε λεία παθογόνα ύστερα από αλληλεπίδραση με τα νεκρά λεία βακτήρια, αλλά δεν μπόρεσε να δώσει ικανοποιητική απάντηση για το πως γίνεται αυτό.

¹ Αποικία είναι ένα σύνολο από μικροοργανισμούς που έχουν προέλθει από διαδοχικές διαιρέσεις ενός κυττάρου, όταν αυτό αναπτύσσεται σε στερεό θρεπτικό υλικό.

1869	Απομονώνεται DNA από τον κυτταρικό πυρήνα
1903	Αποδεικνύεται ότι τα χρωμοσώματα είναι φορείς του κληρονομικού (γενετικού) υλικού
1944	Αποδεικνύεται ότι το DNA είναι το γενετικό υλικό
1952	Οι Hsu και Pomerat χρησιμοποιούν υποτονικό διάλυμα & το αλκαλοειδές κολχικίνη, που σταματά τη διαίρεση στην μετάφαση, για την ανάλυση των χρωματοσωμάτων.
1952	Οι Hershey και Chase εκτελούν σειρά πειραμάτων για την εξακρίβωση του αν οι πρωτεΐνες ή το DNA είναι το υλικό που κληροδοτείται.
1953	Ανακαλύπτεται η δομή της διπλής έλικας του DNA
1960	Οι Nowel , Moorehead , Hungerford επιτυγχάνουν να καλλιεργήσουν λεμφοκύτταρα με χρήση φυτοαιμαγλουτινίνης στο εργαστήριο
1960	Πραγματοποιείται στο Denver διάσκεψη και καθορίζεται το σύστημα ονοματολογίας των ανθρωπίνων μεταφασικών χρωματοσωμάτων.
1970	Ο Caspersson χρησιμοποιώντας τη χρωστική κιν ακρίνη, επιτυγχάνει τη δημιουργία ζωνών στα χρωματοσώματα (ζώνες – Q)
1971	Στη διάσκεψη που πραγματοποιήθηκε στο Παρίσι καθορίζεται το πρότυπο των ζωνών των ανθρωπίνων μεταφασικών χρωμοσωμάτων

Πίνακας 2.1: Χρονικά σημεία – κόμβοι στην ανακάλυψη του γενετικού υλικού

Η απάντηση δόθηκε 16 χρόνια μετά, το 1944 από τους Avery, Mac-Cleod και McCarthy οι οποίοι επανέλαβαν τα πειράματα του Griffith in vitro. Οι ερευνητές διαχώρισαν τα συστατικά των νεκρών λείων βακτηρίων σε υδατάνθρακες, πρωτεΐνες, λιπίδια, RNA, DNA κλπ. Και έλεγξαν ποιο από αυτά είχε την ικανότητα μετασχηματισμού. Διαπίστωσαν ότι το συστατικό που προκαλούσε το μετασχηματισμό των αδρών βακτηρίων σε λεία ήταν το DNA.

Την ίδια εποχή υπήρχαν πολλά βιοχημικά δεδομένα που υποστήριζαν ότι το DNA είναι το γενετικό υλικό.

Η οριστική επιβεβαίωση ότι το DNA είναι το γενετικό υλικό ήρθε το 1952 με τα πειράματα των Hersey και Chase οι οποίοι μελέτησαν τον κύκλο ζωής του βακτηριοφάγου (φάγου) T2.

Οι ερευνητές ιχνηθέτησαν² τους φάγους με ραδιενεργό ³²S που ενσωματώνεται μόνο στις πρωτεΐνες αλλά όχι στο DNA και με ραδιενεργό ³²P που ενσωματώνεται μόνο στο DNA και όχι στις πρωτεΐνες. Στη συνέχεια, με ραδιενεργούς φάγους μόλυναν τα βακτήρια. Τα αποτελέσματα έδειξαν ότι μόνο το DNA εισέρχεται στα βακτηριακά κύτταρα και είναι ικανό «να δώσει τις απαραίτητες εντολές» για να παραχθούν και να πολλαπλασιαστούν οι φάγοι.

Οργανισμός	Συνολικό DNA (σε ζεύγη βάσεων)	Αριθμός Χρωμοσωμάτων (απλοειδές κύτταρο)
Zea Mays (καλαμπόκι)	5,000,000,000	10
Homo Sapiens	3,000,000,000	23
Sacharomyces cerevisiae	140,000,000	16
Drosophila melanogaster	16,000,000	4
Escherichia coli	4,000,000	1

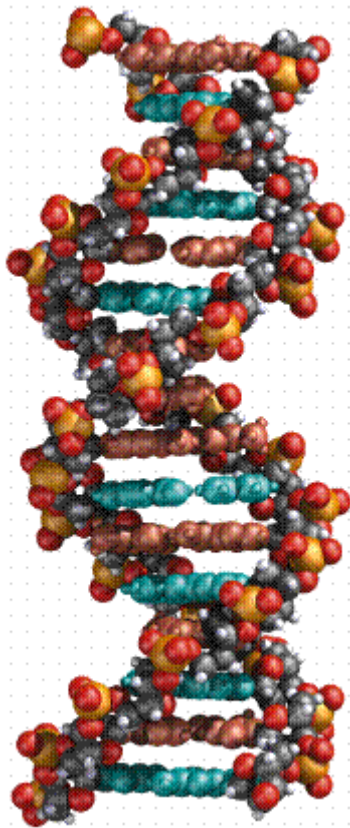
Πίνακας 2.2 : Μέγεθος Γονιδιώματος Διαφόρων Οργανισμών

Η ποσότητα του DNA σε κάθε οργανισμό είναι σταθερή, δεν μεταβάλλεται από αλλαγές στο περιβάλλον και είναι ανάλογη της πολυπλοκότητάς του. Η ποσότητα του DNA είναι επίσης ίδια σε όλα τα είδη κυττάρων ενός οργανισμού δηλαδή στην περίπτωση του ανθρώπου υπάρχει η ίδια ποσότητα DNA στα κύτταρα του σπλήνα, της καρδιάς, του ήπατος κλπ. Μοναδική εξαίρεση, οι γαμέτες των ανώτερων οργανισμών οι οποίοι περιέχουν τη μισή ποσότητα DNA από τα σωματικά κύτταρα.

2.2 Η Δομή του DNA

Ο Erwin Chargaff ανέλυσε τις νιτρώδεις βάσεις σε διάφορες μορφές ζώης και συμπέρανε ότι οι προτάσεις του Levene σχετικά με την ποσόστωση των βάσεων δεν ήταν σωστές. Το DNA έχει πλέον αποδειχτεί ως το γενετικό υλικό από το πείραμα των Hershey-Chase, ωστόσο ο τρόπος με τον οποίο συνδεότανε το DNA και τα γονίδια δεν ήταν ακόμη ξεκάθαρος. Έχει ήδη αποσαφηνιστεί ότι

² Ιχνηθέτηση: Σήμανση χημικών μορίων με τη χρήση ραδιενεργών ισοτόπων, φθοριζουσών ουσιών κτλ. Τυπικό παράδειγμα είναι η χρήση ραδιενεργού φωσφόρου ³²P στα νουκλεοτίδια για την ιχνηθέτηση του DNA



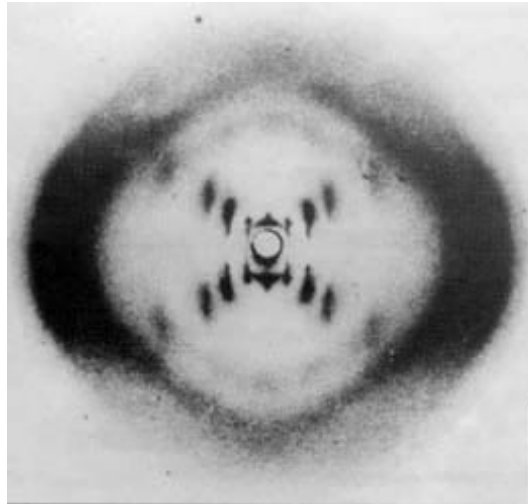
Εικόνα 2.2 : Η διπλή έλικα του DNA

το DNA πρέπει να μεταφέρει πληροφορία από το πατρικό κύτταρο προς το θυγατρικό. Ακόμη, ήταν σαφές ότι πρέπει να περιέχει κάποια πληροφορία για τον αυτοδιπλασιασμό του, πρέπει να είναι χημικώς σταθερό και σχετικά μη μεταβαλλόμενο. Βέβαια, είχε ήδη προσεγγιστεί η έννοια της μεταβολικής αλλαγής καθώς είχε αποδειχτεί ότι χωρίς τις μεταλλάξεις δεν μπορεί να υπάρξει η πρόοδος της εξέλιξης.

Πλήθος επιστημόνων έδειξαν άμεσο ενδιαφέρον για τον καθορισμό της δομής του DNA, μεταξύ των οποίων ήταν οι Francis Crick, James Watson, Rosalind Franklin, και Maurice Wilkens. Οι Watson και Crick συγκέντρωσαν όλα τα διαθέσιμα δεδομένα εκείνης της εποχής σε μια προσπάθεια ανάπτυξης ενός μοντέλου της δομής του DNA. Ο Franklin πήρε φωτομικρογραφίες με τη βοήθεια ακτίνων X υπό διάθλαση (X-ray diffraction) ενός κρυσταλλικού DNA. Τα δεδομένα εκείνης της

εποχής θέλανε το DNA να αποτελείται από ένα μακρύ μόριο και τις πρωτεΐνες να περιελίσσονται ελικοειδώς στο χώρο (όπως ανέφερε ο Linus Pauling).

Το DNA είναι μια διπλή έλικα με βάσεις στο κέντρο της (όπως τα σκαλιά σε μια σκάλα) και σακχαρο - φωσφορικές μονάδες κατά μήκος των πλευρών της έλικας. Τα δύο παράλληλα μέρη της «σκάλας» είναι συμπληρωματικά (κάτι που διαπιστώθηκε από τους Watson και Crick μετά από παρατήρηση των δεδομένων του Chargaff), η αδενίνη (A) ζευγαρώνει αποκλειστικά με θυμίνη (T) και η κυτοσίνη (C) ζευγαρώνει αποκλειστικά με τη γουανίνη (G), τα οποία ζεύγη κρατούνται δεμένα με δεσμούς υδρογόνου. Κάτι ιδιαίτερα σημαντικό για τη συνέχεια είναι η διαπίστωση ότι αν γνωρίζουμε την ακολουθία βάσεων του ενός μέρους της διπλής έλικας μπορούμε να υπολογίσουμε το συμπλήρωμά της και άρα την ακολουθία βάσεων του άλλου μέρους της έλικας.



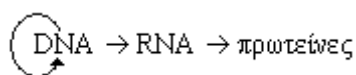
Εικόνα 2.3: Φωτογραφία με χρήση διαθλαστικών ακτίνων -X της διπλής έλικας του DNA

2.3 Η ροή της γενετικής πληροφορίας

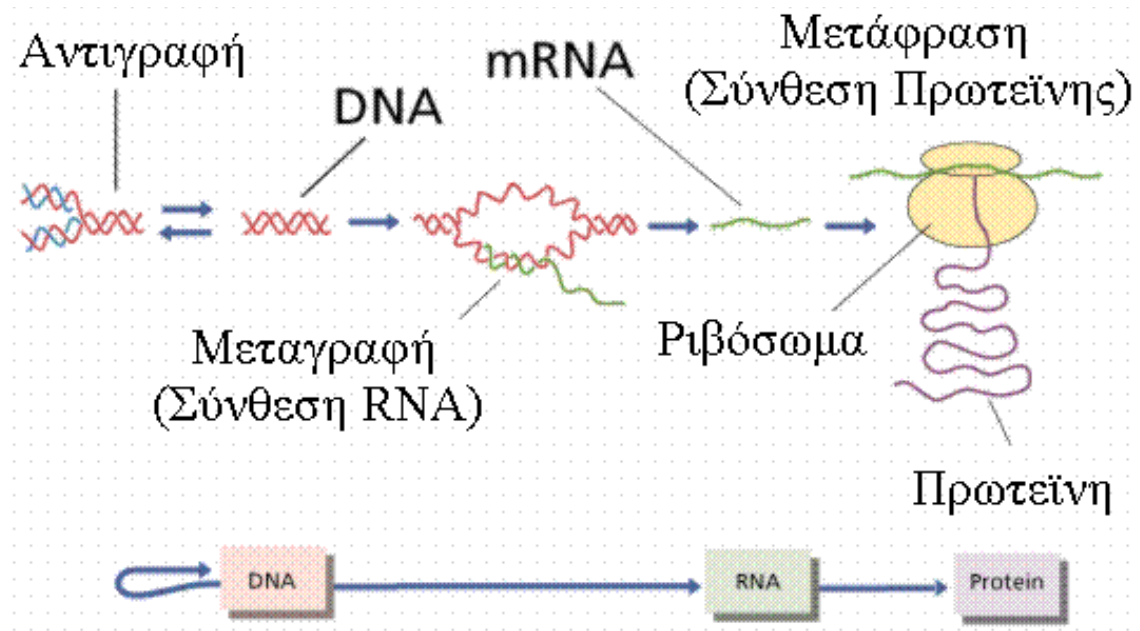
Το DNA ενός οργανισμού είναι ο **μοριακός σκληρός του δίσκος** που περιέχει αποθηκευμένες ακριβείς οδηγίες οι οποίες καθορίζουν τη δομή και τη λειτουργία του οργανισμού. Ταυτόχρονα περιέχει την πληροφορία για τον αυτοδιπλασιασμό του, εξασφαλίζοντας έτσι τη μεταβίβαση των γενετικών οδηγιών από ένα κύτταρο στα θυγατρικά του και από έναν οργανισμό στους απόγονούς του.

Το πρώτο βήμα για την έκφραση της πληροφορίας που υπάρχει στο DNA είναι η μεταφορά της στο RNA με τη διαδικασία της **μεταγραφής**. Το RNA μεταφέρει με τη σειρά του με τη διαδικασία της **μετάφρασης** την πληροφορία στις πρωτεΐνες που είναι υπεύθυνες για τη δομή και τη λειτουργία των κυττάρων και κατ' επέκταση των οργανισμών.

Η σχέση αυτή καθορίζεται στο ακόλουθο σχήμα όπου τα βέλη δείχνουν την κατεύθυνση της μεταφοράς της γενετικής πληροφορίας



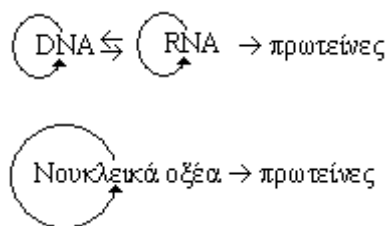
Το σχήμα αυτό αποτελεί το **κεντρικό δόγμα της μοριακής βιολογίας** όπως ονομάστηκε από τον **F.Crick** (1958). Η γενετική πληροφορία είναι η καθορισμένη σειρά των βάσεων, όπως η πληροφορία μιας γραπτής φράσης είναι η σειρά των γραμμάτων που την αποτελούν. Η πληροφορία υπάρχει σε τμήματα του DNA με συγκεκριμένη ακολουθία, τα γονίδια. Αυτά, δια μέσου της μεταγραφής και της μετάφρασης, καθορίζουν τη σειρά των αμινοξέων στην



Εικόνα 2.4: Το κεντρικό δόγμα της μοριακής βιολογίας

πρωτεΐνη (Εικόνα 2.4). Οι πορείες της μεταγραφής και της μετάφρασης των γονιδίων αποτελούν τη **γονιδιακή έκφραση**.

Για αρκετό καιρό οι ερευνητές πίστευαν ότι όλη η ροή της γενετικής πληροφορίας γινόταν προς τη μία μόνο κατεύθυνση, δηλαδή ότι το DNA μεταγράφονταν σε RNA. Σήμερα είναι γνωστό ότι ορισμένοι ιοί έχουν RNA ως γενετικό υλικό. Ένα ένζυμο που υπάρχει στους ίδιους τους ιούς, η αντίστροφη μεταγραφάση χρησιμοποιεί ως καλούπι το RNA για να συνθέσει το DNA. Επιπλέον, σε ορισμένους ιούς το RNA έχει την ικανότητα να αυτοδιπλασιάζεται.



Συνοψίζοντας, διαπιστώνουμε ότι η **αντιγραφή** του DNA διακονίζει τη γενετική πληροφορία, ενώ η **μετάφραση** χρησιμοποιεί αυτή την πληροφορία για να κατασκευάσει ένα πολυπεπτίδιο. Η **μεταγραφή** καθορίζει ποια γονίδια θα εκφραστούν, σε ποιους ιστούς (στους πολυκυτταρικούς ευκαρυωτικούς οργανισμούς) και σε ποια στάδια της ανάπτυξης.

Το απλοειδές ανθρώπινο γονιδίωμα³ έχει μήκος 3×10^9 ζεύγη βάσεων όπως φαίνεται στον Πίνακα 2.2. Όλα τα κύτταρα ενός πολυκυτταρικού οργανισμού έχουν το ίδιο DNA. Το κάθε ανθρώπινο κύτταρο περιέχει σύμφωνα με τις τελευταίες εκτιμήσεις όχι περισσότερα από 40,000 γονίδια.

Τα γονίδια διακρίνονται σε δύο κατηγορίες.

i. Στα γονίδια που μεταφέρονται σε mRNA και μεταφράζονται στη συνέχεια σε πρωτεΐνες

ii. Στα γονίδια που μεταγράφονται και παράγουν tRNA, rRNA και snRNA.

Σε κάθε ομάδα κυττάρων όμως εκφράζονται διαφορετικά γονίδια. Στα πρόδρομα ερυθροκύτταρα για παράδειγμα, εκφράζονται κυρίως τα γονίδια των αιμοσφαιρινών, ενώ στα Β-λεμφοκύτταρα τα γονίδια των αντισωμάτων.

Υπάρχουν **τέσσερα είδη μορίων RNA** που παράγονται με τη μεταγραφή: το αγγελιοφόρο RNA (mRNA), το μεταφορικό RNA (tRNA), το ριβοσωμικό RNA (rRNA) και το μικρό πυρηνικό RNA (snRNA). Τα τρία πρώτα είδη υπάρχουν και στους προκαρυωτικούς και στους ευκαρυωτικούς οργανισμούς, αλλά το τέταρτο υπάρχει μόνο στους ευκαρυωτικούς.

Το **mRNA** (αγγελιοφόρο) αποτελεί το λιγότερο από το 5% του ολικού RNA του κυττάρου. Τα περισσότερα mRNA έχουν μήκος 500-2000 βάσεις.

Το **rRNA** (ριβοσωμικό) είναι το πιο άθρονο από τα είδη RNA και αποτελεί το 80% περίπου του συνολικού RNA του κυττάρου. Έχει μέγεθος από 100-3000 βάσεις.

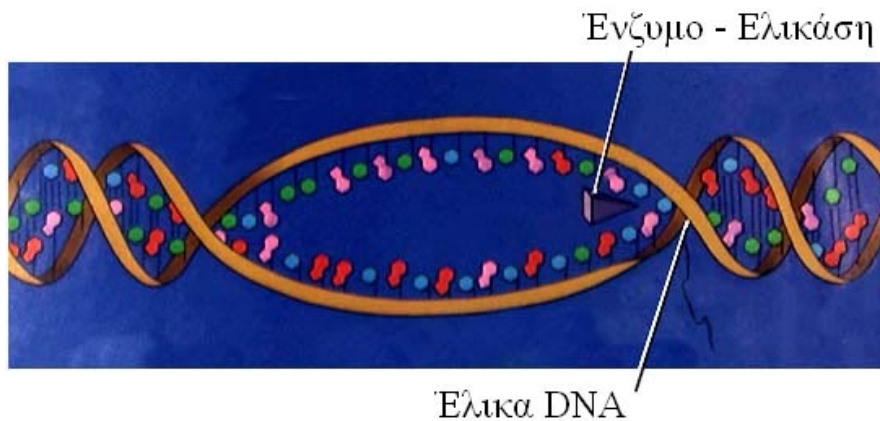
Το **tRNA** (μεταφορικό) αποτελεί το 15% περίπου του συνολικού RNA στο κύτταρο.

Το **snRNA** (μικρό πυρηνικό) είναι μικρά μόρια RNA τα οποία συνδέονται με πρωτεΐνες και σχηματίζουν μικρά ριβονουκλεοπρωτεϊνικά σωματίδια. Τα σωματίδια αυτά καταλύουν την ωρίμανση του mRNA (διαδικασία μόνο στους ευκαρυωτικούς οργανισμούς).

2.4 Το ένζυμο - ελικάση

Η αντιγραφή του DNA αρχίζει από καθορισμένα σημεία που ονομάζονται **θέσεις έναρξης αντιγραφής**. Για να ξεκινήσει η αντιγραφή του

³ **Γονιδίωμα:** το σύνολο του γενετικού υλικού ενός κυττάρου. Συνήθως αναφέρεται στο γενετικό υλικό ενός κυττάρου.



Εικόνα 2.5 : Ο ρόλος ορισμένων ενζύμων στο «άνοιγμα» της αλυσίδας του DNA

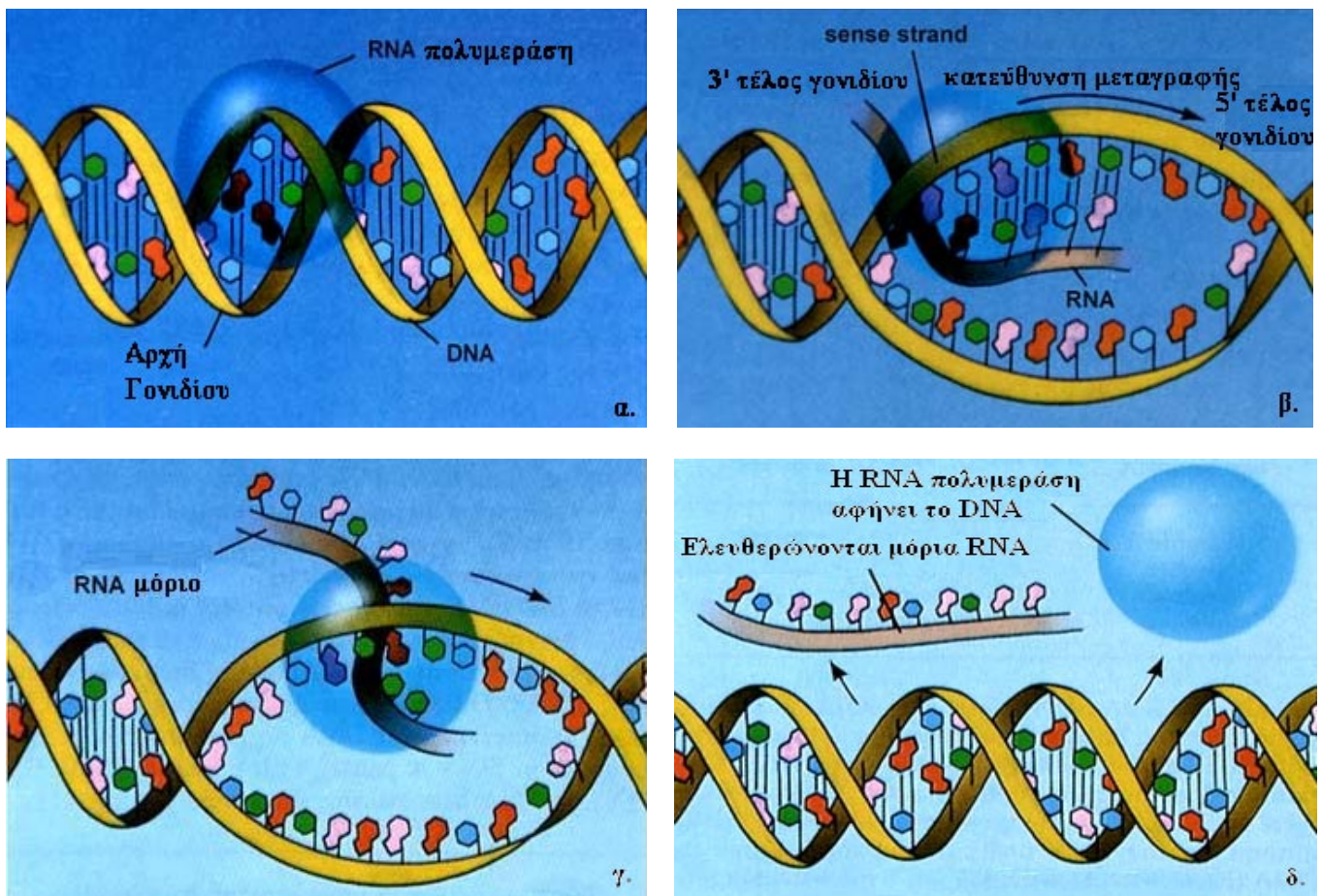
DNA είναι απαραίτητο να ξετυλιχθούν στις θέσεις έναρξης της αντιγραφής οι δύο αλυσίδες.

Όταν το DNA αρχίζει να ξετυλίγεται, δημιουργείται ένα πρόβλημα όμοιο με εκείνο το οποίο αντιμετωπίζουμε όταν προσπαθούμε να ξετυλίξουμε ένα δίκλωνο σκοινί. Το σκοινί, που είναι μια διπλή έλικα, περιστρέφεται γύρω από τον εαυτό του και δημιουργεί υπερέλικες. Οι υπερέλικες εμποδίζουν το ξετύλιγμα του DNA άρα και τη συνέχιση της αντιγραφής. Το κύτταρο λύνει το πρόβλημα αυτό με ειδικά ένζυμα τα οποία καταστρέφουν τις υπερέλικες που δημιουργούνται κατά το ξετύλιγμα. Τα ένζυμα αυτά ονομάζονται **DNA ελικάσες**.

2.5 Μεταγραφή του DNA

Ο μηχανισμός της μεταγραφής είναι ο ίδιος στους προκαρυωτικούς και ευκαρυωτικούς οργανισμούς. Η μεταγραφή καταλύεται από ένα ένζυμο, την **RNA πολυμεράση** (στους ευκαρυωτικούς οργανισμούς υπάρχουν τρία είδη RNA πολυμερασών).

Η RNA πολυμεράση προσδένεται σε ειδικές περιοχές του DNA, που ονομάζονται **υποκινητές**, με την βοήθεια πρωτεϊνών που ονομάζονται **μεταγραφικοί παράγοντες**. Οι υποκινητές και οι μεταγραφικοί παράγοντες αποτελούν τα ρυθμιστικά στοιχεία της μεταγραφής του DNA και επιτρέπουν στην RNA πολυμεράση να αρχίσει σωστά τη μεταγραφή. Οι υποκινητές βρίσκονται πάντοτε πριν από την αρχή κάθε γονιδίου (Εικόνα 2.6.α).



Εικόνα 2.6: Μεταγραφή τμήματος του DNA για το σχηματισμό ενός μορίου RNA

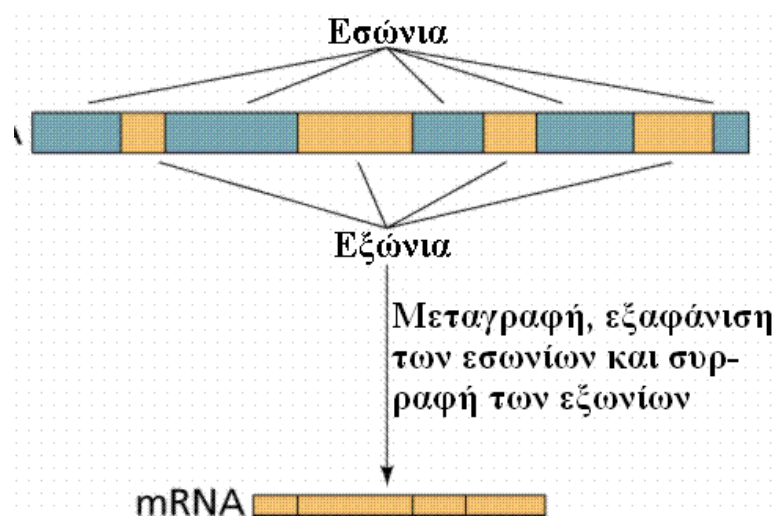
Κατά την έναρξη της μεταγραφής ενός γονιδίου η RNA πολυμεράση προσδένεται στον υποκινητή και προκαλεί τοπικό ξετύλιγμα της διπλής έλικας του DNA (εικόνα 2.6.β). Στη συνέχεια, τοποθετεί τα ριβονουκλεοτίδια απέναντι από τα δεοξυριβονουκλεοτίδια μιας αλυσίδας του DNA σύμφωνα με τον κανόνα της συμπληρωματικότητας των βάσεων, όπως και στην αντιγραφή με τη διαφορά ότι εδώ απέναντι από την αδείνη τοποθετείται ριβονουκλεοτίδιο που περιέχει ουρακίλη (εικόνα 2.6.γ). Η RNA πολυμεράση συνδέει τα ριβονουκλεοτίδια, που προστίθενται το ένα μετά το άλλο, με 3' – 5' φωσφοδιεστερικό δεσμό. Η μεταγραφή έχει προσανατολισμό 5' → 3' όπως και η αντιγραφή. Η σύνθεση του RNA σταματά στο τέλος του γονιδίου όπου ειδικές αλληλουχίες οι οποίες ονομάζονται **αλληλουχίες λήξης της μεταγραφής**, επιτρέπουν την απελευθέρωση του (εικόνα 2.6.δ).

Το μόριο RNA που συντίθεται είναι συμπληρωματικό προς τη μια αλυσίδα της διπλής έλικας του DNA του γονιδίου. Η αλυσίδα αυτή είναι η μεταγραφόμενη και ονομάζεται **μη-κωδική**. Η συμπληρωματική αλυσίδα του DNA του γονιδίου ονομάζεται **κωδική**. **Το RNA είναι το κινητό αντίγραφο της πληροφορίας ενός γονιδίου.**

Στους προκαρυωτικούς οργανισμούς το mRNA αρχίζει να μεταφράζεται σε πρωτεΐνη πριν ακόμη ολοκληρωθεί η μεταγραφή του. Αυτό είναι δυνατό, επειδή δεν υπάρχει πυρηνική μεμβράνη.

Αντίθετα, στους ευκαρυωτικούς οργανισμούς, το RNA που παράγεται κατά τη μεταγραφή ενός γονιδίου συνήθως δεν είναι έτοιμο να μεταφραστεί, αλλά υφίσταται μια πολύπλοκη διαδικασία **ωρίμανσης**. Η διαδικασία αυτή αποτελεί ένα από τα πιο ενδιαφέροντα ευρήματα της Μοριακής Βιολογίας γιατί οδήγησε στο συμπέρασμα ότι τα περισσότερα γονίδια των ευκαρυωτικών οργανισμών (και των ιών που τους προσβάλλουν) είναι **ασυνεχή ή διακεκομμένα**. Δηλαδή η αλληλουχία που μεταφράζεται σε αμινοξέα διακόπτεται από ενδιάμεσες αλληλουχίες οι οποίες δεν μεταφράζονται σε αμινοξέα. Οι αλληλουχίες που μεταφράζονται σε αμινοξέα ονομάζονται **εξώνια** ενώ οι ενδιάμεσες αλληλουχίες ονομάζονται **εσώνια**.

Όταν ένα γονίδιο που περιέχει εσώνια μεταγράφεται, δημιουργείται το **πρόδρομο mRNA** που περιέχει και εξώνια και εσώνια. Το πρόδρομο mRNA μετατρέπεται σε mRNA με τη διαδικασία της ωρίμανσης, κατά την οποία τα εσώνια κόβονται από μικρά ριβονουκλεοπρωτεϊνικά σωματίδια και απομακρύνονται (εικόνα 2.7). Τα ριβονουκλεοπρωτεϊνικά σωματίδια αποτελούνται από snRNA και από πρωτεΐνες και λειτουργούν ως ένζυμα: κόβουν τα εσώνια και συρράπτουν τα εξώνια μεταξύ τους. Έτσι σχηματίζεται το **ώριμο mRNA**. Αυτό, παρότι αποτελείται αποκλειστικά από εξώνια έχει δύο περιοχές που δεν μεταφράζονται σε αμινοξέα. Η μία βρίσκεται στο 5' άκρο και η άλλη στο 3' άκρο. Οι αλληλουχίες αυτές ονομάζονται 5' και 3' αμετάφραστες περιοχές αντίστοιχα. Το mRNA μεταφέρεται από τον πυρήνα στο κυτταρόπλασμα και ειδικότερα στα ριβοσώματα όπου είναι η θέση της πρωτεϊνοσύνθεσης.



Εικόνα 2.7: Εσώνια, Εξώνια και παραγωγή ωρίμου mRNA

2.6 Ο γενετικός κώδικας

Ο γενετικός κώδικας είναι η αντιστοίχιση τριπλέτων βάσεων σε αμινοξέα. Με τη μεταγραφή, οι πληροφορίες που βρίσκονται στα γονίδια μεταφέρονται στο mRNA με βάση τη συμπληρωματικότητα των νουκλεοτιδικών βάσεων. Η αλληλουχία των βάσεων του mRNA καθορίζει την αλληλουχία των αμινοξέων στις πρωτεΐνες με βάση έναν κώδικα αντιστοίχισης νουκλεοτιδίων RNA με αμινοξέα πρωτεϊνών ο οποίος ονομάζεται **γενετικός κώδικας**. Για αυτό η πρωτεϊνοσύνθεση είναι πραγματικά μια διαδικασία «μετάφρασης» από τη γλώσσα των βάσεων στη γλώσσα των αμινοξέων.

Επειδή ο αριθμός των διαφορετικών αμινοξέων που συγκροτούν τις πρωτεΐνες είναι είκοσι και, αντίστοιχα, ο αριθμός των διαφορετικών νουκλεοτιδίων που συγκροτούν το RNA είναι τέσσερα, θεωρήθηκε πιθανό ότι τρία νουκλεοτίδια αντιστοιχούν σε ένα αμινοξύ και για αυτό ο γενετικός κώδικας ονομάστηκε κώδικας τριπλέτας.

Ο κώδικας τριπλέτας είναι φυσική συνέπεια του γεγονότος ότι τέσσερα νουκλεοτίδια αν συνδυαστούν ανά ένα ($4^1 = 4$) ή ανά δύο ($4^2 = 16$), δεν δίνουν αρκετούς συνδυασμούς για να κωδικοποιηθούν τα είκοσι αμινοξέα. Αν όμως συνδυαστούν ανά τρία ($4^3 = 64$) οι συνδυασμοί είναι παραπάνω από αρκετοί.

Τα βασικά χαρακτηριστικά του γενετικού κώδικα είναι:

1. Ο γενετικός κώδικας είναι **κώδικας τριπλέτας**, δηλαδή μια τριάδα νουκλεοτιδίων, που ονομάζεται κωδικόνιο, κωδικοποιεί ένα αμινοξύ.
2. Ο γενετικός κώδικας είναι **συνεχής**, δηλαδή το mRNA διαβάζεται συνεχώς ανά τρία νουκλεοτίδια χωρίς να παραλείπεται κάποιο νουκλεοτίδιο.
3. Ο γενετικός κώδικας είναι **μη επικαλυπτόμενος**, δηλαδή κάθε νουκλεοτίδιο ανήκει σε ένα μόνο κωδικόνιο.
4. Ο γενετικός κώδικας είναι **σχεδόν καθολικός**. Όλοι οι οργανισμοί έχουν τον ίδιο γενετικό κώδικα. Αυτό πρακτικά σημαίνει ότι το mRNA από οποιονδήποτε οργανισμό μπορεί να μεταφραστεί σε εκχυλίσματα φυτικών, ζωικών ή βακτηριακών κυττάρων *in vitro* και να παράγει την ίδια πρωτεΐνη.
5. Ο γενετικός κώδικας χαρακτηρίζεται ως **εκφυλισμένος**. Με εξαίρεση δύο αμινοξέα (μεθειονίνη και τρυπτοφάνη) τα υπόλοιπα δεκαοκτώ κωδικοποιούνται από δύο μέχρι και έξι διαφορετικά κωδικόνια. Τα κωδικόνια που κωδικοποιούν το ίδιο αμινοξύ ονομάζονται **συνώνυμα**.
6. Ο γενετικός κώδικας έχει **κωδικόνιο έναρξης** και **κωδικόνια λήξης**. Το κωδικόνιο έναρξης σε όλους τους οργανισμούς είναι το AUG και κωδικοποιεί το αμινοξύ μεθειονίνη. Υπάρχουν τρία κωδικόνια λήξης, τα UAG, UGA και UAA. Η

παρουσία των κωδικονίων αυτών στο μόριο του mRNA οδηγεί στο τερματισμό της σύνθεσης της πολυπεπτιδικής αλυσίδας.

Ο όρος κωδικόνιο δεν αφορά μόνο στο mRNA αλλά και το γονίδιο από το οποίο παράγεται. Έτσι, για παράδειγμα, το κωδικόνιο έναρξης AUG αντιστοιχεί στο κωδικόνιο έναρξης της κωδικής αλυσίδας του γονιδίου ATG, κοκ.

Η αλληλουχία βάσεων ενός γονιδίου και του mRNA του, που κωδικοποιεί μια πολυπεπτιδική αλυσίδα, αρχίζει με το κωδικόνιο έναρξης και τελειώνει με το κωδικόνιο λήξης. Η διαδρομή με βήμα τριπλέτας από το κωδικόνιο έναρξης μέχρι το κωδικόνιο λήξης ορίζεται ως **ανοικτό πλαίσιο ανάγνωσης** (το κωδικόνιο έναρξης περιλαμβάνεται στο πλαίσιο ανάγνωσης ενώ το κωδικόνιο λήξης όχι).

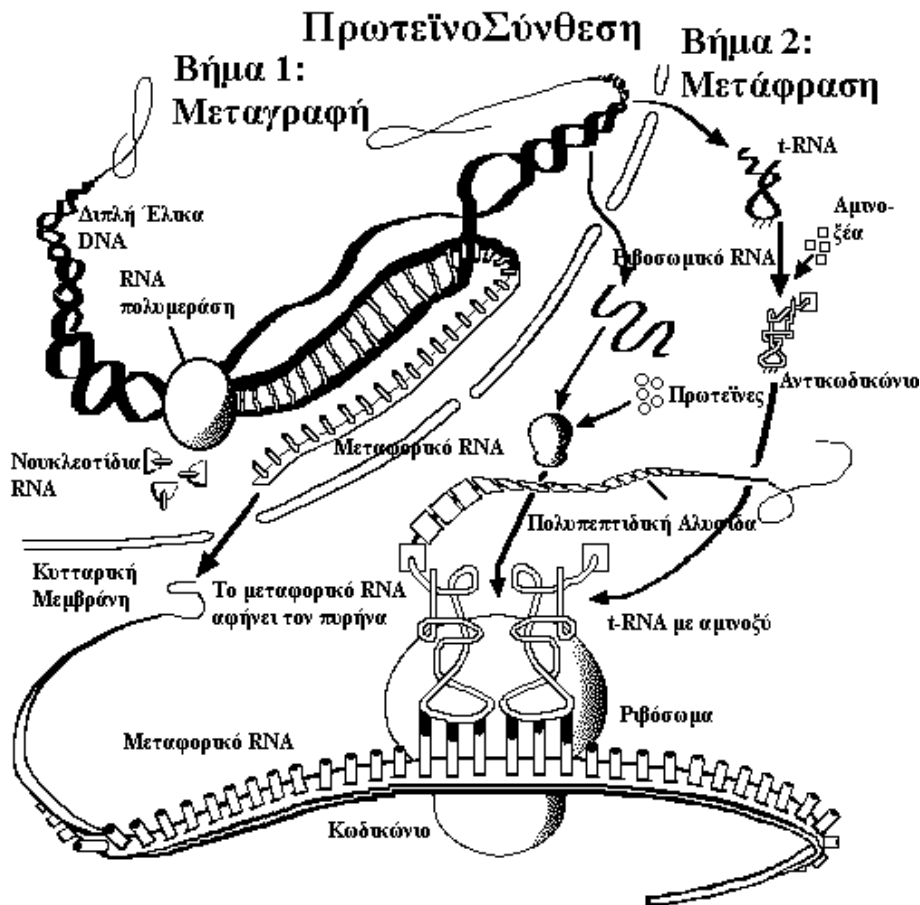
2.7 Μετάφραση

Η μετάφραση του mRNA, δηλαδή η αντιστοίχιση των κωδικονίων σε αμινοξέα και η διαδοχική σύνδεση των αμινοξέων σε πολυπεπτιδική αλυσίδα, πραγματοποιείται στα ριβοσώματα με τη βοήθεια των tRNA και τη συμμετοχή αρκετών πρωτεϊνών και ενέργειας (εικόνα 2.8). Τα ριβοσώματα μπορούν να χρησιμοποιηθούν ως θέση μετάφρασης για οποιοδήποτε mRNA. Αυτό εξηγεί γιατί τα βακτήρια μπορούν να χρησιμοποιηθούν ως εργοστάσια παραγωγής ανθρωπίνων πρωτεϊνών.

Κάθε ριβόσωμα αποτελείται από δύο υπομονάδες, μια μικρή και μια μεγάλη, και έχει μια θέση πρόσδεσης του mRNA στη μικρή υπομονάδα και δύο θέσεις εισδοχής των tRNA στη μεγάλη υπομονάδα. Κάθε μόριο tRNA έχει μια ειδική τριπλέτα νουκλεοτιδίων, το αντικωδικόνιο, με την οποία προσδέεται, λόγω συμπληρωματικότητας, με το αντίστοιχο κωδικόνιο του mRNA. Επιπλέον, κάθε μόριο tRNA διαθέτει μια ειδική θέση σύνδεσης με ένα συγκεκριμένο αμινοξύ.

Η πρωτεϊνοσύνθεση διακρίνεται σε τρία στάδια: την έναρξη, την επιμήκυνση και τη λήξη.

Έναρξη: Κατά την έναρξη της μετάφρασης το mRNA συνδέεται μέσω μιας αλληλουχίας που υπάρχει στην 5' αμετάφραστη περιοχή του, με το ριβοσωμικό RNA της μικρής υπομονάδας του ριβοσώματος, σύμφωνα με τους κανόνες συμπληρωματικότητας των βάσεων. Το πρώτο κωδικόνιο του mRNA είναι πάντοτε AUG και σ' αυτό προσδέεται το tRNA που φέρει το αμινοξύ μεθειονίνη. Όμως, δεν έχουν όλες οι πρωτεΐνες του οργανισμού ως πρώτο αμινοξύ μεθειονίνη. Αυτό συμβαίνει, γιατί σε πολλές πρωτεΐνες, μετά τη σύνθεσή τους απομακρύνονται ορισμένα αμινοξέα από το αρχικό αμινικό άκρο



Εικόνα 2.8: Διάγραμμα Πρωτεϊνοσύνθεσης

τους. Το σύμπλοκο που δημιουργείται μετά την πρόσδεση του mRNA στη μικρή υπομονάδα του ριβοσώματος και του tRNA που μεταφέρει την μεθειονίνη ονομάζεται σύμπλοκο έναρξης της πρωτεϊνοσύνθεσης και στη συνέχεια η μεγάλη υπομονάδα συνδέεται με τη μικρή.

Επιμήκυνση: Κατά την επιμήκυνση ένα δεύτερο μόνο tRNA με αντικωδικόνιο συμπληρωματικό του δεύτερου κωδικονίου του mRNA τοποθετείται στην κατάλληλη εισδοχή του ριβοσώματος, μεταφέροντας το δεύτερο αμινοξύ. Μεταξύ της μεθειονίνης και του δεύτερου αμινοξέως σχηματίζεται πεπτιδικός δεσμός και αμέσως μετά το πρώτο tRNA αποσυνδέεται από το ριβόσωμα και απελευθερώνεται στο κυτταρόπλασμα όπου συνδέεται πάλι με μεθειονίνη, έτοιμο για επόμενη χρήση. Το ριβόσωμα και το mRNA έχουν τώρα ένα tRNA, πάνω στο οποίο είναι προσδεμένα δύο αμινοξέα. Έτσι αρχίζει η επιμήκυνση της πολυπεπτιδικής αλυσίδας.

Στη συνέχεια το ριβόσωμα κινείται κατά μήκος του mRNA κατά ένα κωδικόνιο. Ένα τρίτο tRNA έρχεται να προσδεθεί μεταφέροντας το αμινοξύ του. Ανάμεσα στο δεύτερο και στο τρίτο αμινοξύ σχηματίζεται πεπτιδικός δεσμός. Η

πολυπεπτιδική αλυσίδα συνεχίζει να αναπτύσσεται, καθώς νέα tRNA φέρουν αμινοξέα τα οποία προσδένονται μεταξύ τους.

Λήξη: Η επιμήκυνση σταματά σε ένα κωδικόνιο λήξης (UGA, UAG και UAA) επειδή δεν υπάρχουν tRNA που να αντιστοιχούν σε αυτά. Το τελευταίο tRNA απομακρύνεται από το ριβόσωμα και η πολυπεπτιδική αλυσίδα απελευθερώνεται.

Σημειώνεται ότι πολλά μόρια mRNA μπορούν να μεταγράφονται από ένα μόνο γονίδιο. Πολλά ριβοσώματα μπορούν να μεταφράζουν ταυτόχρονα ένα mRNA, το καθένα σε διαφορετικό σημείο κατά μήκος του μορίου. Αμέσως μόλις το ριβόσωμα έχει μεταφράσει τα πρώτα κωδικόνια, η θέση έναρξης του mRNA είναι ελεύθερη για την πρόσδεση ενός άλλου ριβοσώματος. Το σύμπλεγμα των ριβοσωμάτων με το mRNA ονομάζεται πολύσωμα. Έτσι, η πρωτεϊνοσύνθεση είναι μια οικονομική διαδικασία. Ένα κύτταρο μπορεί να αναπαραγάγει μεγάλα ποσά μιας πρωτεΐνης από ένα ή από δύο αντίγραφα ενός γονιδίου.

2.8 Γονιδιακή Ρύθμιση: Ο έλεγχος της γονιδιακής έκφρασης

Η έκφραση των γονιδίων ρυθμίζεται με διάφορους μηχανισμούς.

Ο όρος γονιδιακή έκφραση αναφέρεται συνήθως σε όλη τη διαδικασία με την οποία ένα γονίδιο ενεργοποιείται, για να παραγάγει μια πρωτεΐνη. Όμως σε κάθε κύτταρο δεν παράγονται όλες οι πρωτεΐνες σε κάθε χρονική στιγμή. Επιπλέον, επειδή το κύτταρο χρειάζεται κάθε πρωτεΐνη σε συγκεκριμένη ποσότητα, οι πρωτεΐνες ενός κυττάρου δεν παράγονται σε ίσες ποσότητες. Αν λοιπόν όλα τα γονίδια δούλευαν με τον ίδιο ρυθμό, ορισμένες πρωτεΐνες θα παράγονταν σε μεγάλες ποσότητες και άλλες σε ποσότητες που δεν θα επαρκούσαν. Έτσι, είναι απαραίτητη η ύπαρξη και η λειτουργία ενός προγράμματος ρύθμισης της γονιδιακής έκφρασης, που παρέχει τις οδηγίες για το είδος και την ποσότητα των πρωτεϊνών οι οποίες πρέπει να παραχθούν σε κάθε χρονική στιγμή.

Στα βακτήρια η ρύθμιση της γονιδιακής έκφρασης αποσκοπεί κυρίως στην προσαρμογή του οργανισμού στις εναλλαγές του περιβάλλοντος, έτσι ώστε να εξασφαλίζονται οι καλύτερες συνθήκες για τη βασική λειτουργία του που είναι η αύξηση και η διαίρεση.

Τα κύτταρα ενός πολυκύτταρου οργανισμού, σε αντίθεση με τα κύτταρα που ανήκουν σε ένα βακτηριακό στέλεχος και είναι πανομοιότυπα μεταξύ τους, διαφέρουν στη δομή και τη λειτουργία τους. Η ζωή αρχίζει όταν ένα

γονιμοποιημένο ωάριο διαιρείται με μίτωση και παράγει τρισεκατομμύρια κύτταρα, που έχουν τα ίδια γονίδια. Στα αρχικά στάδια της εμβρυογένεσης τα κύτταρα εξειδικεύονται, για να εκτελέσουν επιμέρους λειτουργίες, μια διαδικασία που ονομάζεται κυτταρική διαφοροποίηση. Τα κύτταρα ενός πολύπλοκου πολυκύτταρου οργανισμού, όπως τα νευρικά, τα μυϊκά, τα ηπατικά, διαφέρουν στην μορφή και τη λειτουργία τους αλλά έχουν το ίδιο γενετικό υλικό, άρα και τα ίδια γονίδια. Επομένως, πώς διαφέρουν τόσο πολύ μεταξύ τους;

Μολονότι όλα τα κύτταρα έχουν τις ίδιες γενετικές οδηγίες, έχουν αναπτύξει μηχανισμούς που τους επιτρέπουν να εκφράζουν τη γενετική τους πληροφορία επιλεκτικά και να ακολουθούν μόνο τις οδηγίες που χρειάζονται κάθε χρονική στιγμή. Κάθε κυτταρικός τύπος έχει εξειδικευμένη λειτουργία και πρέπει να υπάρχει πλήρης συντονισμός των λειτουργιών όλων των κυττάρων. Για αυτό, η τελειοποίηση των συστημάτων ελέγχου είναι αναγκαία και λόγω της μεγαλύτερης πολυπλοκότητας των ευκαρυωτικών οργανισμών, αλλά και επειδή πρέπει να ελεγχθεί προσεχτικά η ανάπτυξη πολυκύτταρων οργανισμών. Κατά συνέπεια, η ρύθμιση των γονιδίων στα ευκαρυωτικά κύτταρα γίνεται σε πολλά επίπεδα.

2.8 Η γονιδιακή ρύθμιση στους προκαρυωτικούς οργανισμούς

Ένα βακτηριακό κύτταρο E.Coli έχει περίπου 3000 γονίδια. Μερικά γονίδια μεταγράφονται συνεχώς και κωδικοποιούν πρωτεΐνες, που χρειάζονται για τις βασικές λειτουργίες του κυττάρου. Άλλα γονίδια μεταγράφονται μόνον όταν το κύτταρο αναπτύσσεται σε ειδικές περιβαλλοντολογικές συνθήκες, επειδή τα προϊόντα των γονιδίων αυτών είναι απαραίτητα για την επιβίωση του κυττάρου στις συνθήκες αυτές.

Για παράδειγμα, τα βακτήρια E.Coli χρησιμοποιούν ως πηγή άνθρακα το σάκχαρο γλυκόζη. Αν στο περιβάλλον του βακτηρίου υπάρχει ο δισακχαρίτης λακτόζη, το βακτήριο έχει τη δυνατότητα να τον διασπάσει για να επιβιώσει ή θα πεθάνει, μολονότι γύρω του υπάρχει άφθονη τροφή; Το βακτήριο λύνει το πρόβλημα αυτό ρυθμίζοντας την παραγωγή των κατάλληλων ενζύμων, που θα διασπάσουν τη λακτόζη σε γλυκόζη και γαλακτόζη.

Οι μηχανισμοί με τους οποίους ένα κύτταρο «ξυπνά» ένα «κοιμισμένο» γονίδιο είναι οι πιο σημαντικοί και πολύπλοκοι της Μοριακής Βιολογίας. Οι αρχικές μελέτες της ρύθμισης των γονιδίων έγιναν από τους Jacob και Monod, το 1961. Οι ερευνητές περιέγραψαν την ικανότητα του βακτηρίου E.Coli να παραγάγει τα τρία απαραίτητα ένζυμα που χρειάζεται για να μεταβολίσει το δισακχαρίτη λακτόζη, όταν δεν υπάρχει άλλη πηγή άνθρακα στη τροφή του. Οι

Jacob και Monod απέδειξαν με γενετικές μελέτες ότι τα γονίδια που κωδικοποιούν τα τρία αυτά ένζυμα βρίσκονται το ένα δίπλα στο άλλο πάνω στο γονιδίωμα του βακτηρίου και αποτελούν μονάδα που την ονόμασαν οπερόνιο της λακτόζης.

Σε αυτό περιλαμβάνονται εκτός από αυτά τα γονίδια που ονομάζονται δομικά και αλληλουχίες DNA που ρυθμίζουν τη μεταγραφή τους. Οι αλληλουχίες αυτές που βρίσκονται μπροστά από τα δομικά γονίδια είναι κατά σειρά ένα ρυθμιστικό γονίδιο, ο υποκινητής και ο χειριστής.

Το οπερόνιο της λακτόζης δεν μεταγράφεται ούτε μεταφράζεται, όταν απουσιάζει από το θρεπτικό υλικό η λακτόζη. Τότε λέμε ότι τα γονίδια που το αποτελούν βρίσκονται υπό καταστολή. Πώς όμως επιτυγχάνεται η καταστολή; Με δύο ρυθμιστικά μόρια: μια αλληλουχία DNA που ονομάζεται χειριστής και βρίσκεται μεταξύ του υποκινητή και του πρώτου γονιδίου, και μια ρυθμιστική πρωτεΐνη - καταστολέας. Όταν απουσιάζει η λακτόζη ο καταστολέας προσδένεται ισχυρά στο χειριστή και εμποδίζει την RNA πολυμεράση να αρχίσει τη μεταγραφή των γονιδίων του οπερονίου. Ο καταστολέας κωδικοποιείται από ένα ρυθμιστικό γονίδιο, που βρίσκεται μπροστά από τον υποκινητή. Το ρυθμιστικό γονίδιο μεταγράφεται συνεχώς και παράγει λίγα μόρια του καταστολέα. Τα μόρια αυτά προσδένονται συνεχώς στο χειριστή.

Όταν στο θρεπτικό υλικό υπάρχει μόνο λακτόζη, τότε ο ίδιος ο δισακχαρίτης προσδένεται στην καταστολέα και δεν του επιτρέπει να προσδεθεί στο χειριστή. Τότε η RNA πολυμεράση είναι ελεύθερη να αρχίσει τη μεταγραφή. Δηλαδή η λακτόζη λειτουργεί ως επαγωγέας της μεταγραφής των γονιδίων του οπερονίου. Τότε τα γονίδια αρχίζουν να εκφράζονται δηλαδή να μεταγράφονται και να συνθέτουν τα ένζυμα. Τα τρία ένζυμα μεταφράζονται από το ίδιο μόριο mRNA το οποίο περιέχει κωδικόνια έναρξης και λήξης για κάθε ένζυμο. Συμπερασματικά, η ίδια η λακτόζη ενεργοποιεί τη διαδικασία για την αποκωδικοποίησή της. Όταν η λακτόζη διασπαστεί πλήρως, τότε η πρωτεΐνη - καταστολέας είναι ελεύθερη να προσδεθεί στο χειριστή και να καταστείλει τη λειτουργία των τριών γονιδίων.

Στο γονιδίωμα των προκαρυωτικών οργανισμών τα γονίδια των ενζύμων που παίρνουν μέρος σε μια μεταβολική οδό, όπως η διάσπαση της λακτόζης ή η βιοσύνθεση διαφόρων αμινοξέων, οργανώνονται σε υπερόνια, δηλαδή σε ομάδες που υπόκεινται σε κοινό έλεγχο της έκφρασής τους.

2.9 Η γονιδιακή ρύθμιση στους ευκαρυωτικούς οργανισμούς

Η ρύθμιση της έκφρασης των γονιδίων στα ευκαρυωτικά κύτταρα γίνεται με ιδιαίτερα πολύπλοκους μηχανισμούς και αποτελεί σήμερα αντικείμενο εντατικής ερευνητικής μελέτης. Η πλήρης διαλεύκανση των μηχανισμών αυτών θα δώσει απαντήσεις για το πώς, όταν οι μηχανισμοί αυτοί απορυθμίζονται, τα κύτταρα βγαίνουν από το αυστηρό πρόγραμμα της λειτουργίας του και γίνονται καρκινικά.

Στα ευκαρυωτικά κύτταρα η γονιδιακή έκφραση ρυθμίζεται σε τέσσερα επίπεδα:

Στο επίπεδο της μεταγραφής. Ένας αριθμός μηχανισμών ελέγχουν ποια γονίδια θα μεταγραφούν ή / και με ποια ταχύτητα θα γίνει η μεταγραφή. Το DNA των ευκαρυωτικών κυττάρων δεν οργανώνεται σε οπερόνια αλλά κάθε γονίδιο έχει το δικό του υποκινητή και μεταγράφεται αυτόνομα. Η RNA πολυμεράση λειτουργεί (όπως και στους προκαρυωτικούς οργανισμούς) με τη βοήθεια των πρωτεϊνών, που ονομάζονται μεταγραφικοί παράγοντες. Μόνο που στους ευκαρυωτικούς οργανισμούς οι μεταγραφικοί παράγοντες παρουσιάζουν τεράστια ποικιλία. Κάθε κυτταρικός τύπος περιέχει διαφορετικά είδη μεταγραφικών παραγόντων. Διαφορετικός συνδυασμός μεταγραφικών παραγόντων ρυθμίζει τη μεταγραφή κάθε γονιδίου. Μόνο όταν ο σωστός συνδυασμός των μεταγραφικών παραγόντων προσδεθεί στον υποκινητή ενός γονιδίου, αρχίζει η RNA πολυμεράση τη μεταγραφή ενός γονιδίου.

Στο επίπεδο μετά τη μεταγραφή. Περιλαμβάνονται οι μηχανισμοί με τους οποίους γίνεται η ωρίμανση του πρόδρομου mRNA και επίσης η ταχύτητα με την οποία το ώριμο mRNA αφήνει τον πυρήνα και εισέρχεται στο κυτταρόπλασμα.

Στο επίπεδο της μετάφρασης. Ο χρόνος που ζουν τα μόρια mRNA στο κυτταρόπλασμα δεν είναι ο ίδιος για όλα τα είδη RNA, επειδή μετά από κάποιο χρονικό διάστημα αποικοδομούνται. Επίσης, ποικίλει και η ικανότητα πρόσδεσης του mRNA στα ριβοσώματα.

Στο επίπεδο μετά τη μετάφραση. Ακόμα και όταν γίνει η πρωτεϊνοσύνθεση και παραχθεί η κατάλληλη πρωτεΐνη, μπορεί να χρειάζεται να υποστεί τροποποιήσεις για να γίνει βιολογικά λειτουργική.

2.10 Ποσοστό Έκφρασης Γονιδίων

Τα κύτταρα ενός πολυκυτταρικού οργανισμού έχουν στο σύνολό τους τον ίδιο αριθμό γονιδίων. Ωστόσο, δεν εκφράζονται όλα τα γονίδια με τον ίδιο τρόπο αλλά ανάλογα με τον ιστό (κυτταρική διαφοροποίηση) και τις ανάγκες του

κυττάρου. Για να επιτευχθεί η μεταγραφή ενός γονιδίου (δηλαδή για να υπάρξει έκφραση) πρέπει να υπάρχει συγκεκριμένος συνδυασμός μεταγραφικών παραγόντων.

Η διάρκεια ζωής ενός mRNA στο κυτταρόπλασμα θα καθορίζει την ποσότητα από την πρωτεΐνη που θα κατασκευαστεί και καθορίζεται από το μήκος της ουράς των αδενινών οι οποίες προστίθενται κατά την ωρίμανση στο 3' άκρο της αλυσίδας του DNA. Από το συγκεκριμένο μήκος καθορίζεται και το λεγόμενο ποσοστό της έκφρασης.

Συγκεκριμένα, κάθε φορά που ένα ριβόσωμα «περνά» και σχηματίζει ένα μόριο πρωτεΐνης μειώνεται το μήκος της ουράς των αδενινών κατά μία. Όταν οι αδενίνες εξαλειφθούν από το άκρο της αλυσίδας του DNA, το mRNA ανακυκλώνεται σε ριβονουκλεοτίδια.

Κάνοντας χρήση μιας ψευδογλώσσας προγραμματισμού, η παραπάνω διαδικασία μπορεί να γραφεί ως εξής:

```
Input : adenine,mRNA
Output : protein, ribonuclei
while (adenine>0)
    produce 1 protein molecule
    adenine = adenine -1
end
recycle mRNA
```

2.11 Η υβριδοποίηση των νουκλεικών οξέων χρησιμοποιείται για την ανίχνευση κλώνων γονιδιωματικής ή cDNA βιβλιοθήκης

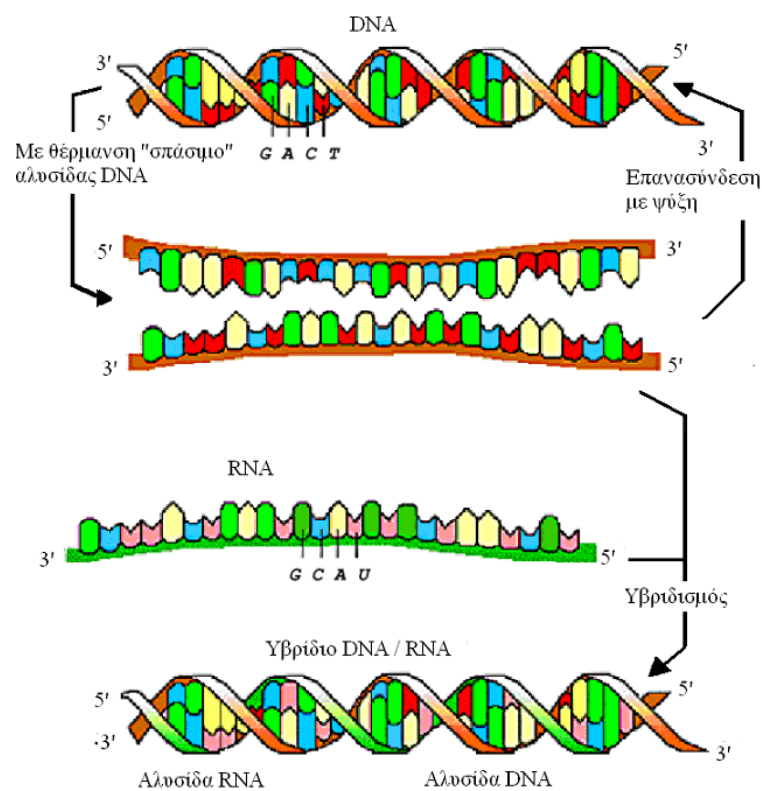
Η απομόνωση του συνολικού DNA από κύτταρα προκαρυωτικού ή ευκαρυωτικού οργανισμού στο δοκιμαστικό σωλήνα είναι υπόθεση ρουτίνας.

Αν επιδράσουμε στο DNA που απομονώθηκε με κατάλληλες χημικές ουσίες ή αυξήσουμε τη θερμοκρασία του τότε σπάζουν οι δεσμοί υδρογόνου μεταξύ των δύο συμπληρωματικών αλυσίδων και οι δύο αλυσίδες αποχωρίζονται η μία από την άλλη. Η διαδικασία αυτή ονομάζεται **αποδιάταξη**.

Οι δύο μονόκλωνες συμπληρωματικές αλυσίδες σε κατάλληλες συνθήκες μπορούν να **επανασυνδεθούν**. Στην ιδιότητα αυτή στηρίζεται η διαδικασία της **υβριδοποίησης**⁴ που είναι η σύνδεση μονόκλωνων συμπληρωματικών αλυσίδων DNA ή συμπληρωματικών DNA –RNA με το DNA της βιβλιοθήκης (το οποίο έχει αποδιαταχθεί) και υβριδοποιούν μόνο το συμπληρωματικό τους DNA.

Η διαδικασία της υβριδοποίησης ακολουθείται και για την απομόνωση ενός συγκεκριμένου γονιδίου από μια cDNA⁵ βιβλιοθήκη.

Η υβριδοποίηση είναι μια πολύ σημαντική ιδιότητα του DNA που μας δίνει τη δυνατότητα **αν έχουμε γνωστό μόριο DNA να το χρησιμοποιήσουμε ως ανιχνευτή** για τον εντοπισμό του συμπληρωματικού του όταν το τελευταίο βρίσκεται μαζί με χιλιάδες άλλα κομμάτια.



Εικόνα 2.9 : Σχηματικό Διάγραμμα Υβριδισμού

⁴ **Υβριδισμός** ή **Υβριδοποίηση** καλείται η σύνδεση δύο μονόκλωνων αλυσίδων DNA, με υδρογονικούς δεσμούς, σύμφωνα με τον κανόνα της συμπληρωματικότητας των βάσεων.

⁵ **cDNA** : συμπληρωματική αλυσίδα DNA (**complementary**). Η σύνθεση του cDNA γίνεται από το ένζυμο αντίστροφη μεταγραφάση. Παράγονται έτσι υβριδικά cDNA – mRNA. Το mRNA διασπάται με κατάλληλες χημικές ουσίες ή αποδιατάσσεται με θέρμανση και τα cDNA χρησιμεύουν σαν καλούπι για τη σύνθεση μιας συμπληρωματικής αλυσίδας DNA.

Μια γονιδιωματική βιβλιοθήκη περιέχει ένα τεράστιο αριθμό από κλωνοποιημένα κομμάτια χρωμοσωμικού DNA τα οποία έχουν παραχθεί με δράση κάποιας περιοριστικής ενδονουκλεάσης. Ορισμένα από τα κομμάτια αυτά περιέχουν **ολόκληρα γονίδια**, άλλα περιέχουν **κομμάτια γονιδίων** και άλλα **τμήματα του DNA που δεν κωδικοποιούν πρωτεΐνες**. Έτσι πρέπει μέσα από όλα αυτά τα κομμάτια να εντοπίσουμε αυτό που θέλουμε να μελετήσουμε.

Η τεχνική που χρησιμοποιείται συνήθως περιλαμβάνει τη χρήση ιχνηθετημένων ανιχνευτών μορίων DNA ή RNA που περιέχουν αλληλουχίες συμπληρωματικές προς το κλωνοποιημένο DNA.

Βιβλιογραφία

- [1] Biology, M.Barbor, M.Boyle,M.Cassidy and K.Senior, Collins International, 1998
- [2] Biology an Exploration of Life, C. McFadden, W. Keeton, Norton, 1995
- [3] Biotechnology from A to Z, 2nd Edition, W.Bains, Oxford University Press, 1998
- [4] Human Genetics, Lewis R., WCB McGraw – Hill, 1999
- [5] Principles of Medical Genetics, T.D. Gelehrter, F.S. Collins and D.Ginsbourg, Williams and Wilkins, 1998
- [6] Understanding Biology, 3rd Edition, P. Raven and G.Johnson, Wm. C. Brown Publishers.,1995

3

Καρκίνος :

Μια πολυγονιδιακή νόσος

3.1 Τι είναι ο Καρκίνος

Καρκίνος ονομάζεται η ασθένεια που χαρακτηρίζεται από ανεξέλεγκτο πολλαπλασιασμό κυττάρων που εμφανίζουν ποικίλο βαθμό πιστότητας με τα προγονικά και δημιουργούν (κακοήθεις) όγκους (μάζες καρκινικών κυττάρων ή αλλιώς νεοπλασίες). [2] Η ονομασία του δόθηκε από τον Ιπποκράτη ο οποίος παρατήρησε ότι η μορφή των όγκων έμοιαζε με καβούρι (καρκίνο). Στην ιατρική ονομάζεται και Νεοπλασματική Νόσος.

3.2 Καρκινογένεση

Ο καρκίνος είναι γενετικό νόσημα: προκύπτει από παθολογικές μεταβολές της αλληλουχίας του DNA. Διαφέρει από άλλα γενετικά νοσήματα κατά το γεγονός ότι οι εμπλεκόμενες μεταλλάξεις⁶ είναι σωματικές, δηλαδή συμβαίνουν σε διάσπαρτα μεμονωμένα σωματικά κύτταρα του ώριμου σώματος , και διαφέρουν από τις μεταλλάξεις της γαμετικής σειράς, που μεταβιβάζονται από

⁶ Μετάλλαξη ή Μεταλλαγή (mutation) ονομάζεται οποιαδήποτε μεταβολή που μπορεί να συμβεί στο γενετικό υλικό ενός οργανισμού. Στα ευκαριωτικά αν η μεταβολή προσβάλλει κύτταραγαμετών τότε χαρακτηρίζεται γενετική μεταλλαγή και μπορεί να κληρονομηθεί. Αντίθετα αν προσβληθούν σωματικά κύτταρα (μη φυλετικά), η μεταλλαγή αυτή ονομάζεται σωματική μεταλλαγή η οποία και δεν κληρονομείται. Οι μεταλλαγές που έχουν ως αποτέλεσμα αλλαγές γονιδίων παράγουν ένα διαφορετικό αλληλόμορφο.

τα γαμετικά κύτταρα από τα οποία αναπτύσσεται ολόκληρος ο πολυκύτταρος οργανισμός [5].

Ένα μεγάλο μέρος της έρευνας κατά του καρκίνου αναζητά τις αιτίες εμφάνισης αυτών των μεταλλάξεων, παρ' ολ' αυτά τα συμπεράσματα είναι ανεπαρκή. Ωστόσο το αποτέλεσμα είναι σαφές: απώλεια ελέγχου της κυτταρικής αύξησης που οδηγεί στον ανεξέλεγκτο πολλαπλασιασμό κυττάρων, ο οποίος χαρακτηρίζει την νόσο.

Ας δούμε όμως την διαδικασία της καρκινογένεσης αναλυτικά:

1. Μετατροπή των πρωτοογκογονιδίων σε ογκογονίδια και απενεργοποίηση των ογκοκατασταλτικών γονιδίων.

Στο γονιδίωμα των φυσιολογικών κυττάρων υπάρχουν γονίδια τα οποία ελέγχουν την φυσιολογική ανάπτυξη τους, την κυτταρική αύξηση και διαφοροποίηση και ονομάζονται πρωτοογκογονίδια. Στο γονιδίωμα επίσης υπάρχουν και τα ογκοκατασταλτικά γονίδια, τα οποία είναι φυσιολογικά γονίδια που τα προϊόντα τους(πρωτεΐνες) αναστέλλουν τον κυτταρικό πολλαπλασιασμό.

Το πρώτο βήμα της καρκινογένεσης είναι η μεταλλαγή των πρωτοογκογονιδίων σε ογκογονίδια, γονίδια δηλαδή που προάγουν τον σχηματισμό όγκου, αλλά και η απενεργοποίηση των ογκοκατασταλτικών γονιδίων.

Υπεύθυνη για αυτές τις αλλαγές δεν είναι μία και μοναδική μετάλλαξη αλλά ένα πλήθος μεταλλάξεων που πιθανώς συσσωρεύονται από γενιά σε γενιά μεταλλάσσοντας τα πρωτοογκογονίδια σε ογκογονίδια και προκαλώντας τέτοιες διαφοροποιήσεις στα ογκοκατασταλτικά γονίδια που τα καθιστούν ανενεργά ([4], [5]).

Τα ογκογονίδια μπορεί να οδηγούν στην παραγωγή μίας μη φυσιολογικής πρωτεΐνης ή ακόμα και σε αφύσικα υψηλά επίπεδα μιας φυσιολογικής πρωτεΐνης. Σε κάθε περίπτωση όμως το κύτταρο οδηγείται σε αύξηση του ρυθμού πολλαπλασιασμού του. Ενώ ταυτόχρονα η απενεργοποίηση των ογκοκατασταλτικών γονιδίων καθιστά το κύτταρο ανίκανο να βάλει τέρμα σε αυτόν τον πολλαπλασιασμό. Έτσι ο πολλαπλασιασμός καθίσταται ανεξέλεγκτος και οδηγεί στο σχηματισμό των όγκων.

2. Μετατροπή των φυσιολογικών κυττάρων σε καρκινικά

Έχει διαπιστωθεί ότι τα καρκινικά κύτταρα προέρχονται από φυσιολογικά κύτταρα ωστόσο έχουν κάποια χαρακτηριστικά που τα διαφοροποιούν από αυτά.

Για το λόγο αυτό θεωρούμε ότι τα καρκινικά κύτταρα έχουν υποστεί αλλαγές. Διαπιστώθηκε ότι οι αλλαγές αυτές περνούν κατά την κυτταρική διαίρεση από ένα καρκινικό κύτταρο στα θυγατρικά του, γεγονός που δείχνει ότι πρόκειται για αλλαγές στο γενετικό υλικό. Τα καρκινικά κύτταρα διαφέρουν από τα φυσιολογικά στα εξής:

- 1) Συνήθως δεν είναι διαφοροποιημένα. Δεν έχουν δηλαδή την εξειδίκευσή των φυσιολογικών κυττάρων με αποτέλεσμα να μην μπορούν να επιτελέσουν τον ρόλο τους μέσα στον ιστό. Πρέπει να αναφερθεί όμως ότι υπάρχουν και καρκίνοι, οι ονομάζονται καλά διαφοροποιημένοι, που τα κύτταρα τους είναι ουσιαστικά πανομοιότυπα με τα φυσιολογικά.
- 2) Εμφανίζουν ανεξέλεγκτο πολλαπλασιασμό. (Μπορούν να) κάνουν επαναλαμβανόμενους κυτταρικούς κύκλους, που δεν ελέγχονται από τους μηχανισμούς που φυσιολογικά περιορίζουν τον πολλαπλασιασμό, γεγονός που οδηγεί σε ανοργάνωτη αύξηση.
- 3) Έχουν μη φυσιολογικούς πυρήνες. Οι πυρήνες είναι διογκωμένοι και μπορεί να έχουν εμφανείς χρωμοσωμικές ανωμαλίες (αλλαγές στον αριθμό και στον τύπο των χρωμοσωμάτων)
- 4) Μπορούν να δημιουργήσουν όγκους. Οι όγκοι είναι συσσωματώματα καρκινικών κυττάρων με την δυνατότητα να προσβάλλουν και να καταστρέψουν γειτονικούς ιστούς. Διακρίνονται σε καλοήθεις, οι οποίοι δεν διηθούν παρακείμενους ιστούς ούτε μεθίστανται σε απομακρυσμένες θέσεις, και σε κακοήθεις ή καρκίνους οι οποίοι και διηθούν και μεθίστανται και στις περιοχές όπου μεθίστανται υποπληθυσμοί καρκινικών κυττάρων τους εγκαθίστανται, αναπτύσσονται εκ νέου και διηθούν ξανά.

Η μετατροπή των φυσιολογικών κυττάρων σε καρκινικά είναι επακόλουθο της μεταλλαγής των πρωτοογκογονιδίων σε ογκογονίδια αλλά και την απενεργοποίηση των ογκοκατασταλτικών γονιδίων.

3. Ανεξέλεγκτος πολλαπλασιασμός των καρκινικών κυττάρων και δημιουργία όγκων.

Τα ογκογονίδια οδηγούν σε αύξηση του ρυθμού πολλαπλασιασμού και διαίρεσης των κυττάρων ενώ η απενεργοποίηση των ογκοκατασταλτικών γονιδίων συνεπάγεται αδυναμία καταστολής του κυτταρικού πολλαπλασιασμού. Έτσι τα καρκινικά κύτταρα πολλαπλασιάζονται ανεξέλεγκτα δημιουργώντας όγκους (ή νεοπλασίες, tumors). Στην περίπτωση

που ο δημιουργηθείς πρωτοπαθής όγκος είναι καλοήθης, συνήθως μπορεί να αφαιρεθεί πλήρως με χειρουργική επέμβαση χωρίς να υποτροπιάζει.

Αν είναι κακοήθης (καρκινικός) δημιουργεί μεταστάσεις⁷, με αποτέλεσμα η αφαίρεση του πρωτοπαθούς όγκου να μην επιφέρει την ίαση, αφού πλέον υπάρχουν άλλες καρκινικές εστίες, σε διάφορα μέρη του σώματος([2],[5]). Οι κακοήθεις όγκοι με την αύξηση του μεγέθους τους και την διεισδυτικότητα των καρκινικών κυττάρων προκαλούν σοβαρά προβλήματα στους ιστούς και στην λειτουργία ολόκληρου του οργανισμού. Ενώ στην περίπτωση των καλοήθων τα προβλήματα που δημιουργούν είναι ουσιαστικά μόνο μηχανικά (πιέζουν τους γύρω ιστούς). Επιπλέον, υπάρχει ο κίνδυνος κακοήθους εξαλλαγής ενός καλοήθους όγκου, δηλαδή να υποπέσει σε κακοήθη όγκο και να αρχίσει να διηθεί τους γύρω ιστούς.

4. Δημιουργία μεταστάσεων.

Τα καρκινικά κύτταρα έχουν την ικανότητα να διεισδύουν στα αιμοφόρα αγγεία και μέσω της ροής του αίματος να μεταφέρονται σε άλλο μέρος του σώματος που δεν βρίσκεται σε επαφή με το προηγούμενο, να εγκαθίστανται και να δημιουργούν νέους καρκινικούς όγκους (αιματογενής διασπορά). Αλλοτε μεταφέρονται μέσω των λεμφαγγείων οπότε έχουμε την λεμφογενή διασπορά. Το αποτέλεσμα είναι να προσβάλλονται και άλλα μέρη και όργανα του σώματος με επακόλουθο την δυσλειτουργία ολόκληρου του οργανισμού και συνήθως, δυστυχώς, τον θάνατο.

3.3 Μορφές Καρκίνου

Κάτω από την ονομασία «Καρκίνος» κρύβεται ένα πλήθος διαφορετικών μορφών της νόσου. Στην ιατρική ο χαρακτηρισμός της κάθε συγκεκριμένης περίπτωσης καρκίνου αποτελεί μία σύνθετη διαδικασία. Γενικότερα υπάρχουν πάνω από 200 διαφορετικά είδη καρκίνου και δεν αντιμετωπίζονται όλοι με τον ίδιο τρόπο. Κάθε ένα είδος έχει τον δικό του τρόπο θεραπευτικής αντιμετώπισης. Οι περισσότερες μορφές καρκίνου είναι στην ουσία όγκοι εκτός από ορισμένους τύπους καρκίνου όπως η λευχαιμία, των οποίων τα καρκινικά κύτταρα κυκλοφορούν μέσα στο αίμα.

Οι περισσότεροι καρκίνοι παίρνουν το όνομά τους από τον τύπο του κυττάρου ή του οργάνου στο οποίο πρωτοεμφανίζονται (πρωτοπαθής όγκος). Αν κάνουν μετάσταση ο νέος όγκος φέρει το ίδιο όνομα με τον αρχικό. Ορισμένοι

⁷ Μετάσταση ονομάζεται η μεταφορά κακοήθων κυττάρων από μια τοποθεσία σε μία άλλη, η οποία δεν βρίσκεται σε επαφή με την προηγούμενη.

όγκοι παίρνουν το όνομά τους από τον επιστήμονα που τους ανακάλυψε (π.χ. Hodgkin, Brenner).

3.4 Εξατομικευμένη Θεραπεία του καρκίνου

Η πρόγνωση, όσο και η έγκαιρη διάγνωση αλλά κυρίως η θεραπεία του καρκίνου αποτελούν προκλήσεις για τους σύγχρονους ερευνητές. Και αυτό όχι μόνο εξαιτίας της ποικιλίας των μορφών του αλλά κυρίως διότι οι μεταλλάξεις συμβαίνουν τυχαία και έτσι κάθε περίπτωση καρκίνου μπορεί να χαρακτηρίζεται από τον δικό της, μοναδικό συνδυασμό μεταλλάξεων [5]. Συνεπώς, φαίνεται ότι καμιά θεραπεία δεν θα είναι αποτελεσματική στο σύνολο των ασθενών. Επιπλέον, ο καρκίνος γενικά δεν ανιχνεύεται προτού ο πρωτοπαθής όγκος αποκτήσει διάμετρο τουλάχιστον 1cm, οπότε αποτελείται ήδη από εκατοντάδες κύτταρα με γενετική ποικιλότητα, τα οποία συχνά έχουν ήδη αρχίσει να μεθίστανται. Έτσι το ερευνητικό ενδιαφέρον εστράφη προς την ανάπτυξη μεθόδων βάσει των οποίων θα είναι δυνατή η εφαρμογή εξατομικευμένης θεραπείας του καρκίνου. [1]

3.5 Η χρήση της Βιοπληροφορικής στην μελέτη του καρκίνου

Όσο απλή και αν παρουσιάζεται η διαδικασία της καρκινογένεσης, ο καρκίνος είναι μία πολυσύνθετη νόσος. Το πλήθος και η τυχαιότητα των μεταλλάξεων που οδηγούν στην εμφάνιση της, και το πλήθος των γονιδίων στα οποία συμβαίνουν, κάνουν το θέμα εξαιρετικά περίπλοκο. Οι μεταλλάξεις αυτές μπορεί να συμβαίνουν είτε στο γονιδίωμα (genome) είτε στο (μεταγράφημα) (transcriptome) ή και να σχετίζονται με το (πρωτεϊνωμα) (proteome) [1].

Ωστόσο η κατανόηση της παθογένειας του καρκίνου σχετίζεται άμεσα με την ταυτοποίηση των γονιδίων που η μετάλλαξη τους συνεπάγεται εμφάνιση καρκίνου αλλά και των συγκεκριμένων μεταλλάξεων, στόχος που φαντάζει απραγματοποίητος, ενώ με τη χρήση της Βιοπληροφορικής ρεαλιστικός.

Η Βιοπληροφορική έδωσε πραγματικά «φτερά» στην έρευνα κατά του καρκίνου παρέχοντας τα κατάλληλα εργαλεία και την υπολογιστική ισχύ ώστε να είναι δυνατή η αποθήκευση, η επεξεργασία και η ανάλυση του τεράστιου πλήθους δεδομένων που σχετίζονται με την νόσο του καρκίνου.

Στη συνέχεια αυτής της διπλωματικής θα προσπαθήσουμε να παρουσιάσουμε σφαιρικά τις μεθόδους της Βιοπληροφορικής που ιδιαίτερος συνεισφέρουν ή στις οποίες θα μπορούσαμε να πούμε ότι στηρίζεται η μελέτη του καρκίνου.

Βιβλιογραφία

- [1] Gavin J. Gordon Bioinformatics in Cancer and Cancer Therapy, Humana Press
- [2] Emanuel Rubin Essential Pathology, Third Edition, Lippincott Williams & Wilkins
- [3] R.J. Trent Μοριακή Ιατρική Εισαγωγικές Έννοιες, μετάφραση-Επιμέλεια Αλέξανδρος Α.Κώτσης
Ιατρικές εκδόσεις Π. Χ. Πασχαλίδης
- [4] Thomas D. Gelehrter Francis S. Collins David Ginsburg Αρχές Ιατρικής Γενετικής Ιατρικές
Εκδόσεις Πασχαλίδης
- [5] Alberts, Bray, Hopkin, Johnson, Lewis, Raff, Roberts, Walter Βασικές Αρχές κυτταρικής Βιολογίας
Ιατρικές Εκδόσεις Π.Χ.Πασχαλίδης

4

Ακολουθιακές Μέθοδοι

4.1 Εισαγωγή

Με την ολοκλήρωση του προγράμματος του ανθρώπινου γονιδιώματος ένα πλήθος δεδομένων έδωσε τροφή στην έρευνα και αποτέλεσε πηγή πληροφοριών. Το πλήθος των δεδομένων προς επεξεργασία οδήγησε στην ανάπτυξη μεθόδων υψηλής απόδοσης. Στο παρόν κεφάλαιο θα παρουσιάσουμε μια κατηγορία τέτοιων βιοπληροφοριακών μεθόδων, τις Ακολουθιακές Μεθόδους.

4.2 Ακολουθιακές μέθοδοι

Οι ακολουθιακές μέθοδοι ή αλλιώς μέθοδοι σύγκρισης ακολουθιών (DNA sequencing) βασίζονται στην απεικόνιση των νουκλεϊκών και πρωτεϊνικών ακολουθιών ως συμβολοσειρές. Εργαλεία τους είναι οι αλγόριθμοι σύγκρισης ακολουθιών, η δημιουργία βάσεων δεδομένων και οι αλγόριθμοι ταχείας αναζήτησης μέσα σε αυτές. Είναι βιοπληροφοριακές μέθοδοι υψηλής απόδοσης και χάρη σε αυτές οι γνώσεις μας γύρω από βιολογικά θέματα γενικότερα αλλά και ειδικότερα για τον καρκίνο έχουν πολλαπλασιαστεί.

4.3 Αλγόριθμοι σύγκρισης ακολουθιών

Ένα από τα βασικά εργαλεία που χρησιμοποιούν οι ακολουθιακές μέθοδοι είναι οι αλγόριθμοι σύγκρισης ακολουθιών.

4.3.1 Αλγόριθμοι Ταιριάσματος Προτύπου

Δοθείσης μιας συμβολοσειράς S n χαρακτήρων και ενός προτύπου αναγνώρισης P m χαρακτήρων, θέλουμε να δούμε αν το P περιέχεται μέσα στην S και αν αυτό συμβαίνει, σε ποια θέση.

Κριτήριο αποδοτικότητας είναι ο αριθμός των συγκρίσεων που πρέπει να γίνουν.

Υπάρχουν δύο κατηγορίες τέτοιων αλγορίθμων: οι *Αλγόριθμοι Ακριβούς Ταιριάσματος Προτύπου* και οι *Αλγόριθμοι Προσεγγιστικού Ταιριάσματος Προτύπου*.

Στο ακριβές ταίριασμα προτύπου ενδιαφερόμαστε να εντοπίσουμε όλες τις εμφανίσεις ενός δοσμένου προτύπου (μοτίβου) P (“δομημένου” ή “μη-δομημένου”) σε μια συμβολοσειρά (βιολογική αλληλουχία) S . Ενώ, στο προσεγγιστικό ταίριασμα προτύπου έχουμε: Για ένα κείμενο S , ένα μοτίβο P , μια παράμετρο k και μια συνάρτηση ομοιότητας $d()$, εντόπισε τις θέσεις i, j στο κείμενο, έτσι ώστε $d(P, S_{i:j}) \leq k$

Αλγόριθμοι ακριβούς ταίριασματος είναι η απλοϊκή μέθοδος (naive method), οι αλγόριθμοι Boyer-Moore, Knuth-Morris-Pratt, Brute-force, Shift Or/Shift And και το αυτόματο Aho-Corasick

4.3.2 Αλγόριθμοι Ευθυγράμμισης Ακολουθιών – Sequence Alignment Algorithms

Ευθυγράμμιση (ή στοίχιση) ακολουθιών είναι η διαδικασία σύγκρισης τους για εύρεση ατομικών χαρακτήρων ή προτύπων χαρακτήρων με την ίδια σειρά στις δύο (ή περισσότερες) ακολουθίες. Πανομοιότυποι χαρακτήρες τοποθετούνται στην ίδια στήλη ενώ μη όμοιοι χαρακτήρες μπορούν να τοποθετηθούν είτε στην ίδια στήλη δηλώνοντας mismatch ή απέναντι από κενό. Υπάρχουν παραπάνω από μία δυνατές ευθυγραμμίσεις. Η βέλτιστη λύση πρέπει να ελαχιστοποιεί τις διαφορές ανάμεσα στις δυο ακολουθίες ή διαφορετικά να μεγιστοποιεί τη συνάρτηση ομοιότητας. Η διαδικασία αυτή στηρίζεται σε πίνακες που βαθμολογούν τις ομοιότητες (matches) και διαφορές (mismatches) μεταξύ διαδοχικών συμβόλων. Τέτοιου τύπου πίνακες είναι οι Dayhoff PAM, BLOSUM κτλ.

Για πολύ λίγες ακολουθίες γνωρίζουμε τη δομή και τη λειτουργία τους. Η ευθυγράμμιση μας παρέχει λειτουργικές, δομικές και εξελικτικές πληροφορίες. Μπορούμε να ευθυγραμμίσουμε δυο ακολουθίες και αν είναι αρκετά όμοιες πιθανόν να έχουν τον ίδιο πρόγονο, επίσης πιθανόν να έχουν την ίδια δομή και λειτουργία οι οργανισμοί. Αν για μία από τις ακολουθίες γνωρίζουμε δομή και λειτουργία και την έχουμε ευθυγραμμίσει με μια άγνωστη ακολουθία μπορούμε να εξαγάγουμε παρόμοια συμπεράσματα και για την άγνωστη. Με την ίδια λογική είναι δυνατός και ο εντοπισμός μεταλλάξεων στα γονίδια, πράγμα πολύ σημαντικό, ιδιαίτερα για την νόσο του καρκίνου.

Μπορούμε να έχουμε ολική ευθυγράμμιση (global alignment) ή τοπική ευθυγράμμιση (local alignment). Στην ολική ευθυγράμμιση, η ευθυγράμμιση γίνεται σε όλο το πεδίο της ακολουθίας και συμπεριλαμβάνει όσα περισσότερα matching ζευγάρια μπορεί. Περιοχές με υψηλή τοπική ομοιότητα αγνοούνται για να επιτευχθεί μεγαλύτερη ολική βαθμολογία. Χαρακτηριστικός αλγόριθμος ολικής ευθυγράμμισης είναι ο Needleman & Wunsch. Στην τοπική ευθυγράμμιση αναζητούμε περιοχές τοπικής ομοιότητας. Η ευθυγράμμιση αυτή σταματά στις περιοχές που είναι ταυτόσημες ή που έχουν μεγάλη ομοιότητα και δίδεται μεγαλύτερη προτεραιότητα στην εύρεση τέτοιων περιοχών δηλαδή προτύπων(patterns) χαρακτήρων παρά για την εύρεση ατομικών ταυτίσεων χαρακτήρων. Αλγόριθμος τοπικής ευθυγράμμισης είναι ο Smith-Waterman.

4.3.3 Αλγόριθμοι Ευθυγράμμισης Πολλαπλών Ακολουθιών

Η ιδέα της ευθυγράμμισης περισσότερων των 2 ακολουθιών οδήγησε στην ανάπτυξη αυτών των αλγορίθμων και αποτελεί φυσική γενίκευση της ευθυγράμμισης 2 ακολουθιών.

Οι αλγόριθμοι FASTA, BLAST και CLUSTALw είναι οι χαρακτηριστικότεροι αλγόριθμοι ευθυγράμμισης πολλαπλών ακολουθιών. Η ταχύτητά τους (περίπου 50 φορές ταχύτεροι από τους αλγόριθμους ολικής ευθυγράμμισης που προαναφέραμε) τους καθιστά και ένα σπουδαίο εργαλείο ταχείας αναζήτησης βιολογικών βάσεων δεδομένων.

Γενικότερα η ευθυγράμμιση Πολλαπλών ακολουθιών χρησιμοποιείται για

- 1) την αναγνώριση και αναπαράσταση πρωτεϊνικών οικογενειών και υπεροικογενειών
- 2) στην αναπαράσταση των χαρακτηριστικών που μεταφέρονται στις ακολουθίες DNA ή στις πρωτεϊνικές ακολουθίες και
- 3) στην αναπαράσταση της εξελικτικής ιστορίας (φυλογενετικά δέντρα) από ακολουθίες DNA ή πρωτεϊνών.

4.3.3.1 Ο αλγόριθμος FASTA

Ο FASTA ([6], [7], [8]) είναι ευριστικός αλγόριθμος τοπικής ευθυγράμμισης και μάλιστα ο πρώτος ευριστικός (heuristic) αλγόριθμος ευρείας χρήσεως αναζήτησης ομοιοτήτων σε βιολογικές βάσεις. Ευριστικός είναι ο αλγόριθμος που βρίσκει λύση γρήγορα και εύκολα χωρίς όμως να εγγυάται ότι είναι και η βέλτιστη. Έτσι είναι περισσότερο προσεγγιστικός αλγόριθμος παρά ακριβής. Συνήθως επιτυγχάνεται η εύρεση λύσεως κοντά στην βέλτιστη.

Ο αλγόριθμος FASTA (fast-a) είναι μια συντομογραφία του FAST-ALL και ονομάστηκε έτσι επειδή χρησιμοποιείται για ταχεία (fast) σύγκριση είτε πρωτεϊνικών ακολουθιών είτε νουκλεοτιδικών.

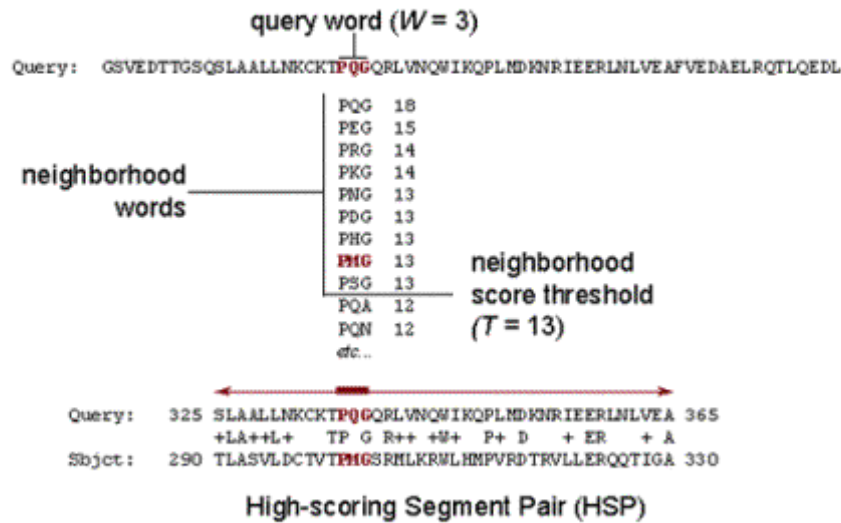
Ψάχνει να βρει τις βέλτιστες τοπικές στοιχίσεις από την ανίχνευση της ακολουθίας για μικρές αντιστοιχίες που ονομάζονται "λέξεις" (words). Αρχικά, υπολογίζεται το σύνολο των τμημάτων ("init1") στα οποία υπάρχουν πολλαπλές "λέξεις" (words). Τα αποτελέσματα διάφορων τμημάτων μπορούν να αθροιστούν για να παράγουν ένα "initn". Μια βελτιστοποιημένη στοίχιση που περιλαμβάνει κενά (gaps) εμφανίζεται ως "opt". Η ευαισθησία και η ταχύτητα της αναζήτησης είναι αντιστρόφως σχετιζόμενες και ελεγχόμενες από τη μεταβλητή "k-tuple" που προσδιορίζει το μέγεθος μιας λέξης (word) [8]

Σαν είσοδο δέχεται δεδομένα σε μορφή FASTA (FASTA format) . Η ακολουθία σε μορφή FASTA ξεκινάει με μια γραμμή περιγραφής (single-line description), και ακολουθείται από γραμμές ακολουθίας με 60 χαρακτήρες για τα αμινοξέα σε κάθε γραμμή. Η γραμμή περιγραφής ξεχωρίζει από τις γραμμές της ακολουθίας επειδή στην αρχή της γραμμής περιγραφής υπάρχει το σύμβολο (">").Περισσότερες πληροφορίες σχετικά με το FASTA υπάρχουν στην ιστοσελίδα <http://www.ebi.ac.uk/Tools/fasta/> και <http://www.ebi.ac.uk/Tools/fasta33/index.html>

4.3.3.2 BLAST - Basic Local Alignment Search Tool

Ο αλγόριθμος BLAST [1][2] χρησιμοποιείται για τη σύγκριση μιας ακολουθίας με μια βάση δεδομένων, έχει αρχικά αναπτυχθεί και διατηρείται από το NCBI (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/BLAST/>). Είναι ένας ευριστικός (heuristic) αλγόριθμος σύγκρισης ακολουθιών βελτιστοποιημένης ταχύτητας που χρησιμοποιείται για να ψάχνει σε βάσεις ακολουθιών την άριστη τοπική στοίχιση με μια αναζήτηση. Η αρχική αναζήτηση γίνεται για μια λέξη μήκους "W" (3 στο blastp) που δίνει αποτελέσματα για τουλάχιστον "T", όταν συγκρίνεται με την ζητούμενη ακολουθία, χρησιμοποιώντας έναν δεδομένο πίνακα υποκαταστάσεων (substitution matrix). Οι επιτυχείς λέξεις που έχουν score T ή μεγαλύτερο επεκτείνονται και προς τις δύο κατευθύνσεις σε μια απόπειρα να παραχθούν στοιχίσεις που να υπερβαίνουν το προκαθορισμένο κατώφλι (threshold) "S". Οι περιοχές που ικανοποιούν αυτή τη συνθήκη ονομάζονται HSP (High-scoring Segment Pair). Η παράμετρος "T" καθορίζει την ταχύτητα και την ευαισθησία της αναζήτησης (Εικόνα 1). Ενώ το BLAST υπολογίζει τις παραμέτρους της EVD (Extreme Value Distribution), από τις οποίες θα υπολογίσει τη στατιστική σημαντικότητα από εξομοιώσεις που έχει

The BLAST Search Algorithm



Εικόνα 1: Αλγόριθμος BLAST- Κατασκευή καταλόγου «λέξεων» που οδηγούν σε «καλές» στοιχίσεις. Για κάθε λέξη (αντίστοιχο του k-tuple) της ακολουθίας με την οποία πραγματοποιείται η αναζήτηση (query) προϋπολογίζονται εκείνες οι λέξεις οι οποίες στοιχιζόμενες έναντι αυτής θα έδιναν score (με βάση το επιλεγμένο σύστημα βαθμονόμησης) μεγαλύτερο ή ίσο από την προκαθορισμένη τιμή κατωφλίου T. Αυτές οι k-tuples όταν εντοπισθούν σε ακολουθίες της βάσης δεδομένων δίνουν πιθανά σημεία έναρξης μιας καλής τοπικής στοιχίσης. Στην ουσία, η διαδικασία αυτή είναι γενίκευση της διαδικασίας εντοπισμού «καλών» διαγωνίων του αλγορίθμου FASTA. Το πλεονέκτημα αυτής της μεθόδου είναι ότι επιτρέπονται στο αρχικό αυτό στάδιο αντιστοιχίσεις και μεταξύ τμημάτων των ακολουθιών που δεν εμφανίζουν απόλυτη ταύτιση.

πραγματοποιήσει από πριν, το FASTA τις υπολογίζει από όλες τις άλλες ακολουθίες της βάσης δεδομένων και για αυτό το λόγο είναι και πιο αργό.

Η περιοχή αναζήτησης ('Search' box) του BLAST δέχεται διαφορετικούς τύπους μορφοποιημένης εισόδου και αυτόματα καθορίζει τη μορφή. Οι τρεις μορφοποιήσεις (formats) με τις οποίες μπορεί να εισαχθεί μια ακολουθία και να γίνει η αναζήτηση είναι FASTA format, σκέτη ακολουθία (bare sequence) δηλαδή μόνο οι γραμμές της ακολουθίας χωρίς τη FASTA γραμμή ορισμού και με τη χρήση προσδιοριστών (identifiers).

Υπάρχουν πολλές παραλλαγές του προγράμματος BLAST [3], οι οποίες εκτελούν διαφορετικές εργασίες.

- blastn - συγκρίνει μια νουκλεοτιδική ακολουθία (DNA) με μια βάση νουκλεοτιδικών ακολουθιών (DNA). Η αναζήτηση γίνεται και στις δύο αλυσίδες. Είναι ένα πρόγραμμα βελτιστοποιημένης ταχύτητας, όχι όμως και ευαισθησίας.

- **blastp** - συγκρίνει την ζητούμενη αμινοξική ακολουθία με μια βάση πρωτεϊνικών ακολουθιών (σύγκριση πρωτεΐνης με πρωτεΐνες).
- **blastx** - συγκρίνει μια άγνωστη νουκλεοτιδική ακολουθία (DNA) μεταφρασμένη σε όλα τα πλαίσια ανάγνωσης (reading frames) με μια βάση πρωτεϊνικών ακολουθιών του NCBI. Χρησιμοποιείται για την εύρεση πιθανών μεταφρασμένων πρωτεϊνικών προϊόντων μιας άγνωστης νουκλεοτιδικής ακολουθίας.
- **tblastn** - συγκρίνει την ζητούμενη πρωτεϊνική ακολουθία με μια βάση νουκλεοτιδικών ακολουθιών (DNA) του NCBI που μεταφράζεται δυναμικά σε όλα τα πλαίσια ανάγνωσης (reading frames).
- **tblastx** - μετατρέπει μια νουκλεοτιδική ακολουθία (DNA) σε μια πρωτεϊνική ακολουθία σε όλα τα πλαίσια ανάγνωσης (reading frames) και μετά τη συγκρίνει με μια βάση νουκλεοτιδικών ακολουθιών του NCBI η οποία έχει μεταφραστεί σε όλα τα πλαίσια ανάγνωσης (reading frames).
- **BLAST2** - Ονομάζεται εξελιγμένο (advanced) BLAST. Εκτελεί στοιχίσεις που περιέχουν κενά (gapped alignments).
- **MEGABLAST** - είναι ένα πρόγραμμα που χρησιμοποιεί έναν πλεονεκτικό ("greedy algorithm") αλγόριθμο [5], για αναζήτηση στοίχισης νουκλεοτιδικών ακολουθιών. Χρησιμοποιείται για στοίχιση ακολουθιών με μικρές διαφορές και είναι 10 φορές γρηγορότερο από παρόμοια προγράμματα. Ενδείκνυται για σύγκριση μεταξύ μεγάλων ακολουθιών.
- **PSI-BLAST** - (Position Specific Iterated BLAST) [3] χρησιμοποιεί σταθερή αναζήτηση, στην οποία οι ακολουθίες που θα βρεθούν στον πρώτο γύρο αναζητήσεων χρησιμοποιούνται για να χτίσουν ένα αποτελεσματικό μοντέλο για τους επόμενους κύκλους αναζητήσεων.
- **PHI-BLAST** - (Pattern Hit Initiated BLAST) συνδυάζει το ταίριασμα ενός πρότυπου φυσιολογικής έκφρασης με μια συγκεκριμένη θέση που επαναλαμβάνεται στην πρωτεϊνική ακολουθία.
- **RPS-BLAST** - συγκρίνει μια πρωτεϊνική ακολουθία ως προς την βάση Conserved Domain Database (CD-Search).

4.3.3.3 CLUSTALw

Το CLUSTALW είναι το πιο διαδεδομένο πρόγραμμα ευθυγράμμισης πολλαπλών βιολογικών ακολουθιών. Το πρόγραμμα χρησιμοποιεί έναν ιδιαίτερα πολύπλοκο αλγόριθμο προοδευτικής ευθυγράμμισης (progressive alignment) για να κάνει σταδιακή ευθυγράμμιση πολλαπλών πρωτεϊνικών ή νουκλεοτιδικών (DNA) ακολουθιών. Όλες οι ακολουθίες πρέπει να είναι ένα αρχείο, η μία μετά την άλλη. Το πρόγραμμα CLUSTALW βρίσκεται στο EBI (European Bioinformatics Institute) (<http://www.ebi.ac.uk/clustalw/>).

4.4 Βιολογικές Βάσεις Δεδομένων

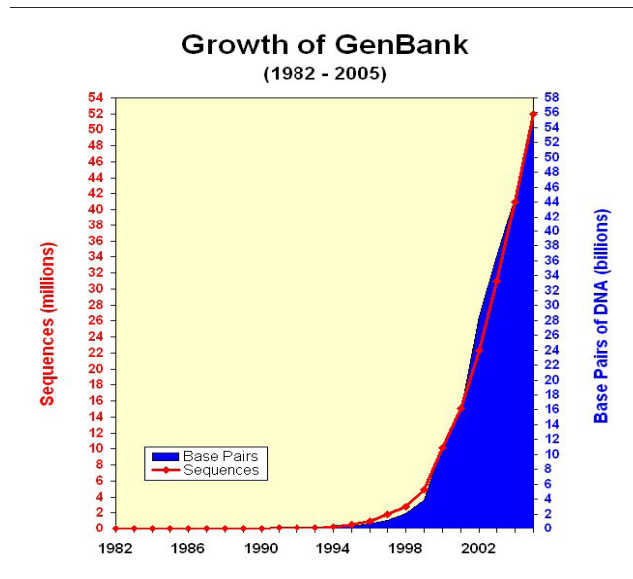
Η συλλογή ακολουθιών DNA με μεθόδους της Μοριακής Βιολογίας (χρήση ειδικών ενζύμων, περιοριστικών ενδονουκλεασών, βακτηρίων ξενιστών, δυνατότητα αντιγραφής, μεταγραφής και μετάφρασης σε δοκιμαστικούς σωλήνες κτλ) αποτελεί διαδικασία ρουτίνας για τους βιολόγους σήμερα. Ωστόσο, η ψηφιοποίηση τους έδωσε τεράστιες δυνατότητες αποθήκευσης, επεξεργασίας και ανάλυσης τους.

Με την απεικόνιση αλληλουχιών σε μορφή αναγνώσιμη από τους υπολογιστές δημιουργήθηκαν βάσεις βιολογικών δεδομένων και συγκεκριμένα *Βάσεις Δεδομένων Νουκλεοτιδικών Ακολουθιών* και *Βάσεις Δεδομένων Πρωτεϊνικών Ακολουθιών*. Αναπτύχθηκαν τόσο Πρωτογενείς βάσεις δεδομένων όπου δίνουν τον χώρο και τον μηχανισμό για την κατάθεση και την πρόσβαση σε πειραματικά δεδομένα που αφορούν αλληλουχίες, όσο και Δευτερογενείς που στηρίζονται σε δεδομένα των πρωτογενών βάσεων αλλά περιέχουν επιπλέον και σχολιασμό ο οποίος δεν έχει (απαραίτητα) πειραματική υποστήριξη. Η ανάπτυξη τους πραγματοποιήθηκε αρχικά σε τρεις μεριές του πλανήτη. Στην Αμερική από το National Center for Biotechnology Information – NCBI (<http://www.ncbi.nlm.nih.gov/>) στην Ιαπωνία από το National Institute of Genetics (<http://www.nig.ac.jp/index-e.html>) και συγκεκριμένα από την DNA Data Bank of Japan (<http://www.ddbj.nig.ac.jp/Welcome-e.html>) και στην Ευρώπη από το European Molecular Biology Laboratory- EMBL (<http://www.embl.org/>). Σήμερα πλέον έχουμε συνεργασία των τριών ερευνητικών κέντρων. Σημαντικό για την έρευνα είναι η ελεύθερη πρόσβαση μέσω των αντίστοιχων ιστοσελίδων στις βάσεις δεδομένων και στα εργαλεία αναζήτησης και σύγκρισης ακολουθιών.

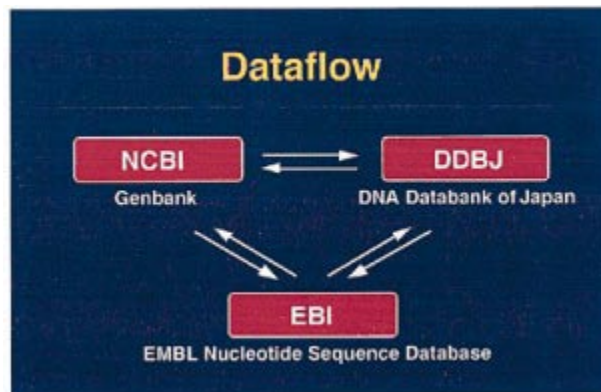
4.4.1 Βάσεις Δεδομένων Νουκλεοτιδικών Ακολουθιών

- 1) GenBank – NCBI (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>)
- 2) EMBL DataBank (<http://www.ebi.ac.uk/embl/>)
- 3) DDBJ (<http://www.ddbj.nig.ac.jp/Welcome-e.html>)

Ενώ πλέον υπάρχει συνεργασία στα πλαίσια της International Nucleotide Sequence Database Collaboration – INSDC (<http://www.insdc.org/page.php?page=home>).



Εικόνα 2: <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>

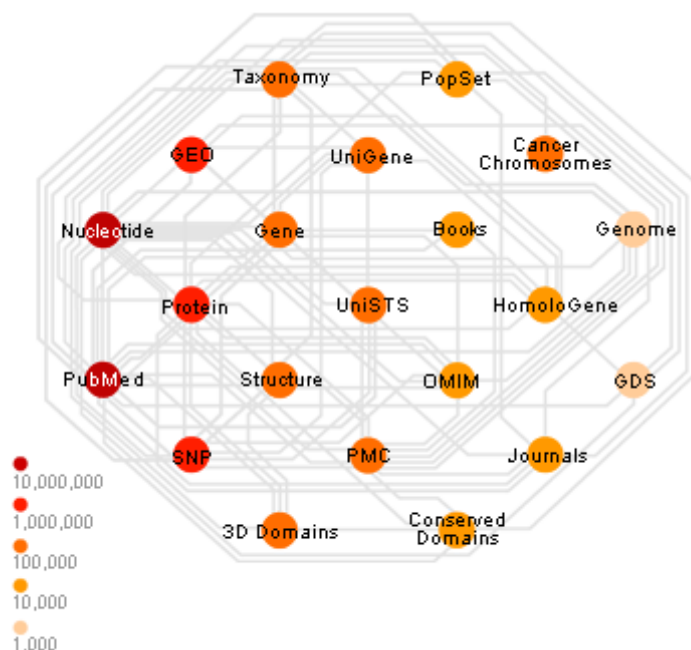


Εικόνα 3: Επικοινωνία μεταξύ των τριών βάσεων δεδομένων και κοινός μηχανισμός αμοιβαίας ενημέρωσης. <http://www.ebi.ac.uk/embl/Contact/dataflow.gif>

4.4.2 Βάσεις Δεδομένων Πρωτεϊνικών Ακολουθιών

Στις Βάσεις δεδομένων πρωτεϊνικών ακολουθιών τα πράγματα είναι κάπως πιο περίπλοκα γιατί κάθε μορφής δεδομένα που αποτελούν προϊόν έρευνας ή υπολογιστικής ανάλυσης κατατίθεται σε βάσεις δεδομένων. Έτσι έχουν δημιουργηθεί ΒΔ σχετικά με την πρωτοταγή δομή των πρωτεϊνών, την δευτεροταγή, με ομολογίες πρωτεϊνών και πολλές άλλες.

Ένας κατάλογος Βάσεων πρωτεϊνικών δεδομένων διατίθεται στην ηλεκτρονική διεύθυνση <http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz> ενώ χαρακτηριστική ΒΔ είναι η UniProt (Universal Protein Resource) <http://www.ebi.ac.uk/uniprot/>. Άλλες βάσεις Δεδομένων βρίσκονται στην διεύθυνση <http://www.ncbi.nlm.nih.gov/Database/>.



Εικόνα 3: Ο χάρτης των βάσεων του NCBI όπως δίνεται στην ιστοσελίδα <http://www.ncbi.nlm.nih.gov/Database/>



http://www.ebi.ac.uk/Information/databases_sitemap.html : Όλες οι βάσεις που διατίθενται από την EBI αλφαβητικά.

Νουκλεοτιδικές βάσεις				
A/A	Όνομασία	Ακρωνύμιο	Ιστοσελίδα	Σύντομη Περιγραφή
1.	Συνολικού Ανθρώπινου Γονιδιώματος	Nucleotide	http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide	Περιλαμβάνει το σύνολο του ανθρώπινου (Homo Sapiens) γονιδιώματος σε ακολουθίες DNA και σχόλια.
2.	Expressed Sequence Tags database	dbEST	http://www.ncbi.nlm.nih.gov/dbEST/index.html	Η βάση δεδομένων EST (dbEST) περιλαμβάνει ESTs από ένα πλήθος οργανισμών και πληροφορίες γύρω από αυτά. Τα ESTs είναι μικρά τμήματα γονιδίων (μήκους 500- 800 νουκλεοτιδίων) τα οποία μεταγράφονται. Τα ESTs παράγονται με τη βοήθεια του υβριδισμού σε cDNA βιβλιοθήκες.
3.	database of "Sequence	dbSTS	http://www.ncbi.nlm.nih.gov/dbSTS/index.html	Περιεχόμενο της συγκεκριμένης βάσης

	Tagged Sites"			είναι Sequence Tagged Sites με τις απαραίτητες ακολουθιακές και χαρτογραφικές πληροφορίες γύρω από αυτά. Τα Sequence Tagged Sites είναι μικρά τμήματα ακολουθιών DNA (200 με 500 ζ.β.) όπου το καθένα υπάρχει μία και μοναδική φορά στο γονιδίωμα και η θέση του στο γονιδίωμα όπως και η ακολουθία των βάσεων του είναι γνωστές.
4.	Single Nucleotide Polymorphism database	dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/ και http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp	Περιλαμβάνει πολυμορφισμούς μονού νουκλεοτιδίου (SNPs) γονιδίων.
5.	Genome Survey Sequences Database	dbGSS	http://www.ncbi.nlm.nih.gov/dbGSS/	Παρόμοια με την dbEST όμως το σύνολο των ακολουθιών που περιλαμβάνει προέρχονται από υβριδισμό γενωμικού DNA και όχι μέσω cDNA βιβλιοθηκών.
6.	Genome database	Genome	http://www.ncbi.nlm.nih.gov/sites/entrez	Είναι μία μεγάλη βάση δεδομένων η οποία περιλαμβάνει 6 μεγάλες κατηγορίες οργανισμών. Περιέχει ποικιλία γονιδιωμάτων, ολόκληρα χρωμοσώματα, χαρτογραφήσεις ακολουθιών και άλλων πληροφοριών.
7.	Gene database	Gene	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene	Περιλαμβάνει ένα πλήθος γονιδίων και παρέχει δυνατότητες επιλογής μορφής αναζήτησης γονιδίου.
8.	Homologous Genes database	HomoloGene	http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene	Ουσιαστικά αποτελεί μία μηχανή αναζήτησης ομολόγων γονιδίων ανάμεσα σε γονιδιώματα
9.	Genome Project database	Genome Project	http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeproj	
10.	RefSeqGene Project	RefSeqGene	http://www.ncbi.nlm.nih.gov/projects/RefSeq/RSC/	Η RefSeqGene αποτελεί υποσύνολο της RefSeq. Περιλαμβάνει γενομικές ακολουθίες καλώς

				ορισμένων γονιδίων, οι οποίες μπορούν να χρησιμοποιηθούν ως αναφορά σε διάφορες ερευνητικές εργασίες. Είναι αρκετά χρήσιμη σε μελέτες που αφορούν μεταλλάξεις και πολυμορφισμούς.
11.	GenBank	GenBank	http://www.ncbi.nlm.nih.gov/Genbank/	Είναι η βάση δεδομένων γονιδιακών ακολουθιών του αμερικανικού National Institutes of Health (NIH). Μια ιδιαίτερα σημαντική συλλογή όλων των ακολουθιών DNA που είναι διαθέσιμες στο ευρύ κοινό.
12.	Mammalian Gene Collection	MGC	http://mgc.nci.nih.gov/	Ο στόχος της MGC (συλλογή γονιδίων θηλαστικών) είναι να παρέχει στους ερευνητές πρόσβαση χωρίς περιορισμούς, σε κλώνους cDNA γονιδίων ανθρώπων, ποντικών και αρουραίων. Οι cDNA κλώνοι που παρέχονται είναι ταυτοποιημένοι ακολουθιακά, πλήρους μήκους κωδικοποιημένοι κατά πρωτεΐνες [full-length protein-coding (FL-CDS)]. Το 2005 το πρόγραμμα προσέθεσε τα cDNAs της αγελάδας, τα οποία προέκυψαν από το πρόγραμμα Genome Canada.
13.	Population Set	PopSet	http://www.ncbi.nlm.nih.gov/sites/entrez?db=popset	Το PopSet είναι ένα σύνολο ακολουθιών DNA που έχουν συλλεχθεί προκειμένου να αναλυθεί η σχετικότητα της εξέλιξης ενός πληθυσμού. Ο πληθυσμός αυτός μπορεί να προέρχεται από διαφορετικά μέλη του ίδιου γένους ή από οργανισμούς διαφορετικού είδους.

14.	Probe Database	Probe	http://www.ncbi.nlm.nih.gov/sites/entrez?db=probe	<p>Η βδ Probe είναι πλήρως προσβάσιμη από το κοινό και περιλαμβάνει ανιχνευτές (probes) νουκλειικού οξέως. Οι ανιχνευτές είναι μόρια που συμπληρώνουν ή υβριδίζουν συγκεκριμένες ακολουθίες mRNA ή DNA και επομένως μπορούν να χρησιμοποιηθούν για τον εντοπισμό γονιδίων και επακόλουθα πληροφοριών γύρω από αυτό (http://www.ncbi.nlm.nih.gov/Web/Newsltr/V14N2/probe.html). Είναι επίσης σχεδιασμένοι για χρήση σε ένα ευρύ φάσμα των εφαρμογών της βιοιατρικής έρευνας. Η βάση διαθέτει και άλλες πληροφορίες όπως για την αποτελεσματικότητα της έρευνας και ακολουθιακές ομοιότητες που έχουν υπολογιστεί.</p>
15.	Short Read Archive	SRA	http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=show&f=mail&m=main&s=main	<p>Η SRA αποθηκεύει μη επεξεργασμένα ακολουθιακά δεδομένα τα οποία προκύπτουν από ακολουθιακές πλατφόρμες που επωνομάζονται "επόμενης" γενιάς όπως οι εξής: Roche 454 GS System, Illumina Genome Analyzer, Applied Biosystems SOLiD System, Helicos Heliscope, Complete Genomics, κ.ο.κ.</p>
16.	Third Party Annotation	TPA	http://www.ncbi.nlm.nih.gov/Genbank/TPA.html	<p>Είναι μία βδ που έχει σχεδιαστεί για την αποθήκευση πειραματικών ή συμπερασματικών αποτελεσμάτων που υποστηρίζουν την θεώρηση των ακολουθιακών δεδομένων η οποία προέκυψε από τα</p>

				αρχικά δεδομένα της GenBank
17.	Online Mendelian Inheritance in Man	OMIM	http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim	Είναι μία περιεκτική έγκυρη και χρονική επιτομή του ανθρώπινου γονιδιώματος και των γενετικών φαινοτύπων. Η OMIM εμπεριέχει πληροφορίες για όλες τις γνωστές μεντελικές δυσλειτουργίες και πάνω από 12000 γονίδια. Επικεντρώνεται στη σχέση μεταξύ φαινοτύπου και γονοτύπου. Ενημερώνεται καθημερινά. Οι εγγραφές της περιέχουν συνοπτικούς συνδέσμους με άλλες πηγές πληροφόρησης για το γονίδιομα.
18.	Cancer Chromosomes		http://www.ncbi.nlm.nih.gov/sites/entrez?db=cancerchromosomes	Τρεις βάσεις δεδομένων NCI/NCBI SKY/M-FISH & CGH Database, the NCI Mitelman Database of Chromosome Aberrations in Cancer, and the NCI Recurrent Aberrations in Cancer έχουν ενσωματωθεί στην Cancer Chromosomes.
Βάσεις Δεδομένων Πρωτεϊνών				
19.	Reference Sequence project collection	RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/	Η συλλογή RefSeq παρέχει ένα κατανόησιμο, ολοκληρωμένο, μη επαναλαμβανόμενο και καλά ορισμένο σύνολο ακολουθιών, συμπεριλαμβανομένου του γονιδιακού DNA, του μεταγραφήματος και πρωτεϊνών.
20.	Molecular Modeling DataBase	MMDB	http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml	Η συγκεκριμένη βδ περιλαμβάνει Τρισδιάστατες δομές βιολογικών μακρομορίων που έχουν οριστεί πειραματικά και αποτελεί υποσύνολο της βδ Protein Data Bank(http://www.rcsb.org/p)

				db/home/home.do) η οποία περιλαμβάνει και δομές θεωρητικά μόνο τεκμηριωμένες.
21.		3D Domains	http://www.ncbi.nlm.nih.gov/sites/entrez?db=domains	<p>Οι Τρισδιάστατοι Χώροι (3D Domains) είναι δομικά συμπαγείς χώροι οι οποίοι αναγνωρίζονται αυτόματα από τη MMDB (μακρόμοριακή βάση τρισδιάστατης δομής του Entrez). Αποτελούν μονάδες σύγκρισης για τους υπολογισμούς δομικής γειτνίασης χρησιμοποιώντας τον αλγόριθμο VAST. Οι σύνδεσμοι στον VAST ή οι περιοχές τρισδιάστατων χώρων παρουσιάζουν τρισδιάστατους χώρους (καθώς και ολοκληρωμένες αλυσίδες πολυπεπτιδίων) με παρόμοιες τρισδιάστατες δομές.</p>
22.	Protein Clusters Database	Protein Clusters	http://www.ncbi.nlm.nih.gov/sites/entrez?db=proteinclusters	<p>Αποτελεί μία συλλογή clusters (δεσμών-ομάδων) πρωτεϊνών τα οποία αποτελούνται από πρωτεϊνικές ακολουθίες αναφοράς(RefSeq), οι οποίες κωδικοποιούνται από ολόκληρα γονιδιώματα. Αυτή η βάση δεδομένων περιέχει εξίσου clusters τα οποία έχουν αυτόματα δημιουργηθεί και δεν περιλαμβάνονται σχόλια ή άλλες πληροφορίες για αυτά (non-curated clusters) και clusters για τα οποία περιλαμβάνονται ένα πλήθος πληροφοριών και αποτελέσματα επεξεργασίας (curated clusters)</p>

23.	Taxonomy Database	Taxonomy	http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy	Σε αυτή την βάση δεδομένων είναι καταγεγραμμένα τα ονόματα όλων των οργανισμών των οποίων υπάρχει τουλάχιστον μία νουκλεοτιδική ή πρωτεϊνική ακολουθία στις γενετικές βάσεις δεδομένων.
24.	Clusters of Orthologous Groups of proteins	COG	http://www.ncbi.nlm.nih.gov/COG/	Η βδ COG περιλαμβάνει clusters συγκεκριμένων ομάδων πρωτεϊνών τα οποία προέκυψαν από την σύγκριση πρωτεϊνικών ακολουθιών οι οποίες κωδικοποιούνται από ολόκληρα γονιδιώματα και αντιπροσωπεύουν πρωτεύουσας σημασίας γενετικές γενεαλογίες. Κάθε cluster αποτελείται από ξεχωριστές πρωτεΐνες ή από ομάδες ανακολουθιών από τουλάχιστον 3 γενεές.
Βάσεις Δεδομένων που προκύπτουν με τη βοήθεια Μικροδιατάξεων (Microarrays)				
25.		ArrayExpress	http://www.ebi.ac.uk/microarray-as/ae/	Η βάση αυτή ουσιαστικά αποτελεί χώρο αποθήκευσης δεδομένων που σχετίζονται με την έκφραση των γονιδίων και προκύπτουν με βάση την τεχνολογία των μικροδιατάξεων
26.		ArrayExpress Warehouse	http://www.ebi.ac.uk/microarray-as/aew/	Μια επιπρόσθετη πολύτιμη βάση δεδομένων εκφράσεων γονιδιακών προφίλ.
27.		MEDLINE	http://www.ebi.ac.uk/Databases/MEDLINE/medline.html	Το MEDLINE είναι η πρώτη βιβλιογραφική βάση δεδομένων του NLM, η οποία καλύπτει τα πεδία της ιατρικής, της νοσηλευτικής, της οδοντιατρικής, της κτηνιατρικής, του

				<p>συστήματος ιατρικής περίθαλψης και των προκλινικών επιστημών. Το MEDLINE περιέχει βιβλιογραφικές παραθέσεις και αποσπάσματα συγγραφέων από περισσότερα από 5000 βιοιατρικά περιοδικά εκδοθέντα στις Η.Π.Α. και σε άλλες 70 χώρες. Τα αρχεία περιλαμβάνουν πάνω από 16 εκατομμύρια παραθέσεις, οι οποίες χρονολογούνται από τα μέσα της δεκαετίας του 60, και ανανεώνονται καθημερινώς. Η κάλυψη είναι παγκόσμια, ωστόσο οι περισσότερες εγγραφές είναι από αγγλόφωνες πηγές ή έχουν αγγλικά αποσπάσματα.</p>
28.		UniProt	<p>http://www.ebi.ac.uk/uniprot/ και http://www.uniprot.org/</p>	<p>Η αποστολή της βδ UniProt είναι να παρέχει στην επιστημονική κοινότητα μία κατανοητή, υψηλής ποιότητας και ελεύθερα προσβάσιμη πηγή πρωτεϊνικών ακολουθιών και λειτουργικών πληροφοριών. Η UniProt αποτελείται από τέσσερα μέρη, το καθένα προορισμένο για διαφορετικές χρήσεις.</p>
29.	Protein knowledgebase	UniProtKB	<p>http://www.uniprot.org/help/uniprotkb</p>	<p>Η βάση πληροφοριών για πρωτεΐνες αποτελείται από δύο μέρη: το Swiss-Prot, το οποίο είναι υπομνηματισμένο με το χέρι και επιθεωρημένο και από το TrEMBL, το οποίο είναι αυτόματα υπομνηματισμένο και δεν είναι επιθεωρημένο.</p>

30.		UniProtKB/ Swiss-Prot	http://au.expasy.org/sprot/relnotes/relstat.html	
31.		UniProtKB/ TrEMBL	http://www.ebi.ac.uk/uniprot/TrEMBLstats/	
32.	UniRef databases	UniRef	http://www.uniprot.org/help/uniref	Οι βάσεις δεδομένων UniRef παρέχουν ομαδοποιημένα σύνολα ακολουθιών από τη βάση UniProt και από επιλεγμένες UniParc εγγραφές, με σκοπό να επιτευχθεί πλήρης κάλυψη του χώρου ακολουθιών ο οποίος παρουσιάζεται από διάφορες πλευρές και έτσι περιλαμβάνονται πλεονάζουσες ακολουθίες όπου η επαναληψή τους δεν είναι εμφανής (δεν περιλαμβάνονται όμως και οι περιγραφές τους).
33.	UniProt Archive	UniParc	http://www.uniprot.org/help/uniparc	Το UniParc είναι μία κατανοητή και μη-πλεονάζουσα βάση δεδομένων η οποία περιέχει τις περισσότερες από τις ελεύθερα διαθέσιμες ακολουθίες πρωτεϊνών στον κόσμο. Οι πρωτεΐνες μπορεί να υπάρχουν σε διαφορετικές πηγές βάσεων δεδομένων και σε πολλαπλά αντίγραφα στην ίδια βάση δεδομένων.
34.		Mutations	http://www.sanbi.ac.za/people/faculty/professors/heikki-lehvaslaiho/	Είναι μία βδ στην οποία ξεκίνησε η καταγραφή μεταλλάξεων από τον Heikki Lehvaslaiho
35.	Sequence Retrieval System	SRS	http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+srsq2+-noSession	Το Sequence Retrieval System (Σύστημα Αποκατάστασης Ακολουθιών) μας παρέχει την δυνατότητα αναζήτησης ποικίλων βιολογικών ακολουθιών όπως και αναζήτησης στις βιβλιογραφικές βάσεις

				δεδομένων που διατίθενται από το EBI.
36.	<i>Stanford Microarray Database</i>		http://smd.stanford.edu/	

Πίνακας 1: Βάσεις Βιολογικών Δεδομένων

4.5 Η χρήση των ακολουθιακών μεθόδων στην μελέτη του Καρκίνου

Η οργάνωση του πλήθους των βιολογικών δεδομένων σε βάσεις δεδομένων σε συνδυασμό με την χρήση αλγορίθμων σύγκρισης αλληλουχιών και ταχείας αναζήτησης βάσεων δεδομένων παρέχει τεράστιες δυνατότητες στην έρευνα του καρκίνου.

Η σύγκριση του γενετικού υλικού καρκινικών κυττάρων με αυτό των υγιών, με ακολουθιακές μεθόδους, κάνει την ταυτοποίηση και τον χαρακτηρισμό των γενετικών αλλαγών που συμβαίνουν κατά τη διάρκεια της εξαλλαγής των κυττάρων από υγιή σε καρκινικά, υλοποιήσιμο στόχο.

Με τον εντοπισμό των συγκεκριμένων γενετικών αλλαγών που οδηγούν στην εμφάνιση της νόσου είναι δυνατή η έγκαιρη διάγνωση της (όπως ήδη έχει πραγματοποιηθεί για τον καρκίνο του παχέος εντέρου) και η δημιουργία φαρμάκων για την αντιμετώπιση της.

Επίσης, η χρήση ακολουθιακών μεθόδων ενδείκνυται για την ανάπτυξη εξατομικευμένης θεραπείας του καρκίνου. Μπορούμε να πούμε πως είναι δυνατός ο εντοπισμός των συγκεκριμένων αλλαγών που οδήγησαν στην εμφάνιση της νόσου σε συγκεκριμένο ασθενή μέσω της σύγκρισης του γενετικού υλικού των καρκινικών του κυττάρων με αυτό των υγιών. Έτσι θα μπορούσαμε να οδηγηθούμε στην δημιουργία των κατάλληλων φαρμάκων αποκλειστικά για αυτόν τον ασθενή.

Ωστόσο αυτό στο οποίο θα λέγαμε πως υστερούν οι ακολουθιακές μέθοδοι είναι η πρόγνωση του καρκίνου. Θα πρέπει να προηγηθούν οι γενετικές αλλαγές προκειμένου να εντοπιστούν, τότε όμως είναι αργά για πρόγνωση.

Ας δούμε συγκεκριμένες μεθόδους που χρησιμοποιούνται:

4.5.1 Expressed Sequence Tags (ESTs) ⁸

Τα ESTs είναι μικρά τμήματα γονιδίων που μεταγράφονται (expressed genes). Αποτελούνται από περίπου 500 έως 800 νουκλεοτίδια και παράγονται με την βοήθεια του υβριδισμού σε cDNA βιβλιοθήκες. Είναι ένας τρόπος να περάσουμε από το σύνολο των γονιδίων (γονιδίωμα) στα γονίδια που μεταγράφονται (μεταγράφημα). Έχουν ήδη δημιουργηθεί βάσεις δεδομένων ESTs (πχ <http://www.ncbi.nlm.nih.gov/dbEST/index.html>).

Τα ESTs χρησιμοποιούνται ευρέως στην έρευνα και κυρίως στην μελέτη γενετικών ασθενειών.

Συγκεκριμένα για τον καρκίνο, ο φαινότυπος της νόσου μελετήθηκε μέσω της σύγκρισης συνόλων ESTs [12][13]. Ένα παράδειγμα είναι η χρήση, των διαθέσιμων εκείνη την εποχή, βάσεων δεδομένων EST για την ηλεκτρονική απεικόνιση των διαφορών μεταξύ των φυσιολογικών και προσβεβλημένων ιστών του προστάτη και την ανάπτυξη δυναμικών μοριακών δεικτών για την ανίχνευση του καρκίνου [14]

4.5.2 Serial Analysis of Gene Expression SAGE

Η μέθοδος σειριακής ανάλυσης της γονιδιακής έκφρασης μπορεί να περιγραφεί απλά ως εξής[15]: Από το σύνολο του μεταγραφώματος (transcript pool) εκφρασμένο σε cDNA πραγματοποιούμε αντίγραφα. Από τα αντίγραφα cDNA αποκόπτουμε με την βοήθεια ενζύμων μικρά τμήματα μήκους 14-21 ζεύγη βάσεων. Συνάπτουμε τα μικρά αυτά τμήματα (tags) το ένα μετά το άλλο δημιουργώντας μία ενιαία αλυσίδα. Αλληλουχούμε (sequencing) την μονόκλωνη αλυσίδα. Οι μικρές ακολουθίες από τις οποίες αποτελείται η αλυσίδα είναι δείκτες που χαρακτηρίζουν συγκεκριμένα τμήματα του μεταγραφώματος. Επομένως η συχνότητα εμφάνισης των συγκεκριμένων tags στην αλυσίδα εκφράζει το πλήθος που θα έχουν τα τμήματα που αντιπροσωπεύουν στο μεταγράφημα (transcriptome).

Η τεχνολογία αυτή έχει χρησιμοποιηθεί προκειμένου να γίνει σύγκριση της έκφρασης γονιδίων υπό διαφορετικές πειραματικές συνθήκες [16][17]. Επίσης η SAGE μπορεί να χρησιμοποιηθεί και για την εξακρίβωση της έκφρασης προηγουμένως άγνωστων γονιδίων καθώς δεν βασίζεται σε κάποια γνωστή ακολουθία. Ωστόσο η μέθοδος απαιτεί διαδικασίες υψηλής απόδοσης και επομένως χαρακτηρίζεται από υψηλό κόστος. Μειονέκτημα της επίσης είναι πως

⁸ <http://www.ncbi.nlm.nih.gov/About/primer/est.html>

αρκετές φορές υπάρχει δυσκολία αναγνώρισης των tags εξαιτίας του πολύ μικρού τους μήκους.

Όσον αφορά τον καρκίνο χρησιμοποιήθηκαν βιβλιοθήκες που προέκυψαν με την μέθοδο **SAGE** και βάσεις δεδομένων ESTs προκειμένου να αναγνωριστούν τα συγκεκριμένα γονίδια τα οποία παρουσιάζουν διαφορετικά επίπεδα έκφρασης στα κύτταρα του μελανώματος και στα φυσιολογικά μελανοκύτταρα [18]. Επίσης η ίδια ομάδα ερευνητών κατάφερε να κατατάξει ασθενείς σε σύνολα σύμφωνα με την ανοσιακή απόκριση που εμφάνιζαν σε πρωτεΐνες που υπερεκφράζονταν με συγκεκριμένο τρόπο. Η μελέτη της συγκεκριμένης απόκρισης ίσως μπορέσει τελικά να μας δώσει μία εξήγηση για την ποικιλία της κλινικής απόκρισης που παρατηρείται στους ασθενείς.

4.5.3 Massively Parallel Signature Sequencing MPSS

Η MPSS λειτουργεί παρόμοια με την μέθοδο **SAGE**, ωστόσο χάρη στην προηγμένη τεχνολογία που χρησιμοποιεί χαρακτηρίζεται από μεγαλύτερη ταχύτητα και ακρίβεια. Ήδη γίνονται προσπάθειες βελτίωσης της μεθόδου και μείωσης του κόστους της.

Η MPSS χρησιμοποιήθηκε στην μελέτη του καρκίνου του προστάτη. Με τη βοήθεια της μεθόδου παρατηρήθηκαν και ταυτοποιήθηκαν παραλλαγές που προκύπτουν κατά τη μεταγραφή του γονιδίου CT45 στον καρκίνο του προστάτη [19]. Ανακαλύφθηκαν συγκεκριμένες παραλλαγές που υπερ εκφράζονταν στα καρκινικά κύτταρα των όγκων συγκριτικά με τα φυσιολογικά κύτταρα του ιστού του προστάτη.

4.5.4 Single Nucleotide Polymorphisms (SNPs)⁹

Ως πολυμορφισμό ενός μόνο νουκλεοτιδίου - SNP ονομάζουμε μία διαφοροποίηση που παρατηρείται σε ένα μόνο νουκλεοτίδιο μίας ακολουθίας DNA. Όταν π.χ. στην σύγκριση 2 ανθρώπινων γονιδιωμάτων παρατηρείται να διαφέρουν σε ένα μόνο νουκλεοτίδιο δηλ εκεί που η μία ακολουθία έχει το ζεύγος θυμίνη- αδερίνη η άλλη να έχει γουανίνη- κυτοσύνη ή ακόμα απλούστερα οι δύο παρακάτω ακολουθίες διαφέρουν κατά ένα μόνο νουκλεοτίδιο: GATCC GTTCC.

Έχουν ήδη δημιουργηθεί βάσεις δεδομένων SNPs (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) οι οποίες είναι ελεύθερα προσβάσιμες και συνεισφέρουν πάρα πολύ στην σύγχρονη έρευνα.

⁹ Πολυμορφισμός μονού νουκλεοτιδίου

Οι πληροφορίες που μας διατίθενται από τις βάσεις δεδομένων SNPs μπορούν να χρησιμοποιηθούν σε ένα πλήθος μεθόδων που προάγουν την γνώση μας για τον καρκίνο.

Μεμονωμένα SNPs μπορούν να χρησιμοποιηθούν ως ενδείξεις στο γονιδίωμα για την αντιστοίχιση συγκεκριμένων γονότυπων με ασθένειες ή με την εμφάνιση συγκεκριμένων μορφών καρκίνου. Επίσης τα δεδομένα που μας προσφέρονται από τις βάσεις δεδομένων SNPs μπορούν να συμβάλουν στον εντοπισμό συγκεκριμένων μεταλλάξεων και ιδιαίτερας απαλείψεων και εισδοχών που συμβαίνουν σε ορισμένα είδη καρκίνων [20]. Διαφοροποιήσεις που ανακαλύφθηκαν στις κωδικές περιοχές κάποιων γονιδίων μπορεί να επηρεάζουν την δραστηριότητα των πρωτεϊνών που κωδικοποιούν και πολυμορφισμοί που ανακαλύφθηκαν στις regulatory περιοχές μπορεί να επηρεάζουν το επίπεδο της γονιδιακής έκφρασης [21]. Ήδη έχουν παρατηρηθεί περιπτώσεις όπου το πλήθος των πρωτεϊνών που παράγονται ή ο συγκεκριμένος πολυμορφισμός που εκφράζεται σχετίζεται με την απόκριση που θα έχει ο ασθενής στην χημειοθεραπεία [22] Ωστόσο, αν και οι ακολουθιακές μέθοδοι χρησιμοποιούνται για τον εντοπισμό και ταυτοποίηση μεμονωμένων σημειακών μεταλλάξεων και SNPs, αν όμως συγκεκριμένες διαφοροποιήσεις γνωρίζουμε ότι υπάρχουν μπορούν να εντοπιστούν με την χρήση της τεχνολογίας των Μικροδιατάξεων (microarray technology) που θα αναπτύξουμε στο επόμενο κεφάλαιο.

Βιβλιογραφία

- [1] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool, *J Mol Biol.*, 215(3), 403-10
- [2] Altschul, S. F. and Gish, W. (1996) Local alignment statistics, *Methods in Enzymology*, 266, 460-480
- [3] Altschul, S. F., Madden, T. L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25, 3389-3402
- [4] Arratia, R. and Waterman, M. S. (1994) A phase transition for the score in matching random sequences allowing deletions *Ann. Appl. Probab.* 4: 200-225
- [5] Miller, W., Zhang, Z., Schwartz, S., Wagner, L. (2000) A greedy algorithm for aligning DNA sequences, *Comput Biol*, 7(1-2), 203-14
- [6] Pearson, W.R. (1990) Rapid and Sensitive Sequence Comparison with FASTP and FASTA, *Methods in Enzymology*, 183, 63- 98
- [7] Pearson, W.R. (1991) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms, *Genomics*, 11(3), 635-50
- [8] Pearson, W.R., Lipman, D.J. (1988) Improved tools for biological sequence comparison, *Proc Natl Acad Sci.*, 85(8), 2444-8

- [9] Pearson, W. R. and Wood, T.C. (2001) Statistical significance in biological sequence comparison. In Handbook of Statistical Genetics. Balding, D.J, Bishop M., Cannings C. (eds). John Wiley and Sons, Ltd. England: 39-65
- [10] Smith, T.F., and Waterman, M.S. (1981) Identification of common molecular subsequences, *J Mol Biol*, 147(1), 195-7
- [11] Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, 22, 4673-4680
- [12] Krizman D.B., Wagner L., Lash A., Strausberg R.L. and Emmert-Buck M.R. 1999. The Cancer Genome Anatomy Project: EST sequencing and the genetics of cancer progression. *Neoplasia* 1: 101–106
- [13] Reis E.M., Ojopi E.P., Alberto F.L., Rahal P., Tsukumo F., Mancini U.M., Guimaraes G.S., Thompson G.M., Camacho C., Miracca E., Carvalho A.L., Machado A.A., Paquola A.C., Cerutti J.M., da Silva A.M., Pereira G.G., Valentini S.R., Nagai M.A., Kowalski L.P., Verjovski-Almeida S., Tajara E.H., Dias-Neto E., Bengtson M.H., Canevari R.A., Carazzolle M.F., Colin C., Costa F.F., Costa M.C., Estecio M.R., Esteves L.I., Federico M.H., Guimaraes P.E., Hackel C., Kimura E.T., Leoni S.G., Maciel R.M., Maistro S., Mangone F.R., Massirer K.B., Matsuo S.E., Nobrega F.G., Nobrega M.P., Nunes D.N., Nunes F., Pandolfi J.R., Pardini M.I., Pasini F.S., Peres T., Rainho C.A., dos Reis P.P., Rodrigues-Lisoni F.C., Rogatto S.R., dos Santos A., dos Santos P.C., Sogayar M.C. and Zanelli C.F. 2005. Large-scale transcriptome analyses reveal new genetic marker candidates of head, neck, and thyroid cancer. *Cancer Res* 65: 1693–1699
- [14] Asmann Y.W., Kosari F., Wang K., Chevillie J.C. and Vasmataz G. 2002. Identification of differentially expressed genes in normal and malignant prostate by electronic profiling of expressed sequence tags. *Cancer Res* 62: 3308–3314
- [15] Velculescu V.E., Zhang L., Vogelstein B. and Kinzler K.W. 1995. Serial analysis of gene expression. *Science* 270: 484–487
- [16] Zucchi I, Mento E, Kuznetsov V.A., Scotti M., Valsecchi V., Simionati B, Vicinanza E, Valle G, Pilotti S, Reinbold R, Vezzoni P, Albertini A. and Dulbecco R. 2004. Gene expression profiles of epithelial cells microscopically isolated from a breast-invasive ductal carcinoma and a nodal metastasis. *Proc Natl Acad Sci USA* 101: 18147–18152
- [17] Sengoele G., Luo W., Fine D., Perschl A.M., Fierlbeck W., Haririan A., Sorensson J., Rehman T.U., Hauser P., Trevick J.S., Kulak S.C., Wegner B. and Ballermann B.J. 2005. A SAGE-based comparison between glomerular and aortic endothelial cells. *Am J Physiol Renal Physiol* 288: F1290–F1300
- [18] Matsuzaki Y., Hashimoto S., Fujita T., Suzuki T., Sakurai T., Matsushima K. and Kawakami Y. 2005. Systematic identification of human melanoma antigens using serial analysis of gene expression (SAGE). *J Immunother* 28: 10–19
- [19] Chen Y.T., Scanlan M.J., Venditti C.A., Chua R., Theiler G., Stevenson B.J., Iseli C., Gure A.O., Vasicek T., Strausberg R.L., Jongeneel C.V., Old L.J. and Simpson A.J. 2005b. Identification of cancer/testis antigen genes by massively parallel signature sequencing. *Proc Natl Acad Sci USA* 102: 7940–7945
- [20] Hoque M.O., Lee C.C., Cairns P., Schoenberg M. and Sidransky D. 2003. Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis. *Cancer Res* 63: 2216–2222
- [21] Marsh S. 2005. Thymidylate synthase pharmacogenetics. *Invest New Drugs* 23: 533–537

- [22] Landi S., Gemignani F., Gioia-Patricola L., Chabrier A. and Canzian F. 2003. Evaluation of a microarray for genotyping polymorphisms related to xenobiotic metabolism and DNA repair. *Biotechniques* 35: 816–820, 822, 824–817

5

Τεχνικές Μικροδιατάξεων

5.1 Εισαγωγή στην τεχνολογία των Μικροδιατάξεων

Η τεχνολογία των μικροδιατάξεων (microarrays) προβάλλει ως μια πλατφόρμα βιολογικών εξερευνήσεων της οποίας οι δυνατότητες δεν έχουν ακόμη προσδιοριστεί εξαιτίας κυρίως του μεγάλου εύρους τους. Οι μικροδιατάξεις αποτελούν μέρος μιας νέας κλάσης της βιοτεχνολογίας η οποία επιτρέπει την παρακολούθηση του επιπέδου έκφρασης χιλιάδων γονιδίων ταυτόχρονα. Η έκφραση αυτού του μεγάλου αριθμού των γονιδίων μπορεί να παρατηρηθεί παράλληλα τόσο στο χώρο όσο και στο χρόνο.

Οι μικροδιατάξεις υβριδισμού γενετικού υλικού αποτελούν ήδη αξιόλογα εργαλεία αναγνώρισης και ποσοτικοποίησης πολλών ακολουθιών DNA που βρίσκονται σε περίπλοκης δομής δείγματα νουκλεϊκού οξέος. Μέχρι στιγμής μάλιστα έχουν χρησιμοποιηθεί σε έρευνες γενετικής αποκωδικοποίησης, σε αναλύσεις μεταλλάξεων καθώς και σε παρακολούθηση της έκφρασης συγκεκριμένων γονιδίων του γονιδιώματος.

Οι μικροδιατάξεις είναι, επί της αρχής αλλά και κατά την πράξη, επεκτάσεις των μεθόδων που βασίζονται στον υβριδισμό και οι οποίες έχουν χρησιμοποιηθεί για δεκαετίες με σκοπό την αναγνώριση και την ποσοτικοποίηση νουκλεϊκών οξέων σε βιολογικά δείγματα (για παραδείγματα, βλ. [1], [2], [3]). Στα υπό εξέταση δείγματα δίδονται χημικές «ετικέτες» και έπειτα υφίστανται υβριδισμό. Μετά από αρκετό χρόνο από τον υβριδισμό και ένα σύνολο βημάτων ξεπλύματος, λαμβάνεται μια εικόνα.

Πολύ γενικά τα πειράματα που διεξάγονται μπορούν να κατηγοριοποιηθούν σε αυτά των μονόχρωμων ραδιενεργών καθετήρων (probes), δίχρωμων (Cy3/Cy5) διατάξεων και υψηλής πιστότητας Affymetrix διατάξεων.

Η σημασία των μικροδιατάξεων, των οποίων η ανακάλυψη για πολλούς σημαίνει ότι ακριβώς σήμαινε και αυτή του μικροσκοπίου, αποτυπώνεται και στον αριθμό των δημοσιεύσεων. Χαρακτηριστικό δείγμα είναι ότι ο αριθμός των εργασιών με λέξεις κλειδιά τις “sequence analysis” αυξήθηκε από 10 σε 10.000 μέσα στη δεκαπενταετία 1967 – 1982 ενώ οι εργασίες των οποίων ο τίτλος περιείχε τη λέξη «microarray» χρειάστηκαν μόνο έξι χρόνια (1996 – 2001) για να συμπληρώσουν το παραπάνω νούμερο.

Οι εφαρμογές τους εκτείνονται από την αναγνώριση και τη τιθάσευση σύνθετων γενετικών ασθενειών (πχ. μεσογειακή αναιμία ή σύνδρομο Down), την ανακάλυψη νέων φαρμάκων, τις τοξικολογικές έρευνες και την ανίχνευση μεταλλάξεων και πολυμορφισμών μέχρι την παρακολούθηση των γονιδίων στο χρόνο, των γονιδίων σε ένα συγκεκριμένο ιστό και τον καθορισμό του σταδίου μιας ασθένειας (κάτι που ως γνωστό σήμερα δεν βασίζεται σε ποσοτικά κριτήρια αλλά επαφίεται στην ικανότητα του παθολογοανατόμου).

Εκτός από τις εφαρμογές, αρκετή συζήτηση έχει ήδη γίνει σχετικά με τις δυνατές επιπτώσεις της τεχνολογίας αυτής. Όλο και πιο έντονα ακούγεται πλέον ο όρος της «προληπτικής ιατρικής» η οποία στο μέλλον δεν θα σχετίζεται μόνο με γενικές συμβουλές βασιζόμενη στις στατιστικές του πληθυσμού αλλά θα δίνει και εξατομικευμένες οδηγίες υγείας βασιζόμενη στην έκφραση των γονιδίων του κάθε ατόμου – εν δυνάμει ασθενή- και άρα στην προδιάθεσή του.

Ακόμη μια συνέπεια της ορθολογικής χρήσης της νέας αυτής τεχνολογίας είναι η παροχή της δυνατότητας υποκατηγοριοποίησης των ασθενειών και ο σχεδιασμός φαρμάκων καταλλήλων για κάθε μια κατηγορία. Το κλινικό ή παρακλινικό σύμπτωμα μπορεί να οφείλεται σε διάφορες παθολογικές αιτίες οι οποίες όμως συχνά δεν αντιμετωπίζονται από τη συμβατική ιατρική. Με τη νέα τεχνολογία όμως, δίνεται η δυνατότητα της θεραπείας των αιτιών και όχι της καταστολής των συμπτωμάτων.

Θίχτηκε σε προηγούμενη παράγραφο αλλά θα αναλυθεί σε αυτή με μεγαλύτερη έκταση η δυνατότητα εξατομικευμένης ιατρικής βοήθειας. Η ανάλυση των γονιδίων ενός ανθρώπου δείχνει τη γενετική του ταυτότητα και κάποτε μπορεί να βοηθήσει ακόμα και στη θεραπεία ασθενειών που δεν έχουν δώσει κλινικά συμπτώματα. Έτσι, ακόμα και σε περιπτώσεις ασθενειών για τις οποίες δεν έχει βρεθεί η θεραπεία όπως η Οξεία Ανεπάρκεια του Ανοσοποιητικού Συστήματος (AIDS) ο ιατρός έχει τη δυνατότητα να γνωρίζει καλύτερα την πραγματική κατάσταση του ασθενούς και να χορηγεί την αντίστοιχη φαρμακευτική αγωγή με λιγότερα ρίσκα και περισσότερες πιθανότητες επιτυχίας.

Ωστόσο, ζώντας μέσα στη γλυκιά ορμή του νέου τομέα, οι ερευνητές της τεχνολογίας των microarrays καλούνται να επιδείξουν και την ανάλογη υπευθυνότητα και ηθική δεοντολογία. Υπάρχουν χρήσεις των microarrays που ενδέχεται να προκαλέσουν είτε το κοινό αίσθημα ή ακόμα και να παραβιάσουν ανθρώπινες ελευθερίες. Τον Απρίλιο του 2003 ο James Watson, ο "πατέρας του DNA", υποστήριξε· "είμαστε προϊόντα των γονιδίων μας και κανένας εκτός από μας δεν έχει το δικαίωμα να μας "φροντίζει" ή να μας επιβάλλει κανόνες συμπεριφοράς... Το κράτος πρέπει να μείνει μακριά από τη Γενετική... δεν μπορεί η κοινωνία να υπαγορεύει στους ανθρώπους τον τρόπο με τον οποίο θα χρησιμοποιούν την τεχνολογία" (Την "απελευθέρωση" της γενετικής ζητεί νομπελίστας, "ΤΟ ΒΗΜΑ", 10.4.2003, σ. 52.). Ο δε συνεργάτης του Francis Crick, προ ολίγων ετών δήλωσε· "Κανένα νεογέννητο παιδί δεν θα έπρεπε να αναγνωρίζεται ως ανθρώπινο πριν περάσει από ορισμένα τεστ για τα γενετικά του χαρίσματα... Αν αποτύχει σε αυτά τα τεστ, χάνει το δικαίωμα στη ζωή".

5.2 Η Τεχνολογία των Μικροδιατάξεων

Γενικά η τεχνολογία των μικροδιατάξεων μπορεί να εφαρμοστεί

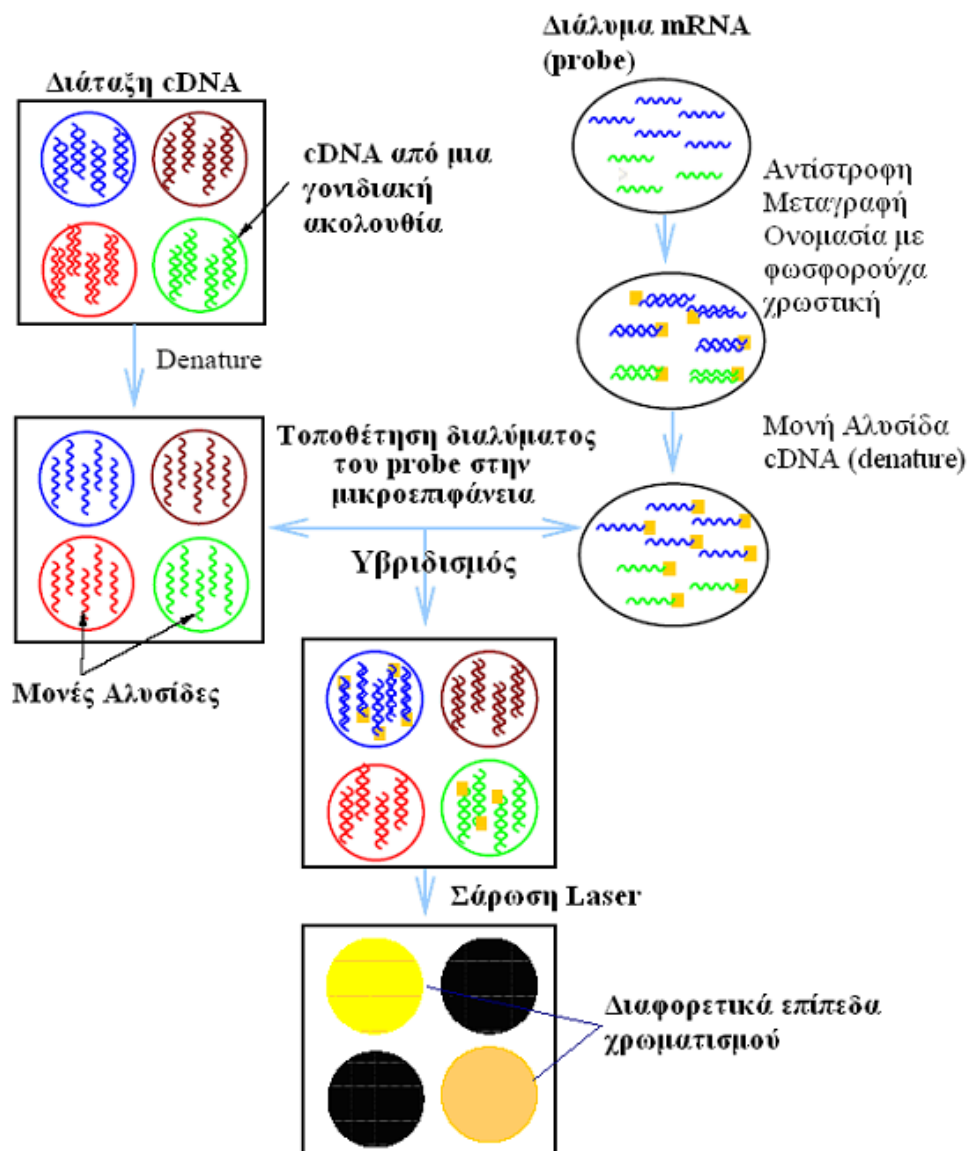
- α) για την κατανόηση των γενετικών ανωμαλιών γονιδιακής έκφρασης που οδηγούν ή συμβάλλουν στην διαδικασία της νεοπλασίας
 - β) στην ανακάλυψη νέων διαγνωστικών και προγνωστικών δεικτών καθώς και δεικτών ανταπόκρισης
 - γ) ταυτοποίηση της εγκυρότητας των μοριακών στόχων
 - δ) ανάπτυξη και παραγωγή νέων φαρμάκων
 - ε) ταυτοποίηση και τελειοποίηση της δομικής δραστηριότητας και αξιολόγηση πρωτεϊνών/παραγόντων
 - στ) πρόληψη παρενεργειών και τοξικότητας
 - ζ) επιβεβαίωση του τρόπου δράσης παραγόντων
 - η) αναγνώριση και ταυτοποίηση γονιδίων που εμπλέκονται στην φαρμακευτική αντίσταση και ευαισθησία
- και
- θ) στην αναγνώριση υποομάδων ασθενών που θα επωφεληθούν από συγκεκριμένες θεραπείες[6].

5.2.1 Μικροδιατάξεις cDNA

Όπως έχει αναλυθεί διεξοδικά στο δεύτερο κεφάλαιο, ένα κύτταρο περιέχει το σύνολο του γενετικού κώδικα του φορέα του το οποίο είναι αποθηκευμένο στο μόριο του DNA του. Ανάλογα με τη λειτουργία του, το κύτταρο χρησιμοποιεί

διαφορετικά γονίδια. Για παράδειγμα, σε ένα εγκεφαλικό κύτταρο εκφράζεται διαφορετικός αριθμός και είδος γονιδίων από ότι εκφράζεται σε ένα ηπατικό.

Έχει αναλυθεί η διαδικασία σύμφωνα με την οποία όταν ένα κύτταρο χρειάζεται να χρησιμοποιήσει ένα γονίδιο, ο κώδικας κατασκευής του συγκεκριμένου γονιδίου αντιγράφεται σε ένα μεταφορικό RNA με τη διαδικασία της μεταγραφής. Το mRNA στη συνέχεια μεταφράζεται σε πρωτεΐνη, που είναι το λειτουργικό προϊόν των περισσότερων γονιδίων. Η μετάφραση συμβαίνει σε κάθε χρονική στιγμή και για όλα τα γονίδια που χρησιμοποιούνται από το κύτταρο, σε διαφορετικά επίπεδα βέβαια για το κάθε γονίδιο.



Εικόνα 5.1 : Εύρεση τρόπου και ποσού γονιδιακής έκφρασης για δεδομένο τύπο κυττάρου

Τα πειράματα υβριδισμού γενετικού υλικού σε μικροδιατάξεις (εικόνα 5.1) μετρούν τη συγκέντρωση του mRNA που βρίσκεται ελεύθερο σε ένα σύνολο κυττάρων και μια υψηλή συγκέντρωση mRNA για ένα δεδομένο γονίδιο αντιστοιχεί σε υψηλό επίπεδο έκφρασης του γονιδίου. Στη πράξη, τα επίπεδα έκφρασης του γονιδίου μετρώνται σε περισσότερα του ενός κυτταρικά δείγματα και συγκρίνονται καθώς δεν καταμετρούνται με απόλυτες τιμές.

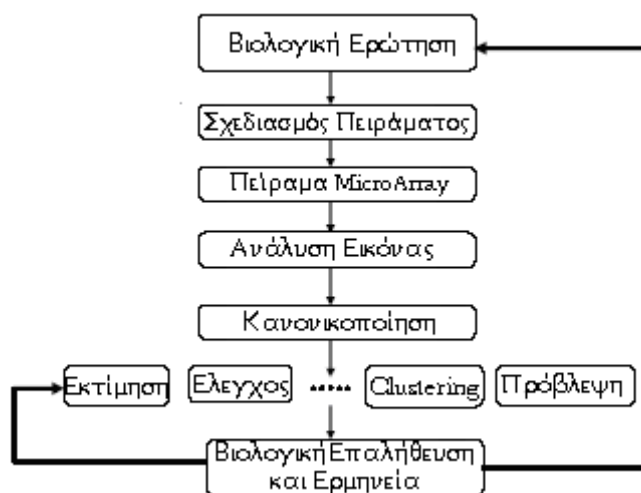
5.2.2 Τα Βήματα ενός Πειράματος Υβριδισμού Μικροδιατάξεων.

Αρχικά, το πείραμα ξεκινά με τη βιολογική ερώτηση. Στη συνέχεια, ο σχεδιασμός του πειράματος με βάση το πρώτο στάδιο οδηγεί, στην περίπτωση που όλα γίνουν καλά, στην εκτέλεση του ίδιου του πειράματος. Στο τέταρτο βήμα, η εικόνα που λαμβάνεται ως η έξοδος του πειράματος αναλύεται. Στο βήμα αυτό ανήκουν τα στάδια του addressing, της τμηματοποίησης (segmentation) και της απλοποίησης (reduction).

Στο στάδιο του addressing βρίσκονται οι περιοχές της εικόνας που ανήκουν (ή καλύτερα φιλοξενούνται) στην κηλίδα. Όλη η περιοχή της κηλίδας και του background της καλείται περιοχή – στόχος (target area). Στο στάδιο της τμηματοποίησης η περιοχή – στόχος διαχωρίζεται σε υποπεριοχή του foreground και σε υποπεριοχή του background. Με την απλοποίηση, τα δύο μονήρη μεγέθη R και G τα οποία αντιστοιχούν στα επίπεδα χρωματισμού των δύο καναλιών αντιστοιχίζονται σε ένα, το $\frac{R}{G}$.

Στο πέμπτο βήμα του πειράματος, τα δεδομένα που προκύπτουν κανονικοποιούνται ενώ στο έκτο τα κανονικοποιημένα επίπεδα χρωματισμού υφίστανται κάποιου είδους στατιστική ανάλυση. Στη στατιστική ανάλυση περιλαμβάνεται η ανάλυση προτύπων (cluster analysis) η οποία απαντά σε ερωτήματα του τύπου «ποια γονίδια ενεργούν μαζί;» ή «όμοια έκφραση γονιδίου σημαίνει αυτόματα και όμοια λειτουργία του κυττάρου του;». Στη στατιστική ανάλυση όμως ανήκει και η διακριτική ανάλυση (discriminant analysis) η οποία για παράδειγμα κατηγοριοποιεί τους τύπους των όγκων αλλά ανήκει και η επιλογή των μεταβλητών (για παράδειγμα ο έλεγχος για το ποια γονίδια είναι υπεύθυνα και επομένως μπορούν να χαρακτηριστούν ως διαφορετικές δύο η περισσότερες κλάσεις). Ακόμη, ανήκει και ο έλεγχος για differentially expressed γονίδια, δηλαδή για γονίδια τα οποία δεν εκφράζονται καθόλου στη φάση του ελέγχου αλλά εκφράζονται πλήρως στην πειραματική φάση και τέλος ανήκει και ο τομέας των γονιδιακών δικτύων.

Έβδομο και τελευταίο βήμα είναι η βιολογική επαλήθευση των αποτελεσμάτων καθώς και η ερμηνεία τους, ένα βήμα που ελπίζουμε ότι θα



Εικόνα 5.2 : Βήματα Εκτέλεσης Βιολογικού Πειράματος σε Microarrays

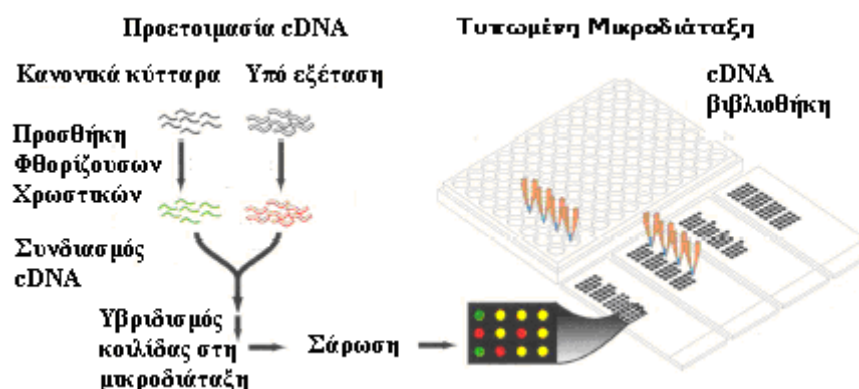
οδηγήσει την έρευνα στο αρχικό της στάδιο, όπως φαίνεται και στην εικόνα 5.2.

5.2.3 Κατασκευή μιας μικροδιάταξης

Μία ίνα (string) του mRNA είναι ένα αντίγραφο της μιας αλυσίδας της ακολουθίας DNA του γονιδίου. Πλέον, είναι εφικτό να κατασκευαστεί ένα συμπληρωματικό αντίγραφο του μορίου mRNA και να είναι δυνατή η περαιτέρω πειραματική εκμετάλλευση του προκύψαντος mRNA μέσω του αντιγράφου του cDNA.

Μια μικροδιάταξη είναι μια γυάλινη επιφάνεια επάνω στην οποία χιλιάδες κηλίδες από cDNA που αντιπροσωπεύουν διαφορετικά γονίδια (ή και τμήματα γονιδίων) τυπώνονται από έναν ρομποτικό βραχίονα. Ο βραχίονας ελέγχει ένα πλέγμα πολύ λεπτών μυτών έγχυσης γενετικού υλικού (print tips) τα οποία μπορούν να γεμίσουν με δείγματα από cDNA κλώνους (διαφορετικό δείγμα για το καθένα tip) και να τα τυπώσουν επάνω στο slide. Κάθε κηλίδα θα περιέχει χιλιάδες αντίγραφα του cDNA τμήματος από ένα γονίδιο. Κανονικά, όλες οι κηλίδες γεμίζουν με cDNA από διαφορετικά γονίδια (ή μέρη τους), ωστόσο αρκετά συχνά υπάρχουν κηλίδες – ρέπλικες για το ίδιο δείγμα επάνω στην ίδια μικροδιάταξη (συνήθως από 1-10 όπως αναφέρεται στο [4]).

Τα δύο κυτταρικά δείγματα που πρόκειται να συγκριθούν μέσω της γονιδιακής έκφρασης είναι συχνά α). κύτταρα που υπόκεινται σε κάποιου είδους θεραπεία και β). κανονικά (δηλαδή αυτά που δεν υπόκεινται σε θεραπεία) ή α). κύτταρα που έχουν δειγματοληφθεί από όγκους και β). από κανονικούς ιστούς ή απλά κύτταρα που προέρχονται από δύο διαφορετικά είδη ιστού. Για κάθε ένα από τα δύο δείγματα, απομονώνεται το mRNA και κάθε μια ακολουθία ονοματίζεται με τη βοήθεια μιας χημικής ετικέτας (χημική φωσφορούχας



Εικόνα 5.3 : Χρωματισμός κυττάρων ανάλογα με το αν είναι φυσιολογικά ή όχι

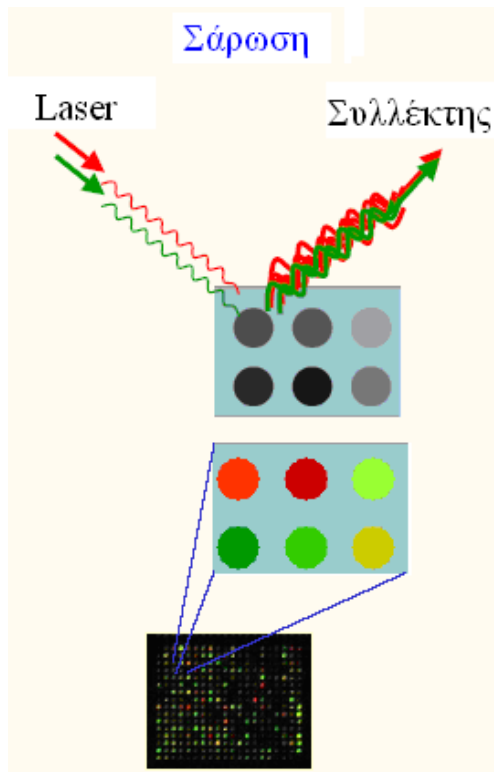
μπογιάς) κατά τη στιγμή της μετατροπής της σε cDNA. Συνήθως, τα κύτταρα που υπόκεινται στην εξέταση (για παράδειγμα καρκινικά) βάφονται κόκκινα και τα κανονικά πράσινα (εικόνα 5.3).

Με την προσθήκη ίσων ποσοτήτων των δύο ονοματισμένων δειγμάτων cDNA στη μικροδιάταξη, το δείγμα του cDNA θα υβριδιστεί στη cDNA κηλίδα του γυάλινου slide, δηλαδή, θα ζευγαρώσει με τα συμπληρωματικά ως προς αυτό κομμάτια του cDNA του slide. Εάν ένα γονίδιο έχει υψηλότερο επίπεδο έκφρασης στο υπό εξέταση δείγμα (για παράδειγμα στο καρκινικό) από ό,τι στο κανονικό δείγμα, θα υπάρχει περισσότερη κόκκινη από πράσινη μπογιά στη συγκεκριμένη κηλίδα.

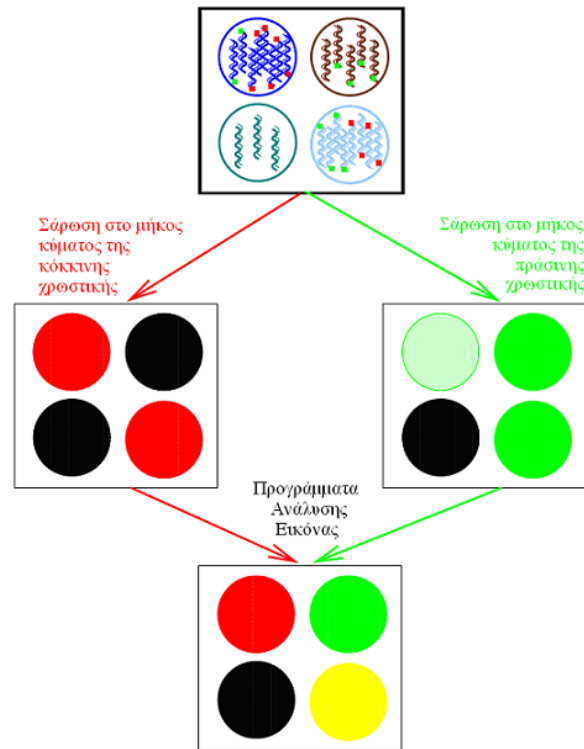
Τα επίπεδα χρωματισμού των κόκκινων και πράσινων μπογιών ανιχνεύονται με τη χρήση ενός σαρωτή laser (εικόνα 5.4 και εικόνα 5.5). Έτσι, τα κόκκινα και πράσινα επίπεδα χρωματισμού είναι οι μετρήσεις από τις οποίες θα ξεκινήσει η οποιαδήποτε στατιστική ανάλυση.

5.3 Η χρήση των Μικροδιατάξεων στη μελέτη του Καρκίνου

Η τεχνολογία των μικροδιατάξεων όπως θα δούμε στη συνέχεια, χρησιμοποιείται ευρέως στην έρευνα του καρκίνου, με μεγάλες προοπτικές ανάπτυξης. Ο συγκριτικός υβριδισμός που βρίσκεται στην βάση του όλου σκεπτικού των μικροδιατάξεων μπορεί να εξυπηρετήσει διαφορετικούς σκοπούς στην μελέτη του καρκίνου: α) μπορεί να υποδείξει τις διαφορές στην μεταγραφή που είναι υπεύθυνες για την αλλαγή ενός υγιούς κυττάρου σε καρκινικό, και β) να διακρίνει μορφές μη φυσιολογικής μετάφρασης σε ετερογενείς καρκίνους. Η κατανόηση της ποικίλης βάσης του καρκίνου είναι κρίσιμη για την εύρεση θεραπειών που στοχεύουν σε διαφορετικές μορφές της ασθένειας, έτσι ώστε κάθε ασθενής να δέχεται την πιο κατάλληλη και αποτελεσματική αγωγή.



Εικόνα 5.4 : Εκπομπή και λήψη φωτονίων κατά τη σάρωση



Εικόνα 5.5 : Σάρωση στο κατάλληλο μήκος κύματος

Οι διάφορες μορφές του καρκίνου είναι συνήθη παραδείγματα γενετικά ετερογενών ασθενειών, αλλά δεν είναι οι μοναδικές. Ο διαβήτης, καρδιακές και άλλες ασθένειες έχουν επίσης ετερογενή γενετικά αίτια[5].

5.3.1 Μικροδιατάξεις cDNA

Η τεχνολογία των μικροδιατάξεων και συγκεκριμένα οι μικροδιατάξεις cDNA χρησιμοποιήθηκαν αφειδώς για την μελέτη του καρκίνου ([7], [8]). Η ογκογένεση έχει ερευνηθεί από οποιαδήποτε πλευρά και πιθανώς για κάθε είδος όγκου ([9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]).

Με την βοήθεια των μικροδιατάξεων έχουν μελετηθεί κάποιες χρωμοσωμικές αναδιατάξεις που συμβαίνουν κατά την εμφάνιση της νόσου. ([22], [23], [24], [25], [26]). Κάποιες ερευνητικές ομάδες ασχολήθηκαν και με τους κληρονομικούς καρκίνους σε επίπεδο μεταγραφήματος [27]. Κάποιες άλλες αναζήτησαν δείκτες που χαρακτηρίζουν την εμφάνιση ή την εξέλιξη της νόσου ενώ άλλες εστίασαν την ερευνά τους στην χρησιμοποίηση του μεταγραφήματος για την κατάταξη σε υποκατηγορίες τόσο τελειώς διαφορετικών όγκων όσο και παρόμοιων ιστολογικά όγκων ([28], [29], [30], [31], [32], [33], [34], [35], [36], [37]). Αρκετοί ερευνητές προσπάθησαν με την βοήθεια των μικροδιατάξεων να κατανοήσουν ή να προβλέψουν την απόκριση συγκεκριμένων όγκων σε

χημειοθεραπευτικούς παράγοντες ή σε άλλες θεραπευτικές μεθόδους ([38], [39], [40], [41], [42], [36], [43], [44], [45], [46], [47]).

Στόχος των περισσότερων από αυτές τις προσπάθειες είναι η ανάπτυξη και βελτίωση της θεραπείας με την απόκτηση γνώσης και κατανόησης του καρκίνου σε μοριακό επίπεδο.

5.3.2 Μικροδιατάξεις Ολιγονουκλεοτιδίων

Ολιγονουκλεοτίδια ονομάζονται μικρές ακολουθίες DNA μήκους 20 έως 70 νουκλεοτίδια. Οι μικροδιατάξεις ολιγονουκλεοτιδίων ακολουθούν ακριβώς την ίδια λογική και έχουν παρόμοια χρήση με τις cDNA μικροδιατάξεις. Έτσι, όπως και στις cDNA μικροδιατάξεις οι ακολουθίες των ολιγονουκλεοτιδίων και οι θέσεις των κηλίδων πάνω στην διάταξη είναι γνωστές. Ωστόσο οι δύο τύποι μικροδιατάξεων παρουσιάζουν και σημαντικές διαφορές. Οι cDNA κλώνοι είναι αρκετά μακρύτεροι από αυτούς των ολιγονουκλεοτιδίων με αποτέλεσμα να δημιουργούν σταθερότερους δεσμούς και εντονότερο συγκριτικά σήμα [48] ωστόσο δεν μπορούν εύκολα να βοηθήσουν στην εξακρίβωση διαφοροποιήσεων του μεταγραφήματος ([49], [50]).

Διατάξεις ολιγονουκλεοτιδίων μπορούν να παρασκευαστούν εφόσον επαρκής πληροφορία για την αλληλουχία του γονιδιώματος ή του μεταγραφήματος ενός οργανισμού είναι γνωστή. Επομένως ο αριθμός των οργανισμών που είναι κατάλληλοι για αυτή την τεχνολογία είναι περιορισμένος. Από την άλλη μεριά, αποτελεί πλεονέκτημα το ότι τα ολιγονουκλεοτίδια μπορεί να αποτελούνται από οποιοδήποτε τμήμα γνωστού γονιδιώματος.

Οι μικροδιατάξεις ολιγονουκλεοτιδίων χρησιμοποιήθηκαν ευρέως στην έρευνα. Χαρακτηριστική είναι η χρήση τους στη μελέτη χρωμοσωμικών ανωμαλιών [51], της ποικιλίας στην συρραφή που παρουσιάζεται στο μεταγράφημα συγκεκριμένου είδους κυττάρων [52] και στην εύρεση γονιδιακών μεταλλάξεων εκ νέου ή γονιδιακών πολυμορφισμών ([53], [54]).

Στη μελέτη του καρκίνου συγκεκριμένα, οι πίνακες ολιγονουκλεοτιδίων χρησιμοποιήθηκαν περίπου κατά τον ίδιο τρόπο όπως και οι πίνακες cDNA. Σχεδόν κάθε όργανο στο οποίο μπορεί να αναπτυχθεί καρκίνος έχει μελετηθεί με αυτή την τεχνολογία, επίσης έχει μελετηθεί από οποιαδήποτε πλευρά η ανάπτυξη όγκου καθώς και η απόκριση του στην θεραπεία ([55], [56], [57], [58], [59], [60], [61], [62], [63]).

Σύντομα παρατηρήθηκε αλληλοεπικάλυψη ερευνητικών δραστηριοτήτων καθώς αρκετές εργασίες μελετούσαν το ίδιο αντικείμενο χρησιμοποιώντας

άλλες τις μικροδιατάξεις cDNA και άλλες ολιγονουκλεοτιδίων. Έτσι πολλά εργαστήρια χρησιμοποίησαν και τις δύο τεχνολογίες με την ελπίδα να παραγάγουν μία περισσότερο ολοκληρωμένη εικόνα της γονιδιακής έκφρασης για διάφορους τύπους καρκίνου ([64], [65], [66], [67]). Ωστόσο οι μικροδιατάξεις ολιγονουκλεοτιδίων θεωρούνται ιδανικές για ορισμένες εφαρμογές. Για παράδειγμα, με τη χρήση πινάκων ολιγονουκλεοτιδίων υψηλής πυκνότητας μελετήθηκαν το γονιδίωμα και το μεταγράφημα ίδιων όγκων και ανακαλύφθηκε πως υπάρχει μία συσχέτιση μεταξύ κάποιων διαφοροποιήσεων στην γονιδιακή έκφραση και στις γονιδιακές ανωμαλίες κυστικό αδenoκαρκίνωμα¹⁰ [68].

Υπάρχουν πίνακες οι οποίοι έχουν προτυποποιηθεί και εμπορευματοποιηθεί και παρουσιάζουν συνέπεια από γενιά σε γενιά. Το σημαντικό είναι ότι μπορούν να χρησιμοποιηθούν σε οποιοδήποτε εργαστήριο. Έτσι αρκετοί ερευνητές στράφηκαν προς τις τυποποιημένες μικροδιατάξεις προκειμένου να έχουν κοινή αναφορά και επομένως να μπορούν να συνδυάσουν ή και να συγκρίνουν απευθείας τα αποτελέσματά τους με τα αποτελέσματα άλλων ερευνητικών ομάδων ([69], [70], [71]). Μεγαλύτεροι τέτοιοι πίνακες χρησιμοποιούνται από ομάδες που στόχο έχουν την ταξινόμηση του είδους όγκου, της φάσης στην οποία βρίσκεται κάθε όγκος και των πιθανοτήτων επιβίωσης με βάση την γονιδιακή έκφραση ([58], [72], [60], [61], [62], [63]).

Τα δεδομένα που μας παρέχουν αφειδώς οι μικροδιατάξεις μπορούν να παρέχουν μία βαθύτερη γνώση του καρκίνου αλλά και της τεχνολογίας αυτής καθαυτής. Διάφορα σύνολα δεδομένων που αφορούν τον καρκίνο έχουν χρησιμοποιηθεί από την επιστημονική κοινότητα της Βιοπληροφορικής για την δοκιμή και σύγκριση αναλυτικών προσεγγίσεων ([56], [73]). Χάρη στη συμβατή φύση τους οι μικροδιατάξεις ολιγονουκλεοτιδίων είναι περισσότερο πιθανό να οδηγήσουν στην κλινική χρήση των μικροδιατάξεων για τη θεραπεία του καρκίνου.

5.3.3 Μικροδιατάξεις SNP

Ο πολυμορφισμός μονού νουκλεοτιδίου (SNP) είναι κάτι φυσικό και είναι αυτό που διαφοροποιεί τον έναν οργανισμό από τον άλλον. Ωστόσο, μπορεί να είναι και η αιτία ανάπτυξης ορισμένων όγκων ή ακόμα και ο λόγος για τον οποίο κάποιοι όγκοι δεν ανταποκρίνονται στην θεραπεία. Μεταλλάξεις που παρατηρούνται στα ογκογονίδια και στα ογκοκατασταλτικά είναι τέτοιου είδους διαφοροποιήσεις. Επίσης πολυμορφισμοί μπορεί να παρατηρηθούν και σε άλλα

¹⁰ adenoid cystic carcinomas

γονίδια και η επίδρασή τους να είναι πολύ σημαντική στην διαδικασία της ογκογένεσης [74] για παράδειγμα όπως αναφέρεται στο παράθεμα [75].

Για τη μελέτη πολυμορφισμών μονού νουκλεοτιδίου συνήθως χρησιμοποιούνται μέθοδοι σύγκρισης ακολουθιών ή τεχνικές που βασίζονται στην αλυσιδωτή αντίδραση πολυμεράσης (PCR¹¹) όταν πρόκειται για την μελέτη ενός συγκεκριμένου πολυμορφισμού τη φορά. Αλλά όταν έχουμε έναν πίνακα πιθανών πολυμορφισμών οι οποίοι έχουν αναγνωριστεί και ταξινομηθεί, η ανάλυση μπορεί να πραγματοποιηθεί με την χρήση των μικροδιατάξεων υβριδισμού για χιλιάδες πολυμορφισμούς (SNPs) ταυτόχρονα. Ένας τέτοιος πίνακας που έχει ήδη βγει στην παραγωγή εξακριβώνει πολυμορφισμούς στα γονίδια Cytochrome P450 που μπορεί να επηρεάζουν το πώς ο κάθε οργανισμός αντιδρά ή απορρίπτει χημειοθεραπευτικούς παράγοντες [76].

Οι SNPs μπορούν επίσης να παίζουν ρόλο δεικτών στο γονιδίωμα. Υψηλής πυκνότητας μικροδιατάξεις SNPs μπορούν να χρησιμοποιηθούν για την εύρεση μεγάλης έκτασης απαλοιφών και προσαυξήσεων στο γονιδίωμα ([77], [78]). Απώλεια της ετεροζυγωτίας συμβαίνει όταν μία περιοχή ενός χρωμοσώματος απαλείφεται. Αυτό είναι κάτι που συμβαίνει πολύ συχνά στην περίπτωση του καρκίνου και οι μικροδιατάξεις SNP μπορούν να χρησιμοποιηθούν ώστε πολύ γρήγορα να εξακριβωθεί πως κάτι τέτοιο συμβαίνει ([79], [80]). Για παράδειγμα έγινε έρευνα ανάμεσα σε 70 ασθενείς που έπασχαν από δύο διαφορετικούς τύπους καρκίνου των πνευμόνων (small cell και nonsmall cell lung cancer). Κατά τη διεξαγωγή αυτής της έρευνας δόθηκε η δυνατότητα να ταυτοποιηθούν ορισμένες περιπτώσεις απώλειας της ετεροζυγωτίας αλλά και παραδείγματα

¹¹ **Polymerase chain reaction (PCR):** μια τεχνική για να παράγονται πολλά αντίγραφα συγκεκριμένου DNA. Η αντίδραση ξεκινά με την χρήση μικρών ακολουθιών (εκκινητές, primer) που ταιριάζουν με τα άκρα των ακολουθιών που πρόκειται να αντιγραφούν. Οι εκκινητές μπορούν να υβριδίσουν με τα αντίγραφα όπως και με τις αρχικές ακολουθίες, ώστε ο συνολικός αριθμός των αντιγράφων αυξάνει εκθετικά με τον χρόνο.

Πολυμεράση: ένα ένζυμο το οποίο συνδέει κατάλοιπα νουκλεϊκών οξέων σε DNA ή RNA πολυμερή. Οι DNA πολυμεράσες αντιγράφουν το DNA σε DNA για να αντιγράψουν πιστά το γονιδίωμα, ενώ οι RNA πολυμεράσες αντιγράφουν το DNA σε RNA ως το πρώτο βήμα στην μεταγραφή των γονιδίων.

γονιδιακής επαύξεσης που πιθανώς εξηγούν την αυξημένη δυνατότητα επέκτασης των συγκεκριμένων όγκων[81].

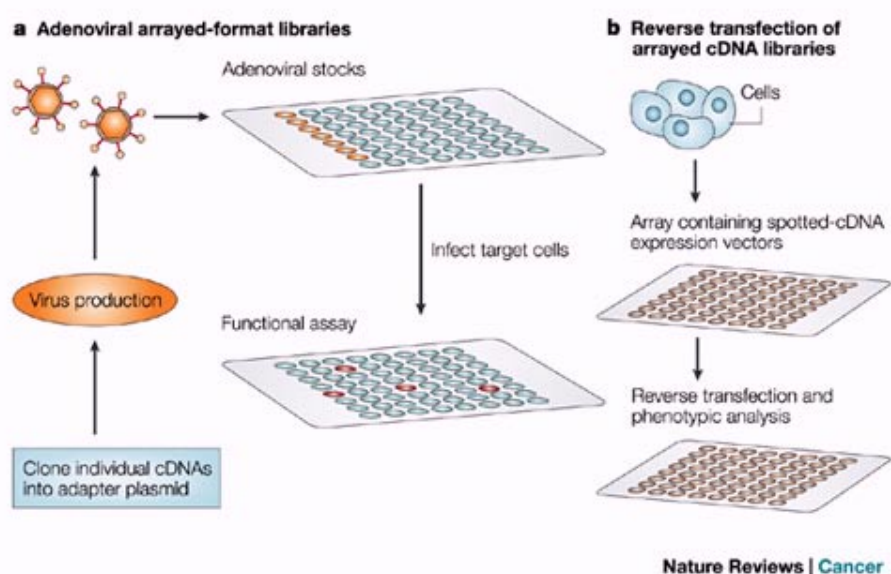
5.3.4 Παράδειγμα Χρήσης Μικροδιατάξεων στην Ανίχνευση του Καρκίνου του Μαστού.

Βιβλιοθήκες με τη μορφή δεδομένων μικροδιατάξεων περιέχουν δείκτες έκφρασης που τοποθετούνται σε προδιατεταγμένες θέσεις εντός του πίνακα. Με τη χρήση αυτής της τεχνικής, αναπτύσσεται μια βιβλιοθήκη έκφρασης συγκεκριμένων αδενοϊών η οποία περιέχει ιούς οι οποίοι μπορούν να προκαλέσουν την έκφραση συγκεκριμένων συμπληρωματικών αλυσίδων του DNA. Αυτά τα αποθέματα ιών μπορούν να χρησιμοποιηθούν για την προσβολή κυττάρων-στόχων. Αυτά τα κύτταρα μπορούν να χρησιμοποιηθούν για την φαινοτυπική ανάλυση σε υψηλούς ρυθμούς ροής αποτελεσμάτων.

Εναλλακτικά, μπορούμε να κάνουμε χρήση της τοποθέτησης νουκλεϊκού οξέος εντός του κυττάρου χωρίς τη χρήση ιών που είναι γνωστή ως διαμόλυνση (transfection). Συγκεκριμένα, μπορεί να γίνει συνδυασμός της αντίστροφης διαμόλυνσης με μικροδιατάξεις που κάνουν χρήση αυτών των ανιχνευτών ανίχνευσης της συμπληρωματικής αλυσίδας του DNA.

Τα κύτταρα μπορούν να διαμολυνθούν (με τη χρήση της αντίστροφης διαμόλυνσης) σε μικροδιατάξεις που κάνουν χρήση συγκεκριμένων συμπληρωματικών αλυσίδων DNA. Το αποτέλεσμα είναι κηλίδες έκφρασης από διαμολυσμένα κύτταρα με συγκεκριμένα cDNA. Τα κύτταρα αυτά μπορούν να χρησιμοποιηθούν για φαινοτυπική ανάλυση.

Η διαδικασία αυτή περιγράφεται στην εικόνα που ακολουθεί.



Εικόνα 5.6 : Παράδειγμα Χρήσης Μικροδιάταξης για την Ανίχνευση Καρκινικών κυττάρων

Βιβλιογραφία

- [1] E. M. Southern, *Journal of Molecular Biology*, 98, 503-517, 1975
- [2] J. C. Alwine, D. J. Kemp and G.R. Stark, *Proceedings of National Academic Science USA*, 74, 5350-4, 1977
- [3] F. C. Kafatos, C. W. Jones and A. Efstatidis, *Nucleic Acid Research*, 7, 1541-52, 1979
- [4] Ingrid Lonnstedt, Terry Speed, *Replicated Microarray data*, 2001
- [5] Αγγελική Κανατά- Συγκριτική Μελέτη Γονιδιακής Εκφρασης
- [6] Samuel Murray, Η τεχνολογία των μικροδιατάξεων (microarrays) και εφαρμογές στην Ογκολογία, Βήμα Κλινικής Ογκολογίας, τόμος 1, τεύχος 1, Ιανουάριος-Μάρτιος 2002
- [7] Bucca G., Carruba G., Saetta A., Muti P., Castagnetta L. and Smith C.P. 2004. Gene expression profiling of human cancers. *Ann NY Acad Sci* 1028: 28-37
- [8] Khan J., Saal L.H., Bittner M.L., Chen Y., Trent J.M. and Meltzer P.S. 1999. Expression profiling in cancer using cDNA microarrays. *Electrophoresis* 20: 223-229
- [9] Al Moustafa A.E., Alaoui-Jamali M.A., Batist G., Hernandez-Perez M., Serruya C., Alpert L., Black M.J., Sladek R. and Foulkes W.D. 2002. Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene* 21: 2634-2640
- [10] Alevizos I., Mahadevappa M., Zhang X., Ohyama H., Kohno Y., Posner M., Gallagher G.T., Varvares M., Cohen D., Kim D., Kent R., Donoff R.B., Todd R., Yung C.M., Warrington J.A. and Wong D.T. 2001. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* 20: 6196-6204
- [11] Baris O., Mirebeau-Prunier D., Savagner F., Rodien P., Ballester B., Loriod B., Granjeaud S., Guyetant S., Franc B., Houlgatte R., Reynier P. and Malthiery Y. 2005. Gene profiling reveals specific oncogenic mechanisms and signaling pathways in oncocytic and papillary thyroid carcinoma. *Oncogene* 24: 4155-4161
- [12] Durkin A.J., Bloomston M., Yeatman T.J., Gilbert-Barnes E., Cojita D., Rosemurgy A.S. and Zervos E.E. 2004. Differential expression of the Tie-2 receptor and its ligands in human pancreatic tumors. *J Am Coll Surg* 199: 724-731
- [13] Elek J., Park K.H. and Narayanan R. 2000. Microarray-based expression profiling in prostate tumors. *In Vivo* 14: 173-182
- [14] Larramendy M.L., Niini T., Elonen E., Nagy B., Ollila J., Vihinen M. and Knuutila S. 2002. Overexpression of translocation-associated fusion genes of FGFRI, MYC, NPM1, and DEK, but absence of the translocations in acute myeloid leukemia. A microarray analysis. *Haematologica* 87: 569-577
- [15] Lee Y.F., John M., Falconer A., Edwards S., Clark J., Flohr P., Roe T., Wang R., Shipley J., Grimer R.J., Mangham D.C., Thomas J.M., Fisher C., Judson I. and Cooper C.S. 2004. A gene expression signature associated with metastatic outcome in human leiomyosarcomas. *Cancer Res* 64: 7201-7204
- [16] Onda M., Emi M., Yoshida A., Miyamoto S., Akaishi J., Asaka S., Mizutani K., Shimizu K., Nagahama M., Ito K., Tanaka T. and Tsunoda T. 2004. Comprehensive gene expression profiling of anaplastic thyroid cancers with cDNA microarray of 25 344 genes. *Endocr Relat Cancer* 11: 843-854
- [17] Rihn B.H., Mohr S., McDowell S.A., Binet S., Loubinoux J., Galateau F., Keith G. and Leikauf G.D. 2000. Differential gene expression in mesothelioma. *FEBS Lett* 480: 95-100

- [18] Selaru F.M., Zou T., Xu Y., Shustova V., Yin J., Mori Y., Sato F., Wang S., Olaru A., Shibata D., Greenwald B.D., Krasna M.J., Abraham J.M. and Meltzer S.J. 2002. Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays. *Oncogene* 21: 475–478
- [19] Sorlie T., Perou C.M., Tibshirani R., Aas T., Geisler S., Johnsen H., Hastie T., Eisen M.B., van de Rijn M., Jeffrey S.S., Thorsen T., Quist H., Matese J.C., Brown P.O., Botstein D., Eystein Lonning P. and Borresen-Dale A.L. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98: 10869–10874
- [20] Wolf M., El-Rifai W., Tarkkanen M., Kononen J., Serra M., Eriksen E.F., Elomaa I., Kallioniemi A., Kallioniemi O.P. and Knuutila S. 2000. Novel findings in gene expression detected in human osteosarcoma by cDNA microarray. *Cancer Genet Cytogenet* 123: 128–132
- [21] Yang L.Y., Wang W., Peng J.X., Yang J.Q. and Huang G.W. 2004. Differentially expressed genes between solitary large hepatocellular carcinoma and nodular hepatocellular carcinoma. *World J Gastroenterol* 10: 3569–3573
- [22] Heiskanen M., Kononen J., Barlund M., Torhorst J., Sauter G., Kallioniemi A. and Kallioniemi O. 2001. CGH, cDNA and tissue microarray analyses implicate FGFR2 amplification in a small subset of breast tumors. *Anal Cell Pathol* 22: 229–234
- [23] Jiang F., Yin Z., Caraway N.P., Li R. and Katz R.L. 2004. Genomic profiles in stage I primary non small cell lung cancer using comparative genomic hybridization analysis of cDNA microarrays. *Neoplasia* 6: 623–635
- [24] Pollack J.R., Perou C.M., Alizadeh A.A., Eisen M.B., Pergamenschikov A., Williams C.F., Jeffrey S.S., Botstein D. and Brown P.O. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 23: 41–46
- [25] Wrobel G., Roerig P., Kokocinski F., Neben K., Hahn M., Reifenberger G. and Lichter P. 2005. Microarraybased gene expression profiling of benign, atypical and anaplastic meningiomas identifies novel genes associated with meningioma progression. *Int J Cancer* 114: 249–256
- [26] Yi Y., Mirosevich J., Shyr Y., Matusik R. and George A.L., Jr. 2005. Coupled analysis of gene expression and chromosomal location. *Genomics* 85: 401–412
- [27] Hedenfalk I., Duggan D., Chen Y., Radmacher M., Bittner M., Simon R., Meltzer P., Gusterson B., Esteller M., Kallioniemi O.P., Wilfond B., Borg A. and Trent J. 2001. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 344: 539–548 Hedenfalk I.A., Ringner M., Trent J.M. and Borg A. 2002. Gene expression in inherited breast cancer. *Adv Cancer Res* 84: 1–34
- [28] Bittner M., Meltzer P., Chen Y., Jiang Y., Seftor E., Hendrix M., Radmacher M., Simon R., Yakhini Z., Ben-Dor A., Sampas N., Dougherty E., Wang E., Marincola F., Gooden C., Lueders J., Glatfelter A., Pollock P., Carpten J., Gillanders E., Leja D., Dietrich K., Beaudry C., Berens M., Alberts D. and Sondak V. 2000. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406: 536–540
- [29] Bull J.H., Ellison G., Patel A., Muir G., Walker M., Underwood M., Khan F. and Paskins L. 2001. Identification of potential diagnostic markers of prostate cancer and prostatic intraepithelial neoplasia using cDNA microarray. *Br J Cancer* 84: 1512–1519
- [30] Cao Q.J., Belbin T., Socci N., Balan R., Prystowsky M.B., Childs G. and Jones J.G. 2004. Distinctive gene expression profiles by cDNA microarrays in endometrioid and serous carcinomas of the endometrium. *Int J Gynecol Pathol* 23: 321–329

- [31] Eschrich S., Yang I., Bloom G., Kwong K.Y., Boulware D., Cantor A., Coppola D., Kruhoffer M., Aaltonen L., Orntoft T.F., Quackenbush J. and Yeatman T.J. 2005. Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol* 23: 3526–3535
- [32] Halvorsen O.J., Oyan A.M., Bo T.H., Olsen S., Rostad K., Haukaas S.A., Bakke A.M., Marzolf B., Dimitrov K., Stordrange L., Lin B., Jonassen I., Hood L., Akslen L.A. and Kalland K.H. 2005. Gene expression profiles in prostate cancer: association with patient subgroups and tumour differentiation. *Int J Oncol* 26: 329–336
- [33] Hegde P., Qi R., Gaspard R., Abernathy K., Dharap S., Earle-Hughes J., Gay C., Nwokekeh N.U., Chen T., Saeed A.I., Sharov V., Lee N.H., Yeatman T.J. and Quackenbush J. 2001. Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res* 61: 7792–7797
- [34] Hu J., Bianchi F., Ferguson M., Cesario A., Margaritora S., Granone P., Goldstraw P., Tetlow M., Ratcliffe C., Nicholson A.G., Harris A., Gatter K. and Pezzella F. 2005. Gene expression signature for angiogenic and nonangiogenic non-small-cell lung cancer. *Oncogene* 24: 1212–1219
- [35] Khanna C., Khan J., Nguyen P., Prehn J., Caylor J., Yeung C., Trepel J., Meltzer P. and Helman L. 2001. Metastasis-associated differences in gene expression in a murine model of osteosarcoma. *Cancer Res* 61: 3750–3759
- [36] Mousses S., Wagner U., Chen Y., Kim J.W., Bubendorf L., Bittner M., Pretlow T., Elkahlon A.G., Trepel J.B. and Kallioniemi O.P. 2001. Failure of hormone therapy in prostate cancer involves systematic restoration of androgen responsive genes and activation of rapamycin sensitive signaling. *Oncogene* 20: 6718–6723
- [37] Ono K., Tanaka T., Tsunoda T., Kitahara O., Kihara C., Okamoto A., Ochiai K., Takagi T. and Nakamura Y. 2000. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Res* 60: 5007–5011
- [38] Akervall J., Guo X., Qian C.N., Schoumans J., Leeser B., Kort E., Cole A., Resau J., Bradford C., Carey T., Wennerberg J., Anderson H., Tennvall J. and Teh B.T. 2004. Genetic and expression profiles of squamous cell carcinoma of the head and neck correlate with cisplatin sensitivity and resistance in cell lines and patients. *Clin Cancer Res* 10: 8204–8213
- [39] Chang B.D., Swift M.E., Shen M., Fang J., Broude E.V. and Roninson I.B. 2002. Molecular determinants of terminal growth arrest induced in tumor cells by a chemotherapeutic agent. *Proc Natl Acad Sci USA* 99: 389–394
- [40] Kihara C., Tsunoda T., Tanaka T., Yamana H., Furukawa Y., Ono K., Kitahara O., Zembutsu H., Yanagawa R., Hirata K., Takagi T. and Nakamura Y. 2001. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res* 61: 6474–6479
- [41] Kudoh K., Ramanna M., Ravatn R., Elkahlon A.G., Bittner M.L., Meltzer P.S., Trent J.M., Dalton W.S. and Chin K.V. 2000. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res* 60: 4161–4166
- [42] Matsuyama R., Togo S., Shimizu D., Momiyama N., Ishikawa T., Ichikawa Y., Endo I., Kunisaki C., Suzuki H., Hayasizaki Y. and Shimada H. 2006. Predicting 5-fluorouracil chemosensitivity of liver metastases from colorectal cancer using primary tumor specimens: three-gene expression model predicts clinical response. *Int J Cancer* 119: 406–413
- [43] Nakeff A., Sahay N., Pisano M. and Subramanian B. 2002. Painting with a molecular brush: genomic/proteomic interfacing to define the drug action profile of novel solid-tumor selective anticancer agents. *Cytometry* 47: 72–79

- [44] Peehl D.M., Shinghal R., Nonn L., Seto E., Krishnan A.V., Brooks J.D. and Feldman D. 2004. Molecular activity of 1,25-dihydroxyvitamin D3 in primary cultures of human prostatic epithelial cells revealed by cDNA microarray analysis. *J Steroid Biochem Mol Biol* 92: 131–141
- [45] Samimi G., Manorek G., Castel R., Breaux J.K., Cheng T.C., Berry C.C., Los G. and Howell S.B. 2005. cDNA microarray-based identification of genes and pathways associated with oxaliplatin resistance. *Cancer Chemother Pharmacol* 55: 1–11
- [46] Smith D.I. 2002. Transcriptional profiling develops molecular signatures for ovarian tumors. *Cytometry* 47: 60–62
- [47] Takata R., Katagiri T., Kanehira M., Tsunoda T., Shuin T., Miki T., Namiki M., Kohri K., Matsushita Y., Fujioka T. and Nakamura Y. 2005. Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling. *Clin Cancer Res* 11: 2625–2636
- [48] Stillman B.A. and Tonkinson J.L. 2001. Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal Biochem* 295:149–157
- [49] Castle J., Garrett-Engele P., Armour C.D., Duenwald S.J., Loerch P.M., Meyer M.R., Schadt E.E., Stoughton R., Parrish M.L., Shoemaker D.D. and Johnson J.M. 2003. Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol* 4: R66
- [50] Kampa D., Cheng J., Kapranov P., Yamanaka M., Brubaker S., Cawley S., Drenkow J., Piccolboni A., Bekiranov S., Helt G., Tammana H. and Gingeras T.R. 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* 14: 331–342
- [51] Roversi G., Pfundt R., Moroni R.F., Magnani I., van Reijmersdal S., Pollo B., Straatman H., Larizza L. and Schoenmakers E.F. 2005. Identification of novel genomic markers related to progression to glioblastoma through genomic profiling of 25 primary glioma cell lines. *Oncogene* 25: 1571–1583
- [52] Nagao K., Togawa N., Fujii K., Uchikawa H., Kohno Y., Yamada M. and Miyashita T. 2005. Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the PTCH gene with exon junction microarrays. *Hum Mol Genet* 14: 3379–3388
- [53] Huang H.Y., Illei P.B., Zhao Z., Mazumdar M., Huvos A.G., Healey J.H., Wexler L.H., Gorlick R., Meyers P. and Ladanyi M. 2005. Ewing sarcomas with p53 mutation or p16/p14ARF homozygous deletion: a highly lethal subset associated with poor chemoresponse. *J Clin Oncol* 23: 548–558
- [54] Wikman F.P., Lu M.L., Thykjaer T., Olesen S.H., Andersen L.D., Cordon-Cardo C. and Orntoft T.F. 2000. Evaluation of the performance of a p53 sequencing microarray chip using 140 previously sequenced bladder tumor samples. *Clin Chem* 46: 1555–1561
- [55] Agrawal D., Chen T., Irby R., Quackenbush J., Chambers A.F., Szabo M., Cantor A., Coppola D. and Yeatman T.J. 2003. Osteopontin identified as colon cancer tumor progression marker. *CR Biol* 326: 1041–1043
- [56] Bhattacharjee A., Richards W.G., Staunton J., Li C., Monti S., Vasa P., Ladd C., Beheshti J., Bueno R., Gillette M., Loda M., Weber G., Mark E.J., Lander E.S., Wong W., Johnson B.E., Golub T.R., Sugarbaker D.J. and Meyerson M. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98: 13790–13795
- [57] Bonome T., Lee J.Y., Park D.C., Radonovich M., Pise-Masison C., Brady J., Gardner G.J., Hao K., Wong W.H., Barrett J.C., Lu K.H., Sood A.K., Gershenson D.M., Mok S.C. and Birrer M.J. 2005.

- Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. *Cancer Res* 65: 10602–10612
- [58] Dressman H.K., Hans C., Bild A., Olson J.A., Rosen E., Marcom P.K., Liotcheva V.B., Jones E.L., Vujaskovic Z., Marks J., Dewhirst M.W., West M., Nevins J.R. and Blackwell K. 2006. Gene expression profiles of multiple breast cancer phenotypes and response to neoadjuvant chemotherapy. *Clin Cancer Res* 12: 819–826
- [59] Frederiksen C.M., Knudsen S., Laurberg S. and Orntoft T.F. 2003. Classification of Dukes's B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol* 129: 263–271
- [60] Freije W.A., Castro-Vargas F.E., Fang Z., Horvath S., Cloughesy T., Liau L.M., Mischel P.S. and Nelson S.F. 2004. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64: 6503–6510
- [61] Ginos M.A., Page G.P., Michalowicz B.S., Patel K.J., Volker S.E., Pambuccian S.E., Ondrey F.G., Adams G.L. and Gaffney P.M. 2004. Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res* 64: 55–63
- [62] Gyorffy B., Surowiak P., Kiesslich O., Denkert C., Schafer R., Dietel M. and Lage H. 2006. Gene expression profiling of 30 cancer cell lines predicts resistance towards 11 anticancer drugs at clinically achieved concentrations. *Int J Cancer* 118: 1699–1712
- [63] Huang C.C., Cutcliffe C., Coffin C., Sorensen P.H., Beckwith J.B. and Perlman E.J. 2006. Classification of malignant pediatric renal tumors by gene expression. *Pediatr Blood Cancer* 46: 728–738
- [64] Bloom G., Yang I.V., Boulware D., Kwong K.Y., Coppola D., Eschrich S., Quackenbush J. and Yeatman T.J. 2004. Multi-platform, multi-site, microarray-based human tumor classification. *Am J Pathol* 164: 9–16
- [65] Carter S.L., Eklund A.C., Meham B.H., Kohane I.S. and Szallasi Z. 2005. Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics* 6: 107
- [66] Kuo W.P., Jenssen T.K., Butte A.J., Ohno-Machado L. and Kohane I.S. 2002. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 18: 405–412
- [67] Warnat P., Eils R. and Brors B. 2005. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics* 6: 265
- [68] Kasamatsu A., Endo Y., Uzawa K., Nakashima D., Koike H., Hashitani S., Numata T., Urade M. and Tanzawa H. 2005. Identification of candidate genes associated with salivary adenoid cystic carcinomas using combined comparative genomic hybridization and oligonucleotide microarray analyses. *Int J Biochem Cell Biol* 37: 1869–1880
- [69] Bammler T., Beyer R.P., Bhattacharya S., Boorman G.A., Boyles A., Bradford B.U., Bumgarner R.E., Bushel P.R., Chaturvedi K., Choi D., Cunningham M.L., Deng S., Dressman H.K., Fannin R.D., Farin F.M., Freedman J.H., Fry R.C., Harper A., Humble M.C., Hurban P., Kavanagh T.J., Kaufmann W.K., Kerr K.F., Jing L., Lapidus J.A., Lasarev M.R., Li J., Li Y.J., Lobenhofer E.K., Lu X., Malek R.L., Milton S., Nagalla S.R., O'Malley J.P., Palmer V.S., Pattee P., Paules R.S., Perou C.M., Phillips K., Qin L.X., Qiu Y., Quigley S.D., Rodland M., Rusyn I., Samson L.D., Schwartz D.A., Shi Y., Shin J.L., Sieber S.O., Slifer S., Speer M.C., Spencer P.S., Sproles D.I., Swenberg J.A., Suk W.A., Sullivan R.C., Tian R., Tennant R.W., Todd S.A., Tucker C.J., Van Houten B., Weis B.K., Xuan S. and Zarbl H. 2005. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2: 351–356

- [70] Dobbin K.K., Beer D.G., Meyerson M., Yeatman T.J., Gerald W.L., Jacobson J.W., Conley B., Buetow K.H., Heiskanen M., Simon R.M., Minna J.D., Girard L., Misek D.E., Taylor J.M., Hanash S., Naoki K., Hayes D.N., Ladd-Acosta C., Enkemann S.A., Viale A. and Giordano T.J. 2005. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res* 11: 565–572
- [71] Irizarry R.A., Warren D., Spencer F., Kim I.F., Biswal S., Frank B.C., Gabrielson E., Garcia J.G., Geoghegan J., Germino G., Griffin C., Hilmer S.C., Hoffman E., Jedlicka A.E., Kawasaki E., Martinez-Murillo F., Morsberger L., Lee H., Petersen D., Quackenbush J., Scott A., Wilson M., Yang Y., Ye S.Q. and Yu W. 2005. Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2: 345–350
- [72] Frederiksen C.M., Knudsen S., Laurberg S. and Orntoft T.F. 2003. Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol* 129: 263–271
- [73] Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Coller H., Loh M.L., Downing J.R., Caligiuri M.A., Bloomfield C.D. and Lander E.S. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531–537
- [74] Turesky R.J. 2004. The role of genetic polymorphisms in metabolism of carcinogenic heterocyclic aromatic amines. *Curr Drug Metab* 5: 169–180
- [75] Hein D.W. 2002. Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutat Res* 506–507: 65–77
- [76] Jain K.K. 2005. Applications of AmpliChip CYP450. *Mol Diagn* 9: 119–127
- [77] Matsuzaki H., Dong S., Loi H., Di X., Liu G., Hubbell E., Law J., Berntsen T., Chadha M., Hui H., Yang G., Kennedy G.C., Webster T.A., Cawley S., Walsh P.S., Jones K.W., Fodor S.P. and Mei R. 2004a. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 1: 109–111
- [78] Matsuzaki H., Loi H., Dong S., Tsai Y.Y., Fang J., Law J., Di X., Liu W.M., Yang G., Liu G., Huang J., Kennedy G.C., Ryder T.B., Marcus G.A., Walsh P.S., Shriver M.D., Puck J.M., Jones K.W. and Mei R. 2004b. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res* 14: 414–425
- [79] Dumur C.I., Dechsukhum C., Ware J.L., Cofield S.S., Best A.M., Wilkinson D.S., Garrett C.T. and Ferreira-Gonzalez A. 2003. Genome-wide detection of LOH in prostate cancer using human SNP microarray technology. *Genomics* 81: 260–269
- [80] Huang J., Wei W., Zhang J., Liu G., Bignell G.R., Stratton M.R., Futreal P.A., Wooster R., Jones K.W. and Shaper M.H. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genomics* 1: 287–299
- [81] Zhao X., Weir B.A., LaFramboise T., Lin M., Beroukhi R., Garraway L., Beheshti J., Lee J.C., Naoki K., Richards W.G., Sugarbaker D., Chen F., Rubin M.A., Janne P.A., Girard L., Minna J., Christiani D., Li C., Sellers W.R. and Meyerson M. 2005. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* 65: 5561–5570

6

Μέθοδοι Πρωτεϊνικής Ανάλυσης

6.1 Εισαγωγή

Μια πλήρης κατανόηση της διαδικασίας που οδηγεί σε ασθένεια απαιτεί προσέγγιση από όλες τις δυνατές πλευρές. Ιδιαίτερα για τον καρκίνο που αποτελεί πολυπαραγοντική νόσο χρειάζεται, για παράδειγμα, μία λεπτομερής γνώση όχι μόνο για το DNA και το RNA, αλλά και για αρκετές πλευρές των παραγόμενων πρωτεϊνών, από την βιοσύνθεσή τους και τις μετα-μεταφραστικές τροποποιήσεις τους (post-translational modifications, PTM) έως τις λειτουργίες και τις αλληλεπιδράσεις τους. Έτσι η έρευνα του καρκίνου στράφηκε και προς τον τομέα της πρωτεωμικής.

6.2 Πρωτεωμική

Ο όρος πρωτέωμα αναφέρεται στο συνολικό πρωτεϊνικό συμπλήρωμα που κωδικοποιείται και εκφράζεται από ένα γονιδίωμα.

Ο όρος πρωτεωμική αναφέρεται στην συστηματική μελέτη του συνολικού πρωτεϊνικού συμπληρώματος που εκφράζεται από το γονιδίωμα ή από συγκεκριμένα κύτταρα ή ιστούς, υγείων ή ασθενών.

Οι στόχοι της πρωτεωμικής μπορούν να χωριστούν σε τρεις κατηγορίες:

- α) προσδιορισμός πρωτεϊνών σε μεγάλη κλίμακα και μετα-μεταφραστικές τροποποιήσεις
- β) συγκρίσεις διαφορικής έκφρασης συγκεκριμένων πρωτεϊνών σε υγιείς και παθολογικές καταστάσεις και
- γ) μελέτες αλληλεπιδράσεων πρωτεϊνών.

Άλλες πλευρές της πρωτεωμικής περιλαμβάνουν την συμμετοχή των πρωτεϊνών σε νέες προσεγγίσεις για την ανάπτυξη φαρμάκων.

6.3 Κλασικές Μέθοδοι Πρωτεϊνικής Ανάλυσης

Στην πρωτεωμική κλασικές μέθοδοι πρωτεϊνικής ανάλυσης συναντούν σύγχρονα εργαλεία βιοπληροφορικής, όπως οι πρωτεϊνικές βάσεις δεδομένων.

Μία από τις κλασικές μεθόδους πρωτεϊνικής ανάλυσης είναι η φασματομετρία μαζών.

Οι δύο συνεισφορές της φασματομετρίας μαζών, που την έχουν κάνει απαραίτητο εργαλείο στην πρωτεωμική, είναι η ικανότητα προσδιορισμού απολύτων μοριακών μαζών καθώς και η ικανότητα παροχής πληροφοριών για αλληλουχίες μερικώς είτε ακόμη και πλήρως. Αρκετές στρατηγικές και τεχνικές φασματομετρίας μαζών έχουν αναπτυχθεί για εφαρμογές που ποικίλουν από τον προσδιορισμό και το χαρακτηρισμό των πεπτιδίων και πρωτεϊνών μέχρι την εξερεύνηση των βιολογικών συνεπειών από τις αλλαγές στις ανώτερες δομές των πρωτεϊνών.

Βασικές όμως για την μελέτη του πρωτεώματος είναι οι τεχνικές διαχωρισμού πρωτεϊνών όπως είναι η υγρή χρωματογραφία και η ηλεκτροφόρηση.

Η ηλεκτροφόρηση είναι μια τεχνική διαχωρισμού φορτισμένων ουσιών και κυρίως πρωτεϊνών. Κατά την ηλεκτροφόρηση ηλεκτρικά φορτισμένα μακρομόρια μετακινούνται προς τον έναν ή τον άλλον πόλο ηλεκτρικού πεδίου. Η μέθοδος χρησιμοποιείται για τον διαχωρισμό, απομόνωση και ανάλυση των πρωτεϊνικών μιγμάτων. Είναι μία κλασική εργαστηριακή μέθοδος τόσο στην έρευνα όσο και στην καθημερινή κλινική πράξη.

Η μέθοδος της υγρής χρωματογραφίας έχει ως αρχή λειτουργίας της (όπως όλες οι μέθοδοι χρωματογραφίας) ότι τα συστατικά του μίγματος που επιθυμούμε να διαχωρίσουμε μετατοπίζονται με διαφορετική ταχύτητα εξαιτίας της διαφορετικής κατανομής τους. Τα σήματα των συστατικών που ανιχνεύονται απεικονίζονται στα χρωματογραφήματα με τη μορφή κορυφών (*peaks*) που μοιάζουν με καμπύλες Gauss.

Οι κλασικές αυτές τεχνικές πρωτεϊνικής ανάλυσης με την βοήθεια των βάσεων δεδομένων και της αναζήτησης μέσα σε αυτές, δίνουν την δυνατότητα ανάλυσης των πρωτεϊνών στα συστατικά τους, αναγνώρισης των πεπτιδίων από τα οποία αποτελούνται και ανοίγουν τον δρόμο για την περαιτέρω μελέτη τους.

Έτσι είναι πλέον δυνατή η δημιουργία ομολογιών πρωτεϊνών, η πρόβλεψη της τριτοταγούς δομής τους μέσω της πρωτοταγούς και ένα πλήθος δεδομένων που εξάγεται από την μελέτη των πρωτεϊνών διατίθεται στην έρευνα.

6.4 Η συνεισφορά της Πρωτεϊνικής Ανάλυσης στη μελέτη του Καρκίνου

Ένας ιδιαίτερα ενδιαφέρον για τον καρκίνο ερευνητικός τομέας πρωτεϊνικής ανάλυσης είναι η αναζήτηση κλινικών βιοδεικτών ([1], [2], [3], [4], [5], [6], [7]). Με την χρήση μεθόδων πρωτεϊνικής ανάλυσης υψηλής απόδοσης στο σέρουμ ασθενών έχει διαπιστωθεί πως είναι δυνατή η ανίχνευση καρκίνου του παγκρέατος[2]. Με την χρήση επίσης παρόμοιων μεθόδων έχει γίνει ταυτοποίηση των πρωτεϊνών που σχετίζονται με τον καρκίνο του παγκρέατος[3]. Σε συνεργασία μεταξύ πολλών ερευνητικών εργαστηρίων έχει πραγματοποιηθεί, επίσης με τη χρήση της πρωτεωμικής, μελέτη σε γυναίκες που εμφάνιζαν καρκίνο των ωοθηκών *benign pelvic masses*, ή ήταν υγιείς[7]. Μέσα από αυτή την μελέτη ανακαλύφθηκαν τρεις ξεχωριστοί πρωτεϊνικοί δείκτες αποκλειστικά για τον καρκίνο των ωοθηκών που θα μπορούσαν προφανώς να αποτελέσουν καρκινικούς δείκτες για την έγκαιρη εύρεση όγκων. Ο συνδυασμός αυτών των 3 πρωτεϊνών με την πρωτεΐνη- δείκτη CA-125 διαπιστώθηκε πως παρουσιάζει μεγαλύτερη ευαισθησία στην ανίχνευση καρκίνου των ωοθηκών απ' ότι η χρήση μόνο του καρκινικού δείκτη CA-125 .

Η πρωτεωμική χρησιμοποιήθηκε επίσης για την κατάταξη των όγκων σε υποκατηγορίες για την πραγματοποίηση ιστολογικής διάγνωσης είτε απευθείας από δείγματα όγκων ή, σε περίπτωση που δεν έχει ανιχνευθεί όγκος, για την έγκαιρη ανίχνευσή του ([8], [9], [10]). Η απευθείας ανάλυση ιστού με τη χρήση φασματομετρίας μαζών είναι εφικτή σε μονο-κυτταρικό επίπεδο (*single cell level*) [11] και έχει χρησιμοποιηθεί σε δείγματα ιστών με σκοπό την πληρέστερη ανάλυση του όγκου ([12], [13]). Η τεχνική αυτή έχει χρησιμοποιηθεί για την μελέτη πολλών περιοχών εγκεφάλου των μικροσκοπικών στρωμάτων των ιστών [14]. Αυτό δίνει την δυνατότητα διερεύνησης τόσο στον όγκο όσο και στον γύρω ιστό για τυχόν ύπαρξη αλλαγών που σχετίζονται με την ανάπτυξη όγκου ή για την αναζήτηση μικροσκοπικών όγκων σε ένα δείγμα ιστού που κατά τα άλλα φαίνεται φυσιολογικό. Τέτοιου είδους μελέτες πιθανότατα θα αναδείξουν αξιολογές πληροφορίες σχετικά με την εμφάνιση όγκων από προσβεβλημένες μικροσκοπικές θέσεις που δεν μπορούν να αποκτηθούν μέσω καμίας άλλης μεθόδου.

Βιβλιογραφία

- [1] Alaiya A., Al-Mohanna M. and Linder S. 2005. Clinical cancer proteomics: promises and pitfalls. *J Proteome Res* 4: 1213–1222
- [2] Bhattacharyya S., Siegel E.R., Petersen G.M., Chari S.T., Suva L.J. and Haun R.S. 2004. Diagnosis of pancreatic cancer using serum proteomic profiling. *Neoplasia* 6: 674–686
- [3] Chen R., Yi E.C., Donohoe S., Pan S., Eng J., Cooke K., Crispin D.A., Lane Z., Goodlett D.R., Bronner M.P., Aebersold R. and Brentnall T.A. 2005a. Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. *Gastroenterology* 129: 1187–1197
- [4] Hondermarck H., Vercoutter-Edouart A.S., Revillion F., Lemoine J., el-Yazidi-Belkoura I., Nurcombe V. and Peyrat J.P. 2001. Proteomics of breast cancer for marker discovery and signal pathway profiling. *Proteomics* 1: 1216–1232
- [5] Srinivas P.R., Kramer B.S. and Srivastava S. 2001. Trends in biomarker research for cancer detection. *Lancet Oncol* 2: 698–704
- [6] Steel L.F., Shumpert D., Trotter M., Seeholzer S.H., Evans A.A., London W.T., Dwek R. and Block T.M. 2003. A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics* 3: 601–609
- [7] Zhang Z., Bast R.C., Jr., Yu Y., Li J., Sokoll L.J., Rai A.J., Rosenzweig J.M., Cameron B., Wang Y.Y., Meng X.Y., Berchuck A., Van Haaften-Day C., Hacker N.F., de Bruijn H.W., van der Zee A.G., Jacobs I.J., Fung E.T. and Chan D.W. 2004. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* 64: 5882–5890
- [8] Borczuk A.C., Shah L., Pearson G.D., Walter K.L., Wang L., Austin J.H., Friedman R.A. and Powell C.A. 2004. Molecular signatures in biopsy specimens of lung cancer. *Am J Respir Crit Care Med* 170: 167–174
- [9] Seike M., Kondo T., Fujii K., Okano T., Yamada T., Matsuno Y., Gemma A., Kudoh S. and Hirohashi S. 2005. Proteomic signatures for histological types of lung cancer. *Proteomics* 5: 2939–2948
- [10] Steel L.F., Shumpert D., Trotter M., Seeholzer S.H., Evans A.A., London W.T., Dwek R. and Block T.M. 2003. A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics* 3: 601–609
- [11] Danna E.A. and Nolan G.P. 2006. Transcending the biomarker mindset: deciphering disease mechanisms at the single cell level. *Curr Opin Chem Biol* 10: 20–27
- [12] Greengauz-Roberts O., Stoppler H., Nomura S., Yamaguchi H., Goldenring J.R., Podolsky R.H., Lee J.R. and Dynan W.S. 2005. Saturation labeling with cysteine-reactive cyanine fluorescent dyes provides increased sensitivity for protein expression profiling of laser-microdissected clinical specimens. *Proteomics* 5: 1746–1757
- [13] Jain K.K. 2002. Recent advances in oncoproteomics. *Curr Opin Mol Ther* 4: 203–209
- [14] Chaurand P., Sanders M.E., Jensen R.A. and Caprioli R.M. 2004. Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections. *Am J Pathol* 165: 1057–1068

7

Λοιπές Μέθοδοι Μελέτης της Νόσου του Καρκίνου

7.1 Εισαγωγή

Η θέληση και η ανάγκη του ανθρώπου να γνωρίζει σε βάθος και από όλες τις πλευρές τις ασθένειες που ταλαιπωρούν την φθαρτή φύση του αλλά και να μαθαίνει τρόπους να τις αντιμετωπίζει και να τις θεραπεύει, τον οδηγούν σε συνεχώς νέες και διαφορετικές ερευνητικές αναζητήσεις. Ένας από τους πρώτους τρόπους που επινοήθηκαν για το σκοπό αυτό ήταν η διεξαγωγή πειραμάτων σε ζώα, κάτι που εδώ και δεκαετίες έχει προκαλέσει αντιδράσεις φιλοζωικών οργανώσεων και όχι μόνο. Παρ' όλ' αυτά, τα πειραματόζωα χρησιμοποιήθηκαν και για την μελέτη του καρκίνου. Η Βιοπληροφορική όπως θα δούμε, έρχεται να συμπληρώσει και κάποτε να αναπληρώσει αυτά τα αμφιλεγόμενα πειράματα.

7.2 Η συμβολή των πειραμάτων σε ζώα στη μελέτη του Καρκίνου.

Η εισαγωγή φυσιολογικών και ανατομικών μοντέλων ζώων στην ιατρική έρευνα βοήθησε στο να ξεπεραστούν δυσκολίες οι οποίες υπάρχουν εκ φύσεως σε μελέτες σχετικά με τον άνθρωπο. Η μεγάλη γενετική βιοποικιλότητα όπως εμφανίζεται στον ανθρώπινο πληθυσμό συχνά λειτουργεί ως περιοριστικός παράγοντας στην ερμηνεία εξειδικευμένων πειραματικών αποτελεσμάτων, κάτι που απαιτεί τις περισσότερες φορές τη δυσβάσταχτη οικονομικά και χρονικά αξιολόγηση μεγάλου αριθμού ανθρωπίνων δειγμάτων.

Από τη στιγμή που είναι δυνατόν να αναπαραχθούν πειραματόζωα τύπου GEM (Genetically-Engineered Mouse) με σαφώς ορισμένο γενετικό υπόβαθρο, οι διαφορές όσον αφορά τη γενετική ετερογένεια είναι δυνατόν να μειωθούν σημαντικά. Εναλλακτικά, η διασταύρωση ποντικών που φέρουν ένα

συγκεκριμένο φαινότυπο με είδη διαφορετικού υποβάθρου στα οποία ο φαινότυπος έχει τροποποιηθεί, αποτελεί μία ισχυρή γενετική προσέγγιση για την αναγνώριση γονιδίων τα οποία προσδιορίζουν τις βιολογικές συμπεριφορές ([1], [2], [3]). Επιπρόσθετα, πολλά γενετικά τροποποιημένα ποντίκια αναπτύσσουν όγκους μετά από κάποιο συγκεκριμένο χρονικό διάστημα. Επομένως, οι διαφοροποιήσεις ανά στάδιο της καρκινογένεσης, όπως επίσης και οι αποκρίσεις σε θεραπευτικούς παράγοντες, μπορούν να αποτιμηθούν και πιθανώς να αντιστοιχηθούν στα αντίστοιχα στάδια της εξέλιξης του καρκίνου στον άνθρωπο. Τέτοιου είδους μελέτες που αφορούν σε κάθε στάδιο του καρκίνου χωριστά είναι γενικά αδύνατον να εφαρμοστούν σε ανθρώπους, δεδομένου ότι η διάγνωση των όγκων γίνεται στα τελευταία στάδια και η πολλαπλή δειγματοληψία των καρκινικών ιστών δεν είναι κλινικά ενδεδειγμένη ([3], [4]). Με αυτό τον τρόπο, τα GEM πειραματόζωα έχουν συμβάλει σημαντικά στον καθορισμό μοριακών μονοπατιών σηματοδότησης τα οποία συνδέονται με την ανάπτυξη του καρκίνου. Έτσι, τα GEM πειραματόζωα έχουν καταστεί σημαντικά στην εκτίμηση των σταδίων σε προκλινικούς ελέγχους και σε μελέτες πρόληψης.

7.3 Εισαγωγή της Βιοπληροφορικής στα πειράματα με GEM πειραματόζωα.

Παρά το ότι η εφαρμογή μεθόδων γενετικά τροποποιημένων ποντικών είναι ένα ισχυρό εργαλείο στη μελέτη παρόμοιων καταστάσεων που εμφανίζονται στον ανθρώπινο καρκίνο, εξακολουθούν να συνοδεύονται από διάφορους περιορισμούς. Τα πειραματόζωα, προκειμένου να φέρουν τις αντιπροσωπευτικές ασθένειες θα έπρεπε να διαθέτουν τόσο τις γενετικές όσο και τις φαινοτυπικές αλλαγές που συνοδεύουν αυτές τις ασθένειες στον άνθρωπο. Δυστυχώς, συνήθως δε συμβαίνει έτσι. Επιπλέον, τα μοντέλα ζώων που φέρουν νεοπλασίες που συναντούνται στους ανθρώπινους οργανισμούς, πρέπει να συγκρίνονται με βάση την ομοιότητά τους με την ανθρώπινη ασθένεια. Επίσης, τα κριτήρια σύγκρισης αυτών των μοντέλων θα πρέπει να βασίζονται σε πλήθος συγκριτικών παραγόντων.

Η έλευση νέων γονιδιακών τεχνολογιών, όπως η απεικόνιση του συνόλου της γονιδιακής έκφρασης και ο συγκριτικός υβριδισμός του γονιδιώματος, παρέχουν ένα σημαντικό μέσο τόσο για τη σύγκριση των ομοιοτήτων και των διαφορών οι οποίες απαντώνται σε επίπεδο γονιδιακής έκφρασης μεταξύ πειραματόζωων (ποντικών) και ανθρώπων όσο και για τη σύγκριση της εξέλιξης του καρκίνου σε καθεμία από τις δύο περιπτώσεις. Ωστόσο, το κατά πόσο αυτές οι GEM μέθοδοι συνοψίζουν σε παγκόσμια κλίμακα τα μοριακά

μονοπάτια τα οποία αφορούν στον ανθρώπινο καρκίνο είναι κάτι που μόλις ξεκίνησε να αξιολογείται. ([5], [6], [7]) .

Έχει παραχθεί μια μεγάλη ποικιλία GEM πειραματόζων με σκοπό να φέρουν συγκεκριμένες μεταλλάξεις οι οποίες λαμβάνουν χώρα στα διάφορα στάδια του ανθρώπινου καρκίνου. Κυρίως, στα GEM πειραματόζωα γίνεται προσπάθεια ώστε να μιμούνται την ανθρώπινη ασθένεια όσον αφορά:

- (1) την υπερέκφραση ή την ενεργοποίηση γονιδίων τα οποία είναι σχετικά με την ανάπτυξη του ανθρώπινου καρκίνου (ογκογονίδια),
- (2) την εξάλειψη των γονιδίων – στόχων με μεθόδους καταστροφής (knock-out strategies),
- (3) την παραγωγή των επικρατέστερων αρνητικών πρωτεϊνών ώστε να διακόψουν τη λειτουργία των ασθενών γονιδίων ([8], [3]), και
- (4) τους συνδυασμούς των παραπάνω.

Η συγκριτική απεικόνιση της έκφρασης των καρκινικών γονιδίων θα διευρύνει την κατανόηση των ογκογενετικών μονοπατιών, μελέτη η οποία είναι δύσκολο να διεξαχθεί στις ανθρώπινες ασθένειες εξαιτίας της πιθανής πολλαπλότητας των γενετικών αλλαγών, οι οποίες σχετίζονται με τον ανθρώπινο καρκίνο [9]. Για τους ανθρώπους που νοσούν από καρκίνο, κάποιες μελέτες έχουν χρησιμοποιήσει παρόμοιες μεθόδους συγκρίνοντας κάθε φορά πολλαπλά πρότυπα γενετικών εκφράσεων και όχι ένα μόνο. Αυτό γίνεται προκειμένου να ενισχυθεί ο καθορισμός των χαρακτηριστικών μιας συγκεκριμένης ασθένειας και να προβλεφθούν κλινικά αποτελέσματα, με σκοπό να επιλυθεί η βιολογική ετερογένεια [10]. Συνδυάζοντας πολλαπλές μικροδιατάξεις γενετικών προτύπων μαζί με κλινικούς παράγοντες επικινδυνότητας, η μελέτη πάνω στα παραπάνω μπορεί να σκιαγραφήσει το προσωπικό ιατρικό προφίλ του ασθενή. Κάτι τέτοιο μάς οδηγεί στο αποκορύφωμα της εξατομικευμένης ιατρικής [10].

7.3.1 Μέθοδοι πρόβλεψης κλάσης

Μέθοδοι της Βιοπληροφορικής ήρθαν να δώσουν νέα ώθηση και στην μελέτη των πειραματικών δεδομένων, τέτοιες μέθοδοι είναι οι μέθοδοι πρόβλεψης κλάσης.

Οι μέθοδοι πρόβλεψης κλάσης χρησιμοποιούνται ευρέως σε βιοϊατρικές και κλινικές μελέτες και περιλαμβάνουν δεδομένα μικροδιατάξεων. Σε αυτές τις μεθόδους μεγάλη σημασία έχει η ανάπτυξη προγνωστικών και διαγνωστικών γενομικών ταξινομητών. Η εφαρμογή ταξινόμησης απεικονίσεων της γονιδιακής έκφρασης σε ποντίκια απαντάται λιγότερο συχνά. Ωστόσο έχει

χρησιμοποιηθεί σε συγκρίσεις διασταυρώσεων διάφορων ειδών ως εργαλείο επικύρωσης του καρκίνου στα πειραματόζωα. Επίσης χρησιμοποιείται για την εξακρίβωση γενετικών υπογραφών οι οποίες διατηρούνται στα είδη και εμφανίζονται επίσης στον καρκίνο ([11], [12]). Η χρήση των μεθόδων πρόβλεψης κλάσης προκειμένου να προβλεφθεί η απόκριση των ποντικών σε συγκεκριμένη φαρμακευτική αγωγή είναι πολύ πιθανόν να αποκτήσει αυξανόμενη σημασία στο μέλλον ([13], [14]).

7.3.2 Μελέτη των Μεταστάσεων

Το θέμα των μεταστάσεων είναι αρκετά περίπλοκο. Η διεξαγωγή πειραμάτων σε ζώα έχοντας ως εργαλείο τις μεθόδους της Βιοπληροφορικής έρχεται να φωτίσει αρκετά σημεία. Πολλά από τα ερευνητικά αποτελέσματα αφορούν κυρίως τον καρκίνο του μαστού ο οποίος κατεξοχήν μελετήθηκε μέσω ζωικών μοντέλων.

Σημαντικές πληροφορίες μπορούν να εξαχθούν από την απεικόνιση της έκφρασης γενετικών διατάξεων συγκρίνοντας μεταστατικούς ιστούς με τον βασικό καρκίνο του μαστού. Παρά το ότι δεν υπάρχουν γονίδια τα οποία να φαίνονται να έχουν κάποιο συγκεκριμένο μεταστατικό φαινότυπο [15], έχει φανεί από τη γονιδιακή απεικόνιση ότι τα πρώιμα καρκινώματα του μαστού μοιάζουν πολύ με τις μακρινές τους μεταστάσεις ([16], [17], [18], [19], [20]).

Μια κοινά αποδεκτή θεωρία πάνω στη μετάσταση υποστηρίζει ότι η μετάσταση αργεί να εμφανιστεί στην ανάπτυξη του καρκίνου και έρχεται ως αποτέλεσμα συσσωρευμένων μεταλλάξεων. Ωστόσο, η μελέτη του προφίλ της έκφρασης των μεταστάσεων δείχνει ότι έμφυτοι μηχανισμοί ίσως να είναι ήδη υπαρκτοί από την αρχή της ασθένειας και αυτοί είναι που προσδιορίζουν εάν ένας όγκος θα εμφανίσει μεταστατική συμπεριφορά ή όχι [19]. Έχει βρεθεί ένα σύνολο από 117 γονίδια, με το οποίο προβλέπεται η πιθανότητα μεταστάσεων σε καρκινώματα του μαστού, χρησιμοποιώντας μικροδιατάξεις έκφρασης γενετικών προτύπων ([21], [2]).

Πρόσφατα, ένα μικρό σύνολο 17 γονιδίων είχε παρατηρηθεί να προβλέπει την πιθανότητα μεταστάσεων σε ένα σύνολο όγκων ([2], [22], [23]). Επομένως, σε αντίθεση με το κλασικό μοντέλο μετάστασης στο οποίο μόνο ένας μικρός πληθυσμός των καρκινικών κυττάρων συσσωρεύει τον απαιτούμενο για την μετάσταση αριθμό μεταλλάξεων, αυτές οι παρατηρήσεις υποδεικνύουν ότι οι πρώιμοι όγκοι ίσως να φέρουν την γονιδιακή «υπογραφή» της μετάστασης ([2], [24]). Επίσης, υποδεικνύεται ότι οι πιθανότητες μεταστάσεων καθορίζονται περισσότερο από κάποιους συνδυασμούς των αρχικών καρκινικών

μεταλλάξεων, παρά από τις διαδικασίες που επιτελούνται κατά τα τελικά στάδια της μετάστασης [25].

Βιβλιογραφία

- [1] Balmain και Nagase, 1998 Balmain, A., and Nagase, H. 1998. Cancer resistance genes in mice: models for the study of tumour modifiers. *Trends Genet* 14 : 139–144
- [2] Hunter, 2003 Hunter, K., Welch, D. R., and Liu, E. T. 2003. Genetic background is an important determinant of metastatic potential. *Nat Genet* 34 : 23–24; author reply 25
- [3] Kavanaugh και Green, 2003 Kavanaugh, C., and Green, J. E. 2003. The use of genetically altered mice for breast cancer prevention studies. *J Nutr* 133 : 2404S–2409S
- [4] Barkan, 2004 Barkan, D., Montagna, C., Ried, T., Green, J. E. 2004. Mammary gland cancer. In: *Mouse Models of Human Cancer*, ed. E. C. Holland, pp. 103–131. Hoboken: Wiley
- [5] Ried et al., 1995 Ried, T., Just, K. E., Holtgreve-Grez, H., du Manoir, S., Speicher, M. R., Schrock, E., Latham, C., Blegen, H., Zetterberg, A., Cremer, T., et al. 1995. Comparative genomic hybridization of formalin-fixed, paraffin-embedded breast tumors reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas. *Cancer Res* 55 : 5415–5423
- [6] Lee et al., 2005 Lee, J. S., Grisham, J. W., and Thorgeirsson, S. S. 2005. Comparative functional genomics for identifying models of human cancer. *Carcinogenesis* 26 : 1013–1020
- [7] Sweet-Cordero et al., 2005 Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J. J., Ladd-Acosta, C., Mesirov, J., Golub, T. R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 37 : 48–55
- [8] Hutchinson and Muller, 2000 Hutchinson, J. N., and Muller, W. J. 2000. Transgenic mouse models of human breast cancer. *Oncogene* 19 : 6130–6137
- [9] Desai et al., 2002b Desai, K. V., Xiao, N., Wang, W., Gangi, L., Greene, J., Powell, J. I., Dickson, R., Furth, P., Hunter, K., Kucherlapati, R., Simon, R., Liu, E. T., and Green, J. E. 2002b. Initiating oncogenic event determines gene-expression patterns of human breast cancer models. *Proc Natl Acad Sci USA* 99 : 6967–6972
- [10] Nevins et al., 2003 Nevins, J. R., Huang, E. S., Dressman, H., Pittman, J., Huang, A. T., and West, M. 2003. Towards integrated clinico-genomic models for personalized medicine: combining gene expression signatures and clinical factors in breast cancer outcomes prediction. *Hum Mol Genet* 12(Spec No. 2) : R153–R157
- [11] Lee, J. S., Grisham, J. W., and Thorgeirsson, S. S. 2005. Comparative functional genomics for identifying models of human cancer. *Carcinogenesis* 26 : 1013–1020
- [12] Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J. J., Ladd-Acosta, C., Mesirov, J., Golub, T. R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 37 : 48–55
- [13] Huang, E., West, M., and Nevins, J. R. 2003b. Gene expression profiling for prediction of clinical characteristics of breast cancer. *Recent Prog Horm Res* 58 : 55–73
- [14] Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., Joshi, M. B., Harpole, D., Lancaster, J. M., Berchuck, A., Olson, J. A., Jr., Marks, J. R., Dressman, H. K., West, M., and Nevins, J. R. 2006. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439 : 353–357

- [15] Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaelin, C. M., Marks, J. R., Riggins, G., and Polyak, K. 2001. A SAGE (serial analysis of gene expression) view of breast tumor progression. *Cancer Res* 61 : 5697–5702
- [16] Weigelt, B., Glas, A. M., Wessels, L. F., Witteveen, A. T., Peterse, J. L., and van't Veer, L. J. 2003. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci USA* 100 : 15901–15905
- [17] Hao, X., Sun, B., Hu, L., Lahdesmaki, H., Dunmire, V., Feng, Y., Zhang, S. W., Wang, H., Wu, C., Wang, H., Fuller, G. N., Symmans, W. F., Shmulevich, I., and Zhang, W. 2004. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *Cancer* 100 : 1110–1122
- [18] Lahdesmaki, H., Hao, X., Sun, B., Hu, L., Yli-Harja, O., Shmulevich, I., and Zhang, W. 2004. Distinguishing key biological pathways between primary breast cancers and their lymph node metastases by gene function-based clustering analysis. *Int J Oncol* 24 : 1589–1596
- [19] Weigelt, B., and van't Veer, L. J. 2004. Hard-wired genotype in metastatic breast cancer. *Cell Cycle* 3:756–757
- [20] Huang, J., Li, X., Hilf, R., Bambara, R. A., and Muyan, M. 2005. Molecular basis of therapeutic strategies for breast cancer. *Curr Drug Targets Immune Endocr Metabol Disord* 5 : 379–396
- [21] van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347 : 1999–2009
- [22] Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. 2003. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33 : 49–54
- [23] Qiu, T. H., Chandramouli, G. V., Hunter, K. W., Alkharouf, N. W., Green, J. E., and Liu, E. T. 2004. Global expression profiling identifies signatures of tumor virulence in MMTV-PyMT-transgenic mice: correlation to human disease. *Cancer Res* 64 : 5973–5981
- [24] Weigelt, B., Glas, A. M., Wessels, L. F., Witteveen, A. T., Peterse, J. L., and van't Veer, L. J. 2003. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci USA* 100 : 15901–15905
- [25] Bernards, R., and Weinberg, R. A. 2002. A progression puzzle. *Nature* 418 : 823

8

Συμπεράσματα για το Μέλλον της Βιοπληροφορικής στην Αντιμετώπιση του Καρκίνου

Η επιστήμη της Βιοπληροφορικής θα μπορούσε να θεωρηθεί ότι συγκεντρώνει τις χρήσεις των μαθηματικών, στατιστικών και υπολογιστικών μεθόδων για την επεξεργασία και την ανάλυση βιολογικών δεδομένων. Η γονιδιακή επανάσταση που συνέβη πριν μια δεκαετία περίπου δεν θα μπορούσε να στεφθεί από επιτυχία (αυτή που την συνέδεσε με την πλήρη αποκωδικοποίηση του ανθρώπινου γονιδιώματος) χωρίς την υποβοήθηση που δέχτηκε ειδικά πριν το τέλος της ιδιαίτερα εξελιγμένων στατιστικών αλγορίθμων, μεταξύ των οποίων βρίσκονται η ανάλυση της ακολουθίας του DNA (DNA sequencing), η γονιδιακή έκφραση σε συστοιχίες μικροδιατάξεων (microarray expression profiling) καθώς και η ακολουθιακή ανάλυση του γονιδιώματος (genomic sequence analysis).

Παρότι η τέχνη της συλλογής των δεδομένων και η ένωσή τους προς τη λύση ενός και μόνο προβλήματος θεωρείται παραδοσιακά ως σημαντικό μέρος της επιστημονικής προσπάθειας, οι πρόσφατες εξελίξεις στο χώρο της Βιοπληροφορικής μας δείχνουν ότι η ανάλυση των δεδομένων και η ερμηνεία τους είναι βήματα που καθορίζουν το συνολικό ρυθμό απόκτησης της βιολογικής γνώσης και κατανόησης των θεραπευτικών διαδικασιών.

Πιστεύουμε ότι προκειμένου να ενωθεί η Βιοπληροφορική με τη Βιολογία (και μάλιστα με προσοδοφόρο και για τις δύο τρόποι) και να εκπαιδευτεί μια νέα γενιά Βιολόγων και επιστημόνων της Στατιστικής Βιοπληροφορικής θα πρέπει να υπάρξει το όραμα τόσο από τα Πανεπιστήμια όσο και από τους πολιτικούς και οικονομικούς θεσμούς κάθε χώρας της Βιολογίας ως μιας Επιστήμης της Πληροφορίας.

Τόσο η ανάπτυξη όσο και η χρήση της Βιοπληροφορικής είναι πλέον σαφές ότι σχετίζεται όχι μόνο με το μέλλον αλλά και με το παρόν της θεραπευτικής του Καρκίνου. Και αυτό γιατί οι περισσότερες θεραπείες κατά του Καρκίνου φαίνεται πως φέρουν αποτελέσματα σε ένα μικρό σύνολο ασθενών, κάτι που είναι πολύ λογικό να παραμείνει έτσι για πολλά από τα νέα αντικαρκινικά φάρμακα που στόχο έχουν το μοριακό επίπεδο του οργανισμού. Κάτι που ήδη αρχίζει και διαφαίνεται σε ένα μεγάλο αριθμό ασθενών οι οποίοι δέχονται θεραπευτικές αγωγές χωρίς αποτέλεσμα αυξάνοντας έτσι, εκτός από τον πόνο και την αγωνία των συγγενών και των ίδιων και το οικονομικό βάρος που θα πρέπει να καλύπτουν τα ταμεία υγείας, ειδικά στον αναπτυγμένο κόσμο. Είναι λοιπόν ιδιαίτερα κεντρικής σημασίας, για τον Άνθρωπο, να αναπτυχθούν εργαλεία ακρίβειας που θα μεταφέρουν την κατάλληλη θεραπεία στον σωστό άρρωστο και θα έχουν ως βάση ένα βιολογικό χαρακτηρισμό καθενός εκ των όγκων των ασθενών.

Η χρήση μικροσυστοιχιών DNA για τον έλεγχο της γονιδιακής έκφρασης των όγκων είναι ένα πολύ δυνατό εργαλείο για τη φαρμακογενομική (pharmacogenomic) στοχοποίηση των θεραπειών. Ένα καλό παράδειγμα είναι το σύστημα Oncotype DX™ (της Genomic Health) το οποίο αναπτύχθηκε πρόσφατα για την ταυτοποίηση των αρνητικών υποδοχέων οιστρογόνου (node-negative estrogen-receptor-positive) ασθενών από καρκίνους του στήθους οι οποίοι δεν χρήζονται χημειοθεραπείας. Αυτή η ανάπτυξη των γονιδιακών ελέγχων που στόχο έχουν την ευρεία κλινική εφαρμογή απαιτεί την επιμελημένη προσπάθεια μιας ομάδας από κλινικούς Ιατρούς, Βιολόγους και επιστήμονες της Βιοστατιστικής.

Στον πόλεμο κατά του Καρκίνου, αν επιθυμούμε τη χρήση ακριβών, επαναλήψιμων διαγνωστικών μεθόδων θα πρέπει να στραφούμε σε κάτι διαφορετικό από τις αδόμητες μελέτες ετερογενών ομάδων του πληθυσμού που συνήθως στοιβάζονται στη βιβλιογραφία και που σχεδόν ποτέ δεν επικυρώνονται από ανεξάρτητες πηγές ή εφαρμόζονται σε ευρεία βάση. Θα πρέπει να στρέψουμε το ενδιαφέρον και την επιστημονική στήριξη μας σε διαγνωστικούς ελέγχους που θα βασίζονται στην έκφραση των γονιδίων προκειμένου να βοηθήσουμε τους ασθενείς αλλά και τους ιατρούς τους στη λήψη ενός μεγάλου εύρους δύσκολων αποφάσεων που αφορούν στη λήψη ή μη των σύγχρονων θεραπευτικών σκευασμάτων ή μεθόδων. Και όχι μόνο αυτό.

Η Βιοπληροφορική είναι ιδιαίτερα χρήσιμη σήμερα στη βελτιστοποίηση της διαδικασίας εύρεσης νέων αντικαρκινικών φαρμάκων. Πολλοί όγκοι αποτελούνται από μια μίξη υποκλώνων και περιέχουν διαφορετικά σύνολα

μεταλλαγμένων, υπερ-εκφραζόμενων και ήσυχων (silenced) γονιδίων. Αυτή η ετερότητα δημιουργεί τις προϋποθέσεις ώστε η διαδικασία της ταυτοποίησης καλών μοριακών στόχων να θεωρείται δύσκολη. Τα περισσότερα υπερ-εκφραζόμενα γονίδια και οι μεταλλάξεις τους συνήθως δεν αποτελούν καλούς μοριακούς στόχους και αυτό λόγω των υπολειπόμενων υποκλώνων οι οποίοι είναι ανθεκτικοί στη συγκεκριμένη θεραπευτική διαδικασία. Ο καλύτερος στόχος είναι ένα γονίδιο «κόκκινης τελείας» (red dot) του οποίου η μετάλλαξη συμβαίνει στα αρχικά στάδια της ογκογένεσης και αποσυντονίζει τη διαδικασία που οδηγεί την αύξηση του όγκου σε όλους τους υποκλώνους του. Παραδείγματα αποτελούν οι μεταλλάξεις στα γονίδια ABL, HER-2, KIT, EGFR και πιθανώς στο BRAF που απαντώνται στη χρόνια μυελογενή λευχαιμία, στον καρκίνο του στήθους, σε όγκους του γαστροισοφαγικού σωλήνα, στον καρκίνο των πνευμόνων και στο μελάνωμα, αντίστοιχα.

Σε αυτές τις περιπτώσεις, η ορθή εφαρμογή της θεραπείας απαιτεί την εύρεση των γονιδίων «κόκκινης τελείας» καθώς και τη διαγνωστική κατηγοριοποίηση των μονοπατιών που οδηγούν τον όγκο κάθε ασθενή στην αύξησή του. Τόσο η ανάπτυξη όσο και η εφαρμογή τέτοιων μεθόδων βιοπληροφορικής από ομάδες διαφορετικών ειδικοτήτων οι οποίες επιτελούν προσανατολισμένες διεθνείς έρευνες είναι ιδιαίτερα σημαντικές για το σύνολο των βημάτων που μπορούμε να ακολουθήσουμε στην καταπολέμηση του Καρκίνου.

Ένα άλλο σημείο, αυτό της χρήσης της γονιδιακής τεχνολογίας για την ανάπτυξη εξατομικευμένων φαρμάκων εξαρτάται από την ορθή χρήση της βιοπληροφορικής, όπως αυτή αναλύθηκε στην παρούσα διπλωματική εργασία. Τα εργαλεία που επιτρέπουν ταχείς προόδους στον τομέα της θεραπευτικής του Καρκίνου είναι πλέον διαθέσιμα. Αυτό που απαιτείται από όλους εμάς είναι η σοφία της διαχείρισής τους, έτσι ώστε να προκύψουν τεχνολογίες και να οργανωθούν οι γνώσεις από διάφορα, κάποτε αλληλοεπικαλυπτόμενα επιστημονικά πεδία για την ανάπτυξη πραγματικά αποτελεσματικών θεραπειών κατά του Καρκίνου.

Μέχρι να βρεθεί όμως το απόλυτο φάρμακο κατά του Καρκίνου (αν ποτέ υπάρξει κάτι τέτοιο), αυτός θα παραμένει ως κύρια αιτία θανάτου στο σύγχρονο κόσμο. Ο καρκίνος είναι πρωταρχικά η ασθένεια του γονιδίου: Ο τρόπος με τον οποίο επιχειρεί την κυριαρχία του επάνω στη ζωή εξαρτάται σε μεγάλο βαθμό από τη μοριακή υπογραφή του καθενός από εμάς. Η Βιοπληροφορική φαίνεται πως θα καθοδηγήσει, θα συντονίσει και στο τέλος θα χαρακτηρίσει τις πολύ-επιστημονικές ερευνητικές προσπάθειες που σκοπό έχουν την πρόγνωση, τη διάγνωση και την εξατομικευμένη ιατρική θεραπεία του Καρκίνου. _

