



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάλυση γονιδιακής έκφρασης μικροσυστοιχιών σε κυτταρικές
σειρές λευχαιμίας**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΑΙΚΑΤΕΡΙΝΗΣ Ε. ΣΚΟΥΤΑ

Επιβλέπων : Κουτσούρης Δημήτριος
Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2009



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάλυση γωνιακής έκφρασης μικροσυστοιχιών σε κυτταρικές
σειρές λευχαιμίας**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΑΙΚΑΤΕΡΙΝΗΣ Ε. ΣΚΟΥΤΑ

Επιβλέπων : Κουτσούρης Δημήτριος
Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 8^η Ιουλίου 2009.

.....
Κουτσούρης Δημήτριος
Καθηγητής ΕΜΠ

.....
Νικήτα Κωνσταντίνα
Καθηγήτρια ΕΜΠ

.....
Τσανάκας Παναγιώτης
Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2009

.....
Αικατερίνη Ε. Σκούτα

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

Copyright © Αικατερίνη Σκούτα, Ιούλιος 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.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή, κ. Δημήτριο Κουτσούρη και την κ. Ανδριάννα Πρέντζα, Επίκουρη Καθηγήτρια στο Πανεπιστήμιο Πειραιώς, για την ευκαιρία που μου έδωσαν να εκπονήσω την παρούσα διπλωματική εργασία σε έναν ενδιαφέροντα τομέα της βιοϊατρικής, καθώς και για την πολύτιμη βοήθειά τους. Επίσης, θα ήθελα να ευχαριστήσω θερμά τον υποψήφιο διδάκτορα κ. Εμμανουήλ Σηφάκη για την καθοδήγηση, τις πολύτιμες συμβουλές και τον χρόνο που μου αφιέρωσε κατά τη διάρκεια της εκπόνησης και συγγραφής της εργασίας αυτή. Τέλος, θα ήθελα να ευχαριστήσω την καθηγήτρια Παιδιατρικής Αιματολογίας/Ογκολογίας του Πανεπιστημίου Αθηνών κ. Φωτεινή Τζωρτζάτου - Σταθοπούλου και τον υποψήφιο διδάκτορά της κ. Γιώργο Λάμπρου, Μονάδα Αιματολογίας/Ογκολογίας, Α' Παιδιατρική Κλινική Πανεπιστημίου Αθηνών, Νοσοκομείο Παίδων «Αγία Σοφία» για τη συνεργασία.

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας εργασίας είναι η ανάπτυξη ενός αλγορίθμου επεξεργασίας και ανάλυσης δεδομένων από τυπωμένες cDNA μικροσυστοιχίες. Η τεχνολογία μικροσυστοιχιών επιτρέπει τη μέτρηση των επιπέδων έκφρασης χιλιάδων γονιδίων μόνο σε ένα πείραμα, δημιουργώντας έτσι μια αφθονία δεδομένων. Για να είναι δυνατή η εξαγωγή πληροφοριών που είναι σημαντικές για το βιολογικό σύστημα που μελετάται, υπάρχουν διάφορες μέθοδοι προεπεξεργασίας και ανάλυσης των δεδομένων. Στο πλαίσιο της ανάπτυξης του λογισμικού αυτού, υλοποιήθηκαν σε περιβάλλον Matlab[®] μια σειρά απαραίτητων βημάτων προεπεξεργασίας, όπως η διόρθωση θορύβου υποβάθρου, το φιλτράρισμα και η κανονικοποίηση και δίνεται στο χρήστη η δυνατότητα επιλογής μεταξύ διαφόρων εναλλακτικών επιλογών ως προς τα επιμέρους βήματα. Ακολουθεί η υλοποίηση αλγορίθμων ομαδοποίησης και συγκεκριμένα του αλγορίθμου k-μέσων, καθώς και ανάλυση πρωτεωουσών συνιστωσών που παρέχει μια ευκολότερη οπτικοποίηση των δεδομένων. Παρουσιάζονται επίσης τα αποτελέσματα της εφαρμογής του αλγορίθμου που αναπτύχθηκε, χρησιμοποιώντας δεδομένα από cDNA μικροσυστοιχίες της λευχαιμικής σειράς CCRFM-CEM που υφίστανται αγωγή με διαφορετικές συγκεντρώσεις πρεδνιζολόνης. Επειδή οι διαφορετικοί αλγόριθμοι μπορεί να παράγουν διαφορετικά αποτελέσματα, για την εξαγωγή συμπερασμάτων, λαμβάνεται υπόψη και το βιολογικό υπόβαθρο κάθε έρευνας.

Λέξεις κλειδιά: γονιδιακή έκφραση, cDNA μικροσυστοιχίες, ανάλυση πρωτεωουσών συνιστωσών, ομαδοποίηση k-μέσων, οξεία λεμφοβλαστική λευχαιμία.

ABSTRACT

The objective of the present thesis is the development of an algorithm such that the data from spotted cDNA microarrays can be processed and analyzed. Microarray is a high-throughput technology allowing the simultaneous screening of the expression levels of thousands of genes in one experiment, creating a vast amount of data. There are several methods for data pre-processing and analysis, in order to extract meaningful information about the biological system studied. We implement different pre-processing methods, such as background correction, filtering and normalization, within Matlab[®] programming environment. The user has also the ability to choose between alternative choices, moreover implement principal component analysis, which reduces the dimensionality of the data and allows a better visualization, and k-means clustering for grouping genes with similar expression patterns. We then applied our algorithm to microarray data from a T-cell leukemia cell line (CCRFM-CEM) treated with different prednisolone concentrations.

Keywords: gene expression, cDNA microarrays, principal component analysis, k-means clustering, acute lymphoblastic leukemia.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ	5
ΠΕΡΙΛΗΨΗ	7
ABSTRACT	9
ΛΙΣΤΑ ΕΙΚΟΝΩΝ	14
ΛΙΣΤΑ ΠΙΝΑΚΩΝ	16
1. ΕΙΣΑΓΩΓΗ	17
2. ΛΕΥΧΑΙΜΙΑ	19
2.1. ΚΑΡΚΙΝΟΣ.....	19
2.2. ΤΟ ΑΙΜΑ ΚΑΙ ΤΟ ΣΥΣΤΗΜΑ ΤΗΣ ΛΕΜΦΟΥ.....	20
2.3. ΛΕΥΧΑΙΜΙΕΣ.....	23
2.3.1. Γενικά.....	23
2.3.2. Τύποι Λευχαιμίας.....	24
2.3.2.1. Οξεία λεμφοβλαστική λευχαιμία (Acute lymphoblastic leukaemia – ALL).....	25
2.3.2.2. Χρόνια λεμφοβλαστική λευχαιμία (Chronic lymphoblastic leukaemia – CLL).....	25
2.3.2.3. Οξεία μυελοειδής λευχαιμία (Acute myeloid leukemia –AML).....	26
2.3.2.4. Χρόνια μυελοειδής λευχαιμία (Chronic myeloid leukemia – CML).....	26
2.4. ΟΞΕΙΑ ΛΕΜΦΟΒΛΑΣΤΙΚΗ ΛΕΥΧΑΙΜΙΑ.....	26
2.4.1. Αιτίες.....	27
2.4.2. Διάγνωση– Πρόγνωση.....	29
2.4.3. Ταξινόμηση.....	29
2.4.3.1. Ταξινόμηση κατά FAB (French American British Cooperative Group).....	30
2.4.3.2. Ταξινόμηση βασισμένη σε ανοσοφαινοτυποποίηση.....	30
2.4.3.3. Ταξινόμηση κατά WHO.....	31
2.4.4. Θεραπευτικές αγωγές.....	33
2.4.4.1. Χημειοθεραπεία.....	33
2.4.4.2. Μεταμόσχευση αρχέγονων κυττάρων ή μυελού των οστών.....	34
3. DNA, RNA, ΠΡΩΤΕΪΝΕΣ ΚΑΙ ΓΟΝΙΔΙΑΚΗ ΕΚΦΡΑΣΗ	35
3.1. ΚΕΝΤΡΙΚΟ ΔΟΓΜΑ ΤΗΣ ΒΙΟΛΟΓΙΑΣ.....	35
3.2. ΓΟΝΙΔΙΑ.....	36
3.3. DNA.....	36
3.3.1. Γενικά.....	36
3.3.2. Το μοντέλο της διπλής έλικας.....	38
3.4. RNA.....	40
3.5. ΓΕΝΕΤΙΚΟΣ ΚΩΔΙΚΑΣ.....	41
3.6. ΠΡΩΤΕΪΝΕΣ.....	41
3.7. ΓΟΝΙΔΙΑΚΗ ΕΚΦΡΑΣΗ.....	43
4. ΜΙΚΡΟΣΥΣΤΟΙΧΙΕΣ	45

4.1.	ΕΙΣΑΓΩΓΗ	45
4.2.	ΕΠΙΛΟΓΗ ΚΑΙ ΣΥΝΘΕΣΗ ΑΚΟΛΟΥΘΙΑΣ ΑΝΙΧΝΕΥΣΗΣ	47
4.3.	ΑΠΟΜΟΝΩΣΗ mRNA ΚΑΙ ΑΝΤΙΣΤΡΟΦΗ ΜΕΤΑΓΡΑΦΗ	48
4.4.	ΦΘΟΡΙΖΟΥΣΑ ΣΗΜΑΝΣΗ ΤΩΝ cDNA	49
4.5.	ΚΑΤΑΣΚΕΥΗ ΤΗΣ cDNA ΜΙΚΡΟΣΥΣΤΟΙΧΙΑΣ	50
4.6.	ΥΒΡΙΔΙΣΜΟΣ.....	50
4.7.	ΣΑΡΩΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΕΙΚΟΝΑΣ.....	51
4.8.	ΕΠΑΛΗΘΕΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	53
5.	ΕΠΙΣΚΟΠΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ	54
5.1.	ΔΙΟΡΘΩΣΗ ΘΟΡΥΒΟΥ ΥΠΟΒΑΘΡΟΥ	54
5.1.1.	<i>Μέθοδοι διόρθωσης υποβάθρου.....</i>	<i>55</i>
5.2.	ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ	56
5.2.1.	<i>Λογαριθμικός μετασχηματισμός.....</i>	<i>56</i>
5.2.2.	<i>Μετασχηματισμός τετραγωνικής ρίζας.....</i>	<i>57</i>
5.2.3.	<i>Box-Cox μετασχηματισμός.....</i>	<i>57</i>
5.2.4.	<i>Μετασχηματισμός ομοιότητας (affine transformation).....</i>	<i>57</i>
5.3.	ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ	58
5.3.1.	<i>Γενικά</i>	<i>58</i>
5.3.2.	<i>Ολική κανονικοποίηση.....</i>	<i>61</i>
5.3.3.	<i>Κανονικοποίηση εξαρτώμενη από την ένταση.....</i>	<i>62</i>
5.3.4.	<i>Κανονικοποίηση σε κάθε υποπλέγμα (print-tip group).....</i>	<i>62</i>
5.3.5.	<i>Κανονικοποίηση σε κάθε υποπλέγμα με κλιμακοποίηση.....</i>	<i>63</i>
5.3.6.	<i>Κανονικοποίηση με εναλλαγή της χρωστικής (dye swap).....</i>	<i>64</i>
5.3.7.	<i>Κανονικοποίηση ανάμεσα στα πλακίδια.....</i>	<i>65</i>
5.4.	ΦΙΛΤΡΑΡΙΣΜΑ ΔΕΔΟΜΕΝΩΝ	65
5.4.1.	<i>Φιλτράρισμα δεδομένων χαμηλής έντασης.....</i>	<i>66</i>
5.4.2.	<i>Καθορισμός κατώτατου και ανώτατου ορίου.....</i>	<i>66</i>
5.4.3.	<i>Εκτύπωση πολλαπλών κουκίδων (multiple spotting).....</i>	<i>66</i>
5.5.	ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΗ ΜΑΘΗΣΗ	67
5.5.1.	<i>Μέτρηση της απόστασης (Αλγόριθμοι Ομαδοποίησης).....</i>	<i>67</i>
5.5.2.	<i>Ιεραρχική ομαδοποίηση (Hierarchical clustering).....</i>	<i>68</i>
5.5.2.1.	<i>Ομαδοποίηση απλής σύνδεσης (single linkage clustering).....</i>	<i>69</i>
5.5.2.2.	<i>Ομαδοποίηση πλήρους σύνδεσης (complete linkage clustering).....</i>	<i>70</i>
5.5.2.3.	<i>Ομαδοποίηση μέσου όρου σύνδεσης (average linkage clustering).....</i>	<i>70</i>
5.5.2.4.	<i>Μέθοδος Ward.....</i>	<i>70</i>
5.5.3.	<i>Ομαδοποίηση k-μέσων (k-means clustering).....</i>	<i>72</i>
5.5.4.	<i>Αυτο-οργανούμενοι χάρτες (Self-organizing maps- SOM).....</i>	<i>73</i>
5.5.5.	<i>Ανάλυση Πρωτευουσών Συνιστωσών (principal components analysis).....</i>	<i>74</i>

5.6.	ΜΕΘΟΔΟΙ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ.....	76
5.6.1.	Γραμμικοί διαχωριστές (<i>linear discriminants</i>).....	77
5.6.2.	Μηχανές υποστήριξης διανυσμάτων (<i>Support vector machines – SVM</i>).....	77
5.6.3.	Μέθοδος <i>k</i> -κοντινότερων γειτόνων (<i>k-nearest neighbors</i>).....	78
6.	ΜΕΘΟΔΟΛΟΓΙΑ.....	79
6.1.	ΕΙΣΑΓΩΓΗ.....	79
6.2.	ΥΛΙΚΟ ΚΑΙ ΜΕΘΟΔΟΙ.....	79
6.2.1.	Συλλογή δεδομένων.....	79
6.2.2.	Σάρωση μικροσυστοιχίας.....	81
6.2.3.	Ανάλυση εικόνας με χρήση του <i>ImaGene</i>	82
6.3.	ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ.....	84
6.3.1.	Διόρθωση υποβάθρου.....	85
6.3.2.	Φιλτράρισμα δεδομένων χαμηλής έντασης.....	85
6.3.3.	Κανονικοποίηση.....	86
6.4.	ΠΕΡΑΙΤΕΡΩ ΕΠΙΛΟΓΗ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ΤΗΣ ΜΙΚΡΟΣΥΣΤΟΙΧΙΑΣ.....	90
6.5.	ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ.....	90
6.5.1.	Ανάλυση πρωτεουσών συνιστωσών (<i>PCA</i>).....	91
6.5.2.	Ομαδοποίηση <i>k-means</i>	92
7.	ΑΠΟΤΕΛΕΣΜΑΤΑ.....	94
8.	ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΕΚΤΑΣΕΙΣ.....	97
9.	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	98

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

Εικόνα 2.1: Αίμα (Παγκύπριος Αντιλευχαιμικός Σύνδεσμος "ΖΩΗ" 2009).....	21
Εικόνα 2.2: Σχηματισμός των διαφοροποιημένων κυττάρων του αίματος από αρχέγονο αιμοποιητικό κύτταρο, στο μυελό των οστών (Medical Library 2008).....	22
Εικόνα 2.3: Κατά τη διάρκεια της ανάπτυξης, τα ευκαρυωτικά κύτταρα συνεχώς περνούν από τις τέσσερις φάσεις του κυτταρικού κύκλου, δημιουργώντας νέα θυγατρικά κύτταρα (Lodish, Darnell & etc. 2003).....	23
Εικόνα 2.4: Οξεία Λεμφοβλαστική Λευχαιμία. Μυελός των οστών. (IBMS - The Institute of Biomedical Science 2009).....	27
Εικόνα 2.5: Μετασχηματισμός των αιματοποιητικών κυττάρων στην παθογένεση της ALL. (Pui et al. 2004).....	28
Εικόνα 2.6: Εκτιμώμενη συχνότητα συγκεκριμένων γονοτύπων ALL σε παιδιά. (Pui, Relling & Downing 2004).....	33
Εικόνα 3.1: Κεντρικό δόγμα της βιολογίας (Κεντρικό δόγμα Μοριακής Βιολογίας - Βικιπαίδεια 2008).....	35
Εικόνα 3.2: Επισκόπηση των τεσσάρων βασικών μοριακών γενετικών διαδικασιών (Lodish, Darnell & etc. 2003).....	36
Εικόνα 3.3: Νουκλεοτίδιο του DNA (Life Sciences Cyberbridge 2007).....	37
Εικόνα 3.4: Το βέλος δείχνει τον φωσφοδιεστερικό σκελετό με προσανατολισμό 5'→3' (Μαρμάρα Β., Λαμπροπούλου- Μαρμάρα Μ. 2005).....	37
Εικόνα 3.5: Μοντέλο διπλής έλικας του DNA (Lodish, Darnell & etc. 2003).....	38
Εικόνα 3.6: Τα χρωματισμένα χρωμοσώματα είναι χρήσιμα στην αποκάλυψη χρωμοσωματικών ανωμαλιών και στη σύγκριση καρυοτύπων διαφορετικών ειδών (Lodish, Darnell & etc. 2003).....	39
Εικόνα 3.7: Μικροσκοπική εμφάνιση ενός μεταφασικού χρωμοσώματος (Lodish, Darnell & etc. 2003).....	39
Εικόνα 3.8: Οι κωδικοποιημένες πληροφορίες στο DNA μετατρέπονται σε μια ακολουθία αμινοξέων στις πρωτεΐνες με μια διαδικασία πολλαπλών βημάτων (Lodish, Darnell & etc. 2003).	40
Εικόνα 3.9: Δομή αμινοξέων (Lodish, Darnell & etc. 2003).....	41
Εικόνα 3.10: Οι τρεις ρόλοι του RNA στην πρωτεϊνοσύνθεση (Lodish, Darnell & etc. 2003).....	42
Εικόνα 3.11: Επτά τύποι πρωτεϊνών που συμμετέχουν στον έλεγχο της αύξησης και του πολλαπλασιασμού κυττάρων (Lodish, Darnell & etc. 2003).....	43
Εικόνα 4.1: Παράδειγμα μικροσυστοιχίας με περίπου 40.000 ακολουθίες ανίχνευσης, με μεγεθυμένο τμήμα της για να φανούν οι λεπτομέρειες (Prasad 2003).....	46
Εικόνα 4.2: Παράδειγμα εκτέλεσης ενός πειράματος μικροσυστοιχίας (Lodish, Darnell & etc. 2003).....	48
Εικόνα 5.1: Τα βήματα της ιεραρχικής ομαδοποίησης. (Causton, Quackenbush & Brazma 2003).	69

Εικόνα 5.2: Απλή σύνδεση, πλήρης σύνδεση και σύνδεση μέσου όρου (Causton, Quackenbush & Brazma 2003).	70
Εικόνα 5.3: Παράδειγμα εφαρμογής Ιεραρχικής ομαδοποίησης σε δεδομένα (Quackenbush 2001).	71
Εικόνα 5.4: Ομαδοποίηση k-μέσων (Causton, Quackenbush & Brazma 2003).	72
Εικόνα 5.5: Αυτο-οργανούμενοι χάρτες (Causton, Quackenbush & Brazma 2003).	74
Εικόνα 5.6: Απεικόνιση της PCA ανάλυσης (Kaviraki 2007).	76
Εικόνα 6.1: IntelliGene™ Human Cancer CHIP Version 4.0 (IntelliGene™ Human Cancer CHIP 2009).	80
Εικόνα 6.2: Packard Bioscience ScanArray 4000XL (GSI Lumonics ScanArray 4000 chip scanner 2009).	82
Εικόνα 6.3: Παράδειγμα Αρχείου εξόδου ImaGene v 6.1.0.	83
Εικόνα 6.4: Διάγραμμα διασποράς του καναλιού Cy3 ως προς το Cy5.	86
Εικόνα 6.5: Η συνάρτηση βάρους για ένα τελικό και ένα εσωτερικό σημείο του πεδίου τιμών (The MathWorks Incorporated 2009).	88
Εικόνα 6.6: Το γράφημα (α) δείχνει ότι η έκτροπη τιμή επηρεάζει τις ομαλοποιημένες τιμές αρκετών κοντινών γειτονικών σημείων. Το γράφημα (β) δείχνει ότι το υπόλοιπο της εκτροπής τιμής (outlier) είναι μεγαλύτερο από 6 MAD. Το γράφημα (γ) δείχνει ότι οι ομαλοποιημένες τιμές των γειτονικών της έκτροπής τιμής σημείων, απεικονίζουν το μεγαλύτερο μέρος των δεδομένων (The MathWorks Incorporated 2009).	89
Εικόνα 7.1: IR-διάγραμμα διασποράς πριν (α) και μετά (β) την loess κανονικοποίηση των δεδομένων του πειράματος «0vs1».	94
Εικόνα 7.2: (α) ομαδοποίηση k-μέσων των διορθωμένων για θόρυβο υποβάθρου και κανονικοποιημένων δεδομένων, (β) προφίλ των γονιδίων που ανήκουν στις τρεις ομάδες και (γ) προφίλ των κεντροειδών των ομάδων.	95
Εικόνα 7.3: (α) Απεικόνιση των δεδομένων μετά την ανάλυση πρωτεωουσών συνιστωσών, ως προς τις δύο κύριες συνιστώσες και με ομαδοποίηση k-μέσων, (β) προφίλ των γονιδίων που ανήκουν στις τρεις ομάδες και (γ) προφίλ των αντίστοιχων κεντροειδών των ομάδων.	96

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

Πίνακας 2.1: Κυτταρογενετικές μετατοπίσεις συνδέονται με συγκεκριμένες μοριακές γενετικές ανωμαλίες στην ALL (Stams et al. 2005, McWhirter et al. 1999, Den Boer et al., Martin-Subero et al. 2005, Caslini et al. 2004).....	31
Πίνακας 2.2: Σχέση της πρόγνωσης με την κυτταρογενετική αλλαγή που εμφανίζεται (Acute lymphoblastic leukemia - Wikipedia, the free encyclopedia 2009).....	32
Πίνακας 2.3: Συσχετισμός της πρόγνωσης με τα κυτταρογενετικά ευρήματα στο μυελών των οστών στην οξεία λεμφοβλαστική λευχαιμία (Den Boer et al.).	32

1. ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια η ανάπτυξη της τεχνολογίας των μικροσυστοιχιών επέτρεψε τον ταυτόχρονο υβριδισμό χιλιάδων γονιδίων σε ένα μόνο πείραμα. Η ανάλυση της γονιδιακής έκφρασης όλων αυτών των γονιδίων, παίζει σημαντικό ρόλο στον προσδιορισμό προτύπων γονιδιακής έκφρασης και κοινών μεταβολικών οδών, καθιστώντας δυνατή την ανάπτυξη διαγνωστικών δεικτών για πιο ακριβή διάγνωση και πρόγνωση ασθενειών, όπως η λευχαιμία, καθώς και την ανάπτυξη νέων θεραπευτικών στόχων, που θα επιτρέψουν την στοχευμένη και ειδική θεραπεία του ασθενούς.

Για τη δημιουργία των cDNA μικροσυστοιχιών, ανάλογα με το βιολογικό σύστημα που μελετάται, επιλέγονται κατάλληλα δείγματα RNA (δείγμα αναφοράς και προς εξέταση δείγμα) και μετατρέπονται σε cDNA με παράλληλη σήμανση τους με χρωστικές ουσίες που έχουν διαφορετικά μήκη κύματος διέγερσης και εκπομπής. Στη συνέχεια τα δείγματα αυτά τυπώνονται στις μικροσυστοιχίες, σε κουκίδες που καθεμία αντιστοιχεί συνήθως σε ένα γονίδιο. Ακολουθεί η σάρωση των μικροσυστοιχιών με συνεστιακό μικροσκόπιο και η εξαγωγή των επιπέδων έκφρασης κάθε γονιδίου.

Τα ακατέργαστα δεδομένα (raw data) που προκύπτουν από τις μικροσυστοιχίες είναι μεγάλα σε αριθμό και απαιτείται η ανάπτυξη μεθόδων επεξεργασίας και ανάλυσης, ώστε να είναι δυνατή η εξαγωγή σημαντικής πληροφορίας σχετικά με το βιολογικό σύστημα που μελετάται. Συνήθως τα δεδομένα αυτά είναι αρκετά θορυβώδη εξαιτίας διαφόρων πηγών διακύμανσης στα πειράματα. Για την απόκτηση αξιόπιστων δεδομένων γονιδιακής έκφρασης που να μπορούν να χρησιμοποιηθούν στην περαιτέρω ανάλυση και ερμηνεία απαιτείται οι πειραματικές διαδικασίες να είναι αυστηρά τυποποιημένες.

Εκτός από την τυποποίηση των πειραματικών διαδικασιών, έχουν αναπτυχθεί διάφορες μέθοδοι επεξεργασίας που ελαχιστοποιούν το θόρυβο και τη διακύμανση που δεν οφείλονται στη διαφορική έκφραση των γονιδίων. Πρόκειται για μεθόδους διόρθωσης του θορύβου υποβάθρου, φιλτραρίσματος και κανονικοποίησης.

Επειδή ο προσδιορισμός προτύπων γονιδιακής έκφρασης και η ταξινόμηση των γονιδίων σε ομάδες έκφρασης παρέχει καλύτερη κατανόηση του βιολογικού τους ρόλου και της μεταξύ τους σχέσης, μετά την προεπεξεργασία των δεδομένων χρησιμοποιούνται διάφορες τεχνικές ανάλυσης των δεδομένων, τόσο μη επιβλεπόμενης μάθησης, όπως η ιεραρχική ομαδοποίηση (hierarchical clustering), η ομαδοποίηση k-μέσων (k-means clustering) και οι αυτο-οργανούμενοι χάρτες (SOM-Self Organizing Maps), όσο και επιβλεπόμενης μάθησης, όπως οι μηχανές υποστήριξης διανυσμάτων (SVM-Support Vector Machines), που βασίζονται στην τεχνολογία των νευρωνικών δικτύων. Επίσης για τη μείωση της διαστατικότητας των δεδομένων, για άμεση οπτικοποίηση και για την ανακάλυψη των γονιδίων που συνεισφέρουν από κοινού περισσότερο στη διακύμανση των δεδομένων, χρησιμοποιείται η ανάλυση πρωτευουσών συνιστωσών (PCA-principal component analysis).

Η ομαδοποίηση των γονιδίων μπορεί να οδηγήσει στον προσδιορισμό των παρόμοια εκφρασμένων γονιδίων. Δεδομένου ότι υπάρχουν ενδείξεις ότι πολλά

λειτουργικά συνδεδεμένα γονίδια εκφράζονται παρόμοια, καθώς επίσης και για το ότι η παρόμοια έκφραση γονιδίων μπορεί να είναι αποτέλεσμα της παρόμοιας ρύθμισης (π.χ. δύο γονίδια που είναι όμοια εκφρασμένα μπορεί να ρυθμίζονται από τον ίδιο μεταγραφικό παράγοντα), είναι σημαντική η ανάπτυξη εργαλείων επεξεργασίας και ανάλυσης δεδομένων, που θα συντελούν στην εξαγωγή ορθών συμπερασμάτων. Η χρήση διαφορετικών τύπων προεπεξεργασίας και ομαδοποίησης, μπορούν να αποκαλύψουν διαφορετικές όψεις της γονιδιακής έκφρασης. Για το λόγο αυτό η επιλογή τους πρέπει να συνδυάζεται με άρτια γνώση του βιολογικού υπόβαθρου.

Στην παρούσα εργασία έχει αναπτυχθεί σε περιβάλλον Matlab, ένα εργαλείο επεξεργασίας και ανάλυσης δεδομένων από τυπωμένες cDNA μικροσυστοιχίες, με δυνατότητα επιλογής από το χρήστη τόσο μεταξύ διαφορετικών αλγορίθμων προεπεξεργασίας και ομαδοποίησης, όσο και διαφορετικών παραμέτρων τους. Τέλος παρουσιάζονται τα αποτελέσματα της εφαρμογής του σε δεδομένα από cDNA μικροσυστοιχίες κυτταρικών σειρών λευχαιμίας T-λεμφοκυττάρων (CCRFM-CEM).

2. ΛΕΥΧΑΙΜΙΑ

2.1. ΚΑΡΚΙΝΟΣ

Ο **καρκίνος** προκαλείται από μεταλλάξεις γονιδίων που ελέγχουν τον κυτταρικό πολλαπλασιασμό και χαρακτηρίζεται από τον ανεξέλεγκτο πολλαπλασιασμό των κυττάρων ενός ιστού. Αυτά σχηματίζουν μάζες κυττάρων (καρκινικοί όγκοι) ή μεταναστεύουν στο αίμα, όπως στις διάφορες μορφές λευχαιμίας (Alison 2001).

Υπάρχουν δύο τύποι γονιδίων που σχετίζονται με την καρκινογένεση. Τα **ογκογονίδια** και τα **ογκοκατασταλτικά γονίδια**. Τα ογκογονίδια "προέρχονται" από γονίδια που υπάρχουν φυσιολογικά στο ανθρώπινο γονιδίωμα και ονομάζονται **πρωτο-ογκογονίδια**. Στα φυσιολογικά κύτταρα, τα πρωτο-ογκογονίδια κωδικοποιούν πρωτεΐνες οι οποίες εμπλέκονται στον έλεγχο της αύξησης. Η ακολουθία των γεγονότων που οδηγούν σε αύξηση του κυττάρου, αρχίζει με την αλληλεπίδραση σηματοδοτικών μορίων (π.χ. αυξητικών παραγόντων, ορμονών, νευροδιαβιβαστών) με ειδικούς υποδοχείς, είτε στην επιφάνεια είτε στο εσωτερικό του κυττάρου. Το κύτταρο ερμηνεύει το σήμα και παράγει μεταγωγικά μόρια (transducers), ή, αλλιώς, δεύτερους αγγελιοφόρους (second messengers), τα οποία τροποποιούν τη μεταγραφή των γονιδίων στον πυρήνα ή επιδρούν με άμεσο τρόπο σε διάφορες κυτταρικές πρωτεΐνες, με αποτέλεσμα να επέρχονται καθορισμένες φαινοτυπικές μεταβολές.

Όμως διάφορα είδη μεταλλάξεων, που μπορεί να προκληθούν από μεταλλαξογόνους παράγοντες, μετατρέπουν τα πρωτο-ογκογονίδια σε ογκογονίδια, τα οποία υπερλειτουργούν και οδηγούν το κύτταρο σε ανεξέλεγκτο πολλαπλασιασμό και στη δημιουργία καρκίνου. Η μετατροπή ενός πρωτο-ογκογονιδίου σε ογκογονίδιο μπορεί να είναι το αποτέλεσμα μιας **σημειακής μετάλλαξης** ή μιας **χρωμοσωμικής ανωμαλίας**, συνηθέστερα μετατόπισης.

Τα ογκοκατασταλτικά γονίδια είναι γονίδια που ελέγχουν την κυτταρική διαίρεση, καταστέλλοντας την, όποτε είναι απαραίτητο. Η αναστολή της δράσης τους που είναι συνήθως αποτέλεσμα μετάλλαξης, κυρίως έλλειψης, αφαιρεί από το κύτταρο τη δυνατότητα ελέγχου του πολλαπλασιασμού και οδηγεί σε καρκινογένεση (Greaves, Wiemels 2003, Mullighan et al. 2007).

Σε γενετικό επίπεδο ο καρκίνος είναι το αποτέλεσμα:

- Μετατροπής πρωτο-ογκογονιδίων σε ογκογονίδια
- Απουσίας λειτουργικότητας ογκοκατασταλτικών γονιδίων και
- Αδρανοποίησης των μηχανισμών επιδιόρθωσης του DNA.

Τα καρκινικά κύτταρα ταξιδεύουν συχνά μέσω της κυκλοφορίας του αίματος ή μέσω του συστήματος λέμφων σε άλλα μέρη του σώματος, όπου αρχίζουν να μεγαλώνουν και να αντικαθιστούν τους φυσιολογικούς ιστούς. Αυτή η διαδικασία εξάπλωσης ονομάζεται **μετάσταση**. Οι καλοήθειες (μη-καρκινοειδείς) όγκοι δεν εξαπλώνονται σε άλλα μέρη του σώματος (μεταστατικοί) και είναι πολύ σπάνια απειλητικοί για τη ζωή.

2.2. ΤΟ ΑΙΜΑ ΚΑΙ ΤΟ ΣΥΣΤΗΜΑ ΤΗΣ ΛΕΜΦΟΥ

Το αίμα είναι το υγρό που κυκλοφορεί στο αγγειακό σύστημα των ανθρώπων και των ζώων. Η κυκλοφορία του διαμέσου της καρδιάς, των αρτηριών, των φλεβών και των τριχοειδών αγγείων επιτελεί διάφορες λειτουργίες όπως η μεταφορά θρεπτικών ουσιών, ορμονών, βιταμινών, θερμότητας και οξυγόνου στους ιστούς και την απομάκρυνση άχρηστων ουσιών που παράγονται κατά τον μεταβολισμό και του διοξειδίου του άνθρακα. Επιπρόσθετα, παρέχει άμυνα κατά των λοιμώξεων μέσω της δράσης των λευκών αιμοσφαιρίων και των αντισωμάτων. Το αίμα αποτελεί έναν εξαιρετικά εξειδικευμένο κυκλοφορούντα ιστό, ο οποίος αποτελείται από διάφορους τύπους κυττάρων που συγκρατούνται μέσα σε ένα υγρό μέσο που ονομάζεται πλάσμα.

Ο μυελός των οστών είναι το μαλακό εσωτερικό τμήμα μερικών οστών όπως το κρανίο, η ωμοπλάτη, τα πλευρά, η λεκάνη, και η σπονδυλική στήλη. Μακροσκοπικά, ανάλογα με την εμφάνισή τους έχουν περιγραφεί δύο τύποι: ο **ερυθρός** ή **ενεργός μυελός των οστών**, που το χρώμα του οφείλεται στην παρουσία πολλών ερυθροκυττάρων και των πρόδρομων μορφών τους και ο κίτρινος μυελός των οστών, πλούσιος σε λιποκύτταρα, που δεν παράγει κύτταρα του αίματος. Ο κίτρινος μυελός των οστών διατηρεί το αιμοποιητικό του δυναμικό.

Ο ερυθρός μυελός των οστών αποτελείται από διακλαδιζόμενα τριχοειδικά κολποειδή και δικτυωτό υπόστρωμα. Οι διάμεσοι χώροι ανάμεσα στα κολποειδή καταλαμβάνονται από τις αιμοποιητικές χορδές, που περιέχουν αρχέγονα και πρόδρομα κύτταρα του αίματος, τα οποία βρίσκονται σε διάφορα στάδια διαφοροποίησης και ωρίμανσης. Το δικτυωτό υπόστρωμα, αποτελεί ένα πλέγμα στήριξης των αιμοποιητικών κυττάρων.

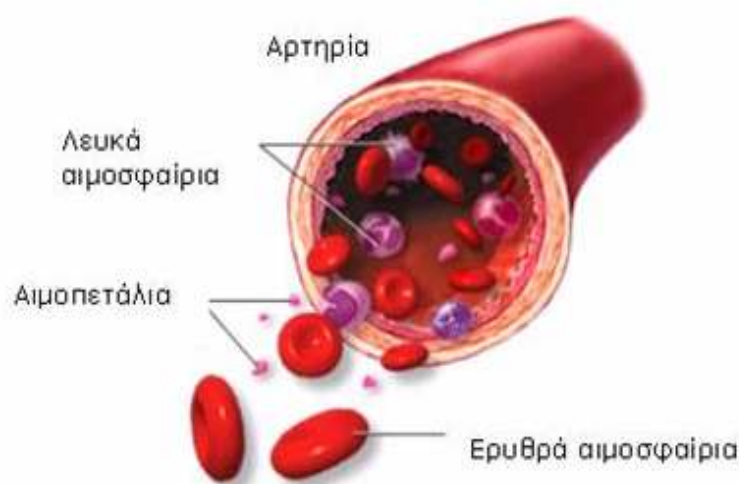
Ο μυελός των οστών παίζει σημαντικό ρόλο στην αιμοποιητική λειτουργία, αφού παράγει αιμοποιητικούς αυξητικούς παράγοντες που είναι σημαντικοί για το μηχανισμό ελέγχου της αύξησης των διαφόρων τύπων αιμοποιητικών αρχέγονων κυττάρων και συμμετέχει ενεργά στο ανοσοποιητικό σύστημα, γιατί αποτελεί τη θέση ωρίμανσης των Β λεμφοκυττάρων (Ιστολογία – Εμβρυολογία 2009).

Όλα τα αιμοσφαίρια προέρχονται από το ίδιο αρχικό αρχέγονο ή πρόδρομο κύτταρο, και νωρίς στην ανάπτυξή τους χωρίζονται σε δύο κύριες οικογένειες - τη **μυελοειδή** (myeloid) και τη **λεμφοειδή** (lymphoid) οικογένεια.

Η μυελοειδής οικογένεια περιλαμβάνει όλα τα ερυθρά αιμοσφαίρια, αιμοπετάλια και ορισμένα λευκά αιμοσφαίρια, τα *κοκκιοκύτταρα* (*granulocytes*) και τα *μονοκύτταρα* (*monocytes*). Η λεμφοειδής οικογένεια περιλαμβάνει όλα τα άλλα λευκά αιμοσφαίρια.

Αν και τα λευκά αιμοσφαίρια αυτών των δύο οικογενειών καταπολεμούν από κοινού τους λοιμώδεις παράγοντες, οι μέθοδοί τους διαφέρουν. Τα μυελοειδή λευκά αιμοσφαίρια είναι η άμεση "υπεράσπιση κατά της λοίμωξης" του σώματος. Τα λεμφοειδή κύτταρα χρειάζονται περισσότερο χρόνο για να επιδράσουν αλλά είναι πιο ειδικευμένα η δράση τους ενάντια σε εισβάλλοντες οργανισμούς. Τα λεμφοειδή κύτταρα απαντώνται επίσης στο λεμφικό ιστό σε όλα τα μέρη του σώματος.

Το **λεμφικό σύστημα** είναι το σύστημα συλλογής, φιλτραρίσματος και παροχέτευσης αποβλήτων του σώματος. Ωστόσο, ακριβώς όπως το κυκλοφορικό σύστημα μεταφέρει το αίμα, το λεμφικό σύστημα μεταφέρει το διαφανές ρευστό που αποκαλείται λέμφος, το οποίο βοηθά στη μεταφορά λεμφοκυττάρων. Έτσι, τα λεμφοκύτταρα απαντώνται και στο αίμα και στη λέμφο (Pui 2006, Keene 2002).



Εικόνα 2.1: Αίμα (Παγκύπριος Αντίλευχαιμικός Σύνδεσμος "ΖΩΗ" 2009).

Τα ερυθρά αιμοσφαίρια περιέχουν αιμοσφαιρίνη για τη μεταφορά οξυγόνου από τους πνεύμονες σε όλα τα μέρη του σώματος. Παίρνουν τα απόβλητα καθώς κυκλοφορούν και τα μεταφέρουν στους πνεύμονες, όπου εκπνέονται ως διοξείδιο του άνθρακα. Η έλλειψη ερυθρών αιμοσφαιρίων αποκαλείται αναιμία.

Τα αιμοπετάλια είναι στην πραγματικότητα κυτταρικά τμήματα που παράγονται από έναν τύπο κυττάρου του μυελού των οστών που ονομάζεται **μεγακαρυοκύτταρο**. Τα αιμοπετάλια είναι σημαντικά για την πήξη του αίματος και την αιμόσταση. Βοηθούν την επούλωση, στα αιμοφόρα αγγεία, τραυμάτων που έχουν προκληθεί από τραύματα ή μώλωπες. Η έλλειψη των αιμοπεταλίων καλείται θρομβοπενία.

Τα λευκά αιμοσφαίρια είναι σημαντικά για την υπεράσπιση του σώματος ενάντια στις μολύνσεις. Τα **λεμφοκύτταρα** είναι ένας τύπος λευκών κυττάρων του αίματος. Οι άλλοι τύποι λευκών κυττάρων του αίματος είναι τα κοκκιοκύτταρα (ουδετερόφιλα, βασεόφιλα, και ηωσινόφιλα) και τα μονοκύτταρα.

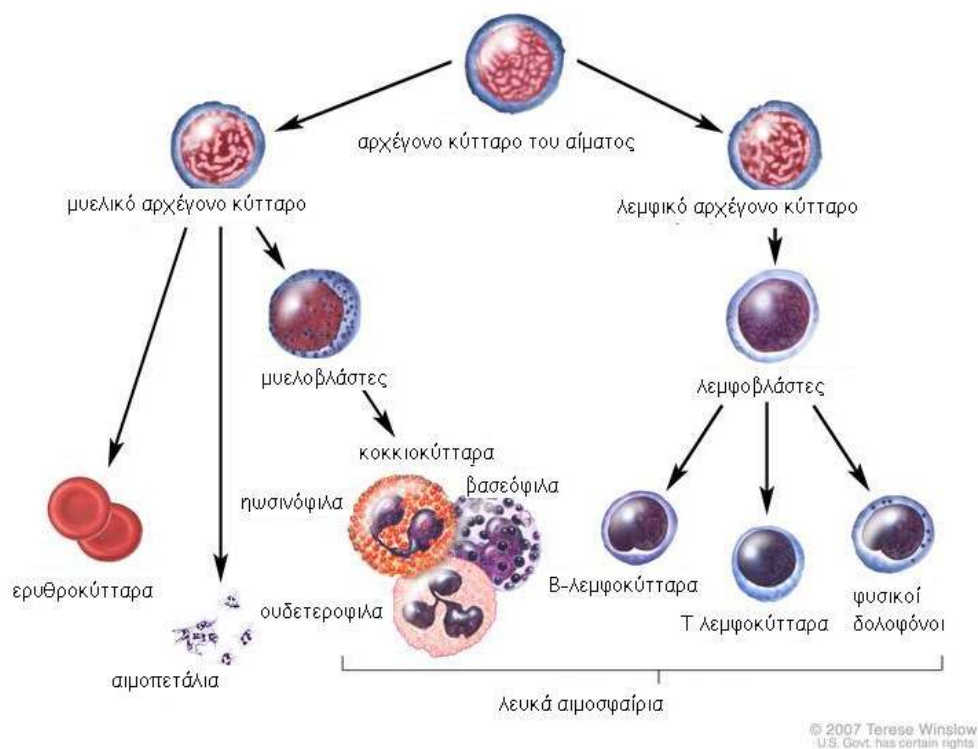
Τα λεμφοκύτταρα είναι τα βασικά κύτταρα του λεμφικού ιστού, ένα σημαντικό τμήμα του ανοσοποιητικού συστήματος. Ο λεμφικός ιστός βρίσκεται στους λεμφαδένες, στο θύμο αδένα, τον σπλήνα, τις αμυγδαλές και τις αδενοειδείς εκβλαστήσεις, και υπάρχει και σε όλο το γαστρεντερικό και αναπνευστικό σύστημα και στο μυελό των οστών.

Τα λεμφοκύτταρα αναπτύσσονται από τα κύτταρα που καλούνται **λεμφοβλάστες** για να γίνουν ώριμα κύτταρα που καταπολεμούν τις μολύνσεις. Οι δύο τύποι

λεμφοκυττάρων είναι **B-λεμφοκύτταρα** και τα **T-λεμφοκύτταρα**. Τα B-λεμφοκύτταρα προστατεύουν το σώμα από την εισβολή μικροβίων με το να αναπτύσσονται (ωριμάζουν) στα κύτταρα του πλάσματος, δημιουργώντας τα αντισώματα. Αυτά τα αντισώματα προσκολλώνται στα μικρόβια, όπως τα βακτηρίδια, τους ιούς, και τους μύκητες. Μόλις το μικρόβιο «περικυκλωθεί» από τα αντισώματα, άλλα λευκά αιμοσφαίρια αποκαλούμενα κοκκιοκύτταρα μπορούν να το αναγνωρίσουν και να το καταστρέψουν. Τα T-λεμφοκύτταρα μπορούν να αναγνωρίσουν τα κύτταρα που μολύνονται από ιούς και να τα καταστρέψουν άμεσα.

Τα κοκκιοκύτταρα είναι λευκά αιμοσφαίρια με εμφανή κυτταροπλασματικά κοκκία, που μπορούν να διακριθούν στο μικροσκόπιο. Αυτά τα κοκκία περιέχουν ένζυμα και άλλες ουσίες που μπορούν να καταστρέψουν μικρόβια, όπως τα βακτηρίδια. Οι τρεις τύποι κοκκιοκυττάρων είναι τα ουδετερόφιλα, τα βασεόφιλα, και τα ηωσινόφιλα, που διακρίνονται από το μέγεθος και το χρώμα των κοκκίων τους. Τα κοκκιοκύτταρα δημιουργούνται από τα αιμοποιητικά κύτταρα που καλούνται **μυελοβλάστες**, για να γίνουν ώριμα κύτταρα ενάντια των μολύνσεων.

Τα μονοκύτταρα, που συσχετίζονται με τα κοκκιοκύτταρα, είναι επίσης σημαντικά για την προστασία του σώματος από τα βακτηρίδια. Αρχίζουν να παράγονται στο μυελό των οστών ως αιμοποιητικοί **μονοβλάστες** και αναπτύσσονται σε ώριμα μονοκύτταρα. Αφού κυκλοφορήσουν στο αίμα για μια ημέρα, τα μονοκύτταρα εισέρχονται στους ιστούς του σώματος και γίνονται μακρόφαγα, τα οποία μπορούν να καταστρέψουν μερικά μικρόβια περικυκλώνοντας τα και αφομοιώνοντας τα. Τα μακρόφαγα επίσης βοηθούν σημαντικά τα λεμφοκύτταρα στο να αναγνωρίζουν τα μικρόβια και να αρχίσουν να παράγουν αντισώματα για να τα καταπολεμήσουν (Pui 2006, Keene 2002).



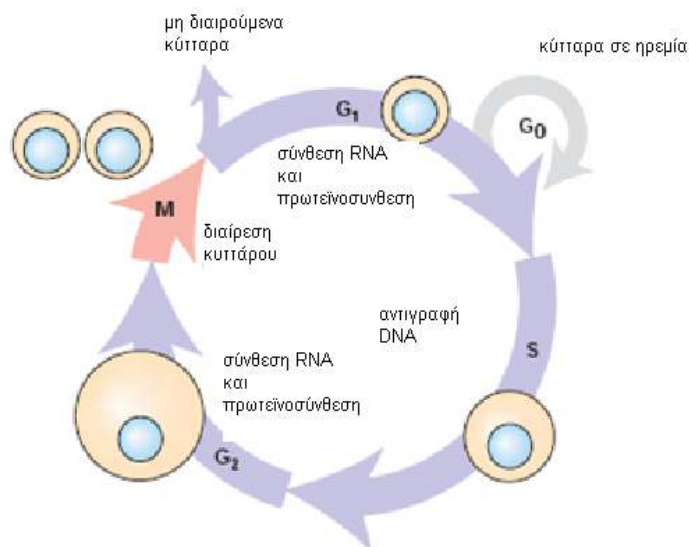
Εικόνα 2.2: Σχηματισμός των διαφοροποιημένων κυττάρων του αίματος από αρχέγονο αιμοποιητικό κύτταρο, στο μυελό των οστών (Medical Library 2008).

2.3. ΛΕΥΧΑΙΜΙΕΣ

2.3.1. Γενικά

Η λευχαιμία εμφανίζεται όταν το σώμα αρχίζει να συσσωρεύει εμφανώς μη φυσιολογικά λευκά αιμοσφαίρια. Τα «ανώμαλα» αυτά λευχαιμικά κύτταρα, επειδή δεν μπορούν να ωριμάσουν κανονικά, συσσωρεύονται στο μυελό των οστών. Τελικά όλα τα φυσιολογικά λευκά και ερυθρά αιμοσφαίρια και αιμοπετάλια εξωθούνται ή δεν αντικαθίστανται. Ο υγιής μυελός των οστών αντικαθίσταται από τα ανώριμα κύτταρα που εισέρχονται τελικά στο αίμα και μεταφέρονται σ' όλο το σώμα. Έτσι, ενώ ο αριθμός ανώριμων κυττάρων στο αίμα αυξάνεται, ο αριθμός φυσιολογικών ερυθρών και λευκών αιμοσφαιρίων και αιμοπεταλίων μειώνεται.

Οι λευχαιμίες και τα λεμφώματα αποτελούν παραδείγματα ελαττωματικής διαφοροποίησης που οδηγεί σε ανάπτυξη κακοήθων όγκων. Δεδομένου ότι όλοι οι διαφορετικοί τύποι κυττάρων αίματος προέρχονται από ένα κοινό αρχέγονο κύτταρο στο μυελό των οστών, οι απόγονοι αυτών των κυττάρων δημιουργούν τις διάφορες λεμφικές σειρές κυττάρων, με τη δέσμευση τους σε συγκεκριμένες διαδικασίες διαφοροποίησης. Τα κύτταρα κάθε τύπου υποβάλλονται σε διαφορετικούς κύκλους διαίρεσης καθώς διαφοροποιούνται, αλλά μόλις διαφοροποιηθούν πλήρως, η κυτταρική διαίρεση παύει.



Εικόνα 2.3: Κατά τη διάρκεια της ανάπτυξης, τα ευκαρυωτικά κύτταρα συνεχώς περνούν από τις τέσσερις φάσεις του κυτταρικού κύκλου, δημιουργώντας νέα θυγατρικά κύτταρα (Lodish, Darnell & etc. 2003).

Αντίθετα, τα λευχαιμικά κύτταρα ή τα κύτταρα λεμφωμάτων αποτυγχάνουν να υποβληθούν στην τελική διαφοροποίηση. Αντ' αυτού, σταματούν στα πρώτα στάδια της ωρίμανσης, και διατηρούν την ικανότητά τους για πολλαπλασιασμό, συνεχίζουν να αναπαράγονται και συσσωρεύονται. Έτσι με την πάροδο του χρόνου, αυτά τα κύτταρα εισέρχονται στην κυκλοφορία του αίματος και διαδίδονται σε άλλα όργανα, όπου μπορούν να αποτρέπουν άλλα κύτταρα του σώματος από τη φυσιολογική τους

λειτουργία. Οποιοδήποτε αιμοποιητικό κύτταρο ή λεμφοκύτταρο από το μυελό των οστών μπορεί να μετατραπεί σε λευχαιμικό κύτταρο (Alison 2001).

Κάθε ένα από αυτά τα κύτταρα έχει τη λειτουργία του, διαφέρει μορφολογικά από τα άλλα και έχει μια διάρκεια ζωής χαρακτηριστική για εκείνο τον ιδιαίτερο τύπο κυττάρων. Στους υγιείς ανθρώπους η καταστροφή και η παραγωγή των λεμφοκυττάρων είναι ισορροπημένες και, επομένως, ο αριθμός κυττάρων που υπάρχουν στο αίμα σε μια συγκεκριμένη χρονική στιγμή είναι σχετικά σταθερός. Σε ένα υγιές άτομο, μόνο τα ώριμα ενήλικα κύτταρα βρίσκονται στο περιφερικό αίμα, ενώ σε πολλές ασθένειες, μπορεί να υπάρχουν ανώριμες και μη φυσιολογικές μορφές των κυττάρων.

Κατά τη διάρκεια της διαφοροποίησης και της ωρίμανσης των αιματοποιητικών αρχέγονων κυττάρων, τα κύτταρα υφίστανται αξιοσημείωτες αλλαγές στη δομή, το μέγεθος και τη μορφή τους. Η μετάβαση από το ένα στάδιο στο άλλο δεν εμφανίζεται να έχει σαφή όρια, αλλά η έκφραση διαφορετικών μορίων στην επιφάνεια των κυττάρων επιτρέπει τη διάκριση των διαφορετικών προδρόμων σταδίων. Επιπλέον, οι μεταβολικές οδοί, το κυτταροπλασματικό περιεχόμενο και η δομή χρωματίνης του πυρήνα διαφέρουν μεταξύ των διαφορετικών σταδίων και εξαρτώνται από τη οικογένεια στην οποία ανήκουν.

Η αιμοποίηση ελέγχεται στενά από πολυάριθμες κυτταροκίνες και αυξητικούς παράγοντες. Στις αιματολογικές κακοήθειες τα πρωτεύοντα και δευτερεύοντα όργανα της αιμοποίησης (περιοχές της εξωμυελικής αιμοποίησης) μπορούν να διεισδυθούν από τα κακοήθη κύτταρα, με συνέπεια την απώλεια της αιμοποιητικής τους δυνατότητας.

Η διαφοροποίηση των αιματοποιητικών κυττάρων σε μια συγκεκριμένη κατηγορία καθορίζεται σε μεγάλο βαθμό από ειδικούς αιμοποιητικούς παράγοντες αύξησης και των κυτταρικών τους υποδοχέων. Μόλις αυτοί οι αιμοποιητικοί αυξητικοί παράγοντες προσδεθούν στους υποδοχείς τους, η ιδιαιτερότητα τους επιτυγχάνεται με την ενεργοποίηση συγκεκριμένων ενδοκυτταρικών σημάτων. Αυτό επιτρέπει μόνο ένας περιορισμένος αριθμός κυττάρων που βρίσκεται σε ένα συγκεκριμένο στάδιο διαφοροποίησης, να ανταποκριθεί στον αυξητικό παράγοντα στο κυτταρικό περιβάλλον. Ο έλεγχος της αιμοποίησης απαιτεί και αυστηρούς κανονισμούς για τον περιορισμό του πολλαπλασιασμού τους. Η μίτωση μπορεί να περιοριστεί από διάφορους ανασταλτικούς παράγοντες, όπως ο μετασχηματισμένος παράγοντας αύξησης -/3 (TGF-/3) (Secker-Walker 1996).

2.3.2. Τύποι Λευχαιμίας

Οι τέσσερις συνήθεις τύποι λευχαιμίας είναι:

1. Οξεία Λεμφοβλαστική Λευχαιμία (*Acute Lymphoblastic Leukaemia - ALL*)
2. Χρόνια Λεμφοβλαστική Λευχαιμία (*Chronic Lymphoblastic Leukaemia - CLL*)
3. Οξεία Μυελοειδής Λευχαιμία (*Acute Myeloid Leukaemia - AML*)

4. Χρόνια Μυελοειδής Λευχαιμία (*Chronic Myeloid Leukaemia - CML*)

Ο πρώτος παράγοντας που εξετάζεται για την ταξινόμηση της λευχαιμίας ενός ασθενή είναι εάν τα περισσότερα από τα ανώμαλα κύτταρα είναι ώριμα (μοιάζουν με κανονικά λευκά αιμοσφαίρια) ή ανώριμα (μοιάζουν περισσότερο με αρχέγονα κύτταρα).

Στην **οξεία λευχαιμία** (Acute leukaemia), τα κύτταρα του μυελού των οστών δεν μπορούν να ωριμάσουν κατάλληλα. Τα ανώριμα κύτταρα λευχαιμίας συνεχίζουν να αναπαράγονται και να αναπτύσσονται. Μερικοί τύποι οξείων λευχαιμιών αποκρίνονται καλά στην θεραπεία, και πολλοί ασθενείς μπορούν να θεραπευτούν. Άλλοι τύποι οξείων λευχαιμιών έχουν μια λιγότερο ευνοϊκή προοπτική.

Στην **χρόνια λευχαιμία** (Chronic leukaemia), τα κύτταρα μπορούν να ωριμάσουν εν μέρει αλλά όχι εντελώς. Αυτά τα κύτταρα δεν είναι πραγματικά κανονικά. Γενικά δεν αντιμετωπίζουν τη μόλυνση τόσο καλά όσο τα κανονικά λευκά αιμοσφαίρια. Και φυσικά, επιζούν περισσότερο, αναπτύσσονται, και υπερνικούν τα κανονικά κύτταρα. Οι χρόνιες λευχαιμίες εξελίσσονται με το πέρασμα του χρόνου και οι περισσότεροι ασθενείς μπορούν να ζήσουν για πολλά έτη. Ωστόσο, είναι γενικά πιο δύσκολο να θεραπευτούν από ότι οι οξείες.

Οι λευχαιμίες είναι είτε **μυελοειδείς** είτε **λεμφοειδείς**. Έτσι, όταν η λευχαιμία προσβάλλει τα κύτταρα, που πρόκειται να σχηματίσουν τελικά αιμοπετάλια, ερυθρά αιμοσφαίρια, κοκκιοκύτταρα και μονοκύτταρα, αποκαλείται μυελοειδής (myeloid), μυελοκυτταρική (myelocytic), μυελογενής (myelogenous), ή κοκκιοκυτταρική (granulocytic) λευχαιμία. Αντιθέτως όταν η λευχαιμία προσβάλλει τα κύτταρα που πρόκειται να γίνουν λεμφοκύτταρα αποκαλείται λεμφοβλαστική (lymphoblastic), λεμφοειδής (lymphoid), λεμφοκυτταρική (lymphocytic), ή λεμφική (lymphatic) λευχαιμία (Keene 2002).

2.3.2.1. Οξεία λεμφοβλαστική λευχαιμία (Acute lymphoblastic leukaemia – ALL)

Αυτός ο τύπος λευχαιμίας είναι η πιο κοινή παιδική λευχαιμία, αλλά εμφανίζεται και σε ενήλικες. Τα μη φυσιολογικά κύτταρα στην ALL είναι ανώριμα λεμφοκύτταρα που αποκαλούνται λεμφοβλάστες (lymphoblasts). Τα λεμφοκύτταρα ενεργούν για να προστατεύουν το σώμα από λοίμωξη. Όταν ο μυελός των οστών προσβάλλεται από ALL, η παραγωγή κυττάρων που καταπολεμούν τη λοίμωξη μειώνεται, και μπορεί να επέλθει σοβαρή λοίμωξη. Επιπλέον, η λευχαιμία μπορεί επίσης να εξωθήσει τα αιμοπετάλια και τα κύτταρα που παράγουν ερυθρά αιμοσφαίρια.

2.3.2.2. Χρόνια λεμφοβλαστική λευχαιμία (Chronic lymphoblastic leukaemia – CLL)

Είναι χρόνιος ή αργά εξελισσόμενος τύπος λευχαιμίας. Όπως και με την ALL, προσβάλλονται τα λεμφοκυτταρικά κύτταρα. Αυτή η ασθένεια εμφανίζεται συνήθως

σε ηλικιωμένους ασθενείς. Επειδή η ασθένεια εξελίσσεται αργά, τα φυσιολογικά λεμφοκύτταρα δεν εξωθούνται τόσο γρήγορα όσο στην οξεία μορφή. Αυτό σημαίνει ότι ασθενείς με CLL μπορεί να μην αισθανθούν καθόλου ασυνήθιστα συμπτώματα μέχρι τα μεταγενέστερα στάδια της ασθένειας.

2.3.2.3. Οξεία μυελοειδής λευχαιμία (Acute myeloid leukemia – AML)

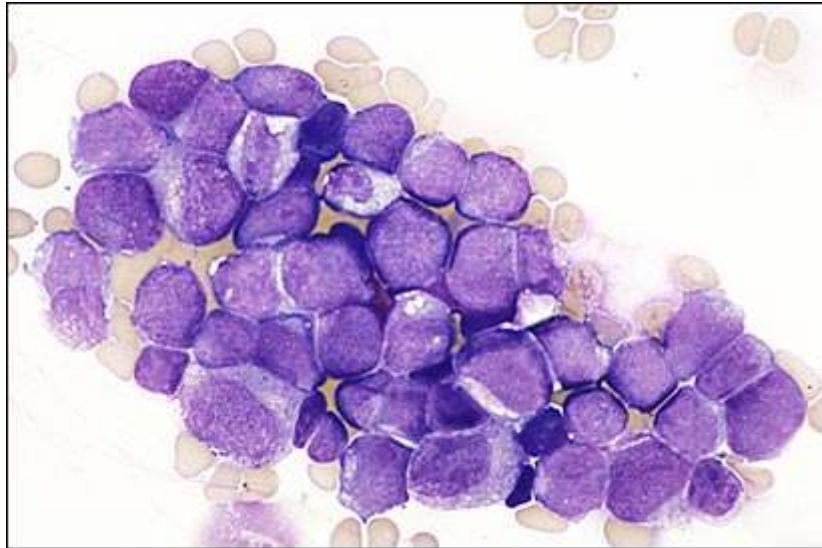
Η οξεία μυελοειδής λευχαιμία προσβάλλει τη μυελοειδή κατηγορία κυττάρων. Σ' αυτό τον τύπο λευχαιμίας προσβάλλονται κυρίως τα κοκκιοκύτταρα. Τα μυελοειδή κύτταρα κυκλοφορούν στο αίμα και ψάχνουν για λοιμώδεις εισβολείς στο αίμα και στο γύρω ιστό. Τα κοκκιοκύτταρα αντιδρούν γρήγορα σ' όλους τους τύπους εισβολέων ενώ τα λεμφοειδή κύτταρα χρειάζονται περισσότερο χρόνο να αντιδράσουν και ο ρόλος τους στην καταπολέμηση της λοίμωξης είναι πιο εξειδικευμένος. Όταν τα μονοκύτταρα της μυελοειδούς οικογένειας ανακαλύπτουν εισβολείς, τους «περικυκλώνουν» ή τους απορροφούν και τους καταστρέφουν. Η συσσώρευση ανώριμων ή "βρεφικών" μυελοειδών κυττάρων μπορεί να προκαλέσει αποφράξεις των αιμοφόρων αγγείων.

2.3.2.4. Χρόνια μυελοειδής λευχαιμία (Chronic myeloid leukemia – CML)

Τα μη φυσιολογικά μυελοειδή κύτταρα αυτού του τύπου λευχαιμίας είναι συνήθως ώριμα στην εμφάνιση αλλά δεν λειτουργούν φυσιολογικά. Αυτή η λευχαιμία συνδέεται με έναν ανώμαλο δείκτη στα κύτταρα που αποκαλείται «**χρωμόσωμα Φιλαδέλφειας**» (Philadelphia chromosome) και προσβάλλει νεαρότερους ασθενείς συχνότερα από την CLL. Η CML διαφέρει επίσης από τους πιο κοινούς τύπους λευχαιμίας επειδή έχει δύο στάδια. Το πρώτο είναι ο αργός, χρόνιος πολλαπλασιασμός (*chronic multiplication*) των μη φυσιολογικών κυττάρων. Ορισμένες φορές, η ασθένεια μπορεί ξαφνικά να μεταστραφεί σε οξύ στάδιο όπου και γίνεται όπως ακριβώς και η οξεία λευχαιμία (Pui 2006, Keene 2002, Pui et al. 2004).

2.4. ΟΞΕΙΑ ΛΕΜΦΟΒΛΑΣΤΙΚΗ ΛΕΥΧΑΙΜΙΑ

Η οξεία λεμφοβλαστική λευχαιμία είναι, κατά κύριο λόγο νόσος της παιδικής ηλικίας και λιγότερο συχνά των ενηλίκων. Όπως αναφέρθηκε, αποτελεί νεοπλασματική νόσο που προέρχεται από τη μη ελεγχόμενη από ρυθμιστικούς μηχανισμούς αύξηση των αώρων κυττάρων της λεμφικής σειράς (λεμφοβλάστες), τα οποία καταλαμβάνουν το μυελό των οστών, εκτοπίζοντας τις φυσιολογικές σειρές του αίματος.



Εικόνα 2.4: Οξεία Λεμφοβλαστική Λευχαιμία. Μυελός των οστών. Πλήρης αντικατάσταση από λεμφοβλάστες μικρού/μεσαίου μεγέθους με το παραγγειακό κυτταρόπλασμα και στρογγυλοί πυρήνες με πυκνή χρωματίνη (L1 τύπος κατά FAB, φαινότυπος κοινής- ALL) (IBMS - The Institute of Biomedical Science 2009).

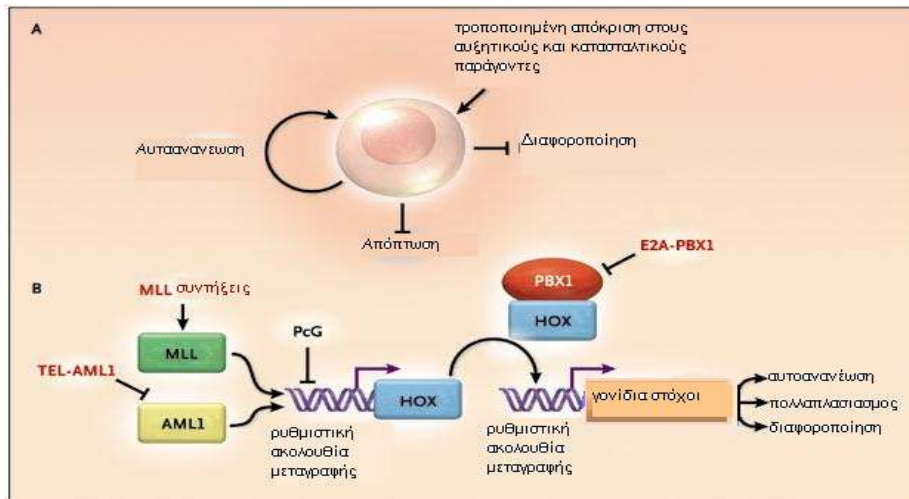
2.4.1. Αιτίες

Οι αιτίες της λευχαιμίας, των λεμφωμάτων, του μυελώματος και των σχετικών αιματολογικών διαταραχών δεν είναι ακριβώς γνωστές. Ωστόσο, παράγοντες όπως η κληρονομικότητα, η ακτινοβολία, διάφορες χημικές ουσίες και κάποιοι ιοί φαίνεται να επηρεάζουν την ανάπτυξή τους. Αυτοί οι παράγοντες κινδύνου αποτελούν μέρος μόνο της αιτίας σε ένα συγκεκριμένο ασθενή και εξηγούν έναν πολύ μικρό αριθμό του συνόλου των κρουσμάτων. Στις περισσότερες περιπτώσεις η αιτία παραμένει άγνωστη (Leukemia & Lymphoma Society 2009).

Οι χρωμοσωμικές μεταθέσεις, αποτελούν συνήθη γενετική βλάβη στις οξείες λεμφοβλαστικές λευχαιμίες. Η μοριακή ανάλυση των γονιδίων που βρίσκονται στα σημεία σύνδεσης των χρωμοσωμάτων στις μεταθέσεις των ALL, απέδειξε την συμμετοχή στην λευχαιμογένεση, ομάδας γονιδίων που παράγουν μεταγραφικούς παράγοντες. Οι μεταγραφικοί παράγοντες παίζουν το ρόλο «διακόπτη», σαν θετικοί ή αρνητικοί ρυθμιστές της λειτουργίας άλλων γονιδίων, ρυθμίζοντας κύρια τη διαφοροποίηση και λιγότερο τον πολλαπλασιασμό των κυττάρων του αίματος. Στην ίδια ομάδα ανήκουν και τα ομειωτικά γονίδια (homeobox genes), στα οποία αποδίδεται η διαφοροποίηση κατά την εμβρυική ανάπτυξη.

Στις μεταθέσεις των λεμφοκυτταρικών λευχαιμιών σταθερός εταίρος εκτός των προαναφερθέντων μεταγραφικών και ομειωτικών γονιδίων, εμφανίζονται τα γονίδια τα απόλυτα ειδικά των λεμφοκυττάρων, δηλαδή των ανοσοσφαιρινών (IgH, IgK, IgA), ή των υποδοχέων των T κυττάρων (TCR) στις λευχαιμίες B και T προέλευσης αντίστοιχα. Η μεταφορά των μεταγραφικών ή ομειωτικών γονιδίων υπό την επιρροή των ισχυρών Ig ή TCR ενισχυτών, οδηγεί στην απορρύθμιση τους και την υπερπαραγωγή ενός κατά τα άλλα φυσιολογικού προϊόντος.

Ένα δεύτερο είδος μετάθεσης οδηγεί στην σύντηξη δύο γονιδίων, μεταγραφικών και/ή ομειωτικών, με αποτέλεσμα την παραγωγή μιας παθολογικής, χιμαιρικής πρωτεΐνης. Στις περιπτώσεις αυτές, η έκτοπη λόγω της μετάθεσης, έκφραση του ενός από τα δύο γονίδια θεωρείται ότι συμβάλλει στην εκτροπή των φυσιολογικών γεγονότων που οδηγούν στην διαφοροποίηση, με αποτέλεσμα την αναστολή της (Λουκόπουλος, Σταματόπουλος 1995).



Εικόνα 2.5: Μετασηματισμός των αιματοποιητικών κυττάρων στην παθογένεση της ALL. Η ανάπτυξη της λευχαιμίας απαιτεί ένα αρχέγονο αιματοποιητικό κύτταρο ή κάποιος από τους εντολοδόχους απογόνους να «ξεφύγει» από τους φυσιολογικούς μηχανισμούς του ομοιοστατικού ελέγχου, που ρυθμίζουν τα σήματα των αυξητικών παραγόντων, της διαφοροποίησης, της απόπτωσης και της αυτο-ανανέωσης (A). Μια κοινή μεταβολική οδός που σηματοδοτείται, από την παραγωγή χιμαιρικών αυξητικών παραγόντων, όπως συντήξεις MLL πρωτεϊνών, TEL-AML1, και E2A-PBX1, εξαιτίας μετατόπισης (B). (Pui et al. 2004).

Οι μεταθέσεις που χαρακτηρίζουν τις οξείες λεμφοβλαστικές λευχαιμίες B ή T προέλευσης διαφοροποιούνται στη βάση του γονιδίου του αντιγονικού υποδοχέα που συμμετέχει στη μετάθεση. Σημαντικές μεταθέσεις στις T-ALL, εμφανίζονται οι t(10;14)(q24;q11), t(1;14)(p32;q11) και η t(11;14)(p13;q11). Και στις τρεις συμμετέχει το γονίδιο του TCRδ (14q11), το οποίο συνδέεται στην πρώτη με το ομειωτικό HOX11 γονίδιο προκαλώντας την έκτοπη μεταγραφή του τελευταίου, στις δε επόμενες δύο, με τα μεταγραφικά TAL1/SCL και TAL2. Μια διαφορετική ομάδα μεταγραφικών γονιδίων, τα RHOMBE1 και RHOMBE2 (Ttg1 και Ttg2) συνδέονται με τα γονίδια του TCRδ και TCRβ στις μεταθέσεις t(11;14)(p15;q11), t(11;14)(p13;q11) και t(7;11)(q35;p13). Στη μετάθεση t(8;14)(q24;q11), το CMYC, πρότυπο της ομάδας γονιδίων που ελέγχουν τον κυτταρικό πολλαπλασιασμό, υπερλειτουργεί λόγω της σύνδεσης του με το TCRα (14q11) γονίδιο.

Εκτός των ειδικών γενετικών βλαβών, παρατηρούνται και άλλες μη ειδικές, όπως ο υπερδιπλοειδισμός. Ανάμεσα στις επιπρόσθετες βλάβες που εμφανίζονται συνήθως κατά την πορεία της νόσου, συχνά περιγράφονται οι σημειακές μεταλλάξεις του p53 γονιδίου. Τέτοιες βλάβες κατά κανόνα σηματοδοτούν εξέλιξη της νόσου και/ή ανάπτυξη αντοχής στη θεραπεία (Λουκόπουλος, Σταματόπουλος 1995).

2.4.2. Διάγνωση– Πρόγνωση

Τα συμπτώματα της λευχαιμίας και των σχετικών διαταραχών του μυελού των οστών είναι χαρακτηριστικά πολλών άλλων ασθενειών. Ορισμένοι τύποι λευχαιμίας δεν προκαλούν καθόλου συμπτώματα μέχρι να φτάσει η ασθένεια σε προχωρημένο στάδιο και μερικές φορές ούτε και τότε. Ωστόσο, αυτό δεν σημαίνει ότι ακόμη και η πολύ έγκαιρη διάγνωση θα απέτρεπε την επέκταση της ασθένειας, επειδή η λευχαιμία επεκτείνεται στο μυελό των οστών από τότε που αρχίζει (American Cancer Society 2007).

Η λευχαιμία εντοπίζεται με δύο εξετάσεις - εξέταση αίματος και βιοψία μυελού των οστών. Κατά την εξέταση αίματος λαμβάνεται δείγμα αίματος, αναλύεται με το μικροσκόπιο και μετριοούνται οι αριθμοί λευκών αιμοσφαιρίων και αιμοπεταλίων στο δείγμα. Αν ο αριθμός λευκών αιμοσφαιρίων είναι μη φυσιολογικός μπορεί να υπάρξει:

1. χαμηλός αριθμός αιμοπεταλίων,
2. χαμηλοί αριθμοί ερυθρών αιμοσφαιρίων,
3. Χαμηλός αριθμός ώριμων λευκών αιμοσφαιρίων ή υψηλοί αριθμοί ανώριμων ή πρόωρα αποδεσμευμένων λευκών αιμοσφαιρίων (που αποκαλούνται βλαστικά (blasts)).

Κατά τη βιοψία μυελού των οστών, λαμβάνεται τμήμα του μυελού των οστών. Το δείγμα μυελού των οστών αναλύεται χρησιμοποιώντας μια ευρεία ποικιλία εξετάσεων, από μικροσκοπική ανάλυση μέχρι ειδικές μελέτες καλλιέργειας μυελού των οστών. Γίνεται έπειτα η διάγνωση.

Αντίθετα από την κοινή αντίληψη, τα λευχαιμικά κύτταρα, δεν αναπτύσσονται πάντα γρηγορότερα από τα φυσιολογικά κύτταρα. Οι υψηλοί αριθμοί στο αίμα προκαλούνται από τη συσσώρευση λευχαιμικών κυττάρων σε αιμοφόρα αγγεία. Τα ανώριμα λευχαιμικά κύτταρα συσσωρεύονται στα αιμοφόρα αγγεία επειδή δεν χρησιμοποιούνται από το σώμα. Επομένως, οι αριθμοί τους όλο και αυξάνονται. Πολύ υψηλοί αριθμοί κυττάρων στο αίμα μπορούν να οδηγήσουν στην απόφραξη των αιμοφόρων αγγείων και σε φθορά των ιστών του σώματος. Το αίμα γίνεται πολύ πυκνό και κολλώδες, καθιστώντας δύσκολη τη φυσιολογική κυκλοφορία. Αυτό το πρόβλημα είναι ένα από τα αποτελέσματα της λευχαιμίας που επιδιώκει να υπερνικήσει η θεραπεία (National Cancer Institute 2008).

2.4.3. Ταξινόμηση

Η παθολογική εξέταση, η κυτταρογενετική (ιδιαίτερα η παρουσία του χρωμοσώματος της Φιλαδέλφειας) και η ανοσοφαινοτυποποίηση καθορίζουν εάν τα λεμφοβλαστικά λευχαιμικά κύτταρα ξεκίνησαν από Β-λεμφοκύτταρα ή Τ-λεμφοκύτταρα. Το τεστ DNA μπορεί να πιστοποιήσει πόσο επιθετική είναι η ασθένεια, διαφορετικές μεταλλάξεις έχουν συνδεθεί με την σύντομη ή μακροχρόνια επιβίωση (Haferlach et al. 2005, Pediatric Oncology Resource Center 2005, Παρασκευόπουλος 1990).

Η μορφολογική μελέτη των βλαστικών κυττάρων και η ταξινόμηση τους με μια από τις σήμερα χρησιμοποιούμενες ταξινομήσεις έχει ουσιαστική αξία, αφού σχετίζεται με την πρόγνωση και την εξέλιξη της νόσου.

2.4.3.1. Ταξινόμηση κατά FAB (French American British Cooperative Group)

Η κατάταξη που χρησιμοποιείται από το **FAB** (French American British Cooperative Group) (L₁, L₂, L₃) έχει προγνωστική μόνο αξία, αφού ξεχωρίζει τα βλαστικά κύτταρα σε μικρά ή μεγάλα (με αντίστοιχα καλή ή κακή πρόγνωση), χωρίς να καλύπτει το σύνολο των ποικιλιών της νόσου.

- L₁, η πιο κοινή υποκατηγορία στα παιδιά. Οι λεμφοβλάστες, είναι μικρά κύτταρα.
- L₂, αποτελεί 10% των ALL περιπτώσεων. Αυτά τα κύτταρα είναι μεγαλύτερα.
- L₃, είναι η σπανιότερη υποκατηγορία.

2.4.3.2. Ταξινόμηση βασισμένη σε ανοσοφαινοτυποποίηση

Οι κυτταρογενετικές μελέτες, η κυτταρομετρία ροής, και οι μοριακές γενετικές μελέτες παρέχουν περισσότερες αναλυτικές πληροφορίες για τις υποκατηγορίες της ALL. Με τη βοήθεια αυτών των τεστ η ALL διαίρεται σε ομάδες, σύμφωνα με την ανοσοφαινοτυποποίηση (Immunophenotype) της λευχαιμίας, η οποία λαμβάνει υπόψη τον τύπο λεμφοκυττάρου (κύτταρο B ή κύτταρο T) και το επίπεδο ωρίμανσης των λευχαιμικών κυττάρων. Αυτές οι ομάδες έχουν αντικαταστήσει κατά ένα μεγάλο μέρος την ταξινόμηση κατά FAB.

Κάθε υποκατηγορία ταξινομείται με τον καθορισμό των δεικτών επιφάνειας (surface markers) των ανώμαλων λεμφοκυττάρων. Υπάρχουν 2 κύριοι ανοσοποιητικοί τύποι: τα **pre-B** και **pre-T** λευχαιμικά λεμφοκύτταρα. Τα ώριμα B λευχαιμικά λεμφοκύτταρα (L₃) ταξινομούνται πλέον ως **λέμφωμα Burkitt** (American Cancer Society 2007).

ALL των B-λεμφοκυττάρων

- πρόιμη pre-B ALL (συντά ονομάζεται pro-B ALL) – περίπου το 10% των περιπτώσεων
- κοινή ALL - περίπου το 50% των περιπτώσεων
- pre-B ALL - περίπου το 10% των περιπτώσεων
- ώριμα B-λεμφοκύτταρα ALL (Burkitt λέμφωμα) - περίπου το 4% των περιπτώσεων

ALL των T-λεμφοκυττάρων

- pre-T ALL - περίπου το 5% έως 10% των περιπτώσεων
- ALL των ώριμων T-λεμφοκυττάρων– περίπου 5% έως 20% των περιπτώσεων

2.4.3.3. Ταξινόμηση κατά WHO

Η κατάταξη του **WHO** είναι περισσότερο κατατοπιστική με τις περισσότερες υποκατηγορίες βλαστικών κυττάρων. Σύμφωνα με την κατάταξη αυτή:

1. Οξεία λεμφοβλαστική λευχαιμία: Προηγούμενοι τύποι κατά FAB L1/L2.
 - a. Οξεία λεμφοβλαστική λευχαιμία των **προδρόμων Β** κυττάρων.
Κυτταρογενετικές υποκατηγορίες:
 - i. t(12;21)(p12,q22) TEL/AML-1
 - ii. t(1;19)(q23;p13) PBX/E2A
 - iii. t(9;22)(q34;q11) ABL/BCR
 - iv. T(V,11)(V;q23) V/MLL
 - b. Οξεία λεμφοβλαστική λευχαιμία των **προδρόμων Τ** κυττάρων.
2. Οξεία λεμφοβλαστική λευχαιμία: Προηγούμενοι τύποι κατά FAB L3.
3. Οξεία λευχαιμία Biphenoctic

Η κυτταρογενετική ανάλυση αποτελεί ανεξάρτητο προγνωστικό δείκτη, συμβάλλοντας στη διάγνωση και τυποποίηση της νόσου, στην επιλογή της θεραπείας και στον έλεγχο του θεραπευτικού αποτελέσματος. Ο προσδιορισμός των ειδικών, κλωνικών χρωμοσωματικών αλλοιώσεων συμβάλει στη διερεύνηση της παθογένεσης, στην ταυτοποίηση των γονιδίων που εμπλέκονται στη λευχαιμογένεση και στην ακριβή ταυτοποίηση ομάδων ασθενών υψηλού κινδύνου (American Cancer Society 2007).

Κυτταρογενετική μετατόπιση	Μοριακή γενετική ανωμαλία
t(12;21)CRYPTIC	TEL-AML1 fusion
t(1;19)(q23;p13)	E2A-PBX (PBX1) fusion
t(9;22)(q34;q11)	BCR-ABL fusion(P185)
t(4;11)(q21;q23)	MLL-AF4 fusion
t(8;14)(q24;q32)	IGH-MYC fusion
t(11;14)(p13;q11)	TCR-RBTN2 fusion

Πίνακας 2.1: Κυτταρογενετικές μετατοπίσεις συνδέονται με συγκεκριμένες μοριακές γενετικές ανωμαλίες στην ALL (Stams et al. 2005, McWhirter et al. 1999, Den Boer et al., Martin-Subero et al. 2005, Caslini et al. 2004).

Η κλασική κυτταρογενετική ανάλυση (καρυότυπος: απεικόνιση των χρωμοσωμάτων σε ζεύγη και κατά ελαττούμενο μέγεθος) αποτελεί την κύρια μέθοδο μελέτης της ύπαρξης χρωμοσωματικών ανωμαλιών (Moortman et al. 2007).

Μερικές κυτταρογενετικές υποκατηγορίες έχουν χειρότερη πρόγνωση από άλλες. Αυτές περιλαμβάνουν:

- Μια μετατόπιση μεταξύ των χρωμοσωμάτων 9 και 22, γνωστή ως **χρωμόσωμα της Φιλαδέλφειας**, που εμφανίζεται σε περίπου 20% των ενηλίκων και 5% των παιδιατρικών περιπτώσεων της ALL.
- Μια μετατόπιση μεταξύ των χρωμοσωμάτων 4 και 11 που εμφανίζεται σε περίπου 4% των περιπτώσεων και είναι η πιο κοινή σε παιδιά μικρότερα των 12 μηνών.
- Δεν έχουν όλες οι περιπτώσεις μετατοπίσεων των χρωμοσωμάτων ανεπαρκή πρόγνωση. Μερικές μετατοπίσεις είναι σχετικά πιο ευνοϊκές. Για παράδειγμα, ο υπερδιπλοειδισμός (>50 χρωμοσώματα) είναι ένας καλός προγνωστικός παράγοντας.

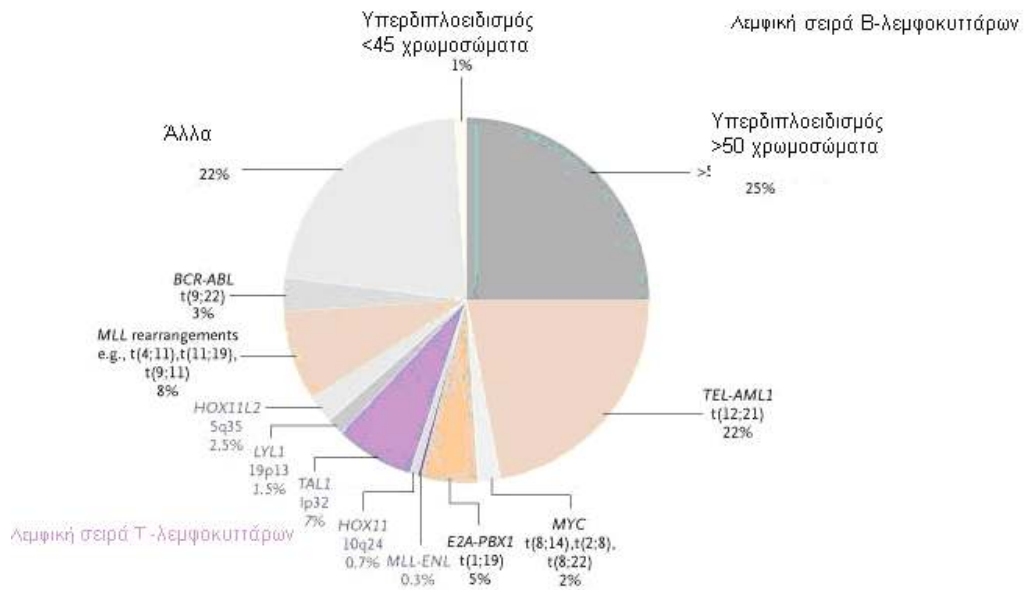
Κυτταρογενετική αλλαγή	Κατηγορία επικινδυνότητας
Χρωμόσωμα Φιλαδέλφειας	Κακή πρόγνωση
t(4;11)(q21;q23)	Κακή πρόγνωση
t(8;14)(q24.1;q32)	Κακή πρόγνωση
Σύνθετος καρυότυπος (περισσότερες από 4 ανωμαλίες)	Κακή πρόγνωση
Χαμηλός Υπερδιπλοειδισμός ή σχεδόν τριπλοειδισμός	Κακή πρόγνωση
Υψηλός Υπερδιπλοειδισμός	Καλή πρόγνωση
απαλοιφή (9p)	Καλή πρόγνωση

Πίνακας 2.2: Σχέση της πρόγνωσης με την κυτταρογενετική αλλαγή που εμφανίζεται (Acute lymphoblastic leukemia - Wikipedia, the free encyclopedia 2009).

Πρόγνωση	Κυτταρογενετικά ευρήματα
Ευνοϊκή	Υπερδιπλοειδισμός > 50, t (12;21)
Ενδιάμεση	Υπερδιπλοειδισμός 47 -50; φυσιολογικά (διπλοειδής); del (6q); Αναδιάταξη του 8q24
Μη ευνοϊκή	Υποδιπλοειδισμός – σχεδόν απλοειδής, σχεδόν τετραπλοειδισμός, απαλοιφή (17p); t (9;22); t (11q23)

Πίνακας 2.3: Συσχετισμός της πρόγνωσης με τα κυτταρογενετικά ευρήματα στο μυελών των οστών στην οξεία λεμφοβλαστική λευχαιμία (Den Boer et al.).

Παιδιά



Εικόνα 2.6: Εκτιμώμενη συχνότητα συγκεκριμένων γονοτύπων ALL σε παιδιά. Οι γενετικές αλλοιώσεις που εμφανίζονται αποκλειστικά σε περιπτώσεις οξείας λευχαιμίας Τ-Λεμφοκυττάρων, φαίνονται με μοβ. Όλες οι άλλες γενετικές υποκατηγορίες, είτε είναι αποκλειστικές είτε εμφανίζονται πρώτιστα σε περιπτώσεις οξείας λευχαιμίας Β-Λεμφοκυττάρων (Pui, Relling & Downing 2004).

Ορισμένοι παράγοντες που επηρεάζουν την πρόγνωση (πιθανότητα αποκατάστασης) και τη θεραπεία είναι οι:

- Η ηλικία και ο αριθμός των λευκών κυττάρων κατά τη διάγνωση.
- Πόσο γρήγορα και πόσο χαμηλά μειώνεται ο αριθμός των λευχαιμικών κυττάρων, μετά από την αρχική θεραπεία.
- Το φύλο και η φυλή.
- Εάν τα λευχαιμικά κύτταρα ξεκίνησαν από τα Β-λεμφοκύτταρα ή τα Τ-λεμφοκύτταρα.
- Εάν υπάρχουν ορισμένες αλλαγές στα χρωμοσώματα των λεμφοκυττάρων.
- Εάν η λευχαιμία έχει κάνει μετάσταση στον εγκέφαλο και στην σπονδυλική στήλη.
- Εάν το παιδί έχει σύνδρομο Down.

2.4.4. Θεραπευτικές αγωγές

Η οξεία λευχαιμία σχεδόν πάντα χρειάζεται θεραπευτική αγωγή με συνδυασμούς αντικαρκινικών φαρμάκων που αποκαλείται **συνδυασμός χημειοθεραπείας**.

2.4.4.1. Χημειοθεραπεία

Η χημειοθεραπεία περιλαμβάνει τη συνεχή χρήση φαρμάκων που καταστρέφουν ή θέτουν υπό έλεγχο την ανάπτυξη κακοήθων κυττάρων. Μπορεί να χορηγείται ως ένα μόνο φάρμακο ή ως συνδυασμός φαρμάκων. Τα φάρμακα πρέπει να κυκλοφορήσουν στο αίμα για να φθάσουν στα μη φυσιολογικά κύτταρα. Τα φάρμακα χημειοθεραπείας παρεμποδίζουν ή καταστρέφουν την αναπαραγωγή και την ανάπτυξη μη φυσιολογικών κυττάρων.

Η θεραπευτική αγωγή έχει αρχικά στόχο να εξαλείψει τα περισσότερα μη φυσιολογικά κύτταρα και να επιτρέψει σε φυσιολογικά κύτταρα να επανέλθουν στο μυελό των οστών και αποκαλείται **θεραπεία εφόδου** (induction therapy). Όταν η χημειοθεραπεία επιφέρει τη φυσιολογική εμφάνιση του αίματος και του μυελού των οστών και ο ασθενής αισθάνεται καλά λέγεται ότι υπάρχει **πλήρης ύφεση**. Πλήρης ύφεση δεν σημαίνει ότι έχει εξαλειφθεί όλη η λευχαιμία και πολλές έρευνες είναι αφιερωμένες στην ανάπτυξη καλύτερων τεχνικών εντοπισμού μικρότερων αριθμών υπολειπόμενων λευχαιμικών κυττάρων στο μυελό των οστών.

Θεραπεία σταθεροποίηση (Consolidation therapy) χορηγείται όταν η ασθένεια δεν είναι πλέον εμφανής ως αποτέλεσμα της εισαγωγικής θεραπείας, αλλά είναι γνωστό ότι είναι παρούσα. Επαναλαμβανόμενοι κύκλοι χημειοθεραπείας χορηγούνται με μειωμένες δόσεις για την περαιτέρω μείωση του αριθμού μη φυσιολογικών κυττάρων. Σε πολλές περιπτώσεις, αν και τα μη φυσιολογικά κύτταρα δεν μπορούν να εντοπιστούν, υπάρχουν ακόμη μικροί αριθμοί στο μυελό των οστών. Αν συμβαίνει αυτό, η λευχαιμία μπορεί να επανεμφανιστεί. Αυτό αποκαλείται **υποτροπή** ή επανεμφάνιση της νόσου. Η επαναληπτική θεραπευτική αγωγή μπορεί να εμφανίσει και άλλη πλήρη ύφεση ή όχι.

Στην προσπάθεια περαιτέρω μείωσης της υπολειπόμενης ασθένειας (minimal residual disease), μπορεί να χορηγούνται χαμηλές δόσεις φαρμάκων, πράγμα που αποκαλείται **συντηρητική χημειοθεραπεία** (maintenance chemotherapy). Σ' αυτό το στάδιο, η χημειοθεραπεία εισαγωγής και σταθεροποίησης έχει μειώσει τον αριθμό μη φυσιολογικών κυττάρων στο ελάχιστο και οι επαναλαμβανόμενοι κύκλοι συντηρητικής θεραπείας στοχεύουν να διατηρήσουν την ασθένεια υπό έλεγχο μέχρι να επιτευχθεί είτε η εξάλειψή της είτε η καταστροφή της από το ανοσοποιητικό σύστημα.

2.4.4.2. Μεταμόσχευση αρχέγονων κυττάρων ή μυελού των οστών

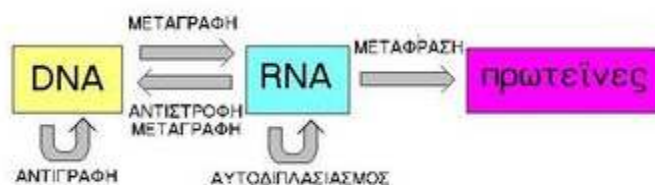
Ο όρος μεταμόσχευση μυελού των οστών χρησιμοποιείται κατά παράδοση για να περιγραφεί η διαδικασία αφαίρεσης κυττάρων από το μυελό των οστών για επανέγχυση στον ασθενή μετά τη χορήγηση υψηλών δόσεων χημειοθεραπείας ή/και ακτινοθεραπείας. Αυτή η διαδικασία αποκαλείται τώρα **μεταμόσχευση αρχέγονων κυττάρων** (stem cell transplantation).

Τα αρχέγονα κύτταρα είναι σημαντικά από όπου κι αν συλλέγονται, είτε στο αίμα όπου κινούνται, είτε στο μυελό των οστών όπου αναπαύονται και διαιρούνται. Τα αρχέγονα κύτταρα μπορούν να συλλέγονται με ποικίλους τρόπους από διαφορετικούς **δότες** (αυτόλογος, αδελφός, μη συγγενής) (Keene 2002, Leukemia & Lymphoma Society 2009, National Cancer Institute 2008).

3. DNA, RNA, ΠΡΩΤΕΪΝΕΣ ΚΑΙ ΓΟΝΙΔΙΑΚΗ ΕΚΦΡΑΣΗ

3.1. ΚΕΝΤΡΙΚΟ ΔΟΓΜΑ ΤΗΣ ΒΙΟΛΟΓΙΑΣ

Το **κύτταρο** είναι η ελάχιστη μονάδα της ζωής. Υπάρχει ένα πλήθος συγκεκριμένων χημικών μετασχηματισμών που όχι μόνο παρέχουν την ενέργεια που απαιτείται από ένα κύτταρο, αλλά και συντονίζουν όλες τις δραστηριότητες του κύτταρο. Η διαδικασία της ζωής περιλαμβάνει μια ποικιλία μορίων, από το νερό ως τις μικρές οργανικές ενώσεις (π.χ., λιπαρά οξέα και σάκχαρα), και τα μακρομόρια (DNA, πρωτεΐνες, και πολυσακχαρίτες) που καθορίζουν τη δομή των κυττάρων. Τα μακρομόρια ελέγχουν και κυβερνούν τις περισσότερες από τις δραστηριότητες της ζωής.

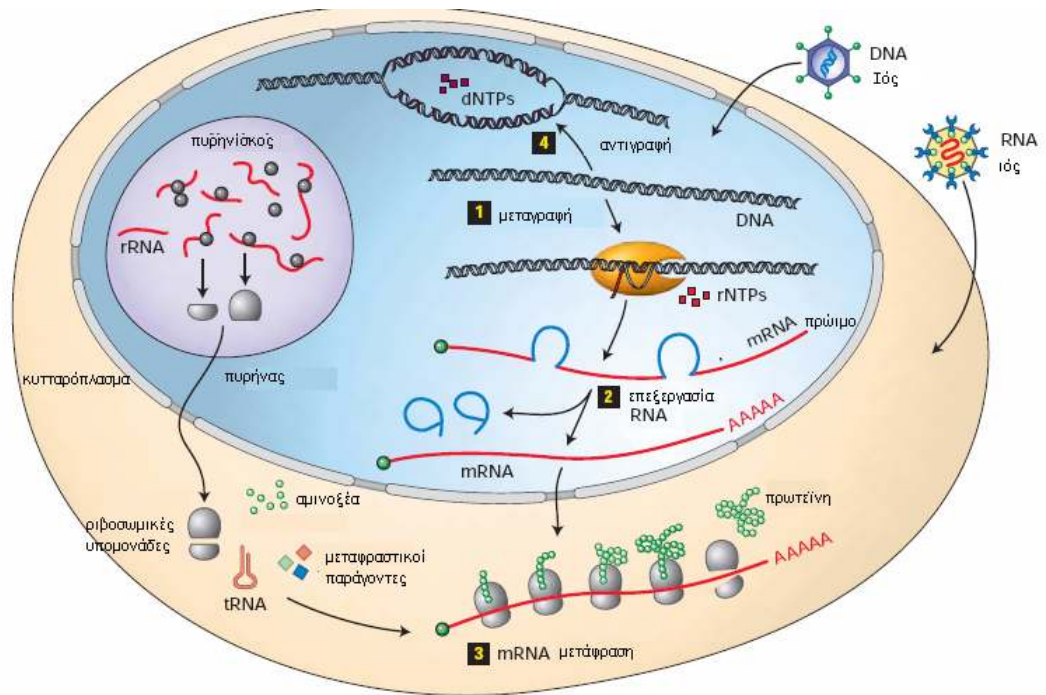


Εικόνα 3.1: Κεντρικό δόγμα της βιολογίας (Κεντρικό δόγμα Μοριακής Βιολογίας - Βικιπαίδεια 2008).

Το **DNA** ενός οργανισμού είναι το μοριακό πρόγραμμα που περιέχει ακριβείς οδηγίες, οι οποίες καθορίζουν τη δομή και τη λειτουργία του κυττάρου. Ταυτόχρονα περιέχει την πληροφορία για τον αυτοδιπλασιασμό του, εξασφαλίζοντας έτσι τη μεταβίβαση των γενετικών πληροφοριών από το μητρικό στα θυγατρικά κύτταρα και από τον οργανισμό στους απογόνους του μέσω της διαδικασίας της **αντιγραφής (replication)**. Οι εντολές που λαμβάνει το κύτταρο για να εκτελέσει ορισμένες διαδικασίες δίδονται μέσω των παραγομένων πρωτεϊνών.

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

Η γενετική πληροφορία είναι η καθορισμένη σειρά των αζωτούχων βάσεων των νουκλεοτιδίων. Η πληροφορία υπάρχει σε τμήματα του DNA με συγκεκριμένη ακολουθία, τα **γονίδια**. Αυτά, διαμέσου της μεταγραφής και της μετάφρασης, καθορίζουν τη σειρά των αμινοξέων στις πολυπεπτιδικές αλυσίδες. Οι πορείες της μεταγραφής και της μετάφρασης των γονιδίων αποτελούν τη γονιδιακή έκφραση (Lee 2004).



Εικόνα 3.2: Επισκόπηση των τεσσάρων βασικών μοριακών γενετικών διαδικασιών (Lodish, Darnell & etc. 2003).

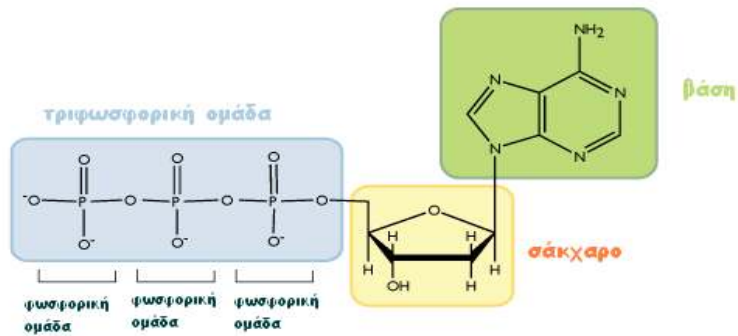
3.2. ΓΟΝΙΔΙΑ

Γονίδια ονομάζονται οι αλληλουχίες νουκλεοτιδίων τμήματος του DNA, οι οποίες ελέγχουν τα κληρονομικά γνωρίσματα ενός οργανισμού. Ένα γονίδιο αποτελεί πρότυπο για τη σύνθεση ενός λειτουργικού μορίου RNA. Το γενετικό υλικό ενός κυττάρου αποτελεί το γονιδιώμα του. Το γονιδιώμα καθορίζει τη γενετική κατασκευή ενός οργανισμού ή ενός κυττάρου, το **γονότυπο** του. Ο **φαινότυπος**, είναι το σύνολο των χαρακτηριστικών που εμφανίζει ένας οργανισμός υπό την επιρροή ενός συνόλου περιβαλλοντικών παραγόντων (Lee 2004).

3.3. DNA

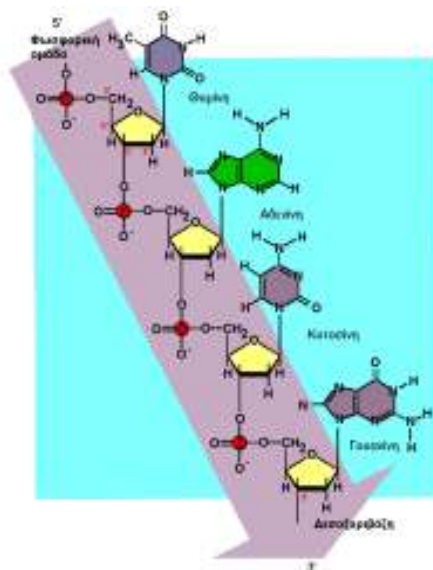
3.3.1. Γενικά

Το DNA, όπως και το RNA, είναι ένα μακρομόριο, που αποτελείται από νουκλεοτίδια. Κάθε νουκλεοτίδιο του DNA αποτελείται από μία πεντόζη, τη δεοξυριβόζη, ενωμένη με μια φωσφορική ομάδα και μια αζωτούχο βάση. Στα νουκλεοτίδια του DNA η αζωτούχος βάση μπορεί να είναι μια από τις: αδενίνη (A), γουανίνη (G), κυτοσίνη (C) και θυμίνη (T). Σε κάθε νουκλεοτίδιο η αζωτούχος βάση συνδέεται με τον 1' άνθρακα της δεοξυριβόζης και η φωσφορική ομάδα με τον 5' άνθρακα.



Εικόνα 3.3: Νουκλεοτίδιο του DNA (Life Sciences Cyberbridge 2007).

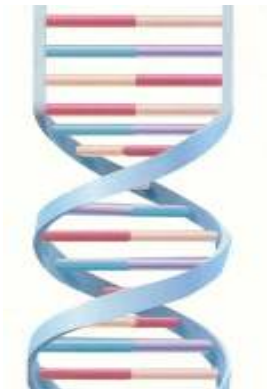
Μια πολυνουκλεοτιδική αλυσίδα σχηματίζεται από την ένωση πολλών νουκλεοτιδίων με ομοιοπολικό δεσμό. Ο δεσμός αυτός δημιουργείται μεταξύ του υδροξυλίου του 3' άνθρακα της πεντόζης του πρώτου νουκλεοτιδίου και της φωσφορικής ομάδας που είναι συνδεδεμένη στον 5' άνθρακα της πεντόζης του επόμενου νουκλεοτιδίου. Ο δεσμός αυτός ονομάζεται **3' - 5' φωσφοδιεστερικός δεσμός**. Με τον τρόπο αυτό η πολυνουκλεοτιδική αλυσίδα που δημιουργείται έχει ένα σκελετό, που αποτελείται από επανάληψη των μορίων φωσφορική ομάδα-πεντόζη-φωσφορική ομάδα-πεντόζη. Ανεξάρτητα από τον αριθμό των νουκλεοτιδίων από τα οποία αποτελείται η πολυνουκλεοτιδική αλυσίδα, το πρώτο της νουκλεοτίδιο έχει πάντα μια ελεύθερη φωσφορική ομάδα συνδεδεμένη στον 5' άνθρακα της πεντόζης του και το τελευταίο νουκλεοτίδιο της έχει ελεύθερο το υδροξύλιο του 3' άνθρακα της πεντόζης του. Για το αυτό αναφέρεται ότι ο προσανατολισμός της πολυνουκλεοτιδικής αλυσίδας είναι 5'→3'.



Εικόνα 3.4: Το βέλος δείχνει τον φωσφοδιεστερικό σκελετό με προσανατολισμό 5'→3' (Μαρμάρα Β., Λαμπροπούλου-Μαρμάρα Μ. 2005).

3.3.2. Το μοντέλο της διπλής έλικας

Το 1953 οι Watson και Crick διατύπωσαν το μοντέλο της διπλής έλικας του DNA, που αναφέρεται στη δομή του DNA στο χώρο. Σύμφωνα με το μοντέλο αυτό το DNA αποτελείται από δύο πολυνουκλεοτιδικές αλυσίδες που σχηματίζουν στο χώρο μία δεξιόστροφη διπλή έλικα. Η διπλή έλικα έχει ένα σταθερό σκελετό, που αποτελείται από επαναλαμβανόμενα μόρια φωσφορικής ομάδας-δεοξυριβόζης ενωμένων με φωσφοδιεστερικό δεσμό. Ο σκελετός αυτός είναι υδρόφιλος και βρίσκεται στο εξωτερικό του μορίου. Προς το εσωτερικό του σταθερού αυτού σκελετού βρίσκονται οι αζωτούχες βάσεις που είναι υδρόφοβες. Οι αζωτούχες βάσεις της μιας αλυσίδας συνδέονται με δεσμούς υδρογόνου με τις αζωτούχες βάσεις της απέναντι αλυσίδας με βάση τον κανόνα της **συμπληρωματικότητας**. Η αδενίνη συνδέεται μόνο με θυμίνη με δύο δεσμούς υδρογόνου και αντίστροφα, ενώ η κυτοσίνη μόνο με γουανίνη με τρεις δεσμούς και αντίστροφα. Οι δεσμοί υδρογόνου που αναπτύσσονται μεταξύ των βάσεων σταθεροποιούν τη δευτεροταγή δομή του μορίου. Οι δύο αλυσίδες ενός μορίου DNA είναι συμπληρωματικές, και αυτό υποδηλώνει ότι η αλληλουχία της μιας καθορίζει την αλληλουχία της άλλης. Η συμπληρωματικότητα έχει τεράστια σημασία για τον αυτοδιπλασιασμό του DNA, μια ιδιότητα που το καθιστά το καταλληλότερο μόριο για τη διατήρηση και τη μεταβίβαση της γενετικής πληροφορίας (Lee 2004, Alberts 2008, Μαρμάρα Β., Λαμπροπούλου-Μαρμάρα Μ. 2005).

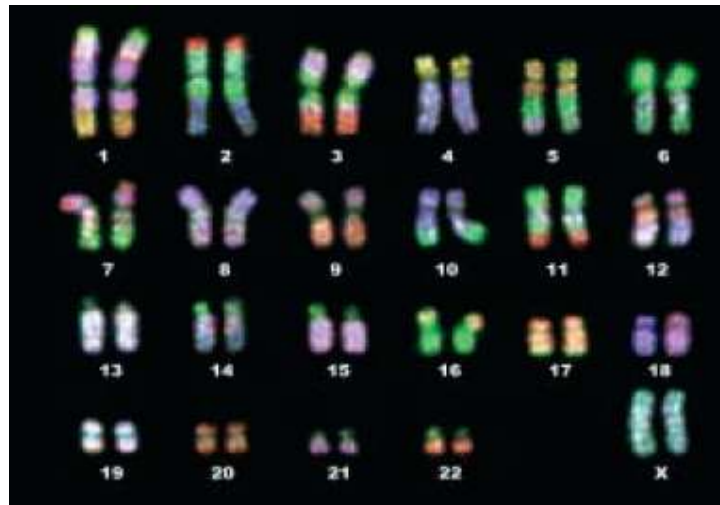


Εικόνα 3.5: Μοντέλο διπλής έλικας του DNA (Lodish, Darnell & etc. 2003).

Το ανθρώπινο γονιδίωμα σε ένα απλοειδές κύτταρο (γαμέτη) αποτελείται από περίπου 3×10^9 ζεύγη βάσεων DNA, που είναι οργανωμένα σε 23 χρωμοσώματα. Τα κύτταρα στα οποία το γονιδίωμα υπάρχει σε ένα μόνο αντίγραφο, όπως είναι τα προκαρυωτικά κύτταρα και οι γαμέτες των διπλοειδών οργανισμών, ονομάζονται **απλοειδή**. Τα κύτταρα στα οποία το γονιδίωμα υπάρχει σε δύο αντίγραφα, όπως είναι τα σωματικά κύτταρα των ανώτερων ευκαρυωτικών οργανισμών, ονομάζονται **διπλοειδή**.

Στον άνθρωπο τα φυσιολογικά αρσενικά και θηλυκά άτομα έχουν στον πυρήνα των σωματικών τους κυττάρων 23 ζεύγη χρωμοσωμάτων. Το ένα χρωμόσωμα κάθε ζεύγους είναι πατρικής και το άλλο μητρικής προέλευσης και ελέγχουν τις ίδιες ιδιότητες. Από τα 23 ζεύγη τα 22 είναι μορφολογικά ίδια στα αρσενικά και στα

θηλυκά άτομα και ονομάζονται **αυτοσωμικά** χρωμοσώματα. Το 23ο ζεύγος στα θηλυκά άτομα αποτελείται από δύο X χρωμοσώματα, ενώ στα αρσενικά από ένα X και ένα Y χρωμόσωμα.



Εικόνα 3.6: Τα χρωματισμένα χρωμοσώματα είναι χρήσιμα στην αποκάλυψη χρωμοσωματικών ανωμαλιών και στη σύγκριση καρυοτύπων διαφορετικών ειδών (Lodish, Darnell & etc. 2003).

Τα ζευγάρια χρωμοσωμάτων στα διπλοειδή κύτταρα μπορούν να μην είναι ίδια και μπορούν να περιέχουν παραλλαγές του ίδιου γονιδίου. Οι παραλλαγές έχουν δημιουργηθεί από τυχαίες μεταλλάξεις με την πάροδο του χρόνου. Μια μετάλλαξη αποτελείται από μια αλλαγή στην ακολουθία των ζευγαριών βάσεων στο DNA. Μια μετάλλαξη σε μια ακολουθία σε περιοχή κωδικοποίησης μπορεί να αλλάξει την ακολουθία αμινοξέων στην πρωτεΐνη. Παραλλαγές γονιδίων όπως σε αυτά που καταλαμβάνουν την ίδια θέση σε ένα χρωμόσωμα καλούνται **αλληλόμορφα γονίδια**. Δύο χρωμοσώματα με τα ίδια τμήματα γονιδίων θεωρούνται **ομόλογα**.

Η μονάδα της αντιγραφής είναι το **χρωμόσωμα**. Όταν ένα κύτταρο διαιρείται, όλα τα χρωμοσώματα και επομένως τα όλα τα γονίδια του αντιγράφονται. Εκτός από τις πληροφορίες για τη σύνθεση των πρωτεϊνών, τα γονίδια περιλαμβάνουν επίσης ρυθμιστικά στοιχεία που καθορίζουν πότε και που το εν λόγω γονίδιο πρέπει να μεταγραφεί και να μεταφραστεί (Alberts 2008).

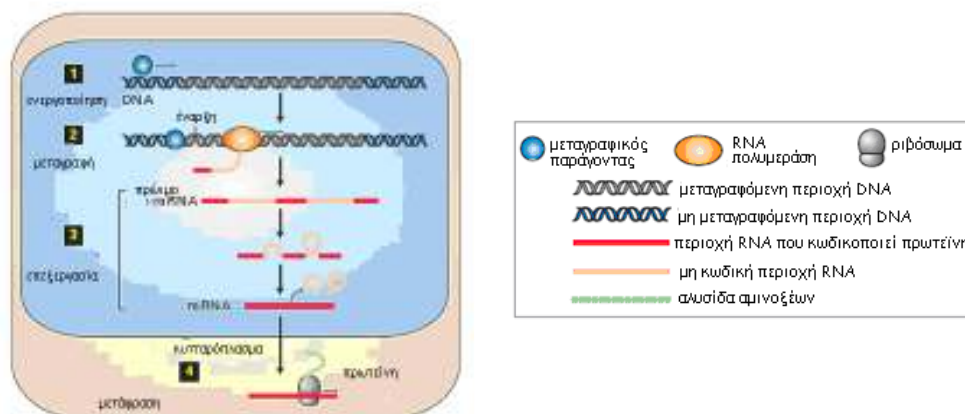


Εικόνα 3.7: Μικροσκοπική εμφάνιση ενός μεταφασικού χρωμοσώματος (Lodish, Darnell & etc. 2003).

3.4. RNA

Το **RNA** ή **Ριβονουκλεϊκό οξύ** είναι ένα πολυμερές νουκλεϊκό οξύ που αποτελείται από μονομερή νουκλεοτίδια που παίζουν σημαντικό ρόλο στη διαδικασία της μετάφρασης από το δεοξυριβονουκλεϊκό οξύ (DNA) σε πρωτεϊνικά προϊόντα. Το RNA λειτουργεί ως ο αγγελιοφόρος του DNA (Αγγελιοφόρο RNA ή mRNA). Το RNA αποτελεί δομικό συστατικό των ριβοσωμάτων (rRNA) και επίσης αποτελεί σημαντικό μεταφορέα αμινοξέων (tRNA - διαθέτει ειδικούς υποδοχείς για την πρόσδεση αμινοξέων και μια περιοχή αντικωδικονίου για την αναγνώριση του κωδικονίου που δεσμεύεται με την ειδική ακολουθία του αγγελιοφόρου RNA με ισχυρούς δεσμούς υδρογόνου) που προορίζονται για την διαδικασία της πρωτεϊνσύνθεσης. Το RNA έχει δομή παρόμοια με το DNA, αλλά διαφέρει σε ορισμένα δομικά στοιχεία: τα μόρια RNA περιέχουν ριβόζη αντί για δεοξυριβόζη σαν το κύριο σάκχαρο και επίσης περιέχει τη βάση ουρακίλη αντί για τη θυμίνη που απαντάται στο DNA. Το RNA **μεταγράφεται** από το DNA με τη βοήθεια κυρίως ενός ενζύμου που ονομάζεται RNA πολυμεράση και στη συνέχεια επεξεργάζεται με έναν αριθμό άλλων δευτερευόντων ενζύμων. Στη συνέχεια χρησιμοποιείται σαν βάση για τη μετάφραση των γονιδίων σε πρωτεΐνες, μεταφέροντας αμινοξέα στα ριβοσώματα για να δημιουργηθούν πρωτεΐνες (πάντα βάσει της αρχής της συμπληρωματικότητας).

Στους ευκαρυωτικούς οργανισμούς το mRNA που παράγεται με τη μεταγραφή υφίσταται μια διαδικασία **ωρίμανσης** προτού να είναι έτοιμο να προχωρήσει στα ριβοσώματα για τη μετάφραση. Αυτό γίνεται γιατί τα περισσότερα γονίδια είναι ασυνεχή και εκτός των αλληλουχιών που μεταφράζονται υπάρχουν αλληλουχίες οι οποίες δε μεταφράζονται σε αμινοξέα. Οι αλληλουχίες που μεταφράζονται είναι τα **εξώνια** και εκείνες που δε μεταφράζονται είναι τα **εσώνια**. Έτσι, το mRNA που μόλις έχει σχηματιστεί από τη μεταγραφή ενός γονιδίου ονομάζεται **πρόδρομο mRNA** και περιέχει εσώνια και εξώνια. Το πρόδρομο mRNA μετατρέπεται σε mRNA με τη διαδικασία της ωρίμανσης κατά την οποία τα εσώνια κόβονται από μικρά ριβοζονουκλεοπρωτεϊνικά σωματίδια (snRNPs), που λειτουργούν ως ένζυμα. Τα εξώνια που απομένουν συρράπτονται μεταξύ τους και με αυτό τον τρόπο σχηματίζεται το **ώριμο mRNA** που μεταφέρεται στα ριβοσώματα για την **πρωτεϊνσύνθεση**. Δύο περιοχές του ώριμου mRNA δε μεταφράζονται σε αμινοξέα, η μια βρίσκεται στο άκρο 5' και η άλλη στο άκρο 3' (Lodish, Darnell & etc. 2003).



Εικόνα 3.8: Οι κωδικοποιημένες πληροφορίες στο DNA μετατρέπονται σε μια ακολουθία αμινοξέων στις πρωτεΐνες με μια διαδικασία πολλαπλών βημάτων (Lodish, Darnell & etc. 2003).

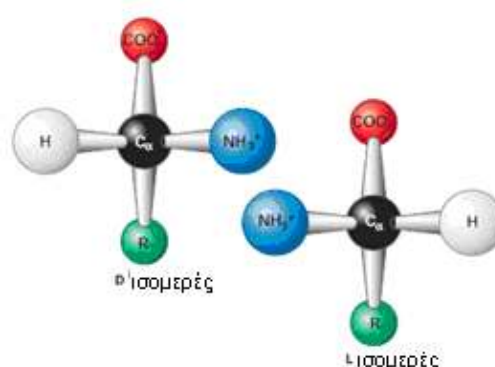
3.5. ΓΕΝΕΤΙΚΟΣ ΚΩΔΙΚΑΣ

Η αλληλουχία των βάσεων του mRNA καθορίζει την αλληλουχία των αμινοξέων στις πρωτεΐνες με βάση ένα κώδικα αντιστοίχισης νουκλεοτιδίων RNA με αμινοξέα πρωτεϊνών ο οποίος ονομάζεται **γενετικός κώδικας**.

Ο γενετικός κώδικας είναι **κώδικας τριπλέτας**, δηλαδή μια τριάδα νουκλεοτιδίων του mRNA, **κωδικόνιο**, κωδικοποιεί ένα αμινοξύ (συνολικά είκοσι αμινοξέα). Επίσης είναι **μη επικαλυπτόμενος**, δηλαδή κάθε νουκλεοτίδιο ανήκει σε ένα μόνο κωδικόνιο. Η αλληλουχία βάσεων ενός γονιδίου, και του mRNA του, που κωδικοποιεί μια πολυπεπτιδική αλυσίδα, αρχίζει με το **κωδικόνιο έναρξης** (το AUG και κωδικοποιεί το αμινοξύ μεθειονίνη) και τελειώνει με το **κωδικόνιο λήξης** (ένα από τα UAG, UGA και UAA). 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) (Lee 2004).

3.6. ΠΡΩΤΕΪΝΕΣ

Οι **πρωτεΐνες** είναι μεγάλα σύνθετα βιομόρια, αποτελούμενα από αμινοξέα, τα οποία ενώνονται μεταξύ τους με **πεπτιδικούς δεσμούς** σχηματίζοντας μια γραμμική αλυσίδα, καλούμενη **αλυσίδα πολυπεπτιδίων**. Ένας πεπτιδικός δεσμός δημιουργείται μέσω αντίδρασης κατά τη διάρκεια της οποίας ένα μόριο νερού αφαιρείται. Λόγω του τρόπου με τον οποίο ο πεπτιδικός δεσμός δημιουργείται, μια πολυπεπτιδική αλυσίδα έχει πάντα ένα άμινο (NH_2) και ένα καρβοξυλικό (COOH) άκρο.



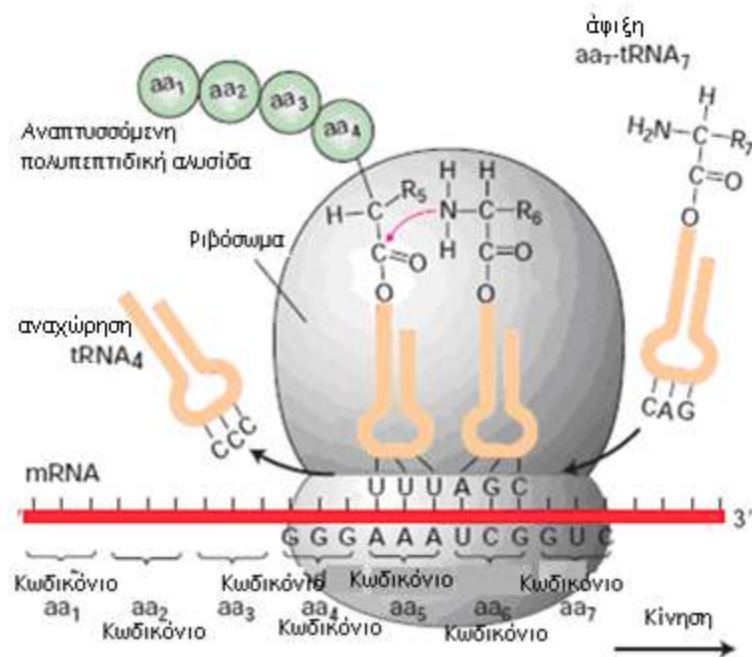
Εικόνα 3.9: Δομή αμινοξέων (Lodish, Darnell & etc. 2003).

Η ακολουθία αμινοξέων σε μια πρωτεΐνη καθορίζεται από ένα γονίδιο και κωδικοποιείται κατά τον γενετικό κώδικα DNA. Παρόλο που ο γενετικός κώδικας κωδικοποιεί 20 αμινοξέα, τα αμινοξέα που συνιστούν την πρωτεΐνη συχνά υφίστανται χημικές αλλαγές κατά τη μετα-μεταγραφική τροποποίηση: είτε προτού να μπορέσει η

πρωτεΐνη να λειτουργήσει στο κύτταρο, είτε ως τμήμα των μηχανισμών ελέγχου. Οι πρωτεΐνες αποτελούν είτε το δομικό συστατικό του κυττάρου είτε συνεργούν σε κάποια συγκεκριμένη λειτουργία. Ο δε βιολογικός τους ρόλος καθορίζεται κάθε φορά από την τρισδιάστατη δομή τους που είναι συνέπεια της αλληλουχίας των αμινοξέων, η οποία και ξεκινά από την πρωτοταγή δομή.

Η **μετάφραση** του mRNA, δηλαδή η αντιστοίχιση των κωδικονίων σε αμινοξέα και η σύνδεση των αμινοξέων σε πολυπεπτιδική αλυσίδα, πραγματοποιείται στα ριβοσώματα με τη βοήθεια των tRNA, αρκετών πρωτεϊνών και ενέργειας.

Η πρωτεϊνοσύνθεση διακρίνεται σε τρία στάδια: την **έναρξη**, την **επιμήκυνση** και τη **λήξη**. 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 tRNA διαθέτει μια ειδική θέση σύνδεσης με ένα συγκεκριμένο αμινοξύ.



Εικόνα 3.10: Οι τρεις ρόλοι του RNA στην πρωτεϊνοσύνθεση (Lodish, Darnell & etc. 2003).

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 της συμπληρωματικότητας των βάσεων.

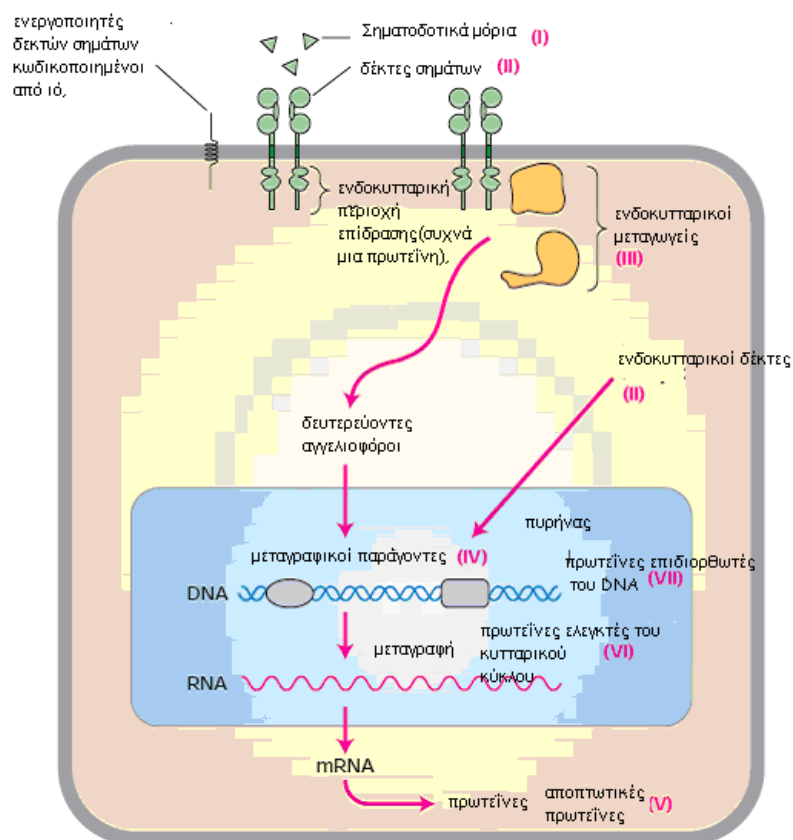
Κατά την **επιμήκυνση** ένα δεύτερο μόριο tRNA με αντικωδικόνιο συμπληρωματικό του δεύτερου κωδικονίου του mRNA τοποθετείται στην κατάλληλη εισδοχή του ριβοσώματος, μεταφέροντας το δεύτερο αμινοξύ. Μεταξύ της

μεθειονίνης και του δεύτερου αμινοξέος σχηματίζεται πεπτιδικός δεσμός. Η πολυπεπτιδική αλυσίδα συνεχίζει να επιμηκώνεται, καθώς νέα tRNA φέρουν αμινοξέα τα οποία προσδένονται μεταξύ τους.

Η επιμήκυνση σταματά σε ένα κωδικό λήξης (UGA, UAG ή UAA), επειδή δεν υπάρχουν tRNA που να αντιστοιχούν σε αυτά. Το κωδικόνιο λήξης αναγνωρίζεται από τον παράγοντα απελευθέρωσης ο οποίος προκαλεί τη λήξη και την απελευθέρωση της πολυπεπτιδικής αλυσίδας καθώς και τον αποχωρισμό των δύο υπομονάδων του ριβοσώματος (Lodish, Darnell & etc. 2003, Lee 2004, Alberts 2008).

3.7. ΓΟΝΙΔΙΑΚΗ ΕΚΦΡΑΣΗ

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. 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 συγκεκριμένη χρονική στιγμή.



Εικόνα 3.11: Επτά τύποι πρωτεϊνών που συμμετέχουν στον έλεγχο της αύξησης και του πολλαπλασιασμού κυττάρων (Lodish, Darnell & etc. 2003).

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 πρωτεϊνών, που ονομάζονται **μεταγραφικοί παράγοντες**. 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, ποικίλλει και η ικανότητα πρόσδεσης του mRNA στα ριβοσώματα.
- **Στο επίπεδο μετά τη μετάφραση:** Ακόμη και όταν γίνει η πρωτεϊνοσύνθεση και παραχθεί η κατάλληλη πρωτεΐνη, μπορεί να πρέπει να υποστεί τροποποιήσεις, για να γίνει βιολογικά λειτουργική.

Μερικά γονίδια που κωδικοποιούν πρωτεΐνες μεταγράφονται λιγότερο ή περισσότερο συχνά, και ονομάζονται «**housekeeping**» γονίδια και απαιτούνται πάντα για τις βασικές αντιδράσεις. Άλλα γονίδια δεν μεταγράφονται ή, μεταγράφονται για συγκεκριμένες λειτουργίες του οργανισμού, μόνο σε ιδιαίτερες στιγμές και κάτω από ιδιαίτερες εξωτερικές συνθήκες. Το σήμα που «καλύπτει» ή «αποκαλύπτει» ένα γονίδιο μπορεί να προέλθει από το εξωτερικού του κυττάρου, όπως μια θρεπτική ουσία ή μια ορμόνη. Πρόσθετες ρυθμιστικές ακολουθίες στο DNA υπαγορεύουν εάν ένα γονίδιο θα ανταποκριθεί στα σήματα και στη συνέχεια επηρεάζουν την μεταγραφή του γονιδίου που κωδικοποιεί την πρωτεΐνη (Lodish, Darnell & etc. 2003).

4. ΜΙΚΡΟΣΥΣΤΟΙΧΙΕΣ

4.1. ΕΙΣΑΓΩΓΗ

Οι **μικροσυστοιχίες γονιδίων** (αλλιώς γνωστές σαν γονιδιακό ή γενωμικό τσιπ, DNA chip είτε διάταξη γονιδίων) είναι μία διάταξη μικροσκοπικών κουκίδων που αντιπροσωπεύουν μοναδικά γονίδια ακινητοποιημένα με ομοιοπολικούς δεσμούς σε μία στερεή επιφάνεια (συνήθως γυάλινη). Χρησιμοποιούνται για τη μέτρηση DNA ή χρησιμοποιούν DNA για το σύστημα ανίχνευσής τους. Ποσοτικές και ποιοτικές μετρήσεις με μικροσυστοιχίες γονιδίων εκμεταλλεύονται την εκλεκτική φύση της αρχής της συμπληρωματικότητας μεταξύ νουκλεϊκών οξέων DNA-DNA ή DNA-RNA, υπό αυστηρά ελεγχόμενες συνθήκες θερμοκρασίας και με τη χρήση φθορίζουσών ουσιών.

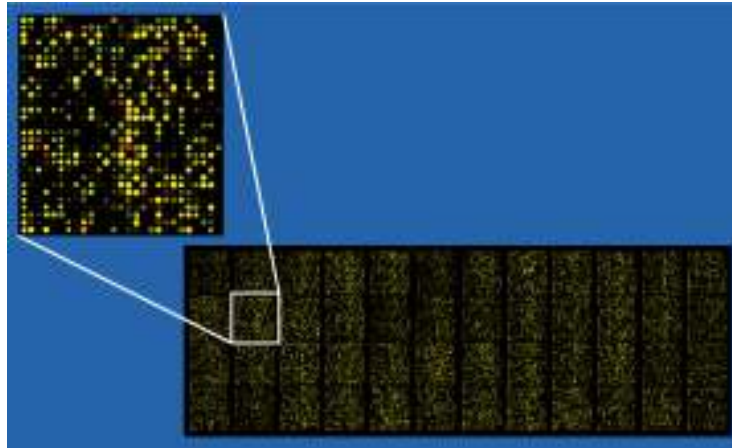
Η τεχνολογία των μικροσυστοιχιών επιτρέπει την μέτρηση των επιπέδων έκφρασης χιλιάδων διαφορετικών μορίων RNA σε μια συγκεκριμένη στιγμή της ζωής ενός οργανισμού, ενός ιστού, ή ενός κυττάρου. Η σύγκριση των επιπέδων των μορίων RNA μπορεί να χρησιμοποιηθεί για να αποκρυπτογραφηθούν χιλιάδες διαδικασίες που συμβαίνουν ταυτόχρονα στους ζωντανούς οργανισμούς.

Μπορεί να χρησιμοποιηθεί σε κλινικές συνθήκες προκειμένου:

1. Να προσδιοριστούν διαγνωστικοί ή προγνωστικοί βιοδείκτες
2. Να ταξινομηθούν ασθένειες (π.χ., καρκινικοί όγκοι με διαφορετική πρόγνωση που δεν μπορούν να διακριθούν με μικροσκοπική εξέταση)
3. Να ελεγχθεί η απόκριση στη θεραπεία και
4. Να γίνουν κατανοητοί οι μηχανισμοί που σχετίζονται με την έναρξη της ασθένειας.

Τα πειράματα με χρήση μικροσυστοιχιών περιέχουν πολλά βήματα, κάθε ένα από τα οποία εισάγει το δικό του θόρυβο και απόκλιση. Το τελικό αποτέλεσμα μπορεί να επηρεαστεί από οποιαδήποτε από τα βήματα της διαδικασίας και απαιτείται καλός πειραματικός σχεδιασμός και προσεκτική στατιστική ανάλυση για την επιτυχή ερμηνεία των δεδομένων.

Οι μικροσυστοιχίες DNA αποτελούνται από «κουκίδες» (spots) μεγέθους 5-150μm πάνω σε μια στέρεα επιφάνεια, κυρίως γυάλινη. Σε αυτές τις κουκίδες, επικολλώνται τμήματα μονόκλωνου DNA μήκους από 20 έως 1000 ή ακόμα και περισσότερες βάσεις. Αυτές οι μικροσυστοιχίες προσδιορίζουν την ακολουθία βάσεων DNA ενός γονιδίου σε ένα δείγμα χρησιμοποιώντας mRNA ή cDNA που έχει σημειωθεί με φθορίζουσες ουσίες. Στη συνέχεια τα ιχνηθετημένα mRNAs ή cDNAs υβριδοποιούνται με τα ακινητοποιημένα τμήματα DNA στο πλακίδιο. Τα πλακίδια στη συνέχεια σαρώνονται με υψηλής ταχύτητας ανιχνευτές φθορισμού (Prasad 2003).



Εικόνα 4.1: Παράδειγμα μικροσυστοιχίας με περίπου 40.000 ακολουθίες ανίχνευσης, με μεγεθυμένο τμήμα της για να φανούν οι λεπτομέρειες (Prasad 2003).

Υπάρχουν τρεις βασικές τεχνικές δημιουργίας μικροσυστοιχιών:

1. τυπωμένες cDNA μικροσυστοιχίες
2. τυπωμένες ολιγονουκλεοτιδικές μικροσυστοιχίες
3. in situ ολιγονουκλεοτιδικές μικροσυστοιχίες

Οι δύο κύριες πλατφόρμες για την κατασκευή υψηλής πυκνότητας μικροσυστοιχίες (>70,000 ξεχωριστοί κλώνοι) είναι οι τυπωμένες cDNA μικροσυστοιχίες και τα GeneChips. Οι πρώτες αφορούν την ρομποτική εκτύπωση μορίων DNA (robotic spotting) που συχνά απομονώνονται από αλληλουχίες EST (με τεχνικές αλυσιδωτής αντίδρασης πολυμεράσης, polymerase chain reaction, PCR, - υπάρχουν περίπου 1,000,000 EST's που αντιπροσωπεύουν το 60-90% όλων των γονιδίων) πάνω σε ένα κατάλληλο συμπαγές υπόστρωμα. Το πλεονέκτημα της πλατφόρμας αυτής είναι ότι μπορούν να τυπωθούν και άγνωστες αλληλουχίες.

Τα GeneChips αφορούν την in-situ σύνθεση ομάδων ειδικών για γονίδια ολιγονουκλεοτιδίων πάνω σε ένα παρόμοιο υπόστρωμα μέσω μίας φωτολιθογραφικής διαδικασίας, ή μίας άλλης διαδικασίας σύνθεσης, γνωστής ως in-situ light-directed combinatorial synthesis, ή μέσω συμβατικής σύνθεσης ακολουθούμενης από καθήλωση σε υπόστρωμα.

Να σημειωθεί ότι στην παρούσα εργασία, ως **ανιχνευτές DNA (probes)** αναφέρονται οι γνωστές ακολουθίες DNA που βρίσκονται ακινητοποιημένες στη μικροσυστοιχία σαν «μονάδες ανίχνευσης», ενώ ως **στόχοι (targets)**, το σύνολο των RNA που είναι προς εξέταση, το οποίο ιχνηθετείται με φθορίζουσες χρωστικές ουσίες και στη συνέχεια υβριδίζεται με τους ανιχνευτές (Lee 2004).

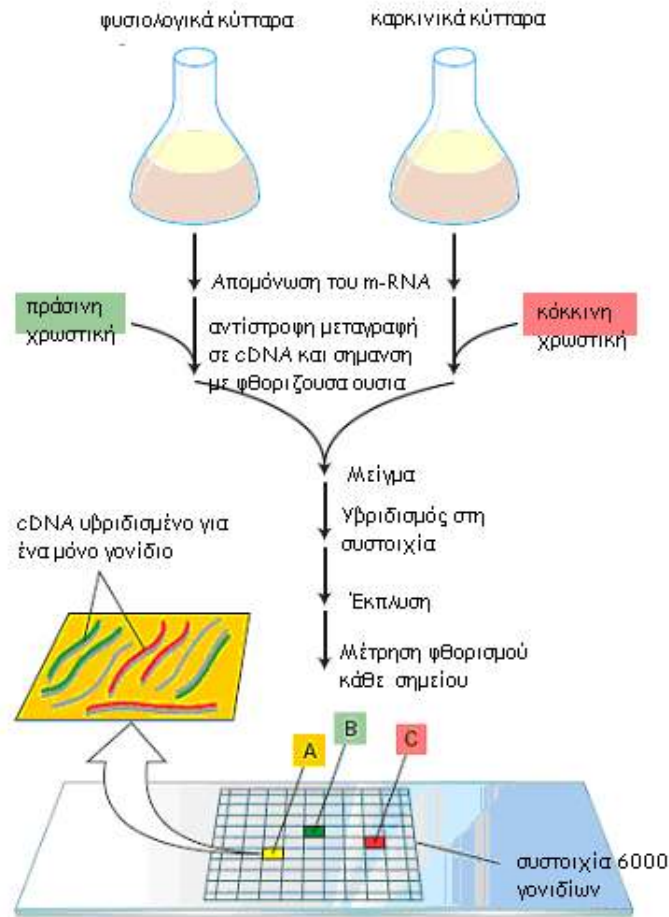
4.2. ΕΠΙΛΟΓΗ ΚΑΙ ΣΥΝΘΕΣΗ ΑΚΟΛΟΥΘΙΑΣ ΑΝΙΧΝΕΥΣΗΣ

Για μεγάλης κλίμακας μελέτες έκφρασης γονιδίων, το πρώτο βήμα είναι να επιλεγούν και να προετοιμαστούν οι συγκεκριμένοι ανιχνευτές υβριδοποίησης.

Η τεχνολογία μικροσυστοιχιών cDNA παρέχει μεγάλη ευελιξία στην επιλογή και την παραγωγή των σειρών, δεδομένου ότι οι αλληλουχίες ελέγχου μπορούν να δημιουργηθούν από τις βιβλιοθήκες cDNA. Μια **cDNA βιβλιοθήκη** είναι ένα σύνολο πλασμιδίων με ενσωματωμένα κομμάτια mRNA που έχουν μετατραπεί σε cDNA με τη διαδικασία της αντίστροφης μεταγραφής. Η ακολουθία ανίχνευσης που χρησιμοποιείται σε μια μικροσυστοιχία cDNA μπορεί να ενισχυθεί από μια τέτοια βιβλιοθήκη cDNA με τη μέθοδο PCR (αλυσιδωτή αντίδραση πολυμεράσης). Η επιλογή των DNA που θα χρησιμοποιηθούν στις κουκίδες της μικροσυστοιχίας καθορίζει το ποια γονίδια θα μπορέσουν να ανιχνευθούν σε μια ανάλυση συγκριτικού υβριδισμού (Prasad 2003).

Τα κύρια βήματα προετοιμασίας των δειγμάτων και εκτέλεσης ενός πειράματος cDNA μικροσυστοιχίας είναι:

1. Απομόνωση του mRNA και πολυμερισμός του με τη γνωστή τεχνική **αλυσιδωτής αντίδρασης πολυμεράσης (PCR)**, η οποία παράγει πολλά αντίγραφα των συγκεκριμένων τμημάτων mRNA.
2. Μετατροπή του mRNA σε cDNA με τη μέθοδο **αντίστροφης μεταγραφής**, μια διαδικασία που είναι η αντίστροφη της κανονικής διαδικασίας μεταγραφής, που παράγει mRNA από το DNA.
3. Σήμανση του προς εξέταση δείγματος cDNA με κόκκινη (Cy5) και του δείγματος ελέγχου (αναφορά) με πράσινη (Cy3) χρωστική ουσία φθορισμού.
4. Υβριδισμός με τις γνωστές αλληλουχίες βάσεων της μικροσυστοιχίας.
5. Έκπλυση των αδέσμευτων κλώνων.
6. Σάρωση της μικροσυστοιχίας.
7. Ανάλυση των δεδομένων.



Εικόνα 4.2: Παράδειγμα εκτέλεσης ενός πειράματος μικροσυστοιχίας (Lodish, Darnell & etc. 2003).

4.3. ΑΠΟΜΟΝΩΣΗ mRNA και ΑΝΤΙΣΤΡΟΦΗ ΜΕΤΑΓΡΑΦΗ

Ένα μόριο αγγελιοφόρου RNA περιλαμβάνει μια ακολουθία νουκλεοτιδίων που αντιστοιχεί στην ακολουθία αμινοξέων στην πρωτεΐνη. Αυτό το τμήμα του νουκλεϊνικού οξέος καλείται **περιοχή κωδικοποίησης (coding region)**. Επειδή το mRNA, είναι ένα ακριβές αντίγραφο των περιοχών κωδικοποίησης του DNA, η ανάλυση του mRNA μπορεί να χρησιμοποιηθεί για να προσδιορίσει πολυμορφισμούς στις περιοχές κωδικοποίησης του DNA. Ένας **πολυμορφισμός** είναι μια περιοχή DNA για την οποία υπάρχουν παραλλαγές στις ακολουθίες νουκλεοτιδίων σε έναν πληθυσμό οργανισμών. Τέτοιες παραλλαγές μπορούν μερικές φορές να εξηγήσουν την ύπαρξη μιας ασθένειας ή την έλλειψη ενός ενζύμου μέσα σε έναν πληθυσμό.

Η προετοιμασία του mRNA που θα χρησιμοποιηθεί σε μια μικροσυστοιχία απαιτεί την απομόνωση του από το υπόλοιπο mRNA. Το mRNA αποτελεί μόνο το 3% του ολικού RNA ενός κυττάρου, οπότε η απομόνωσή του σε ικανοποιητικές ποσότητες για την εκτέλεση ενός πειράματος (1-2 μg) μπορεί να είναι δύσκολη. Επίσης τα μόρια mRNA είναι δύσκολο να χρησιμοποιηθούν λόγω της ευαισθησίας τους που τα καθιστά ευπαθή. Το περιβάλλον είναι γεμάτο από ένζυμα που αποικοδομούν μόρια RNA. Για να αποτραπεί η απώλεια των δειγμάτων, αυτά

υπόκεινται σε αντίστροφη μεταγραφή, ώστε να ληφθούν σε DNA μορφή που είναι πιο σταθερή. Τα προϊόντα αυτής της αντίδρασης ονομάζονται συμπληρωματικά DNA (cDNA), διότι οι αλληλουχίες τους είναι συμπληρωματικές προς τα αρχικά mRNA.

Ένα πρόβλημα με την σύνθεση του cDNA είναι ότι δεν μπορούν όλα τα mRNA να μεταγραφούν αντίστροφα με την ίδια αποτελεσματικότητα. Το γεγονός αυτό μπορεί να οδηγήσει σε αλλαγή των σχετικών ποσοτήτων των διαφορετικών mRNA που ανιχνεύονται από την μικροσυστοιχία. Ένα πρόσθετο πρόβλημα αποτελεί το ότι μερικά mRNA μεταγράφονται αντίστροφα μόνο για μερικό μήκος τους, οπότε είναι πιο δύσκολο να δεσμευτούν στα συμπληρωματικά τους DNA στην μικροσυστοιχία (Prasad 2003).

4.4. ΦΘΟΡΙΖΟΥΣΑ ΣΗΜΑΝΣΗ ΤΩΝ cDNA

Για να ανιχνευθούν τα cDNA που έχουν δεσμευθεί στη μικροσυστοιχία πρέπει να σημειωθούν με ένα μόριο που να προσδιορίζει την παρουσία τους. Τέτοιου είδους μόρια που χρησιμοποιούνται στον συγκριτικό υβριδισμό με μικροσυστοιχίες είναι **φθορίζουσες χρωστικές**. Μια χρωστική με διαφορετικό χρώμα συνήθως χρησιμοποιείται για κάθε δείγμα, ώστε να μπορούν να διακριθούν τα δύο δείγματα πάνω στην μικροσυστοιχία.

Οι φθορίζουσες χρωστικές εμφανίζουν χρώμα μόνο όταν διεγερθούν από συγκεκριμένης συχνότητας φως από λέιζερ. Ακόμη όμως και τότε τα χρώματα δεν είναι άμεσα παρατηρήσιμα: το μήκος κύματος του εκπεμπόμενου φωτός χρησιμοποιείται για να ρυθμίσει έναν ανιχνευτή που μετρά τον φθορισμό.

Ο αριθμός των φθορίζόντων μορίων που σημαίνουν κάθε cDNA εξαρτάται από το μήκος του και πιθανόν από την αλληλουχία του, τα οποία συχνά δεν είναι γνωστά. Αυτός είναι ένας λόγος που εξηγεί γιατί οι εντάσεις φθορισμού για διαφορετικά cDNA δεν μπορούν να συγκριθούν ποσοτικά. Ωστόσο, πανομοιότυπα cDNA από κάθε κλώνο μπορούν να συγκριθούν αρκεί ο ίδιος αριθμός μορίων σήμανσης να έχει προστεθεί στην DNA ακολουθία του κάθε κλώνου.

Για να εξισωθούν οι ολικές συγκεντρώσεις των δύο cDNA, τα διαλύματα αραιώνονται πριν να εφαρμοστούν στη μικροσυστοιχία ώστε να έχουν την ίδια συνολική ένταση φθορισμού. Η διαδικασία κάνει δύο πιθανώς αδικαιολόγητες υποθέσεις: πρώτον ότι το συνολικό ποσό του mRNA σε κάθε τύπο κυττάρου που εξετάζεται είναι το ίδιο και δεύτερον ότι κάθε χρωστική εκπέμπει το ίδιο ποσό φωτός σε σχέση με την συγκέντρωσή της. Η δεύτερη υπόθεση μπορεί να εξαλειφθεί με κατάλληλη κανονικοποίηση, αλλά η πρώτη είναι δύσκολο να ελεγχθεί.

Υπάρχει ποικιλία στις μεθόδους σήμανσης του στόχου και οι πιο συχνά χρησιμοποιούμενες είναι:

1. RNA μπορεί να σημειωθεί μέσω της αντίστροφης μεταγραφής (RT) με την ενσωμάτωση τροποποιημένων νουκλεοτιδίων (nt), είτε άμεσα ιχνηθετημένων με δείκτες φθορισμού είτε αργότερα χημικά επικολώμενων.

2. Μια πολύ διαφορετική στρατηγική περιλαμβάνει την ενίσχυση του αγγελιοφόρου RNA με *in vitro* μεταγραφή (IVT) από μια ακολουθία υποκινητών που ενσωματώνεται στο cDNA κατά την αντίστροφη μεταγραφή RT.
3. Νουκλεοτίδια τροποποιημένα με βιοτίνη ενσωματώνονται στην *in vitro* μεταγραφή και χρησιμοποιούνται στη συνέχεια σαν βάσεις για την σήμανση με την φθορίζουσα ουσία.

Συνήθως τα δύο δείγματα (ένα προς έλεγχο και ένα ως αναφορά) υβριδίζονται σε κάθε αντικειμενοφόρο πλάκα για να αντισταθμίσουν τυχόν διαφορές στη διαδικασία επισήμανσης. Για το προς εξέταση δείγμα και το δείγμα αναφοράς (για παράδειγμα, δείγμα ασθενούς και υγιούς ιστού) με τη διαδικασία της αντίστροφης μεταγραφής δημιουργούνται cDNA μόρια που ιχνηθετούνται με δύο διαφορετικές χρωστικές ουσίες φθορισμού σε χωριστούς δοκιμαστικούς σωλήνες και στη συνέχεια αναμειγνύονται για ταυτόχρονο υβριδισμό με τις γνωστές ακολουθίες DNA της μικροσυστοιχίας. Οι δημοφιλέστερες επιλογές για τις χρωστικές ουσίες φθορισμού είναι Cy3 (πράσινο) και Cy5 (κόκκινο). Το φάσμα εκπομπής όλων των χρωστικών ουσιών που χρησιμοποιούνται, βρίσκεται σε επαρκώς διακριτά μήκη κύματος για να είναι δυνατή η ανίχνευσή τους (Lee 2004).

4.5. ΚΑΤΑΣΚΕΥΗ ΤΗΣ cDNA ΜΙΚΡΟΣΥΣΤΟΙΧΙΑΣ

Στις τυπωμένες cDNA μικροσυστοιχίες, οι DNA ανιχνευτές τοποθετούνται σαν διάλυμα στην επιφάνεια της αντικειμενοφόρου πλάκας και έπειτα προσδένονται. Υπάρχουν διάφορες επιλογές των υλικών υποστήριξης, από πλαστικά πολυμερή μέχρι γυαλί. Η επιφάνεια των μικροσκοπικών γυάλινων πλακών επικαλύπτεται συνήθως με χημικές ουσίες για να μειωθεί ο φθορισμός του υποβάθρου από το μη ειδικό υβριδισμό του σημασμένου στόχου.

Υπάρχουν δύο βασικές μέθοδοι για εκτύπωση των cDNA επάνω στην αντικειμενοφόρο πλάκα.

1. Η **εκτύπωση με επαφή**, στηρίζεται σε ακίδες (τριχοειδή) που παίρνουν διάλυμα και το τοποθετούν σαν σταγόνες πάνω στην πλάκα.
2. Η **εκτύπωση χωρίς επαφή**, όπου χρησιμοποιούνται παραλλαγές της τεχνολογίας έγχυσης, στην οποία μικρές σταγόνες διαλύματος μπορούν να ψεκαστούν με υψηλή ακρίβεια επάνω σε μια επιφάνεια (Prasad 2003).

4.6. ΥΒΡΙΔΙΣΜΟΣ

Ο υβριδισμός του ιχνηθετημένου στόχου με τους ανιχνευτές DNA στη μικροσυστοιχία πραγματοποιείται με την προσθήκη των στόχων σε διαλυτή μορφή μέσω ρυθμιστικού διαλύματος στη αντικειμενοφόρο πλάκα, μέσα σε ένα περιορισμένο χώρο, που ακολουθείται από επώαση για ένα δεδομένο χρονικό διάστημα σε ορισμένη θερμοκρασία.

Οι συνθήκες υβριδοποίησης πρέπει να είναι τέτοιες ώστε να προωθηθούν οι **ειδικές υβριδοποιήσεις** (specific hybridizations) μεταξύ του στόχου και των μεμονωμένων ακολουθιών ανίχνευσης και να περιοριστούν οι μη ειδικές υβριδοποιήσεις (non specific hybridizations) με την ίδια τη βάση ή με άλλες ακολουθίες ανίχνευσης.

Η μικροσυστοιχία διαθέτει εκατοντάδες ή χιλιάδες κουκίδες, καθεμία από τις οποίες περιέχει DNA με διαφορετική αλληλουχία. Εάν ένας κλώνος περιέχει cDNA το οποίο είναι συμπληρωματικό ως προς το DNA μιας κουκίδας, το cDNA αυτό θα υβριδιστεί στην κουκίδα και θα είναι ανιχνεύσιμο λόγω του φθορισμού του. Με τον τρόπο αυτό κάθε κουκίδα στη μικροσυστοιχία δείχνει την παρουσία ή μη ενός συγκεκριμένου cDNA. Σε κάθε κουκίδα υπάρχει αρκετό DNA ώστε και οι δύο κλώνοι να μπορούν να υβριδιστούν σε αυτή.

Μετά την υβριδοποίηση και έπειτα από αρκετές ώρες, το διάλυμα προς υβριδοποίηση απορρίπτεται και οι αντικειμενοφόρες πλάκες υποβάλλονται σε έκπλυση για να αφαιρεθούν οι μη ειδικοί υβριδισμοί των στόχων. Μετά από τη διαδικασία έκπλυσης, η αντικειμενοφόρος πλάκα είναι ξηρή και έτοιμη να σαρωθεί. Οι χρησιμοποιούμενες χρωστικές ουσίες εξασθενούν με την πάροδο του χρόνου από την έκθεση στο φως. Έτσι τα υβριδοποιημένα πλακίδια και τα ιχνηθετημένα διαλύματα των στόχων πρέπει να αποθηκεύονται στο σκοτάδι (Prasad 2003, Quackenbush 2001).

4.7. ΣΑΡΩΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΕΙΚΟΝΑΣ

Όταν οι cDNA κλώνοι έχουν υβριδιστεί στη μικροσυστοιχία και κάθε αδέσμευτος κλώνος έχει εκπλυθεί, η μικροσυστοιχία σαρώνεται για να καθοριστεί το ποσοστό του κάθε κλώνου που έχει προσδεθεί σε κάθε κουκίδα. Οι κλώνοι είναι σημασμένοι με φθορίζοντα μόρια που εκπέμπουν ανιχνεύσιμο φως όταν διεγερθούν από λέιζερ. Οι ανιχνευτές των μικροσυστοιχιών χρησιμοποιούν συνεστιακά λέιζερ ή άλλες πηγές φωτός για να παραγάγουν φως στο μήκος κύματος που διεγείρει τις χρωστικές ουσίες φθορισμού. Το εκπεμπόμενο φως ανιχνεύεται από τον ανιχνευτή, ο οποίος καταγράφει και την έντασή του.

Κάθε ένα από τα φθορίζοντα μόρια που έχουν χρησιμοποιηθεί για να σημειωθούν τα cDNA έχει ένα χαρακτηριστικό μήκος κύματος διέγερσης: μόνο φως με το συγκεκριμένο μήκος κύματος μπορεί να προκαλέσει τον φθορισμό του μορίου. Το εκπεμπόμενο φως έχει επίσης χαρακτηριστικό μήκος κύματος το οποίο είναι διαφορετικό από το μήκος κύματος διέγερσης. Ο ανιχνευτής, που χρησιμοποιείται για τον εκπεμπόμενο από τη μικροσυστοιχία φθορισμό, είναι ευαίσθητος στο μήκος κύματος εκπομπής, αλλά φιλτράρει και το μήκος κύματος διέγερσης: με τον τρόπο αυτό το φθορίζον φως που μας ενδιαφέρει μπορεί να διαχωριστεί από το φως του λέιζερ που διαχέεται από το πλακίδιο.

Ένα καλό ζευγάρι ανιχνευτών για ένα πείραμα συγκριτικού υβριδισμού θα πρέπει να διαθέτει πολλά διαφορετικά μήκη κύματος εκπομπής ή διέγερσης. Φως το οποίο εκπέμπεται από τις δύο χρωστικές, μπορεί επιλεκτικά να φιλτραριστεί, και να μετρηθεί ξεχωριστά το ποσό που εκπέμπεται από κάθε χρωστική. Τα μήκη κύματος

διέγερσης είναι διαφορετικά, καθιστώντας δυνατή την διέγερση κάθε χρωστικής και στη συνέχεια της σάρωσης της κάθε μιας, ξεχωριστά.

Με τη σάρωση της μικροσυστοιχίας αποκτάται μια εικόνα της μικροσυστοιχίας με τις υβριδοποιημένες χρωστικές ουσίες φθορισμού. Η ένταση του φθορισμού εκπομπής των χρωστικών ουσιών αποτυπώνεται σε υψηλής ευκρίνειας μονοχρωματικές εικόνες που αποκτώνται για κάθε φθορίζουσα χρωστική ουσία. Το λογισμικό των σαρωτών δημιουργεί έπειτα μια σύνθετη χρωματισμένη εικόνα για υβριδοποιήσεις που περιλαμβάνουν πολλές χρωστικές ουσίες. Ο στόχος είναι να μετρηθεί, για κάθε κουκίδα της μικροσυστοιχίας, το σχετικό ποσό φθορισμού από κάθε χρωστική ουσία που υβριδοποιείται με το στόχο της. Έπειτα, προσδιορίζονται οι ακολουθίες ανίχνευσης DNA και ποσοτικοποιούνται οι εντάσεις από την υψηλής ευκρίνειας εικόνα (Lee 2004, Prasad 2003, Causton, Quackenbush & Brazma 2003).

Η διαδικασία σάρωσης των εικόνων από τυπωμένες cDNA μικροσυστοιχίες μπορεί να χωριστεί σε τρία τμήματα, συμπεριλαμβανομένου της εφαρμογής πλέγματος (gridding), της κατάτμησης (segmentation), και της εκτίμησης της έντασης (intensity extraction).

Η **εφαρμογή πλέγματος** είναι η διαδικασία απόδοσης συντεταγμένων σε κάθε μία από τις κουκίδες. Ένα πλέγμα τοποθετείται κατά προσέγγιση στην θέση των κουκίδων, και ένας αλγόριθμος χρησιμοποιείται για να καθορίσει με ακρίβεια τη θέση των κουκίδων και να ταξινομήσει τα εικονοστοιχεία (pixels) ως τμήμα σημείου ενδιαφέροντος ή του υποβάθρου. Επίσης, το αρχείο πλέγματος περιέχει πληροφορίες για την ταυτότητα της κάθε κουκίδας.

Η **διαδικασία κατάτμησης**, επιτρέπει την ταξινόμηση των εικονοστοιχείων είτε ως σημείο ενδιαφέροντος είτε ως υπόβαθρο. Οι αλγόριθμοι κατάτμησης σχεδιασμένοι να ανιχνεύουν τα όρια των κουκίδων, περιλαμβάνουν τους αλγορίθμους ανίχνευσης ακμών, τις μεθόδους ιστογράμματος και την κατάτμηση με βάση καθορισμένους και προσαρμοσμένους κύκλους.

Μόλις αποφασιστούν οι θέσεις των κουκίδων, υπολογίζονται οι εντάσεις φθορισμού όλων των εικονοστοιχείων μέσα σε κάθε κουκίδα και έξω από αυτή (δηλ., το τοπικό υπόβαθρο) και δημιουργείται ένα αρχείο αποτελεσμάτων, σε ένα βήμα που καλείται **εκτίμηση έντασης**. Αυτό περιλαμβάνει τον υπολογισμό, για κάθε κουκίδα της μικροσυστοιχίας, της εντάσεως φθορισμού του σήματος ενδιαφέροντος, της εντάσεως φθορισμού του υποβάθρου, των θέσεων, των μέσων όρων, των διαμέσων, των τυπικών αποκλίσεων, κ.α. και ενδεχομένως ποιοτικών μέτρων.

Εάν ένα συγκεκριμένο γονίδιο είναι πολύ ενεργό, παράγει πολλά μόρια αγγελιοφόρου RNA, τα οποία υβριδοποιούνται με το DNA στη μικροσυστοιχία και παράγουν μια πολύ φωτεινή περιοχή φθορισμού. Γονίδια που είναι λιγότερο ενεργά παράγουν λιγότερα μόρια mRNA και οδηγούν σε πιο αμυδρά σήματα φθορισμού. Εάν δεν υπάρχει φθορισμός, κανένα από τα μόρια mRNA δεν έχει υβριδοποιηθεί στο DNA, δείχνοντας ότι το γονίδιο είναι ανενεργό. Οι ερευνητές χρησιμοποιούν συχνά αυτήν την τεχνική για να εξετάσουν τη δραστηριότητα των διάφορων γονιδίων σε διαφορετικές στιγμές.

Χρησιμοποιώντας το λογισμικό ανάλυσης εικόνας, καθορίζονται οι εντάσεις των σημάτων κάθε χρωστικής ουσίας σε κάθε κουκίδα της μικροσυστοιχίας και προσδιορίζεται ο λόγος (Cy5/Cy3). Μια θετική τιμή για το λόγο (Cy5/Cy3) δείχνει μια σχετική περίσσεια του αντιγράφου που έχει χρωστεί με Cy5 στο δείγμα, ενώ μια αρνητική τιμή του λόγου (Cy5/Cy3) είναι ενδεικτική μιας σχετικής περίσσειας των αντιγράφων του δείγματος που έχουν σημειωθεί με Cy3. Στη συνέχεια τα δεδομένα αναλύονται (Lee 2004, Prasad 2003, Causton, Quackenbush & Brazma 2003).

4.8. ΕΠΑΛΗΘΕΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Η τεχνολογία μικροσυστοιχιών χρησιμοποιήθηκε περισσότερο ως ερευνητική μεθοδολογία και τα αποτελέσματα της επαληθεύονται με πιο ακριβείς μεθόδους. Η δυναμική της έγκειται στο γεγονός ότι δίνει τη δυνατότητα να ερευνηθεί ολόκληρο σχεδόν το γονιδίωμα σε ένα μόνο πείραμα, προσφέροντας μια ολιστική προσέγγιση που μπορεί να αποκαλύψει πρότυπα έκφρασης που οι παραδοσιακές μέθοδοι δεν μπορούν.

Η ανάλυση των μικροσυστοιχιών χρησιμοποιείται συχνά για την ανακάλυψη των γονιδίων που είναι υποψήφια για περαιτέρω ανάλυση. Επειδή το κόστος κατασκευής των μικροσυστοιχιών είναι μεγάλο στις μέρες μας, πριν την εκτέλεση εκτενέστερων πειραμάτων, απαιτείται επαλήθευση των αποτελεσμάτων με ανεξάρτητες μεθόδους, όπως η Northern blot ανάλυση και η ποσοτική μέθοδος αντίστροφης μεταγραφής PCR πραγματικού χρόνου (quantitative real-time RT-PCR- qRT-PCR).

Η Northern blot ανάλυση είναι μια μέθοδος στην οποία RNA ανιχνεύεται με υβριδισμό με ραδιοσημασμένο cDNA. Πρόκειται για μεταφορά μορίων RNA από γέλη αгарόζης, όπου και προηγήθηκε διαχωρισμός, σε μια μεμβράνη με τριχοειδή προσρόφηση. Μετά τη μεταφορά στη μεμβράνη, πραγματοποιείται υβριδισμός με ραδιοσημασμένο cDNA συμπληρωματικό του υπό μελέτη γονιδίου, και με τη βοήθεια ακτίνων-X λαμβάνεται μια εκτίμηση της σχετικής έκφρασης αυτού.

Στην μέθοδο qRT-PCR αρχικά το RNA μεταγράφεται αντίστροφα σε cDNA και στη συνέχεια η ενίσχυση καταγράφεται με φθορισμό σε πραγματικό χρόνο σε κάθε κύκλο επαύξησης. Οι τιμές φθορισμού που καταγράφονται κατά τη διάρκεια κάθε κύκλου αντιπροσωπεύουν την ποσότητα των επαυξημένων προϊόντων με PCR. Το σημείο στο οποίο καταγράφεται στατιστικά σημαντικό σήμα σε σχέση με το υπόβαθρο, αποκαλείται κύκλος κατώφλι. Η ποσότητα που καταγράφεται κατά τον κύκλο κατώφλι είναι αντιστρόφως ανάλογη του λογαρίθμου της αρχικής ποσότητας. Κατασκευάζονται έτσι τυπικές καμπύλες γνωστών ποσοτήτων αρχικών προτύπων, που χρησιμοποιούνται στη συνέχεια για να μπορούν με ακρίβεια να ποσοτικοποιηθούν και τα υπό μελέτη άγνωστα δείγματα (Lee 2004).

5. ΕΠΙΣΚΟΠΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΕΠΕΞΕΡΓΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα που λαμβάνονται από τις μικροσυστοιχίες είναι θορυβώδη, εξαιτίας διαφόρων πηγών διακύμανσης στα πειράματα. Προκειμένου να αποκτηθούν αξιόπιστα δεδομένα γονιδιακής έκφρασης για περαιτέρω ανάλυση (π.χ. ομαδοποίηση) και ερμηνεία, οι πειραματικές διαδικασίες πρέπει να είναι αυστηρά τυποποιημένες, ώστε να ελαχιστοποιηθεί ο θόρυβος και έχουν αναπτυχθεί διαφορετικές μέθοδοι για τη διόρθωση θορύβου υποβάθρου, το φιλτράρισμα των «κακής» ποιότητας σημείων και για την κανονικοποίηση δεδομένων.

5.1. ΔΙΟΡΘΩΣΗ ΘΟΡΥΒΟΥ ΥΠΟΒΑΘΡΟΥ

Η μέση ένταση μιας κουκίδας, υπολογίζεται από το μέσο όρο όλων των εικονοστοιχείων της και αποτελείται από το σήμα φθορισμού του cDNA που υβριδίζεται στην κουκίδα, το φθορισμό υποβάθρου που προκύπτει από τη μη ειδική υβριδοποίηση και από ανεπιθύμητες συνεισφορές από το θόρυβο και το σφάλμα μέτρησης φωτονίων (Lee 2004).

Έτσι τα δεδομένα των μικροσυστοιχιών πρέπει να διορθωθούν όσον αφορά στο θόρυβο του υποβάθρου για να εξαλειφθούν τα φαινόμενα μη ειδικής υβριδοποίησης ή της χωρικής ετερογένειας κατά μήκος της συστοιχίας.

Το λογισμικό ανάλυσης εικόνας, μέτρα τις εντάσεις στα κανάλια CH1 και CH2, οι οποίες είναι μη διορθωμένες μέσες εντάσεις των εικονοστοιχείων για κάθε κουκίδα της μικροσυστοιχίας για τα σήματα φθορισμού από τον υβριδισμό των δύο επισημασμένων δειγμάτων και επιστρέφει τις εντάσεις του σήματος ενδιαφέροντος και του υποβάθρου για κάθε κουκίδα. Το **σήμα ενδιαφέροντος (foreground)** είναι μια μέτρηση της συνολικής έντασης της κουκίδας, ενώ το **υπόβαθρο** είναι η μέτρηση του περιβαλλοντικού σήματος (Quackenbush 2001).

Ο φθορισμός του υποβάθρου μπορεί να προκύψει από πολλές πηγές, όπως τη μη ειδική υβριδοποίηση του επισημασμένου δείγματος στην επιφάνεια της συστοιχίας, τον φυσικό φθορισμό της γυάλινης πλάκας, τον οπτικό θόρυβο από τον ανιχνευτή ή να είναι αποτέλεσμα της επεξεργασίας. Η αφαίρεση του περιβαλλοντικού, μη ειδικού σήματος υβριδοποίησης από τη συνολική ένταση είναι γνωστή ως διόρθωση υποβάθρου. Απαιτείται για το λόγο αυτό μια εκτίμηση του φθορισμού του υποβάθρου.

Υπάρχουν διάφορες μέθοδοι εκτίμησής του. Οι τιμές έντασης του τοπικού υποβάθρου αποτελούν μια εκτίμηση του υποβάθρου, αν ληφθούν υπόψη τα εικονοστοιχεία που περιβάλλουν την κάθε κουκίδα. Είναι ο μέσος όρος ή η μεσαία τιμή των εντάσεων αυτών των εικονοστοιχείων. Αυτή η διαδικασία έχει το πλεονέκτημα ότι αντιμετωπίζει κάθε κουκίδα ξεχωριστά, επιτρέποντας να εκτιμηθούν διακυμάνσεις στο υπόβαθρο.

Μια άλλη προσέγγιση είναι η εκτίμηση του ολικού (global) υποβάθρου, που αποτελεί ουσιαστικά μια μόνο μέτρηση του υποβάθρου, που χρησιμοποιείται στη

συνέχεια σε όλη τη μικροσυστοιχία. Η προσέγγιση αυτή προϋποθέτει την επιλογή μιας μεγάλης περιοχής που δεν περιέχει στοιχεία ενδιαφέροντος. Το μειονέκτημα της είναι ότι δεν λαμβάνονται υπόψη διακυμάνσεις στις εντάσεις του τοπικού υποβάθρου και μπορεί έτσι να υπάρξει υπέρ ή υπό-εκτίμηση του υποβάθρου.

Αν υπάρχουν μη ειδικοί υβριδισμοί στην επιφάνεια της μικροσυστοιχίας τόσο η μέθοδος «τοπικού» όσο και «ολικού» υποβάθρου, υπερεκτιμούν το υπόβαθρο. Γι' αυτό εμφανίζονται στις μικροσυστοιχίες «αρνητικά σημεία» (negative spots) ή «μαύρες τρύπες» (Lee 2004, Causton, Quackenbush & Brazma 2003).

5.1.1. Μέθοδοι διόρθωσης υποβάθρου

Η συνηθέστερη πρακτική διόρθωσης του υποβάθρου είναι η αφαίρεση των εκτιμήσεων του τοπικού υποβάθρου. Η βασική υπόθεση στη διόρθωση υποβάθρου, για μικροσυστοιχίες με δύο χρωστικές ουσίες, είναι ότι τα σήματα υποβάθρου, Rb και Gb, είναι προσθετικά στα αληθινά σήματα έντασης, R και G στα ακατέργαστα δεδομένα. Δεδομένων των εντάσεων των σημάτων ενδιαφέροντος, Rf και Gf, το πραγματικό σήμα υπολογίζεται με την αφαίρεση των τιμών τους, έτσι ώστε $R = Rf - Rb$ και $G = Gf - Gb$. Οι διορθωμένες εντάσεις χρησιμοποιούνται έπειτα για να διαμορφώσουν τις λογαριθμικές αναλογίες

$$M = \log_2(R/G) \quad \text{και} \quad A = \frac{1}{2} \log_2 RG$$

για κάθε κουκίδα.

Η παραδοσιακή αυτή προσέγγιση όμως δημιουργεί διορθωμένες εντάσεις με ανεπιθύμητες στατιστικές ιδιότητες. Η διαδικασία διορθώσεων μπορεί να παραγάγει αρνητικές τιμές για την γονιδιακή έκφραση, ειδικά όταν το στοιχείο είναι μικρό (ή μηδέν) ή όταν η εκτίμηση του υποβάθρου είναι μεγάλη, οδηγώντας σε απώλεια λόγων έκφρασης, μερικές φορές για ένα σημαντικό ποσοστό αλληλουχιών ελέγχου μιας μικροσυστοιχίας. Επίσης προκαλεί υψηλή διακύμανση στους λόγους έκφρασης των χαμηλών εντάσεων.

Μια εναλλακτική επιλογή αποτελεί η μη διόρθωση του υποβάθρου, κατά την οποία δε λαμβάνονται υπόψη τα σήματα υποβάθρου, δηλαδή $Rb = Gb = 0$, οπότε τα σήματα ενδιαφέροντος είναι οι ακατέργαστες τιμές των εντάσεων που λαμβάνονται για τα δύο κανάλια.

Υπάρχουν νέες μέθοδοι για τη διόρθωση του υποβάθρου που κάποιες δίνουν καλύτερα και κάποιες χειρότερα αποτελέσματα σε σχέση με την παραδοσιακή μέθοδο. Η εγκατάλειψη της παραδοσιακής μεθόδου αφαίρεσης του υποβάθρου, θα είναι δυνατή όταν οι μέθοδοι αυτές βελτιωθούν και απαλειφθεί το χρόνιο πρόβλημα με τις τιμές που «λείπουν», που περιπλέκουν την ανάλυση των δεδομένων.

Η απλή αφαίρεση του θορύβου υποβάθρου που παρέχεται από το λογισμικό των οργάνων δεν παράγει απαραίτητα μια καλύτερη ανάγνωση της γονιδιακής

έκφρασης. Πρόκειται για μια εκτίμηση του θορύβου του υποβάθρου. Και μια κακή εκτίμηση μπορεί να είναι χειρότερη από καμία εκτίμηση (Ritchie et al. 2007).

Διαγνωστικοί έλεγχοι είναι διαθέσιμοι για να ελέγξουν εάν η διόρθωση του θορύβου του υποβάθρου είναι αποτελεσματική. Οι έλεγχοι δεν είναι καταληκτικοί αλλά μπορούν να είναι χρήσιμοι. Οι αποτελεσματικότεροι έλεγχοι περιλαμβάνουν την εξέταση γονιδίων για τα οποία υπάρχει υποψία ότι δεν είναι εκφρασμένα. Αυτά θα είναι γονίδια τα οποία έχουν πολύ μικρή ένταση φθορισμού (Lee 2004).

5.2. ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ

Τα δεδομένα από κάθε γονίδιο αναφέρονται ως **λόγοι έκφρασης** (expression ratios). Ο λόγος έκφρασης είναι η τιμή του επιπέδου έκφρασης ενός συγκεκριμένου γονιδίου στο προς εξέταση δείγμα διαιρούμενο με την τιμή του στο δείγμα ελέγχου. Οι γονιδιακές εκφράσεις δεν θα είχαν τις επιθυμητές στατιστικές ιδιότητες αν δεν μετασχηματιζόνταν. Υπάρχουν διάφορες μέθοδοι μετασχηματισμού.

5.2.1. Λογαριθμικός μετασχηματισμός

Ο πιο κοινός μετασχηματισμός είναι ο λογαριθμικός μετασχηματισμός.

$$y_{gn} = \log w_{gn}$$

όπου y_{gn} είναι η μετασχηματισμένη τιμή της έντασης και w_{gn} είναι η αρχική τιμή έντασης.

Η βάση του λογαρίθμου μπορεί να είναι 2, 10, ή η φυσική λογαριθμική σταθερά e. Εάν τα δεδομένα έχουν υποστεί διόρθωση υποβάθρου, τα μετασχηματισμένα δεδομένα δεν μπορούν να προσδιοριστούν αν οι αρχικές τους τιμές είναι μηδενικές ή αρνητικές. Άρα αυτός ο τύπος μετασχηματισμού είναι δυνατόν να χρησιμοποιηθεί όταν οι εντάσεις σήματος ξεπερνούν τις εντάσεις του υποβάθρου.

Το πλεονέκτημα της χρήσης λογαρίθμων ως λόγο έκφρασης είναι κατανοητή, αφού γονίδια που είναι υπερεκφρασμένα κατά έναν παράγοντα 2 έχουν λόγο έκφρασης 2, ενώ τα υποεκφρασμένα κατά τον ίδιο παράγοντα έχουν λόγο έκφρασης 0.5 – έτσι τα υποεκφρασμένα γονίδια θα «συνωστίζονταν» μεταξύ των τιμών 1 και 0.5. Αντιθέτως, με τη χρήση λογαρίθμων τα υπερεκφρασμένα γονίδια κατά έναν παράγοντα 2 έχουν λογαριθμικό λόγο 1, ενώ τα υποεκφρασμένα κατά τον ίδιο παράγοντα έχουν λογαριθμικό λόγο - 1, και γονίδια με σταθερή έκφραση (δηλαδή με λόγο 1) έχουν λογαριθμικό λόγο 0. Έτσι προκύπτουν τιμές που ακολουθούν κατά προσέγγιση κανονική κατανομή.

Οι περισσότερες δημοσιευμένες μελέτες έχουν χρησιμοποιήσει για την εύρεση των διαφορικά εκφρασμένων γονιδίων, ένα κατώφλι διπλάσιας αύξησης ή μείωσης στα μετρημένα επίπεδα έκφρασης, αν και δεν υπάρχει θεωρητική βάση για αυτή την

επιλογή. Εναλλακτικά, υπάρχουν προσεγγίσεις όπως αυτή από των Chen et al., που χρησιμοποιούν διαστήματα εμπιστοσύνης για το σκοπό αυτό.

5.2.2. Μετασχηματισμός τετραγωνικής ρίζας

Τα δεδομένα των μικροσυστοιχιών συχνά έχουν διακύμανση που τείνει να αυξάνεται καθώς αυξάνονται τα επίπεδα έκφρασης. Αν αυτή η σχέση είναι τέτοια που η διακύμανση να αυξάνεται σε αναλογία με τη μέση τιμή, τότε το μοντέλο κατανομής Poisson είναι εφαρμόσιμο και η συνάρτηση τετραγωνικής ρίζας

$$y_{gn} = \sqrt{w_{gn}}$$

είναι ένας μετασχηματισμός που σταθεροποιεί τη διακύμανση. Συνεπώς οι μετασχηματισμένες τιμές y_{gn} τείνουν να έχουν μια συνεχή διακύμανση.

5.2.3. Box-Cox μετασχηματισμός

Σε αυτή την οικογένεια μετασχηματισμών που προσδιορίζεται από τη σχέση

$$y_{gn} = \frac{w_{gn}^d - 1}{d}$$

ανήκουν και οι προηγούμενοι δύο τύποι μετασχηματισμών. Όταν $d=1/2$ έχουμε μετασχηματισμό τετραγωνικής ρίζας, ενώ ο λογαριθμικός μετασχηματισμός είναι το όριο της παραπάνω σχέσης με $d \rightarrow 0$. Όταν $d=1$ δεν έχουμε μετασχηματισμό. Αυτός ο τύπος μετασχηματισμού δίνει ένα εύρος μετασχηματισμών που επιτρέπει με κατάλληλη επιλογή του d να προκύπτουν μετασχηματισμένες τιμές με επιθυμητές στατιστικές ιδιότητες.

5.2.4. Μετασχηματισμός ομοιότητας (affine transformation)

Κάποιες φορές είναι απαραίτητο να αλλάξουμε την αρχή (origin) της κλίμακας των μετρούμενων εντάσεων κατά μια σταθερή ποσότητα a . Έτσι αν w_{gn} είναι οι αρχικές εντάσεις, $a+w_{gn}$ είναι οι προσαρμοσμένες. Το a αναφέρεται σαν **παράμετρος μετατόπισης ή αντιστάθμισης** (offset). Μπορεί να χρησιμοποιηθεί σε συνδυασμό με άλλους μετασχηματισμούς, όπως τον λογαριθμικό μετασχηματισμό για την δημιουργία των θετικών ορισμάτων που απαιτούνται. Επίσης σε μικροσυστοιχίες cDNA δύο χρωστικών ουσιών, χρησιμοποιώντας το σχετικό μετασχηματισμό,

$$y_{gn} = \log \left(\frac{w_{gn}^{(R)} - a_n}{w_{gn}^{(G)} + a_n} \right)$$

είναι δυνατό να περιοριστεί η μη γραμμικότητα στο διάγραμμα διασποράς. Η παράμετρος α_n επιλέγεται ξεχωριστά για κάθε μικροσυστοιχία και εφαρμόζεται σε όλα τα γονίδια αυτής. Η παράμετρος αυτή αντιστάθμισης διατηρεί τη μέση ένταση των δύο χρωστικών σχεδόν αμετάβλητη, αλλά μετατοπίζει τα δεδομένα ανάγνωσης των εντάσεων φθορισμού το ένα σε σχέση με το άλλο. Αυτή η αλλαγή είναι καταφανής σε χαμηλές εντάσεις (Lee 2004).

Υπάρχουν όμως και μειονεκτήματα από τη χρήση μόνο λόγων έκφρασης για την ανάλυση των δεδομένων. Αν και οι λόγοι βοηθούν στην αποκάλυψη προτύπων (patterns) στα δεδομένα, αφαιρούν όλες τις πληροφορίες σχετικά με τις απόλυτες τιμές των επιπέδων γονιδιακής έκφρασης. Ποικίλες παράμετροι εξαρτώνται από τις μετρούμενες εντάσεις, συμπεριλαμβανομένων των διαστημάτων εμπιστοσύνης. Αν και οι περισσότερες τεχνικές που έχουν αναπτυχθεί για την ανάλυση των δεδομένων από μικροσυστοιχίες χρησιμοποιούν λόγους έκφρασης, πολλές από αυτές μπορούν να προσαρμοστούν για να χρησιμοποιηθούν με απόλυτες τιμές έντασης (Quackenbush 2001).

5.3. ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

5.3.1. Γενικά

Σκοπός της κανονικοποίησης είναι η ελαχιστοποίηση της επιρροής εξωγενών παραγόντων στα μετρούμενα επίπεδα γονιδιακής έκφρασης των υβριδισμένων δειγμάτων mRNA, έτσι ώστε να είναι πιο εύκολη η διάκριση βιολογικών διαφορών (διαφορικά εκφρασμένα γονίδια - differentially expressed genes).

Η κανονικοποίηση συνήθως βασίζεται στα επίπεδα έκφρασης όλων των γονιδίων της μικροσυστοιχίας. Υποθέτοντας ότι τα περισσότερα γονίδια είναι όμοια εκφρασμένα τόσο στο προς εξέταση δείγμα όσο και στο δείγμα αναφοράς και ότι ο αριθμός των υπερεκφρασμένων γονιδίων ισούται με αυτό των υποεκφρασμένων, η κανονικοποίηση μπορεί να βασιστεί στη συνολική ένταση φθορισμού, στους λόγους έκφρασης και στην εμπειρική σχέση μεταξύ των μεταβλητών (regression analysis).

Η κανονικοποίηση διορθώνει διαφορές κατά τη σήμανση των δειγμάτων, ατέλειες κατά την ανίχνευση του φθορισμού των χρωστικών ουσιών και διαφορές που οφείλονται στην αρχική ποσότητα RNA των δύο δειγμάτων που εξετάζονται. Αυτά τα προβλήματα μπορούν να προκαλέσουν αλλαγή στη μέση τιμή του λόγου Cy5/Cy3 και οι εντάσεις πρέπει να επαναταξινομηθούν πριν αναλυθεί το πείραμα (Quackenbush 2001).

Υπάρχουν τρεις ευρέως χρησιμοποιημένες τεχνικές που μπορούν να χρησιμοποιηθούν για να κανονικοποιήσουν τα δεδομένα γονιδιακής έκφρασης από υβριδισμό σε μια μικροσυστοιχία. Όλες υποθέτουν ότι όλα (ή τα περισσότερα) από τα γονίδια στη μικροσυστοιχία, κάποια υποσύνολα των γονιδίων, ή ένα σύνολο εξωγενών ελέγχων που έχουν εμβολιαστεί (spiked) στο RNA πριν την ιχνηθέτηση, έχουν μέση αναλογία έκφρασης ίση με 1. Προσδιορίζεται έτσι ένας παράγοντας

κανονικοποίησης N_{total} και χρησιμοποιείται έπειτα για να ρυθμίσει τα δεδομένα ώστε να αντισταθμιστούν διαφοροποιήσεις κατά την πειραματική διαδικασία και να ισοσκελιστούν τα σήματα φθορισμού από τα δύο δείγματα που συγκρίνονται.

Υπάρχουν πολλές πηγές συστηματικής απόκλισης και τυχαίων λαθών σε πειράματα cDNA μικροσυστοιχιών, που έχουν επιπτώσεις στα μετρούμενα επίπεδα της έντασης φθορισμού. Μια γνωστή πηγή συστηματικής απόκλισης αποτελεί η μεροληψία που σχετίζεται με τις διάφορες χρωστικές ουσίες φθορισμού. Αυτό είναι εμφανές σε πειράματα που δύο ακριβώς ίδια δείγματα mRNA ιχνηθετούνται με διαφορετικές χρωστικές και στη συνέχεια υβριδίζονται στο ίδιο πλακίδιο. Σε αυτήν την περίπτωση, είναι σπάνιο να προκύψουν εντάσεις που είναι ίδιες για τις δύο φθορίζουσες ουσίες κατά μήκος όλων των κουκίδων. Αν και τέτοιες συστηματικές αποκλίσεις είναι μικρές, μπορούν να δημιουργήσουν σύγχυση όταν αναζητούνται «δυσδιάκριτες» βιολογικές διαφορές.

Η μεροληψία λόγω των χρωστικών, όπως ονομάζεται, μπορεί να προκύψει από μια ποικιλία παραγόντων συμπεριλαμβανομένων των φυσικών ιδιοτήτων των χρωστικών ουσιών (θερμότητα, ευαισθησία στο φως, χρόνος ημίσεως ζωής), της απόδοσης ενσωμάτωσης των χρωστικών ουσιών, της πειραματικής μεταβλητότητας κατά τον υβριδισμό και τις διαδικασίες επεξεργασίας, και των ρυθμίσεων σάρωσης κατά το βήμα συλλογής των δεδομένων (Yang et al. 2002, Quackenbush 2002).

Στα πειράματα βαθμονόμησης-ρύθμισης των μετρητικών διατάξεων (calibration experiments), το ίδιο δείγμα ιχνηθετείται και με τις δύο φθορίζουσες χρωστικές ουσίες και υβριδίζεται στο ίδιο πλακίδιο. Τα πειράματα βαθμονόμησης παρέχουν τη δυνατότητα ελέγχου για την εξερεύνηση πιθανών συστηματικών επιδράσεων, όπως η επίδραση της αυτής μικροσυστοιχίας (slide effect), η επίδραση των χρωστικών και οι αλληλεπιδράσεις των ιχνηθετημένων γονιδίων (Tseng et al. 2001, Yang, Y.H. 2001).

Τα διαφορετικά πλακίδια εισάγουν παραλλαγές τόσο στον υβριδισμό όσο και στην απεικόνιση. Οι παράγοντες που έχουν επιπτώσεις στον υβριδισμό περιλαμβάνουν την ποσότητα των αλληλουχιών ανίχνευσης DNA που ακινητοποιείται στο πλακίδιο κατά τη διάρκεια της επεξεργασίας της μικροσυστοιχίας, την ποσότητα ιχνηθετημένων cDNA που προστίθενται στο πλακίδιο και το τοπικό περιβάλλον σε κάθε αίθουσα υβριδοποίησης.

Οι παραλλαγές μέσα στο ίδιο πλακίδιο, οφείλονται σε περιοχές επιμόλυνσης, σε υψηλής έντασης υπόβαθρο ή στον μη ομοιόμορφο υβριδισμό στην επιφάνεια των πλακιδίων. Στα πειράματα που χρησιμοποιούν πολλαπλές ακίδες για την ακινητοποίηση των αλληλουχιών ανίχνευσης DNA, η παραλλαγή από ακίδα σε ακίδα μπορεί να είναι αξιοσημείωτη.

Κατά τη διάρκεια της απεικόνισης ο παρασιτικός θόρυβος του πλακιδίου και η τοπική κυρτότητα της επιφανείας μπορούν να έχουν επιπτώσεις στην ανίχνευση του φθορισμού. Οι συνεστιακοί ανιχνευτές είναι ιδιαίτερα ευαίσθητοι στην εστίαση. Η επίδραση όλων αυτών των παραγόντων στη μέτρηση της έντασης, ορίζεται ως **επίδραση της αυτής μικροσυστοιχίας** (slide effects) (Tseng et al. 2001, Yang, Y.H. 2001).

Οι συνηθέστερα χρησιμοποιημένες χρωστικές ουσίες φθορισμού, Cy3 και Cy5, είναι σχετικά ασταθείς. Επιπλέον, αυτές οι χρωστικές ουσίες μπορούν να επηρεάσουν διαφορετικά την απόδοση ενσωμάτωσης κατά τη διάρκεια της ιχνηθέτησης, να έχουν διαφορετική κβαντική απόδοση και να ανιχνεύονται από τον ανιχνευτή με διαφορετική απόδοση. Η επίδραση αυτών των παραγόντων στις μετρούμενες εντάσεις, ορίζεται ως **επίδραση επισήμανσης** (label effect).

Στα πειράματα συγκριτικού υβριδισμού οι δύο διαφορεικά εκφρασμένες «δεξαμενές» mRNA ιχνηθετούνται χωριστά με Cy3 ή Cy5 και υβριδίζονται στο ίδιο πλακίδιο. Για να μελετήσουμε την επίδραση της επισήμανσης, σχεδιάζουμε τις γραφικές παραστάσεις M-A σε πειράματα βαθμονόμησης. Επειδή τα δύο διαλύματα cDNA προέρχονται από την ίδια ομάδα RNA, οι μετρήσεις στα Cy3 και Cy5 κανάλια θα έπρεπε να είναι ίδιες, αν η επίδραση της επισήμανσης είναι αμελητέα. Σε αυτήν την ιδανική περίπτωση η γραφική παράσταση M-A βρίσκεται γύρω από τη γραμμή $M = 0$.

Η κανονικοποίηση όσον αφορά στην επίδραση της επισήμανσης είναι μη γραμμική και εξαρτάται από το πλακίδιο. Για να πραγματοποιηθεί η κανονικοποίηση επισήμανσης σε ένα συγκριτικό πείραμα, πρέπει να προσδιοριστεί ένας ικανοποιητικός αριθμός μη-διαφορικά εκφρασμένων γονιδίων σε κάθε πλακίδιο και να χρησιμοποιηθούν για την κατασκευή της καμπύλης κανονικοποίησης (Tseng et al. 2001, Lee 2004, Quackenbush 2002).

Η χρήση δύο χρωστικών μπορεί επίσης να εισαγάγει αλληλεπιδράσεις στα ιχνηθετημένα γονίδια. Για παράδειγμα τα Cy3-dCTP μπορεί να ενσωματωθούν κατά προτίμηση σε μια συγκεκριμένη ακολουθία, σε σχέση με τα Cy5-dCTP. Εάν υπάρξει μια τέτοια αλληλεπίδραση, ορισμένα γονίδια θα παρουσιάζουν πάντα υψηλότερη ένταση σε ένα από τα δύο κανάλια, ακόμη και κάτω από συνθήκες μη διαφορικής έκφρασης και μετά από την κανονικοποίηση. Θεωρητικά αυτή η αλληλεπίδραση θα υπάρχει σε κάποιο βαθμό, ωστόσο εμφανίζεται να είναι ασήμαντη σε μέγεθος, έναντι άλλων πηγών θορύβου (Tseng et al. 2001, Yang, Y.H. 2001).

Επιπλέον, τα σχετικά επίπεδα γονιδιακής έκφρασης από πανομοιότυπα πειράματα μπορούν να έχουν διαφορετικές διακυμάνσεις στα δείγματα, λόγω διαφορών στις πειραματικές συνθήκες. Μια προσαρμογή κλίμακας απαιτείται έτσι ώστε τα σχετικά επίπεδα έκφρασης από ένα συγκεκριμένο πείραμα να μην υπερσχύουν στα μέσα σχετικά επίπεδα έκφρασης σε πανομοιότυπα πειράματα. Ωστόσο μια τέτοια **ολική κανονικοποίηση** (global normalization) δεν είναι κατάλληλη για περιπτώσεις όπου η μεροληψία που υπεισέρχεται λόγω των χρωστικών ουσιών εξαρτάται από τη **συνολική ένταση της κουκίδας** και υπάρχει **χωρική εξάρτηση** για διαφορετικούς τύπους πειραμάτων.

Πολλοί από αυτούς τους παράγοντες, είτε εσωτερικοί είτε εξωτερικοί για το δείγμα, κάνουν τη διάκριση μεταξύ των διαφορικά και σταθερά εκφρασμένων γονιδίων δύσκολη.

Η **κανονικοποίηση** μπορεί να γίνει με διάφορους τρόπους, ανάλογα με το σχεδιασμό του πειράματος: (i) κανονικοποίηση μέσα στο ίδιο πλακίδιο, (ii) κανονικοποίηση σε ζευγάρια πλακιδίων, για πειράματα με εναλλαγή της χρωστικής ουσίας (dye-swap), και (iii) κανονικοποίηση μεταξύ πολλαπλών πλακιδίων. Σε κάθε

περίπτωση πρέπει να βρεθεί το σύνολο γονιδίων που πρέπει να χρησιμοποιηθούν για την κανονικοποίηση. Η επιλογή αυτή εξαρτάται από παράγοντες, όπως το ποσοστό των γονιδίων που αναμένεται να είναι διαφορετικά εκφρασμένα στα ιχνηθετημένα δείγματα με κόκκινη ή πράσινη χρωστική ουσία και η διαθεσιμότητα ακολουθιών ελέγχου (Yang et al. 2002, Yang, Y.H. 2001).

Επειδή τις περισσότερες φορές ένα μικρό ποσοστό των γονιδίων αναμένεται να είναι διαφορετικά εκφρασμένο, τα υπόλοιπα γονίδια αναμένεται να έχουν παρόμοια έκφραση και έτσι μπορούν να χρησιμοποιηθούν ως δείκτες των σχετικών εντάσεων των δύο χρωστικών ουσιών. Δηλαδή χρησιμοποιούμε όλα τα γονίδια για κανονικοποίηση, εάν (i) αναμένεται μόνο ένα σχετικά μικρό ποσοστό των γονιδίων να εκφραστεί διαφορετικά, ή (ii) υπάρχει συμμετρία στα επίπεδα έκφρασης των υπέρ και υπό εκφρασμένων γονιδίων.

Εναλλακτικά, αντί να χρησιμοποιηθούν όλα τα γονίδια της συστοιχίας, μπορεί να χρησιμοποιηθεί ένα υποσύνολο των γονιδίων, που συχνά ονομάζονται γονίδια «σταθερής έκφρασης» (housekeeping genes), που θεωρείται ότι έχουν σταθερή έκφραση κατά τη διάρκεια ποικίλων συνθηκών (π.χ. β-actin). Αν και είναι δύσκολο να προσδιοριστεί ένα τέτοιο σύνολο γονιδίων που δεν αλλάζει σημαντικά κάτω από οποιοδήποτε συνθήκες, είναι δυνατόν να βρεθούν γονίδια που είναι «προσωρινά» γονίδια σταθερής έκφρασης, για συγκεκριμένες πειραματικές συνθήκες. Ένας περιορισμός των γονιδίων σταθερής έκφρασης είναι ότι τείνουν να είναι υψηλά εκφρασμένα και συνεπώς δεν είναι αντιπροσωπευτικά των άλλων γονιδίων ενδιαφέροντος (Yang, Y.H. 2001, Yang et al. 2002).

Μια άλλη εναλλακτική, είναι η χρήση εμβολιασμένων ελέγχων ή μιας σειράς τιτλοδοτημένων ακολουθιών ελέγχου (titration series). Στη μέθοδο εμβολιασμού, συνθετικές ακολουθίες DNA ή οι ακολουθίες DNA από ένα διαφορετικό οργανισμό από αυτόν που μελετάται επικολλώνται στη συστοιχία (με πιθανά αντίγραφα) και συμπεριλαμβάνονται στα δύο διαφορετικά δείγματα mRNA σε ίσες ποσότητες. Συνεπώς αυτές οι επισημασμένες ακολουθίες ελέγχου θα έπρεπε να έχουν ίδιες εντάσεις φθορισμού για την κόκκινη και πράσινη χρωστική και θα μπορούσαν να χρησιμοποιηθούν για την κανονικοποίηση. Στη μέθοδο με τιτλοδοτημένες σειρές, κουκίδες που αποτελούνται από διαφορετικές συγκεντρώσεις του ίδιου γονιδίου ή EST τυπώνονται στη μικροσυστοιχία. Αυτές οι κουκίδες αναμένεται να έχουν ίδιες εντάσεις φθορισμού για την κόκκινη και πράσινη χρωστική για ένα εύρος εντάσεων (Yang, Y.H. 2001, Yang et al. 2002).

5.3.2. Ολική κανονικοποίηση

Σε αυτή την περίπτωση η κανονικοποίηση γίνεται ξεχωριστά σε κάθε πλακίδιο χρησιμοποιώντας μόνο τις εντάσεις φθορισμού της κόκκινης και της πράσινης χρωστικής αυτού του πλακιδίου.

Η μέθοδος της ολικής κανονικοποίησης υποθέτει ότι οι εντάσεις της κόκκινης και της πράσινης χρωστικής συνδέονται με ένα σταθερό παράγοντα, δηλ. $R = kG$ και πρακτικά το κέντρο της διασποράς των λογαριθμικών λόγων μετατοπίζεται στο μηδέν.

$$\log_2 R/G \rightarrow \log_2 R/G - c = \log_2 R/(kG)$$

Μια κοινή επιλογή για την παράμετρο θέσης $c = \log_2 k$ είναι η διάμεσος ή ο μέσος όρος των λογαριθμικών αναλογιών για μια συγκεκριμένη ομάδα γονιδίων. Οι μέθοδοι ολικής κανονικοποίησης αναφέρονται σαν βήματα προεπεξεργασίας των δεδομένων για τον προσδιορισμό διαφορικά εκφρασμένων γονιδίων σε πειράματα cDNA μικροσυστοιχιών ενός πλακιδίου. Υποθέτοντας ότι $R = kG$ προτείνεται μια επαναληπτική μέθοδος για την εκτίμηση του σταθερού παράγοντα κανονικοποίησης k και σημεία αποκοπής για τις αναλογίες R/G .

5.3.3. Κανονικοποίηση εξαρτώμενη από την ένταση

Όπως μπορεί να αποκαλυφθεί από τη γραφική παράσταση του λόγου $\log_2 R/G$ για κάθε στοιχείο της μικροσυστοιχίας ως προς το γινόμενο των εντάσεων $\log_{10} RG$, μπορεί να υπάρξει εξάρτηση από την ένταση. Η 'R-I' γραφική, όπως ονομάζεται μπορεί να αποκαλύψει ατέλειες που σχετίζονται με την ένταση. Η lowess ανιχνεύει συστηματικές αποκλίσεις στην 'R-I' γραφική ή αντίστοιχα στην 'M-A' γραφική και τις εξομαλύνει εφαρμόζοντας γραμμική παλινδρόμηση με τοπικά βάρη και αφαιρώντας την εκτιμώμενη, σύμφωνα με την καλύτερη προσαρμογή, τιμή έντασης από την πειραματικά παρατηρούμενη τιμή σε κάθε σημείο. Η συνάρτηση αυτή χρησιμοποιεί συναρτήσεις βάρους, οι οποίες μειώνουν τη συνεισφορά δεδομένων που βρίσκονται σε ακραίες θέσεις στη γραφική, αφού η τιμή των βαρών είναι μεγάλη όταν το σημείο βρίσκεται κοντά στην καμπύλη προσαρμογής των δεδομένων και μικρή αν βρίσκεται μακριά από αυτή.

Έτσι έχουμε μια A-εξαρτώμενη κανονικοποίηση

$$\log_2 R/G \rightarrow \log_2 R/G - c(A) = \log_2 R/(k(A)G)$$

όπου $c(A)$ είναι η lowess προσαρμογή στην M-A γραφική παράσταση.

Ο χρήστης καθορίζει μια παράμετρο f , για το ποσοστό των δεδομένων που χρησιμοποιούνται για την εξομάλυνση κάθε σημείου. Όσο μεγαλύτερη είναι η τιμή της παραμέτρου τόσο μεγαλύτερη είναι η εξομάλυνση (Yang, Y.H. 2001, Yang et al. 2002).

5.3.4. Κανονικοποίηση σε κάθε υποπλέγμα (print-tip group)

Κάθε υποπλέγμα μιας μικροσυστοιχίας εκτυπώνεται χρησιμοποιώντας την ίδια κεφαλή. Διαφορετικά πειράματα μπορούν να διεξαχθούν χρησιμοποιώντας διαφορετικές διατάξεις εκτύπωσης, που εξαρτώνται από το σχεδιάγραμμα των ακίδων της κεφαλής. (π.χ. 4x4 ή 2x2 κεφαλές εκτύπωσης).

Μπορούν να υπάρξουν συστηματικές διαφορές μεταξύ των ακίδων εκτύπωσης, όπως μικρές διαφορές στο μήκος ή στο άνοιγμα των ακίδων και παραμόρφωσης μετά από πολλές ώρες εκτύπωσης. Εναλλακτικά, οι ομάδες των ακίδων αντιπροσωπεύουν τις χωρικές επιδράσεις στο πλακίδιο. Η κανονικοποίηση μέσα στην ίδια ομάδα ακίδων, είναι μια (print-tip+A) εξαρτώμενη κανονικοποίηση, δηλαδή

$$\log_2 R/G \rightarrow \log_2 R/G - c_i(A) = \log_2 R/(k_i(A)G)$$

όπου $c_i(A)$ είναι η lowess προσαρμογή στην M-A γραφική παράσταση για το i -ιοστο πλέγμα, όπου $i=1, \dots, I$ αντιπροσωπεύει τον αριθμό των ακίδων εκτύπωσης (Yang, Y.H. 2001, Yang et al. 2002).

5.3.5. Κανονικοποίηση σε κάθε υποπλέγμα με κλιμακοποίηση

Μετά από την κανονικοποίηση σε κάθε υποπλέγμα, όλοι οι κανονικοποιημένοι λόγοι έκφρασης από τις διαφορετικές ομάδες ακίδων εκτύπωσης θα κεντραριστούν γύρω από το μηδέν. Ωστόσο είναι δυνατόν οι λόγοι έκφρασης από τις διάφορες ομάδες ακίδων εκτύπωσης να έχουν διαφορετική διασπορά και απαιτείται προσαρμογή κλίμακας.

Μια προσέγγιση είναι να υποθεθεί ότι οι λογαριθμικοί λόγοι έκφρασης από την i -ιοστη ομάδα ακίδων εκτύπωσης ακολουθούν μια κανονική κατανομή με μέση τιμή μηδέν και διασπορά $\alpha_i^2 \sigma^2$, όπου σ^2 είναι η διασπορά των πραγματικών λογαριθμικών λόγων και α_i^2 είναι ένας παράγοντας κλίμακας για την i -ιοστη ομάδα. Προκειμένου να πραγματοποιηθεί κανονικοποίηση κλίμακας, ο παράγοντας α_i για τις διαφορετικές ομάδες ακίδων εκτύπωσης πρέπει να εκτιμηθεί.

Επιβάλλοντας το φυσικό περιορισμό

$$\sum_{i=1}^I \log \alpha_i^2 = 0$$

όπου το I δηλώνει το συνολικό αριθμό των ομάδων στη μικροσυστοιχία, η εκτίμηση της μέγιστης πιθανότητας για το α_i είναι

$$\hat{\alpha}_i = \frac{\sum_{j=1}^{n_i} M_{ij}^2}{\sqrt{\prod_{k=1}^I \sum_{j=1}^{n_k} M_{kj}^2}}$$

όπου M_{ij} δηλώνει την j -ιοστή λογαριθμική αναλογία στην i -ιοστή ομάδα ακίδων εκτύπωσης, $j=1, \dots, n_i$.

Μια εναλλακτική λύση αυτής της εκτίμησης, που θεωρείται προτιμότερη είναι

$$\hat{\alpha}_i = \frac{MAD_i}{\sqrt{\prod_{k=1}^I MAD}}$$

όπου η μέση απόλυτη απόκλιση MAD προσδιορίζεται από τη σχέση

$$MAD_i = \text{median}_j \{ |M_{ij} - \text{median}_j(M_{ij})| \}$$

Αυτή η διαδικασία υποθέτει ότι ένα σχετικά μικρό ποσοστό των γονιδίων θα διαφέρει σημαντικά στην έκφραση του στα δύο δείγματα mRNA. Επιπλέον, υποθέτει ότι η διασπορά της διανομής των λογαριθμικών λόγων έκφρασης θα είναι σχεδόν ίδια για όλα τα υποπλέγματα. Η απόκλιση MAD, δεν επηρεάζεται από ένα μικρό ποσοστό διαφορετικά εκφρασμένων γονιδίων που εμφανίζονται σε ακραίες θέσεις (Yang, Y.H. 2001, Yang et al. 2002).

5.3.6. Κανονικοποίηση με εναλλαγή της χρωστικής (dye swap)

Η κανονικοποίηση με εναλλαγή της χρωστικής έχει ως εξής: δύο υβριδισμοί για τα δύο δείγματα mRNA με εναλλαγή των χρωστικών ουσιών στο δεύτερο υβριδισμό. Έστω ότι οι κανονικοποιημένες τιμές για το πρώτο πλακίδιο δίνονται από τη σχέση $\log_2 R/G - c$ και στην δεύτερη από την $\log_2 R'/G' - c'$. Εδώ τα c και c' αποτελούν τις συναρτήσεις κανονικοποίησης για τα δύο πλακίδια (μπορούν να έχουν ληφθεί με οποιαδήποτε μέθοδο κανονικοποίησης στο πλακίδιο). Αν $c \approx c'$ τότε

$$\begin{aligned} & \frac{1}{2} \left[\log_2 R/G - c - (\log_2 R'/G' - c') \right] \\ & \approx \frac{1}{2} \left[\log_2 R/G + \log_2 G'/R' \right] = \frac{1}{2} \log_2 RG'/GR' = \frac{1}{2} (M - M') \end{aligned}$$

Συνεπώς, μπορούμε να συνδυάσουμε τα σχετικά επίπεδα έκφρασης για τα δύο πλακίδια χωρίς ρητή κανονικοποίηση. Αναφερόμαστε σε αυτή τη διαδικασία ως **αυτο-κανονικοποίηση** (*self-normalization*). Η βασική υπόθεση εδώ είναι ότι το $c \approx c'$ και αυτή η μέθοδος μπορεί να εφαρμοστεί σε όλα τα γονίδια, ακόμα κι αν εκφράζονται διαφορετικά. Σύμφωνα με την προσέγγιση αυτή, τα γονίδια που δεν αναμένεται να εκφράζονται διαφορετικά θα είχαν

$$(M - M')/2 = \frac{1}{2} \log_2 RG'/GR' \approx 0$$

Η ισχύς αυτής της υπόθεσης μπορεί να ελεγχθεί με τη χρήση ενός συνόλου γονιδίων που αναμένεται να έχουν σταθερά επίπεδα έκφρασης (π.χ. γονίδια σταθερής έκφρασης), αν ένα τέτοιο σύνολο είναι διαθέσιμο.

Δεδομένου ότι οι χρωστικές ουσίες αντιστρέφονται στα δύο πειράματα, αναμένεται ότι οι κανονικοποιημένοι λόγοι έκφρασης στα δύο πλακίδια θα είναι ίσου μεγέθους και αντίθετου προσήμου, δηλαδή

$$\log_2 R/G - c \approx - (\log_2 R'/G' - c').$$

Συνεπώς, ρυθμίζοντας εκ νέου την εξίσωση και υποθέτοντας πάλι $c \approx c'$, μπορεί να εκτιμηθεί η συνάρτηση κανονικοποίησης c από τη σχέση :

$$c = \frac{1}{2} [\log_2 R/G + \log_2 R'/G'] = \frac{1}{2} (M + M')$$

Πρακτικά, η $c = c(A)$ εκτιμάται από την lowess προσαρμογή της παράστασης

$$\frac{1}{2} (M + M') = \frac{1}{2} \log_2 RR'/GG' \text{ vs } \frac{1}{2} (A + A'),$$

όπου όλα τα γονίδια χρησιμοποιούνται αυτή τη φορά.

Η παραπάνω μέθοδος κανονικοποίησης προσαρμόζει για χωρικές διαφορές μόνο και υποθέτει ότι η διασπορά των λογαριθμικών λόγων είναι ίδια και στα δύο πλακίδια (Yang, Y.H. 2001, Yang et al. 2002).

5.3.7. Κανονικοποίηση ανάμεσα στα πλακίδια

Μετά από την κανονικοποίηση μέσα στο ίδιο πλακίδιο, όλοι οι κανονικοποιημένοι λογαριθμικοί λόγοι, θα έχουν κεντραριστεί στο μηδέν, ανεξάρτητα από τη μέθοδο κανονικοποίησης. Η μέθοδος κανονικοποίησης μεταξύ των πλακιδίων, που στοχεύει στο να επιτρέψει τη σύγκριση μεταξύ πειραμάτων, μπορεί επίσης να απαιτεί ρύθμιση ως προς την κλίμακα, όταν τα διαφορετικά πλακίδια έχουν ουσιαστικές διαφορές στη διασπορά των λογαριθμικών τους λόγων.

Η αποτυχία πραγματοποίησης κανονικοποίησης κλίμακας θα μπορούσε να οδηγήσει ένα ή περισσότερα πλακίδια να έχουν αδικαιολόγητο βάρος κατά τον υπολογισμό του μέσου όρου των λογαριθμικών λόγων έκφρασης στα πειράματα (Causton, Quackenbush & Brazma 2003, Yang, Y.H. 2001, Yang et al. 2002).

5.4. ΦΙΛΤΡΑΡΙΣΜΑ ΔΕΔΟΜΕΝΩΝ

Η κανονικοποίηση είναι ένας από τους πολλούς μετασχηματισμούς που μπορεί να εφαρμοστούν στα δεδομένα. Ο σκοπός των περισσότερων άλλων μετασχηματισμών είναι το φιλτράρισμα των δεδομένων ώστε να μειωθεί η πολυπλοκότητα τους και να αυξηθεί η ποιότητά τους γενικά. Πολλοί είναι σχεδιασμένοι να επισημαίνουν αμφισβητήσιμα ή χαμηλής ποιότητας δεδομένα, ενώ άλλοι χρησιμοποιούνται για να προσδιορίσουν διαφορικά εκφρασμένα γονίδια ή για να ενισχύσουν ένα συγκεκριμένο χαρακτηριστικό των δεδομένων. Η επιλογή εξαρτάται από το σχεδιασμό του συγκεκριμένου πειράματος και από τα βιολογικά ερωτήματα που πρέπει να απαντηθούν (Causton, Quackenbush & Brazma 2003).

5.4.1. Φιλτράρισμα δεδομένων χαμηλής έντασης

Ένα φιλτράρισμα που εφαρμόζεται ευρέως στα δεδομένα περιορίζει τα χαρακτηριστικά της μικροσυστοιχίας που έχουν εντάσεις φθορισμού ελαφρώς υψηλότερες από αυτές του υποβάθρου. Η εξήγηση αυτής της προσέγγισης είναι απλή: αυτά τα σημεία με εντάσεις φθορισμού κοντά σε αυτές του υποβάθρου είναι τα περισσότερο μετρούμενα χωρίς ακρίβεια και συνεπώς είναι πιθανό να είναι «φτωχής» ή αμφισβητήσιμης ποιότητας. Τυπικά υπολογίζεται το τοπικό υπόβαθρο για κάθε κουκίδα της μικροσυστοιχίας και η τυπική του απόκλιση, σ , σε κάθε κανάλι. Αν χρησιμοποιούμε σαν τιμή έντασης, για τις εντάσεις φθορισμού του υποβάθρου, το μέσο, αυτές που είναι δύο φορές μικρότερες από την τυπική απόκλιση σ επισημαίνονται και απαλείφονται, έτσι ώστε η συνθήκη που ικανοποιείται να είναι

$$G_{median}^{spot} > 2 \sigma(G^{background}) \text{ και } R_{median}^{spot} > 2 \sigma(R^{background}).$$

Σαν εναλλακτική, μπορεί να χρησιμοποιηθεί ένα ακρότατο όριο (cut-off) που βασίζεται σε κάποιο ποσοστό, όπου αφαιρούνται κουκίδες αν έχουν εντάσεις μικρότερες από ένα αυθαίρετο κλάσμα του μέσου όρου του υποβάθρου. Κάποιες ομάδες ερευνητών χρησιμοποιούν καθορισμένες ελάχιστες εντάσεις για να προσδιορίσουν και να εξαλείψουν στοιχεία που είναι αμφισβητήσιμα, ή εξαλείφουν σημεία που ανήκουν κάτω από το 5 ή 10% της έντασης σε κάθε κανάλι (Causton, Quackenbush & Brazma 2003).

5.4.2. Καθορισμός κατώτατου και ανώτατου ορίου

Μια άλλη προσέγγιση είναι ο καθορισμός κατώτατου και ανώτατου ορίου, δηλαδή ελάχιστες και μέγιστες ανεκτές τιμές για τις εντάσεις. Τα δεδομένα που έχουν τιμές κάτω από το όριο προσαρμόζονται προς τα πάνω και τίθενται ίσες με το κατώτερο όριο και αυτές που βρίσκονται πάνω από το ανώτερο όριο εξισώνονται με τις τιμές του ανωτάτου ορίου. Χρησιμοποιούνται κυρίως για ανάλυση γονιδίων με χαμηλά επίπεδα έκφρασης, καθώς αυτές οι μετρήσεις είναι πιθανό να μην είναι ακριβείς. Είναι καλή πρακτική η επισήμανση των γονιδίων που είναι προσαρμοσμένα στις τιμές ανωτάτου και κατώτατου ορίου, καθώς αυτοί οι λόγοι έκφρασης είναι τεχνητά κατασκευασμένοι και δεν αντικατοπτρίζουν πειραματικές μετρήσεις. Ένας καλός τρόπος επιλογής των τιμών των ορίων είναι η χρήση μετρούμενων τιμών στις οποίες η σχέση μεταξύ των γονιδιακών εκφράσεων των δειγμάτων που συγκρίνονται, αρχίζουν να παρεκκλίνουν από τη γραμμικότητα (Causton, Quackenbush & Brazma 2003).

5.4.3. Εκτύπωση πολλαπλών κουκίδων (multiple spotting)

Η εκτύπωση πολλαπλών κουκίδων παρέχει χρήσιμες πληροφορίες για την ποιότητα των δεδομένων. Επιτρέπει το ποιοτικό φιλτράρισμα, δηλαδή την αφαίρεση των απομακρυσμένων σημείων και των αναξιόπιστων γονιδίων και τον προσδιορισμό των προβληματικών πλακιδίων (Causton, Quackenbush & Brazma 2003).

5.5. ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΗ ΜΑΘΗΣΗ

5.5.1. Μέτρηση της απόστασης (Αλγόριθμοι Ομαδοποίησης)

Οι αλγόριθμοι **ομαδοποίησης** χρησιμοποιούνται για την ταξινόμηση των δεδομένων και την ομαδοποίηση των γονιδίων βάση της διασποράς τους στο χώρο έκφρασης. Διάφορες τεχνικές ομαδοποίησης έχουν εφαρμοστεί για τον προσδιορισμό προτύπων στα δεδομένα γονιδιακής έκφρασης. Οι περισσότερες τεχνικές είναι **ιεραρχικές (hierarchical)**. Σε αυτές η ταξινόμηση που προκύπτει έχει έναν αυξανόμενο αριθμό «φωλιασμένων» κατηγοριών και το τελικό αποτέλεσμα μοιάζει με μια φυλογενετική ταξινόμηση. Υπάρχουν επίσης **μη-ιεραρχικές** τεχνικές ομαδοποίησης (flat), όπως η k-μέσων (k-means), οι οποίες απλά διανέμουν τα αντικείμενα σε ξεχωριστές ομάδες χωρίς προσπάθεια να διευκρινιστεί η σχέση μεταξύ των μεμονωμένων στοιχείων. Οι τεχνικές ομαδοποίησης ταξινομούνται επίσης σε **διαχωριστικές ή συσσωρευτικές**. Μια **διαχωριστική** μέθοδος ξεκινά θεωρώντας το σύνολο των δεδομένων ως μια μεγάλη ομάδα και προχωρά χωρίζοντας τα αντικείμενα σε όλο και μικρότερες ομάδες, ξεκινώντας από αυτές που είναι περισσότερο ανόμοιες. Οι **συσσωρευτικές** τεχνικές ξεκινούν με ομάδες ενός μέλους και σταδιακά με μια επαναληπτική διαδικασία συγχωνεύουν τις περισσότερο όμοιες, μέχρι να συμπεριληφθούν όλα τα δεδομένα. Τέλος, η ομαδοποίηση μπορεί να είναι είτε **επιβλεπόμενη** είτε **μη επιβλεπόμενη**. Οι επιβλεπόμενες μέθοδοι χρησιμοποιούν, για την καθοδήγηση των αλγορίθμων ομαδοποίησης, υπάρχουσες βιολογικές πληροφορίες σχετικά με συγκεκριμένα γονίδια που συσχετίζονται λειτουργικά (Causton, Quackenbush & Brazma 2003).

Πριν την επιλογή του αλγορίθμου ομαδοποίησης των δεδομένων πρέπει να καθοριστεί ο τρόπος **«μέτρησης της απόστασης»** που θα χρησιμοποιηθεί. Ο υπολογισμός της απόστασης μεταξύ δύο αντικειμένων είναι βασικός για την τοποθέτηση τους σε ομάδες. Η ανάλυση δεδομένων από μικροσυστοιχίες εντοπίζει τις ομάδες «παρόμοιων» γονιδίων, βρίσκοντας και ομαδοποιώντας αυτά που είναι σε μικρή απόσταση μεταξύ τους. Για να γίνει αυτό πρέπει να προσδιοριστούν οι αποστάσεις μεταξύ των διανυσμάτων γονιδιακής έκφρασης.

Κάθε πείραμα αντιπροσωπεύει ένα ξεχωριστό διακριτό άξονα στο χώρο και ο λογαριθμικός λόγος ενός γονιδίου στο συγκεκριμένο πείραμα αντιπροσωπεύει τη γεωμετρική συντεταγμένη του. Για παράδειγμα, αν έχουμε τρία πειράματα, ο λογαριθμικός λόγος ενός συγκεκριμένου γονιδίου στο πείραμα 1, αντιπροσωπεύει την x-συντεταγμένη, ο λόγος του στο πείραμα 2 την y-συντεταγμένη και ο λόγος του στο τρίτο, την z-συντεταγμένη. Έτσι όλες οι πληροφορίες σχετικά με το γονίδιο μπορούν να αντιπροσωπευτούν από ένα σημείο x-y-z στο χώρο έκφρασης. Ένα δεύτερο γονίδιο με τον ίδιο περίπου λογαριθμικό λόγο για κάθε πείραμα θα αντιπροσωπεύει ένα κοντινό σημείο (χωρικά), ενώ ένα γονίδιο με πολύ διαφορετικό πρότυπο έκφρασης θα βρίσκεται μακριά από το αρχικό γονίδιο. Η διαστατικότητα του χώρου έκφρασης αυξάνεται για να είναι ίση με τον αριθμό των πειραμάτων. Έτσι τα δεδομένα έκφρασης μπορούν να αντιπροσωπευτούν από ένα n-διάστατο χώρο έκφρασης, όπου n είναι ο αριθμός των πειραμάτων και κάθε διάνυσμα γονιδιακής έκφρασης αντιπροσωπεύεται από ένα σημείο στο χώρο αυτό.

Για να ερμηνεύσουμε τα αποτελέσματα από την ανάλυση πολλαπλών πειραμάτων δημιουργούμε μια «**μήτρα εκφράσεων**» στην οποία η κάθε στήλη αντιπροσωπεύει ένα πείραμα και η κάθε γραμμή το διάνυσμα έκφρασης για ένα συγκεκριμένο γονίδιο (Quackenbush 2001, Eisen et al. 1998).

Υπάρχουν διάφορες συναρτήσεις υπολογισμού της απόστασης (distance metrics) που ανήκουν σε δύο κατηγορίες: τις μετρικές και τις ημιμετρικές (metrics και semi metrics).

Για να ταξινομηθεί ως μετρική, μια συνάρτηση απόστασης d_{ij} μεταξύ δύο διανυσμάτων, i και j , πρέπει να υπακούει στους παρακάτω κανόνες:

- Η απόσταση πρέπει να είναι θετική καθορισμένη, $d_{ij} \geq 0$ (δηλαδή πρέπει να είναι μηδέν ή θετική).
- Η απόσταση πρέπει να είναι συμμετρική, $d_{ij} = d_{ji}$, έτσι ώστε η απόσταση από το i στο j να είναι ίδια με την απόσταση από το j στο i .
- Ένα αντικείμενο έχει μηδενική απόσταση από τον εαυτό του, $d_{ii} = 0$.
- Κατά την εξέταση τριών αντικειμένων, i , j και k , η απόσταση από το i στο k είναι πάντα μικρότερη ή ίση από το άθροισμα των αποστάσεων από το i στο j και από το j στο k , δηλαδή $d_{ik} \leq d_{ij} + d_{jk}$. Αυτή η σχέση αποτελεί την τριγωνική ιδιότητα.

Η πιο συνηθισμένη είναι η **ευκλείδεια απόσταση**, που αποτελεί γενίκευση του Πυθαγορείου θεωρήματος. Για ένα n -διάστατο διάνυσμα έκφρασης, η ευκλείδεια απόσταση δίνεται από τη σχέση:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

όπου x_i και y_i είναι οι μετρούμενες τιμές έκφρασης για τα γονίδια x και y αντίστοιχα, στο πείραμα i , και η άθροιση γίνεται για τα n πειράματα που αναλύονται.

Τα μέτρα απόστασης που υπακούν στους πρώτους τρεις κανόνες, αλλά αποτυγχάνουν να εξασφαλίσουν την τριγωνική ιδιότητα αναφέρονται ως ημι-μετρικά. Υπάρχει ένας μεγάλος αριθμός ημι-μετρικών μέτρων απόστασης και αυτά χρησιμοποιούνται συχνά στην ανάλυση έκφρασης (Quackenbush 2001).

5.5.2. Ιεραρχική ομαδοποίηση (Hierarchical clustering)

Η **ιεραρχική ομαδοποίηση** έχει το πλεονέκτημα ότι είναι απλή και το αποτέλεσμα μπορεί εύκολα να οπτικοποιηθεί. Η ιεραρχική ομαδοποίηση είναι μια συσσωρευτική προσέγγιση στην οποία απλά προφίλ έκφρασης ενώνονται για να διαμορφώσουν τις ομάδες, στις οποίες προσχωρούν ακόμα περισσότερα γονίδια μέχρι να ολοκληρωθεί η διαδικασία. Δημιουργούν ένα ιεραρχικό δέντρο, που ονομάζεται δενδροδιάγραμμα και το μήκος των κλαδιών του αντιπροσωπεύουν το βαθμό ομοιότητας των δεδομένων.

Η διαδικασία της ιεραρχικής ομαδοποίησης προχωρά με απλό τρόπο.

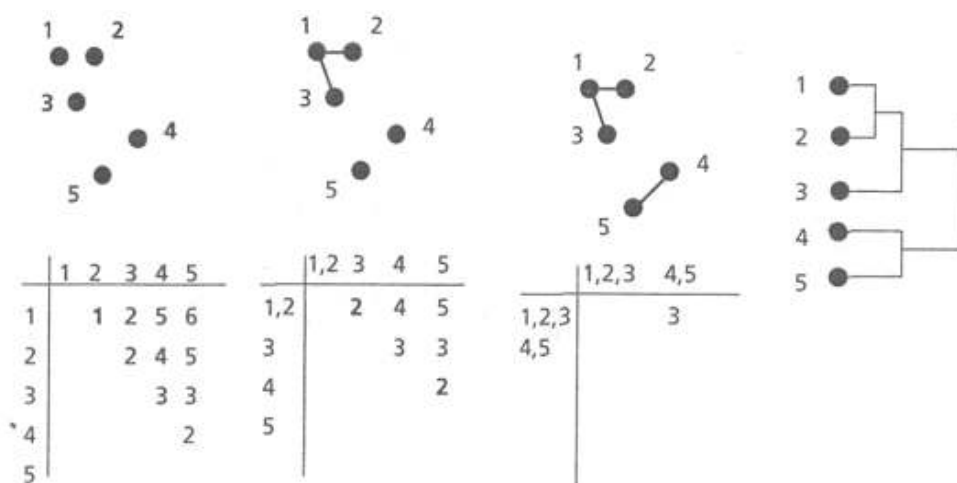
Αρχικά, υπολογίζεται η μήτρα των αποστάσεων κατά ζευγάρια (pairwise) για όλα τα γονίδια που πρέπει να ομαδοποιηθούν.

Δεύτερον, η μήτρα των αποστάσεων διατρέχεται για να βρεθούν τα δύο πιο όμοια γονίδια (ή ομάδες). Αρχικά κάθε ομάδα αποτελείται από ένα μόνο γονίδιο. Εάν διάφορα ζευγάρια έχουν την ίδια απόσταση, ένας προκαθορισμένος κανόνας χρησιμοποιείται για να αποφασίσει μεταξύ των εναλλακτικών λύσεων.

Τρίτον, οι δύο επιλεγμένες ομάδες συγχωνεύονται για να δημιουργήσουν μια νέα ομάδα, που τώρα περιέχει τουλάχιστον δύο αντικείμενα.

Τέταρτο, υπολογίζονται οι αποστάσεις μεταξύ αυτής της νέας ομάδας και των άλλων ομάδων. Δεν χρειάζεται να υπολογιστούν εκ νέου όλες οι αποστάσεις, δεδομένου ότι μόνο αυτές που περιλαμβάνουν τη νέα ομάδα έχουν αλλάξει.

Στο τέλος, τα βήματα 2-4 επαναλαμβάνονται μέχρι όλα τα αντικείμενα εκχωρηθούν σε μια ομάδα.



Εικόνα 5.1: Τα βήματα της ιεραρχικής ομαδοποίησης. Τα αντικείμενα με την ελάχιστη απόσταση μεταξύ τους, ομαδοποιούνται για να σχηματίσουν μια νέα ομάδα. Η απόσταση μεταξύ όλων των ομάδων ή αντικειμένων επαναυπολογίζονται με μια επαναληπτική διαδικασία (Causton, Quackenbush & Brazma 2003).

Υπάρχουν διάφορες παραλλαγές της ιεραρχικής ομαδοποίησης που διαφέρουν στον τρόπο με τον οποίο υπολογίζονται οι αποστάσεις μεταξύ των αυξανόμενων ομάδων και των υπολοίπων μελών του συνόλου δεδομένων, συμπεριλαμβανομένων των άλλων ομάδων. Κάθε μια από αυτές παραγάγει ελαφρώς διαφορετικά αποτελέσματα (Quackenbush 2001, Causton, Quackenbush & Brazma 2003).

5.5.2.1. Ομαδοποίηση απλής σύνδεσης (single linkage clustering)

Η απόσταση μεταξύ δύο ομάδων, i και j , υπολογίζεται ως η ελάχιστη απόσταση μεταξύ ενός μέλους της ομάδας i και ενός μέλους της ομάδας j . Αυτή η τεχνική αναφέρεται επίσης ως μέθοδος ελάχιστης απόστασης, ή «κοντινότερου γείτονα».

Αυτή η μέθοδος τείνει να παραγάγει ομάδες που είναι «χαλαρές», επειδή οι ομάδες μπορούν να ενωθούν αν οποιαδήποτε δύο μέλη είναι κοντά. Ειδικά, η μέθοδος αυτή οδηγεί συχνά σε αλυσιδωτή, ή διαδοχική προσθήκη των απλών δειγμάτων σε μια ήδη υπάρχουσα ομάδα.

5.5.2.2.Ομαδοποίηση πλήρους σύνδεσης (complete linkage clustering)

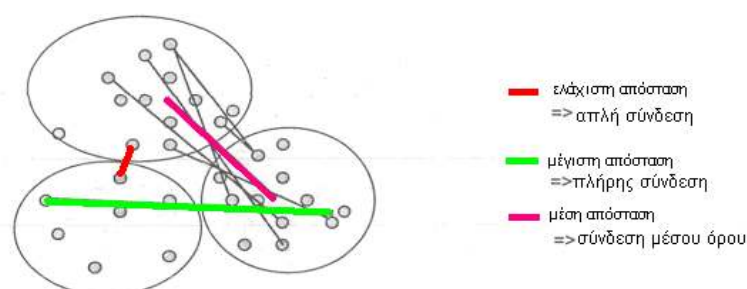
Η μέθοδος αυτή είναι επίσης γνωστή και ως μέθοδος μέγιστης απόστασης ή «μακρύτερου γείτονα». Η απόσταση μεταξύ δύο ομάδων υπολογίζεται ως η μέγιστη απόσταση μεταξύ των μελών των σχετικών ομάδων. Έτσι δημιουργούνται πιο συμπαγείς ομάδες και είναι συχνά παρόμοιες σε μέγεθος. Η μέθοδος αυτή είναι ευαίσθητη στις ακραίες τιμές.

5.5.2.3.Ομαδοποίηση μέσου όρου σύνδεσης (average linkage clustering)

Η απόσταση μεταξύ των ομάδων υπολογίζεται χρησιμοποιώντας τις μέσες τιμές. Υπάρχουν διάφορες μέθοδοι υπολογισμού του μέσου όρου. Η πιο κοινή είναι η μέθοδος ομάδων ζευγαριών μη σταθμισμένων μέσων (UPGMA- unweighted pair-group method average). Ο μέσος όρος της απόστασης υπολογίζεται ως η απόσταση μεταξύ κάθε σημείου της ομάδας από όλα τα σημεία μιας άλλης ομάδας. Οι δύο ομάδες με τη μικρότερη μέση απόσταση ενώνονται μαζί για να διαμορφώσουν μια νέα ομάδα. Σχετικές μέθοδοι αντικαθιστούν το μέσο όρο με το κεντροειδές (centroid).

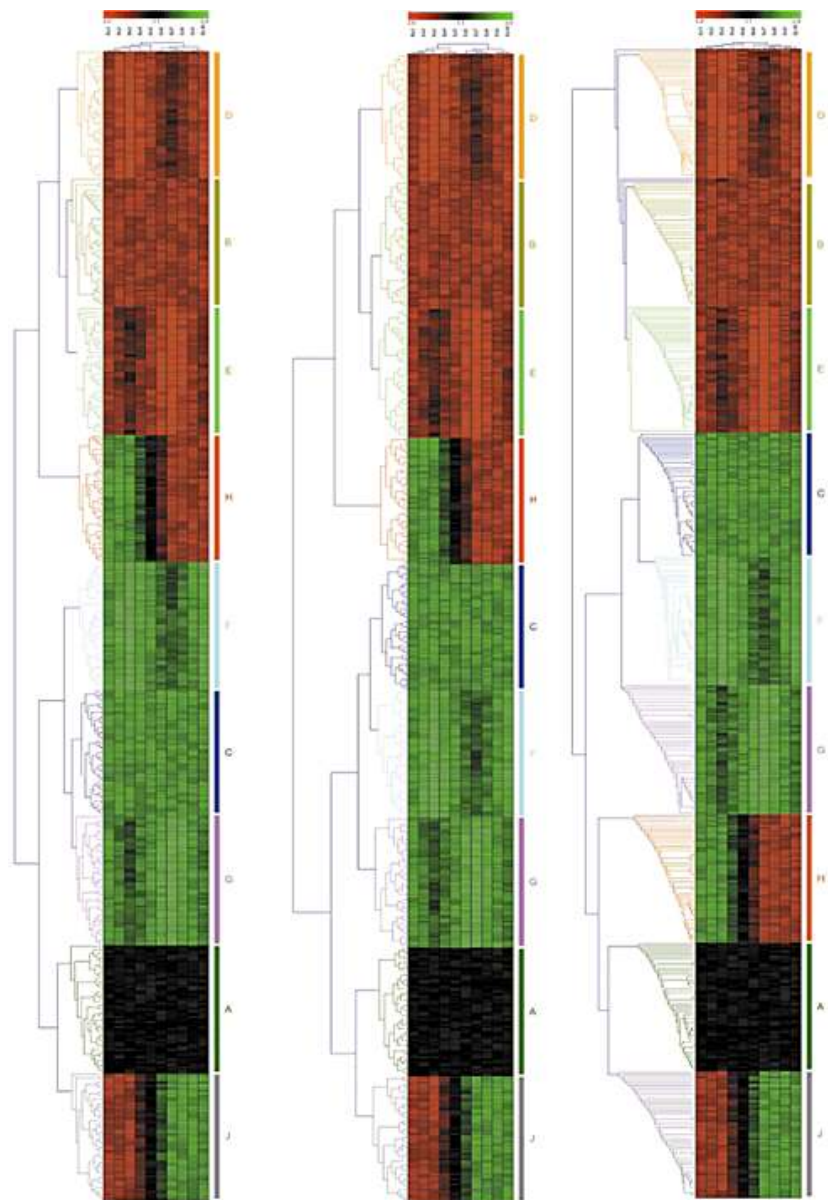
5.5.2.4.Μέθοδος Ward.

Στη μέθοδο Ward, γνωστή επίσης και ως ομαδοποίηση ελάχιστης διασποράς, η συμμετοχή σε μια ομάδα καθορίζεται από τον υπολογισμό του συνολικού άθροισματος της τετραγωνικής απόκλισης από το μέσο όρο μιας ομάδας. Η ένωση σε ομάδες γίνεται με τέτοιο τρόπο ώστε να παράγει τη μικρότερη δυνατή αύξηση στο άθροισμα των τετραγωνικών σφαλμάτων. Η μέθοδος αυτή ευνοεί τη δημιουργία μικρών ομάδων, ίδιου μεγέθους, με την ελάχιστη διασπορά μέσα στην ομάδα, ωστόσο είναι υπολογιστικά απαιτητική.



Εικόνα 5.2:Απλή σύνδεση, πλήρης σύνδεση και σύνδεση μέσου όρου (Causton, Quackenbush & Brazma 2003).

Ένα πιθανό πρόβλημα στις περισσότερες ιεραρχικές μεθόδους ομαδοποίησης είναι ότι, καθώς οι ομάδες αυξάνονται σε μέγεθος, το διάλυμα έκφρασης που αντιπροσωπεύει την ομάδα μπορεί πλέον να μην αντιπροσωπεύσει κανένα από τα γονίδια της. Κατά συνέπεια, καθώς η ομαδοποίηση προχωρά, τα πραγματικά πρότυπα γονιδιακής έκφρασης γίνονται όλο και λιγότερο σχετικά. Επιπλέον, αν γίνει μια «κακή» ανάθεση χωρίς στη διαδικασία, δεν μπορεί να διορθωθεί. Μια εναλλακτική που μπορεί να περιορίσει τέτοιες ατέλειες, είναι η χρήση προσεγγίσεων ομαδοποίησης, όπως η k-μέσων και η SOMs για τη διανομή των δεδομένων σε ομάδες που έχουν παρόμοια πρότυπα έκφρασης (Quackenbush 2001, Grant 2004, Causton, Quackenbush & Brazma 2003).



Εικόνα 5.3: Παράδειγμα εφαρμογής Ιεραρχικής ομαδοποίησης σε δεδομένα (Quackenbush 2001).

5.5.3. Ομαδοποίηση k-μέσων (k-means clustering)

Αν υπάρχει εκ των προτέρων γνώση για τον αριθμό των ομάδων που αντιπροσωπεύουν τα δεδομένα, η μέθοδος των k-μέσων είναι μια καλή εναλλακτική λύση σε σχέση με τις ιεραρχικές μεθόδους. Στη μέθοδο k-μέσων, τα αντικείμενα χωρίζονται σε ένα καθορισμένο αριθμό (k) ομάδων, έτσι ώστε οι ομάδες να είναι εσωτερικά όμοιες αλλά εξωτερικά ανόμοιες (δεν παράγεται δενδροδιάγραμμα).

Η διαδικασία που περιλαμβάνεται στην ομαδοποίηση k-μέσων είναι εννοιολογικά απλή, αλλά μπορεί να είναι υπολογιστικά απαιτητική.

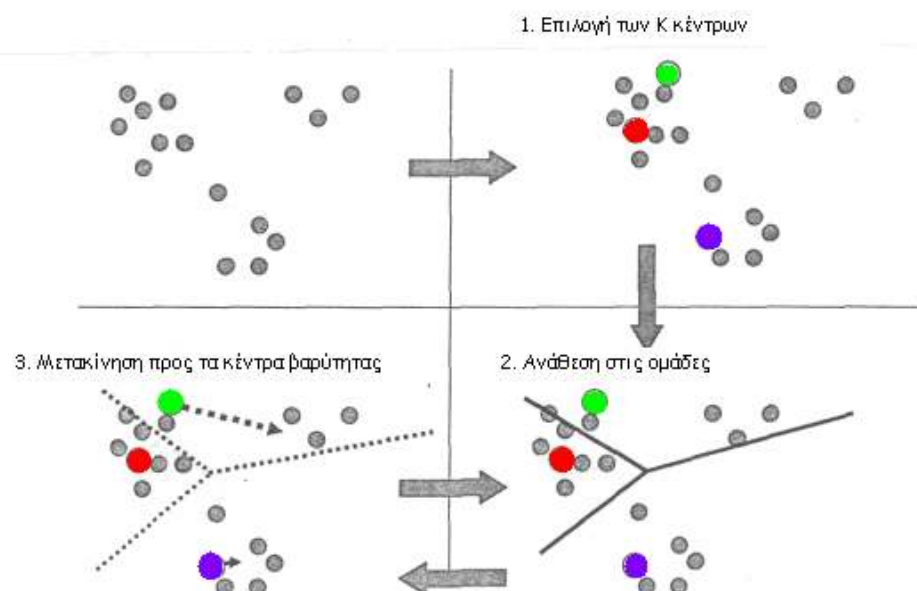
Αρχικά, όλα τα αρχικά αντικείμενα αποδίδονται τυχαία σε μια από τις k ομάδες (όπου το k διευκρινίζεται από το χρήστη).

Δεύτερον, ένα μέσο διάνυσμα έκφρασης υπολογίζεται για κάθε ομάδα και αυτό χρησιμοποιείται για να υπολογιστούν οι αποστάσεις μεταξύ των ομάδων.

Τρίτον, χρησιμοποιώντας μια επαναληπτική μέθοδο, τα αντικείμενα μετακινούνται μεταξύ των ομάδων και οι αποστάσεις μέσα και έξω από τις ομάδες υπολογίζονται σε κάθε μετακίνηση. Τα αντικείμενα επιτρέπεται να παραμείνουν στη νέα ομάδα μόνο εάν είναι πιο κοντά σε αυτή απ' ό,τι στην προηγούμενη ομάδα τους.

Τέταρτον, μετά από κάθε μετακίνηση, το διάνυσμα έκφρασης για κάθε ομάδα υπολογίζεται εκ νέου.

Στο τέλος, η μετάθεση προχωρά μέχρις ότου η μετακίνηση άλλων αντικειμένων, θα έκανε τις ομάδες πιο μεταβλητές, αυξάνοντας τις αποστάσεις στο εσωτερικό τους και μειώνοντας τη μεταξύ τους ανομοιότητα. Κάποιες εφαρμογές της μεθόδου k-μέσων χρειάζονται όχι μόνο τον αριθμό των ομάδων, αλλά και τους κόμβους για κάθε ομάδα. Αυτό επιτρέπει τη χρήση προηγούμενων γνώσεων του συστήματος για την πραγματοποίηση της ομαδοποίησης (Lee 2004, Quackenbush 2001, Causton, Quackenbush & Brazma 2003).



Εικόνα 5.4: Ομαδοποίηση K -μέσων. Οι μεγάλοι κύκλοι αντιπροσωπεύουν τα κέντρα μάζας και οι μικροί τα αντικείμενα (Causton, Quackenbush & Brazma 2003).

Η ομαδοποίηση k-means, όπως και άλλοι αλγόριθμοι, έχει αδυναμίες:

- Όταν ο αριθμός των δεδομένων δεν είναι πολύ μεγάλος, η αρχική ομαδοποίηση καθορίζει σημαντικά τις ομάδες.
- Ο αριθμός των ομάδων, k, πρέπει να καθοριστεί εκ των προτέρων.
- Δεν γνωρίζουμε ποτέ την πραγματική ομάδα, αφού χρησιμοποιώντας τα ίδια δεδομένα, εάν αυτά εισαχθούν με διαφορετική σειρά, μπορεί για μικρό αριθμό δεδομένων, να παραχθούν διαφορετικές ομάδες.
- Είναι ευαίσθητη στις αρχικές συνθήκες. Διαφορετικές αρχικές συνθήκες μπορεί να παράγουν διαφορετικά αποτελέσματα ομαδοποίησης. Ο αλγόριθμος μπορεί να παγιδευτεί στο τοπικό βέλτιστο.
- Δεν ξέρουμε ποτέ ποια ιδιότητα συμβάλλει περισσότερο στη διαδικασία ομαδοποίησης, αφού υποθέτουμε ότι κάθε ιδιότητα έχει το ίδιο βάρος.
- Το γεγονός ότι ο αριθμητικός μέσος δεν είναι εύρωστος στις έκτροπες τιμές έχει σαν αποτέλεσμα στοιχεία που βρίσκονται μακριά από το κεντροειδές να μπορούν να μετατοπίσουν τα κεντροειδή μακριά από τα πραγματικά. Χρησιμοποιώντας αντί του μέσου την διάμεσο, το πρόβλημα αυτό μπορεί να αντιμετωπισθεί.

5.5.4. Αυτο-οργανούμενοι χάρτες (Self-organizing maps- SOM).

Ένας αυτο-οργανούμενος χάρτης (SOM) είναι μια διαχωριστική μέθοδος ομαδοποίησης που βασίζεται στα νευρωνικά δίκτυα. Αυτή η μέθοδος αναθέτει τα γονίδια σε μια σειρά ομάδων βάσει της ομοιότητας του διανύσματος έκφρασής τους, με **το διάνυσμα αναφοράς** που είναι ορισμένο για κάθε ομάδα. Αυτό που διαχωρίζει την ομαδοποίηση SOM από την ομαδοποίηση k-means, είναι η διαδικασία προσδιορισμού αυτών των διανυσμάτων αναφοράς. Πριν ξεκινήσει η ανάλυση, ο χρήστης καθορίζει το γεωμετρικό σχηματισμό των ομάδων, συνήθως ένα δισδιάστατο ορθογώνιο ή εξαγωνικό πλέγμα. Τυχαία διανύσματα δημιουργούνται για κάθε ομάδα, αλλά πριν τα γονίδια ξεκινήσουν να αποδίδονται στις ομάδες, τα διανύσματα «εκπαιδεύονται» με μια επαναληπτική διαδικασία που συνεχίζεται μέχρι να συγκλίνει, έτσι ώστε τα δεδομένα να διαχωριστούν αποτελεσματικότερα.

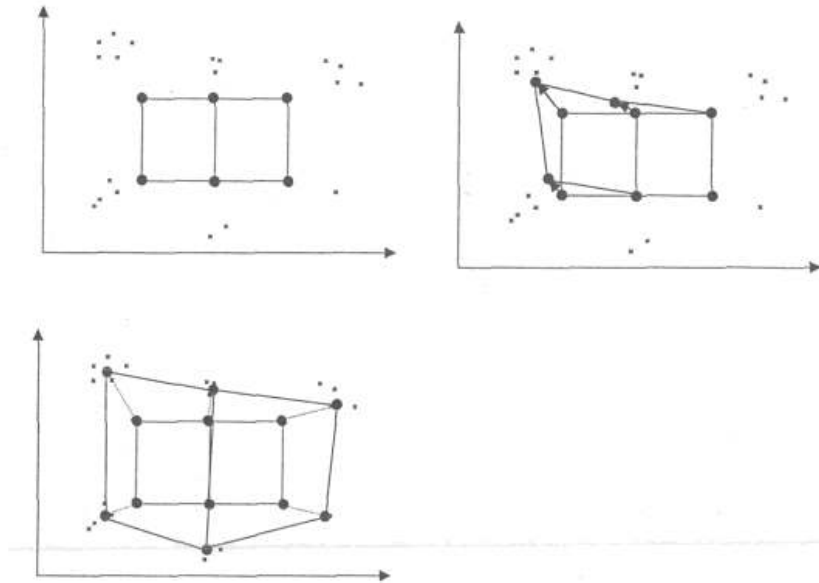
Αρχικά, τυχαία διανύσματα κατασκευάζονται και αποδίδονται σε κάθε διαμέριση.

Δεύτερον, ένα γονίδιο επιλέγεται τυχαία και χρησιμοποιώντας ένα επιλεγμένο μέτρο απόστασης, προσδιορίζεται το διάνυσμα αναφοράς που είναι πιο κοντά στο γονίδιο.

Τρίτον, το διάνυσμα αναφοράς προσαρμόζεται έτσι ώστε να είναι περισσότερο όμοιο με το διάνυσμα του συγκεκριμένου γονιδίου. Τα διανύσματα αναφοράς που είναι «γειτονικά» στο δισδιάστατο πλέγμα του συγκεκριμένου γονιδίου προσαρμόζονται επίσης, ώστε να είναι και αυτά περισσότερο όμοια με το διάνυσμα έκφρασης του τυχαία επιλεγμένου γονιδίου.

Τέταρτο, τα βήματα 2 και 3 επαναλαμβάνονται αρκετές χιλιάδες φορές, μειώνοντας την ποσότητα που τα διανύσματα αναφοράς προσαρμόζονται και αυξάνοντας την αυστηρότητα που χρησιμοποιείται για τον προσδιορισμό της «εγγύτητας» σε κάθε βήμα. Καθώς η διαδικασία συνεχίζεται, τα διανύσματα αναφοράς συγκλίνουν σε καθορισμένες τιμές.

Στο τέλος, τα γονίδια χαρτογραφούνται στις σχετικές ομάδες, ανάλογα με το διάνυσμα αναφοράς με το οποίο είναι πιο όμοια.



Εικόνα 5.5: Αυτο-οργανούμενοι χάρτες. Οι αρχικές θέσεις των διανυσμάτων αναφοράς, περιγράφονται από τους σκιασμένους κύκλους (Causton, Quackenbush & Brazma 2003).

Για την επιλογή του γεωμετρικού σχηματισμού των ομάδων είναι αποτελεσματικό να διευκρινιστεί ο αριθμός των ομάδων στα οποία θα διαιρεθούν τα δεδομένα. Η μέθοδος αυτή δουλεύει καλά με θορυβώδη δεδομένα (Quackenbush 2001, Causton, Quackenbush & Brazma 2003).

5.5.5. Ανάλυση Πρωτεουσών Συνιστωσών (principal components analysis)

Τα δεδομένα γονιδιακής έκφρασης έχουν πολύ υψηλές διαστάσεις. Όταν αναλύουμε τέτοιου τύπου δεδομένα, είναι δύσκολο να σχηματιστούν οι ομάδες. Επίσης σε δεδομένα γονιδιακής έκφρασης με χρήση μικροσυστοιχιών, σε αντίθεση με τα κλασσικά σύνολα έχουμε πολύ μεγαλύτερο αριθμό μεταβλητών σε σχέση με τον αριθμό των παρατηρήσεων (Jolliffe 2002).

Η **ανάλυση πρωτεουσών συνιστωσών (PCA)** είναι μια μέθοδος που χρησιμοποιείται στην ανάλυση δεδομένων γονιδιακής έκφρασης, κυρίως για να μειώσει τη «διαστατικότητα» των δεδομένων και να ανακαλύψει συνδυασμούς πειραμάτων ή γονιδίων που από κοινού συνεισφέρουν περισσότερο στη διακύμανση των δεδομένων.

Η κεντρική ιδέα της ανάλυσης πρωτεουσών συνιστωσών (PCA) είναι να μειωθεί η διαστατικότητα του συνόλου δεδομένων που αποτελείται από έναν μεγάλο αριθμό αλληλένδετων μεταβλητών, διατηρώντας όσο το δυνατόν περισσότερο τη διακύμανση που εμφανίζεται στο σύνολο των δεδομένων.

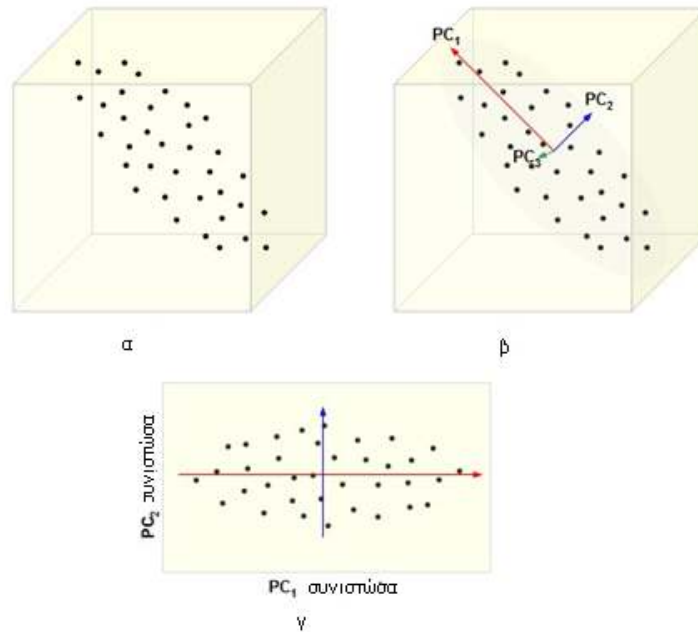
Αυτό επιτυγχάνεται με το μετασχηματισμό σε ένα νέο σύνολο μεταβλητών, τις πρωτεύουσες συνιστώσες (PCs), οι οποίες είναι ασυσχέτιστες και ταξινομούνται έτσι ώστε οι πρώτες να διατηρούν το μεγαλύτερο μέρος της διακύμανσης που παρουσιάζεται στις αρχικές μεταβλητές. Κάθε πρωτεύουσα συνιστώσα είναι ένας γραμμικός συνδυασμός των αρχικών μεταβλητών. Όλες οι πρωτεύουσες συνιστώσες είναι ορθογώνιες μεταξύ τους και έτσι δεν υπάρχει καμία περιττή πληροφορία. Οι πρωτεύουσες συνιστώσες σαν σύνολο, αποτελούν μια ορθογώνια βάση στο χώρο των δεδομένων.

Η πρώτη πρωτεύουσα συνιστώσα είναι ένας ενιαίος άξονας στο χώρο. Όταν προβάλλετε κάθε παρατήρηση σε αυτό τον άξονα, οι προκύπτουσες τιμές διαμορφώνουν μια νέα μεταβλητή. Και η διακύμανση αυτής της μεταβλητής είναι η μέγιστη μεταξύ όλων των πιθανών επιλογών του πρώτου άξονα. Η δεύτερη πρωτεύουσα συνιστώσα είναι ένας άλλος άξονας στο χώρο, κάθετος στον πρώτο. Η προβολή των παρατηρήσεων σε αυτόν τον άξονα παράγει μια άλλη νέα μεταβλητή. Η διακύμανση αυτής της μεταβλητής είναι η μέγιστη μεταξύ όλων των πιθανών επιλογών αυτού του δεύτερου άξονα.

Το πλήρες σύνολο των κύριων συνιστωσών είναι τόσο μεγάλο όσο το αρχικό σύνολο μεταβλητών. Αλλά το άθροισμα της διακύμανσης των πρώτων πρωτευουσών συνιστωσών συνήθως υπερβαίνει το 80% της συνολικής διακύμανσης των αρχικών δεδομένων. Με την εξέταση των γραφημάτων αυτών των λίγων νέων μεταβλητών, οι ερευνητές αναπτύσσουν συχνά μια βαθύτερη κατανόηση των κινητήριων δυνάμεων που παρήγαγαν τα αρχικά δεδομένα (Quackenbush 2001, Causton, Quackenbush & Brazma 2003).

Η μέθοδος ανάλυσης πρωτευουσών συνιστωσών επιτρέπει την οπτική εκτίμηση του αριθμού ομάδων που αντιπροσωπεύουν τα δεδομένα. Έτσι μπορεί να χρησιμοποιηθεί για να διευκρινίσει τον αριθμό k των ομάδων και για να ομαδοποιήσει τα γονίδια (ή τα πειράματα) στις σχετικές ομάδες.

Η μέθοδος PCA (επίσης αποκαλούμενη ανάλυση ιδιζουσών τιμών-singular value decomposition) είναι μια μαθηματική τεχνική που παρέχει μια προβολή των σύνθετων συνόλων δεδομένων σε ένα μειωμένο, εύκολα απεικονίσιμο χώρο. Η PCA βρίσκει εκείνες τις όψεις που δίνουν τον καλύτερο χωρισμό των δεδομένων.



Εικόνα 5.6: Απεικόνιση της PCA ανάλυσης. α. Ένα σύνολο δεδομένων στον τρισδιάστατο χώρο, β. Οι τρεις ορθογώνιες πρωτεύουσες συνιστώσες (PCs) των δεδομένων, βάση της διακύμανσης, γ. Η προβολή του συνόλου δεδομένων στις δύο πρώτες συνιστώσες, απορρίπτοντας την τρίτη (Kavtaki 2007).

Στις περισσότερες εφαρμογές της PCA, είναι δύσκολο να καθοριστούν με ακρίβεια τα ακριβή όρια των διακριτών ομάδων, ή να καθοριστούν τα γονίδια (ή τα πειράματα) που ανήκουν σε κάθε ομάδα. Ωστόσο, η PCA είναι μια ισχυρή τεχνική για την ανάλυση των δεδομένων γονιδιακής έκφρασης όταν χρησιμοποιείται με μια άλλη τεχνική ομαδοποίησης, όπως η k-μέσων ή η SOM, που απαιτούν να ορίσει ο χρήστης τον αριθμό των ομάδων (Lee 2004, Causton, Quackenbush & Brazma 2003, Raychaudhuri, Stuart & Altman 2000).

5.6. ΜΕΘΟΔΟΙ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ

Οι μέθοδοι επιβλεπόμενης ταξινόμησης αντιπροσωπεύουν μια ισχυρή εναλλακτική λύση που μπορεί να εφαρμοστεί, αν υπάρχει προηγούμενη γνώση σχετικά με τα γονίδια που αναμένεται να συγκεντρωθούν από κοινού. Οι πιο απλές μέθοδοι είναι η μέθοδος γραμμικών διαχωριστών και η μέθοδος k-κοντινότερων γειτόνων (k-nearest neighbors). Μια ευρέως χρησιμοποιημένη μέθοδος είναι οι μηχανές υποστήριξης διανυσμάτων (SVM) (Brown et al. 2000). Τυπικά αυτοί οι «ταξινομητές» εκπαιδεύονται με ένα υποσύνολο δεδομένων, που έχουν μια εκ των προτέρων γνωστή ταξινόμηση (σύνολο εκπαίδευσης) και ελέγχονται ως προς την ακρίβειά τους από ένα άλλο υποσύνολο γνωστής ταξινόμησης (σύνολο ελέγχου). Μετά την αξιολόγηση της ποιότητας της πρόβλεψης, ο ταξινομητής μπορεί να εφαρμοσθεί σε νέα δεδομένα (Quackenbush 2001, Causton, Quackenbush & Brazma 2003).

5.6.1. Γραμμικοί διαχωριστές (linear discriminants)

Οι γραμμικοί διαχωριστές, σε ένα δισδιάστατο χώρο αποτελούν μια ευθεία γραμμή που διαχωρίζει τα σημεία που είναι σημασμένα με την ετικέτα 1, από εκείνα που έχουν την ετικέτα -1. Στις δύο διαστάσεις η ευθεία αυτή δίνεται από τη σχέση $x_2 = ax_1 + b$ και για την εύρεση της χρησιμοποιείται η μέθοδος ελαχίστων τετραγώνων (Causton, Quackenbush & Brazma 2003).

5.6.2. Μηχανές υποστήριξης διανυσμάτων (Support vector machines – SVM)

Οι SVM χρησιμοποιούν ένα σύνολο εκπαίδευσης, στο οποίο, όταν τα γονίδια είναι γνωστό ότι συνδέονται μεταξύ τους, παρέχονται σε θετικά παραδείγματα και τα γονίδια που είναι γνωστό ότι δεν είναι μέλη αυτής της τάξης, αποτελούν αρνητικά παραδείγματα. Αυτά συνδυάζονται σε ένα σύνολο παραδειγμάτων εκπαίδευσης που χρησιμοποιείται από τις SVM για να «εκπαιδευτούν», ώστε να μπορούν να διακρίνουν μεταξύ των μελών και των μη μελών της κατηγορίας, βάσει των δεδομένων έκφρασης. Έχοντας «μάθει» τα χαρακτηριστικά γνωρίσματα έκφρασης κάθε κατηγορίας, οι SVM μπορούν έπειτα να χρησιμοποιηθούν για να αναγνωρίσουν και να ταξινομήσουν τα γονίδια του συνόλου δεδομένων βάσει της έκφρασής τους. Με αυτόν τον τρόπο, οι SVMs χρησιμοποιούν βιολογικές πληροφορίες για να καθορίσουν τα χαρακτηριστικά γνωρίσματα έκφρασης, που είναι χαρακτηριστικά μιας ομάδας και τοποθετούν τα γονίδια στη συγκεκριμένη ομάδα. Οι SVMs μπορούν επίσης να προσδιορίσουν γονίδια του συνόλου εκπαίδευσης που έχουν ακραίες τιμές ή που έχουν τοποθετηθεί προηγουμένως σε λάθος κατηγορία.

Τα δεδομένα γονιδιακής έκφρασης μπορούν να θεωρηθούν σαν ένας m-διάστατος χώρος, όπου τα διανύσματα έκφρασης αντιπροσωπεύονται ως σημεία σε αυτό το χώρο. Οι SVMs είναι ένας δυαδικός ταξινομητής (binary classifier) που προσπαθεί να χωρίσει τα γονίδια σε δύο κατηγορίες (στο θετικό σύνολο εκπαίδευσης, ή έξω από αυτό) με τον καθορισμό ενός βέλτιστου υπερεπιπέδου που χωρίζει τα μέλη της κατηγορίας από τα μη μέλη. Οι SVM λύνουν το πρόβλημα της χαρτογράφησης των διανυσμάτων γονιδιακής έκφρασης από τον αρχικό διανυσματικό χώρο έκφρασης σε ένα υψηλότερων διαστάσεων «χώρο χαρακτηριστικών», στον οποίο η απόσταση μετριέται χρησιμοποιώντας μια μαθηματική συνάρτηση, γνωστή ως συνάρτηση kernel και στη συνέχεια τα δεδομένα μπορούν να χωριστούν σε δύο κατηγορίες.

Για μερικά σύνολα δεδομένων, οι SVMs μπορεί να μην επιτύχουν σαφή διαχωρισμό, είτε λόγω λαθών στην ταξινόμηση του συνόλου εκπαίδευσης, είτε λόγω θορύβου στα δεδομένα, είτε από μια εσφαλμένη επιλογή της συνάρτησης kernel. Για αυτό το λόγο, οι περισσότερες εφαρμογές επιτρέπουν στους χρήστες να διευκρινίσουν ένα «μαλακό» περιθώριο, που επιτρέπει σε μερικά εκπαιδευτικά παραδείγματα να «πέσουν» σε λάθος πλευρά του διαχωριστικού υπερεπιπέδου.

Ο πλήρης προσδιορισμός των SVMs απαιτεί συνεπώς τον καθορισμό τόσο της συνάρτησης kernel όσο και το μέγεθος της ρήτρας που πρέπει να επιβληθεί για την παραβίαση του μαλακού περιθωρίου. Είναι συχνά δύσκολη η επιλογή της καλύτερης συνάρτησης kernel, των παραμέτρων και των ρητρών. Διαφορετικές παράμετροι

παράγουν συχνά διαφορετικές ταξινομήσεις. Είναι επομένως συχνά απαραίτητο να αυξηθεί διαδοχικά η πολυπλοκότητα της συνάρτησης kernel μέχρι να επιτευχθεί η κατάλληλη ταξινόμηση (Brown et al. 2000).

Οι SVMs είναι ένας από τους επιβλεπόμενους αλγορίθμους ταξινόμησης που έχει εφαρμοστεί για τον προσδιορισμό των γονιδίων που μοιράζονται κοινά πρότυπα έκφρασης, αλλά κυρίως χρησιμοποιούνται για την ταξινόμηση δειγμάτων. Αν μετράται η γονιδιακή έκφραση μορίων RNA που προέρχεται από διάφορους ασθενείς, για τα οποία υπάρχει για παράδειγμα, ταξινόμηση με βάση τη φάση της ασθένειας ή δεδομένων επιβίωσης, μπορούμε να χρησιμοποιήσουμε τα δεδομένα των μικροσυστοιχιών για να εκπαιδύσουμε έναν αλγόριθμο που μπορεί έπειτα να εφαρμοστεί στην ταξινόμηση άλλων προηγούμενων μη ταξινομημένων δειγμάτων (Furey et al. 2000, Golub et al. 1999). Αυτή η προσέγγιση θα μπορούσε να χρησιμοποιηθεί για την ανάπτυξη «μοριακών αποτυπωμάτων έκφρασης» για την ταξινόμηση ασθενειών. Στη διάγνωση του καρκίνου, η δυνατότητα να παραχθεί ένα μοριακό αποτύπωμα της έκφρασης για κάθε όγκο, θα μπορούσε να αποδειχθεί εξαιρετικά σημαντική καθώς ιστολογικά παρόμοιοι όγκοι μπορεί στην πραγματικότητα να είναι αποτέλεσμα ουσιαστικών διαφορετικών γενετικών αλλαγών, οι οποίες μπορούν να επηρεάσουν σημαντικά την εξέλιξη του όγκου και την απόκριση στη θεραπεία.

5.6.3. Μέθοδος k-κοντινότερων γειτόνων (k-nearest neighbors)

Η μέθοδος γραμμικών διαχωριστών υποθέτει ότι τα σημασμένα σημεία μπορούν να διαχωριστούν από ένα υπερεπίπεδο. Ένας παραβολικός διαχωριστής μπορεί να επιτύχει καλύτερα αποτελέσματα, ενώ οι SVMs είναι πιο κατάλληλη μέθοδος. Μία απλή εναλλακτική των SVMs είναι η μέθοδος k-κοντινότερων γειτόνων.

Σε αυτή τη μέθοδο, η ετικέτα ενός μη εκ των προτέρων σημασμένου σημείου, εκτιμάται από τις ετικέτες των k κοντινότερων σημείων. Για τη μέτρηση της απόστασης χρησιμοποιείται ένα καθορισμένο μέτρο, για παράδειγμα, η ευκλείδεια απόσταση. Κάποιες φορές το k ερμηνεύεται όχι ως ο αριθμός των σημείων, αλλά η ακτίνα γύρω από το συγκεκριμένο σημείο. Ο πιο απλός τρόπος για να αποδοθεί ετικέτα σε ένα νέο σημείο, ονομάζεται «ψηφοφορία πλειοψηφίας». Σε αυτή παίρνουμε τα k κοντινότερα σημεία και μετρούμε πόσα από αυτά είναι σημασμένα με κάθε τρόπο. Η «σταθμισμένη ψηφοφορία πλειοψηφίας» αποδίδει βάρη σε διαφορετικές κατευθύνσεις. Σε αυτή την περίπτωση η διαδικασία εκπαίδευσης του αλγορίθμου είναι η εύρεση του βέλτιστου βάρους για κάθε διάσταση (Causton, Quackenbush & Brazma 2003).

6. ΜΕΘΟΔΟΛΟΓΙΑ

6.1. ΕΙΣΑΓΩΓΗ

Η τεχνολογία των μικροσυστοιχιών επιτρέπει την ταυτόχρονη εξέταση των επιπέδων έκφρασης χιλιάδων γονιδίων σε ένα μόνο πείραμα. Το πλήθος των δεδομένων που προκύπτουν είναι τεράστιο και υπάρχει ένα ουσιαστικό ζήτημα σχετικά με το πως μπορεί κανείς να εξάγει συμπεράσματα και σημαντικές πληροφορίες, σχετικά με το βιολογικό σύστημα που μελετάται.

Τα δεδομένα που λαμβάνονται από τις μικροσυστοιχίες είναι θορυβώδη, εξαιτίας διαφόρων πηγών διακύμανσης στα πειράματα. Προκειμένου να αποκτηθούν αξιόπιστα δεδομένα γονιδιακής έκφρασης για περαιτέρω ανάλυση (π.χ. ομαδοποίηση) και ερμηνεία, οι πειραματικές διαδικασίες πρέπει να είναι αυστηρά ελεγχόμενες και έχουν αναπτυχθεί διάφορες μέθοδοι για τη διόρθωση θορύβου υποβάθρου, το φιλτράρισμα των «κακής» ποιότητας σημείων και για την κανονικοποίηση των δεδομένων. Ωστόσο διαφορετικές μέθοδοι επεξεργασίας των δεδομένων παράγουν διαφορετικά αποτελέσματα και καθιστούν την ανάλυση των δεδομένων περισσότερο δύσκολη.

Στην παρούσα εργασία χρησιμοποιούμε δεδομένα γονιδιακής έκφρασης από κυτταρική σειρά λευχαιμίας T-λεμφοκυττάρων (συγκεκριμένα την CCRF-CEM), που υφίσταται αγωγή με διαφορετικές συγκεντρώσεις πρεδνιζολόνης (Lambrou et al.). Ο πειραματικός σχεδιασμός, η ανάλυση των μικροσυστοιχιών, καθώς και η ανάλυση των εικόνων που προκύπτουν πραγματοποιήθηκε στη Μονάδα Αιματολογίας-Ογκολογίας, Α' Παιδιατρική Κλινική Πανεπιστημίου Αθηνών, Νοσοκομείο Παίδων «Αγία Σοφία». Η προεπεξεργασία και ανάλυση των δεδομένων (Sifakis, E.G. 2008) πραγματοποιήθηκε στο Matlab[®], ένα προγραμματιστικό περιβάλλον που είναι φιλικό προς το χρήστη και συγχρόνως αποτελεί ένα ισχυρό εργαλείο για την ανάπτυξη και εφαρμογή αλγορίθμων για την ανάλυση δεδομένων μικροσυστοιχιών.

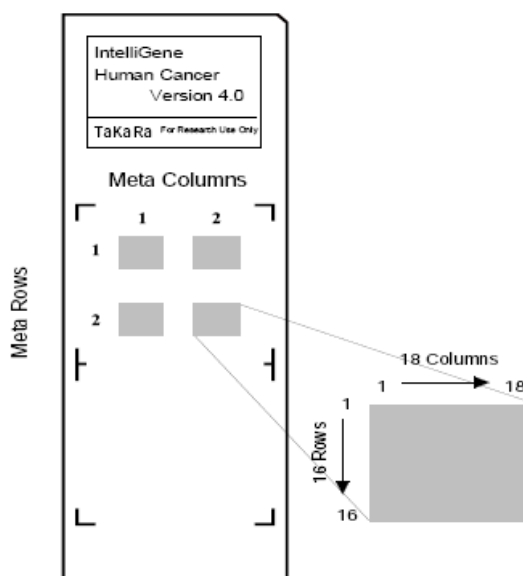
6.2. ΥΛΙΚΟ ΚΑΙ ΜΕΘΟΔΟΙ

6.2.1. Συλλογή δεδομένων

Η λευχαιμική σειρά που χρησιμοποιείται, είναι η **CCRF-CEM**. Προέρχεται από λευχαιμικά κύτταρα της CEM, ενός κοριτσιού που σε ηλικία 2 11/12 χρόνων ανέπτυξε μια μάζα του μεσοθωρακίου, η οποία διαγνώστηκε ως λεμφοσάρκωμα με βιοψία του λεμφαδένα της αντίστοιχης τραχηλικής χώρας. Η ασθένεια της εξελίχθηκε σε οξεία λευχαιμία με αριθμό λευκών αιμοσφαιρίων στο περιφερικό αίμα 492.000, από τα οποία το 87% ήταν λεμφοβλάστες. Αν και για κάποια διαστήματα με χορήγηση μεθοτρεξάτης και πρεδνιζολόνης παρατηρήθηκαν αιματολογικές και κλινικές υφέσεις, η ασθένεια της συνέχισε να εξελίσσεται. Σε ηλικία 3 10/12

επανεισήχθει στην κλινική θεραπείας καρκίνου, με αριθμό λεμφοκυττάρων στο περιφερικό αίμα 177.000 από τα οποία το 98% ήταν λεμφοβλάστες. Παρά την εντατική χημειοθεραπεία η υποτροπή συνέχισε. Πριν το θάνατό της στις 13 Νοεμβρίου 1964, 4 δείγματα των 20ml του περιφερικού της αίματος χρησιμοποιήθηκαν για κυτταρική καλλιέργεια, σε αποστειρωμένους φυγόκεντρους σωλήνες που περιείχαν κιτρικό οξύ 50mg/10ml αίματος (Foley et al. 1965). Η σειρά αυτή εκτίθεται σε διαφορετικές συγκεντρώσεις πρεδνιζολόνης και στη συνέχεια οι γονιδιακές εκφράσεις αναλύονται, χρησιμοποιώντας μικροσυστοιχίες cDNA 1,2K.

Η μικροσυστοιχία που χρησιμοποιείται είναι η IntelliGene™ Human Cancer CHIP Version 4.0. Είναι μια τυπωμένη cDNA μικροσυστοιχία με ακινητοποιημένα 886 τμήματα cDNA ανθρώπινων γονιδίων, από τα οποία τα 588 είναι τμήματα ανθρώπινων γονιδίων που σχετίζονται με τον καρκίνο και τα 298 είδη cDNA είναι τμήματα επιλεγμένα με την τεχνική διαφορικής παράθεσης (differential display method-DD method), μεταξύ καρκινικού και φυσιολογικού ιστού. Για τα σχετιζόμενα με τον καρκίνο γονίδια, τμήματα cDNA περίπου 300 ζευγών βάσεων για κάθε γονιδίου, που έχουν ελάχιστη ομολογία με άλλα γονίδια καταχωρημένα στη βάση δεδομένων cDNA, σημαίνονται στη μικροσυστοιχία. Τα τμήματα cDNA, 250 - 1.200 ζευγών βάσεων, που επιλέχθηκαν με τη μέθοδο DD ως γονίδια που έχουν διαφορά στα επίπεδα έκφρασης, κλωνοποιήθηκαν και προσδιορίστηκε η αλληλουχία βάσεων τους. Αυτή η μικροσυστοιχία DNA είναι κατάλληλη για την έρευνα του καρκίνου, αφού το σύνολο των γονιδίων που χρησιμοποιείται, είναι γνωστό ότι συμμετέχουν στην εξέλιξη του καρκίνου, στην απόπτωση, στον κυτταρικό κύκλο και την ογκοκαταστολή (IntelliGene™ Human Cancer CHIP 2009).



Εικόνα 6.1: IntelliGene™ Human Cancer CHIP Version 4.0 (IntelliGene™ Human Cancer CHIP 2009).

Η έκφραση κάθε γονιδίου μπορεί να αναλυθεί ταυτόχρονα μέσω του υβριδισμού των RNA στόχων που έχουν ληφθεί από ανθρώπινα κύτταρα. Αυτή η μικροσυστοιχία, είναι επίσης επισημασμένη με τμήματα DNA, που λειτουργούν ως έλεγχοι. Αυτές οι κουκίδες έλεγχου μπορούν να χρησιμοποιηθούν για να

ομαλοποιήσουν τα σήματα έντασης των φθορίζουσών ουσιών, καθορίζοντας το επίπεδο του υποβάθρου.

Κουκίδες θετικού ελέγχου: Τμήματα cDNA που προέρχονται από τον άνθρωπο (ως γονίδια σταθερής έκφρασης).

- 1: beta-actin
- 2: ATP synthase, H⁺ transporting, mitochondrial F₀ complex, subunit b, isoform 1
- 3: glucose-6-phosphate dehydrogenase
- 4: glyceraldehyde-3-phosphate dehydrogenase
- 5: hexokinase 1
- 6: ribosomal protein S5
- 7: transferrin receptor
- 8: tubulin, alpha2

Κουκίδες αρνητικού ελέγχου

- 9: lambda-A
- 10: lambda-B
- 11: pUC19
- 12: Arabidopsis chlorophyll ab binding protein

Τα τμήματα DNA από γονίδια που συνδέονται με τον καρκίνο και οι αλληλουχίες έλεγχου παρατάσσονται στη χρωματισμένη περιοχή της γυάλινης πλάκας. Η περιοχή των κουκίδων αποτελείται από 4 μεταπλέγματα, κάθε ένα από τα οποία αποτελείται από υπο-πλέγματα 16 σειρών x 18 στηλών. Η θέση κάθε μεταπλέγματος μπορεί να προσδιοριστεί από τον αριθμό της μεταγραμμής και της μεταστήλης που βρίσκεται.

Η μικροσυστοιχία υβριδίζεται με cDNA από μη επεξεργασμένα κύτταρα και από κύτταρα που έχουν υποστεί αγωγή με 10nM, και 701 μM πρεδνιζολόνης. Τα ακόλουθα ζευγάρια των δειγμάτων συγκρίνονται σε πειράματα:

1. Μη επεξεργασμένα κύτταρα ως προς κύτταρα που έχουν υποστεί αγωγή με 10nM πρεδνιζολόνης (αναφέρεται ως «Ovs1»).
2. Κύτταρα υπό αγωγή με 10nM ως προς κύτταρα υπό αγωγή με 701 μM πρεδνιζολόνης (αναφέρεται ως «1vs3»).
3. Μη επεξεργασμένα κύτταρα ως προς κύτταρα υπό αγωγή με 701 μM πρεδνιζολόνης (αναφέρεται ως «Ovs3»).

6.2.2. Σάρωση μικροσυστοιχίας

Οι μικροσυστοιχίες στη συνέχεια σαρώνονται με το **ScanArray 4000XL** (GSI Lumonics, USA) και παράγονται οι εικόνες των μικροσυστοιχιών. Το ScanArray 4000XL είναι ένα σύστημα ανάλυσης μικροσυστοιχιών τριών χρωμάτων, που συνδυάζει ευελιξία στην ανάλυση μικροσυστοιχιών και δυνατότητα αναβάθμισης.

Χρησιμοποιείται για την ανάγνωση των επιπέδων έκφρασης από τον υβριδισμό των στόχων που έχουν επικολληθεί στις μικροσυστοιχίες.

Χαρακτηριστικά

1. Τρία λέιζερ με δυνατότητα επιλογής από το χρήστη και εύρος διέγερσης από 488nm ως 633nm
2. Έως επτά φίλτρα εκπομπής, με εύρος ανίχνευσης από 500nm ως 700nm
3. Ανάλυση εικονοστοιχείων 5 μm
4. Ευαισθησία ≤ 0.1 μικρών /μόριο φθοριτή
5. Περιοχή σάρωσης 22mm X 73mm
6. Γρηγορότερες ταχύτητες σάρωσης (< 5 λεπτά για 22mm X 73mm)
7. Επιλογή ανάγνωσης ραβδοκώδικα
8. Δυνατότητα αναβάθμισης



Εικόνα 6.2: Packard Bioscience ScanArray 4000XL (GSI Lumonics ScanArray 4000 chip scanner 2009).

6.2.3. Ανάλυση εικόνας με χρήση του ImaGene

Το ImaGene v 6.1.0 (Biodiscovery Inc.,USA) είναι το λογισμικό ανάλυσης εικόνας που χρησιμοποιείται για την ανάλυση των εικόνων των μικροσυστοιχιών, που έχουν παραχθεί από το σαρωτή και τελικά τη δημιουργία των ακατέργαστων δεδομένων (raw data) για μετέπειτα ανάλυση. Πρόκειται για τιμές έντασης φθορισμού σημάτων ενδιαφέροντος και υποβάθρου από τα κανάλια των Cy3 και Cy5.

Το ImaGene αποθηκεύει τα ποσοτικοποιημένα δεδομένα από κάθε εικόνα που επεξεργάζεται, σε ξεχωριστά αρχεία κειμένων. Λόγω της ικανότητας επεξεργασίας πολλαπλών εικόνων, το ImaGene δεν εξάγει πλέον τους λόγους έκφρασης των δεδομένων από δύο κανάλια. Οι αναλυτές πρέπει είτε να παραγάγουν οι ίδιοι τους επιθυμητούς λόγους, είτε να εισαγάγουν τα ακατέργαστα δεδομένα σε προγράμματα ανάλυσης δεδομένων.

Το αρχείο δεδομένων, που είναι το αρχείο εξόδου του ImaGene, διαιρείται σε δύο κύρια μέρη. Το πρώτο μέρος είναι η επικεφαλίδα με πληροφορίες σχετικά με τις παραμέτρους που τίθενται στην ανάλυση. Το δεύτερο μέρος του αρχείου περιλαμβάνει τα δεδομένα που εξάγονται από τις εικόνες.

Το αρχείο εξόδου έχει την παρακάτω μορφή:

```

Begin Header
version 6.1.0
Date Sat May 12 22:43:12 EEST 2007
Image File D:\My Documents\Academic\Scientific
Work\Projects\Leukemia\Experiments\Microarrays\Chemotherapy\CCRFCCEM\Prednisolone-1.2k Chip\Control vs
0,00001nM-4h\Exp-1\CCRFCCEM (Cy3)\Predni0+Cy5\Predni1-Takara_90 100_021106.Tif
Page 0
Page Name
Inverted false
Begin Field Dimensions
Field Metarows Metacols Rows Cols
A 2 16 18
End Field Dimensions
Begin Measurement parameters
Segmentation Method auto
Signal Low 0.0
Signal High 0.0
Background Low 0.0
Background High 0.0
Background Buffer 2.0
Background Width 5.0
End Measurement parameters
Begin Alerts
Control Type Minimum threshold If tested Percentage allowed If failed Maximum
IF tested Percentage allowed If failed CV threshold If tested If failed If failed
BLANK 0.0 false 1.0% false 500.0 true 10.0% true 1.0 true true 1.0 false
POSITIVE 1000.0 true 0.1% false 100000.0 false 1.0% false 1.0 false 1.0 false
End Alerts
Begin Quality Flags
Begin Flagging Settings
Empty Spots true Threshold: 2.0
Poor Spots true
Begin Poor Spots Parameters
Background contamination flag true Threshold: 0.9995
Background tested against subgrid data only true
Signal contamination flag false Threshold: 0.9995
Signal contamination test connected to background contamination threshold false
Ignored percentage flag true Threshold: 25.0
Open perimeter flag true Threshold: 25.0
Shape regularity flag true Threshold: 0.65
Area To Perimeter Ratio flag false Threshold: 0.65
Offset flag true Threshold: 60.0
End Poor Spots Parameters
Negative Spots true
End Flagging Settings
Begin Flagged spots
# of Empty Spots: 1056
# of Poor Spots: 0
# of Negative Spots: 0
# of Manually Flagged Spots: 0
End Flagged spots
End Quality Flags
End Header
Begin Raw Data
Field Meta Row Meta Column Row Column Gene ID Flag Signal Mean Background Mean Signal Mec
Background Median Signal Mode Background Mode Signal Area Background Area Signal Total Background Total
Signal Stdev Background Stdev Shape Regularity Ignored Area Spot Area Ignored Median Area To
Perimeter Open Perimeter XCoord YCoord Diameter Position offset Offset X Offset Y Expected >
Expected Y CM-X CM-Y CM Offset CM Offset-X CM Offset-Y Min Diam Max Diam Control
Failed Control Background contamination present Signal contamination present Ignored % failed Negative Open
perimeter failed Shape regularity failed Perim-to-area failed Offset failed Empty spot Negative spot
Selected spot Saturated spot
A 1 1 1 1 position_marker 0 3671.1621 3554.9182 3289.0 3100.5
4122.1459 2697.9882 216.0 1368.0 792971.0 4863128.0 2541.7485 2517.5549 0.8437 1.
217.0 14177.0 1.0 0.0 109.5 245.5 17.0 3.2764 -2.354 -2.2789 111.854 247.7789 109.5 245.5
3.2764 -2.354 -2.2789 16.6177 16.6177 POSITIVE 0 0 0 0 0 0 0 0 0 0
0 0 0

```

Εικόνα 6.3: Παράδειγμα αρχείου εξόδου ImaGene v6.1.0.

Όπως φαίνεται παραπάνω, το αρχείο εξόδου παρέχει για κάθε κουκίδα της μικροσυστοιχίας, ένα πλήθος χαρακτηριστικών της. Κάποια από αυτά είναι τα:

1. Πεδίο (field): το όνομα του πεδίου στο οποίο βρίσκεται η κουκίδα.
2. Μεταγραμμή (Metarow): ο αριθμός της μεταγραμμής του μεταπλέγματος που βρίσκεται η κουκίδα.
3. Μεταστήλη (Metacolumn): ο αριθμός της μεταστήλης του μεταπλέγματος που βρίσκεται η κουκίδα.
4. Γραμμή (row): ο αριθμός της γραμμής στο υποπλέγμα, όπου βρίσκεται η κουκίδα.
5. Στήλη (column): ο αριθμός της στήλης στο υποπλέγμα, όπου βρίσκεται η κουκίδα.
6. Ταυτότητα του γονιδίου (GeneID): πληροφορίες σχετικά με το γενετικό υλικό που έχει σημειωθεί στην κουκίδα.

7. Σημαία (flag): ο αριθμητικός κωδικός της σημαίας για την κουκίδα (0 καμία σημαία, κωδικοί σημαίας 1,..., 7).
8. Μέση τιμή του σήματος (Signal Mean): ο μέσος όρος της έντασης των εικονοστοιχείων του σήματος ενδιαφέροντος.
9. Μέση τιμή του υποβάθρου (Background Mean): ο μέσος όρος της έντασης των εικονοστοιχείων του τοπικού υποβάθρου.
10. Διάμεσος σημάτων (Signal Median): η μεσαία τιμή των εντάσεων των εικονοστοιχείων του σήματος ενδιαφέροντος.
11. Διάμεσος υποβάθρου (Background Median): η μεσαία τιμή των εντάσεων των εικονοστοιχείων του τοπικού υποβάθρου.
12. Επικρατούσα τιμή των σημάτων (Signal Mode): η επικρατούσα τιμή των εντάσεων των εικονοστοιχείων του σήματος ενδιαφέροντος.
13. Επικρατούσα τιμή του υποβάθρου (Background Mode): η επικρατούσα τιμή των εντάσεων των εικονοστοιχείων του τοπικού υποβάθρου.
14. Περιοχή σήματος (Signal area): ο αριθμός των εικονοστοιχείων του σήματος ενδιαφέροντος.
15. Περιοχή υποβάθρου (Background area): ο αριθμός των εικονοστοιχείων του τοπικού υποβάθρου.
16. Συνολικό σήμα (signal total): η συνολική ένταση των εικονοστοιχείων του σήματος ενδιαφέροντος .
17. Συνολικό υπόβαθρο (Background total): η συνολική ένταση των εικονοστοιχείων του τοπικού υποβάθρου.
18. Τυπική απόκλιση σήματος (Signal Stdev): η σταθερή απόκλιση των εντάσεων των εικονοστοιχείων του σήματος ενδιαφέροντος.
19. Τυπική απόκλιση υποβάθρου (Background Stdev): η σταθερή απόκλιση των εντάσεων των εικονοστοιχείων του τοπικού υποβάθρου.
20. Κανονικότητα σχήματος (Shape Regularity): πρώτα η περιοχή σήματος εγγράφεται σε κύκλο. Ο αριθμός των εικονοστοιχείων που δεν έχουν σήμα και εμπίπτουν σε αυτόν τον κύκλο υπολογίζεται και διαιρείται με το εμβαδό του κύκλου. Αυτός ο λόγος αφαιρείται από το 1 και καλείται «κανονικότητα σχήματος».

6.3. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ

Η ανάλυση των ακατέργαστων δεδομένων από τις μικροσυστοιχίες, πραγματοποιείται στο Matlab[®] v 7.6.0, που παρέχει ένα ολοκληρωμένο περιβάλλον για την ανάλυση δεδομένων γονιδιακής έκφρασης και επιτρέπει στο χρήστη να εκτελέσει συγκεκριμένους αλγορίθμους, με ευκολία και ευελιξία, προσαρμοσμένους στις ανάγκες του (Venet 2003, Knowlton, Dozmorov & Centola 2004, Wang et al. 2002).

Σε ολόκληρη την προεπεξεργασία χρησιμοποιούμε για την ένταση τόσο του σήματος όσο και του υποβάθρου, την τιμή της διαμέσου (median) των εντάσεων κάθε καναλιού, καθώς η διάμεσος είναι ένα εύρωστο μέτρο της κεντρικής τάσης των δεδομένων, σε σχέση με τη μέση τιμή (mean) (Samuels, Witmer 2003).

6.3.1. Διόρθωση υποβάθρου

Τα δεδομένα που λαμβάνονται από τις μικροσυστοιχίες είναι θορυβώδη, εξαιτίας διαφόρων πηγών διακύμανσης στα πειράματα. Έτσι για τη διόρθωση του θορύβου υλοποιούνται δύο εναλλακτικές:

1. Αφαίρεση της διαμέσου των εντάσεων τοπικού υποβάθρου από τη διάμεσο των εντάσεων του σήματος ενδιαφέροντος (median background correction).
2. Μη διόρθωση του υποβάθρου (no background correction).

Στην πρώτη περίπτωση, για να υπολογίσουμε την διορθωμένη τιμή έντασης, χρησιμοποιείται η σχέση

$$I_{median} = I_{median}^{foreground} - I_{median}^{background}$$

δηλαδή αφαιρείται μια προσέγγιση της επίδρασης του μη ειδικού υβριδισμού και της χωρικής ετερογένειας της μικροσυστοιχίας.

Οι τιμές $I_{median}^{foreground}$ και $I_{median}^{background}$ αντιπροσωπεύουν τις διαμέσους των τιμών των εντάσεων των εικονοστοιχείων του σήματος ενδιαφέροντος και του υποβάθρου αντίστοιχα, όπως αυτές προσδιορίζονται από το πρόγραμμα επεξεργασίας της εικόνας για κάθε γονίδιο και για κάθε κανάλι.

Στη δεύτερη περίπτωση, δεν πραγματοποιείται διόρθωση υποβάθρου και άρα

$$I_{median} = I_{median}^{foreground}$$

Αυτή η προσέγγιση εφαρμόζεται, γιατί υπάρχουν αρκετές ανησυχίες για την αποτελεσματικότητα της διόρθωσης θορύβου, μέσω της μέτρησης του υποβάθρου που λαμβάνεται από το μηχάνημα. Η απλή αφαίρεση του θορύβου υποβάθρου που παρέχεται από το λογισμικό των οργάνων δεν παράγει απαραίτητως μια καλύτερη ανάγνωση της γονιδιακής έκφρασης. Πρόκειται για μια εκτίμηση του θορύβου. Και μια κακή εκτίμηση μπορεί να είναι χειρότερη από καμία εκτίμηση (Ritchie et al. 2007).

6.3.2. Φιλτράρισμα δεδομένων χαμηλής έντασης

Μετά τη διόρθωση υποβάθρου πραγματοποιείται φιλτράρισμα των δεδομένων, βάση της έντασης των στοιχείων της μικροσυστοιχίας. Η διαδικασία αυτή φιλτράρει τα δεδομένα χαμηλής έντασης τα οποία συνδέονται με «κακά σημεία». Τα δεδομένα χαμηλής έντασης αντιστοιχούν σε κουκίδες στις οποίες δεν υπάρχει σημαντική διαφορά μεταξύ των εντάσεων του σήματος ενδιαφέροντος και του υποβάθρου και για τα οποία υπάρχει αβεβαιότητα. Μετά τη διόρθωση υποβάθρου είναι δυνατόν να προκύψουν και σημεία με αρνητική τιμή έντασης, όταν η ένταση του υποβάθρου είναι μεγαλύτερη από αυτή του σήματος ενδιαφέροντος. Στη συγκεκριμένη εργασία χρησιμοποιείται ως απολύτως κατώτερη τιμή έντασης σημείου, το κατώφλι ίσο με 10 και για τα δύο κανάλια. Η τιμή αυτή επιλέγεται τέτοια ώστε να απορριφθούν οι

αρνητικές και οι πολύ μικρές τιμές έντασης που υπάρχουν στα «κακής ποιότητας» δεδομένα που έχουν προκύψει κυρίως από τη διόρθωση υποβάθρου (Yang et al. 2002).

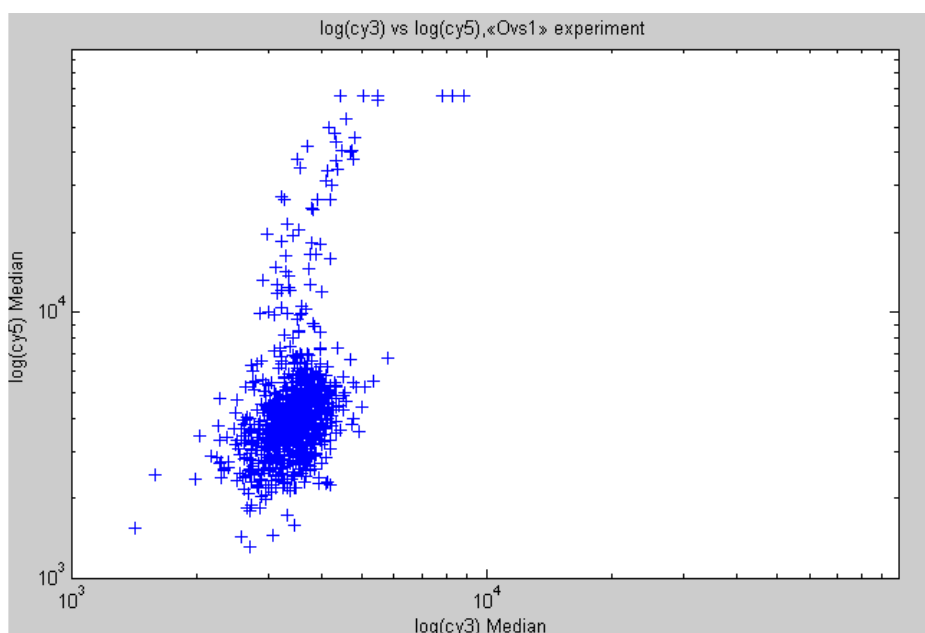
6.3.3. Κανονικοποίηση

Στη συνέχεια τα δεδομένα υποβάλλονται σε **κανονικοποίηση**. Σκοπός της κανονικοποίησης είναι η ελαχιστοποίηση της επιρροής των εξωγενών παραγόντων, που προκαλούν συστηματική διακύμανση στα μετρούμενα επίπεδα γονιδιακής έκφρασης των δύο δειγμάτων mRNA που έχουν υβριδιστεί από κοινού στη συστοιχία, προκειμένου να είναι δυνατές οι συγκρίσεις βιολογικής σημασίας. Η κανονικοποίηση διορθώνει διαφορές κατά τη σήμανση των δειγμάτων, ατέλειες κατά την ανίχνευση του φθορισμού των χρωστικών ουσιών και διαφορές που οφείλονται στην αρχική ποσότητα RNA των δύο δειγμάτων που εξετάζονται (Quackenbush 2001).

Στην παρούσα εργασία για την κανονικοποίηση, χρησιμοποιούνται τα επίπεδα έκφρασης όλων των γονιδίων της μικροσυστοιχίας, λόγω της υπόθεσης ότι μόνο ένα μικρό ποσοστό των γονιδίων θα διαφοροποιείται σημαντικά όσον αφορά στην έκφραση τους στα δείγματα και υλοποιούνται τρεις εναλλακτικές κανονικοποίησης.

Υπάρχει η δυνατότητα κανονικοποίησης των δεδομένων, χρησιμοποιώντας **λογαριθμικό μετασχηματισμό** με βάση το 2. Είναι ο πιο συχνά χρησιμοποιούμενος μετασχηματισμός, λόγω της ιδιότητάς του να κατανέμει τα δεδομένα ομοιόμορφα.

Χρησιμοποιώντας την συνάρτηση `maloglog` στο Matlab προκύπτει το διάγραμμα διασποράς του καναλιού Cy3 ως προς το Cy5.



Εικόνα 6.4: Διάγραμμα διασποράς του καναλιού Cy3 ως προς το Cy5.

Μια δεύτερη εναλλακτική κανονικοποίησης των δεδομένων είναι **κανονικοποίηση εξαρτώμενη από την ένταση** (Yang et al. 2002, Quackenbush 2002, Park et al. 2003). Εφαρμόζεται «ομαλοποίηση τοπικής παλινδρόμησης» (local regression smoothing) και συγκεκριμένα τοπικά σταθμισμένη εξομάλυνση του διαγράμματος διασποράς (Locally Weighted scatter plot smoother), που χρησιμοποιεί τοπικά σταθμισμένη γραμμική παλινδρόμηση για να ομαλοποιήσει τα δεδομένα (Cleveland 1979). Η παράμετρος f , που δηλώνει το ποσοστό των δεδομένων που θα χρησιμοποιηθούν για την κανονικοποίηση τίθεται ίση με 40%. Όσο μεγαλύτερη είναι η τιμή της παραμέτρου, τόσο καλύτερη είναι η εξομάλυνση (Yang et al. 2002). Η διαδικασία ομαλοποίησης θεωρείται τοπική, γιατί κάθε εξομαλυμένη τιμή καθορίζεται από τις τιμές των γειτονικών σημείων, που ανήκουν στο πεδίο τιμών (span). Είναι σταθμισμένη, επειδή ορίζεται συνάρτηση βάρους παλινδρόμησης για τα δεδομένα που ανήκουν στο πεδίο τιμών. Επιπλέον της συνάρτησης βάρους, στη συνάρτηση παλινδρόμησης χρησιμοποιείται μια εύρωστη (robust) συνάρτηση βάρους, που καθιστά τη διαδικασία ανεκτική σε έκτροπες τιμές (outliers).

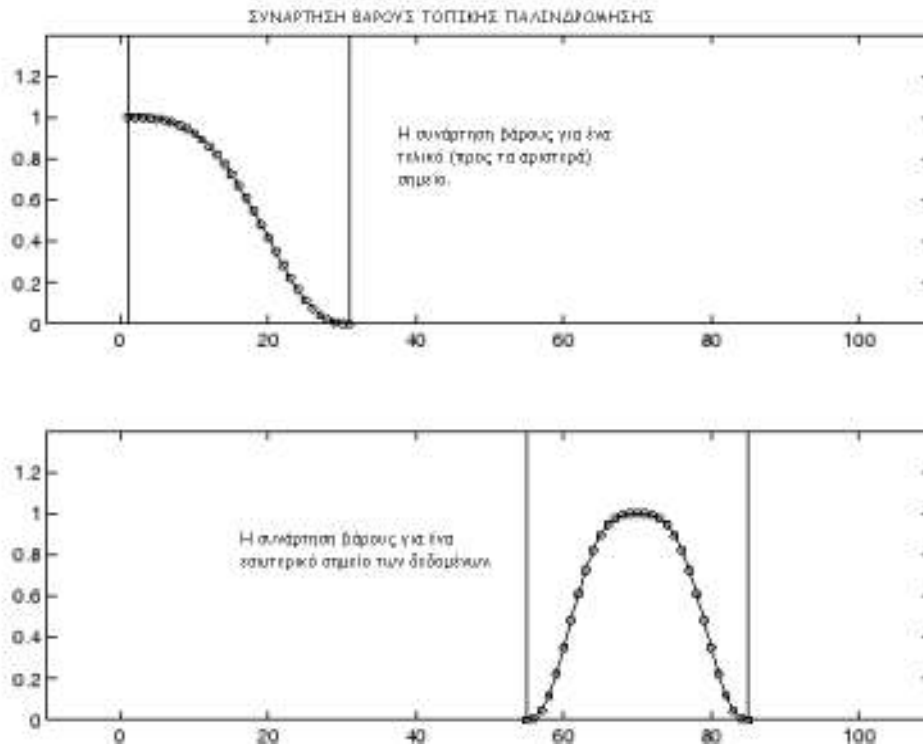
Στο Matlab υπάρχουν δύο συναρτήσεις τοπικής παλινδρόμησης που διαφοροποιούνται στο μοντέλο της παλινδρόμησης που χρησιμοποιούν. Αυτές είναι η **lowess**, που χρησιμοποιεί πολυώνυμο πρώτου βαθμού και η **loess**, που χρησιμοποιεί πολυώνυμο δευτέρου βαθμού, για την παλινδρόμηση ελαχίστων τετραγώνων με βάρη (Cleveland 1979).

Η Lowess μέθοδος κανονικοποίησης, πραγματοποιεί ένα μεγάλο αριθμό γραμμικών παλινδρομήσεων σε επικαλυπτόμενα παράθυρα και στη συνέχεια οι επιμέρους παλινδρομήσεις συνδυάζονται για να σχηματιστεί η καμπύλη παλινδρόμησης. Για κάθε σημείο δεδομένων που βρίσκεται στο πεδίο τιμών υπολογίζονται τα βάρη παλινδρόμησης w_{re_i} , χρησιμοποιώντας τη σχέση

$$w_{re_i} = \left(1 - \left|\frac{x-x_i}{d(x)}\right|^3\right)^3$$

όπου x είναι η εκτιμώμενη τιμή, που σχετίζεται με την εξαρτημένη μεταβλητή που πρέπει να ομαλοποιηθεί, x_i είναι οι κοντινότεροι γείτονες του x , όπως ορίζονται από το πεδίο τιμών και $d(x)$ είναι η απόσταση της πιο απομακρυσμένης εκτιμώμενης τιμής, μέσα στο πεδίο τιμών, από τον άξονα των x . Τα σημεία που πρέπει να ομαλοποιηθούν έχουν το μεγαλύτερο βάρος και την μεγαλύτερη επίδραση στην προσαρμογή. Τα σημεία εκτός του πεδίου τιμών έχουν μηδενικό βάρος και καθόλου επίδραση στην προσαρμογή.

Στη συνέχεια εφαρμόζεται μια σταθμισμένη γραμμική παλινδρόμηση ελαχίστων τετραγώνων και οι κανονικοποιημένες τιμές αποδίδονται μέσω της σταθμισμένης παλινδρόμησης στις εκτιμώμενες τιμές ενδιαφέροντος. Αν ο αριθμός των γειτονικών, στο προς ομαλοποίηση σημείο, σημείων είναι ίδιος και στις δύο πλευρές, η συνάρτηση βάρους είναι συμμετρική. Διαφορετικά η συνάρτηση βάρους δεν είναι συμμετρική.

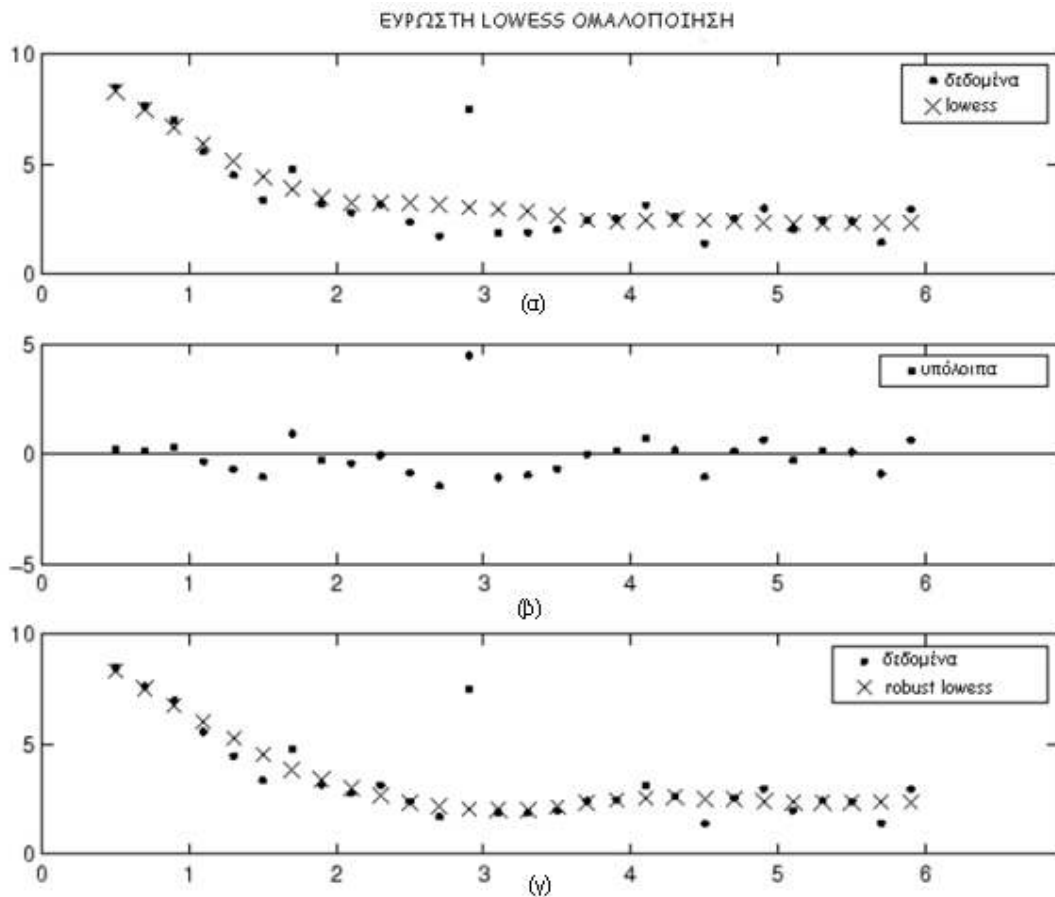


Εικόνα 6.5: Η συνάρτηση βάρους για ένα τελικό και ένα εσωτερικό σημείο του πεδίου τιμών (The MathWorks Incorporated 2009).

Αν τα δεδομένα περιέχουν έκτροπες τιμές, χρησιμοποιείται η εύρωστη (robust) μέθοδος. Η εύρωστη διαδικασία προσαρμογής επιτηρεί για σημεία που αποκλίνουν στο διάγραμμα διασποράς και υπολογίζει για τα σημεία αυτά, τα εύρωστα βάρη w_{ro_i} από τη σχέση

$$w_{ro_i} = \begin{cases} \left(1 - (r_i/6MAD)^2\right)^2, & |r_i| < 6MAD \\ 0, & |r_i| > 6MAD \end{cases}$$

όπου r_i είναι το υπόλοιπο για το i -οστό σημείο και ορίζεται ως η διαφορά μεταξύ της εξαρτημένης παρατηρούμενης τιμής και της εκτιμώμενης εξαρτημένης τιμής που παράγεται από την παλινδρόμηση, και MAD είναι η μέση απόλυτη απόκλιση των υπολοίπων r ($MAD = \text{median}(|r|)$).



Εικόνα 6.6: Το γράφημα (α) δείχνει ότι η έκτροπη τιμή επηρεάζει τις ομαλοποιημένες τιμές αρκετών κοντινών γειτονικών σημείων. Το γράφημα (β) δείχνει ότι το υπόλοιπο της εκτροπής τιμής (outlier) είναι μεγαλύτερο από 6 MAD. Επομένως, το εύρωστο βάρος είναι μηδέν για αυτό το σημείο δεδομένων. Το γράφημα (γ) δείχνει ότι οι ομαλοποιημένες τιμές των γειτονικών της έκτροπής τιμής σημείων, απεικονίζουν το μεγαλύτερο μέρος των δεδομένων (The MathWorks Incorporated 2009).

Κατά τη διάρκεια της ανάλυσης, όσον αφορά στις συναρτήσεις παλινδρόμησης που χρησιμοποιούνται, η ανεξάρτητη παράμετρος (εκτιμώμενη τιμή x) είναι η συνολική ένταση και η εξαρτημένη είναι η λόγος των εντάσεων στα δύο κανάλια.

Με τη συνάρτηση `mairplot` στο Matlab, δημιουργούνται οι 'R-I' γραφικές παραστάσεις (γραφικές παραστάσεις του λόγου $\log_2 R/G$ για κάθε στοιχείο της μικροσυστοιχίας ως προς το γινόμενο των εντάσεων $\log_{10} RG$) των δεδομένων, οι οποίες μπορούν να αποκαλύψουν ατέλειες που σχετίζονται με την ένταση. Η lowess εξομαλύνει τις συστηματικές αποκλίσεις στις 'R-I' γραφικές παραστάσεις των δεδομένων (ratio-intensity). Η συνάρτηση `mairplot` δίνει στο χρήστη τη δυνατότητα να ορίσει τις παραμέτρους της lowess κανονικοποίησης. Μέσω της παραμέτρου 'order', ορίζεται αν η προσαρμογή θα γίνει με πολυώνυμο 1^{ου} ή 2^{ου} βαθμού, μέσω της 'robust', αν θα πραγματοποιηθεί εύρωστη κανονικοποίηση των δεδομένων και με την παράμετρο 'span', ορίζεται το μέγεθος του παραθύρου εφαρμογής (πεδίο τιμών) (The MathWorks Incorporated 2009).

Για κανονικοποίηση των δεδομένων με χρήση όλων των γονιδίων της συστοιχίας, εφαρμόζεται η συνάρτηση `loess`, δηλαδή χρησιμοποιείται πολυώνυμο 2^{ου} βαθμού. Επειδή μπορεί να υπάρξουν συστηματικές διαφορές μεταξύ των ακίδων εκτύπωσης,

όπως μικρές διαφορές στο μήκος ή στο άνοιγμα των ακίδων και παραμόρφωση μετά από πολλές ώρες εκτύπωσης, πραγματοποιείται και κανονικοποίηση με χρήση των δεδομένων που ανήκουν μέσα στην ίδια ομάδα ακίδων, δηλαδή μια (print-tip+A) εξαρτώμενη κανονικοποίηση. Για το λόγο αυτό χρησιμοποιείται πληροφορία από τα πεδία metarow και metacolumn του αρχείου εξόδου του ImaGene v 6.1.0 και τα δεδομένα χωρίζονται σε τμήματα ανάλογα με τη θέση των κουκίδων στη μικροσυστοιχία, δηλαδή χωρίζονται σε τμήματα που ανήκουν στο υποπλέγμα που έχει εκτυπωθεί με την ίδια κεφαλή (print tip). Η συνάρτηση που χρησιμοποιείται για την κανονικοποίηση αυτή, είναι η **lowess**.

6.4. ΠΕΡΑΙΤΕΡΩ ΕΠΙΛΟΓΗ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ΤΗΣ ΜΙΚΡΟΣΥΣΤΟΙΧΙΑΣ

Όπως αναφέρθηκε και παραπάνω, όταν στη διαδικασία επεξεργασίας των δεδομένων περιλαμβάνεται και η διόρθωση υποβάθρου, οι εντάσεις του τοπικού υποβάθρου ποικίλουν κατά μήκος μιας μικροσυστοιχίας και για κάθε κανάλι και ένας σημαντικός αριθμός σημείων με διορθωμένες τιμές έντασης είναι είτε αρνητικές ή μηδενικές, συνεπώς αυτά τα σημεία εξαιρούνται από την μετέπειτα ανάλυση. Επιπλέον αυτών των σημείων, εξαιτίας του αποκλεισμού διαφορετικών γονιδίων σε καθένα από τα τρία πειράματα, η ανάλυση περιλαμβάνει μόνο γονίδια που εμφανίζονται να είναι κοινά και στα τρία σύνολα πειραματικών δεδομένων.

6.5. ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Αφού επιλεγούν κάποιες από τις παραπάνω εναλλακτικές για την προεπεξεργασία των ακατέργαστων δεδομένων, υπολογίζονται οι λόγοι έκφρασης για κάθε κουκίδα της μικροσυστοιχίας και για κάθε πειραματικό σχεδιασμό και οργανώνονται σε ένα πίνακα κατάλληλων διαστάσεων, όπου οι γραμμές αντιπροσωπεύουν τα μεμονωμένα γονίδια και οι στήλες συγκεκριμένες πειραματικές συνθήκες. Το στοιχείο στην i γραμμή και j στήλη, είναι η πληροφορία της έκφρασης του γονιδίου i , κάτω από τη συνθήκη j . Κάθε εγγραφή στον πίνακα αντιπροσωπεύει τη σχετική ένταση φθορισμού του γονιδίου στο προς εξέταση δείγμα και στο δείγμα αναφοράς. Σημαντικές βιολογικές πληροφορίες μπορούν να προκύψουν συγκρίνοντας τα επίπεδα γονιδιακής έκφρασης κάτω από διαφορετικές συνθήκες. Καθώς δεν υπάρχει εκ των προτέρων γνώση των προτύπων γονιδιακής έκφρασης όλων των πειραμάτων, χρησιμοποιείται μη επιβλεπόμενη μέθοδος ομαδοποίησης (Eisen et al. 1998). Συγκεκριμένα οι λόγοι έκφρασης ταξινομούνται χρησιμοποιώντας ομαδοποίηση k -μέσων. Επειδή η ομαδοποίηση k -μέσων, απαιτεί από το χρήστη να προσδιορίσει τον αριθμό των ομάδων στις οποίες θα διαχωριστούν τα δεδομένα, εφαρμόζεται αρχικά ανάλυση πρωτευουσών συνιστωσών, που παρέχει τις κατευθύνσεις μεγαλύτερης μεταβλητότητας και βοηθά στην ομαδοποίηση των δεδομένων.

6.5.1. Ανάλυση πρωτεύουσών συνιστωσών (PCA)

Η μέθοδος PCA (principal component analysis), είναι μια τεχνική που χρησιμοποιείται για τη μείωση της διαστατικότητας ενός μεγάλου συνόλου δεδομένων, βασισμένη στο μετασχηματισμό των αρχικών μεταβλητών σε ένα νέο σύνολο μεταβλητών, τις πρωτεύουσες συνιστώσες (PCs), οι οποίες είναι ασυσχέτιστες και ταξινομούνται έτσι ώστε οι πρώτες να διατηρούν το μεγαλύτερο μέρος της διακύμανσης που παρουσιάζεται στις αρχικές μεταβλητές.

Ο χρήστης μπορεί να επιλέξει την εφαρμογή ανάλυσης πρωτεύουσών συνιστωσών στα δεδομένα είτε να μην πραγματοποιήσει αυτή την ανάλυση.

Η συνάρτηση που χρησιμοποιείται στο Matlab είναι η `princomp`, που υπολογίζει τις πρωτεύουσες συνιστώσες του συνόλου δεδομένων. Ορίζεται :

$$[\text{COEFF}, \text{SCORE}, \text{latent}, \text{tsquare}] = \text{princomp}(X)$$

Η `princomp(X)` εκτελεί ανάλυση πρωτεύουσών συνιστωσών στην $n \times p$ μήτρα δεδομένων X , και επιστρέφει τους συντελεστές κύριων συνιστωσών, επίσης γνωστούς ως φορτία (loadings). Οι γραμμές του X αντιστοιχούν στις παρατηρήσεις και οι στήλες στις μεταβλητές. Η `COEFF` είναι μια $p \times p$ μήτρα, που κάθε στήλη της περιέχει τους συντελεστές μιας πρωτεύουσας συνιστώσας. Οι στήλες είναι κατά σειρά συνιστωσών μειωμένης διακύμανσης. Η μεταβλητή `SCORE` περιέχει τις τιμές των πρωτεύουσών συνιστωσών, δηλαδή την αντιστοίχιση των στοιχείων της μήτρας X στο χώρο των πρωτεύουσών συνιστωσών. Οι γραμμές του `SCORE` αντιστοιχούν στις παρατηρήσεις και οι στήλες στις συνιστώσες. Το όρισμα `latent` είναι ένα διάνυσμα που περιέχει τις ιδιοτιμές της μήτρας συνδιακύμανσης του X , ενώ το όρισμα `tsquare`, περιέχει Hotelling's T στατιστικά για κάθε σημείο των δεδομένων και είναι ένα μέτρο της απόστασης των πολυμεταβλητών από το κέντρο του συνόλου δεδομένων.

Μπορεί να χρησιμοποιηθεί η συνάρτηση `cumsum` για να εμφανιστεί το συνολικό άθροισμα της διακύμανσης.

$$\text{cumsum}(\text{pcvars} ./ \text{sum}(\text{pcvars})) * 100$$

Όταν το ποσοστό αυτό δείχνει ότι σχεδόν το 90% της διακύμανσης αποδίδεται από τις δύο πρώτες πρωτεύουσες συνιστώσες, μπορούμε να έχουμε την αναπαράσταση των δεδομένων σε μια γραφική παράσταση ως προς τις δύο αυτές κύριες συνιστώσες (The MathWorks Incorporated 2009).

6.5.2. Ομαδοποίηση k-means

Μετά την ανάλυση πρωτευουσών συνιστωσών, μπορεί να πραγματοποιηθεί ομαδοποίηση k-μέσων.

Στη στατιστική εργαλειοθήκη του Matlab υπάρχει η συνάρτηση k-means.

$$[IDX,C,sumd,D] = kmeans(X,k)$$

Η συνάρτηση αυτή χρησιμοποιεί τους λόγους γονιδιακής έκφρασης, διαχωρίζει τα δεδομένα σε k ομάδες και επιστρέφει για κάθε σημείο ένα δείκτη με την ομάδα στην οποία ανήκει. Ο χρήστης προσδιορίζει τον αριθμό των ομάδων που θα διαχωριστούν τα δεδομένα, καθώς επίσης και το μέτρο που θα χρησιμοποιηθεί για τη μέτρηση της απόστασης, στο p-διάστατο χώρο. Μπορεί να επιλέξει ανάμεσα σε διαφορετικά μέτρα απόστασης, ανάλογα με τον τύπο των δεδομένων που πρέπει να ομαδοποιηθούν.

Η συνάρτηση k-means, βασίζεται στον αλγόριθμο ομαδοποίησης k-μέσων και διαχωρίζει τα δεδομένα σε ομάδες, ελαχιστοποιώντας την διακύμανση στο εσωτερικό τους και μεγιστοποιώντας τη μεταξύ τους μεταβλητότητα. Η kmeans αντιμετωπίζει κάθε γονίδιο ως αντικείμενο που έχει μια θέση στο χώρο. Βρίσκει ένα διαχωρισμό στον οποίο τα αντικείμενα μέσα σε μια ομάδα βρίσκονται το ένα όσο το δυνατόν πιο κοντά στο άλλο και όσο το δυνατόν πιο μακριά από τα αντικείμενα άλλων ομάδων.

Κάθε ομάδα προσδιορίζεται από τα αντικείμενα – μέλη της και από το κεντροειδές της. Το κεντροειδές κάθε ομάδας είναι το σημείο για το οποίο το άθροισμα των αποστάσεων όλων των αντικειμένων της ομάδας από αυτό, ελαχιστοποιείται. Η kmeans υπολογίζει διαφορετικά κεντροειδή για κάθε μέτρο απόστασης και ελαχιστοποιεί το άθροισμα σύμφωνα με το μέτρο αυτό.

Αφού επιλεγούν οι θέσεις των κεντροειδών των ομάδων, είτε με τυχαίο τρόπο, είτε σύμφωνα με συγκεκριμένη ανάθεση μέσω των ορισμάτων εισόδου, ο αλγόριθμος ομαδοποίησης προχωρά βάση των παρακάτω βημάτων:

1. Χρησιμοποιώντας τις θέσεις των κεντροειδών, κάθε γονίδιο (ή πείραμα) τοποθετείται στην ομάδα που αντιπροσωπεύεται από το κοντινότερο στο σημείο αυτό, κέντρο.
2. Ο αλγόριθμος επαναυπολογίζει τα κεντροειδή όλων των ομάδων.

Η επαναληπτική διαδικασία συγκλίνει όταν όλα τα κεντροειδή παραμένουν σταθερά.

Η kmeans ελαχιστοποιεί το άθροισμα των αποστάσεων κάθε αντικειμένου από το κεντροειδές της ομάδας, για όλες τις ομάδες. Ο αλγόριθμος μετακινεί αντικείμενα μεταξύ των ομάδων μέχρι το άθροισμα να μην μπορεί να μειωθεί περαιτέρω. Το αποτέλεσμα είναι ένα σύνολο ομάδων που είναι συμπαγείς και σαφώς διαχωρισμένες (Azuaje, Dopazo 2005, Quackenbush 2001, Causton, Quackenbush & Brazma 2003).

Μπορεί ο χρήστης να ελέγξει λεπτομέρειες της ελαχιστοποίησης χρησιμοποιώντας προαιρετικές παραμέτρους εισόδου, συμπεριλαμβανομένου των αρχικών τιμών των κεντροειδών των ομάδων και του μέγιστου αριθμού επαναλήψεων. Μέσω της παραμέτρου **'start'**, ο χρήστης μπορεί να επιλέξει τις τιμές

‘sample’, οπότε επιλέγονται τυχαία οι k θέσεις των κεντροειδών από τον πίνακα των δεδομένων, **‘uniform’**, οπότε επιλέγονται ομοιόμορφα σημεία με τυχαίο τρόπο, **‘cluster’** οπότε εκτελείται μια προκαταρκτική φάση ομαδοποίησης σε ένα τυχαίο υποσύνολο της τάξης του 10% (αρχικοποιείται σύμφωνα με τυχαία επιλογή των θέσεων) και **‘matrix’**, οπότε οι θέσεις των κεντροειδών προσδιορίζονται από ένα πίνακα εισόδου που περιέχει τις θέσεις αυτές.

Η λύση στην οποία καταλήγει η k means, εξαρτάται από την επιλογή των αρχικών σημείων που λειτουργούν ως κεντροειδή. Είναι έτσι πιθανό να καταλήξει σε ένα τοπικό ελάχιστο, όπου η τοποθέτηση οποιουδήποτε σημείου σε μια νέα ομάδα θα αύξανε το συνολικό άθροισμα των αποστάσεων από το κεντροειδές, ενώ υπάρχει μια καλύτερη λύση. Δίνεται όμως η δυνατότητα στο χρήστη να χρησιμοποιήσει την προαιρετική παράμετρο **‘replicates’** και να ορίσει τον αριθμό των επαναλήψεων του αλγορίθμου ομαδοποίησης, κάθε φορά με διαφορετικές αρχικές θέσεις για τα κεντροειδή και να ξεπεραστεί το πρόβλημα. Η συνάρτηση k means επιστρέφει τη λύση με τη μικρότερη τιμή $sumd$.

Όσον αφορά στο μέτρο της απόστασης που θα χρησιμοποιηθεί για την ομαδοποίηση, ο χρήστης το ορίζει μέσω της παραμέτρου **‘distance’**. Για την παράμετρο αυτή χρησιμοποιείται η τιμή **‘sqEuclidean’** (τετραγωνική ευκλείδεια απόσταση, όπου κάθε κεντροειδές υπολογίζεται από τα σημεία της ομάδας). Επίσης ο χρήστης ορίζει την αντιμετώπιση μιας ομάδας που κατά την επαναληπτική διαδικασία θα χάσει όλα τα μέλη της, μέσω τις παραμέτρου **‘emptyaction’**, με τιμές **‘error’**, οπότε η άδεια ομάδα αντιμετωπίζεται ως λάθος, **‘drop’** οπότε αφαιρούνται οι ομάδες που είναι άδειες και **‘singleton’**, οπότε σχηματίζεται μια ομάδα που αποτελείται από το σημείο που βρίσκεται μακρύτερα από το κεντροειδές (The MathWorks Incorporated 2009).

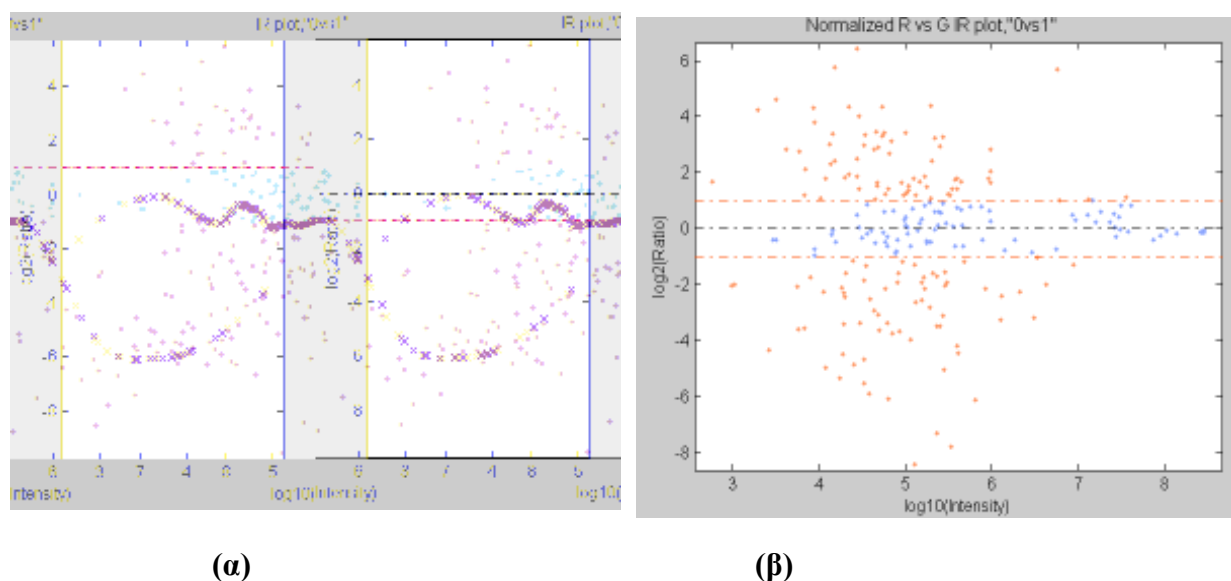
Είναι δυνατόν ο χρήστης να επιλέξει την εφαρμογή της ομαδοποίησης k -μέσων τόσο σε δεδομένα που έχουν επεξεργαστεί με τις μεθόδους προεπεξεργασίας που αναφέρονται παραπάνω χωρίς να έχουν υποστεί ανάλυση πρωτεουσών συνιστωσών, όσο και σε δεδομένα που έχουν προκύψει μετά την ανάλυση πρωτεουσών συνιστωσών. Στη δεύτερη περίπτωση με χρήση της συνάρτησης $gscatter$, δημιουργείται μια γραφική αναπαράσταση των ομάδων, όπου σημεία από κάθε ομάδα έχουν διαφορετικό χρώμα.

7. ΑΠΟΤΕΛΕΣΜΑΤΑ

Η παραπάνω μεθοδολογία προεπεξεργασίας και ανάλυσης δεδομένων, εφαρμόζεται σε δεδομένα που προέρχονται από τη γονιδιακή έκφραση κυτταρικής σειράς λευχαιμίας T-λεμφοκυττάρων που υφίσταται αγωγή με διαφορετικές συγκεντρώσεις πρεδνιζολόνης στα πειράματα «0vs1», «0vs3» και «1vs3».

Τα δεδομένα υφίστανται διόρθωση θορύβου με αφαίρεση της διαμέσου των εντάσεων τοπικού υποβάθρου από τη διάμεσο των εντάσεων του σήματος ενδιαφέροντος και loess κανονικοποίηση. Πιο συγκεκριμένα, για την loess κανονικοποίηση χρησιμοποιείται πολυώνυμο 2^{ου} βαθμού για την καμπύλη προσαρμογής και επίσης η παράμετρος f, που δηλώνει το ποσοστό των δεδομένων που θα χρησιμοποιηθούν για την κανονικοποίηση του κάθε σημείου, τίθεται ίση με 40%.

Παρακάτω παρουσιάζεται το IR-διάγραμμα διασποράς για το πρώτο πείραμα «0vs1», πριν και μετά την κανονικοποίηση. Πρόκειται για γραφική παράσταση του λογαριθμικού λόγου $\log_2 \frac{I^R}{I^G}$ για κάθε στοιχείο του πίνακα συναρτήσει του γινομένου των εντάσεων $\log_{10}(I^R * I^G)$, όπου I^R και I^G είναι οι τιμές έντασης του κόκκινου και του πράσινου καναλιού, αντίστοιχα. Στο IR-διάγραμμα διασποράς πριν την κανονικοποίηση, είναι σχεδιασμένη και η καμπύλη προσαρμογής.

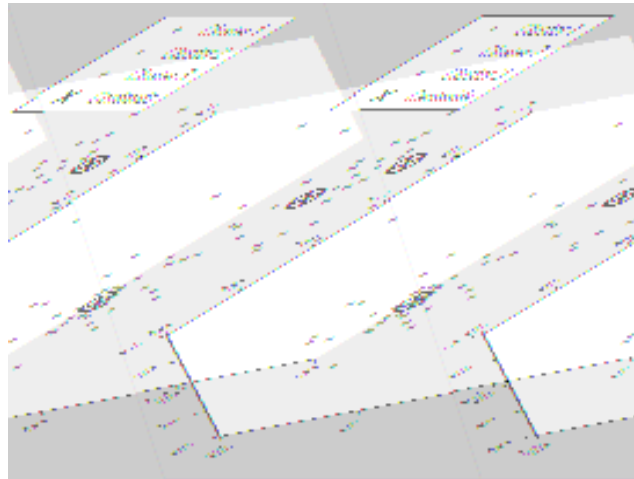


Εικόνα 7.1: IR-διάγραμμα διασποράς πριν (α) και μετά (β) την loess κανονικοποίηση των δεδομένων του πειράματος «0vs1». Τα δεδομένα είναι διορθωμένα για θόρυβο υποβάθρου. Όλες οι κουκίδες μπλε χρώματος είναι μέσα στις γραμμές διπλάσιας μεταβολής (two-fold change). Οι κουκίδες κόκκινου χρώματος αντιπροσωπεύουν δεδομένα έξω από τις γραμμές διπλάσιας μεταβολής, δηλαδή είναι είτε υπέρ είτε υπό εκφρασμένα.

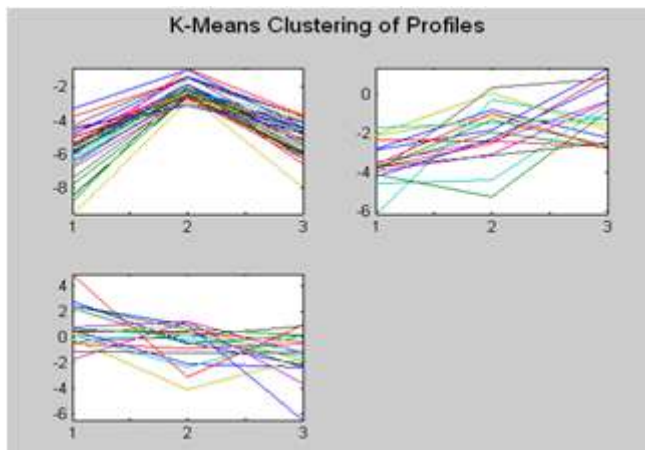
Η πραγματοποίηση ομαδοποίησης k-μέσων στα κανονικοποιημένα δεδομένα διαχωρίζει τα δεδομένα σε ομάδες. Ως μέτρο της απόστασης των σημείων των ομάδων χρησιμοποιείται η ευκλείδεια απόσταση, οι θέσεις των κεντροειδών υπολογίζονται με εκτέλεση μιας προκαταρκτικής φάσης ομαδοποίησης σε ένα τυχαίο υποσύνολο της τάξης του 10% και για την αποφυγή του τοπικού ελαχίστου, ορίζεται

ο αριθμός των επαναλήψεων του αλγορίθμου ομαδοποίησης ίσος με 5. Ο αριθμός των ομάδων k , στις οποίες διαχωρίζονται τα δεδομένα τίθεται ίσος με 3.

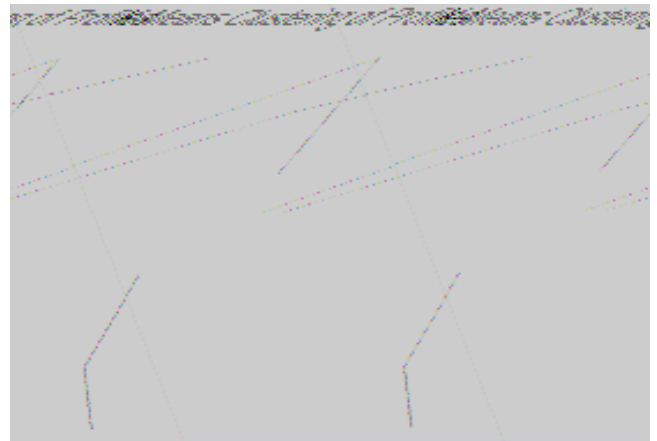
Παρακάτω φαίνεται ο διαχωρισμός των δεδομένων σε ομάδες, τα προφίλ των γονιδίων κάθε ομάδας, καθώς και τα προφίλ των αντίστοιχων κεντροειδών.



(α)



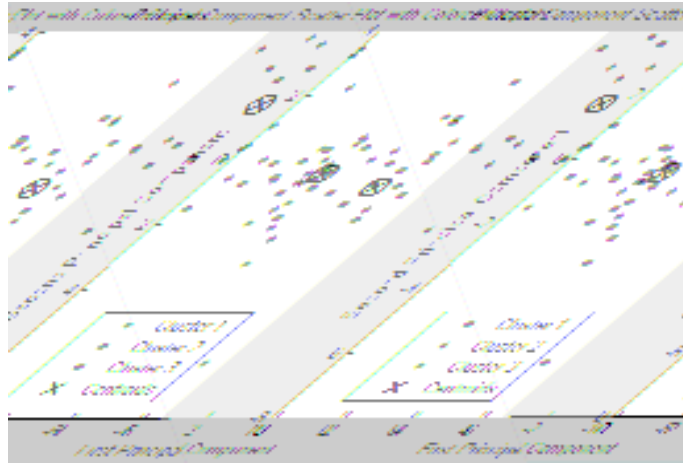
(β)



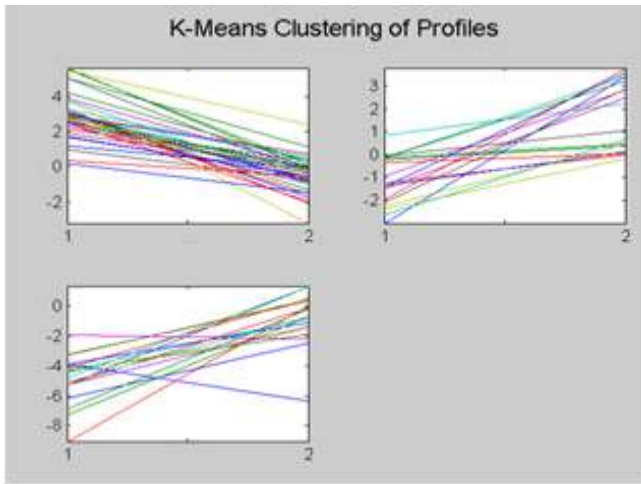
(γ)

Εικόνα 7.2: (α) ομαδοποίηση k -μέσων των διορθωμένων για θόρυβο υποβάθρου και κανονικοποιημένων δεδομένων, (β) προφίλ των γονιδίων που ανήκουν στις τρεις ομάδες και (γ) προφίλ των κεντροειδών των ομάδων.

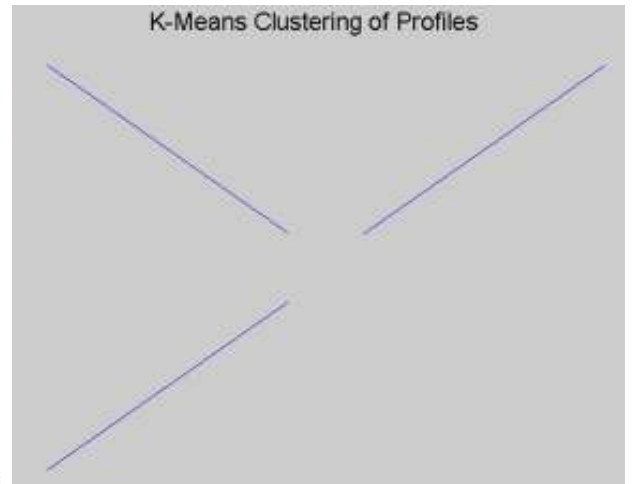
Η πραγματοποίηση ανάλυσης πρωτεουσών συνιστωσών στα δεδομένα βοηθά στη μείωση της διαστατικότητας και την αποτελεσματικότερη ανάλυση, εφόσον τα μειωμένα δεδομένα διατηρούν τις σημαντικές πληροφορίες των αρχικών δεδομένων. Η συνολική διακύμανση των δύο πρώτων πρωτεουσών συνιστωσών υπολογίστηκε ίση με 93.08, ποσοστό που είναι μεγαλύτερο του 90% της συνολικής διακύμανσης. Επιπλέον, μπορούμε να έχουμε καλύτερη οπτικοποίηση των δεδομένων, όπως φαίνεται στην εικόνα 7.3, όπου παρουσιάζεται η ομαδοποίηση k -μέσων μετά την εφαρμογή ανάλυσης πρωτεουσών συνιστωσών.



(α)



(β)



(γ)

Εικόνα 7.3: (α) Απεικόνιση των δεδομένων μετά την ανάλυση πρωτενουσών συνιστωσών, ως προς τις δύο κύριες συνιστώσες και με ομαδοποίηση k-μέσων, (β) προφίλ των γονιδίων που ανήκουν στις τρεις ομάδες και (γ) προφίλ των αντίστοιχων κεντροειδών των ομάδων.

Η οπτικοποίηση των δεδομένων σε ομάδες μετά την πραγματοποίηση ανάλυσης πρωτενουσών συνιστωσών εμφανίζεται να έχει καλύτερα αποτελέσματα σε σχέση με την οπτικοποίηση που προκύπτει με εφαρμογή της ομαδοποίησης k-μέσων απευθείας στα αρχικά δεδομένα. Όπως αναφέρθηκε και παραπάνω το ποσοστό συνολικής διακύμανσης των δύο πρώτων πρωτενουσών συνιστωσών είναι μεγαλύτερο του 90%, γεγονός που εξασφαλίζει ότι μετά την μείωση των διαστάσεων διατηρείται το σημαντικότερο μέρος της πληροφορίας των αρχικών δεδομένων. Οι πρωτεύουσες συνιστώσες είναι μεταξύ τους κάθετες, οπότε χρησιμοποιώντας την αναπαράσταση των δεδομένων ως προς τις δύο πρώτες αφαιρείται η περιττή πληροφορία και παρέχεται καλύτερη οπτικοποίηση.

8. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΕΚΤΑΣΕΙΣ

Στην παρούσα εργασία αναπτύχθηκε ένας αλγόριθμος επεξεργασίας και ανάλυσης δεδομένων από τυπωμένες cDNA μικροσυστοιχίες. Πρόκειται για ένα εργαλείο που μπορεί να βοηθήσει στην επιλογή των γονιδίων που είναι κατάλληλα για περαιτέρω ανάλυση, ώστε να είναι δυνατή η εξαγωγή συμπερασμάτων βιολογικής σημασίας.

Ο χρήστης έχει τη δυνατότητα να επιλέξει μεταξύ διαφόρων εναλλακτικών όσον αφορά στα βήματα που θα ακολουθήσει κατά την προεπεξεργασία των δεδομένων, όπως για παράδειγμα ως προς τη διόρθωση θορύβου και την κανονικοποίηση ή την περαιτέρω ανάλυση που θα χρησιμοποιηθεί. Επιπλέον, δίνεται στο χρήστη η δυνατότητα να επιλέξει μεταξύ διαφόρων τιμών για τις παραμέτρους των συναρτήσεων που έχουν χρησιμοποιηθεί κατά την υλοποίηση.

Η δυνατότητα επιλογής μεταξύ διαφορετικών αλγορίθμων και παραμέτρων, είναι ιδιαίτερα σημαντική, καθώς η χρήση συγκεκριμένων μεθόδων επεξεργασίας, ο τρόπος με τον οποίο αντιμετωπίζονται τα δεδομένα ή ο τρόπος με τον οποίο μετράται η ομοιότητα της γονιδιακής έκφρασης μπορεί να έχει τεράστια επιρροή στα συμπεράσματα της ανάλυσης, αποκαλύπτοντας διαφορετικές όψεις για τη σχέση των γονιδίων.

Επιπροσθέτως, ο αλγόριθμος που αναπτύχθηκε παρέχει μια άμεση οπτικοποίηση των ομάδων στις οποίες διαχωρίζονται τα δεδομένα. Η ομαδοποίηση είναι δυνατόν να βοηθήσει στην ανακάλυψη γονιδίων που είναι παρόμοια εκφρασμένα και πιθανόν να συνδέονται όσον αφορά στη γονιδιακή τους ρύθμιση. Η επιλογή της εφαρμογής ανάλυσης πρωτεουσών συνιστωσών, μειώνει τη διαστατικότητα των δεδομένων και αφαιρεί την περιττή πληροφορία, περιορίζοντας ακόμα περισσότερο τον αριθμό των γονιδίων που θα χρησιμοποιηθούν στην περαιτέρω ανάλυση για την εξαγωγή των βιολογικών συμπερασμάτων.

Ολοκληρώνοντας θα μπορούσαμε να προσθέσουμε πως πρόκειται για έναν αλγόριθμο που είναι επεκτάσιμος. Πιο συγκεκριμένα, δύναται κάποιος αναπτύσσοντας τον κατάλληλο κώδικα ή με απλές προσθήκες ή διαφοροποιήσεις, να δώσει και άλλες εναλλακτικές όσον αφορά στα βήματα προεπεξεργασίας ή ανάλυσης, έτσι ώστε συγκρίνοντας τα αποτελέσματα των διαφορετικών αναλύσεων να είναι δυνατή η εξαγωγή συγκριτικών συμπερασμάτων.

9. ΒΙΒΛΙΟΓΡΑΦΙΑ

- Acute lymphoblastic leukemia - Wikipedia, the free encyclopedia 2009, 16/5/2009-last update. Available: http://en.wikipedia.org/wiki/Acute_lymphoblastic_leukaemia [2009, 2/4/2009].
- Alberts, B. 2008, *Molecular biology of the cell*, New York : Garland Science, c2008.
- Alison, M. 2001, *The Cancer Handbook*, Nature Publishing Group.
- American Cancer Society 2007, 08/03/2007-last update, *ACS : How Is Acute Lymphocytic Leukemia Classified?*. Available: http://www.cancer.org/docroot/CRI/content/CRI_2_4_3X_How_Is_Acute_Lymphocytic_Leukemia_Classified.asp?nav=crl [2009, 6/3/2009].
- Azuaje, F. & Dopazo, J. 2005, *Data Analysis and Visualization in Genomics and Proteomics*, 1st edn, Wiley.
- Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Jr & Haussler, D. 2000, "Knowledge-based analysis of microarray gene expression data by using support vector machines", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 1, pp. 262-267.
- Caslini, C., Serna, A., Rossi, V., Introna, M. & Biondi, A. 2004, "Modulation of cell cycle by graded expression of MLL-AF4 fusion oncoprotein", *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K.*, vol. 18, no. 6, pp. 1064-1071.
- Causton, H.C., Quackenbush, J. & Brazma, A. 2003, *Microarray gene expression data analysis : a beginner's guide*, Blackwell Pub., Malden, MA.
- Cleveland, W.S. 1979, "Robust Locally Weighted Regression and Smoothing Scatterplots", *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829.
- Den Boer, M.L., van Slegtenhorst, M., De Menezes, R.X., Cheok, M.H., Buijs-Gladdines, J.G., Peters, S.T., Van Zutven, L.J., Beverloo, H.B., Van der Spek, P.J., Escherich, G., Horstmann, M.A., Janka-Schaub, G.E., Kamps, W.A., Evans, W.E. & Pieters, R. "A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study", [Online], vol. 10, no. 2, pp. 125-134.
- Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. 1998, "Cluster analysis and display of genome-wide expression patterns", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 95, no. 25, pp. 14863-14868.
- Foley, G.E., Lazarus, H., Farber, S., Uzman, B.G., Boone, B.A. & MCCarthy, R.E. 1965, "Continuous Culture of Human Lymphoblasts from Peripheral Blood of a Child with Acute Leukemia", *Cancer*, vol. 18, pp. 522-529.

- Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Michel, G. & Haussler, D. 2000, "Support vector machine classification and validation of cancer tissue samples using microarray expression data", *Bioinformatics*, vol. 16, no. 10, pp. 906.
- 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. & Lander, E.S. 1999, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", *Science*, vol. 286, no. 5439, pp. 531-537.
- Grant, R.P. 2004, *Computational genomics : theory and application*, Horizon Bioscience, Wymondham.
- Greaves, M.F. & Wiemels, J. 2003, "Origins of chromosome translocations in childhood leukaemia", *Nature reviews.Cancer*, vol. 3, no. 9, pp. 639-649.
- GSI Lumonics ScanArray 4000 chip scanner 2009 . Available: [http://www.gmi-inc.com/BioTechLab/GSI Lumonics ScanArray 4000.html](http://www.gmi-inc.com/BioTechLab/GSI_Lumonics_ScanArray_4000.html) [2009, 13/5/2009].
- Haferlach, T., Kohlmann, A., Schnittger, S., Dugas, M., Hiddemann, W., Kern, W. & Schoch, C. 2005, "Global approach to the diagnosis of leukemia using gene expression profiling", *Blood*, vol. 106, no. 4, pp. 1189-1198.
- Jolliffe, I.T. 2002, *Principal Component Analysis (Springer Series in Statistics)*, 2nd ed. edn, Springer.
- IBMS - *The Institute of Biomedical Science* 2009d, . Available: http://www.ibms.org/index.cfm?method=science.images&subpage=image_galleries_haematology [2009, 17/5/2009].
- IntelliGene™ Human Cancer CHIP Ver. 4.0 | TAKARA BIO INC. Genetic Engineering Research 2009 . Available: http://catalog.takara-bio.co.jp/en/product/manual_info.asp?catcd=B1000540&subcatcd=B1000541&unitid=U100005732 [2009, 6/4/2009].
- Kavraki, L.E. 2007, 11/6/2007-last update, *Dimensionality Reduction Methods for Molecular Motion*. Available: <http://cnx.org/content/m11461/latest/> [2009, 15/3/2009].
- Keene, N.B. 2002, *Childhood leukemia: a guide for families, friends & caregivers*, Sebastopol, CA : O'Reilly, c2002.
- Knowlton, N., Dozmorov, I.M. & Centola, M. 2004, "Microarray Data Analysis Toolbox (MDAT): for normalization, adjustment and analysis of gene expression data", *Bioinformatics*, vol. 20, no. 18, pp. 3687-3690.
- Lambrou, G.I., Vlahopoulos, S., Papathanasiou, C., Papanikolaou, M., Karpusas, M., Zoumakis, E. & Tzortzatos-Stathopoulou, F. "Prednisolone exerts late mitogenic

and biphasic effects on resistant acute lymphoblastic leukemia cells: Relation to early gene expression", *Leukemia research*, vol. In Press, Corrected Proof.

Life Sciences Cyberbridge 2007. Available:

http://cyberbridge.mcb.harvard.edu/dna_1.html [2009, 18/5/2009] .

Lee, M.T. 2004, *Analysis of microarray gene expression data*, Kluwer Academic, Boston.

Leukemia & Lymphoma Society - Acute Lymphocytic Leukemia 2009, 30/4/2009-last update. Available: http://www.leukemia-lymphoma.org/all_page?item_id=7049 [2009, 6/3/2009].

Lodish, H., Darnell, J.E. & etc. 2003, *Molecular Cell Biology*, 5th Revised edition edn, W.H.Freeman & Co Ltd.

Martin-Subero, J.I., Odero, M.D., Hernandez, R., Cigudosa, J.C., Agirre, X., Saez, B., Sanz-Garcia, E., Ardanaz, M.T., Novo, F.J., Gascoyne, R.D., Calasanz, M.J. & Siebert, R. 2005, "Amplification of IGH/MYC fusion in clinically aggressive IGH/BCL2-positive germinal center B-cell lymphomas", *Genes, chromosomes & cancer*, vol. 43, no. 4, pp. 414-423.

McWhirter, J.R., Neuteboom, S.T., Wancewicz, E.V., Monia, B.P., Downing, J.R. & Murre, C. 1999, "Oncogenic homeodomain transcription factor E2A-Pbx1 activates a novel WNT gene in pre-B acute lymphoblastoid leukemia", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 20, pp. 11464-11469.

Medical Library | Medem.com 2008, 14/2/2008-last update. Available:

<http://www.medem.com/medlib/article/ZZZHE608LOE> [2009, 15/3/2009].

Moorman, A.V., Harrison, C.J., Buck, G.A., Richards, S.M., Secker-Walker, L.M., Martineau, M., Vance, G.H., Cherry, A.M., Higgins, R.R., Fielding, A.K., Foroni, L., Paietta, E., Tallman, M.S., Litzow, M.R., Wiernik, P.H., Rowe, J.M., Goldstone, A.H., Dewald, G.W. & Adult Leukaemia Working Party, Medical Research Council/National Cancer Research Institute 2007, "Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial", *Blood*, vol. 109, no. 8, pp. 3189-3197.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., Su, X., Pui, C.H., Relling, M.V., Evans, W.E., Shurtleff, S.A. & Downing, J.R. 2007, "Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia", *Nature*, vol. 446, no. 7137, pp. 758-764.

National Cancer Institute, a, , *Acute Lymphoblastic Leukemia in Children*. Available:

<http://www.cancer.gov/cancertopics/factsheet/ALLinchildren> [2009, 4/3/2009] .

- National Cancer Institute, b, , *Childhood Acute Lymphoblastic Leukemia Treatment - National Cancer Institute*. Available: <http://www.cancer.gov/cancertopics/pdq/treatment/childALL> [2009, 3/3/2009] .
- Park, T., Yi, S.G., Kang, S.H., Lee, S., Lee, Y.S. & Simon, R. 2003, "Evaluation of normalization methods for microarray data", *BMC Bioinformatics*, vol. 4, pp. 33.
- Pediatric Oncology Resource Center 2005, 07/2005-last update, *ALL: Molecular Characterization*. Available: <http://www.acor.org/ped-onc/diseases/ALLmolchar.html> [2009, 6/3/2009] .
- Prasad, P.N. 2003, *Introduction to biophotonics*, Hoboken, N.J. : Wiley-Interscience, c2003.
- Pui, C. 2006, *Childhood Leukemias*, 2nd edn, Cambridge University Press.
- Pui, C.H., Relling, M.V. & Downing, J.R. 2004, "Acute lymphoblastic leukemia", *The New England journal of medicine*, vol. 350, no. 15, pp. 1535-1548.
- Pui, C.H., Schrappe, M., Ribeiro, R.C. & Niemeyer, C.M. 2004, "Childhood and adolescent lymphoid and myeloid leukemia", *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program*, , pp. 118-145.
- Quackenbush, J. 2002, "Microarray data normalization and transformation", *Nat. Genet.*, vol. 32 Suppl, pp. 496-501.
- Quackenbush, J. 2001, "Computational analysis of microarray data", *Nat. Rev. Genet.*, vol. 2, no. 6, pp. 418-427.
- Raychaudhuri, S., Stuart, J.M. & Altman, R.B. 2000, "Principal components analysis to summarize microarray experiments: application to sporulation time series", *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, , pp. 455-466.
- Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A. & Smyth, G.K. 2007, "A comparison of background correction methods for two-colour microarrays", *Bioinformatics (Oxford, England)*, vol. 23, no. 20, pp. 2700-2707.
- Samuels, M.L. & Witmer, J.A. 2003, *Statistics for the life sciences*, 3rd edn, Prentice Hall, Pearson Education, Upper Saddle River, N.J.
- Secker-Walker, L.M. 1996, *Chromosomes and Genes in Acute Lymphoblastic Leukemia (Medical Intelligence Unit)*, Landes Bioscience.
- Sifakis, E.G., Lambrou, G.I., Prentza, A., Koutsouris, D. & Tzortzatou-Stathopoulou, F. 2008b, *cDNA microarray analysis of a glucocorticoid treated acute lymphoblastic leukemia cell line*, *BioInformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*.

- Stams, W.A., den Boer, M.L., Beverloo, H.B., Meijerink, J.P., van Wering, E.R., Janka-Schaub, G.E. & Pieters, R. 2005, "Expression levels of TEL, AML1, and the fusion products TEL-AML1 and AML1-TEL versus drug sensitivity and clinical outcome in t(12;21)-positive pediatric acute lymphoblastic leukemia", *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 11, no. 8, pp. 2974-2980.
- The MathWorks Incorporated 2009, *MATLAB - The Language of Technical Computing*. Available: <http://www.mathworks.com/products/matlab/> [2008, 11/7/2008] .
- Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C. & Wong, W.H. 2001, "Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects", *Nucleic Acids Research*, vol. 29, no. 12, pp. 2549.
- Venet, D. 2003, "MatArray: a Matlab toolbox for microarray data", *Bioinformatics*, vol. 19, no. 5, pp. 659-660.
- Wang, J., Nygaard, V., Smith-Sorensen, B., Hovig, E. & Myklebost, O. 2002, "MArray: analysing single, replicated or reversed microarray experiments", *Bioinformatics*, vol. 18, no. 8, pp. 1139-1140.
- Yang, Y.H., Dudoit, S., Luu, P. & Speed, T.P. 2001, "Normalization for cDNA microarray data", *SPIE*, eds. M.L. Bittner, Y. Chen, A.N. Dorsel & E.R. Dougherty, San Jose, CA.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. & Speed, T.P. 2002, "Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation", *Nucleic Acids Res.*, vol. 30, no. 4, pp. e15.
- Ιστολογία - Εμβρυολογία Ι 2009. Available: http://emed.med.uoa.gr/application/syllabus_I/aimopiisi/aimopiisi/mielos/index.htm [2009, 20/5/2009].
- Κεντρικό δόγμα Μοριακής Βιολογίας - Βικιπαίδεια 2008, 25/12/2008-last update. Available: <http://el.wikipedia.org/wiki> [2009, 17/5/2009].
- Μαρμάρα, Β., Λαμπροπούλου-Μαρμάρα, Μ. 2005, *Βιολογία Κυττάρου: "ΜΟΡΙΑΚΗ ΠΡΟΣΕΓΓΙΣΗ"*, 5η έκδοση, Τυρογάμα, ΑΘΗΝΑ.
- Λουκόπουλος, Δ & Σταματόπουλος, Κ 1995, "Λευχαιμογένεση." in *Ετήσιο μετεκπαιδευτικό σεμινάριο της Αιματολογικής εταιρίας 1995, "Λευχαιμίες"*, pp. 29-37-43-49.
- Παγκύπριος Αντιλευχαιμικός Σύνδεσμος "ΖΩΗ" 2009. Available: <http://www.antileukemia-zoe.org.cy/cgi-bin/banner.cgi?url=http://www.antileukemia-zoe.org.cy/> [2009, 13/5/2009] .

Παρασκευόπουλος, Π. Σ., (ed) 1990, *Λευχαιμίες , Νέες απόψεις για την παθογένεια, τη διάγνωση και τη θεραπεία τους.*, University Studio Press, Θεσσαλονίκη.