

**Revealing Temporal & Functional Relations in Breast
Cancer Expression Profile Using Biclustering-Based
Analysis**

by

Güneş Gündem

**A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of**

**Master of Science
in
Computational Sciences and Engineering**

Koç University

November 2006

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Güneş Gündem

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assist. Prof. Özlem Keskin

Assoc. Prof. Attila Gürsoy

Assist. Prof. Halil Kavaklı

Assist. Prof. Engin Erzin

Assist. Prof. Mehmet Sayar

Date:

ABSTRACT

The emerging microarray technology has useful applications in cancer research. Reports detailing the expression profiles of various types of cancers have pointed to the utility of this approach in defining cancer classes and subclasses with distinct molecular configurations and clinical behaviour. Another crucial application of microarrays to cancer research is the detection and selection of diagnostic marker genes. These molecular markers give valuable additional information about tumour diagnosis, prognosis and therapy development. In the case of breast cancer, estrogen receptor (ER) and progesterone receptor (PR) have been used as prognostic markers in clinical management of breast cancer patients. Patients with ER negative breast tumors have a poorer prognosis than patients with ER positive tumors. Expression profiling of the mammary gland to identify tumor associated genes differentially expressed in breast tumors regarding the presence or absence of ER and PR lead to the identification of genes, the function of some of which is unknown. However most of these previous studies uses clustering methods which focus on the global pattern seen in the expression profiles. On the other hand, in this study, we used biclustering, which captures the local expression changes better to improve the understanding on the molecular mechanisms which underlie the response to estrogen in breast cancer cells. In addition, functional annotation enrichment, pathway information, transcriptional factor binding site analysis were also utilized. The results show that estrogen responsive genes act in groups of genes or the so-called ‘modules’ which follow an order in time. Furthermore the temporal order is associated with a biologically very meaningful functional order, which is supported by the transcription factor analysis. The pathway analysis indicates that estrogen is responsible for eliciting mechanisms required for tumorigenesis. All these results show that biclustering is very useful in the analysis of time series data and it is capable of elucidating the underlying functional and temporal relationship between genes better since it catches local expression behaviours better than the global ones.

ÖZETÇE

Gelişmekte olan gen yongası (microarray chip) teknolojisinin kanser araştırmalarında yararlı uygulamaları vardır. Kanser türlerinin ifade davranışları ile ilgili raporlar bu teknolojinin farklı moleküler konfigürasyon ve klinik davranışlara sahip kanser sınıflarını belirlemede yararlı olduğunu göstermektedir. Gen yongalarının kanser araştırmalarında diğer bir önemli uygulaması da tanısal işaretçi (marker) genlerin bulunması ve seçilmesidir. Bu moleküler işaretçi genler, tümör tanısı, öngörücü sına (prognosis) ve tedavi geliştirilmesinde değerli ek bilgiler vermiştir. Göğüs kanserine gelince, östrojen ve projesteron reseptörleri, göğüs kanseri hastalarının klinik incelemesinde tanısal işaretçi olarak kullanılmaktadır. Östrojen reseptörü pozitif hastalar, östrojen reseptörü negatif hastalardan daha iyi bir sonuç tahmini göstermektedir. Gen yongaları, östrojen ve projesteron reseptörlerinin varlığında ve yokluğunda farklı olarak ifade edilen genleri bulmak için kullanılmıştır. Bu çalışmalar aralarında fonksiyonları bilinmeyenler de dahil olmak üzere pekçok yeni genin belirlenmesini sağlamıştır. Bu çalışmaların pekçoğunda, ifade davranışlarındaki genel örüntüye (pattern) yoğunlaşan kümeleme (clustering) yöntemleri kullanılmıştır. Bu çalışmada ise, bölgesel ifade örüntülerini iyi yakalayan 'iki boyutlu kümeleme' (biclustering) yöntemi kullanıldı. Bölgesel ifade benzerliklerinin, göğüs kanseri hücrelerinin östrojene tepki verirken kullandığı önemli moleküler mekanizmaları daha iyi bulabileceği düşünülmüştür. Bu amacı gerçekleştirmek için, 'iki boyutlu kümeleme' ve ek olarak işlevsel açıklayıcı bilgi zenginleştirilmesi (functional annotation enrichment), yolak bilgisi, transkripsiyonel faktör bağlanma bölgesi analizinden de faydalanıldı. Sonuçlar, östrojene tepki veren genlerin, zaman içinde bir sırayı takip eden gen grupları (gen modülleri) şeklinde çalıştığını gösterdi. Dahası, aralarında zamansal sıra bulunan gen gruplarının, aynı zamanda işlevsel olarak da birbirlerini takip ettikleri bulunmuştur. Bu sonuç, transkripsiyonel faktör bağlanma bölgesi analizi tarafından da desteklendi. Yolak analizi ise, östrojenin tümör oluşumu için gerekli olan

mekanizmaları aktive etmekte yeterli olduğunu gösterdi. Bütün bu sonuçlar, iki boyutlu kümelemenin zaman serisi verilerinin analizinde çok kullanışlı olduğunu gösterdi. Bölgesel ifade davranışlarını genel olandan daha iyi belirleyen bu yöntem, genler arasındaki işlevsel ve zamansal ilişkileri açığa çıkarabilmektedir.

ACKNOWLEDGEMENTS

This thesis is the result of an intensive two-year study, which included many difficult times. There were many people to be grateful to. I should begin expressing my sincere thanks and appreciation to my advisor Özlem Keskin and co-advisor Attila Gürsoy for their patience, guidance and support throughout the whole study. It would not be possible to accomplish it without the positive energy they gave me.

I also want to thank my parents, Sedat and Hidayet Gundem for being the best parents one can ever have. Everything I do was and is to get their appreciation even though I do not want to accept this fact most of the times.

I should not forget to thank my little brother Ugur and my sister Ziyne for loving me and never expecting anything in return.

And finally I know that it is impossible to get by with a little help from my friends.

He that is thy friend indeed,
He will help thee in thy need:
If thou sorrow, he will weep;
If thou wake, he cannot sleep:
Thus of every grief in heart
He with thee does bear a part.
These are certain signs to know
Faithful friend from flattering foe.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	VI
LIST OF TABLES	X
LIST OF FIGURES	XI
NOMENCLATURE	XII
CHAPTER 1 INTRODUCTION	1
1.1 BACKGROUND	1
1.2 THESIS CONTRIBUTION	4
CHAPTER 2 LITERATURE REVIEW.....	7
2.1 BREAST CANCER	7
2.1.1 Estrogen Receptor-Related Signaling in Breast Tissue	8
2.1.2 Types of Breast Cancer.....	12
2.1.3 The Risk Factors in Breast Cancer	15
2.1.4 Stages of Breast Cancer	17
2.1.5 Treatment	18
2.2 MICROARRAY TECHNOLOGY	19
2.2.1 Microarrays	19
2.2.2 Analysis of Microarray Data.....	23
2.3 Application of Microarrays in Cancer Research.....	26
2.4 MICROARRAY STUDIES ON BREAST CANCER	29

CHAPTER 3	METHODS AND MATERIALS.....	32
3.1	THE MICROARRAY DATA USED.....	34
3.2	THE CLUSTERING METHOD USED	36
3.2.1	Biclustering	36
3.3	PREPROCESSING OF THE DATA	36
3.4	BICLUSTERING OF THE DATA	36
3.5	POST-PROCESSING OF THE RESULTS	36
3.5.1	Functional Enrichment Analysis	36
3.5.2	Transcriptional Factor Binding Site Analysis:	37
3.5.3	Other additional analysis	37
CHAPTER 4	TEMPORAL AND FUNCTIONAL ANALYSIS	39
4.1	GENERAL FUNCTIONAL PATTERN OF BC CELLS IN RESPONSE TO ESTROGEN STIMULATION	40
4.2	BICLUSTERING ANALYSIS OF GENE EXPRESSION DATA OF CELLS AFTER ESTROGEN STIMULATION.....	40
4.3	COMPARISON OF BICLUSTERS WITH RESPECT TO CONDITIONS	44
4.4	COMPARISON OF BICLUSTERS WITH RESPECT TO GENE CONTENT.....	49
4.4.1	Functional relationship between biclusters	54
4.4.2	The 12 hours of estrogen exposure is critical	54
4.4.3	Biclustering detects the timely transition in functions in time-series data	59
4.5	COMPARISON OF BICLUSTERING WITH HIERARCHICAL CLUSTERING.....	61
4.5.1	Timely transition in function cannot be detected by hierarchical clustering 63	
4.5.2	Similar functional enrichments are seen between the two methods.....	66
CHAPTER 5	PATHWAY ANALYSIS.....	67

5.1 Six Different barriers to develop cancer.....	67
5.1.1 Response to Estrogen Exposure: Self-sufficiency in Growth Signals.....	72
5.1.2 Response to Estrogen Exposure: Insensitivity to Antigrowth Signals	72
5.1.3 Response to Estrogen Exposure: Evasion of Programmed Cell Death and Escape from Immune System	75
5.1.4 Response to Estrogen Exposure: Limitless Replicative Potential.....	77
5.1.5 Response to Estrogen Exposure: Sustained Angiogenesis	78
5.1.6 Response to Estrogen Exposure: Tissue invasion and angiogenesis.....	79
5.2 DOWN-REGULATED AND UP-REGULATED BICLUSTERS HAVE THE SAME EFFECTS ON THE MAPK PATHWAY.....	80
 CHAPTER 6 GENE FREQUENCY AND TF ANALYSIS	84
6.1 THE OCCURRENCE FREQUENCY OF GENES IN BICLUSTERS	84
6.2 TRANSCRIPTIONAL FACTOR BINDING SITE ANALYSIS	87
 CHAPTER 7 DISCUSSION.....	90
 APPENDIX	94
 BIBLIOGRAPHY	134

LIST OF TABLES

2.1	The most common cancer types in women and men from.	8
4.1	The result of biclustering analysis of the full data set.	43
4.2	The functional categories enriched in different biclusters.	49
4.3	The GO categories enriched in different biclusters.	56
4.4	The list of genes which are present in bicluster 9 only.	57
4.5	The functional categories enriched in different hierarchical clusters	65
5.1	The Six Barriers to Develop Cancer.	69
5.2	The list of genes from different bicluster, which are related to apoptosis	75
5.3	The list of genes from different biclusters included in MAPK signaling	82

LIST OF FIGURES

1.1	One of the mechanisms of estrogen action.	2
2.1	Breast tissue.	9
2.2	Different mechanisms of estrogen action in breast tissue.	10
2.3	Production of oligonucleotide microarrays by photolithography.	19
2.4	Components required for RNA-expression analysis using GeneChip	20
3.1	Conceptual scheme of biclustering algorithm.	34
4.1	Functional enrichment results for Experiment-1.	41
4.2	Functional enrichment results for Experiment-2.	42
4.3	The bicluster vs. Time matrix.	45
4.4	The hierarchical tree for biclustering	47
4.5	Biclusters vs. Biclusters Matrix.	53
4.6	The Venn scheme of the relationship among the biclusters 6, 7 and 8.	60
4.7	The change in functional enrichment with respect to the time.	61
4.8	Clusters vs. Time Matrix.	63
4.9	The hierarchical tree for by hierarchical clustering.	64
5.1	The MAPK signaling pathway from KEGG.	70
5.2	The cell cycle map from KEGG.	71
5.3	'Role of stathmin in microtubule destabilization' from BioCarta	73
5.4	'Role of Ran in mitotic spindle regulation' from BioCarta	74
5.5	The genes from included in the toll-like receptor pathway from KEGG	77
5.6	Axon guidance pathway from KEGG	79
5.7	Adhesion molecules for endothelial cells	80
5.8	MAPK signaling pathway from KEGG	83
6.1	The role of CCND1 and RB in passage from G1 to S phase of cell cycle	85
6.2	The transcriptional factors and the biclusters in which they are enriched	87

NOMENCLATURE

BC	Breast Cancer
PWM	Position Weight Matrix
ER	Estrogen Receptor
E2	Estrogen
ERE	Eestrogen Response Elements
GO	Gene Ontology
MAPK	Mitogen Activated Proteins Kinase
TANGO	Tool for Analysis of GO enrichments
KEGG	Kyoto Encyclopedia of Genes and Genomes
OMIM	Online Mendelian Inheritance in Men
GOTM	GO Tree Machines
PRIMA	Promoter Integration in Microarray Analysis

Chapter 1

Introduction

1.1 Background

Breast cancer is the most common tumor among the women living in developed countries, and it is the most important cause of cancer-associated morbidity and mortality. In recent years, new molecular markers have been discovered as important targets for diagnosis and therapy and new drugs are discovered. Thus due to this improvements and early diagnosis, cancer-associated mortality has decreased [1, 2]. Today, different molecular factors are being used in clinical practice to predict the prognosis and response of breast cancer patients to therapy. The absence and/or expression of estrogen and progesterone receptors, up-regulation of *Ki-67* (Mib-1) [3], osteopontin [4] and mutations in the genes *BRCA1* and *BRCA2* [5, 6] are examples to such factors. Estrogen receptors (ERs) are members of the nuclear receptor (NR) superfamily which mediates the effects of the steroid hormones in a diverse range of physiological processes [7]. Estrogens have different effects on somatic and reproductive cells. They are very important for the regulation, growth, differentiation, and homeostasis of many different tissues such as breast, heart, bone, but they have also been associated pathologically with a higher risk for breast and endometrial cancer [8]. Due to these effects, some women entering menopause are prescribed to take hormone replacement therapy to decrease the unpleasant symptoms of menopause and their risk of cardiac failure and to prevent loss of bone mass. In addition,

estrogens are also included in oral contraceptives. Therefore it is very important to understand fully how estrogen acts on breast tissue both at the gene and the pathway level in breast cancer cells. ER uses mainly 3 mechanisms to exert its effects on its target genes: 1) by binding of the estrogen-bound ER to DNA at estrogen response elements (EREs) (Figure 1.1); 2) by binding of the estrogen-occupied ER to other transcription factors, such as activating protein-1 (AP-1) [9], nuclear factor- κ B (NF- κ B) etc. [10]; and 3) by effecting transcription indirectly via sequestration of other transcriptional components [11].

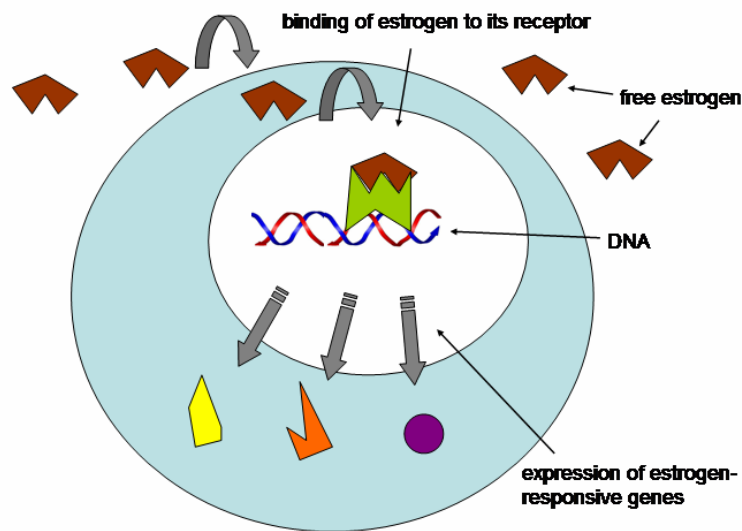


Figure 1.1. One of the mechanisms of estrogen action. Estrogen binds to its receptor, then the estrogen-bound receptor binds to the estrogen-responsive elements (EREs) found in the promoter regions of estrogen-responsive genes and leads to their expression.

To increase the knowledge about breast cancer, microarrays have been used many times in the literature, which gave valuable information on novel marker genes which may be prognostically important and the functional groups of genes which can be used to differentiate different subtypes of the disease. These works helped us differentiate breast

cancer patients with *BRCA1* and *BRCA2* tumor suppressor mutations from those with sporadic disease [12] or differentiate patients with different subtypes with respect to the outcome.

To understand the mechanisms of estrogen actions, several microarray experiments have been carried out in breast tumour cell lines to study alterations in gene-expression profiles in response to estrogen treatment [13-21]. In one of these studies, the authors presented a method to characterize estrogen sensitivity of large numbers of genes expressed in cell culture and used this method on ER α -positive human breast cancer cells to measure estrogen sensitivity of estrogen-responsive genes [14]. This study showed that biological actions of estrogen depend on its concentration since the expression of estrogen-responsive genes have significantly different estrogen sensitivities. In another microarray study, using different methods including hierarchical clustering, it was shown that a substantially greater proportion of genes involved in cell cycle and nucleotide processing, such as DNA repair or RNA splicing, were stimulated by estrogen, whereas a greater proportion of cell adhesion and extracellular matrix genes as well as enzymes were inhibited by estrogen [17]. In another study, Creighton et al. [15] performed DNA microarray-based gene expression experiments of three estrogen receptor α (ER α)-positive breast cancer cell lines stimulated by estrogen over a time course. They used a supervised clustering method and obtained 8 clusters (A-H) each one being induced or repressed at different time points. They found that in cluster B, which is up-regulated beginning at 4 hrs and cluster C, which is up-regulated beginning from 8 hrs, were enriched in functional categories related to 'RNA metabolism' and 'cell cycle and mitosis' respectively. This functional order is in good agreement with the one of the results we obtained (in section 4.4.2). Cicatiello et al. [13] performed a microarray analysis of breast cancer cells exposed to estrogen. They collected data at different time points. Analysis of the resulting time series data led to the identification of 344 differentially expressed genes. They then applied hierarchical

clustering to these genes using GeneCluster [22] and found 8 clusters corresponding to 8 main patterns of gene activation/inhibition. They showed that estrogen appears to affect a broad range of functions related to metabolic status, proliferation, survival, differentiation and resistance to stress and chemotherapy, RNA and protein synthesis and metabolism. Furthermore, by analysing the chromosomal locations of the genes, they showed that chromosome 17 is a site particularly rich in genes activated by estrogen, suggesting the likely existence of estrogen-responsive domains in the human genome.

1.2 Thesis Contribution

To understand the time course of the events that takes place as a response to estrogen treatment, time series genes expression data is utilized. The previous studies, which make use of classical clustering methods give information only about what functions are turned on or off as a response to estrogen. However, in time series data there are more information than that obtained with classical clustering methods. But with methods like hierarchical clustering, which is used in most of the previous works, it is not possible to reveal the relationship between time and function. Based on the assumption that related genes behave similarly under all conditions, clustering methods partition the genes into disjoint sets. However, the assumption may hold for situations in which the dataset contains few, homogeneous conditions but does not hold for datasets with many, heterogeneous conditions [23]. Furthermore, having a gene in one group only means that the gene is involved in one biological process only, which is usually not the case in biological systems. In most cases, genes are multifunctional. Since the same gene may be involved in more than one function, it may be grouped with a different set of genes under some conditions and with another set of genes under other conditions. Additionally, genes do not have to be active at all times meaning that they do not have to exhibit a distinctive expression pattern over all the conditions. To be able to exploit the data as best as we can, we used a different clustering method called ‘biclustering’. This method searches for subsets of genes having

similar expression patterns over subsets of conditions. Thus biclusters do not have to cover all conditions and genes can be in more than one bicluster. Because of that biclustering can capture the local expression changes better. Making use of that property, we showed that estrogen-responsive genes act in a sequential order. There is a relationship between different biclusters with respect to the time intervals in which their expression profile changes and with respect to the functions they perform. The biclusters which follow an order in time also showed a biologically very meaningful order in function. For example, a group of transcriptional factor genes were shown to be up-regulated before another group of genes which are related to functional categories such as cell cycle, protein folding and nucleic acid metabolism (in section 4.4.1). This shows that biclustering can discern the genes whose expression exhibits significant changes locally i.e. over narrower time intervals. This notion was strengthened by another example, in which the functional order and the temporal order was even more detailed (in section 4.4.2).

In addition to these results, we also analyzed the biclusters with respect to the pathways to which they are annotated and with respect to the transcriptional factor binding sites enriched in the promoter regions of genes the biclusters contain. In the pathway analysis part, we showed that estrogen affects a broad range of functions in breast cancer cells and is actually capable of eliciting enough mechanisms to overcome the 6 barriers to develop cancer proposed by Hanahan and Weinberg [24]. Focusing on a specific pathway (MAPK-Mitogen Activated Proteins Kinase signaling), we also show that down-regulated and up-regulated biclusters are in good agreement with each other. More interpretations can be made as pathway information increases. In the transcriptional factor binding site analysis, we show that there are 2 transcriptional factors whose binding sites were enriched in the promoters of the genes belonging to different biclusters: E2F and SRF. The transcription factors were complementary to the results of the functional enrichment analysis (functional and temporal order).

All these results indicate that biclustering is very useful in analyzing time series data. Although we used it on breast cancer, it can be used with any other time series data since it is very successful at grouping the genes with respect to the specific time interval(s) in which they are functionally very important.

Chapter 2

LITERATURE REVIEW

2.1 Breast Cancer

Cancer cells are defined by two heritable properties: they and their progeny 1) reproduce in defiance of the normal restraints and 2) invade and colonize territories normally reserved for others. Irrespective of the cancer type, it is the combination of these features that makes malignant tumours peculiarly dangerous, which is reflected by the fact that in the USA and other developed countries, cancer is presently responsible for about 25% of all deaths among human[1]. Thus research on this area is quite important. Particularly important for women health is the research going on on breast cancer, since breast cancer is the most common type of cancer seen among women and it is only second to lung cancer in terms of mortality (Table 2.1). Although breast cancer-associated mortality has decreased due to early diagnosis and the discovery of new drugs in the last years, research on this area is still a very hot topic.

Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple (Figure 2.1). The remainder of the breast is made up of fatty, connective, and lymphatic tissue. Most breast abnormalities are not cancerous. Still, many need to be biopsied to prove they are not cancer. Most breast lumps turn out to be fibrocystic changes, which happens just before a

menstrual period is about to begin. However some of the lumps are benign breast tumours such as fibroadenomas or papillomas, which are abnormal growths, but are not cancer and cannot spread outside of the breast to other organs. They are not life threatening. But some benign breast conditions such as papillomas and atypical hyperplasia are important because women with these conditions have a higher risk of developing breast cancer.

Table 2.1. The most common cancer types in women and men from [1].

Male-most common	Male-cause of death	Female-most common	Female
prostate cancer (33%)	lung cancer (31%)	breast cancer (32%)	lung cancer (27%)
lung cancer (13%)	prostate cancer (10%)	lung cancer (12%)	breast cancer (15%)
colorectal cancer (10%)	colorectal cancer (10%)	colorectal cancer (11%)	colorectal cancer (10%)
bladder cancer (7%)	pancreatic cancer (5%)	endometrial cancer (6%, uterus)	ovarian cancer (6%)
cutaneous melanoma (5%)	leukemia (4%)	non-Hodgkin's lymphoma (4%)	pancreatic cancer (6%)

2.1.1 Estrogen Receptor-Related Signaling in Breast Tissue

Estrogens have different effects on somatic and reproductive cells. They are very important for the regulation, growth, differentiation, and homeostasis of many different tissues such as breast, heart, bone, but as stated before they have also been associated pathologically with a higher risk for breast and endometrial cancer [8]. Estrogen exerts its effects through two main mechanisms: genomic and non-genomic. Figure 2.2 summarizes the multiple estrogen-receptor-related signal-transduction pathways, emphasizing effects associated with increased proliferation and inhibition of apoptosis.

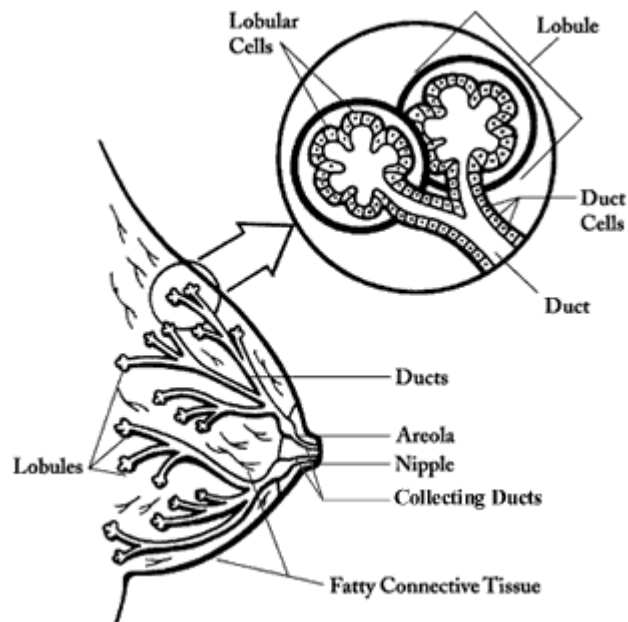


Figure 2.1. Breast tissue

Genomic action occurs in 3 different ways. The first of these mechanisms involves the binding of the hormone to nuclear estrogen receptors, which then bind as dimers to estrogen-response elements in the regulatory regions of estrogen-responsive genes. The second and the third mechanisms are the association of ligand-occupied estrogen receptor with basal transcription factors [10] and coactivators, corepressors [11] respectively to alter gene expression. Since the initial discovery and characterization of the estrogen receptor- α (ER α) in the 1960s, research on mechanisms of estrogen-receptor signaling has revealed its complexity, which is represented by the discovery of estrogen receptor- β (ER β) and of signaling pathways mediated by estrogen receptors that become associated with mitochondria and the plasma membrane [25-29]. ER α and ER β have 96 percent amino-acid identity in their DNA-binding domains, whereas there is only 53 percent homology in their ligand-binding domains, which accounts for differences in the responses of the two

receptors to various ligands. For example, tamoxifen has been reported to be both an agonist and an antagonist for ER α but only an antagonist for ER β . The receptors also differ in their activation domains, which suggests that they may recruit different proteins to the transcription complexes, thereby altering the specificity of their genomic transcriptional effects.

Also seen in Figure 2.2 is the effect of estrogen on nonnuclear genomic DNA. The presence of specific, high-affinity estrogen binding in nonnuclear subcellular fractions, including mitochondria implies that the estrogen receptor could be located

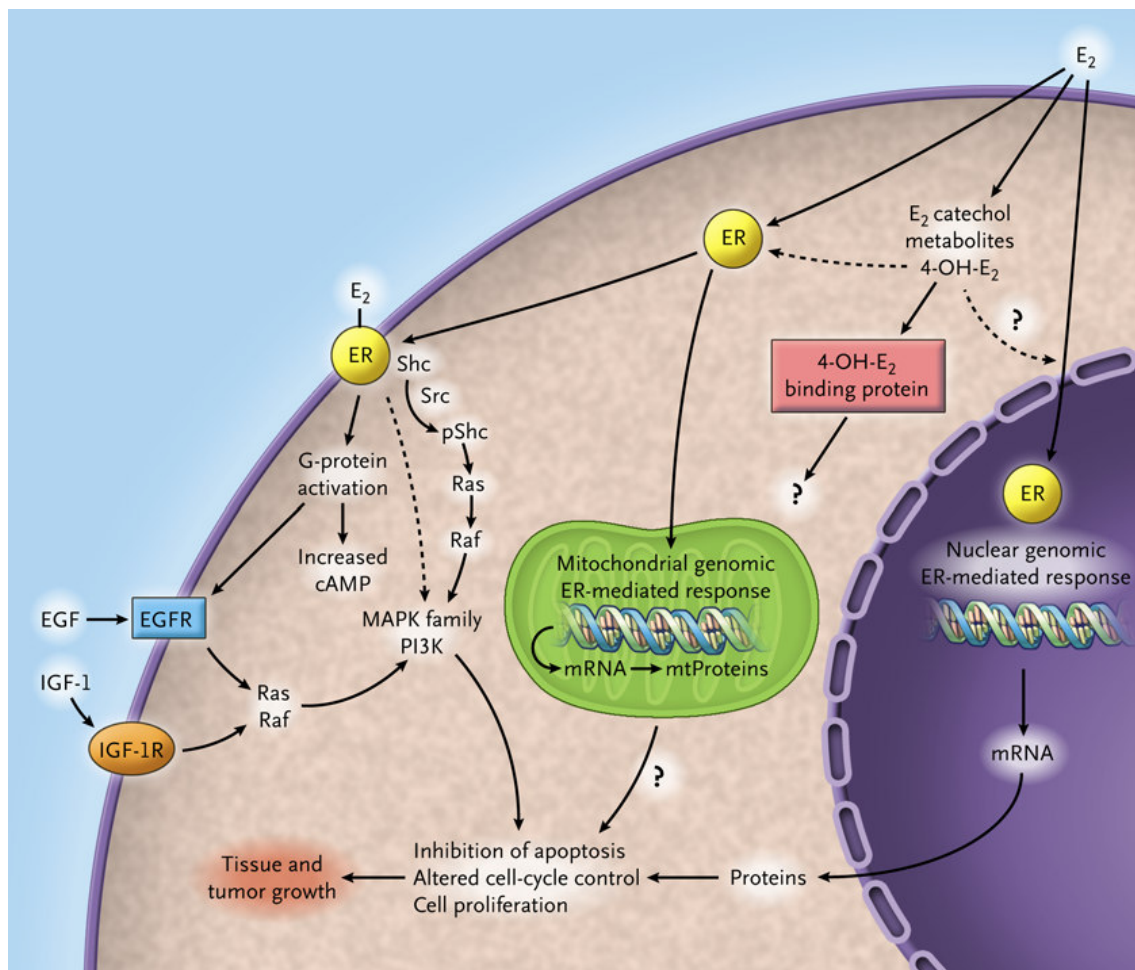


Figure 2.2. Different mechanisms of estrogen action in breast tissue.

at these sites [25]. Indeed, recent studies have definitively shown the presence of ER α , ER β , or both in mitochondria of various cells and tissues [30]. Mitochondrial genome contains potentially estrogen-responsive sequences [31, 32], and estrogen increases mitochondrial DNA-encoded gene transcript levels. The mechanisms of mitochondrial estrogen-receptor import are not known, although ER β contains a putative mitochondrial targeting peptide domain [30].

The nongenomic mechanism of estrogen action is as follows: Like peptide growth factors, estrogens also cause activation of various protein kinases, such as mitogen-activated protein kinases, and increase levels of second messengers, such as cyclic AMP (cAMP), within minutes. These nongenomic, nontranscriptional effects involve a membrane-bound form of ER α , ER β , or both [25, 26] and facilitate cross-talk between the membrane, estrogen receptor-related signaling process and other signal-transduction pathways, such as the epidermal growth factor receptor and insulin-like growth factor 1 receptor signaling pathways (Figure 2.1). [33] In addition, cross-talk between the genomic and second-messenger pathways probably has important roles in estrogenic control of cell proliferation and inhibition of apoptosis and may have implications for therapy [25, 26, 33].

2.1.2 Types of Breast Cancer

It is important to understand some of the key words used to describe different types of breast cancer. It is not unusual for a single breast tumor to be a combination of these types and to have a mixture of invasive and in situ cancer.

Adenocarcinoma: Nearly all breast cancers start in the ducts or lobules of the breast. Because this is glandular tissue, they are called adenocarcinomas, a term applied to cancers

of glandular tissue anywhere in the body. The 2 main types of breast adenocarcinomas are ductal carcinomas and lobular carcinomas.

In situ: This term is used for the early stage of cancer, when it is confined to the immediate area where it began. Specifically in breast cancer, *in situ* means that the cancer remains confined to ducts (ductal carcinoma in situ) or lobules (lobular carcinoma in situ). It has not invaded surrounding fatty tissues in the breast nor spread to other organs in the body.

Ductal carcinoma in situ (DCIS): Ductal carcinoma in situ (also known as intraductal carcinoma) is the most common type of noninvasive breast cancer. DCIS means that the cancer cells are inside the ducts but have not spread through the walls of the ducts into the surrounding breast tissue. About 20% of new breast cancer cases will be DCIS. Nearly all women diagnosed at this early stage of breast cancer can be cured. A mammogram is the best way to find DCIS early. When DCIS is diagnosed, the pathologist will consider whether an area of dead or degenerating cancer cells, called *tumor necrosis*, is present. If necrosis is present, the tumor is considered more aggressive. The term *comedocarcinoma* is often used to describe DCIS with necrosis.

Lobular carcinoma in situ (LCIS): Although not a true cancer, LCIS (also called lobular neoplasia) is sometimes classified as a type of noninvasive breast cancer. It begins in the milk-producing glands but does not grow through the wall of the lobules. Most breast cancer specialists think that LCIS itself does not usually become an invasive cancer, but women with this condition do have a higher risk of developing an invasive breast cancer in the same breast or in the opposite breast. For this reason, women with LCIS, in particular, should pay close attention to having regular mammograms.

Infiltrating (or invasive) ductal carcinoma (IDC): This is the most common breast cancer. It starts in a milk passage, or duct, of the breast, has broken through the wall of the duct, and invaded the fatty tissue of the breast. At this point, it can metastasize, or spread to

other parts of the body through the lymphatic system and bloodstream. About 80% of invasive breast cancers are infiltrating ductal carcinomas.

Infiltrating (or invasive) lobular carcinoma (ILC): Infiltrating lobular carcinoma starts in the milk-producing glands, or lobules. Similar to IDC, it also can spread (metastasize) to other parts of the body. About 10% of invasive breast cancers are ILCs. Invasive lobular carcinoma may be harder to detect by a mammogram than invasive ductal carcinoma.

Inflammatory breast cancer: This uncommon type of invasive breast cancer accounts for about 1% to 3% of all breast cancers. It makes the skin of the breast look red and feel warm and gives the skin a thick, pitted appearance. Doctors now know that these changes are not caused by inflammation or infection, but by cancer cells blocking lymph vessels or channels in the skin.

Medullary carcinoma: This special type of infiltrating breast cancer has a rather well-defined, distinct boundary between tumor tissue and normal tissue. It also has some other special features, including the large size of the cancer cells and the presence of immune system cells at the edges of the tumor. Medullary carcinoma accounts for about 5% of breast cancers. The outlook, or prognosis, for this kind of breast cancer is better than for other types of invasive breast cancer. But these are often hard to distinguish from infiltrating ductal carcinoma and are treated the same way.

Mucinous carcinoma: This rare type of invasive breast cancer is formed by mucus-producing cancer cells. The prognosis for mucinous carcinoma is better than for the more common types of invasive breast cancer. *Colloid carcinoma* is another name for this type of breast cancer.

Paget disease of the nipple: This type of breast cancer starts in the breast ducts and spreads to the skin of the nipple and then to the areola, the dark circle around the nipple. It is rare, accounting for only 1% of all cases of breast cancer. The skin of the nipple and areola often appears crusted, scaly, and red, with areas of bleeding or oozing. The woman

may notice burning or itching. Paget disease may be associated with in situ carcinoma or with infiltrating breast carcinoma. If no lump can be felt in the breast tissue and the biopsy shows DCIS but no invasive cancer, the prognosis is excellent.

Phyllodes tumor: This very rare breast tumor develops in the stroma (connective tissue) of the breast, in contrast to carcinomas, which develop in the ducts or lobules. Phyllodes (also spelled *phylloides*) tumors are usually benign but on rare occasions may be malignant.

Benign phyllodes tumors are treated by removing the mass and a narrow margin of normal breast tissue. A malignant phyllodes tumor is treated by removing it along with a wider margin of normal tissue, or by mastectomy. These cancers do not respond to the usual treatments for invasive ductal or lobular breast cancer. In the past, both benign and malignant phyllodes tumors were referred to as cystosarcoma phyllodes.

Tubular carcinoma: Tubular carcinomas are another special type of infiltrating breast carcinoma. It was named tubular because of the way the cells look under the microscope. Tubular carcinomas account for about 2% of all breast cancers and have a better prognosis than infiltrating ductal or lobular carcinomas.

2.1.3 The Risk Factors in Breast Cancer

Risk factors for breast cancer include [2]:

- **Age and Gender:** As with most cancers, age is a significant factor. In fact, 77% of new cases and 84% of breast cancer deaths occur in women aged 50 and older. More than 80% of breast cancer cases occur in women over 50. Less than 1% of breast cancers occur in men. The risk of breast cancer is clearly related to hormonal influences, but how these affect the disease and particularly types of the disease is not yet clear.
- **Genetic Factors and Family History of Breast Cancer:** Some families appear to have a genetic tendency for breast cancer. Two variant genes have been found that appear to account for this: *BRCA1* and *BRCA2*. The genes *p53* and *BARD1* also appear to be

important. Researchers have identified several other defective genes that may cause breast cancer, including *BRCA3* and *Noey2* (which is a disease inherited only from the father's side of the family). These discoveries suggest that breast cancer occurs when the body's anti-cancer surveillance and control systems, which normally get rid of abnormal cells, fail to work. The body's reduced ability to get rid of abnormal cells leads to damage that gradually accumulates. Women carrying mutated *BRCA1* and/or *BRCA2* genes start with pre-existing dysfunction of this system and have a "head start" in this damaging process. Hormones are important because they encourage cell growth. High levels of hormones during a woman's reproductive years, especially when they are not interrupted by the hormonal changes of pregnancy, appear to increase the chances that genetically damaged cells will grow and cause cancer.

- **Early Menstruation and Late Menopause:** Women who get their periods early (before age 12) or went through menopause late (after age 55) are at higher risk. Also, women who have never had children or who had them only after the age of 30 have an increased risk.
- **Oral Contraceptives (birth control pills):** Birth control pills may slightly increase the risk for breast cancer, depending on age, length of use, and other factors. No one knows how long the effects of the pill last after stopping it.
- **Hormone Replacement Therapy (HRT):** Use of HRT has been shown to increase the risk of breast cancer.
- **Obesity:** Obesity is controversial as a risk factor. Some studies report obesity as a risk of breast cancer, possibly associated with higher levels of estrogen production in obese women.
- **Alcohol Consumption:** Significant alcohol use (more than 1-2 drinks a day) has been associated with an increased risk of breast cancer.

- **Chemicals:** Some studies have pointed to exposure to estrogen-like chemicals that are found in pesticides and other industrial products as a possible increased risk of breast cancer.
- **DES:** Women who took diethylstilbestrol (DES) to prevent miscarriage may have an increased risk of breast cancer after age 40.
- **Radiation:** People exposed to radiation, particularly during childhood, may face an increased risk for breast cancer in adulthood. Especially at risk are those that received chest irradiation for prior cancers.
- **Additional Risk Factors:** Some studies show previous breast, uterine, ovarian, or colon cancer, and a strong history of cancer in the family may increase the risk for breast cancer. Such history may indicate genetic factors described above.

2.1.4 Stages of Breast Cancer

There are different stages in breast cancer (from the American Joint Committee on Cancer):

- **STAGE 0.** In Situ ("in place") disease in which the cancerous cells are in their original location within normal breast tissue. Known as either DCIS (ductal carcinoma in situ) or LCIS (lobular carcinoma in situ) depending on the type of cells involved and the location, this is a pre-cancerous condition, and only a small percentage of DCIS tumors progress to become invasive cancers. There is some controversy within the medical community on how to best treat DCIS.
- **STAGE I.** Tumor less than 2 cm in diameter with no spread beyond the breast
- **STAGE IIA.** Tumor 2 to 5 cm in size without spread to axillary (armpit) lymph nodes or tumor less than 2 cm in size with spread to axillary lymph nodes
- **STAGE IIB.** Tumor greater than 5 cm in size without spread to axillary lymph nodes or tumor 2 to 5 cm in size with spread to axillary lymph nodes

- STAGE IIIA. Tumor smaller than 5 cm in size with spread to axillary lymph nodes which are attached to each other or to other structures, or tumor larger than 5 cm in size with spread to axillary lymph nodes
- STAGE IIIB. The tumor has penetrated outside the breast to the skin of the breast or of the chest wall or has spread to lymph nodes inside the chest wall along the sternum
- STAGE IV. A tumor of any size with spread beyond the region of the breast and chest wall, such as to liver, bone, or lungs

2.1.5 Treatment

The choice of initial treatment is based on many factors. For stage I, II, or III cancers, the main considerations are to adequately treat the cancer and prevent a recurrence either at the place of the original tumor (local) or elsewhere in the body (metastatic). For stage IV cancer, the goal is to improve symptoms and prolong survival. However, in most cases, stage IV breast cancer cannot be cured.

- Surgery may consist only of breast lump removal (lumpectomy), or partial, total, or radical mastectomy, usually with the removal of one or more lymph nodes from the armpit (axilla). Special procedures to find the most likely lymph nodes to which cancer may have spread (sentinel nodes) are often used.
- Radiation therapy can be directed at the tumor, the breast, the chest wall, or other tissues known or suspected to have remaining cancer cells.
- Chemotherapy is often used to kill cancer cells that may still remain in the breast or that may have already spread to other parts of the body.
- Biologicals are an entirely new type of anti-cancer drug. Biologicals can be used alone or with chemotherapy. Trastuzumab (Herceptin) is an example of this class of drugs. It affects how cancer cells function and grow. Some 20 - 25% of breast cancers respond to trastuzumab. Trastuzumab is not chemotherapy, but it may be combined with

chemotherapy. Recent studies show that adding trastuzumab to chemotherapy or treating with trastuzumab after chemotherapy helps prevent the cancer from coming back and can make people who had *HER2*-positive breast cancer live longer.

- Hormonal therapy with tamoxifen is used to block the effects of estrogen that may otherwise help breast cancer cells to survive and grow. Most women with breast cancers which express estrogen or progesterone on their surface benefit from treatment with tamoxifen. A new class of medicines called aromatase inhibitors, such as Aromasin, have been shown to be as good as or possibly even better than tamoxifen in women with stage IV breast cancer.

2.2 Microarray Technology

2.2.1 Microarrays

A microarray can be defined as an ordered collection of microspots, each spot containing a single defined species of a nucleic acid. The microarray technique is based on hybridisation of nucleic acids. In this technique, sequence complementarity leads to the hybridisation between two single-stranded nucleic acid molecules, one of which is immobilised on a matrix [34]. There exist two variants of the chips: cDNA microarrays and oligonucleotide arrays [35-37]. Although both the DNA and oligonucleotide chips can be used to analyse patterns of gene expression, fundamental differences exist between these methods [38]. Two commonly used types of chips differ in the size of the arrayed nucleic acids. In cDNA microarrays, relatively long DNA molecules are immobilised by high-speed robots on a solid surface such as membranes, glass or silicon chips [36]. Sample DNAs are amplified by the polymerase chain reaction (PCR) and usually are longer than 100 nts. This type of arrays is used mostly for large-scale screening and expression studies. The oligonucleotide arrays are fabricated either by in situ light-directed chemical synthesis or by conventional synthesis followed by immobilisation on a glass substrate [39, 40].

Those with short nucleic acids (oligonucleotides up to 25 nt) are useful for the detection of mutations and expression monitoring, gene discovery and mapping. In the procedure of genomic analysis, both types of microarrays are exposed to a labelled sample, hybridised, and complementary sequences are determined.

Oligonucleotide chips are produced by adapting semiconductor photolithography to synthesise oligonucleotide probes in situ on glass or membrane substrate. These chips are designed and produced on the basis of sequence information alone, without the need for any clones, PCR products, DNA and so on. Probe arrays are manufactured by light-directed chemical synthesis, which combines solid-phase chemical synthesis with photolithographic fabrication techniques employed in the semiconductor industry [41, 42]. Using a series of

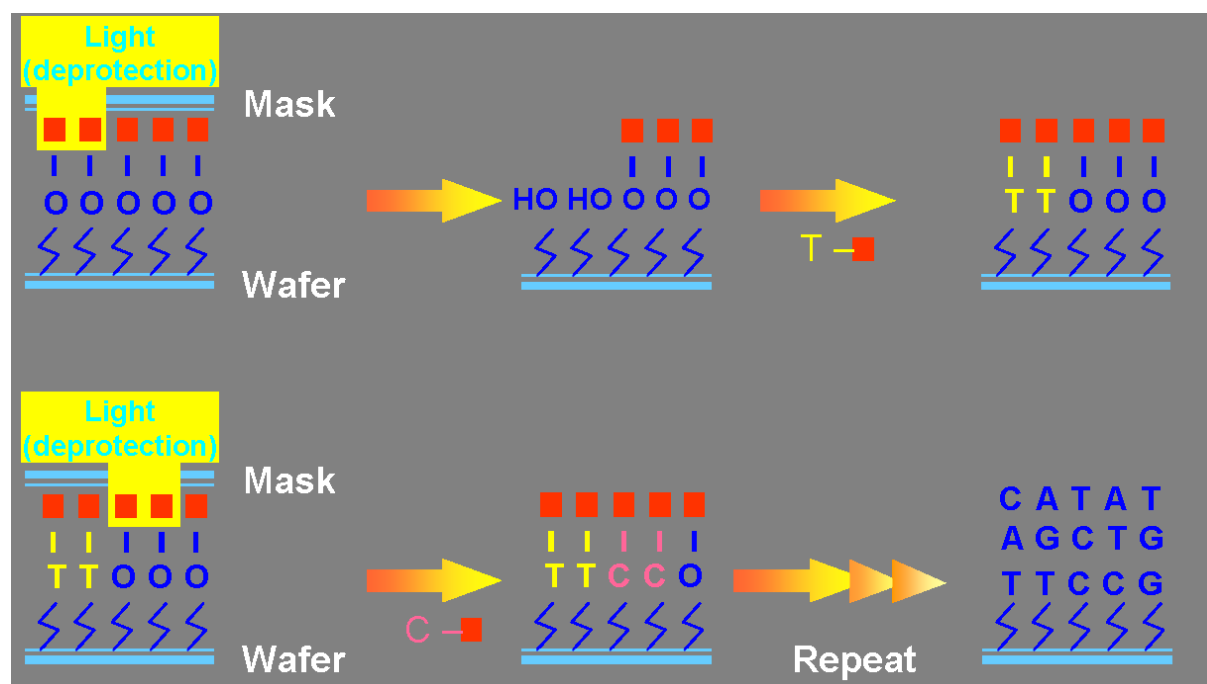


Figure 2.3. Production of oligonucleotide microarrays by photolithography. Taken from [43]

photolithographic masks to define chip exposure sites, followed by specific chemical synthesis steps, the process constructs high-density arrays of oligonucleotides, with each probe in a predefined position in the array (Figure 2.3).

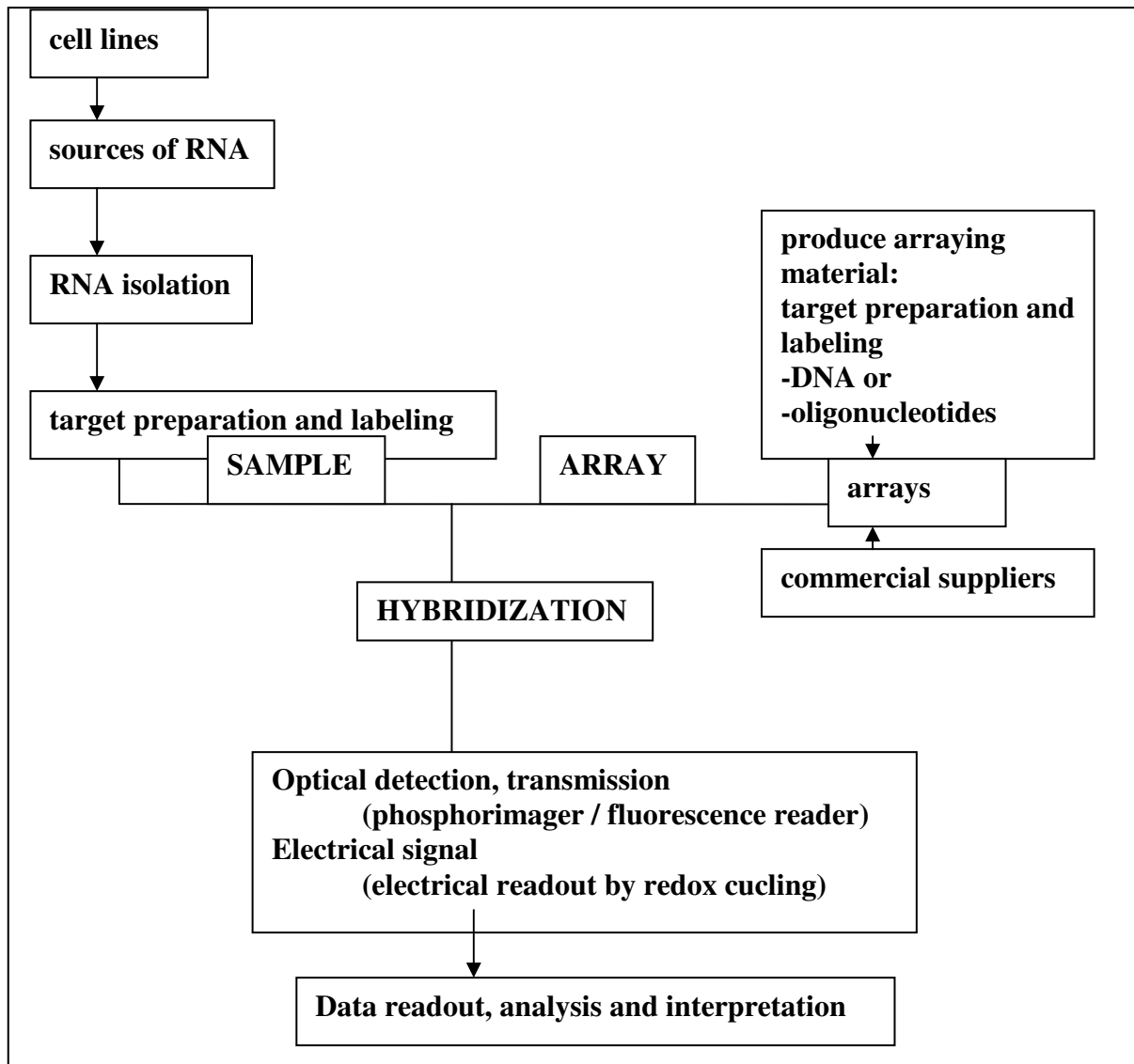


Figure 2.4. Components required for RNA-expression analysis using GeneChip-microarrays

In the case of cDNA microarrays (Figure 2.4), probes are usually generated by amplifying genomic DNA with gene-specific primers. Spotting PCR products (of approximately 0.6–2.4 kb) representing specific genes onto a matrix produces DNA arrays. These PCR products are generated using chromosomal DNA as a template, and subsequently purified by precipitation or gel-filtration, or both. For both, glass and membrane, each array dot is generated by depositing a few nanoliters of purified PCR product, usually at 100–500 µg/ml. The printing is carried out by a robot (arrayer) that spots a sample of each gene product onto a number of matrices in a serial operation. The membranes commonly used are commercially available nitrocellulose and charged nylon which are employed in standard blotting assays (Southern blot, colony and plaque blot, dot and slot blot). The disadvantages of this method are that the genetic material is non-covalently attached which may result in its loss from the support, and that only a small amount of the DNA is available for hybridization [44]. Glass-based arrays are most often made on microscope slides. They are coated with poly-lysine, amino silanes or amino-reactive silanes [36], which enhance both the hydrophobicity of the slide and the adherence of the deposited DNA. In most cases, DNA is cross-linked to the matrix by ultraviolet irradiation. After fixation, residual amines on the slide surface are reacted with succinic anhydride to reduce the positive charge at the surface. As the final step, the deposited DNA is split single-stranded by heat or alkali.

The dominant application of microarrays has been in measuring gene expression in different situations, including analysis of diseased versus normal tissues [45] profiling tumors and predicting outcomes [12, 46-53], studying gene regulation during development [54], and following the stimulation of cells in vitro [55]. Other array applications include comparative genomic hybridization [55], chromatin immunoprecipitation [56, 57], mutation detection [58], genotyping [59, 60] and microarray-mediated localized cell transfection [61].

2.2.2 Analysis of Microarray Data

The power of microarray analysis is that it is capable of distinguishing and quantifying thousands to tens-of-thousands of array elements (genes) simultaneously. It is likely that in the near future, analysis of the complete human transcriptome will be possible. This power creates one of the greatest challenges of microarray experimentation and analysis: the vast amount of data generated. There are numerous problems associated with analysis including data adjustment for background, normalization, and scaling. While these adjustments are necessary for basic analysis, there are no consensus adjustment procedures in order to make validation of microarray data analysis.

The methods used for microarray data analysis can be organized into two main categories, supervised and unsupervised learning [62]. Unsupervised approaches rely on the data set itself to contain all of the structure necessary to discriminate the relevant cohorts therein, without prior assumptions. Supervised approaches generally use “pre”-classified cohorts to identify candidate discriminating genes. For example, an investigator may wish to identify genes whose expression distinguishes two different histologically defined subtypes of a particular cancer. This set of discriminating genes is then tested by applying it to an independent set of samples to determine its predictive accuracy. There are advantages and disadvantages to both approaches. Supervised methods may not allow the discovery of novel, previously unrecognized, patient cohorts. However, many relationships may exist in complex datasets, and those that are uncovered with unsupervised approaches may not be biologically relevant. Therefore, many studies use a combination of unsupervised and supervised methodologies to determine relationships between gene expression and disease cohorts.

2.2.2.1 Unsupervised analysis

The goal of clustering is to group together objects (genes or samples) with similar properties. This can also be viewed as the reduction of the dimensionality of the system.

Clustering is not a new technique, many algorithms have been developed for it and many of these algorithms have been applied to analyze expression data. The hierarchical [63] and K-means clustering algorithms [64] as well as self-organizing maps [22] have all been used for clustering expression profiles. Even a simple clustering algorithm based on binning (i.e. discretizing the expression profile space and clustering together the profiles that map into the same bin) has been shown to be useful for clustering genes and subsequent discovering of transcription factor binding sites [65]. More recently new algorithms have been developed specifically for gene expression profile clustering, for instance based on finding approximate cliques in graphs [66].

Hierarchical clustering works by iteratively joining the two closest clusters starting from singleton clusters [63] or iteratively partitioning clusters starting with the complete set [67]. After each joining of two clusters, the distances between all the other clusters and a new joined cluster are recalculated. The complete linkage, average linkage, and single linkage methods use maximum, average, and minimum distances between the members of two clusters respectively. Note that to obtain a particular partitioning into clusters, the threshold distance should be chosen by independent means (typically by the user himself).

The K-means clustering algorithm typically uses the Euclidean properties of the vector space. The desired number of clusters K has to be chosen a priori. After the initial partitioning of the vector space into K parts, the algorithm calculates the center points in each subspace and adjusts the partition so that each vector is assigned to the cluster the center of which is the closest. This is repeated iteratively until either the partitioning stabilizes or the given number of iterations is exceeded. The approaches for the initial selection of the first set of K cluster centers can vary.

Eisen et al. [63] have developed a hierarchical clustering-based algorithm and visualization software package, which is currently one of the most frequently used tools for expression profile clustering and data visualization. They applied their software to gene

expression matrices obtained by combining 80 different yeast samples studied in various hybridization experiments at Stanford University.

Hierarchical clustering [63] has also been used for sample clustering. An interesting application of this approach is the clustering of tumors to find new possible tumor subclasses. In a recent paper by Alizadeh et al. [46], diffuse large B-cell lymphoma (DLBCL) was studied using 96 samples of normal and malignant lymphocytes. Applying a hierarchical clustering algorithm to these samples they showed that there is a diversity in gene expression among the tumors of DLBCL patients. They identified two molecularly distinct forms of DLBCL, which had gene expression patterns indicative of different stages of B-cell differentiation. Interestingly, these two groups correlated well with patient survival rates, thus confirming that the clusters are meaningful.

Sample clustering has been combined with gene clustering to identify which genes are the most important for sample clustering [46, 68]. Alon et al. [68] have applied a partitioning- based clustering algorithm to study 6500 genes of 40 tumor and 22 normal colon tissues for clustering both genes and samples. They call this method two-way clustering.

2.2.2.2 Supervised analysis

One of the goals of supervised expression data analysis is to construct classifiers, such as linear discriminants, decision trees or support vector machines (SVM), which assign pre-defined classes to a given expression profile. For instance, if a classifier can be constructed based on gene expression profiles which are able to distinguish between two different, but morphologically closely related tumor tissues, such a classifier can be used for diagnostics. Moreover, if such a classifier is based on a set of relatively simple rules, it can help to understand what the mechanisms involved in each tumor are. Typically, such classifiers are trained on a subset of data with a priori given classification and tested on

another subset with known classification. After assessing the quality of the prediction they can be applied to data the classification of which is unknown.

Brown et al. [69] have applied various supervised learning algorithms to six functional classes of yeast genes using gene expression matrices from 79 samples. Genes from some of the classes, such as ribosomal proteins and histones, are expected to be co-expressed. For these classes a good classification accuracy was achieved. Some other functional classes, such as protein kinases, are not expected to have distinct gene expression profiles. It was shown that SVM provides the best prediction accuracy for the functional classes that are expected to be co-regulated.

Golub et al. [49] applied neighborhood analysis to construct class predictors for samples, concretely for leukemias. They were looking for genes the expression of which is best correlated with two known classes of leukemias, acute myeloid leukemia and acute lymphoblastic leukemia. They constructed a classifier based on 50 genes (out of 6817) using 38 samples and applied it to a collection of 34 new samples. The classifier correctly predicted 29 of these 34 samples.

2.3 Application of Microarrays in Cancer Research

The ways microarrays are used in cancer research can be summarized in 4 parts:

2.3.1 Marker identification

Unsupervised analysis can be most useful in discovering novel markers for diagnosis. By comparing cancerous tissue to normal tissue, novel markers for cancer diagnosis have identified in the literature. Hierarchical clustering is often used to identify these markers, which include genes such as *NM23* (colon cancer) [70], *Kallikrein 11* (lung cancer) [71], *MUC1* (prostate cancer) [72], *RACK1* (gastric cancer), prostate stem cell antigen (pancreatic adenocarcinoma) [73] among others. Some of these studies have been

extended further to find the function of these genes in the specific tumor type. Identifying the presence or absence of these novel markers may assist in a more efficient diagnosis of cancer, as well as help to stage them.

2.3.2 Tumor Classification

One of the most challenging problems in tumor biology is the classification of the tumors. Thus, here also, unsupervised analysis becomes quite useful. In one landmark study, Thomas et al. [74] were able to differentiate acute myeloid leukemia from acute lymphoblastic leukemia. This distinction was a previously difficult diagnosis to make, and has significant implications for patient therapy. In addition to classifying tumors, gene expression profiling can also find new classifications for tumors, based on their molecular signatures. In one such study, a new molecular subgroup of B-cell lymphomas was identified, where the different molecular signatures corresponded to different clinical outcome [73].

Similarly, in another study, a small subset of genes derived from a previous SAGE data was used to study the gene expression signatures of ovarian cancers. Ovarian cancers are hard to interpret pathologically, and are often confused with tumors of other origins, such as colon tumors. Using this specialized array, called an OvaChip, Sawiris et al. [75] were able to distinguish ovarian samples from colon tumor samples. Such chips have also been developed to study lymphomas and other blood disorders (LymphoChip) [76] as well as non-cancer diseases such as heart disease (CardioChip) [77]. Due to their ability to efficiently subclassify tumor types, these types of specialized chips may demonstrate significant clinical advantage in the future.

2.3.3 Response to Therapy

Another challenging problem in cancer research is the difference in success of therapy among patients. It is very important to know for which patients therapy would be most advantageous. Because therapy is often expensive and quite harsh. To address this, several groups have used gene expression profiling to analyze differences between responder and non-responder patients, or sensitive cells and resistant cells. In one study, laser capture microdissection (LCM) was used to isolate cells from responder and non-responder bladder cancer samples [78]. Microarray analysis identified 14 genes that accurately predicted the response status of these patients. Other studies have used different chemotherapeutic agents to test different types of cancers to reveal common changes in gene expression [79]. Still others utilized the use of drug resistant cell lines to provide insights into the mechanics of drug resistance. In interesting study, needles containing epidermal growth factor was used ‘to attract metastatic cells in a primary breast cancer implanted as a rat allograft’ [80]. The analysis of these metastatic cells showed that they have greater drug resistance due to the downregulation of pro-apoptotic genes. In another study, the analysis of the response of colon carcinoma cells to one drug resulted in a molecular signature that could be used to predict response for a second drug which acts through a completely different mechanism [81].

While identifying responder status is very important, it is also useful to identify markers of therapeutic response to assess efficacy. In a study of gastric cancer, genes involved in the metabolism of 5-FU were found to be upregulated in both sensitive and resistant cell lines, which showed that the drug was being processed by all the tumor cells equally [82]. In this analysis, another set of genes was shown to be differentially expressed between the two types of cell lines and these genes could be used as makers of sensitivity. This demonstrated that even though cell types were processing the drug equally, other

mechanisms dictated sensitivity. Other studies have identified changes in receptor levels in response to hormone therapy such as androgen receptor levels in prostate cancer [83].

2.3.4 Outcome prediction

Perhaps one of the most interesting, and most promising, applications of gene expression profiling in cancer research is the prediction of patient survival based on their gene expression profiles. Early studies in prostate cancer showed that antibodies specific to genes identified from microarray analysis could be used to stain tissue arrays and that the expression pattern of these proteins could be linked to patient outcome [48]. Another study showed that the outcome of children with medulloblastomas depended highly on the molecular profile of those tumors [84]. In breast cancer research, two important studies showed *BRCA1* mutations carriers could be identified from other breast cancer patients [52, 85]. Furthermore, van't Veer et al. showed that outcome of primary, lymph node negative breast cancers could be predicted using a molecular signature comprised of genes involved in metastasis, invasion and cell cycle, and that this molecular prediction far outperformed the current clinical parameters [52, 85]. This observation is also true for other cancers, including gliomas [86] and lymphomas [87]. Similar determinants of outcome have been identified for melanoma, lung cancer and colon cancer.

2.4 Microarray Studies on Breast Cancer

Breast cancer was among the first malignant diseases that underwent large-scale gene expression analysis. Given the large variability between breast cancer patients with respect to prognosis and treatment response, a search for predictive and prognostic markers has been going on for a long time and clinical methods in present use include histological grade, age, steroid receptors and various markers of proliferation. However, these methods need further refinement and breast cancer was an obvious choice when the first large-scale gene expression analyses of solid malignant tumours were done. Therefore, microarrays

have been utilized in many different ways in breast cancer research. For example, it has been very useful in both classifying known subgroups of patients and uncovering previously unrecognized and often prognostically significant, subsets of different cancer types. A previous study identified distinguishing profiles that successfully differentiated known groups of patients with breast cancer, namely, those with *BRCA1* and *BRCA2* tumor suppressor mutations from those with sporadic disease [12]. Several laboratories have reported that unsupervised clustering of microarray-generated gene expression data from breast cancer consistently separates estrogen receptor positive and negative tumors into two clusters [88, 89], strongly suggesting that genome-wide gene expression analysis can indeed identify clinically important molecular markers. In addition, in one of these expression profiling studies on breast cancer, four patient subgroups, reflecting different molecular features of mammary epithelial cell biology were identified [51]. This study was extended to 5 subtypes [90]. In [85], focusing on 2 of these 5 subtypes, it was found out that distinct molecular mechanisms might have been pre-programmed at an early stage in different subtypes of breast cancer, which points out that different subtypes may require different therapeutic strategies. An alternative expression array platform identified a 70-gene ‘prognostic signature’ in a cohort of 78 young breast cancer patients [52]. This prognostic signature correctly predicted disease outcome in 83% of patients showing that this gene prognostic signature was superior to the classic prognostic factors for breast cancer such as high grade, large tumor size, invasion of vessels and negative estrogen receptor status [52, 91].

In addition to these studies, to understand the mechanisms of estrogen actions (stated above), several microarray experiments have been carried out in breast tumour cell lines to study alterations in gene-expression profiles in response to estrogen treatment [13-21]. In one of these studies, a method for characterizing estrogen sensitivity of large numbers of genes expressed in cell culture by using DNA microarray was devised and used

on ER α -positive human breast cancer cells to measure estrogen sensitivity of estrogen-responsive genes [14]. They defined the estrogen sensitivity of a gene based on three properties, namely, (i) the minimum estrogen concentration required to induce or suppress expression of the gene significantly, (ii) the estrogen concentration that saturates estrogen effects on mRNA expression, and (iii) the shape of the estrogen dose–response curve for mRNA expression (i.e., linear, convex, or concave). The study showed that biological actions of estrogen depend on its concentration since the expression of estrogen-responsive genes have significantly different estrogen sensitivities.

Cunliffe et al. [16] used hierarchical clustering to study the transcriptional effects elicited in 3 different types of breast cancer cell lines by nine regulators of growth and differentiation including estrogen, a number of antiestrogens and epidermal growth factor (EGF). The time series data included 2, 8, and 24 hrs of treatment with one of the 8 chemical agents. Gene related to the regulation of DNA and RNA metabolism and cell cycle genes, as well as cell motility and adhesion genes were enriched as a response to estrogen. When comparing the results for estrogen and EGF responses, the analysis revealed differences between these two mitogens. For example, both ligands lead to differential regulation of receptor signaling and for adhesion-related genes, but these genes were regulated in opposite directions by these two treatments.

Frasor et al. [17] used global gene expression profiling to identify patterns and time courses of genes that are either stimulated or inhibited by estrogen in ER-positive human breast cancer cells. showed that a substantially greater proportion of genes involved in cell cycle and nucleotide processing, such as DNA repair or RNA splicing, were stimulated by estrogen, whereas a greater proportion of cell adhesion and extracellular matrix genes as well as enzymes were inhibited by estrogen. They observed a general up-regulation of positive proliferation regulators, multiple growth factors, genes involved in cell cycle progression, and regulatory factor-receptor loops, and the down-regulation of

transcriptional repressors, such as Mad4 and JunB, and of antiproliferative and proapoptotic genes.

Creighton et al. [15] performed DNA microarray-based gene expression experiments of three estrogen receptor α (ER α)-positive breast cancer cell lines stimulated by estrogen over a time course. They found that there was a remarkable overlap (over 40%) of the regulation patterns of genes in microarray experiment and those obtained from xenografts. They used a supervised clustering method and found that the estrogen-regulated genes were significantly enriched for transcriptional targets of the myc oncogene and were found to be coordinately expressed with Myc in human tumors, reflecting the importance of Myc estrogen response.

Inoue et al. [18] carried out estrogen-responsive expression profiling of MCF-7 breast cancer cells. Based on the results selecting the estrogen-responsive genes, they prepared a custom-made cDNA microarray. Using the narrowed-down gene subset, they analyzed the time course of the estrogen-responsive gene expression profiles, which resulted in subdivision of the genes up-regulated by estrogen into early-responsive and late-responsive genes.

Cicatiello et al. [13] collected gene expression data at different time points over a time course of up to 32 hrs of estrogen exposure on breast cancer cells. They identified 344 differentially expressed genes and then applied hierarchical clustering to these genes using GeneCluster [22] and found 8 clusters corresponding to 8 main patterns of gene activation/inhibition. They showed that estrogen appears to affect several key features of breast cancer cells, including their metabolic status, proliferation, survival, differentiation and resistance to stress and chemotherapy, as well as RNA and protein synthesis, maturation and turn-over rates. Furthermore, by analysing the chromosomal locations of the genes, they showed that chromosome 17 is a site particularly rich in genes activated by

estrogen, suggesting the likely existence of estrogen-responsive domains in the human genome.

Chapter 3

METHODS AND MATERIALS

3.1 The Microarray Data Used

The gene expression data set by Cicatiello et al. (Series GSE1864) was downloaded from the Gene Expression Omnibus Database of NCBI (GEO) [13]. The data comprise a comparative cDNA microarray mRNA analysis of hormone-responsive ZR-75•1 and MCF-7 cells before and after stimulation with a physiological dose of 17β -estradiol (estrogen). Gene expression changes were monitored in hormone starved cells at 11 consecutive times ranging from 1 to 32 h into hormonal stimulation corresponding to a whole human cell cycle. In the time-series data obtained from the GEO database, all data points were normalized with respect to 0-th time expression value (no hormone exposure). The cDNA microarray grid contains 9182 genes of which 8372 are unique. During the experiment, the fluorescent dyes were exchanged between the two cDNAs in parallel to provide a further control. Therefore the data downloaded from GEO contains 25 arrays instead of 11 arrays for the 9182 genes. In the first 13 arrays sample RNAs were labeled with Cy5 while the control cDNAs were labeled with Cy3 (the first dying scheme, experiment-1). This array set correspond to 1 hr, 2hrs, 4 hrs, 6 hrs, 8 hrs, 8 hrs (duplicate), 12 hrs, 16 hrs, 16 hrs (duplicate), 20 hrs, 24 hrs, 28 hrs and 32 hrs of estrogen stimulation. In the last 12 arrays, the dyes were swapped meaning that the sample RNAs were labeled with Cy3 and the control cDNAs with Cy5 (the second dying scheme, experiment-2). This array set

corresponds to 1 hr, 2hrs, 2 hrs (duplicate), 4 hrs, 6 hrs, 8 hrs, 12 hrs, 16 hrs, 20 hrs, 24 hrs, 28 hrs and 32 hrs of estrogen stimulation. In this study, we performed analysis first on the data of experiment-1 and experiment-2 separately and then on the whole expression matrix (all of the 25 arrays), and compared the results. In order to find the genes underlying estrogen response, we applied biclustering to the two experiments only and to the full data obtained by merging the two experiments together. Different functional enrichment results were obtained with Experiment-1 and Experiment-2. The functionally enriched categories of experiment-1 were only protein folding, cell cycle and receptor binding (not shown). In contrast, in Experiment-2 several functional categories were enriched: like mitotic cell cycle, DNA metabolism/replication, unfolded protein binding, cell differentiation, nucleic acid binding and biopolymer metabolism between others (data not shown). Finally, merged data (including both experiments) gave results similar to experiment-2, in which the enriched functional categories were: mitotic cell cycle, DNA metabolism/replication, unfolded protein binding and biopolymer metabolism, M phase, spindle organization and biogenesis, transcription factor activity, nucleic acid binding, etc. Thus not only the results obtained with the full dataset were very close to the results obtained with experiment-2, they add further functional categories. Furthermore the functional relationship between biclusters obtained with the full dataset was biologically more meaningful, as stated below. This means that though Experiment-1 does not give as informative results as Experiment-2, it adds valuable information to the results of experiment-2 when it is merged to experiment-2. We therefore chose to perform the analysis with the full dataset in order to strengthen our analysis. It should also be stated here that functional categories similar to the full-dataset were obtained in similar studies done by other groups [7, 37, 92]. Therefore, from here on, all further study and interpretations were performed on the full dataset.

3.2 The Clustering Method Used

3.2.1 Biclustering

To identify the groups of genes which have similar expression profiles, usually different clustering methods are used, hierarchical clustering being the simplest and most widely used one. Based on the assumption that related genes behave similarly under all conditions, clustering methods partition the genes into disjoint sets. However, the assumption may hold for situations in which the dataset contains few, homogeneous conditions but does not hold for datasets with many, heterogeneous conditions [23]. Furthermore, having a gene in one group only means that the gene is involved in one biological process only, which is usually not the case in biological systems. Since the same

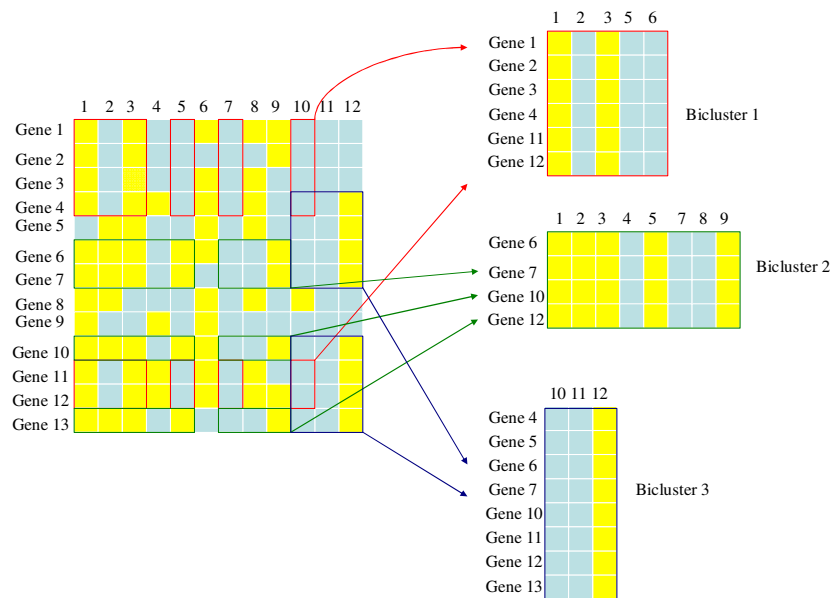


Figure 3.1. Conceptual scheme of biclustering algorithm. In this figure, there are 13 genes (the rows) and 12 conditions (the columns). Biclustering identifies subsets of genes sharing compatible expression patterns across subsets of samples. Thus, biclusters do not have to cover all conditions,

instead the method searches for subsets of genes which show similar expression patterns under a subset of conditions. In addition genes can appear in more than one bicluster. In this figure, 3 biclusters are given as examples. Genes 4, 11 and 12 are shared by biclusters 1 and 3. Conditions 1, 2, 3, 5 and 6 are shared by biclusters 1 and 2. shown which cover the Genes 1, 2, 3, 5, 7 and 9 and conditions 1, 2, 3, 5, 7 and 9 since this genes behave in the same way only over these conditions.

gene may be involved in more than one function, it may be grouped with a different set of genes under some conditions and with another set of genes under other conditions. Additionally, genes do not have to be active at all times meaning that they do not have to exhibit a distinctive expression pattern over all the conditions. Therefore instead of other classical clustering methods, to overcome such difficulties we used biclustering, which clusters the genes and the conditions simultaneously.

A bicluster is a subset of genes which have similar expression patterns under a subset of conditions (Figure 3.1). Biclustering searches for biclusters in the input expression matrix. In this method, all of the genes do not have to behave similarly under different conditions, genes can be in more than one bicluster. As stated above, this final concept may lead to biologically more meaningful results.

Biclustering was introduced in the seventies. It has first been applied to biological data by Cheng and Church [93]. Since then many different methods were proposed in the literature such as SAMBA [23], Order Preserving Submatrix Algorithm [94], Iterative Signature Algorithm [95, 96] and xMotif [97]. Some of these methods were compared with both synthetic and real gene expression data sets in another study [98]. According to this study, SAMBA, graph-theoretic approach provides good results and is slightly more robust regarding regulatory complexity. It is implemented in Expander along with other clustering algorithms.

3.3 Preprocessing of the Data

Short scripts in Python were written to process this initial data to obtain the tab-delimited file which is required by Expander. In the original data, all expression values were already normalized with respect to 0-th time expression value (no estrogen treatment). Before biclustering, genes which do not show >1.5 fold change in expression over 6 of the 25 conditions were discarded. 4097 genes (out of 4319, see below) survived after the filtering. Therefore biclustering was applied to these 4097 genes.

3.4 Biclustering of the Data

As stated above the data set contains 25 conditions, the first 13 columns belonging to experiment-1 and the second one to experiment-2. Both of these data parts correspond to a separate cell cycle. To be able to decide which of the experiments to choose, we applied biclustering to both of them and in addition to the data obtained by merging the two parts. Therefore this merged data corresponds to 2 complete cell cycles. A similar case was seen in a previous work on cell cycle [99]. In addition, there are many examples to cases in which different microarray data were combined to study gene expression on cell cycle genes [100-102].

3.5 Post-processing of the Results

After biclustering, functional enrichment analysis, pathway analysis and transcriptional factor binding site analysis were used to be able to interpret the results of the biclustering. Detailed information about each of these analysis is given in the corresponding chapter.

3.5.1 Functional Enrichment Analysis

Biological meanings of the biclusters were assessed using two different functional enrichment analysis tools: TANGO, which is implemented in Expander, and GOTM [103]. Both methods search for functional enrichments from GO categories for the genes in the

given biclusters. In order to be able to use TANGO and GOTM, Locus Link IDs (GeneBank gene ids) for all the genes had to be obtained. To do that, Gene2accession file was downloaded from <ftp://ftp.ncbi.nih.gov/gene/DATA/>. This file contains all the protein, RNA and genomic accession numbers associated with each gene ID. Using this file, the gene ids which correspond to the cDNA clones on the microarray were extracted. For 4319 genes (out of 8372) the Locus Link IDs could be extracted. That is the reason why 4319 entered filtering process (see above). TANGO algorithm was used with default parameters with all genes as the background set.

3.5.2 Transcriptional Factor Binding Site Analysis:

Genes which exhibit similar expression profiles may be transcriptionally related to each other. The reason for that might be that these genes are under the control of same transcription factors. Therefore analyzing the transcriptional factor binding sites in the promoter regions of genes may give valuable information about the underlying regulatory mechanisms responsible for estrogen response in breast cancer cells. For the transcriptional factor binding site analysis PRIMA algorithm implemented in Expander was used with default parameters. Detailed information about this analysis is given in the chapter 6.

3.5.3 Other additional analysis

Observing the functional and transcriptional site-related properties of the resulting biclusters is very useful for their interpretation. In addition, biclustering can provide more information about the interdependences of different biclusters and the relationship between them, since genes can be in more than one bicluster. However, this final advantage makes the analysis of the resulting biclusters more difficult. In addition to analyzing each bicluster separately from the others, one should observe other patterns seen in the results such as the frequency of occurrence of genes among different biclusters, relationship between the

biclusters and the time points covered by them, etc. To utilize the results as much as possible different matrices were generated, which will be explained in detail in chapter 4.

Chapter 4

TEMPORAL AND FUNCTIONAL ANALYSIS OF BICLUSTERING RESULTS

In this chapter, first the general results obtained using biclustering are discussed. Then the resulting biclusters are compared with respect to their expression behaviours under the conditions they cover and with respect to the genes they contain. Two case studies are given as examples to show how biclustering is helpful in analyzing gene expression time series data. In the last part of this section, a comparison between biclustering and hierarchical clustering is performed.

4.1 General functional pattern of BC Cells in Response to Estrogen Stimulation

In order to find the genes underlying estrogen response, time series gene expression data of breast cancer cells exposed to estrogen were analyzed. The data set used in this study contains two repetitions of the same gene expression experiment, Experiment-1 and Experiment-2, each one giving a whole cell cycle. We applied biclustering to the experiments only and to the full data obtained by merging the two experiments together. Different functional enrichment results were obtained with experiment-1 and experiment-2. The functionally enriched categories of Experiment-1 were only protein folding, cell cycle, mitotic cell cycle and receptor binding (Figure 4.1). In contrast, in Experiment-2 several functional categories were enriched: like mitotic cell cycle, DNA metabolism/replication, unfolded protein binding, cell differentiation, nucleic acid binding and biopolymer

metabolism between others (Figure 4.2). Finally, the merged data (including both experiments) gave results similar to Experiment-2, in which the enriched functional categories were: mitotic cell cycle, DNA metabolism/replication, unfolded protein binding and biopolymer metabolism, M phase, spindle organization and biogenesis, transcription factor activity, nucleic acid binding, etc. Thus the results obtained with the full dataset were very close to the results obtained with Experiment-2, but also they added further functional categories. Furthermore the functional relationship between biclusters obtained with the full dataset was biologically more meaningful, as stated below. This means that though Experiment-1 does not give as good results as Experiment-2, it adds valuable information to the final results when combined with Experiment-2. We therefore chose to perform the analysis with the full dataset in order to strengthen our analysis. It should also be stated here that functional categories similar to those obtained with the full-dataset were obtained in similar studies done by other groups [7, 37, 92]. Therefore, from here on, all further

4.2 Biclustering analysis of gene expression data of cells after estrogen stimulation

Cicatiello et al. [13] showed that in addition to the genes belonging to the transcriptional program of the cell division cycle of eukaryotes, genes which affect several key features of BC cells were also differentially expressed. Therefore estrogen-responsive transcriptome represents more than a simple cell cycle response, which was confirmed again in this study. We used biclustering, which divides the genes into joint subsets of genes and conditions, i.e. biclusters. Therefore each bicluster covers only a subset of both the genes and the conditions. Table 4.1 gives the 20 resulting biclusters and their overall response to estrogen treatment. When the expression profiles of individual clusters are analyzed, we observe that ten of these biclusters show down-regulation (1-5, 16-20) while the rest are up-regulated. The second column shows how many of the conditions

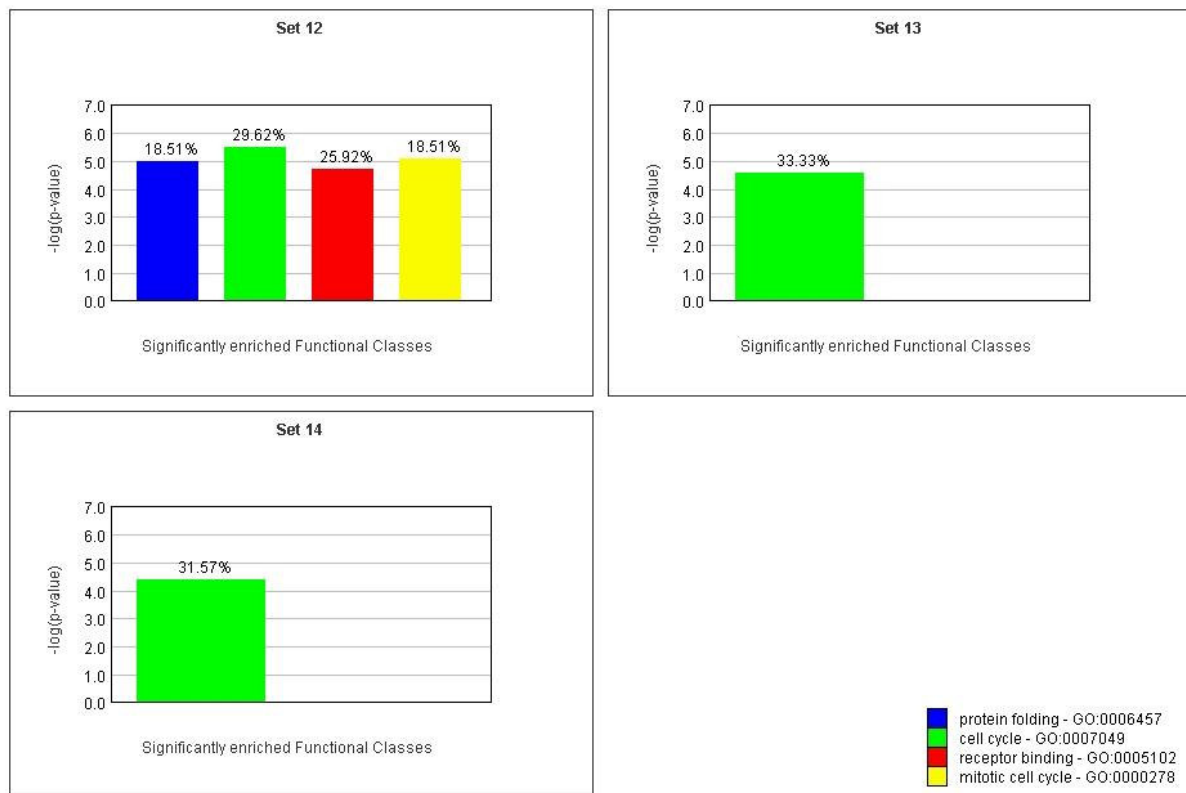


Figure 4.1. Functional enrichment results for Experiment-1. There were 14 biclusters, only 3 (biclusters 12, 13 and 14, represented by Set 12, 13 and 14 in the figure) of which were found to be enriched in 4 GO categories at most. The color codes are in the lower part of the figure.

study and interpretations were performed on the full dataset.

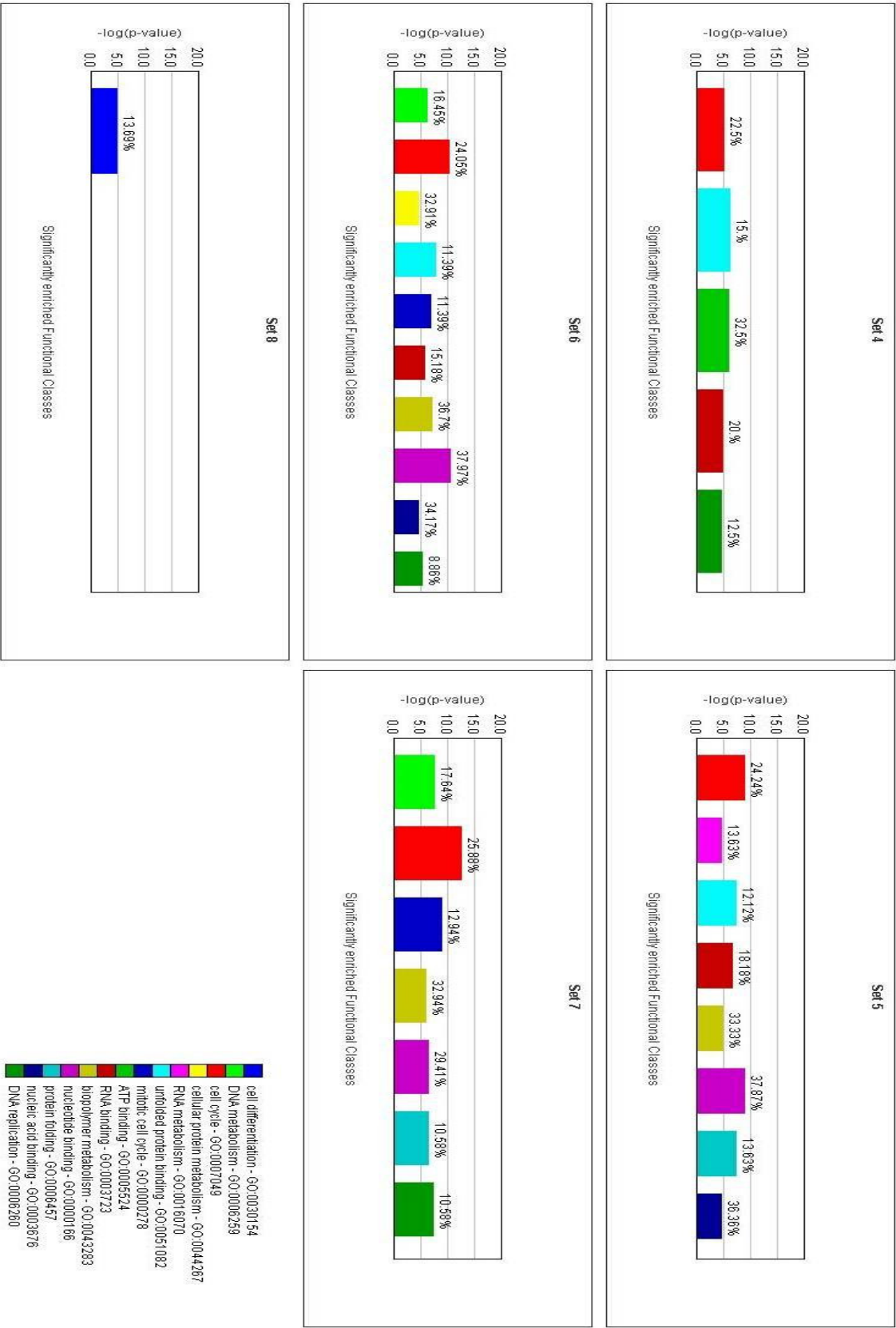


Figure 4.2. Functional enrichment results for Experiment-2. There were 10 biclusters, 6 (biclusters 4, 5, 6, 7 and 8, represented by Sets 4, 5, 6, 7 and 8 in the figure) of which were found to be enriched in 14 GO categories. The color codes are in the lower part of the figure.

(out of 25) were used in the formation of the biclusters. The third column is the number of genes in the cluster. The most crowded cluster contains 121 genes which display similar patterns over seven conditions (time points). Among the up-regulated and down-regulated genes, many were known to be stimulated or repressed by estrogen from previous studies: Examples to the up-regulated ones *TFF1* [104], *PDZK1* [105], *TPD52L1* [18], *SLC7A5* [20], *IGFBP4* [106], *B4GALT1* [18], *Cathepsin* [107], *CDC2* [17], *CDC20* [17] and *KLF10* [18]. Examples to the down-regulated ones *IGFBP5* [108], *ENO1* [18], *HSPD1* [109], *NME1* [109] and *ESR* [18], *STC2* (Stanniocalcin 2) [18].

Table 4.1 The result of biclustering analysis of the full data set. The fifth column gives the general expression direction of the bicluster. The sixth column gives the number of genes which are unique to the corresponding bicluster, i.e not shared with others. The last column gives the number of biclusters with which genes are shared. The number of genes shared with each bicluster changes from one bicluster to another. The list of biclusters when merged data set was used.

Biclusters	Conditions	Genes	Behaviour	Number of Unique Genes	Number of Biclus.
1	10	9	Down-regulated	None	9
2	22	10	Down-regulated	None	9
3	11	47	Down-regulated	7	9
4	10	66	Down-regulated	14	9
5	8	81	Down-regulated	17	8
6	9	58	Up-regulated	17	9
7	8	71	Up-regulated	3	9
8	7	121	Up-regulated	9	9
9	7	27	Up-regulated	20	9
10	18	34	Up-regulated	4	9
11	5	39	Up-regulated	9	9
12	10	71	Up-regulated	4	9
13	10	99	Up-regulated	None	9

14	12	70	Up-regulated	5	9
15	9	110	Up-regulated	6	9
16	9	39	Down-regulated	34	7
17	6	119	Down-regulated	15	9
18	7	93	Down-regulated	7	9
19	12	64	Down-regulated	8	8
20	7	28	Down-regulated	3	9

4.3 Comparison of Biclusters with Respect to Conditions

As stated before, a bicluster is a subset of genes with similar expression patterns over a subset of conditions. To able to interpret the results of the biclustering, we compared the biclusters in two different ways: with respect to the genes they cover and with respect to the conditions. First, to be able see the expression behaviour of the biclusters over the time points they cover, we generated the Bicluster vs. Time matrix (Figure 4.3). Bicluster vs. Time matrix is an $N \times T$ matrix, where N is the number of biclusters and T is the number of time points. The i, j -th entry is 1 if the i -th bicluster is up-regulated in the j -th time point, -1 if the i -th bicluster is down-regulated in the j -th time point and is 0 if the i -th bicluster does not include the j -th time point. Down-regulation and up-regulation are indicated by the blue and red colors respectively. The green color indicates that the expression of the genes in the corresponding bicluster do not coincide with each other, therefore the corresponding bicluster does not contain that condition (time point).

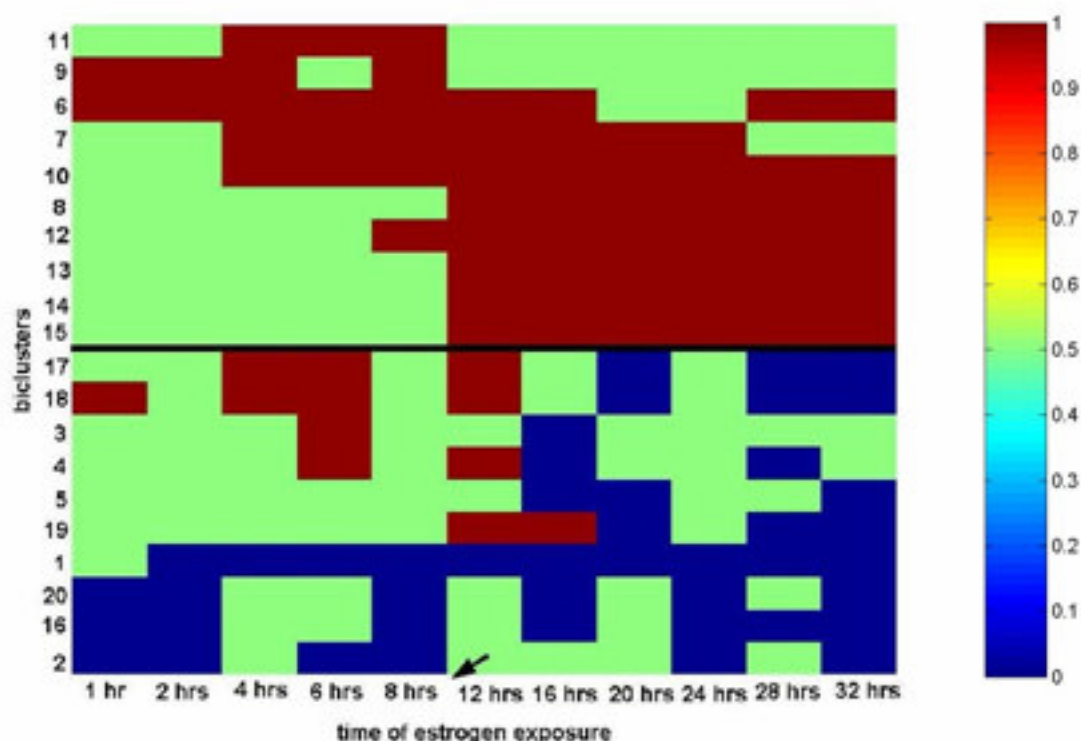


Figure 4.3. The biclusters vs. time matrix (of Experiment-2, see methods and materials for details). The original data set contained 25 conditions. All the 25 arrays were included in biclustering analysis. Since in half of the arrays the dyes were swapped to provide a further control, there were two time vs. biclusters matrices. However, to avoid misunderstanding here only the second experiment is shown. Red color indicates the up-regulation of the corresponding bicluster in the corresponding time point. Blue color indicates down-regulation. Green indicates the absence of that time point in the corresponding bicluster, which means that the genes in the bicluster do not have similar expression patterns in these time points and thus that time point is not included in the bicluster.

As can be seen from the figure, none of the biclusters cover all of the conditions. Moreover, biclusters share only some of the conditions they cover with the other biclusters. There are some expression patterns that can be deduced from this matrix. Some of the biclusters (namely, bicluster 9 and 11) cover time points before 12 hrs of estrogen exposure only while some (such as biclusters 8, 13, 14 and 15) cover time points after 12 hrs, which

may indicate a relationship between the time points and the functions of the genes included in the biclusters which cover them. Further interpretations and analysis on this is done in the following sections.

In order to be able to compare the biclusters better with each other, Bicluster vs. Time matrix (Figure 4.3) was hierarchically clustered using the hierarchical clustering function in MatLab, which resulted in the tree structure in Figure 4.4. This tree shows the closeness of different biclusters to each other with respect to their behaviour (up- or down-regulation) in the conditions they cover. Table 4.2 gives the functional categories enriched in different biclusters. As seen in Table 4.2, the functional categories enriched in these biclusters were in agreement with the results of Cicatiello et al. The enriched functional categories support the idea that response to estrogen treatment represents more than more than a simple cell cycle response, but affects different properties of BC cells including the general metabolic status. Cicatiello et al. [13] obtained 8 clusters using hierarchical clustering. In biclustering analysis, which gave 20 biclusters, genes from different hierarchical clusters are grouped into the same bicluster and most of the time this new mode of grouping together gives biologically more meaningful gene sets as discussed below. For example, three microtubule-related genes, *MAD2L1*, *AURKA* and *PTTG1*, all of which are in the 8th cluster of the Cicatiello et al. are known to interact with *CDC20*, which is in the 7th cluster of the Cicatiello et al. In our analysis, all of them are in the same bicluster, 8th bicluster. Moreover, there are other genes which are found to be involved in microtubule-related functions from OMIM: *DLG7* (absent), *KIF22* (cluster 8), *BRRN1* (absent), *XPOT* (cluster 6), *MAD2L1* (cluster 8), *CDC20* (cluster 7), *PTTG1* (cluster 8), *KPNA2* (cluster 8), *CENPA* (cluster 7), *KIF2C* (cluster 8), *KIF20A* (cluster 8). All these genes except *DLG7* and *BRRN1* are scattered to clusters 6, 7 and 8 in the hierarchical analysis of Cicatiello et al. [13]. Indeed two of them, *DLG7* and *BRRN1* are absent from their results, meaning that the initial filtering procedures in their methods labelled these 2

genes as not being differentially expressed. However in biclustering they are grouped together in bicluster 8. Another example is the splicing-related genes: *SFRS1*, *SNRPF*, *SFB3B3* are in cluster 7, cluster 6 and cluster 8 of Cicatiello et al., while all of them are in bicluster 8 in this analysis. Moreover another splicing-related gene, which was not found to be differentially regulated by Cicatiello et al., is found to be regulated in the same bicluster as the other microtubule-related genes. Examples can also be given from down-regulated biclusters. *ERBB3* and *CITED2* are in the cluster 1 and cluster 2 of Cicatiello et al. respectively. *CITED2* is necessary for *TFAP2* transactivation, and *Cited2*^{-/-} mice embryos have reduced expression of the *TFAP2* target gene *ErbB3* in the neural crest [110]. Therefore, one would expect them to be down-regulated in the same bicluster. This is the case in this analysis, both *CITED2* and *ERBB3* are down-regulated in bicluster 19.

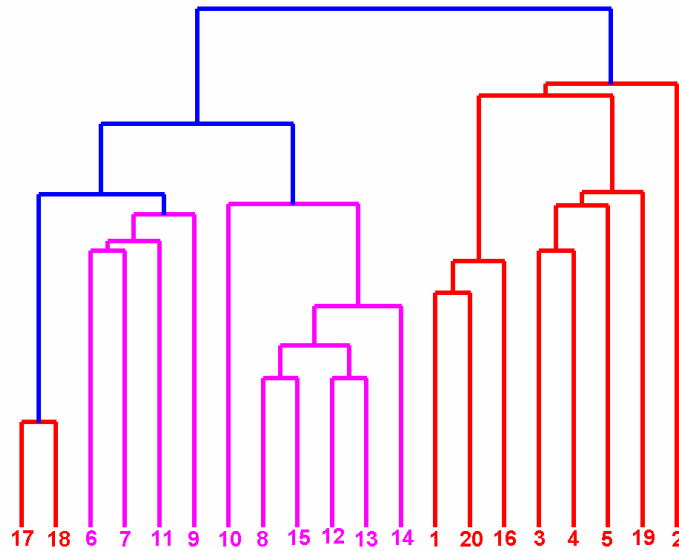


Figure 4.4. The tree obtained by hierarchical clustering of the matrix in Figure 4.3. The tree shows the similarity between the biclusters with respect to the time points they comprise. Shown in purple

are the up-regulated biclusters and in red are the down-regulated ones. Following is the functional categories related with each bicluster are given in Table 4.2.

In Figure 4.4, down- and up-regulated biclusters are separated from each other, they are on different parts of the tree. All down-regulated biclusters except biclusters 17 and 18 are grouped together. The reason why bicluster 17 and 18 are grouped together with the up-regulated ones is because these 2 biclusters have somewhat different expression profiles in before 12 hours of estrogen exposure. Since biclustering is able to catch local behaviors, these down-regulated biclusters seem more similar to the up-regulated ones. Among the up-regulated biclusters there are two groups: the first one which consists of biclusters 6, 7, 11 and 9, and the second one consists of biclusters 8, 15, 12, 13, 14 and 10. First group corresponds to those biclusters which are up-regulated before 12 hrs of estrogen exposure. The second group on the other hand contains biclusters which are up-regulated after 12 hrs generally, except the bicluster 10. This bicluster, which is the most different among biclusters 8, 15, 12, 13, 14 and 10, is like a transitory bicluster between the first group (biclusters 6, 7, 9 and 11) and the second, (8, 12, 13, 14, 15). In the first group, functional categories like protein folding, DNA replication, amino acid activation, ribosome biogenesis, RNA processing, regulation of cell growth, regulation of transcription are seen, while the second group is enriched for more specific functional categories such as G2/M transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis. Therefore it can be said that biclustering separates the genes with respect to the time interval in which they are differentially expressed and the resulting functional assignments to time intervals makes biological sense.

4.4 Comparison of Biclusters with Respect to Gene Content

In the previous part, we analyzed the biclusters with respect to the conditions they cover. However, biclusters consist of conditions and genes. Therefore to be able to compare the biclusters fully, we have to take into account the genes they contain.

Table 4.2. The functional categories enriched in different biclusters. First column contains the bicluster names and next it the upward and downward arrow indicating up- and down-regulation. If there are any functional categories shared by different clusters this is indicated as a separate row

BICLUSTERS	FUNCTIONAL CATEGORIES
bicluster 1 ↓	no functional enrichment
bicluster 2 ↓	no functional enrichment
bicluster 3 ↓	steroid receptor signaling pathway, JAK-STAT signaling cascade, apoptotic program, hexose metabolism
bicluster 4 ↓	axon guidance
bicluster 5 ↓	induction of programmed cell death, regulation of cell migration
bicluster 17 ↓	cell-mediated immune response, I-kappaB kinase/NF-kappaB cascade, regulation of protein biosynthesis, regulation of phosphate metabolism, regulation of protein kinase activity
bicluster 18 ↓	cell-mediated immune response, I-kappaB kinase/NF-kappaB cascade, regulation of protein biosynthesis, regulation of phosphate metabolism, regulation of protein kinase activity, cytokine biosynthesis and metabolism, regulation of protein import into nucleus, retrograde vesicle-mediated transport, Golgi to ER, interleukin-6 biosynthesis
biclusters 17 and 18 ↓	cytokine biosynthesis and metabolism, regulation of protein import into nucleus, retrograde vesicle-mediated transport, Golgi to ER, interleukin-6 biosynthesis
bicluster 19 ↓	→ apoptotic program, hexose metabolism, fatty acid oxidation, main pathways of carbohydrate metabolism

bicluster 20 ↓	no functional enrichment
bicluster 6 ↑	protein folding, DNA replication, ribosome biogenesis, RNA processing, rRNA metabolism, ribosome biogenesis, RNA processing, rRNA metabolism
bicluster 7 ↑	protein, ribosome biogenesis, RNA processing, rRNA metabolism, folding, DNA replication, carboxylic acid transport
biclusters 6 and 7 ↑	protein folding, DNA replication
biclusters 6, 7, 10 and 11 ↑	ribosome biogenesis, RNA processing, rRNA metabolism
bicluster 8 ↑	biogenic amine biosynthesis, amino acid activation, polyamine metabolism, amino acid derivative biosynthesis, nucleoside monophosphate metabolism, tRNA metabolism, translation, microtubule-based processes, DNA recombination, G2/M transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis
bicluster 9 ↑	regulation of cell growth, regulation of transcription, carboxylic acid transport
bicluster 10 ↑	ribosome biogenesis, RNA processing, rRNA metabolism, rRNA metabolism
bicluster 11 ↑	ribosome biogenesis, RNA processing, rRNA metabolism
bicluster 12 ↑	phosphoinositide-mediated signaling, folic acid and derivative biosynthesis, DNA ligation, negative regulation of programmed cell death, regulation of apoptosis, ribosome biogenesis, protein folding, DNA repair, DNA replication, rRNA metabolism, G2/M transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis
bicluster 13 ↑	regulation of apoptosis, ribosome biogenesis, protein folding, DNA repair, DNA replication, rRNA metabolism, microtubule-based processes
bicluster 14 ↑	regulation of apoptosis, negative regulation of programmed cell death, microtubule-based processes, sterol biosynthesis, isoprenoid biosynthesis, protein folding, DNA replication, DNA recombination, G2/M

	transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis
bicluster 15 ↑	microtubule-based processes, DNA recombination, G2/M transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis
biclusters 8 and 15 ↑	microtubule-based processes, DNA recombination
biclusters 12 and 13 ↑	regulation of apoptosis, ribosome biogenesis, protein folding, DNA repair, DNA replication, rRNA metabolism
biclusters 8, 12, 13, 14 and 15 ↑	G2/M transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis
only bicluster 10 ↑	ribosome biogenesis, RNA processing, rRNA metabolism, amino acid transport
only bicluster 14 ↑	regulation of apoptosis, negative regulation of programmed cell death, microtubule-based processes, sterol biosynthesis, isoprenoid biosynthesis, protein folding, DNA replication, DNA recombination
only bicluster 13 ↑	microtubule-based processes
only bicluster 12 ↑	phosphoinositide-mediated signaling, folic acid and derivative biosynthesis, DNA ligation, negative regulation of programmed cell death
only bicluster 15 ↑	DNA repair, DNA replication, mRNA processing, microtubule-based movement, regulation of apoptosis
only bicluster 10 ↑	rRNA metabolism
only bicluster 9 ↑	regulation of cell growth, regulation of transcription, carboxylic acid transport
only bicluster 8 ↑	biogenic amine biosynthesis, amino acid activation, polyamine metabolism, amino acid derivative biosynthesis, nucleoside monophosphate metabolism, tRNA metabolism, translation
only bicluster 7 ↑	carboxylic acid transport

To this aim, we generated the Bicluster vs. Bicluster matrix (Figure 4.5). Short scripts were written in Python to generate this matrix. Bicluster vs. Bicluster matrix is generated

using another matrix M , an $N \times N$ matrix, where N is the number of resulting biclusters (which is 20 in this case.). The element $M(i,j)$ is the number of genes which are shared by the i -th and j -th biclusters. Then M matrix is processed by using the formula; $B(i,j) = [M(i,j)]^2 / [M(i,i) \times M(j,j)]$, where $B(i,j)$ is the corresponding element in the Bicluster vs. Bicluster matrix. As a result each element in the Bicluster vs. Bicluster matrix is a number between 0 and 1, where 0 (blue color in the figure) indicates that they share no genes and 1 indicates that the two biclusters share all of their genes. Figure 4.5 shows the resulting Bicluster vs. Bicluster matrix, which compares 20 biclusters with respect to the genes they contain. Red color indicates that the corresponding biclusters share all of their genes, while blue color indicates that there are no genes shared by the two biclusters. In this figure, the matrix was re-arranged so that biclusters which share more genes could be identified more easily.

As seen in the figure, 20 biclusters fall into either of 2 categories, one consisting of the 1st, 2nd, 3rd, 4th, 5th, 16th, 17th, 18th, 19th and 20th biclusters (the upper-left square) and the other consisting of the rest (the lower-right one). The first category corresponds to the biclusters which have been down-regulated when the cells are exposed to estrogen and the second to the up-regulated ones. The up-regulated and down-regulated biclusters do not have any genes in common. Additionally, the up-regulated biclusters share more genes between each other than the down-regulated biclusters.

As seen from Figure 4.5, some biclusters are more similar to each other (shares genes) while others have totally different genes. For example, bicluster 8 (contains 7 conditions and 127 genes) shares an important number of genes with bicluster 12 (contains 10 conditions and 72 genes), bicluster 13 (contains 10 conditions and 99 genes), bicluster 14 (12 conditions and 70 genes) and bicluster 15 (contains 9 conditions and 108 genes) while it shares only 4 genes with bicluster 9.

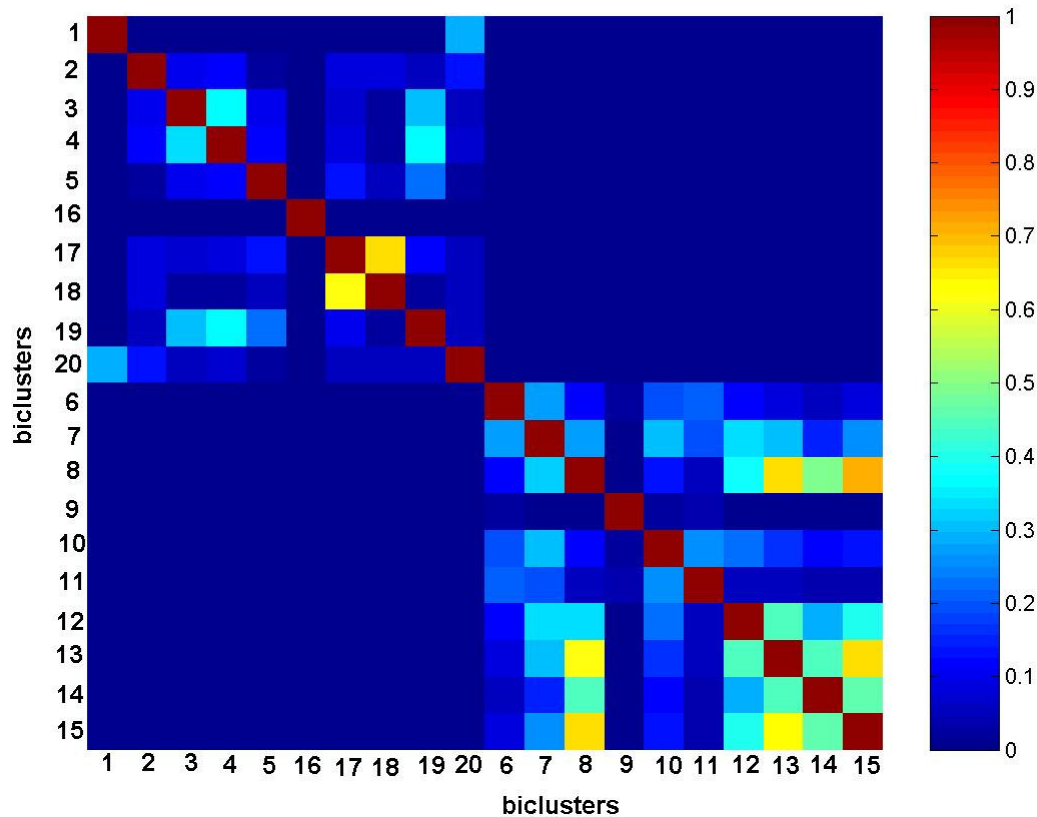


Figure 4.5. Biclusters vs. Biclusters Matrix: 20x20 matrix summarizing the comparison of the biclusters with respect to the genes they contain by color from red (contains common genes) to blue (contains different genes). The diagonal gives the comparison of the biclusters with themselves. Every entry in the matrix is calculated according to the formula given in post-processing part of the methods. The value of the entries are between 1 and 0, 1 indicating that two biclusters all of their genes and 0 indicating they do not share any genes.

The same thing can be seen in Figure 4.4: Bicluster 8 is closer to the biclusters with which it shares genes (on the same part of the tree) than bicluster 9 (on a different part of the tree). This agreement between the two matrices, Bicluster vs. Time (Figure 4.3) and Bicluster vs. Bicluster (Figure 4.5) indicates that there is a strong relationship between the functions of the genes and the time points in which the genes are coregulated.

4.4.1 Functional relationship between biclusters

Since biclustering focus on local expression patterns more than the global expression, it is able to identify genes whose functions are intensified over some time points only. Therefore in this way, we can attribute specific group of genes to specific functions and specific time points. Analysis of different groups of genes showed that genes are ordered both in time and function. In the following part, two examples are given to show the functional relationships between biclusters and sequential up-regulation of the genes included in our biclusters. In the first case study, biclusters 9 is compared with the large group formed by biclusters 8, 12, 13, 14 and 15. The common genes between these two groups are very few and the time points they cover do not overlap (before and after 12 hours). The second case study on the other hand makes the same analysis on bicluster 6, 7 and 8; which share a high number of genes and the time points of which are overlapping. Each of these cases give a good example to how biclustering can give interesting results with time series data and provide insights into how different functional categories of genes act in order to lead to the final changes which are visible at cellular level.

4.4.2 The 12 hours of estrogen exposure is critical

As seen from Figure 4.4, the up-regulated biclusters (which have red color) are divided further among themselves. The 12 hours of estrogen exposure (indicated by an arrow in the figure) seems to be important [13] because some biclusters such as 9 and 11 have values before this second time only, while others such as 8, 13, 14 and 15 are up-regulated afterwards while they do not exhibit any similar pattern before that time point. Moreover, biclusters 9 and 11 are closer to each other in the tree in Figure 4.4 whereas biclusters 8, 12, 13, 14 and 15 are closer to each other. In addition to that, the first 2 biclusters share more genes with each other than they share with biclusters 8, 12, 13, 14 and 15 (Figure 4.5). All these clearly indicate that 12 hrs of estrogen exposure is very important since this is the time that separates the two groups of biclusters from each other.

This can be explained by the previously stated notion that 12 hrs corresponds to the S and G2/M transition of cell cycle . The importance of this time is further highlighted by the following idea: The up-regulation of genes in bicluster 9 before this time point is required for the up-regulation of a second group of genes (biclusters 8, 12, 13, 14, 15; see below). Furthermore, the genes in bicluster 9 (Table 4.4) are functionally enriched in categories like regulation of cell growth, regulation of transcription and transcription from RNA polymerase II promoter (Table 4.3). Bicluster 9 contains many immediate-early genes such as *EGR1*, *EGR2*, *EGR3* (these three are C2H2-type zinc finger proteins), *FOS* and *FOSL2*. On the other hand, the intersection of the biclusters 8, 12, 13, 14 and 15 is enriched in functional categories related to M phase of the cell cycle such as G2/M transition of cell cycle, and regulation of apoptosis, protein folding (Table 4.3). A similar functional pattern was seen in a previous study about temporal gene expression changes in rat uterus as a response to estrogen [92]. In this study, the clustering analysis based on the self-organizing map algorithm revealed that there are at least two phases of gene activation. One contained immediate-early genes, which included certain transcription factors and growth factors as well as oncogenes. The other contained early-late genes, which included genes related to RNA and protein synthesis. In the case of breast cancer cells, when the cells are exposed to estrogen, the accumulation of the products of the genes in bicluster 9 takes place in the first 12 hours. It could be that their up-regulation leads to the up-regulation of the second group (8, 12, 13, 14 and 15 biclusters). This may indicate a regulatory relationship between these two groups of genes. This may also help us make some predictions about the functions of those genes in the two groups which do not have GO annotation. For example, *ZNF703* (zinc finger protein 703) does not have any biological process GO annotation. However it is clustered with some zinc finger proteins (*EGR1*, 2 and 3). Keeping in mind the fact that zinc finger proteins are the most frequently used class of transcription factors, we may assume that *ZNF703* is also a transcription factor important for cellular proliferation.

There are some more relationships worth mentioning about some of the genes in bicluster 9. *CYR61* (cystein-rich angiogenic inducer 61) is differentially overexpressed in breast cancer cells exhibiting high levels of Heregulin (*HRG*), a growth factor closely associated with a metastatic breast cancer phenotype [106]. In addition, *CYR61* is an *IGF* (insulin-like growth factor) binding protein. Another gene in bicluster 9, *EHD1* is a protein thought to play a role in the endocytosis of *IGF* receptors. Therefore this pair of genes is interesting due to their relationship with *IGF*, which may indicate interactive and/or regulatory relationship although the genes are not annotated to be doing so. Another interesting relationship is between *MXD1* and *MYC*. The *c-MYC* proto-oncoprotein, in bicluster 9, plays a pivotal role in cell growth control. The forced expression of this protein promotes progression into S phase [111] and inhibits differentiation and entry into a quiescent state [112-115]. To perform these functions, *c-MYC* have to heterodimerize with *Max* [112-115]. In addition to *c-Myc*, *Max* also heterodimerizes with *Mad*, another member of the bicluster 9. Therefore it is very interesting to find two members of the *Myc/Max/Mad* transcription network in bicluster 9 because this network affect different aspects of cell behavior, including proliferation, differentiation, and apoptosis, by modulating distinct target genes.

Table 4.3. The GO categories enriched in different biclusters. statistical significance (p-value) of that enrichment. The third column gives the observed number of genes having the corresponding functional annotation which are found in the indicated bicluster(s). The fourth column gives the expected number of genes. The last column gives the statistical significance (p-value) of that enrichment.

Biclusters	GO Caretory	# of Genes Observed	# of Genes Expected	Significance of Enrichment
bilcluster 9 only	regulation of cell growth	2	0.11	0.00539
	regulation of transcription\, DNA-dependent	7	1.95	0.00125

	regulation of transcription from RNA polymerase II promoter	5	0.27	4.19748 E-06
8, 12, 13, 14 and 15	G2/M transition of mitotic cell cycle	2	0.03	0.0003
	mitosis	5	0.37	2.8576 E-05
	regulation of mitosis	2	0.1	0.0046
	negative regulation of programmed cell death	4	0.36	0.0004
	protein folding	4	0.61	0.0031
	DNA replication	3	0.45	0.0098
	negative regulation of apoptosis	4	0.36	0.0004
	amino acid transport	2	0.14	0.0084
bicluster 6 only compared to biclusters 7 and 8	nuclear import	2	0.11	0.0058
	protein import into nucleus	2	0.11	0.0056
	Protein import	2	0.11	0.0056
biclusters 6 and 7 only	ribosome biogenesis	2	0.04	0.0006
	RNA processing	3	0.21	0.00096
	rRNA processing	2	0.03	0.0003
biclusters 7 and 8	protein folding	8	0.89	2.5566 E-06
	DNA replication	5	0.65	0.0005

Table 4.4 The list of genes which are present in bicluster 9 only. The third column gives the observed number of genes having the corresponding functional annotation which are found in the indicated bicluster(s). The fourth column gives the expected number of genes. The last column gives the statistical significance (p-value) of that enrichment.

Locus Link	Gene Symbol		GO Annotation
1543	CYP1A1	Cystein-rich angiogenic inducer 61	electron transport
			oxidoreductase activity
1876	E2F6		regulation of progression through cell

			cycle
			negative regulation of transcription from RNA polymerase II promoter
			transcription factor activity
			transcription corepressor activity
1959	EGR2	Early growth response 2	transcription factor activity
			regulation of transcription, DNA-dependent
2353	FOS		transcription factor activity
			regulation of transcription from RNA polymerase II promoter
			cell death
2355	FOSL2		transcription factor activity
			regulation of transcription from RNA polymerase II promoter
			cell death
3491	CYR61		regulation of cell growth
			insulin-like growth factor binding
4084	MXD1		transcription factor activity
			transcription corepressor activity
			cell proliferation
4609	MYC		transcription factor activity
			regulation of transcription from RNA polymerase II promoter
			cell cycle arrest
			positive regulation of cell proliferation
7022	TFAP2C		transcription factor activity
			regulation of transcription from RNA polymerase II promoter
7071	KLF10		negative regulation of transcription from RNA polymerase II promoter
			transcription factor activity
			negative regulation of cell proliferation
7169	TPM2		actin binding
			Cytoskeleton
8870	IER3		anti-apoptosis
			molecular function unknown

9021	SOCS3	regulation of cell growth
		anti-apoptosis
9026	HIP1R	biological process unknown
10395	DLC1	negative regulation of cell growth
		cytoskeleton organization and biogenesis
10938	EHD1	biological process and function unknown
80139	ZNF703	nucleic acid binding

4.4.3 Biclustering detects the timely transition in functions in time-series data

Biclusters 6, 7 and 8 are quite interesting since they exhibit a shift in the time points they contain (Figure 4.3). Bicluster 6 has a much broader time spectrum from 1 hr of exposure to 16 hrs. Bicluster 7 is up-regulated after the 4 hrs up to 24 hrs while bicluster 8 extends from 12 hrs to 32 hrs. Although they first seem to be shifted in time, they also contain some common time points. That is the reason why they share some genes. Figure 4.6 gives the number of genes shared by these 3 clusters. When these genes are analyzed, it is seen that bicluster 6 is up-regulated before bicluster 7, which precedes bicluster 8. The result of the functional enrichment analysis of the genes which are present only in bicluster 6, those which are shared by biclusters 6 and 7, those which are shared by biclusters 7 and 8 and finally those which are exclusively in bicluster 8 support this idea also. The result of the functional enrichment analysis of the 22 genes, which are present in only 6 of these biclusters, the result for the 9 genes shared by biclusters 6 and 7 only, the result for the genes shared by biclusters 7 and 8 only are given in Table 4.3. As the time goes from 1 hr to 32 hrs, the functional categories enriched in the genes changes from nuclear transport → ribosome-related genes → protein folding + DNA replication → mitosis + nucleotide biosynthesis. This result is consistent with current knowledge of the mitogenic effects of estrogen [116].

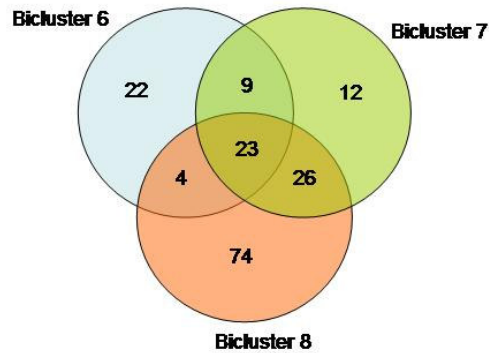


Figure 4.6. The Venn scheme showing the relationship between the biclusters 6, 7 and 8. The numbers in the intersection of different biclusters are the number of genes shared by the corresponding biclusters and the numbers belonging to the a single biclusters are the number of genes present in that bicluster only.

During the first 12 hrs, the transcriptional regulatory requirements of the cell is met in bicluster 9 (see above), then cells enter the S phase, which is marked by an increase in protein synthesis. The functional transition observed between biclusters 6, 7 and 8 corresponds to the natural course of events which takes place as the cell enter S and subsequently M phase: first an increase in protein synthesis, DNA replication then mitosis (Figure 4.7).

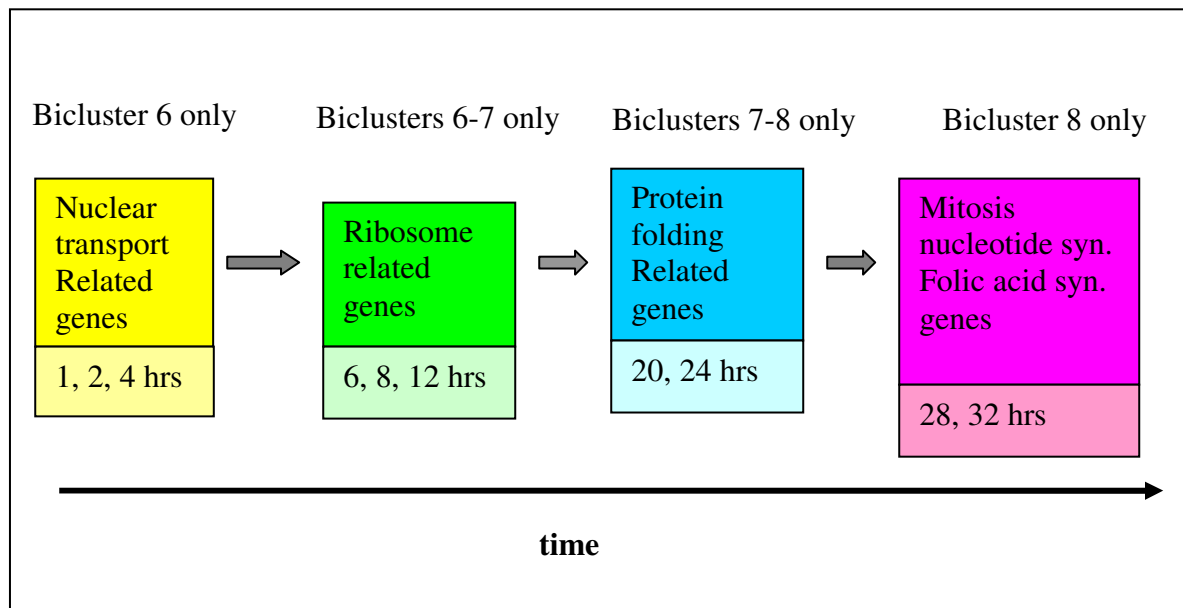


Figure 4.7. The change in functional enrichment with respect to the time.

4.5 Comparison of Biclustering with Hierarchical Clustering

This section compares the performance of biclustering with the results of a hierarchical clustering analysis done using the 477 genes which are found to be differentially expressed in biclustering analysis. The biclusters vs. time matrix in Figure 4.3 is hierarchically clustered using the hierarchical clustering function in MatLab, which gives the tree structure in Figure 4.3. This tree shows which biclusters are more similar to each other with respect to the time points they comprise along with the functional categories enriched within each bicluster. All down-regulated biclusters except biclusters 17 and 18 are grouped together. The reason why bicluster 17 and 18 are grouped together with the up-regulated ones is because these 2 biclusters have somewhat different expression profiles in Experiment-1. Since biclustering is able to catch local behaviors, these down-regulated biclusters seem more similar to the up-regulated ones. Among the up-regulated biclusters there are two groups: the first one which consists of biclusters 6, 7, 11 and 9, and the

second one consists of biclusters 8, 15, 12, 13, 14 and 10. First group corresponds to those biclusters which are up-regulated before 12 hrs of estrogen exposure. The second group on the other hand contains biclusters which are up-regulated after 12 hrs generally, except the bicluster 10. This bicluster, which is the most different among biclusters 8, 15, 12, 13, 14 and 10, is like a transitory bicluster between the first group (biclusters 6, 7, 9 and 11) and the second (8, 12, 13, 14, 15). In the first group, functional categories like protein folding, DNA replication, amino acid activation, ribosome biogenesis, RNA processing, regulation of cell growth, regulation of transcription are seen while the second group is enriched for more specific functional categories such as G2/M transition of cell cycle, regulation of mitosis, folic acid and derivative biosynthesis, nucleotide metabolism and biosynthesis. Therefore it can be said that biclustering separates the genes with respect to the time interval in which they are differentially expressed and the resulting functional assignments to time intervals makes biological sense. To compare this performance of biclustering with hierarchical clustering, the 477 genes which were found to be differentially expressed in biclustering were hierarchically clustered using Cluster [63]. Figure 4.8 shows the Cluster vs. Time of estrogen exposure matrix for this final analysis. This matrix is generated in the same way as the previous Bicluster vs. Time matrix (Figure 4.3). In this figure, again red color indicates up-regulation of the corresponding cluster in the corresponding time point, blue indicates down-regulation. Green indicates that there is no change in expression of genes in the cluster in the corresponding time point. Then this matrix was hierarchically clustered using Matlab as before resulting in the tree structure seen in Figure 4.9. This figure gives the GO categories associated with each cluster. The results from here on should be considered keeping in the mind the fact that biclustering was applied to 4320 genes and 477 genes were found to be differentially expressed. On the other hand hierarchical clustering was applied to these 477 genes. Therefore the results of the hierarchical analysis are inevitably biased towards giving similar functional enrichment.

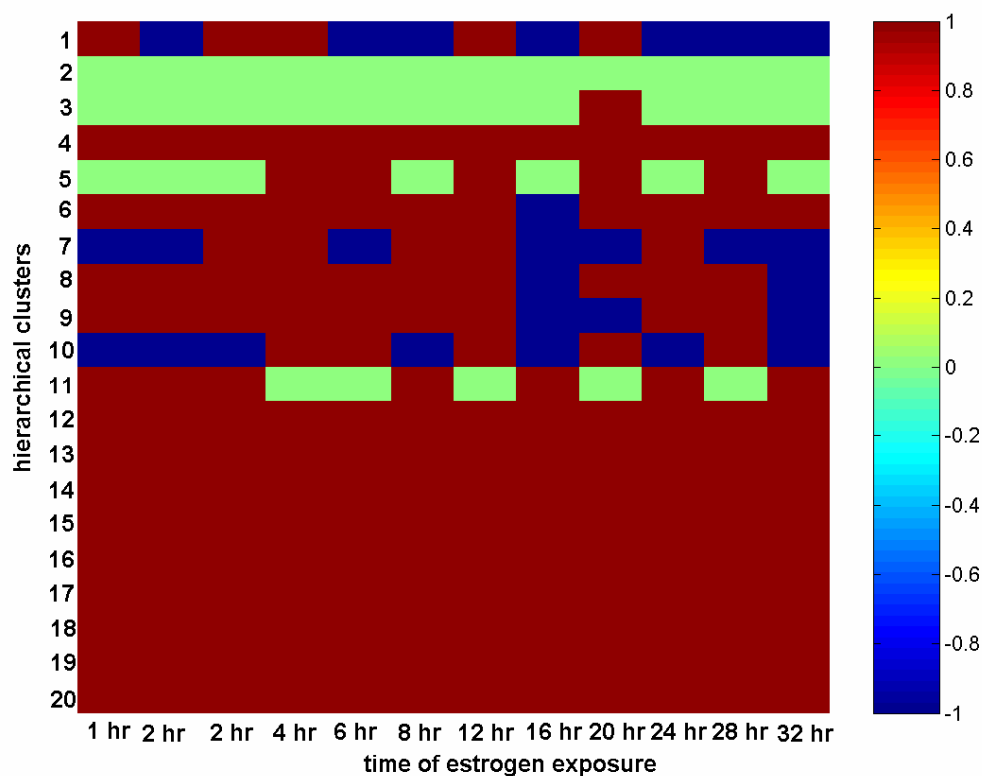


Figure 4.8. The condition (time points) vs. biclusters matrix for the hierarchical clustering of Experiment-2. Each column contains the time points included in the cluster. Red color indicates up-regulation of the corresponding cluster in the corresponding time point, blue indicates down-regulation. Green indicates that there is no change in expression of genes in the cluster in the corresponding time point.

4.5.1 Timely transition in function cannot be detected by hierarchical clustering

The behavior of the same 477 genes is very different in biclustering (Figure 4.9) and in hierarchical clustering (Figure 4.9). The most important advantage of biclustering over hierarchical clustering may be that genes are separated into disjoint clusters in the second

one. Therefore since hierarchical clusters do not share any genes, the timely transition in function between the genes in the biclusters 6, 7 and 8 cannot be detected by hierarchical method. In addition to this, the up-regulation of the genes in bicluster 9 before the genes in biclusters 8, 12, 13, 14 and 15 cannot be detected from (Figure 4.9) Therefore, it is not possible to be able to make comments about the functional relationship between different clusters.

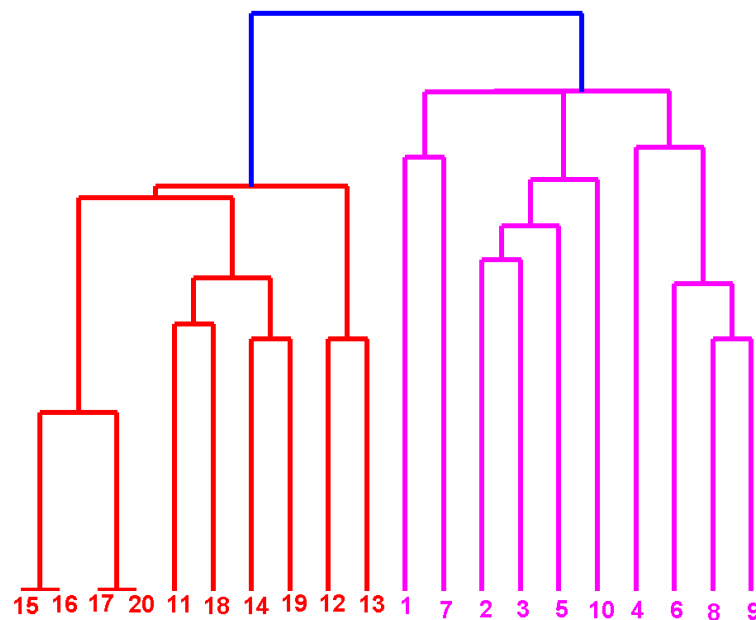


Figure 4.9. The tree obtained by hierarchical clustering of the Cluster vs. Time matrix for hierarchical clustering analysis. The tree shows the similarity between the clusters with respect to the behaviour they exhibit at each time point. Shown in green are the down-regulated clusters and in purple are the up-regulated ones. Table 4.5 gives the functional categories related with each cluster.

Table 4.5. The functional categories enriched in different hierarchical clusters. First column contains the cluster names and next to it the upward and downward arrow indicating up- and down-regulation. If there are any functional categories shared by different clusters this is indicated as a separate row.

CLUSTERS	FUNCTION
cluster 15	phosphoinositide-mediated signaling, DNA repair, DNA replication
cluster 16	ribosome biogenesis, amine biosynthesis, RNA processing, rRNA metabolism
cluster 17	no functional enrichment
cluster 20	ribosome biogenesis, rRNA metabolism
cluster 11	no functional enrichment
cluster 18	cell growth, regulation of cell growth
cluster 14	mRNA metabolism, RNA processing
cluster 19	nucleotide metabolism
cluster 12	phosphoinositide-mediated signaling, mitosis, regulation of mitosis, microtubule-based processes, nucleotide metabolism, sterol biosynthesis, isoprenoid metabolism, protein folding, DNA replication, nucleosome assembly
cluster 13	no functional enrichment
cluster 1	no functional enrichment
cluster 7	no functional enrichment
cluster 2	positive regulation of protein metabolism, positive regulation of protein biosynthesis, interleukin-6 biosynthesis, positive regulation of protein kinase production, positive regulation of immune response
cluster 3	G-protein signaling, coupled to cyclic nucleotide second messenger
cluster 5	carboxylic acid transport
cluster 10	DNA replication
cluster 6	monosaccharide metabolism, cellular carbohydrate metabolism
cluster 8	membrane receptor protein serine/threonine kinase signaling pathway, negative regulation of apoptosis, metal ion transport, regulation of cell migration
cluster 9	glycosaminoglycan metabolism, retrograde vesicle-mediated transport, Golgi to ER

4.5.2 Similar functional enrichments are seen between the two methods.

Similar functional enrichments are seen in the two methods, however this result is inevitable since hierarchical analysis is biased towards the results of biclustering because the 477 genes are those genes which are found to be differentially expressed by biclustering. Another reason for this resemblance is the fact that even if different clustering methods are used, it is impossible to obtain results very different from each other if the same data set is used. Different methods are bound to and furthermore should give similar results. Even in this case, some functional categories like regulation of cell growth, regulation of transcription, G2/M transition of cell cycle, DNA recombination, steroid receptor signaling, axon guidance cannot be detected by the hierarchical method. This shows that biclustering groups the genes better with respect to the functions they have.

Chapter 5

PATHWAY ANALYSIS

This chapter contains the results of the pathway analysis, which consists of two parts. In the first part, the mechanisms of tumorigenesis which are common in many cancer types are studied at the level pathways. In the second part, how the results from up-regulated and down-regulated biclusters are in good harmony with respect their effect on pathways is explained by focusing on a specific pathway, MAPK signaling. The pathways used are from KEGG [117-119] and BioCarta [120]. In some cases, when pathway informations for some genes is not known, OMIM (Online Mendelian Inheritance in Man) information is used. In most of the cases, genes from different clusters in the hierarchical clustering analysis of Cicatiello et al. [13] are grouped in the same bicluster reflecting the usefulness of the last method in helping us understand the functional relationships between genes. The hierarchical clusters in which the genes are found in Ciatello's study [13] are given in parenthesis after the genes names. If the gene is not included in this hierarchical clustering results this is indicated with the word 'absent'.

5.1 Six Different barriers to develop cancer

The underlying mechanism of estrogen action on breast cancer cells can be appreciated fully if the genes in every bicluster are analyzed with respect to the pathway(s) in which they are involved, which will provide a better understanding of the direction the fate of the cell takes upon estrogen treatment. Hanahan and Weinberg [24] suggested that

cells have to overcome 6 barriers to develop cancer, which they summarized to take place in 6 different steps: self-sufficiency in growth signals, insensitivity to antigrowth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. One immediately thinks that estrogen, if it promotes cancer progression, should elicit pathways to achieve these requirements. Actually this is the case in this data set: estrogen exposure affects all 6 mechanisms although possibly not in the same magnitude. Table 5.1 gives a summary of the 6 barriers, the genes associated to each of them and the pathway(s) to which the genes are annotated. Biclusters 8 is up-regulated while biclusters 3, 4 and 17 are down-regulated, which is denoted by the upward arrow right next to the bicluster number in the second column. All the genes in the third column belong to the respective bicluster on the second column. Therefore the genes follow the expression pattern of the corresponding biclusters. Since some genes may have inhibitory effects on the pathway and other genes have stimulatory effects, this is denoted by the downward and upward arrows next to the gene name. Thus if the genes has a positive effect on the pathway (has upward arrow next to its name) and is up-regulated (has upward arrow next to the bicluster it belongs), its expression behaviour is appropriate to overcome the barrier in the first column. Similarly, if the genes has negative effects on the pathway (downward arrow next to it) and is down-regulated (downward arrow next to the respective bicluster number), its expression behaviour is appropriate to overcome the corresponding barrier in the first column. As can be seen from the table, response to estrogen treatment takes place at each of the barriers so as to overcome them although possibly not with the same magnitude. Therefore the expression behaviour of the genes in different biclusters are in agreement with each other and the total expression behaviour represents a full transcriptomic response.

Table 5.1. The Six Barriers to Develop Cancer: This table gives the 6 of the barriers, the genes associated to each of them and the pathway to which the genes are annotated. Bicluster 8 is up-

regulated while biclusters 3, 4 and 17 are down-regulated, which is denoted by the upward arrow right next to the bicluster number in the second column. All the genes in the third column belong to the respective bicluster on the second column. Therefore the genes follow the expression pattern of the corresponding biclusters. Since some genes may have inhibitory effects on the pathway and other genes have stimulatory effects, this is denoted by the downward and upward arrows next to the gene name. Thus if the genes has a positive effect on the pathway (has upward arrow next to its name) and is up-regulated (has upward arrow next to the bicluster it belongs), its expression behaviour is appropriate to overcome the barrier in the first column. Similarly, if the genes has negative effects on the pathway (downward arrow next to it) and is down-regulated (downward arrow next to the respective bicluster number), its expression behaviour is appropriate to overcome the corresponding barrier in the first column.

Barriers to Develop Cancer	Bicluster	Genes	Pathways
Self-sufficiency in Growth Signals	8↑	RET ↑, NTRK1 ↑, STMN1 ↑, RAP1A ↑, HSPA5 ↑	MAPK pathway
Insensitivity to Antigrowth Signals	8↑	CCND1 ↑, CDC2 ↑, CDC20 ↑, CKS2 ↑, MCM2 ↑, MCM7 ↑, PTTG1 ↑	Cell cycle
	8↑	STMN1 ↑, HSPA5 ↑, CDC2 ↑, TUBA1 ↑	Stathmin and breast cancer resistance to antimicrotubule agents
	8↑	RAN ↑, Importin- α ↑, Importin- β ↑, STK6 ↑	Role of Ran in mitotic spindle regulation
Evasion of Programmed Cell Death and Escape from Immune System	8↑	HSPA5 ↑, TPD52L1 ↑, HSPD1 ↑, BIRC5 ↑, MYBL2 ↑, NME1 ↑	Apoptosis-related
	17↓	TLR4 ↓, NF- κ B2 ↓, STAT1 ↓	Toll-like receptor signaling pathway
Limitless Replicative Potential	8↑	RAD51 ↑, EXO1 ↑, KPNA2 ↑, RAD51C ↑	Homologous-recombination
Sustained Angiogenesis	8↑	NRP1, SEMA3A, SEMA5	Axon guidance
Tissue invasion and angiogenesis	4↓	CLDN ↓, OCLN ↓	Cellular adhesion molecules of endothelial cells
	3↓	OCLN ↓, PECAM1 ↓	Cellular adhesion

			molecules of endothelial cells
--	--	--	--------------------------------

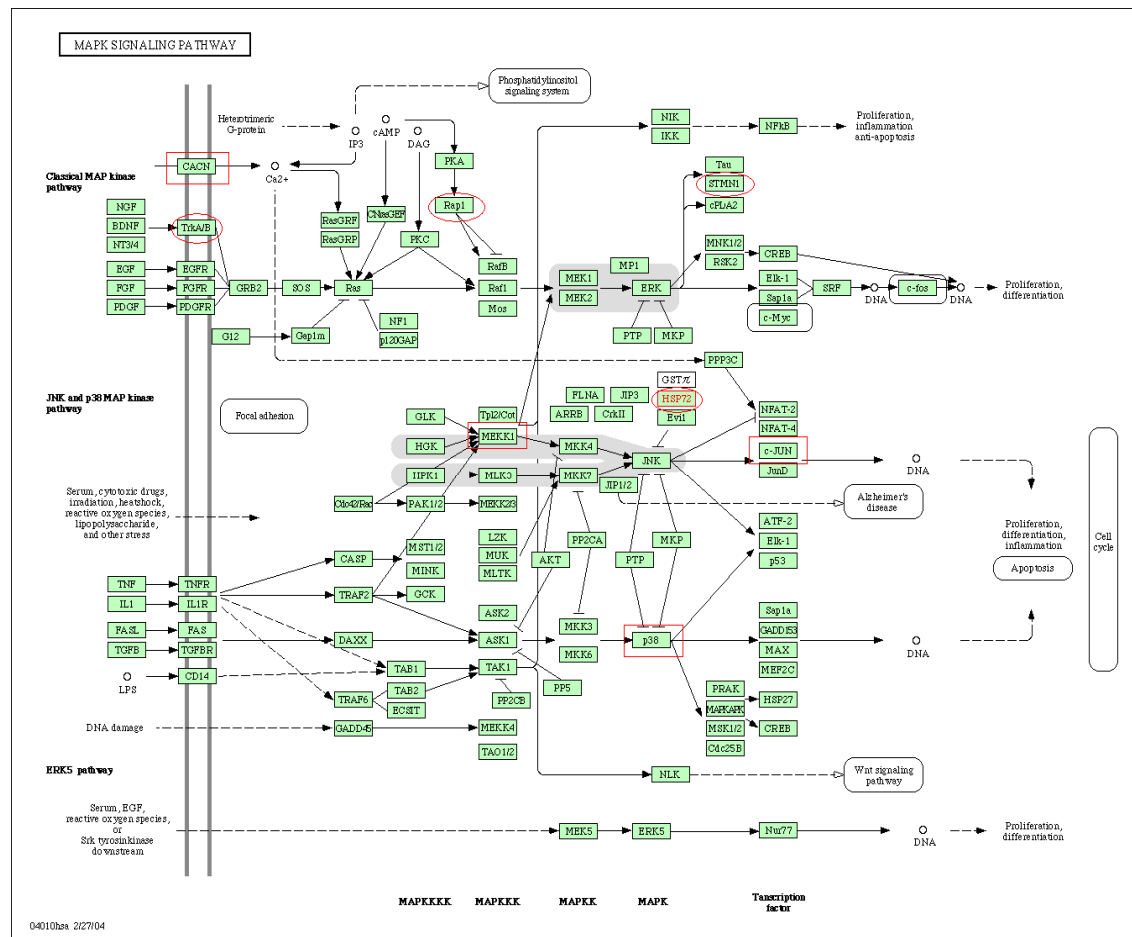


Figure 5.1. The MAPK signaling pathway from KEGG. The genes marked in red circle are from bicluster 8, those in red rectangles are from bicluster 5 and those in black rectangles are from bicluster 9

In the following sections, the barriers and the pathways and genes associated with each of them are given in detail. The genes in each pathway occur in the same bicluster. In most of

Figure 5.2. The cell cycle map from KEGG. The genes from bicluster 8 which are also found in this pathway are shown in red ellipses.

5.1.1 Response to Estrogen Exposure: Self-sufficiency in Growth Signals

This property can be acquired in three different ways: (1) over-expression or mimicking of the growth factor, (2) over-expression or mimicking of the cell surface receptor for that growth factor, (3) alteration of the intracellular pathway elicited by that growth factor. *AREG* and *VGF*, which are up-regulated in bicluster 8 are examples to the first way, while *RET* and *NTRK1*, receptor kinases (again up-regulated in bicluster 8) are examples to the second way. In most of the cancers, however, in addition to deregulation of these receptors and growth factors, the down-stream circuitry that processes the signal received by the receptor is also altered (the third way). Bicluster 8 contains some genes involved in MAPK signaling such as *RET* (cluster 5), *TPD52L1* (cluster 7), *NTRK1* (cluster 8), *STMN1* (absent), *RAP1A* (cluster 8) and *HSPA5* (cluster 5). Figure 5.1 shows genes from this bicluster which are included in the MAPK signaling pathway from KEGG: *NTRK1*, *STMN1*, *RAP1A* and *HSPA5*.

5.1.2 Response to Estrogen Exposure: Insensitivity to Antigrowth Signals

Much of the circuitry that enables normal cells to respond to antigrowth signals is associated with the cell cycle clock, specifically the components governing the transit of the cell through the G1 phase of its growth cycle. In bicluster 8, 3 pathways, cell cycle (Figure 5.2), Stathmin and breast cancer resistance to antimicrotubule agents' (Figure 5.3) and 'role of RAN in mitotic spindle regulation' (Figure 5.4) are examples to such deregulation of cell cycle and mitosis-related pathways. The cell cycle related genes from bicluster 8 are shown in Table A.1. 7 of these genes are included in the cell cycle from KEGG (Figure 5.2).

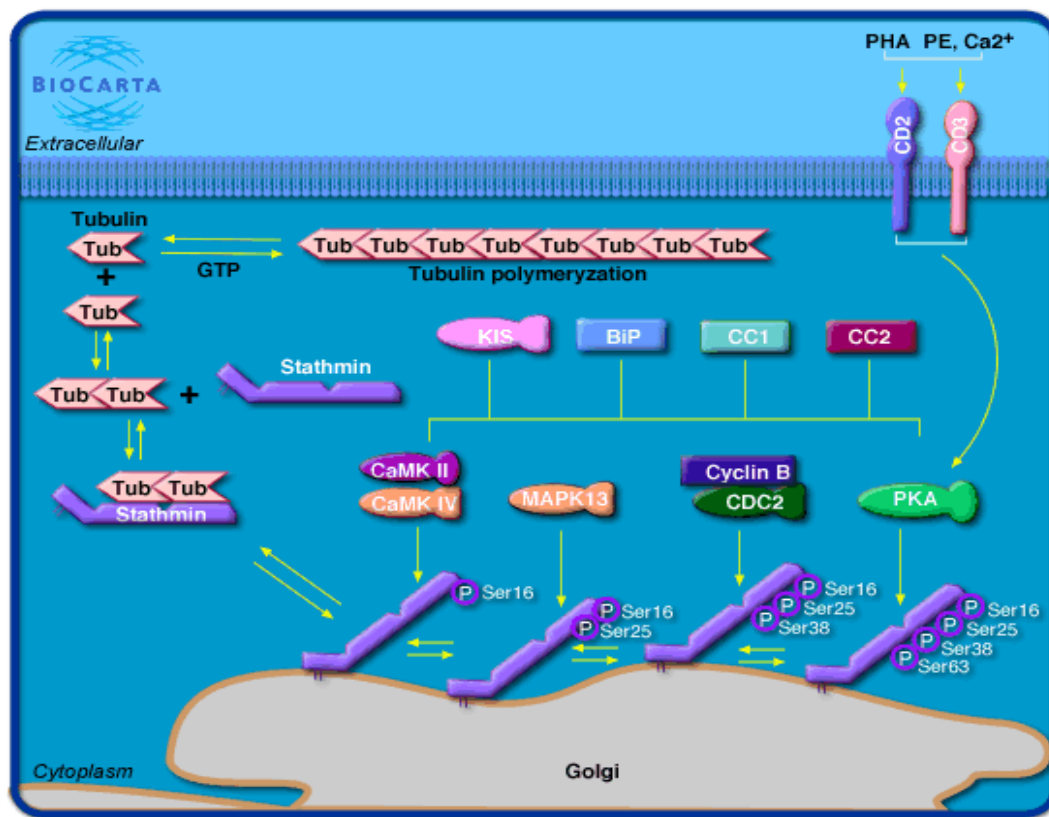


Figure 5.3. ‘Role of stathmin (STMN1) in microtubule destabilization’ from BioCarta

Since some antigrowth signals attack spindle/microtubule-related functions, deregulation of these systems is also associated to escape from growth inhibiting effects. Figure 5.3 shows the ‘Stathmin and breast cancer resistance to antimicrotubule agents’ pathway from Biocarta. In addition to *STMN1*, some other genes in this pathway like *HSPA5* (cluster 5, designated with BiP in the pathway), *CDC2* (cluster 6) and *TUBA1* (cluster 7) are also found to be up-regulated in the same bicluster.

Figure 5.4 illustrates the important role *RAN* (cluster 8) has in spindle stabilization. In addition to *RAN* and importins alpha/beta, *STK6* (aurora kinase A) (cluster 7) from this Biocarta pathway is found in the same bicluster. *STK6* is known to be involved in

microtubule formation and/or stabilization at the spindle pole during chromosome segregation. In addition to these genes involved in the pathways in Figure 5.3 and Figure 5.4, there are other genes which are found to be involved in microtubule-related functions from OMIM: *DLG7* (absent), *KIF22* (cluster 8), *BRRN1* (absent), *XPOT* (cluster 6), *MAD2L1* (cluster 8), *CDC20* (cluster 7), *PTTG1* (cluster 8), *KPNA2* (cluster 8), *CENPA* (cluster 7), *KIF2C* (cluster 8), *KIF20A* (cluster 8). All these genes are scattered to clusters 6, 7 and 8 in the hierarchical analysis of Cicatiello et al. [13]. However in biclustering they are grouped together in bicluster 8.

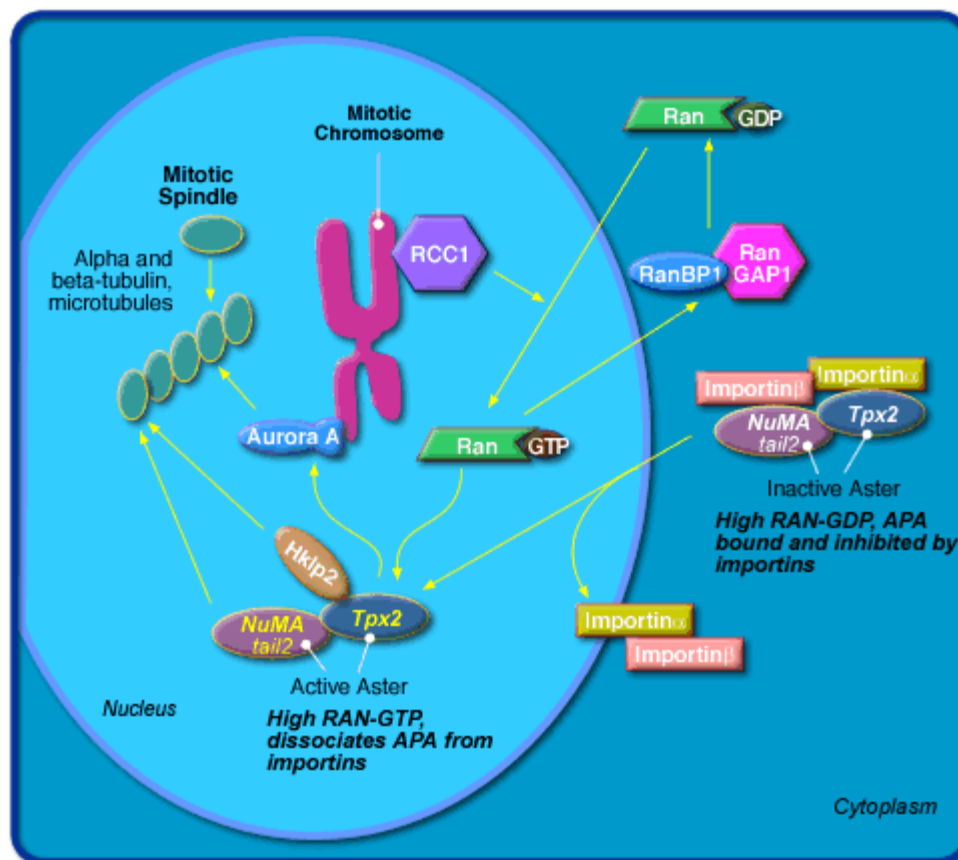


Figure 5.4. 'Role of Ran in mitotic spindle regulation' from BioCarta.

5.1.3 Response to Estrogen Exposure: Evasion of Programmed Cell Death and Escape from Immune System

Tumor progression cannot be achieved only by an increase in the proliferative capacity of the cell. Acquired resistance to apoptosis is a crucial step taken toward progressive tumor behavior. There are many apoptosis regulated genes, which are differentially expressed. Due to the mitogenic effect of estrogen many anti-apoptotic genes are up-regulated while proapoptotic genes are down-regulated (Table 5.2). In addition to activating mechanisms to avoid apoptosis, cancer cells should also be able to escape from the immune system. shows the genes from bicluster 17 included in the toll-like receptor signaling pathway from KEGG: *TLR4* (absent), *NF- κ B2* (absent) and *STAT1* (absent). *TLR4* is the receptor via which some adjuvants induce its in vivo anti-tumor effect [121]. It is therefore very important for the induction of proinflammatory effects. Unlike most other *NF- κ B* transcription factors, activation and nuclear translocation of *NF- κ B2* (p52) [122] can result in repression of *NF- κ B*-dependent genes, which inhibits the antiapoptotic, proliferation-, motility- and invasion-promoting roles of *NF- κ B*. The final gene, *STAT1*, has negative effect on tumor formation and progression [123].

Table 5.2. The list of genes from different bicluster, which are related to apoptosis

Locus Link	Gene Symbol	Function	Bicluster
578	<i>SCARF1</i>	induction of apoptosis	Bic. 3,17,18,19,20
1191	<i>CLU</i>	apoptosis	Bic 3
7185	<i>TRAF1</i>	regulation of apoptosis	Bic 3
10114	<i>HIPK3</i>	anti-apoptosis	Bic 4
573	<i>PTGER2</i>	anti-apoptosis	Bic 5

23273	<i>KIAA0367</i>	induction of apoptosis	Bic 16
247	<i>TSN</i>	induction of apoptosis	Bic 17,18
2537	<i>IFI6</i>	anti-apoptosis	Bic 17,19
8740	<i>TNFSF14</i>	induction of apoptosis	Bic 17,20
2149	<i>F2R</i>	apoptosis	Bic 18
6997	<i>TDGF1</i>	negative regulation of apoptosis	Bic18
598	<i>BCL2L1</i>	anti-apoptosis	Bic 19
3309	<i>HSPA5</i>	caspase inhibitor activity	Bic 6,7,8
7164	<i>TPD52L1</i>	DNA fragmentation during apoptosis	Bic 6,7,8
7187	<i>TRAF3</i>	induction of apoptosis	Bic 6
8565	<i>YARS</i>	apoptosis	Bic 6,7
3329	<i>HSPD1</i>	regulation of apoptosis	Bic 7,8
332	<i>BIRC5</i>	anti-apoptosis	Bic 8
4605	<i>MYBL2</i>	anti-apoptosis	Bic 8
4830	<i>NME1</i>	regulation of apoptosis	Bic 8



Many if not all types of mammalian cells carry an intrinsic, cell-autonomous program that limits the number of multiplicative cycles they can accomplish. Telomeres, the ends of chromosomes are the means of counting the number of cell generations. Because of the inability of DNA polymerase to completely replicate the 3' ends of chromosomal DNA during each S phase, telomeres are progressively shortened with each replication cycle. 85-90% of cancer cells overcome this barrier by up-regulating the expression of the telomerase enzyme [124]. The second way for the maintenance of telomeres is through recombination-based interchromosomal exchanges [125]. Upon exposure to estrogen, breast cells seem to acquire both of these capabilities. An example to the first mechanism is the up-regulation of the *DKC1* (cluster 7), found in most of the up-

regulated biclusters. The product of this gene has telomerase activity and is a part of the telomerase holoenzyme complex. The second mechanism also seem to be employed since some genes involved in homologous-recombination such as *RAD51* (cluster 8), *EXO1* (cluster 6), *KPNA2* (cluster 8) and *RAD51C* (cluster 8) are found up-regulated in bicluster 8.

5.1.5 Response to Estrogen Exposure: Sustained Angiogenesis

As cancerous cell mass increase in size, it becomes difficult to maintain viability for those cells which are at distant places to the capillary blood vessels. Therefore in order to progress to a larger size, cancerous tissue should acquire angiogenetic ability. Figure 5.6 shows the genes included in the axon guidance pathway from KEGG. One of these genes is *NRP1* (cluster 7), neuropilin-1 which is a membrane-bound coreceptor to a tyrosine kinase receptor for both vascular endothelial growth factor (*VEGF*) and semaphorin (*SEMA3A*) family members. *VEGF* is an important angiogenesis-initiating signal in endothelial cells. Therefore *NRP1* plays versatile roles in angiogenesis, axon guidance, cell survival, migration, and invasion and it is up-regulated in the bicluster 13.

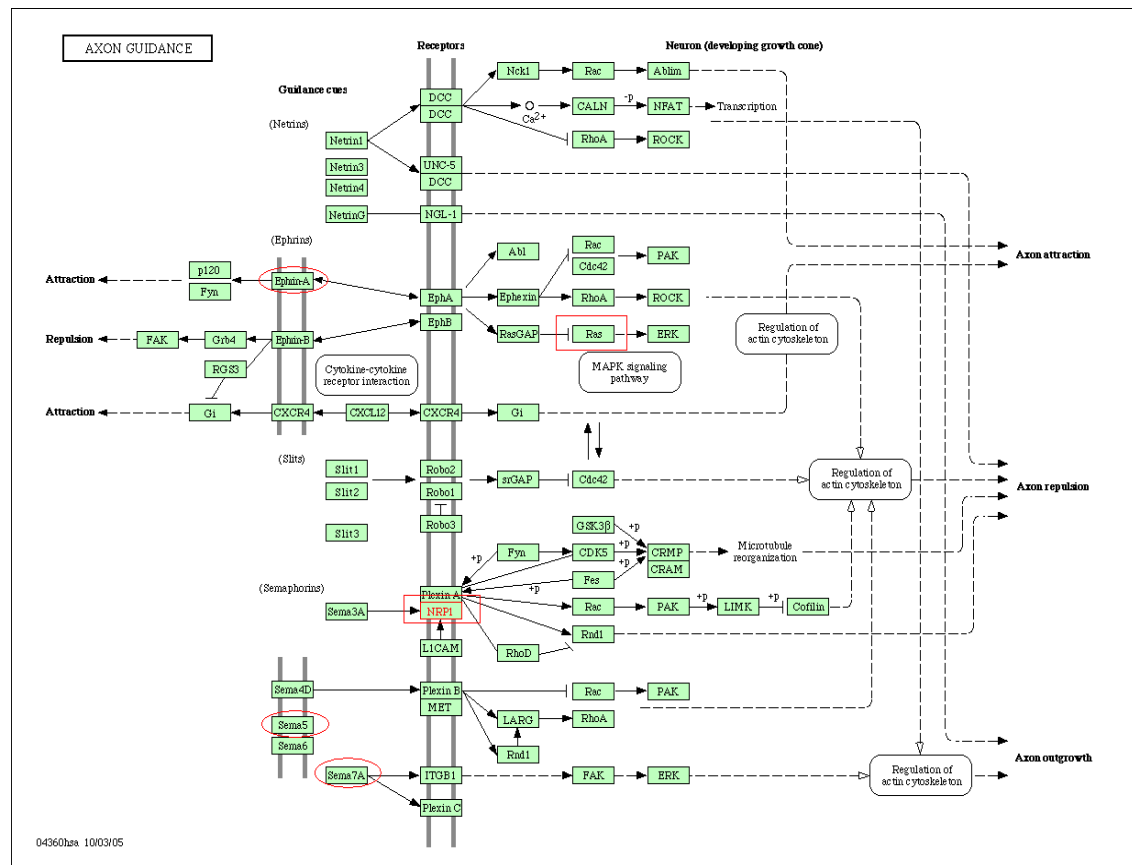


Figure 5.6. Axon guidance pathway from KEGG. The genes in red ellipse are down-regulated in bicluster 16 and those in red rectangle are up-regulated in bicluster 13.

5.1.6 Response to Estrogen Exposure: Tissue invasion and angiogenesis

Several classes of proteins involved in the tethering of cells to their surroundings in a tissue are altered in cells possessing invasive or metastatic capabilities. One of the ways towards acquiring invasive ability is through the affected cell-cell adhesion molecules (CAMs). Fig. 22 shows the cell-cell adhesion molecules in endothelial cells. 3 of these genes are down-regulated in biclusters 3 and 4, which are in a similar time interval (Figure 5.7). *OCN* is in bicluster 3 and 4, *CLDN* is in bicluster 4 only and *PECAM1* is in bicluster 3 only. Down-regulation of these adhesion molecules enables the cancerous cells to move more freely.

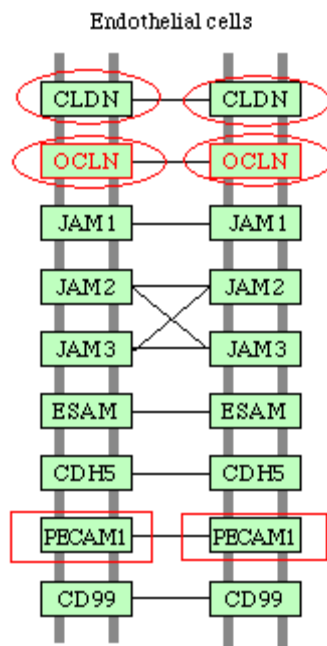


Figure 5.7. Adhesion molecules for endothelial cells: a part of the Cellular Adhesion Molecules pathway from KEGG.

5.2 Down-regulated and up-regulated biclusters have the same effects on the MAPK pathway.

Relating the biological functions of genes in each bicluster may be useful in terms of understanding which processes are dominant at what time intervals. However, to appreciate fully the underlying mechanism of estrogen action on breast cancer cells, one should take into consideration the pathway information available for the genes. In this way, cross-talks between different pathways or between the components of a single pathway can be deduced. Here MAPK signaling pathway from KEGG is chosen as an example to how such interpretations can be made from biclustering results due to its significance in cancer cells. Different biclusters contain genes included in this pathway. Table 5.3 gives the list of genes included in the corresponding map from KEGG (Figure 5.8). The genes are indicated

in red ellipses and the bicluster numbers are written next to them. Some genes are shared by more than one bicluster, hence more than one number are attached to such genes. When different components of a protein complex are included in the same bicluster, the bicluster number is written more than once, each time separated by a comma (as in the case of *HSP72*). There are three major subgroups in the MAPK family: extracellular signal-regulated kinase (ERK), p38 MAPK, and *c-Jun* N-terminal kinase (JNK)/stress-activated protein kinase (SAPK) [126]. Therefore the genes in this pathway should be studied in those three parts: The first group of genes is in the classical ERK/MAPK cascade which is mainly activated by mitogenic stimuli (the upper part of the figure). All of these genes except *CACNB3* are from up-regulated biclusters, which leads to the activation of the MAPK cascade and therefore cell proliferation. The second group of genes, is related to the JNK cascade (in the middle of the figure). While activation of the ERK forms of MAPK causes survival responses, activation of JNK and p38 promote cell death [127]. As expected, the up-regulation of *FLNB*, *HSPA5* and *HSPA1A* (all in bicluster 6) leads to the inactivation of the JNK cascade because these are inhibitors of JNK as seen from the figure. On the other hand, another gene, *MAP3K1* (bicluster 5) is down-regulated, which will have the same effect on JNK cascade. Thus it can be said that the expression profiles of these genes from different biclusters (3 from an up-regulated and one from a down-regulated bicluster) genes are complementary to each other in their effect to JNK cascade. The last group contains *MAPK12/p38* (bicluster 5), which is at the center of the p38 pathway (at the bottom of the figure). It is also down-regulated. The two genes, p38 and MAP3K1 have an apoptotic effect on breast cancer cells. Therefore their down-regulation along with the up-regulation of the inactivators of JNK leads to suppression of apoptosis. This kind of thinking is very useful to fully understand the underlying mechanisms of estrogen action. Unfortunately since pathway information is far from being complete, the interpretations that can be made are limited. However, since genes, which occur in the

same biclusters may be functionally related to each other (see below), one can still make some interesting observations. For example, of 12 genes which are found in the intersection of the biclusters 6, 7, 8 and 14, 4 genes are either directly or indirectly related to the MAPK pathway although they do not exist in the corresponding KEGG map: *MYB* [128], *CCND1* [129], *RAD51* [130] and *MCM7* [131]. Therefore, this analysis may prove useful to overcome the incompleteness of the pathway information and help us make a prediction about the functions of unannotated genes. By either making such observations or by experimental techniques, as pathway annotation information grows, things will get clearer at the pathway level.

Table 5.3 The list of genes from different biclusters which are included in the MAPK signaling map from KEGG.

Locus Link ID	Gene Symbol	Function	Biclusters	
3309	HSPA5	anti-apoptosis	15,14,13,12,11,10,8,7,6	JNK cascade
3925	STMN1	mitotic spindle organization and biogenesis	15, 14,8	MAPK cascade
5906	RAP1A	cell cycle	15,13,8,6	MAPK cascade
4893	NRAS	regulation of progression through cell cycle	14	MAPK cascade
2353	FOS	specific RNA polymerase II transcription factor activity	9	MAPK cascade
4609	MYC	regulation of transcription from RNA polymerase II promoter	9	MAPK cascade
4914	NTRK1	receptor activity	8	MAPK cascade
2317	FLNB	actin cytoskeleton organization and biogenesis	6	JNK cascade
3303	HSPA1A	response to unfolded protein	6	JNK cascade
3725	JUN	RNA polymerase II transcription factor activity	5,19	JNK cascade
4214	MAP3K1	MAP kinase kinase kinase	5,17	JNK

		activity		cascade
6300	MAPK12	cell cycle	5	P38 cascade
784	CACNB3	calcium ion transport	5	MAPK cascade

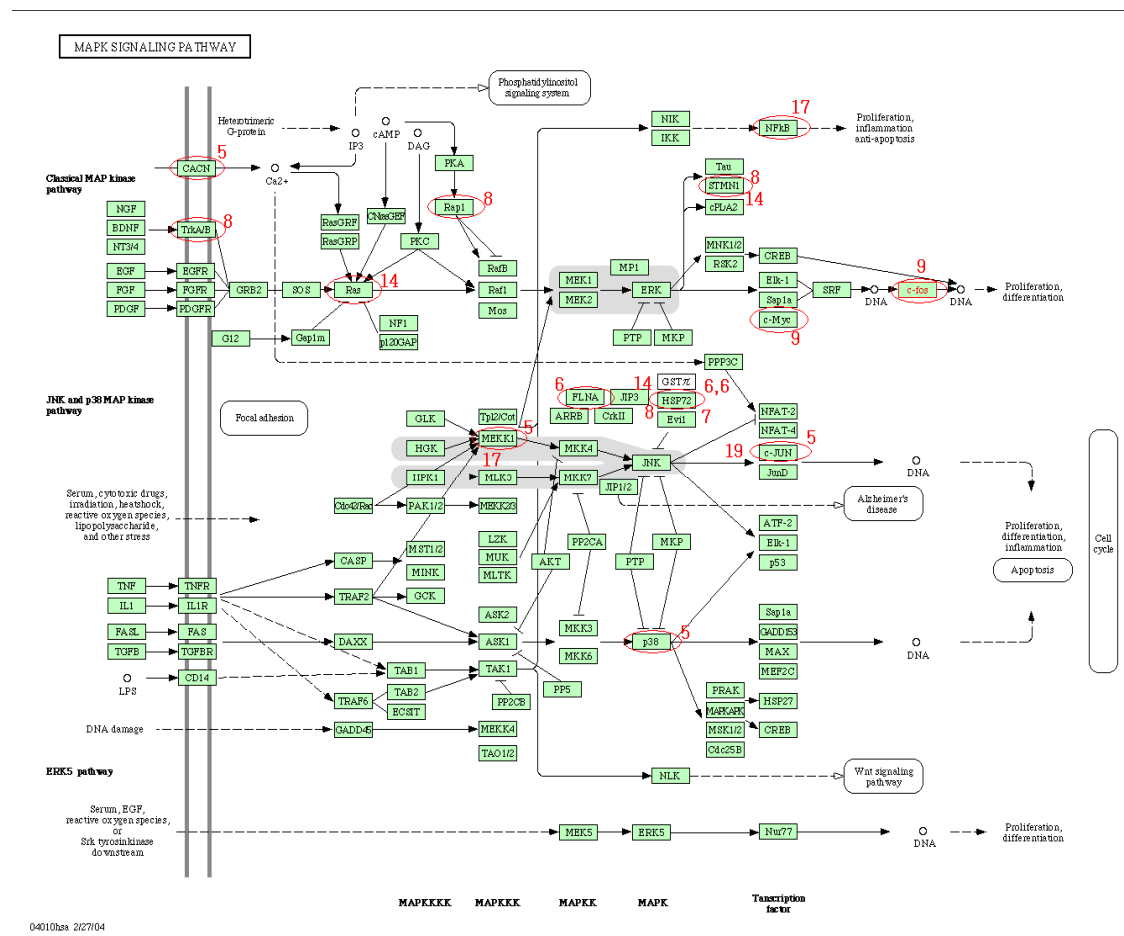


Figure 5.8. MAPK signaling pathway from KEGG. The symbol $\rightarrow$ indicate molecular interaction or relation. The symbol $\dashv$ indicates inhibition. The genes are indicated in red ellipses and the bicluster numbers are written next to them. Some genes are shared by more than one bicluster, hence more than one number are attached to such genes. When different components of a protein complex are included in the same bicluster, the bicluster number is written more than once, each time separated by a comma (as in the case of HSP72)

Chapter 6

THE GENE FREQUENCY AND TRANSCRIPTION FACTOR ANALYSIS

6.1 The Occurrence Frequency of Genes in Biclusters

Since biclustering separates the genes into joint sets, genes can be in more than one bicluster. Some genes are shared by more biclusters than the others, therefore the occurrence frequency of genes in the biclusters differ from gene to gene. This occurrence frequency may give clues about the importance of the functions the genes perform across all the arrays or the relevance of the functions of different genes. Since each bicluster covers a different subset of conditions (time points), the highly frequent genes may be very important for the functions seen across all arrays, which may indicate that common functions are seen in most of the arrays. In our case, such frequently seen genes may correspond to genes which are very important for cell cycle or which have a key role for response to estrogen. Supplementary Table 2 gives the genes which are shared by different biclusters from the most frequent ones to those included in one bicluster only. Since the up-regulated biclusters share more genes among each other than the down-regulated biclusters, the highest number of occurrence for up-regulated genes is two times the highest number of occurrence of down-regulated genes. The maximum number of biclusters which share genes is 10 in the case of up-regulated biclusters and 5 in the case of down-regulated biclusters. In this analysis not only the most frequent genes are import, the least frequent

genes, which are specific to a single bicluster may also indicate the importance of that genes in the specific time points covered by that bicluster. A very good example was given in section 4.4.2: In bicluster 9 there are 21 genes which do not occur in any of the other biclusters. These genes are enriched in transcriptional factors (see section 4.4.2) and the activation of these genes precedes another group of genes which are related to functions such cell cycle, protein folding, etc. Therefore such a frequency analysis may help us have an idea about the importance of genes in the global response or in specific time intervals.

6.1.1 Genes shared by many biclusters are usually functionally important

The more common the genes are between different biclusters, the more crucial the functions they perform are. Two examples are *SLC7A5* and *CCND1*, which are shared by all of the up-regulated biclusters and whose functions are key to the viability of cancer genes. *CCND1* is very important for the G1/S transition of cell cycle. *SLC7A5* (*LAT1*)

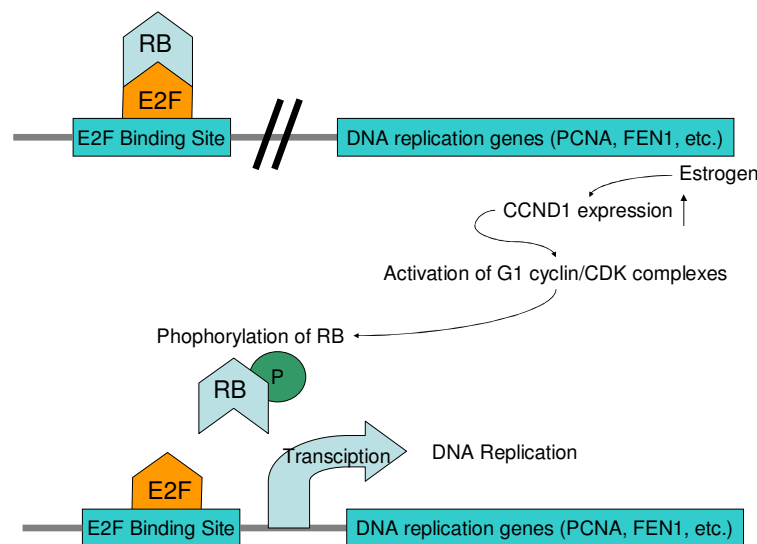


Figure 6.1 The role of CCND1 and RB in passage from G1 to S phase of cell cycle

mediates sodium independent transport of large neutral amino acids including several essential amino acids in a variety of cancer cells including breast [132] and is a major route for providing tumor cells with branched and aromatic amino acids. The resemblance of its expression pattern to *CCND1* shows that it is as indispensable as *CCND1* for growth and proliferation of breast cells. Some of the other highly frequent genes are *FKBP4* (FK506 binding protein 4), *FKBP5* (FK506 binding protein 5), *RBBP6* (Retinoblastoma binding protein 6) and *HSPA5* (Heat shock protein A5). They are found in 9/10 of the up-regulated biclusters. The first two, *FKBP4* and *FKBP5* are large molecular weight immunophilins, which bind to the immunosuppressant drugs FK506 and cyclosporin A. The large molecular weight immunophilins are capable of a number of diverse interactions and functions. The most well studied and interesting is their association with native steroid hormone receptor complexes including glucocorticoid receptor, progesterone receptor and estrogen receptor. The expression of these two proteins is estrogen regulated [133] and they act on the estrogen receptor affecting its transcriptional activity [134]. The function of the third gene, *RBBP6*, is unknown but it is known to bind to retinoblastoma (RB) protein. RB protein is an important regulator of cell cycle progression serving as a gatekeeper of G1 phase [Figure 6.1]. Its inhibiting effect on G1/S transition is inactivated by *CCND1*, leading to passage through the restriction point leads to the transcription of genes involved in DNA replication which takes place in the S phase [135]. The high occurrence frequency of this gene indicates the importance of the function it does although what this function exactly is unknown. Finally the last genes, *HSPA5*, also referred to as GRP78, is a central regulator of endoplasmic reticulum (ER) function due to its roles in protein folding and assembly, targeting misfolded protein for degradation, ER Ca(2+)-binding and controlling the activation of trans-membrane ER stress sensors. Further, due to its anti-apoptotic property, stress and cancer induction of *HSPA5* represents an important pro-survival component of the unfolded protein response [136].

6.2 Transcriptional Factor Binding Site Analysis

For the transcriptional factor binding site analysis PRIMA algorithm implemented in Expander was used. This method identifies putative binding sites of transcription factors and evaluates the statistical significance of their prevalence in a given set of promoters. Binding sites (motifs) that are recognized and bound by TFs are modelled by consensus sequences or position weight matrices (PWMs). The algorithm scans a given set of promoters for TF-binding sites and identifies PWMs that are significantly overrepresented in the examined set in comparison with a background set of promoters. In this section, the results of the transcriptional factor analysis is given. There are 2 transcriptional factors whose binding sites were enriched in the promoters of the genes belonging to different biclusters: E2F and SRF. Before giving more detail about both of them,

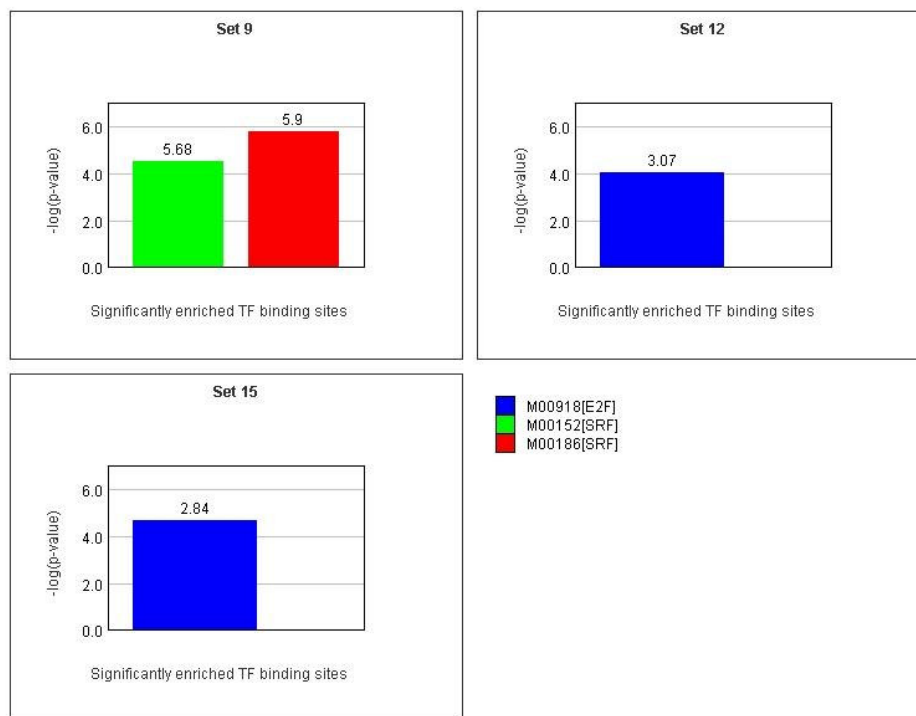


Figure 6.2. The transcriptional factors and the biclusters in which they are enriched.

the biclusters in which they are found to have putative binding sites should be emphasized further. Bicluster 9 (which contains mostly transcription factors) is enriched for SRF (Figure 6.) while biclusters 12 and 15 (which contain many broad GO categories like cell cycle, protein folding and so on) are enriched for E2F (Figure 6.2). Therefore it is very meaningful to see that bicluster 9, which is activated before biclusters 12 and 15 with respect to both time and function, may be regulated by SRF, which is responsible for the regulation of immediate response growth response genes while biclusters 12 and 15, which mark the passage from G1 to S phase (12 hrs of exposure), is regulated by E2F, whose role in G1/S transition is well known [137].

SRF (Serum Response Factors):

This gene encodes a ubiquitous nuclear protein that stimulates both cell proliferation and differentiation. It is a member of the MADS (MCM1, Agamous, Deficiens, and SRF) box superfamily of transcription factors. This protein binds to the serum response element (SRE) in the promoter region of target genes. This protein regulates the activity of many immediate-early genes, for example *FOS* [138] and *FOSL1*, and thereby participates in cell cycle regulation, apoptosis, cell growth, and cell differentiation. This gene is the downstream target of many pathways; for example, the mitogen-activated protein kinase pathway (MAPK) that acts through the ternary complex factors (TCFs). 6 genes are found to contain the binding site of SRF: *EGR3*, *EGR2*, *EGR1*, *FOS*, *FOSL2* and *TPM2*. Of these, only *FOS* is shown to be regulated by this transcription factor [138]. However, the other genes are also interesting candidates to be regulated by SRF. For example, *FOSL2* is one of the members of *FOS* gene family. *EGR1*, 2 and 3 are all the so-called ‘immediate-early growth response’ genes. Of these 6 genes only *TPM2*

seems unrelated to the others. However, an interesting common characteristic of these 6 genes is that the products of all of them are shown to interact with JUN [139].

E2F:

The E2F family of proteins plays a pivotal role in the regulation of cellular proliferation by controlling the expression of genes that are essential for either entry into, or passage through, the cell cycle (Figure 6.1). 25 genes were found to contain E2F binding site: *MYB*, *SFRS1*, *SF3B3*, *RAD51*, *PTDSS1*, *MTHFD1*, *FOXMI*, *DTYMK*, *CCT5*, *CACYBP*, *MYBL2*, *MAD2L1*, *SLC3A2*, *EXO1*, *NCL*, *SNRPF*, *GTF3A*, *MCM2*, *PPIB*, *HSP90B1*, *EMG1*, *PDZK1*, *PAICS*, *AREG*, *ITGB1BP1*. 4 (out of 25) genes were experimentally shown to be regulated by E2F: *RAD51* [140], *MYBL2* [141], *MAD2L1* [142], *SLC3A2* [143].

Chapter 7

DISCUSSION

Although it has been known for many years, biclustering has only recently been applied to microarray data. Unlike other clustering methods, in this method genes can be in more than one bicluster, which is in accordance with the idea that genes usually perform more than one function. Thus biclustering is biologically more meaningful. However, that makes the interpretation of the biclustering results more complicated since the same gene can be in more than one bicluster. In this study, biclustering was used on time series data of estrogen exposed breast cancer cells. The data downloaded from GEO contained 25 arrays, the first 13 and the last 12 of which have different dyeing schemes and correspond to a complete cell cycle each. Biclustering analysis was applied to both of these parts. Additionally, the two parts were merged to obtain a full dataset which corresponds to 2 complete cell cycles. This final merged dataset was also biclustered. When the results for all three datasets were compared, it was seen that the full dataset and the experiment-2 were very similar to each other and they were in agreement with previous results obtained. Furthermore, the full dataset gave more functional categories than both experiment-1 and experiment-2. This means that experiment-1 helped cluster genes better when combined with experiment-2, showing that analysis of two cell cycle data combined together may be more advantageous. Therefore, the full dataset results were used during further analysis.

Since in biclustering, a subset of genes is considered over a subset of conditions, it gives good information about the local expression behavior of genes (over a narrower time window only); the global perspective is lost. Some genes may be working in correlation over a subset of time points (for some part of the cell cycle only), biclustering proves useful for such cases. The grouping together of the genes in bicluster 9 (which are mostly related to transcriptional regulation) is a good example, so are the biclusters 8, 12, 13, 14 and 15. In addition to its advantage in catching local expression patterns, biclustering is also very useful in relating the functions of the genes in one bicluster to the functions of genes in another bicluster (which covers an earlier or later time window). Bicluster 9 (which is up-regulated earlier) and biclusters 8, 12, 13, 14 and 15 follow each other. Another example was the change in function as time points change from bicluster 6, to bicluster 7 and bicluster 8. This example shows that estrogen-responsive genes act in a sequential order. There is a relationship between different biclusters with respect to the time intervals in which their expression profile changes and with respect to the functions they perform. The biclusters which follow an order in time also showed a biologically very meaningful order in function. These examples indicate the power of biclustering in analyzing time series gene expression data.

Biclustering was able to group the genes according to their function and the pathways in which they are involved. During the pathway analysis, it was shown that genes which are in the same pathway were grouped to the similar bicluster while they were in different clusters in the results of the Cicatiello et al. Therefore it was seen that biclustering has an advantage over hierarchical clustering: when it comes to time-series data local behaviors may be very important to understand the pathways relationships that lie in the data.

Biclustering may also prove useful in understanding which pathways or parts of pathways are affected first and which of them are affected later as the cells are

continuously exposed to estrogen. It was seen that estrogen affects a broad range of functions in breast cancer cells and is actually capable of eliciting enough mechanisms to overcome the 6 barriers to develop cancer proposed by Hanahan and Weinberg. This broad activity range indicates that one should be very cautious about uptake of estrogens during menopause or as oral contraceptives taking into account these effects. In the second part of the pathways analysis, some genes belonging to two parts of MAPK signaling pathway (ERK/MAPK cascade and JNK cascade) could be separated from each other. It was shown that the expression behaviors of the up-regulated and down-regulated biclusters are in agreement to each other with respect to the effects they have on JNK cascade (one activating and the other repressing). As pathway information available for genes increases, the results may lead to more elaborate conclusions.

Biclustering results were meaningful when transcriptional factor analysis of the biclusters was done. The analysis gave two TFs: E2F and SRF. SRF was enriched in bicluster 9 and E2F in biclusters 8, 12, 13, 14 and 15. The order of up-regulation of the biclusters and their regulation by E2F and SRF were in agreement with the order of the events that takes place during estrogen exposure response, SRF activating transcription factors and E2F activating the genes responsible for G1/S transition.

Finally, the occurrence frequency of the genes in different biclusters also may give information about the importance of the functions they perform or may help us make predictions about the unannotated genes more confidently since genes whose expression patterns are similar across different biclusters (thus different time points) are most probably functionally more related to each other.

In this study, biclustering was used in time series gene expression data of estrogen exposed breast cancer cells. Interesting relationships between different groups of genes were seen since the genes were better clustered with respect to the specific time intervals in which they are differentially regulated. In addition, this study provides an example to the

use of biclustering with time series data. It is not restricted to this specific data set but can be used with any time series data.

Appendix

GENE LISTS FOR BICLUSTERS

Table A.1 Genes from bicluster 8 involved in cell cycle-related processes, which are inferred from OMIM. The last column gives the number of the hierarchical cluster in the analysis in [13]. The word ‘absent’ indicates that the genes is not included in the hierarchical results since it is eliminated in one of the filtering procedures [13].

Locus Link	Gene Symbol	Function	Hierarchical Cluster
332	<i>BIRC5</i>	G2/M transition of mitotic cell cycle	8
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle	5
983	<i>CDC2</i>	regulation of progression through cell cycle	8
991	<i>CDC20</i>	mitosis	8
1164	<i>CKS2</i>	regulation of cyclin dependent protein kinase activity	8
1736	<i>DKC1</i>	regulation of progression through cell cycle	7
1841	<i>DTYMK</i>	cell cycle	8
3838	<i>KPNA2</i>	G2 phase of mitotic cell cycle	8
4085	<i>MAD2L1</i>	mitotic checkpoint	8
4171	<i>MCM2</i>	cell cycle	7
4176	<i>MCM7</i>	cell cycle	7
4288	<i>MKI67</i>	regulation of progression through cell cycle	absent
4602	<i>MYB</i>	cell cycle	7
4605	<i>MYBL2</i>	regulation of progression through cell cycle	7
4830	<i>NME1</i>	cell cycle	8
5036	<i>PA2G4</i>	cell cycle arrest	7
5901	<i>RAN</i>	mitotic spindle organization and biogenesis	8
5906	<i>RAP1A</i>	negative regulation of progression through cell cycle	8
6790	<i>STK6</i>	spindle organization and biogenesis	8
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle	7
9232	<i>PTTG1</i>	chromosome segregation	7
9787	<i>DLG7</i>	positive regulation of mitotic	absent

		metaphase/anaphase transition	
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle	8
10574	<i>CCT7</i>	regulation of progression through cell cycle	7
10575	<i>CCT4</i>	regulation of progression through cell cycle	7
23397	<i>BRRN1</i>	mitotic chromosome condensation	absent
10112	<i>KIF20A</i>	microtubule motor activity	8
11004	<i>KIF2C</i>	mitosis	8
1058	<i>CENPA</i>	chromosome organization and biogenesis	7
3925	<i>STMN1</i>	mitotic spindle organization and biogenesis	absent

Table A.2 The occurrence frequency of genes in different biclusters.

Locus Link	Gene Symbol	Function	# of Biclusters	Biclusters
8140	<i>SLC7A5</i>	amino acid metabolism and transport	6,7,8,9,10,11,12,13,14,15	10
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle	6,7,8,9,10,11,12,13,14,15	10
6520	<i>SLC3A2</i>	carbohydrate metabolism	6,7,8,10,11,12,13,14,15	9
		amino acid transport		
		cell growth		
5930	<i>RBBP6</i>	nucleic acid binding	6,7,8,10,11,12,13,14,15	9
2288	<i>FKBP4</i>	peptidyl-prolyl cis-trans isomerase activity	6,7,8,10,11,12,13,14,15	9
2289	<i>FKBP5</i>	peptidyl-prolyl cis-trans isomerase activity	6,7,8,10,11,12,13,14,15	9
3309	<i>HSPA5</i>	Anti-apoptosis	6,7,8,10,11,12,13,14,15	9
		unfolded protein binding		
4176	<i>MCM7</i>	transcription	6,7,8,10,11,12,13,14,15	9
		regulation of transcription, DNA-dependent		
		cell cycle		
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle	6,7,8,10,11,12,13,14,15	9
		induction of apoptosis		
6059	<i>ABCE1</i>	unfolded protein binding	6,7,8,10,12,13,15	7
1736	<i>DKC1</i>	regulation of progression through cell cycle	6,7,8,10,12,13,15	7
		telomerase activity		
4288	<i>MKI67</i>	regulation of progression through cell cycle	6,7,8,10,12,13,15	7

5888	<i>RAD51</i>	double-strand break repair via homologous recombination		
		transcription regulator activity		
3939	<i>LDHA</i>	anaerobic glycolysis	6,7,8,12,13,14,15	7
7031	<i>TFF1</i>	carbohydrate metabolism	7,8,10,12,13,14,15	7
		growth factor activity		
7277	<i>TUBA1</i>	microtubule-based movement	7,8,10,12,13,14,15	7
2237	<i>FEN1</i>	double-strand break repair	7,8,12,13,14,15	6
2971	<i>GTF3A</i>	transcription from RNA polymerase III promoter	7,8,12,13,14,15	6
10202	<i>DHRS2</i>	C21-steroid hormone metabolism	7,8,12,13,14,15	6
		negative regulation of progression through cell cycle		
11222	<i>MRPL3</i>	ribosome	7,8,12,13,14,15	6
		protein biosynthesis		
12	<i>SERPINA3</i>	inflammatory response	7,8,12,13,14,15	6
		regulation of lipid metabolism		
7203	<i>CCT3</i>	unfolded protein binding	7,8,12,13,14,15	6
374	<i>AREG</i>	cytokine activity	7,8,10,12,13,14	6
		growth factor activity		
1058	<i>CENPA</i>	chromosome organization and biogenesis		5
5901	<i>RAN</i>	regulation of progression through cell cycle	7,8,13,14,15	5
		mitotic spindle organization and biogenesis		
4141	<i>MARS</i>	methionyl-tRNA aminoacylation	7,8,13,14,15	5
2224	<i>FDPS</i>	cholesterol biosynthesis	7,8,13,14,15	5
		isoprenoid biosynthesis		
4085	<i>MAD2L1</i>	mitotic checkpoint	8,12,13,14,15	5
6790	<i>STK6</i>	cell cycle	8,12,13,14,15	5
		spindle organization and biogenesis		
23397	<i>BRRN1</i>	mitotic cell cycle	8,12,13,14,15	5
		mitotic chromosome condensation		
9141	<i>PDCD5</i>	apoptosis	8,12,13,14,15	5
10797	<i>MTHFD2</i>	folic acid and derivative biosynthesis	8,12,13,14,15	5
27101	<i>CACYBP</i>	ubiquitin cycle	8,12,13,14,15	5

4605	<i>MYBL2</i>	regulation of progression through cell cycle	8,12,13,14,15	5
		transcription factor activity		
332	<i>BIRC5</i>	G2/M transition of mitotic cell cycle	8,12,13,14,15	5
		anti-apoptosis		
991	<i>CDC20</i>	regulation of progression through cell cycle	8,12,13,14,15	5
		ubiquitin cycle		
10419	<i>PRMT5</i>	regulation of mitosis	8,12,13,14,15	5
85004	<i>REERG</i>	estrogen receptor binding	8,12,13,14,15	5
		negative regulation of cell proliferation		
9086	<i>EIF1AY</i>	translational initiation	7,8,12,13,15	5
9156	<i>EXO1</i>	DNA recombination	7,8,12,13,15	5
4171	<i>MCM2</i>	regulation of transcription, DNA-dependent	6,7,8,13,15	5
		cell cycle		
23160	<i>WDR43</i>	Un known function	6,7,8,13,15	5
10439	<i>OLFM1</i>	nervous system development	7,8,13,15	4
27338	<i>UBE2S</i>	ubiquitin cycle	7,8,13,15	4
4522	<i>MTHFD1</i>	folic acid and derivative biosynthesis	8,12,14,15	4
1841	<i>DTYMK</i>	cell cycle	8,12,14,15	4
		cell proliferation		
3329	<i>HSPD1</i>	regulation of apoptosis	7,8,12,13	4
		protein folding		
2618	<i>GART</i>	purine base biosynthesis	7,8,12,13	4
3838	<i>KPNA2</i>	G2 phase of mitotic cell cycle	8,13,14,15	4
9768	<i>KIAA0101</i>	Unknown function	8,13,14,15	4
6652	<i>SORD</i>	sorbitol metabolism	8,13,14,15	4
7083	<i>TK1</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	8,13,14,15	4
1854	<i>DUT</i>	DNA replication	8,13,14,15	4
10606	<i>PAICS</i>	purine nucleotide biosynthesis	8,13,14,15	4
4199	<i>ME1</i>	carbohydrate metabolism	8,13,14,15	4
10112	<i>KIF20A</i>	microtubule-based movement	8,13,14,15	4
4501	<i>MTIX</i>	response to metal ion	8,13,14,15	4
94239	<i>H2AFV</i>	chromosome organization and biogenesis	8,13,14,15	4
5985	<i>RFC5</i>	DNA replication	8,13,14,15	4

		DNA repair		
4830	<i>NME1</i>	regulation of apoptosis	8,13,14,15	4
9787	<i>DLG7</i>	mitotic chromosome movement towards spindle pole	8,12,13,15	4
		cell proliferation		
6633	<i>SNRPD2</i>	nuclear mRNA splicing, via spliceosome	8,12,13,15	4
6636	<i>SNRPF</i>	nuclear mRNA splicing, via spliceosome	8,12,13,15	4
5174	<i>PDZK1</i>	Unknown function	8,12,13,15	4
10436	<i>EMG1</i>	translational elongation	8,12,13,15	4
9232	<i>PTTG1</i>	transcription factor activity	8,12,13,15	4
		chromosome organization and biogenesis		
		Cell cycle	8,13,15	3
5889	<i>RAD51C</i>	DNA repair	8,13,15	3
		DNA recombination		
6119	<i>RPA3</i>	DNA replication	8,13,15	3
		DNA repair		
5479	<i>PPIB</i>	protein folding	8,13,15	3
11004	<i>KIF2C</i>	microtubule-based movement	8,13,15	3
		mitosis		
		cell proliferation		
2674	<i>GFRA1</i>	glial cell line-derived neurotrophic factor receptor activity	8,13,15	3
7162	<i>TPBG</i>	cell motility	8,13,15	3
		cell adhesion		
8533	<i>COP9</i>	signalosome complex	8,13,15	3
5906	<i>RAP1A</i>	cell cycle	8,13,15	3
2305	<i>FOXM1</i>	regulation of transcription, DNA-dependent	8,13,15	3
10575	<i>CCT4</i>	regulation of progression through cell cycle	8,13,15	3
		protein folding		
705	<i>BYSL</i>	cell adhesion	11 only	1
87	<i>ACTN1</i>	structural constituent of cytoskeleton	11 only	1
8314	<i>BAP1</i>	ubiquitin cycle	11 only	1
6432	<i>SFRS7</i>	nuclear mRNA splicing, via spliceosome	11 only	1
6387	<i>CXCL12</i>	cell adhesion	11 only	1

		growth factor activity		
		Chemokin activity		
10561	<i>IFI44</i>	Unknown function	11 only	1
22859	<i>LPHN1</i>	G-protein coupled receptor activity	11 only	1
22984	<i>PDCD11</i>	apoptosis	11 only	1
		regulation of transcription, DNA-dependent		
8165	<i>AKAP1</i>	Unknown function	11 only	1
602	<i>BCL3</i>	regulation of progression through cell cycle	6 only	1
10098	<i>TSPAN5</i>	electron transport	6 only	1
9811	<i>KIAA0427</i>	Unknown function	6 only	1
4057	<i>LTF</i>	humoral immune response	6 only	1
2317	<i>FLNB</i>	actin cytoskeleton organization and biogenesis	6 only	1
3669	<i>ISG20</i>	cell proliferation	6 only	1
		DNA catabolism, exonucleolytic		
5822	<i>PWP2H</i>	signal transduction	6 only	1
9026	<i>HIP1R</i>	biological process unknown	6 only	1
10849	<i>CD3EAP</i>	rRNA transcription	6 only	1
7372	<i>UMPS</i>	nucleoside metabolism	6 only	1
		RNA polymerase I transcription factor activity		
87	<i>ACTN1</i>	actin cytoskeleton	6 only	1
59	<i>ACTA2</i>	actin filament	6 only	1
4928	<i>NUP98</i>	nuclear pore organization and biogenesis	6 only	1
4839	<i>NOL1</i>	regulation of progression through cell cycle	6 only	1
5366	<i>PMAIP1</i>	Unknown function	6 only	1
3303	<i>HSPA1A</i>	protein folding	6 only	1
203068	<i>TUBB</i>	microtubule-based movement	15 only	1
372	<i>ARCN1</i>	intracellular protein transport	15 only	1
9741	<i>LAPTM4A</i>	transport	15 only	1
6880	<i>TAF9</i>	transcription from RNA polymerase II promoter	15 only	1
6628	<i>SNRPB</i>	spliceosome complex	15 only	1
2222	<i>FDFT1</i>	cholesterol biosynthesis	14 only	1
		isoprenoid biosynthesis		
4232	<i>MEST</i>	mesoderm development	14 only	1
4893	<i>NRAS</i>	regulation of progression	14 only	1

		through cell cycle		
7247	<i>TSN</i>	DNA recombination	14 only	1
9256	<i>BZRAP1</i>	biological process unknown	14 only	1
9368	<i>SLC9A3R1</i>	actin cytoskeleton	10 only	1
109	<i>ADCY3</i>	cAMP biosynthesis	10 only	1
4957	<i>ODF2</i>	biological process unknown	10 only	1
2030	<i>SLC29A1</i>	nucleoside transport	10 only	1
10237	<i>SLC35B1</i>	transport	8 only	1
4716	<i>NDUFB10</i>	NADH dehydrogenase activity	8 only	1
4914	<i>NTRK1</i>	transmembrane receptor protein tyrosine kinase signaling pathway	8 only	1
10720	<i>UGT2B11</i>	estrogen metabolism	8 only	1
1973	<i>EIF4A1</i>	translation initiation factor	8 only	1
10237	<i>SLC35B1</i>	transport	8 only	1
7086	<i>TKT</i>	transketolase activity	8 only	1
1164	<i>CKS2</i>	spindle organization and biogenesis	8 only	1
		cell proliferation		
3148	<i>HMGB2</i>	DNA replication	12 only	1
		DNA repair		
4928	<i>NUP98</i>	nuclear pore organization and biogenesis	12 only	1
10213	<i>PSMD14</i>	proteasome complex	12 only	1
9270	<i>ITGB1BP1</i>	cell adhesion	12 only	1

Table A.3. Bicluster 1; list of genes in bicluster 1. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the last column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
654	<i>BMP6</i>	cytokine activity, growth factor activity
1811	<i>SLC26A3</i>	transcription factor activity, transporter activity
4295	<i>MLN</i>	hormone activity, regulation of physiological process
5236	<i>PGM1</i>	carbohydrate metabolism
5289	<i>PIK3C3</i>	inositol or phosphatidylinositol kinase activity
7024	<i>TFCP2</i>	transcription factor activity
7298	<i>TYMS</i>	DNA replication and repair
7764	<i>ZNF217</i>	transcription factor activity
10768	<i>AHCYL1</i>	one-carbon compound metabolism

10798	<i>OR511</i>	olfactory receptor activity
-------	--------------	-----------------------------

Table A.4. Bicluster 2; list of genes in bicluster 2. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
987	<i>LRBA</i>	biological process unknown
1942	<i>EFNA1</i>	cell-cell signaling
4430	<i>MYO1B</i>	myosin, motor activity
4983	<i>OPHN1</i>	axon guidance
7267	<i>TTC3</i>	biological process unknown
7764	<i>ZNF217</i>	transcription factor activity
23158	<i>TBC1D9</i>	calcium ion binding
26959	<i>HBPI</i>	transcription, cell cycle arrest,
51474	<i>LIMA1</i>	focal adhesion
1038	<i>CDR1</i>	biological process unknown

Table A.5. Bicluster 3; list of genes in bicluster 3. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
43	<i>KIAA0143</i>	DNA replication, cell adhesion
347	<i>IGBP1</i>	lipid transporter activity
578	<i>SCARF1</i>	induction of apoptosis
901	<i>Ara.24</i>	regulation of progression through cell cycle
934	<i>RAB7L1</i>	humoral immune response
1153	<i>CIRBP</i>	response to cold
1191	<i>CLU</i>	apoptosis
1374	<i>CPT1A</i>	fatty acid metabolism
1942	<i>EFNA1</i>	cell-cell signaling
2023	<i>ENO1</i>	glycolysis, transcription
2027	<i>ENO3</i>	glycolysis
2065	<i>ERBB3</i>	epidermal growth factor receptor activity
2099	<i>ESR1</i>	steroid hormone receptor activity
2804	<i>GOLGB1</i>	Golgi organization and biogenesis
3488	<i>IGFBP5</i>	regulation of cell growth
4430	<i>MYO1B</i>	motor activity, myosin
4582	<i>MUC1</i>	hormone activity
4680	<i>CEACAM6</i>	cell-cell signaling
4950	<i>OCLN</i>	tight junction
4983	<i>OPHN1</i>	axon guidance

5046	<i>PCSK6</i>	cell-cell signaling
5239	<i>PGM5</i>	carbohydrate metabolism, cell adhesion
5796	<i>PTPRK</i>	receptor activity
6559	<i>SLC12A3</i>	transporter activity
6772	<i>STAT1</i>	regulation of progression through cell cycle, transcription factor activity
6774	<i>STAT3</i>	transcription factor activity
7113	<i>TMPRSS2</i>	scavenger receptor activity
7185	<i>TRAF1</i>	regulation of apoptosis
7253	<i>TSHR</i>	cell-cell signaling
7267	<i>TTC3</i>	biological process unknown
7327	<i>UBE2G2</i>	ubiquitin cycle
7764	<i>ZNF217</i>	transcription factor activity
7903	<i>ST8SIA4</i>	oligosaccharide metabolism
8848	<i>TSC22D1</i>	transcription factor activity
8942	<i>KYNU</i>	tryptophan catabolism
9915	<i>ARNT2</i>	transcription factor activity, response to hypoxia
9969	<i>THRAP1</i>	transcription, androgen receptor signaling pathway
23199	<i>KIAA0182</i>	biological process unknown
25805	<i>BAMBI</i>	biological process unknown
26036	<i>ZNF451</i>	transcription
26959	<i>HBP1</i>	transcription, cell cycle arrest
29799	<i>YPEL1</i>	biological process unknown
55764	<i>IFT122</i>	biological process unknown
283131	<i>TncRNA2</i>	biological process unknown
1038	<i>CDR1</i>	biological process unknown

Table A.6. Biclust 4; list of genes in biclust 4. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
43	<i>KIAA0143</i>	DNA replication, cell adhesion
347	<i>IGBP1</i>	lipid transporter activity
901	<i>45627</i>	regulation of progression through cell cycle
934	<i>RAB7L1</i>	humoral immune response
987	<i>LRBA</i>	biological process unknown
1153	<i>CIRBP</i>	response to cold
1191	<i>CLU</i>	apoptosis
1364	<i>CLDN4</i>	tight junction
1374	<i>CPT1A</i>	fatty acid metabolism
1512	<i>CTSH</i>	proteolysis

1942	<i>EFNA1</i>	cell-cell signaling
2023	<i>ENO1</i>	glycolysis, transcription
2065	<i>ERBB3</i>	epidermal growth factor receptor activity
2099	<i>ESR1</i>	steroid hormone receptor activity
2537	<i>IFI6</i>	anti-apoptosis
2804	<i>GOLGB1</i>	Golgi organization and biogenesis
2896	<i>GRN</i>	cytokine activity
3157	<i>HMGCS1</i>	cholesterol biosynthesis
3488	<i>IGFBP5</i>	insulin-like growth factor binding
4052	<i>LTBP1</i>	growth factor binding
4070	<i>TACSTD2</i>	cell proliferation
4162	<i>MCAM</i>	cell adhesion
4311	<i>MME</i>	cell-cell signaling
4430	<i>MYO1B</i>	motor activity, myosin
4582	<i>MUC1</i>	hormone activity
4950	<i>OCN</i>	tight junction
4958	<i>OMD</i>	cell adhesion
4983	<i>OPHN1</i>	axon guidance
5046	<i>PCSK6</i>	cell-cell signaling
5239	<i>PGM5</i>	cell adhesion
5407	<i>PNLIPRP1</i>	lipid catabolism
5796	<i>PTPRK</i>	receptor activity
6559	<i>SLC12A3</i>	transporter activity
6772	<i>STAT1</i>	regulation of progression through cell cycle, transcription factor activity
6892	<i>TAPBP</i>	immune response
7267	<i>TTC3</i>	biological process unknown
7327	<i>UBE2G2</i>	ubiquitin cycle
7464	<i>CORO2A</i>	glutamine biosynthesis
7764	<i>ZNF217</i>	transcription factor activity
8848	<i>TSC22D1</i>	transcription factor activity
8934	<i>RAB7L1</i>	biological process unknown
8991	<i>SELENBP1</i>	selenium binding
9168	<i>TMSB10</i>	cytoskeleton organization and biogenesis
9518	<i>GDF15</i>	cell-cell signaling
9637	<i>FEZ2</i>	axon guidance
9780	<i>FAM38A</i>	biological process unknown
9891	<i>NUAK1</i>	biological process unknown
9915	<i>ARNT2</i>	response to hypoxia, transcription factor activity
9969	<i>THRAP1</i>	regulation of transcription, DNA-dependent
10114	<i>HIPK3</i>	anti-apoptosis

10370	<i>CITED2</i>	transcription factor activity
10513	<i>APPBP2</i>	microtubule motor activity
10957	<i>PNRC1</i>	transcription
23158	<i>TBC1D9</i>	calcium ion binding
25805	<i>BAMBI</i>	biological process unknown
26036	<i>ZNF451</i>	transcription
26959	<i>HBP1</i>	cell cycle arrest, transcription
29799	<i>YPEL1</i>	biological process unknown
30836	<i>DNTTIP2</i>	receptor activity
51474	<i>LIMA1</i>	focal adhesion
57007	<i>CMKOR1</i>	biological process unknown
54735	<i>HUR7</i>	biological process unknown
83641	<i>FAM107B</i>	biological process unknown
79441	<i>C4orf15</i>	biological process unknown
30836	<i>DNTTIP2</i>	biological process unknown

Table A.7. Bicluster 5; list of genes in bicluster 5. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
43	<i>KIAA0143</i>	DNA replication, cell adhesion
247	<i>TSN</i>	negative regulation of progression through cell cycle
302	<i>ANXA2</i>	skeletal development
347	<i>IGBP1</i>	lipid transporter activity
489	<i>NRAS</i>	metabolism
549	<i>LGR5</i>	mRNA catabolism
573	<i>PTGER2</i>	anti-apoptosis
578	<i>SCARF1</i>	induction of apoptosis
843	<i>CASP10</i>	induction of apoptosis
934	<i>RAB7L1</i>	humoral immune response
1015	<i>CDH17</i>	cell adhesion
1191	<i>CLU</i>	apoptosis
1314	<i>COPA</i>	hormone activity
1376	<i>CPT2</i>	lipid metabolism
1438	<i>CSF2RA</i>	receptor activity
1954	<i>MEGF8</i>	biological process unknown
1993	<i>ELAVL2</i>	regulation of transcription, DNA-dependent
2027	<i>ENO3</i>	glycolysis
2065	<i>ERBB3</i>	epidermal growth factor receptor activity
2537	<i>IFI6</i>	anti-apoptosis
2804	<i>GOLGB1</i>	Golgi organization and biogenesis

2966	<i>GTF2H2</i>	transcription factor activity
2990	<i>GUSB</i>	glycosaminoglycan catabolism
3423	<i>IDS</i>	glycosaminoglycan metabolism
3437	<i>IFIT3</i>	biological process unknown
3633	<i>INPP5B</i>	biological process unknown
3646	<i>EIF3S6</i>	translation initiation factor activity
3725	<i>JUN</i>	transcription factor activity
3779	<i>KCNMB1</i>	ion channel activity
3786	<i>KCNQ3</i>	voltage-gated potassium channel activity
4070	<i>TACSTD2</i>	cell proliferation
4090	<i>SMAD5</i>	transcription factor activity
4163	<i>MCC</i>	negative regulation of progression through cell cycle
4214	<i>MAP3K1</i>	MAP kinase kinase kinase activity
4582	<i>MUC1</i>	hormone activity
4680	<i>CEACAM6</i>	cell-cell signaling
4983	<i>OPHN1</i>	axon guidance
5046	<i>PCSK6</i>	proteolysis
5108	<i>PCM1</i>	transcription
5239	<i>PGM5</i>	cell adhesion
5481	<i>PPID</i>	FK506 binding
6300	<i>MAPK12</i>	cell cycle arrest
6558	<i>SLC12A2</i>	transporter activity
6760	<i>SS18</i>	biological process unknown
6772	<i>STAT1</i>	regulation of progression through cell cycle, transcription factor activity
6932	<i>TCF7</i>	transcription
6997	<i>TDGF1</i>	growth factor activity
7764	<i>ZNF217</i>	transcription factor activity
7903	<i>ST8SIA4</i>	nervous system development
8416	<i>ANXA9</i>	cell-cell adhesion
8578	<i>SCARF1</i>	cell adhesion
8731	<i>RNMT</i>	methyltransferase activity
8740	<i>TNFSF14</i>	induction of apoptosis
8808	<i>IL1RL2</i>	transmembrane receptor activity
8991	<i>SELENBP1</i>	biological process unknown
9168	<i>TMSB10</i>	cytoskeleton organization and biogenesis
9283	<i>GPR37L1</i>	biological process unknown
9308	<i>CD83</i>	humoral immune response
9637	<i>FEZ2</i>	axon guidance
10611	<i>PDLIM5</i>	receptor signaling complex scaffold activity
10787	<i>NCKAP1</i>	apoptosis

10949	<i>HNRPA0</i>	heterogeneous nuclear ribonucleoprotein complex
23098	<i>SARM1</i>	innate immune response
23125	<i>CAMTA2</i>	transcription
23138	<i>N4BP32</i>	biological process unknown
23175	<i>LPIN1</i>	fat cell differentiation
23221	<i>RHOBTB2</i>	small GTPase mediated signal transduction
25805	<i>BAMBI</i>	biological process unknown
26959	<i>HBP1</i>	cell cycle arrest
29799	<i>YPEL1</i>	biological process unknown
30836	<i>DNTTIP2</i>	receptor activity
51474	<i>LIMA1</i>	focal adhesion
57007	<i>CMKOR1</i>	biological process unknown
65125	<i>WNK1</i>	regulation of cellular process
84830	<i>C6orf105</i>	biological process unknown
283131	<i>TncRNA2</i>	biological process unknown
54735	<i>HUR7</i>	biological process unknown
57238	<i>KIAA0492</i>	biological process unknown
784	<i>CACNB3</i>	biological process unknown
9710	<i>KIAA0355</i>	biological process unknown
9747	<i>KIAA0738</i>	biological process unknown

Table A.8. Bicluster 6; list of genes in bicluster 6. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the last column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
59	<i>GARNL1</i>	structural constituent of cytoskeleton
87	<i>E2F6</i>	transcription
262	<i>AMD1</i>	spermidine biosynthesis
595	<i>PSCDBP</i>	G1/S transition of mitotic cell cycle
602	<i>BCL3</i>	regulation of progression through cell cycle
1478	<i>CSTF2</i>	mRNA cleavage
1736	<i>DKC1</i>	regulation of progression through cell cycle
1836	<i>SLC26A2</i>	transporter activity
1960	<i>EGR3</i>	transcription factor activity
2289	<i>FKBP5</i>	FK506 binding
2317	<i>FLNB</i>	cell differentiation
3309	<i>HSPA5</i>	anti-apoptosis
3310	<i>HSPA6</i>	response to unfolded protein
3487	<i>IGFBP4</i>	regulation of cell growth
3669	<i>ISG20</i>	cell proliferation
3692	<i>ITGB4BP</i>	translational initiation

3939	<i>LDHA</i>	anaerobic glycolysis
4057	<i>LTF</i>	humoral immune response
4171	<i>MCM2</i>	cell cycle
4176	<i>MCM7</i>	cell cycle
4288	<i>MKI67</i>	regulation of progression through cell cycle
4602	<i>MYB</i>	cell cycle
4839	<i>NOL1</i>	regulation of progression through cell cycle
4928	<i>NUP98</i>	DNA replication, nucleocytoplasmic transport
5352	<i>PLOD2</i>	protein metabolism
5366	<i>PMAIP1</i>	biological process unknown
5822	<i>PWP2H</i>	signal transduction
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5930	<i>RBBP6</i>	biological process unknown
5979	<i>RET</i>	receptor activity
6059	<i>ABCE1</i>	unfolded protein binding
6426	<i>SFRS1</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	cell growth
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7187	<i>TRAF3</i>	induction of apoptosis
7343	<i>UBTF</i>	transcription
7425	<i>VGF</i>	growth factor activity
8140	<i>SLC7A5</i>	amino acid transport
8565	<i>YARS</i>	protein biosynthesis
8614	<i>STC2</i>	cell-cell signaling
9120	<i>SLC16A6</i>	transporter activity
9221	<i>NOLC1</i>	cell cycle
9775	<i>DDX48</i>	rRNA processing
9791	<i>PTDSS1</i>	phospholipid biosynthesis
9811	<i>KIAA0427</i>	biological process unknown
10098	<i>TSPAN5</i>	electron transport
10163	<i>WASF2</i>	cytoskeleton
10197	<i>PSME3</i>	proteasome activator activity
10236	<i>HNRPR</i>	mRNA processing
10525	<i>HYOU1</i>	response to stress
10849	<i>CD3EAP</i>	rRNA transcription
23160	<i>WDR43</i>	biological process unknown
51727	<i>CMPK</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
92856	<i>IMP4</i>	rRNA processing
10429	<i>Ara.24</i>	biological process unknown
2288	<i>FKBP4</i>	biological process unknown
23524	<i>SRRM2</i>	biological process unknown

3303	<i>HSPA1A</i>	biological process unknown
------	---------------	----------------------------

Table A.9. Bicluster 7; list of genes in bicluster 7. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
12	<i>SERPINA3</i>	inflammatory response
262	<i>AMD1</i>	spermine biosynthesis
374	<i>AREG</i>	growth factor activity
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle
1058	<i>CENPA</i>	chromosome organization and biogenesis (sensu Eukaryota)
1736	<i>DKC1</i>	regulation of progression through cell cycle
1836	<i>SLC26A2</i>	transporter activity
2224	<i>FDPS</i>	isoprenoid biosynthesis
2237	<i>FEN1</i>	double-strand break repair
2289	<i>FKBP5</i>	FK506 binding
2618	<i>GART</i>	purine base biosynthesis
2971	<i>GTF3A</i>	rRNA transcription
3301	<i>DNAJA1</i>	response to unfolded protein
3309	<i>HSPA5</i>	anti-apoptosis
3310	<i>HSPA6</i>	response to unfolded protein
3329	<i>HSPD1</i>	regulation of apoptosis
3376	<i>IARS</i>	protein biosynthesis
3455	<i>IFNAR2</i>	JAK-STAT cascade
3842	<i>TNPO1</i>	protein import into nucleus, translocation
3939	<i>LDHA</i>	anaerobic glycolysis
4141	<i>MARS</i>	protein biosynthesis
4171	<i>MCM2</i>	cell cycle
4176	<i>MCM7</i>	cell cycle
4288	<i>MKI67</i>	regulation of progression through cell cycle
4602	<i>MYB</i>	cell cycle
4691	<i>NCL</i>	biological process unknown
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
5036	<i>PA2G4</i>	cell cycle arrest
5214	<i>PFKP</i>	glycolysis
5352	<i>PLOD2</i>	protein metabolism
5717	<i>COP9</i>	proteasome complex (sensu Eukaryota)
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5901	<i>RAN</i>	regulation of progression through cell cycle
5930	<i>RBBP6</i>	biological process unknown

5979	<i>RET</i>	receptor activity
6059	<i>ABCE1</i>	electron carrier activity
6434	<i>SFRS10</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	amino acid transport
7031	<i>TFF1</i>	growth factor activity
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7277	<i>TUBA1</i>	microtubule-based movement
7372	<i>UMPS</i>	nucleoside metabolism
7425	<i>VGF</i>	growth factor activity
8140	<i>SLC7A5</i>	amino acid transport
8565	<i>YARS</i>	protein biosynthesis
8614	<i>STC2</i>	hormone activity
8662	<i>EIF3S9</i>	translation initiation factor activity
8829	<i>NRP1</i>	axon guidance
9086	<i>EIF1AY</i>	translation initiation factor activity
9156	<i>EXO1</i>	mismatch repair, DNA recombination
9221	<i>NOLC1</i>	cell cycle
9775	<i>DDX48</i>	rRNA processing
10197	<i>PSME3</i>	proteasome activator activity
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle
10236	<i>HNRPR</i>	mRNA processing
10439	<i>OLFM1</i>	development
10525	<i>HYOU1</i>	response to stress
10528	<i>NOL5A</i>	rRNA processing
10574	<i>CCT7</i>	regulation of progression through cell cycle
10963	<i>STIP1</i>	response to stress
11222	<i>MRPL3</i>	protein biosynthesis
11260	<i>XPOT</i>	transport
22948	<i>CCT5</i>	unfolded protein binding
23160	<i>WDR43</i>	biological process unknown
27338	<i>UBE2S</i>	ubiquitin cycle
51582	<i>AZIN1</i>	polyamine biosynthesis
51727	<i>CMPK</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
92856	<i>IMP4</i>	rRNA processing
6426	<i>SFRS1</i>	biological process unknown
2288	<i>FKBP4</i>	biological process unknown

Table A.10. Bicluster 8; list of genes in bicluster 8. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the last column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
---------	-------------	--------------------------

12	<i>SERPINA3</i>	regulation of lipid metabolism
262	<i>AMD1</i>	spermidine biosynthesis
292	<i>SLC25A5</i>	transporter activity
332	<i>BIRC5</i>	G2/M transition of mitotic cell cycle, anti-apoptosis
374	<i>AREG</i>	growth factor activity
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle
645	<i>BLVRB</i>	cellular metabolism
708	<i>CIQBP</i>	immune response
983	<i>CDC2</i>	regulation of progression through cell cycle
991	<i>CDC20</i>	regulation of progression through cell cycle
1058	<i>CENPA</i>	chromosome organization and biogenesis (sensu Eukaryota)
1164	<i>CKS2</i>	spindle organization and biogenesis
1478	<i>CSTF2</i>	mRNA polyadenylation
1509	<i>CTSD</i>	Proteolysis
1736	<i>DKC1</i>	regulation of progression through cell cycle, telomerase activity
1841	<i>DTYMK</i>	cell cycle
1854	<i>DUT</i>	DNA replication
1960	<i>EGR3</i>	transcription factor activity
1973	<i>EIF4A1</i>	translation initiation factor activity
2224	<i>FDPS</i>	isoprenoid biosynthesis
2237	<i>FEN1</i>	double-strand break repair
2288	<i>FKBP4</i>	protein folding
2289	<i>FKBP5</i>	protein folding
2305	<i>FOXO1</i>	transcription factor activity
2618	<i>GART</i>	purine base biosynthesis
2674	<i>GFRA1</i>	glial cell line-derived neurotrophic factor receptor activity
2971	<i>GTF3A</i>	rRNA transcription
3309	<i>HSPA5</i>	anti-apoptosis
3329	<i>HSPD1</i>	regulation of apoptosis
3692	<i>ITGB4BP</i>	translational initiation
3835	<i>KIF22</i>	microtubule-based movement
3838	<i>KPNA2</i>	G2 phase of mitotic cell cycle
3921	<i>RPSA</i>	protein biosynthesis
3925	<i>STMN1</i>	mitotic spindle organization and biogenesis
3939	<i>LDHA</i>	anaerobic glycolysis
4085	<i>MAD2L1</i>	mitotic checkpoint
4141	<i>MARS</i>	protein biosynthesis
4171	<i>MCM2</i>	cell cycle
4176	<i>MCM7</i>	cell cycle
4199	<i>ME1</i>	biological process unknown
4288	<i>MKI67</i>	regulation of progression through cell cycle

4501	<i>MTIX</i>	response to metal ion
4522	<i>MTHFD1</i>	one-carbon compound metabolism
4602	<i>MYB</i>	cell cycle
4605	<i>MYBL2</i>	regulation of progression through cell cycle
4716	<i>NDUFB10</i>	NADH dehydrogenase (ubiquinone) activity
4830	<i>NME1</i>	regulation of apoptosis
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
4914	<i>NTRK1</i>	cell differentiation
5036	<i>PA2G4</i>	cell cycle arrest, rRNA processing
5174	<i>PDZK1</i>	biological process unknown
5315	<i>PKM2</i>	Glycolysis
5479	<i>PIIB</i>	unfolded protein binding
5694	<i>PSMB6</i>	ubiquitin-dependent protein catabolism
5717	<i>COP9</i>	proteasome complex (sensu Eukaryota)
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5889	<i>RAD51C</i>	DNA recombination and repair
5901	<i>RAN</i>	regulation of progression through cell cycle
5906	<i>RAP1A</i>	cell cycle
5930	<i>RBBP6</i>	biological process unknown
5979	<i>RET</i>	protein-tyrosine kinase activity
5985	<i>RFC5</i>	DNA recombination and repair
6059	<i>ABCE1</i>	unfolded protein binding
6119	<i>RPA3</i>	DNA recombination and repair
6426	<i>SFRS1</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	cell growth
6633	<i>SNRPD2</i>	nuclear mRNA splicing, via spliceosome
6636	<i>SNRPF</i>	nuclear mRNA splicing, via spliceosome
6652	<i>SORD</i>	sorbitol metabolism
6790	<i>STK6</i>	spindle organization and biogenesis
7031	<i>TFF1</i>	growth factor activity
7083	<i>TK1</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
7162	<i>TPBG</i>	cell adhesion
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7203	<i>CCT3</i>	unfolded protein binding
7277	<i>TUBA1</i>	microtubule-based movement
7425	<i>VGF</i>	growth factor activity
8140	<i>SLC7A5</i>	amino acid transport
8533	<i>COPS3</i>	signalosome complex
8829	<i>NRP1</i>	axon guidance
8914	<i>TIMELESS</i>	circadian rhythm
9086	<i>EIF1AY</i>	translational initiation

9141	<i>PDCD5</i>	induction of apoptosis
9156	<i>EXO1</i>	DNA recombination and repair
9232	<i>PTTG1</i>	chromosome organization and biogenesis
9446	<i>GSTO1</i>	biological process unknown
9601	<i>PDIA4</i>	protein secretion
9619	<i>ABCG1</i>	cholesterol metabolism
9768	<i>KIAA0101</i>	biological process unknown
9787	<i>DLG7</i>	positive regulation of mitotic metaphase/anaphase transition
9791	<i>PTDSS1</i>	phospholipid biosynthesis
10112	<i>KIF20A</i>	microtubule-based movement
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle
10237	<i>SLC35B1</i>	sugar porter activity
10419	<i>PRMT5</i>	cell proliferation
10436	<i>EMG1</i>	translational elongation
10439	<i>OLFM1</i>	Development
10525	<i>HYOU1</i>	unfolded protein binding
10574	<i>CCT7</i>	regulation of progression through cell cycle
10575	<i>CCT4</i>	regulation of progression through cell cycle
10606	<i>PAICS</i>	purine nucleotide biosynthesis
10720	<i>UGT2B11</i>	estrogen metabolism
10797	<i>MTHFD2</i>	one-carbon compound metabolism
10963	<i>STIP1</i>	response to stres
11004	<i>KIF2C</i>	microtubule-based movement
11222	<i>MRPL3</i>	protein biosynthesis
11260	<i>XPOT</i>	Transport
22948	<i>CCT5</i>	unfolded protein binding
23160	<i>WDR43</i>	biological process unknown
23306	<i>KIAA0286</i>	biological process unknown
23397	<i>BRRN1</i>	mitotic chromosome condensation
23450	<i>SF3B3</i>	nuclear mRNA splicing, via spliceosome
27101	<i>CACYBP</i>	ubiquitin cycle
27338	<i>UBE2S</i>	ubiquitin cycle
51727	<i>CMPK</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
60481	<i>ELOVL5</i>	biological process unknown
85004	<i>RERG</i>	estrogen receptor binding
92856	<i>IMP4</i>	rRNA processing
94239	<i>H2AFV</i>	chromosome organization and biogenesis (sensu Eukaryota)
10551	<i>AGR2</i>	biological process unknown

Table A.11. Bicluster 9; list of genes in bicluster 9. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle
1543	<i>CYP11A1</i>	electron transport
1876	<i>E2F6</i>	regulation of progression through cell cycle
1958	<i>EGR1</i>	transcription factor activity
1960	<i>EGR3</i>	transcription factor activity
2353	<i>FOS</i>	transcription factor activity
2355	<i>FOSL2</i>	transcription factor activity
2683	<i>B4GALT1</i>	carbohydrate metabolism
3491	<i>CYR61</i>	regulation of cell growth, cell adhesion
4084	<i>MXD1</i>	transcription factor activity
4609	<i>MYC</i>	transcription factor activity
5979	<i>RET</i>	receptor activity
7022	<i>TFAP2C</i>	transcription factor activity
7071	<i>KLF10</i>	transcription factor activity
7169	<i>TPM2</i>	cytoskeleton
8140	<i>SLC7A5</i>	amino acid metabolism
8614	<i>STC2</i>	hormone activity
8870	<i>IER3</i>	anti-apoptosis
9021	<i>SOCS3</i>	anti-apoptosis
9026	<i>HIP1R</i>	biological process unknown
9120	<i>SLC16A6</i>	transporter activity
10395	<i>DLC1</i>	cytoskeleton organization and biogenesis
10938	<i>EHD1</i>	biological process unknown
80139	<i>ZNF703</i>	biological process unknown

Table A.12. Bicluster 10; list of genes in bicluster 10. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
109	<i>ADCY3</i>	intracellular signaling cascade
374	<i>AREG</i>	growth factor activity
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle
1509	<i>CTSD</i>	proteolysis
1736	<i>DKC1</i>	regulation of progression through cell cycle, telomerase activity
1960	<i>EGR3</i>	transcription factor activity
2030	<i>SLC29A1</i>	nucleoside transport
2288	<i>FKBP4</i>	FK506 binding

2289	<i>FKBP5</i>	FK506 binding
3309	<i>HSPA5</i>	anti-apoptosis
3487	<i>IGFBP4</i>	cell proliferation
4176	<i>MCM7</i>	cell cycle
4288	<i>MKI67</i>	regulation of progression through cell cycle
4602	<i>MYB</i>	cell cycle
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
4957	<i>ODF2</i>	biological process unknown
5352	<i>PLOD2</i>	protein metabolism
5930	<i>RBBP6</i>	biological process unknown
5979	<i>RET</i>	cell surface receptor linked signal transduction
6059	<i>ABCE1</i>	unfolded protein binding
6434	<i>SFRS10</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	amino acid transport
7031	<i>TFF1</i>	growth factor activity
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7277	<i>TUBA1</i>	microtubule-based movement
7425	<i>VGF</i>	growth factor activity
8140	<i>SLC7A5</i>	amino acid transport
8662	<i>EIF3S9</i>	translational initiation
9221	<i>NOLC1</i>	cell cycle
9368	<i>SLC9A3R1</i>	Wnt receptor signaling pathway
10197	<i>PSME3</i>	proteasome activator activity
10963	<i>STIP1</i>	response to stress
11260	<i>XPOT</i>	transport
10528	<i>NOL5A</i>	biological process unknown

Table A.13. Bicluster 11; list of genes in bicluster 11. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
71	<i>ACTG1</i>	structural constituent of cytoskeleton
87	<i>ACTN1</i>	structural constituent of cytoskeleton
595	<i>CCND1</i>	G1/S transition of mitotic cell cycle
705	<i>BYSL</i>	cell adhesion
2288	<i>FKBP4</i>	FK506 binding
2289	<i>FKBP5</i>	FK506 binding
2683	<i>B4GALT1</i>	carbohydrate metabolism
3309	<i>HSPA5</i>	anti-apoptosis
3487	<i>IGFBP4</i>	regulation of cell growth
4176	<i>MCM7</i>	cell cycle

4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
5036	<i>PA2G4</i>	cell cycle arrest
5930	<i>RBBP6</i>	biological process unknown
5979	<i>RET</i>	cell surface receptor linked signal transduction
6387	<i>CXCL12</i>	growth factor activity
6426	<i>SFRS1</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	amino acid transport
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7425	<i>VGF</i>	growth factor activity
8140	<i>SLC7A5</i>	amino acid metabolism
8165	<i>AKAP1</i>	mitochondrion
8314	<i>BAP1</i>	ubiquitin cycle
8565	<i>YARS</i>	protein biosynthesis
8614	<i>STC2</i>	cell-cell signaling
8662	<i>EIF3S9</i>	translational initiation
9221	<i>NOLC1</i>	cell cycle
10163	<i>WASF2</i>	actin cytoskeleton organization and biogenesis
10197	<i>PSME3</i>	proteasome activator complex
10528	<i>NOL5A</i>	rRNA processing
10561	<i>IFI44</i>	cytoplasm
10963	<i>STIP1</i>	response to stress
22859	<i>LPHN1</i>	receptor activity
22984	<i>PDCD11</i>	apoptosis, rRNA processing
92856	<i>IMP4</i>	rRNA processing

Table A.14. Biclusters 12; list of genes in bicluster 12. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
12	<i>CNR1</i>	regulation of lipid metabolism
262	<i>AMD1</i>	spermidine biosynthesis
332	<i>HSPD1</i>	G2/M transition of mitotic cell cycle
374	<i>CPT1A</i>	growth factor activity
595	<i>PSCDBP</i>	G1/S transition of mitotic cell cycle
991	<i>SELENBP1</i>	regulation of progression through cell cycle
1509	<i>CTSD</i>	proteolysis
1736	<i>DKC1</i>	regulation of progression through cell cycle
1836	<i>SLC26A2</i>	transporter activity
1841	<i>DTYMK</i>	cell cycle
2237	<i>FEN1</i>	double-strand break repair
2288	<i>FKBP4</i>	FK506 binding

2289	<i>FKBP5</i>	FK506 binding
2618	<i>GART</i>	purine base biosynthesis
2971	<i>GTF3A</i>	rRNA transcription
3148	<i>HMGB2</i>	base-excision repair, DNA ligation
3148	<i>HMGB2</i>	DNA replication and repair
3301	<i>DNAJA1</i>	response to unfolded protein
3309	<i>HSPA5</i>	anti-apoptosis
3329	<i>HSPD1</i>	regulation of apoptosis
3692	<i>ITGB4BP</i>	translational initiation
3939	<i>LDHA</i>	anaerobic glycolysis
4085	<i>MAD2L1</i>	cell cycle
4176	<i>MCM7</i>	cell cycle
4288	<i>MKI67</i>	regulation of progression through cell cycle
4522	<i>MTHFD1</i>	one-carbon compound metabolism
4602	<i>MYB</i>	cell cycle
4605	<i>MYBL2</i>	regulation of progression through cell cycle
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
4927	<i>NUP88</i>	protein transport
5352	<i>PLOD2</i>	protein metabolism
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5930	<i>RBBP6</i>	
6059	<i>ABCE1</i>	electron transport
6434	<i>SFRS10</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	cell growth
6633	<i>SNRPD2</i>	nuclear mRNA splicing, via spliceosome
6636	<i>SNRPF</i>	nuclear mRNA splicing, via spliceosome
6790	<i>STK6</i>	spindle organization and biogenesis
7031	<i>TFF1</i>	growth factor activity
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7184	<i>HSP90B1</i>	anti-apoptosis
7277	<i>TUBA1</i>	microtubule-based movement
7372	<i>UMPS</i>	nucleoside metabolism
7425	<i>VGF</i>	growth factor activity
7443	<i>VRK1</i>	
8140	<i>SLC7A5</i>	amino acid transport
9086	<i>EIF1AY</i>	translational initiation
9141	<i>PDCD5</i>	induction of apoptosis
9156	<i>EXO1</i>	nucleotide-excision repair
9221	<i>NOLC1</i>	cell cycle
9232	<i>PTTG1</i>	cell cycle
9270	<i>ITGB1BP1</i>	cell adhesion

9787	<i>DLG7</i>	positive regulation of mitotic metaphase/anaphase transition
10131	<i>TRAP1</i>	unfolded protein binding
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle
10213	<i>PSMD14</i>	proteasome complex (sensu Eukaryota)
10236	<i>HNRPR</i>	mRNA processing
10419	<i>PRMT5</i>	cell proliferation
10436	<i>EMG1</i>	translational elongation
10574	<i>CCT7</i>	regulation of progression through cell cycle
10797	<i>MTHFD2</i>	one-carbon compound metabolism
11222	<i>MRPL3</i>	protein biosynthesis
11260	<i>XPOT</i>	transport
22948	<i>CCT5</i>	unfolded protein binding
23306	<i>KIAA0286</i>	biological process unknown
23397	<i>BRRN1</i>	mitotic cell cycle
27101	<i>CACYBP</i>	ubiquitin cycle
51765	<i>MST4</i>	regulation of apoptosis
85004	<i>RERG</i>	negative regulation of cell proliferation

Table A.15. Biclusters 13; list of genes in bicluster 13. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
12	<i>CNR1</i>	regulation of lipid metabolism
332	<i>HSPD1</i>	G2/M transition of mitotic cell cycle
374	<i>CPT1A</i>	growth factor activity
595	<i>PSCDBP</i>	G1/S transition of mitotic cell cycle
645	<i>BLVRB</i>	cellular metabolism
701	<i>BUB1B</i>	spindle organization and biogenesis, mitotic checkpoint
983	<i>KIAA0125</i>	regulation of progression through cell cycle
991	<i>SELENBP1</i>	regulation of progression through cell cycle
1058	<i>CENPA</i>	chromosome organization and biogenesis (sensu Eukaryota)
1509	<i>CTSD</i>	proteolysis
1736	<i>DKC1</i>	regulation of progression through cell cycle
1854	<i>DUT</i>	nucleotide metabolism
1960	<i>EGR3</i>	transcription factor activity
2224	<i>FDPS</i>	cholesterol biosynthesis
2237	<i>FEN1</i>	double-strand break repair
2288	<i>FKBP4</i>	FK506 binding
2289	<i>FKBP5</i>	FK506 binding
2305	<i>FOXMI</i>	transcription factor activity
2618	<i>GART</i>	purine base biosynthesis

2971	<i>GTF3A</i>	rRNA transcription
3309	<i>HSPA5</i>	anti-apoptosis
3329	<i>HSPD1</i>	regulation of apoptosis
3376	<i>IARS</i>	protein biosynthesis
3838	<i>KPNA2</i>	G2 phase of mitotic cell cycle
3939	<i>LDHA</i>	anaerobic glycolysis
4085	<i>MAD2L1</i>	mitotic checkpoint
4141	<i>MARS</i>	protein biosynthesis
4171	<i>MCM2</i>	cell cycle
4176	<i>MCM7</i>	cell cycle
4199	<i>ME1</i>	carbohydrate metabolism
4288	<i>MKI67</i>	regulation of progression through cell cycle
4501	<i>MT1X</i>	response to metal ion
4602	<i>MYB</i>	cell cycle
4605	<i>MYBL2</i>	regulation of progression through cell cycle
4830	<i>C6orf105</i>	regulation of apoptosis
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
5214	<i>PFKP</i>	glycolysis
5479	<i>PPIB</i>	unfolded protein binding
5717	<i>PSMD11</i>	proteasome complex (sensu Eukaryota)
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5889	<i>RAD51C</i>	DNA recombination and repair
5901	<i>Ara.24</i>	regulation of progression through cell cycle
5906	<i>RAP1A</i>	cell cycle
5930	<i>RBBP6</i>	biological process unknown
5985	<i>RFC5</i>	DNA replication and repair
6059	<i>ABCE1</i>	protein folding
6059	<i>ABCE1</i>	protein folding
6119	<i>RPA3</i>	DNA replication and repair
6520	<i>SLC3A2</i>	amino acid transport
6633	<i>SNRPD2</i>	nuclear mRNA splicing, via spliceosome
6636	<i>SNRPF</i>	nuclear mRNA splicing, via spliceosome
6652	<i>SORD</i>	sorbitol metabolism
6713	<i>SQLE</i>	sterol biosynthesis
6790	<i>STK6</i>	spindle organization and biogenesis
7031	<i>TFF1</i>	growth factor activity
7083	<i>TK1</i>	DNA replication
7162	<i>TPBG</i>	cell adhesion
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7184	<i>HSP90B1</i>	anti-apoptosis
7277	<i>TUBA1</i>	microtubule-based movement

7425	<i>VGF</i>	growth factor activity
7443	<i>VRK1</i>	biological process unknown
8140	<i>SLC7A5</i>	amino acid metabolism
8533	<i>COP9</i>	signalosome complex
8829	<i>NRP1</i>	axon guidance
8914	<i>TIMELESS</i>	circadian rhythm
9086	<i>EIF1AY</i>	translational initiation
9141	<i>PDCD5</i>	induction of apoptosis
9156	<i>EXO1</i>	nucleotide-excision repair
9221	<i>NOLC1</i>	cell cycle
9232	<i>PTTG1</i>	chromosome organization and biogenesis
9768	<i>KIAA0101</i>	biological process unknown
9787	<i>DLG7</i>	positive regulation of mitotic metaphase/anaphase transition
10112	<i>KIF20A</i>	microtubule-based movement
10131	<i>TRAP1</i>	unfolded protein binding
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle
10236	<i>HNRPR</i>	mRNA processing
10419	<i>PRMT5</i>	regulation of mitosis
10436	<i>EMG1</i>	translational elongation
10439	<i>OLFM1</i>	development
10575	<i>CCT4</i>	regulation of progression through cell cycle
10606	<i>PAICS</i>	purine base biosynthesis
10797	<i>MTHFD2</i>	one-carbon compound metabolism
10963	<i>STIP1</i>	response to stress
11004	<i>KIF2C</i>	microtubule-based movement
11222	<i>MRPL3</i>	protein biosynthesis
11260	<i>XPOT</i>	transport
22948	<i>CCT5</i>	unfolded protein binding
23160	<i>WDR43</i>	biological process unknown
23397	<i>BRRN1</i>	mitotic cell cycle
27101	<i>CACYBP</i>	ubiquitin cycle
27338	<i>UBE2S</i>	ubiquitin cycle
51765	<i>MST4</i>	regulation of apoptosis
85004	<i>RERG</i>	negative regulation of cell proliferation
94239	<i>H2AFV</i>	chromosome organization and biogenesis (sensu Eukaryota)

Table A.16. Bicluster 14; list of genes in bicluster 14. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
12	<i>CNR1</i>	regulation of lipid metabolism

292	<i>SLC25A5</i>	mitochondrial transport
332	<i>HSPD1</i>	G2/M transition of mitotic cell cycle
374	<i>CPT1A</i>	growth factor activity
595	<i>PSCDBP</i>	G1/S transition of mitotic cell cycle
645	<i>BLVRB</i>	cellular metabolism
708	<i>TKT</i>	immune response
1058	<i>CENPA</i>	chromosome organization and biogenesis (sensu Eukaryota)
991	<i>SELENBP1</i>	regulation of progression through cell cycle
1478	<i>CSTF2</i>	mRNA polyadenylation
1509	<i>CTSD</i>	proteolysis
1841	<i>DTYMK</i>	cell cycle
1854	<i>DUT</i>	DNA replication
1960	<i>EGR3</i>	transcription factor activity
2222	<i>FDFT1</i>	cholesterol biosynthesis
2224	<i>FDPS</i>	isoprenoid biosynthesis
2237	<i>FEN1</i>	double-strand break repair
2288	<i>FKBP4</i>	FK506 binding
2289	<i>FKBP5</i>	FK506 binding
2971	<i>GTF3A</i>	rRNA transcription
3309	<i>HSPA5</i>	anti-apoptosis
3838	<i>KPNA2</i>	G2 phase of mitotic cell cycle
3925	<i>STMN1</i>	mitotic spindle organization and biogenesis
3939	<i>LDHA</i>	anaerobic glycolysis
4085	<i>MAD2L1</i>	mitotic checkpoint
4141	<i>MARS</i>	protein biosynthesis
4176	<i>MCM7</i>	cell cycle
4199	<i>ME1</i>	malate metabolism
4232	<i>MEST</i>	mesoderm development
4501	<i>MT1X</i>	response to metal ion
4522	<i>MTHFD1</i>	one-carbon compound metabolism
4602	<i>MYB</i>	cell cycle
4605	<i>MYBL2</i>	regulation of progression through cell cycle
4830	<i>C6orf105</i>	negative regulation of cell proliferation
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
4893	<i>NRAS</i>	regulation of progression through cell cycle
5315	<i>PKM2</i>	glycolysis
5694	<i>PSMB6</i>	ubiquitin-dependent protein catabolism
5717	<i>PSMD11</i>	proteasome complex (sensu Eukaryota)
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5901	<i>Ara.24</i>	regulation of progression through cell cycle
5930	<i>RBBP6</i>	biological process unknown

5985	<i>RFC5</i>	DNA replication and repair
6520	<i>SLC3A2</i>	amino acid transport
6652	<i>SORD</i>	sorbitol metabolism
6790	<i>STK6</i>	spindle organization and biogenesis
7031	<i>TFF1</i>	growth factor activity
7083	<i>TK1</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7184	<i>HSP90B1</i>	anti-apoptosis
7247	<i>TSN</i>	DNA recombination
7277	<i>TUBA1</i>	microtubule-based movement
8140	<i>SLC7A5</i>	amino acid transport
9141	<i>PDCD5</i>	induction of apoptosis
9256	<i>BZRAP1</i>	biological process unknown
9768	<i>KIAA0101</i>	biological process unknown
10112	<i>KIF20A</i>	microtubule-based movement
10131	<i>TRAP1</i>	unfolded protein binding
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle
10419	<i>PRMT5</i>	cell proliferation
10606	<i>PAICS</i>	purine base biosynthesis
10797	<i>MTHFD2</i>	one-carbon compound metabolism
11222	<i>MRPL3</i>	protein biosynthesis
22948	<i>CCT5</i>	unfolded protein binding
23397	<i>BRRN1</i>	mitotic chromosome condensation
23450	<i>SF3B3</i>	nuclear mRNA splicing, via spliceosome
27101	<i>CACYBP</i>	ubiquitin cycle
85004	<i>RERG</i>	estrogen receptor binding
94239	<i>H2AFV</i>	chromosome organization and biogenesis (sensu Eukaryota)

Table A.17. Bicluster 15; list of genes in bicluster 15. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
12	<i>CNR1</i>	regulation of lipid metabolism
332	<i>HSPD1</i>	G2/M transition of mitotic cell cycle
372	<i>ARCN1</i>	retrograde vesicle-mediated transport, Golgi to ER
595	<i>PSCDBP</i>	G1/S transition of mitotic cell cycle
708	<i>TKT</i>	immune response
991	<i>SELENBP1</i>	regulation of progression through cell cycle
1058	<i>CENPA</i>	chromosome organization and biogenesis (sensu Eukaryota)
1478	<i>CSTF2</i>	mRNA polyadenylation
1509	<i>CTSD</i>	proteolysis

1736	<i>DKC1</i>	regulation of progression through cell cycle
1841	<i>DTYMK</i>	cell cycle
1854	<i>DUT</i>	DNA replication
2224	<i>FDPS</i>	cholesterol biosynthesis
2237	<i>FEN1</i>	DNA replication and repair
2288	<i>FKBP4</i>	FK506 binding
2289	<i>FKBP5</i>	FK506 binding
2305	<i>FOXM1</i>	transcription factor activity
2674	<i>GFRA1</i>	cell surface receptor linked signal transduction
2971	<i>GTF3A</i>	rRNA transcription
3309	<i>HSPA5</i>	anti-apoptosis
3835	<i>KIF22</i>	microtubule-based movement
3838	<i>KPNA2</i>	M phase specific microtubule process
3925	<i>STMN1</i>	mitotic spindle organization and biogenesis
3939	<i>LDHA</i>	anaerobic glycolysis
4085	<i>MAD2L1</i>	mitotic checkpoint
4141	<i>MARS</i>	protein biosynthesis
4171	<i>MCM2</i>	cell cycle
4176	<i>MCM7</i>	cell cycle
4199	<i>ME1</i>	malate metabolism
4288	<i>MKI67</i>	regulation of progression through cell cycle
4501	<i>MTIX</i>	response to metal ion
4522	<i>MTHFD1</i>	one-carbon compound metabolism
4602	<i>MYB</i>	cell cycle
4605	<i>MYBL2</i>	regulation of progression through cell cycle
4691	<i>NCL</i>	nucleotide binding
4830	<i>C6orf105</i>	negative regulation of cell proliferation
4860	<i>NP</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
5174	<i>PDZK1</i>	Biological process unknown
5315	<i>PKM2</i>	glycolysis
5352	<i>PLOD2</i>	protein metabolism
5479	<i>PPIB</i>	unfolded protein binding
5694	<i>PSMB6</i>	ubiquitin-dependent protein catabolism
5717	<i>PSMD11</i>	proteasome complex (sensu Eukaryota)
5888	<i>RAD51</i>	double-strand break repair via homologous recombination
5889	<i>RAD51C</i>	DNA recombination and repair
5901	<i>Ara.24</i>	regulation of progression through cell cycle
5906	<i>RAP1A</i>	negative regulation of progression through cell cycle
5930	<i>RBBP6</i>	Biological process unknown
5985	<i>RFC5</i>	DNA replication and repair
6059	<i>ABCE1</i>	unfolded protein binding

6119	<i>RPA3</i>	DNA replication and repair
6426	<i>SFRS1</i>	nuclear mRNA splicing, via spliceosome
6434	<i>SFRS10</i>	nuclear mRNA splicing, via spliceosome
6520	<i>SLC3A2</i>	amino acid transport
6628	<i>SNRPB</i>	nuclear mRNA splicing, via spliceosome
6633	<i>SNRPD2</i>	nuclear mRNA splicing, via spliceosome
6636	<i>SNRPF</i>	nuclear mRNA splicing, via spliceosome
6652	<i>SORD</i>	sorbitol metabolism
6713	<i>SQLE</i>	electron transport
6790	<i>STK6</i>	spindle organization and biogenesis
6880	<i>TAF9</i>	amino acid biosynthesis
7031	<i>TFF1</i>	growth factor activity
7083	<i>TK1</i>	DNA replication
7162	<i>TPBG</i>	cell adhesion
7164	<i>TPD52L1</i>	G2/M transition of mitotic cell cycle
7184	<i>HSP90B1</i>	anti-apoptosis
7203	<i>CCT3</i>	unfolded protein binding
7277	<i>TUBA1</i>	microtubule-based movement
7425	<i>VGF</i>	growth factor activity
8140	<i>SLC7A5</i>	amino acid transport
8533	<i>COP9</i>	signalosome complex
9086	<i>EIF1AY</i>	translational initiation
9141	<i>PDCD5</i>	induction of apoptosis
9156	<i>EXO1</i>	DNA recombination and repair
9232	<i>PTTG1</i>	chromosome organization and biogenesis
9446	<i>GSTO1</i>	biological process unknown
9601	<i>PDIA4</i>	protein secretion
9619	<i>ABCG1</i>	cholesterol metabolism
9741	<i>LAPTM4A</i>	transport
9787	<i>DLG7</i>	positive regulation of mitotic metaphase/anaphase transition
10112	<i>KIF20A</i>	microtubule-based movement
10202	<i>DHRS2</i>	negative regulation of progression through cell cycle
10236	<i>HNRPR</i>	mRNA processing
10419	<i>PRMT5</i>	cell proliferation
10436	<i>EMG1</i>	translational elongation
10439	<i>OLFM1</i>	development
10574	<i>CCT7</i>	regulation of progression through cell cycle
10575	<i>CCT4</i>	unfolded protein binding
10606	<i>PAICS</i>	purine nucleotide biosynthesis
10797	<i>MTHFD2</i>	one-carbon compound metabolism
10963	<i>STIP1</i>	response to stress

11004	<i>KIF2C</i>	microtubule-based movement
11222	<i>MRPL3</i>	protein biosynthesis
11260	<i>XPOT</i>	transport
22948	<i>CCT5</i>	unfolded protein binding
23160	<i>WDR43</i>	nucleus
23306	<i>KIAA0286</i>	biological process unknown
23397	<i>BRRN1</i>	mitotic cell cycle
23450	<i>SF3B3</i>	spliceosome complex
27101	<i>CACYBP</i>	ubiquitin cycle
27338	<i>UBE2S</i>	ubiquitin cycle
51765	<i>MST4</i>	regulation of apoptosis
85004	<i>RERG</i>	estrogen receptor binding
94239	<i>H2AFV</i>	chromosome organization and biogenesis (sensu Eukaryota)
203068	<i>TUBB</i>	microtubule-based movement

Table a.18. Bicluster 16; list of genes in bicluster 16. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
273	<i>AMPH</i>	endocytosis
366	<i>AQP9</i>	nutrient import
732	<i>C8B</i>	cytolysis
797	<i>CALCB</i>	signal transduction
954	<i>ENTPD2</i>	hydrolase activity
1179	<i>CLCA1</i>	chloride transport
1268	<i>CNR1</i>	receptor activity
1618	<i>DAZL</i>	regulation of translation
2568	<i>GABRP</i>	ion transport
3612	<i>IMPA1</i>	carbohydrate metabolism
3620	<i>INDO</i>	immune response
3675	<i>ITGA3</i>	cell adhesion
4327	<i>MMP19</i>	cell differentiation
5244	<i>ABCB4</i>	lipid metabolism
5567	<i>PRKACB</i>	signal transduction
5588	<i>PRKCQ</i>	regulation of cell growth
5732	<i>PTGER2</i>	signal transduction
5999	<i>RGS4</i>	inactivation of MAPK activity
6297	<i>SALL2</i>	transcription factor activity
7267	<i>TTC3</i>	biological process unknown
7298	<i>TYMS</i>	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
7786	<i>MAP3K12</i>	JNK cascade

8505	<i>PARG</i>	hydrolase activity
9056	<i>SLC7A7</i>	amino acid transport
9541	<i>CIR</i>	transcription corepressor activity
9705	<i>ST18</i>	transcription
10643	<i>IGF2BP3</i>	protein biosynthesis
11081	<i>KERA</i>	eye development (sensu Mammalia)
11342	<i>RNF13</i>	proteolysis
23273	<i>KIAA0367</i>	induction of apoptosis
23284	<i>LPHN3</i>	signal transduction
51043	<i>ZBTB7B</i>	transcription
253959	<i>GARNL1</i>	Biological process unknown

Table A.19. Bicluster 17; list of genes in bicluster 17. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
43	<i>KIAA0143</i>	cell adhesion
156	<i>ADRBK1</i>	signal transduction
244	<i>ABCB4</i>	
247	<i>TSN</i>	induction of apoptosis
578	<i>SCARF1</i>	induction of apoptosis
594	<i>BCKDHB</i>	branched chain family amino acid catabolism
706	<i>BZRP</i>	biological process unknown
726	<i>CAPN5</i>	proteolysis
753	<i>C18orf1</i>	biological process unknown
901	<i>Ara.24</i>	regulation of progression through cell cycle
934	<i>RAB7L1</i>	humoral immune response
987	<i>LRBA</i>	biological process unknown
997	<i>CDC34</i>	G1/S transition of mitotic cell cycle
1314	<i>COPA</i>	intracellular protein transport
1556	<i>CYP2B7P1</i>	electron transport
1942	<i>EFNA1</i>	cell-cell signaling
1954	<i>MEGF8</i>	biological process unknown
1993	<i>ELAVL2</i>	regulation of transcription, DNA-dependent
2065	<i>ERBB3</i>	epidermal growth factor receptor activity
2069	<i>EREG</i>	regulation of progression through cell cycle
2530	<i>FUT8</i>	oligosaccharide biosynthesis
2537	<i>IFI6</i>	anti-apoptosis
2566	<i>GABRG2</i>	neurotransmitter receptor activity
2911	<i>GRM1</i>	signal transduction
2730	<i>GCLM</i>	cysteine metabolism

2966	<i>GTF2H2</i>	transcription factor activity
3347	<i>HTN3</i>	xenobiotic metabolism
3423	<i>IDS</i>	glycosaminoglycan metabolism
3434	<i>IFIT1</i>	biological process unknown
3586	<i>IL10</i>	cytokine activity
3672	<i>ITGA1</i>	cell adhesion
3762	<i>KCNJ5</i>	ion transport
3762	<i>KCNJ5</i>	potassium ion transport
4034	<i>LRCH4</i>	nervous system development
4065	<i>LY75</i>	endocytosis
4163	<i>MCC</i>	cell cycle
4214	<i>MAP3K1</i>	MAP kinase kinase kinase activity
4286	<i>MITF</i>	regulation of transcription, DNA-dependent
4430	<i>MYO1B</i>	motor activity, myosin
4520	<i>MTF1</i>	transcription factor activity
4582	<i>MUC1</i>	hormone activity
4644	<i>MYO5A</i>	motor activity, myosin
4791	<i>NFKB2</i>	regulation of transcription, DNA-dependent
4842	<i>NOS1</i>	cell-cell signaling
4886	<i>NPY1R</i>	signal transduction
4983	<i>OPHN1</i>	axon guidance
5046	<i>PCSK6</i>	cell-cell signaling
5108	<i>PCM1</i>	transcription
5175	<i>PECAM1</i>	cell adhesion
5239	<i>PGM5</i>	cell adhesion
5307	<i>PITX1</i>	regulation of transcription, DNA-dependent
5431	<i>POLR2B</i>	transcription
5431	<i>POLR2B</i>	transcription
5449	<i>POU1F1</i>	negative regulation of cell proliferation
5481	<i>PPID</i>	FK506 binding
5557	<i>PRIM1</i>	DNA replication
5649	<i>RELN</i>	cell adhesion
6439	<i>SFTPB</i>	lipid metabolism
6558	<i>SLC12A2</i>	transporter activity
6597	<i>SMARCA4</i>	transcription factor activity
6760	<i>SS18</i>	biological process unknown
6772	<i>STAT1</i>	regulation of progression through cell cycle
6774	<i>STAT3</i>	regulation of transcription, DNA-dependent
6872	<i>TAF1</i>	G1 phase of mitotic cell cycle
6892	<i>TAPBP</i>	antigen processing, endogenous antigen via MHC class I
6899	<i>TBX1</i>	transcription factor activity

6932	<i>TCF7</i>	regulation of transcription, DNA-dependent
6992	<i>PPP1R11</i>	biological process unknown
7051	<i>TGM1</i>	cell envelope organization and biogenesis
7099	<i>TLR4</i>	signal transduction
7267	<i>TTC3</i>	biological process unknown
7294	<i>TXK</i>	intracellular signaling cascade
7373	<i>COL14A1</i>	cell adhesion
7581	<i>ZNF33A</i>	regulation of transcription, DNA-dependent
7764	<i>ZNF217</i>	regulation of transcription, DNA-dependent
7855	<i>FZD5</i>	development
8091	<i>HMGA2</i>	regulation of transcription, DNA-dependent
8482	<i>SEMA7A</i>	cell differentiation
8492	<i>PRSS12</i>	proteolysis
8541	<i>PPFIA3</i>	
8549	<i>LGR5</i>	signal transduction
8556	<i>CDC14A</i>	regulation of progression through cell cycle
8578	<i>SCARF1</i>	cell adhesion
8740	<i>TNFSF14</i>	signal transduction
9037	<i>SEMA5A</i>	cell adhesion
9168	<i>TMSB10</i>	cytoskeleton organization and biogenesis
9219	<i>MTA2</i>	transcription factor activity
9371	<i>KIF3B</i>	microtubule-based movement
9725	<i>TMEM63A</i>	biological process unknown
9891	<i>NUAK1</i>	biological process unknown
10147	<i>SFRS14</i>	nuclear mRNA splicing, via spliceosome
10149	<i>GPR64</i>	signal transduction
10253	<i>SPRY2</i>	cell-cell signaling
10370	<i>CITED2</i>	transcription factor activity
10513	<i>APPBP2</i>	microtubule motor activity
10555	<i>AGPAT2</i>	metabolism
10609	<i>SC65</i>	synaptonemal complex formation
10611	<i>PDLIM5</i>	biological process unknown
10866	<i>HCP5</i>	defense response
23125	<i>CAMTA2</i>	regulation of transcription, DNA-dependent
23158	<i>TBC1D9</i>	biological process unknown
23503	<i>ZFYVE26</i>	biological process unknown
23759	<i>PPIL2</i>	biological process unknown
25778	<i>RIPK5</i>	protein serine/threonine kinase activity
26503	<i>SLC17A5</i>	transport
26959	<i>HBPI</i>	regulation of transcription, DNA-dependent
29799	<i>YPEL1</i>	biological process unknown

30836	<i>DNTTIP2</i>	biological process unknown
51043	<i>ZBTB7B</i>	regulation of transcription, DNA-dependent
51474	<i>LIMA1</i>	focal adhesion
54739	<i>BIRC4BP</i>	biological process unknown
54739	<i>BIRC4BP</i>	biological process unknown
55764	<i>IFT122</i>	biological process unknown
65125	<i>WNK1</i>	ion transport
65268	<i>WNK2</i>	protein kinase cascade
84830	<i>C6orf105</i>	biological process unknown

Table A.20. Biclusters 18; list of genes in bicluster 18. The first column gives the NCBI Gene ID of the genes, the second column gives the gene symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
156	<i>ADRBK1</i>	signal transduction
244	<i>ABCB4</i>	blood coagulation
247	<i>TSN</i>	induction of apoptosis
578	<i>SCARF1</i>	induction of apoptosis
594	<i>BCKDHB</i>	branched chain family amino acid catabolism
706	<i>BZRP</i>	signal transduction
726	<i>CAPN5</i>	proteolysis
901	<i>Ara.24</i>	regulation of progression through cell cycle
934	<i>RAB7L1</i>	humoral immune response
987	<i>LRBA</i>	biological process unknown
997	<i>CDC34</i>	G1/S transition of mitotic cell cycle
1314	<i>COPA</i>	retrograde vesicle-mediated transport, Golgi to ER
1942	<i>EFNA1</i>	cell-cell signaling
1954	<i>MEGF8</i>	development
2069	<i>EREG</i>	regulation of progression through cell cycle
2149	<i>F2R</i>	regulation of progression through cell cycle
2566	<i>GABRG2</i>	neurotransmitter receptor activity
2730	<i>GCLM</i>	cysteine metabolism
2911	<i>GRM1</i>	signal transduction
2966	<i>GTF2H2</i>	transcription factor activity
3347	<i>HTN3</i>	xenobiotic metabolism
3434	<i>IFIT1</i>	immune response
3423	<i>IDS</i>	metabolism
3586	<i>IL10</i>	cytokine activity
3672	<i>ITGA1</i>	cell adhesion
3769	<i>KCNJ13</i>	ion transport
4034	<i>LRCH4</i>	nervous system development

4065	<i>LY75</i>	inflammatory response
4214	<i>MAP3K1</i>	MAP kinase kinase kinase activity
4286	<i>MITF</i>	regulation of transcription, DNA-dependent
4322	<i>MMP13</i>	proteolysis
4430	<i>MYO1B</i>	motor activity, myosin
4520	<i>MTF1</i>	transcription factor activity
4644	<i>MYO5A</i>	motor activity, myosin
4886	<i>NPY1R</i>	signal transduction
4983	<i>OPHN1</i>	axon guidance
5108	<i>PCM1</i>	transcription
5307	<i>PITX1</i>	regulation of transcription, DNA-dependent
5431	<i>POLR2B</i>	transcription from RNA polymerase II promoter
5449	<i>POU1F1</i>	transcription from RNA polymerase II promoter
5481	<i>PPID</i>	FK506 binding
5557	<i>PRIM1</i>	DNA replication, synthesis of RNA primer
5639	<i>PRRG2</i>	hormone activity
6018	<i>RLF</i>	regulation of transcription, DNA-dependent
6439	<i>SFTPB</i>	lipid metabolism
6558	<i>SLC12A2</i>	ion transport
6597	<i>SMARCA4</i>	regulation of transcription from RNA polymerase II promoter
6774	<i>STAT3</i>	negative regulation of transcription from RNA polymerase II promoter
6872	<i>TAF1</i>	G1 phase of mitotic cell cycle
6892	<i>TAPBP</i>	immune response
6899	<i>TBX1</i>	regulation of transcription from RNA polymerase II promoter
6997	<i>TDGF1</i>	negative regulation of apoptosis
7051	<i>TGM1</i>	cell envelope organization and biogenesis
7099	<i>TLR4</i>	signal transduction
7267	<i>TTC3</i>	biological process unknown
7294	<i>TXK</i>	intracellular signaling cascade
7373	<i>COL14A1</i>	cell adhesion
7764	<i>ZNF217</i>	transcription factor activity
7855	<i>FZD5</i>	Wnt receptor signaling pathway
8091	<i>HMGA2</i>	regulation of transcription, DNA-dependent
8482	<i>SEMA7A</i>	GPI anchor binding
8492	<i>PRSS12</i>	proteolysis
8541	<i>PPFIA3</i>	biological process unknown
8556	<i>CDC14A</i>	regulation of progression through cell cycle
8578	<i>SCARF1</i>	cell adhesion
8740	<i>TNFSF14</i>	induction of apoptosis
9037	<i>SEMA5A</i>	cell adhesion
9168	<i>TMSB10</i>	cytoskeleton organization and biogenesis

9219	<i>MTA2</i>	transcription factor activity
9595	<i>PSCDBP</i>	regulation of cell adhesion
9725	<i>TMEM63A</i>	biological process unknown
9731	<i>KIAA0562</i>	biological process unknown
9849	<i>ZNF518</i>	nucleic acid binding
10147	<i>SFRS14</i>	nuclear mRNA splicing, via spliceosome
10253	<i>SPRY2</i>	cell-cell signaling
10513	<i>APBP2</i>	intracellular protein transport
10555	<i>AGPAT2</i>	phosphatidic acid biosynthesis
10609	<i>SC65</i>	synaptonemal complex formation
10866	<i>HCP5</i>	defense response
22915	<i>MMRN1</i>	cell adhesion
23125	<i>CAMTA2</i>	regulation of transcription, DNA-dependent
23158	<i>TBC1D9</i>	biological process unknown
23503	<i>ZFYVE26</i>	nucleic acid binding
23759	<i>PPIL2</i>	biological process unknown
26959	<i>HBPI</i>	regulation of transcription, DNA-dependent
51043	<i>ZBTB7B</i>	regulation of transcription, DNA-dependent
54739	<i>BIRC4BP</i>	biological process unknown
65268	<i>WNK2</i>	protein kinase cascade
54739	<i>BIRC4BP</i>	zinc ion binding
65268	<i>WNK2</i>	nucleotide binding
84830	<i>C6orf105</i>	protein serine/threonine kinase activity

Table A.21. Bicluster 19; list of genes in bicluster 19. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
43	<i>KIAA0143</i>	cell adhesion
302	<i>ANXA2</i>	cytoskeletal protein binding
347	<i>IGBP1</i>	lipid metabolism
489	<i>NRAS</i>	proton transport
578	<i>SCARF1</i>	induction of apoptosis
598	<i>BCL2L1</i>	anti-apoptosis
598	<i>BCL2L1</i>	anti-apoptosis
901	<i>Ara.24</i>	regulation of progression through cell cycle
934	<i>RAB7L1</i>	humoral immune response
1153	<i>CIRBP</i>	response to cold
1191	<i>CLU</i>	apoptosis
1368	<i>CPM</i>	proteolysis

1374	<i>CPT1A</i>	lipid metabolism
1376	<i>CPT2</i>	fatty acid metabolism
1942	<i>EFNA1</i>	cell-cell signaling
2023	<i>ENO1</i>	regulation of transcription, DNA-dependent
2027	<i>ENO3</i>	glycolysis
2065	<i>ERBB3</i>	epidermal growth factor receptor activity
2099	<i>ESR1</i>	steroid hormone receptor activity
2203	<i>FBP1</i>	fructose metabolism
2537	<i>IFI6</i>	anti-apoptosis
2730	<i>GCLM</i>	cysteine metabolism
2804	<i>GOLGB1</i>	Golgi organization and biogenesis
2966	<i>GTF2H2</i>	transcription factor activity
3488	<i>IGFBP5</i>	insulin-like growth factor binding
3586	<i>IL10</i>	cytokine activity
3725	<i>JUN</i>	transcription factor activity
4070	<i>TACSTD2</i>	cell proliferation
4151	<i>MB</i>	transport
4582	<i>MUC1</i>	cytoskeleton
4680	<i>CEACAM6</i>	cell-cell signaling
4950	<i>OCLN</i>	tight junction
4983	<i>OPHN1</i>	axon guidance
5046	<i>PCSK6</i>	cell-cell signaling
5239	<i>PGM5</i>	cell adhesion
5407	<i>PNLIPRP1</i>	lipid catabolism
5557	<i>PRIM1</i>	DNA replication, synthesis of RNA primer
5896	<i>RAG1</i>	DNA recombination
6760	<i>SS18</i>	biological process unknown
6772	<i>STAT1</i>	regulation of progression through cell cycle
6892	<i>TAPBP</i>	immune response
7089	<i>TLE2</i>	regulation of transcription, DNA-dependent
7644	<i>ZNF91</i>	regulation of transcription, DNA-dependent
7764	<i>ZNF217</i>	regulation of transcription, DNA-dependent
8416	<i>ANXA9</i>	cell-cell adhesion
8991	<i>SELENBP1</i>	biological process unknown
9780	<i>FAM38A</i>	biological process unknown
9780	<i>FAM38A</i>	integral to membrane
9891	<i>NUAK1</i>	nucleotide binding
9915	<i>ARNT2</i>	regulation of transcription, DNA-dependent, response to xenobiotic stimulus
10103	<i>TSPAN1</i>	cell adhesion
10370	<i>CITED2</i>	transcription factor activity

10611	<i>PDLIM5</i>	receptor signaling complex scaffold activity
23138	<i>N4BP32</i>	biological process unknown
23158	<i>TBC1D9</i>	biological process unknown
25805	<i>BAMBI</i>	biological process unknown
26036	<i>ZNF451</i>	regulation of transcription, DNA-dependent
26959	<i>HBP1</i>	regulation of transcription, DNA-dependent
29799	<i>YPEL1</i>	biological process unknown
30836	<i>DNTTIP2</i>	biological process unknown
51474	<i>LIMA1</i>	focal adhesion
55764	<i>IFT122</i>	biological process unknown
57007	<i>CMKOR1</i>	signal transduction

Table A.22. Bicluster 20; list of genes in bicluster 20. The first column gives the NCBI Gene ID of the genes, the second column gives the genes symbol and the third column gives the GO annotation for the corresponding gene.

Gene ID	Gene Symbol	GO Functional Annotation
43	<i>KIAA0143</i>	cell adhesion
578	<i>SCARF1</i>	induction of apoptosis
654	<i>BMP6</i>	cytokine activity
901	<i>Ara.24</i>	regulation of progression through cell cycle
987	<i>LRBA</i>	biological process unknown
1811	<i>SLC26A3</i>	transcription factor activity
1942	<i>EFNA1</i>	cell-cell signaling
4295	<i>MLN</i>	cell-cell signaling
4925	<i>NUCB2</i>	DNA binding
4983	<i>OPHN1</i>	axon guidance
5239	<i>PGM5</i>	cell adhesion
5289	<i>PIK3C3</i>	inositol or phosphatidylinositol kinase activity
6300	<i>MAPK12</i>	cell cycle arrest
6892	<i>TAPBP</i>	immune response
7024	<i>TFCP2</i>	regulation of transcription from RNA polymerase II promoter
7267	<i>TTC3</i>	biological process unknown
7298	<i>TYMS</i>	DNA replication and repair
7764	<i>ZNF217</i>	regulation of transcription, DNA-dependent
8740	<i>TNFSF14</i>	induction of apoptosis
9760	<i>TOX</i>	regulation of transcription, DNA-dependent
10513	<i>APPBP2</i>	intracellular protein transport
10768	<i>AHCYL1</i>	one-carbon compound metabolism
10949	<i>HNRPA0</i>	mRNA processing
23167	<i>KIAA0143</i>	biological process unknown
23503	<i>ZFYVE26</i>	nucleic acid binding

51474	<i>LIMA1</i>	focal adhesion
54739	<i>BIRC4BP</i>	zinc ion binding

BIBLIOGRAPHY

1. Jemal, A., et al., *Cancer statistics, 2002*. CA Cancer J Clin, 2002. 52(1): p. 23-47.
2. Kubista, E., [*Breast cancer: figures and facts*]. Wien Med Wochenschr, 2001. 151(21-23): p. 548-51.
3. McGuire, W.L., et al., *How to use prognostic factors in axillary node-negative breast cancer patients*. J Natl Cancer Inst, 1990. 82(12): p. 1006-15.
4. Rittling, S.R. and A.F. Chambers, *Role of osteopontin in tumour progression*. Br J Cancer, 2004. 90(10): p. 1877-81.
5. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1*. Science, 1994. 266(5182): p. 66-71.
6. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2*. Nature, 1995. 378(6559): p. 789-92.
7. Abba, M.C., et al., *Gene expression signature of estrogen receptor alpha status in breast cancer*. BMC Genomics, 2005. 6(1): p. 37.
8. Buettner, R., L.B. Mora, and R. Jove, *Activated STAT signaling in human tumors provides novel molecular targets for therapeutic intervention*. Clin Cancer Res, 2002. 8(4): p. 945-54.
9. Kushner, P.J., et al., *Estrogen receptor pathways to AP-1*. J Steroid Biochem Mol Biol, 2000. 74(5): p. 311-7.
10. McKay, L.I. and J.A. Cidlowski, *Molecular control of immune/inflammatory responses: interactions between nuclear factor-kappa B and steroid receptor-signaling pathways*. Endocr Rev, 1999. 20(4): p. 435-59.
11. Harnish, D.C., et al., *The role of CBP in estrogen receptor cross-talk with nuclear factor-kappaB in HepG2 cells*. Endocrinology, 2000. 141(9): p. 3403-11.
12. Hedenfalk, I., et al., *Gene-expression profiles in hereditary breast cancer*. N Engl J Med, 2001. 344(8): p. 539-48.
13. Cicatiello, L., et al., *A genomic view of estrogen actions in human breast cancer cells by expression profiling of the hormone-responsive transcriptome*. J Mol Endocrinol, 2004. 32(3): p. 719-75.
14. Coser, K.R., et al., *Global analysis of ligand sensitivity of estrogen inducible and suppressible genes in MCF7/BUS breast cancer cells by DNA microarray*. Proc Natl Acad Sci U S A, 2003. 100(24): p. 13994-9.
15. Creighton, C.J., et al., *Genes regulated by estrogen in breast tumor cells in vitro are similarly regulated in vivo in tumor xenografts and human breast tumors*. Genome Biol, 2006. 7(4): p. R28.

16. Cunliffe, H.E., et al., *The gene expression response of breast cancer to growth regulators: patterns and correlation with tumor expression profiles*. Cancer Res, 2003. 63(21): p. 7158-66.
17. Frasor, J., et al., *Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype*. Endocrinology, 2003. 144(10): p. 4562-74.
18. Inoue, A., et al., *Development of cDNA microarray for expression profiling of estrogen-responsive genes*. J Mol Endocrinol, 2002. 29(2): p. 175-92.
19. Lobenhofer, E.K., et al., *Regulation of DNA replication fork genes by 17beta-estradiol*. Mol Endocrinol, 2002. 16(6): p. 1215-29.
20. Soulez, M. and M.G. Parker, *Identification of novel oestrogen receptor target genes in human ZR75-1 breast cancer cells by expression profiling*. J Mol Endocrinol, 2001. 27(3): p. 259-74.
21. Wang, D.Y., et al., *Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1*. Mol Endocrinol, 2004. 18(2): p. 402-11.
22. Tamayo, P., et al., *Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation*. Proc Natl Acad Sci U S A, 1999. 96(6): p. 2907-12.
23. Tanay, A., R. Sharan, and R. Shamir, *Discovering statistically significant biclusters in gene expression data*. Bioinformatics, 2002. 18 Suppl 1: p. S136-44.
24. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. 100(1): p. 57-70.
25. Govind, A.P. and R.V. Thampan, *Membrane associated estrogen receptors and related proteins: localization at the plasma membrane and the endoplasmic reticulum*. Mol Cell Biochem, 2003. 253(1-2): p. 233-40.
26. Levin, E.R., *Cell localization, physiology, and nongenomic actions of estrogen receptors*. J Appl Physiol, 2001. 91(4): p. 1860-7.
27. Levin, E.R., *Cellular functions of plasma membrane estrogen receptors*. Steroids, 2002. 67(6): p. 471-5.
28. Pearce, S.T. and V.C. Jordan, *The biological role of estrogen receptors alpha and beta in cancer*. Crit Rev Oncol Hematol, 2004. 50(1): p. 3-22.
29. Weihua, Z., et al., *Update on estrogen signaling*. FEBS Lett, 2003. 546(1): p. 17-24.
30. Chen, J.Q., et al., *Mitochondrial localization of ERalpha and ERbeta in human MCF7 cells*. Am J Physiol Endocrinol Metab, 2004. 286(6): p. E1011-22.
31. Demonacos, C.V., et al., *Mitochondrial genes as sites of primary action of steroid hormones*. Steroids, 1996. 61(4): p. 226-32.

32. Sekeris, C.E., *The mitochondrial genome: a possible primary site of action of steroid hormones*. In Vivo, 1990. 4(5): p. 317-20.
33. Levin, E.R., *Bidirectional signaling between the estrogen receptor and the epidermal growth factor receptor*. Mol Endocrinol, 2003. 17(3): p. 309-17.
34. Southern, E., K. Mir, and M. Shchepinov, *Molecular interactions on microarrays*. Nat Genet, 1999. 21(1 Suppl): p. 5-9.
35. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays*. Nat Biotechnol, 1996. 14(13): p. 1675-80.
36. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, 1995. 270(5235): p. 467-70.
37. Schena, M., et al., *Parallel human genome analysis: microarray-based expression monitoring of 1000 genes*. Proc Natl Acad Sci U S A, 1996. 93(20): p. 10614-9.
38. Lipshutz, R.J., et al., *High density synthetic oligonucleotide arrays*. Nat Genet, 1999. 21(1 Suppl): p. 20-4.
39. McGall, G.H., et al., *The Efficiency of Light-Directed Synthesis of DNA Arrays on Glass Substrates*. J. Am. Chem. Soc., 1997. 119(22): p. 5081-5090.
40. Wodicka, L., et al., *Genome-wide expression monitoring in Saccharomyces cerevisiae*. Nat Biotechnol, 1997. 15(13): p. 1359-67.
41. Fodor, S.P., et al., *Light-directed, spatially addressable parallel chemical synthesis*. Science, 1991. 251(4995): p. 767-73.
42. McGall, G., et al., *Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists*. Proc Natl Acad Sci U S A, 1996. 93(24): p. 13555-60.
43. Shamir, R., *Analysis of Gene Expression Data Lecture Notes*. 2002.
44. Gingeras, T.R., D.Y. Kwoh, and G.R. Davis, *Hybridization properties of immobilized nucleic acids*. Nucleic Acids Res, 1987. 15(13): p. 5373-90.
45. Lock, C., et al., *Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis*. Nat Med, 2002. 8(5): p. 500-8.
46. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. Nature, 2000. 403(6769): p. 503-11.
47. Bittner, M., et al., *Molecular classification of cutaneous malignant melanoma by gene expression profiling*. Nature, 2000. 406(6795): p. 536-40.
48. Dhanasekaran, S.M., et al., *Delineation of prognostic biomarkers in prostate cancer*. Nature, 2001. 412(6849): p. 822-6.
49. Golub, T.R., et al., *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*. Science, 1999. 286(5439): p. 531-7.
50. Shipp, M.A., et al., *Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning*. Nat Med, 2002. 8(1): p. 68-74.

51. Sorlie, T., et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications*. Proc Natl Acad Sci U S A, 2001. 98(19): p. 10869-74.
52. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer*. Nature, 2002. 415(6871): p. 530-6.
53. Volm, M., et al., *Expression profile of genes in non-small cell lung carcinomas from long-term surviving patients*. Clin Cancer Res, 2002. 8(6): p. 1843-8.
54. Miki, R., et al., *Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays*. Proc Natl Acad Sci U S A, 2001. 98(5): p. 2199-204.
55. Pollack, J.R., et al., *Genome-wide analysis of DNA copy-number changes using cDNA microarrays*. Nat Genet, 1999. 23(1): p. 41-6.
56. Lo, A.W., et al., *A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA*. Genome Res, 2001. 11(3): p. 448-57.
57. Shannon, M.F. and S. Rao, *Transcription. Of chips and ChIPs*. Science, 2002. 296(5568): p. 666-9.
58. Ahrendt, S.A., et al., *Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array*. Proc Natl Acad Sci U S A, 1999. 96(13): p. 7382-7.
59. Lindblad-Toh, K., et al., *Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays*. Nat Biotechnol, 2000. 18(9): p. 1001-5.
60. Lindblad-Toh, K., et al., *Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse*. Nat Genet, 2000. 24(4): p. 381-6.
61. Ziauddin, J. and D.M. Sabatini, *Microarrays of cells expressing defined cDNAs*. Nature, 2001. 411(6833): p. 107-10.
62. Brazma, A. and J. Vilo, *Gene expression data analysis*. FEBS Lett, 2000. 480(1): p. 17-24.
63. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns*. Proc Natl Acad Sci U S A, 1998. 95(25): p. 14863-8.
64. Hartigan, J.A., *Clustering Algorithms*. 1975: John Wiley and Sons.
65. Brazma, A., et al., *Predicting gene regulatory elements in silico on a genomic scale*. Genome Res, 1998. 8(11): p. 1202-15.
66. Ben-Dor, A., R. Shamir, and Z. Yakhini, *Clustering gene expression patterns*. J Comput Biol, 1999. 6(3-4): p. 281-97.
67. DeRisi, J.L., V.R. Iyer, and P.O. Brown, *Exploring the metabolic and genetic control of gene expression on a genomic scale*. Science, 1997. 278(5338): p. 680-6.

68. Alon, U., et al., *Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays*. Proc Natl Acad Sci U S A, 1999. 96(12): p. 6745-50.
69. Brown, M.P., et al., *Knowledge-based analysis of microarray gene expression data by using support vector machines*. Proc Natl Acad Sci U S A, 2000. 97(1): p. 262-7.
70. Bertucci, F., et al., *Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters*. Oncogene, 2004. 23(7): p. 1377-91.
71. Bhattacharjee, A., et al., *Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses*. Proc Natl Acad Sci U S A, 2001. 98(24): p. 13790-5.
72. Lapointe, J., et al., *Gene expression profiling identifies clinically relevant subtypes of prostate cancer*. Proc Natl Acad Sci U S A, 2004. 101(3): p. 811-6.
73. Alizadeh, A.A., et al., *Towards a novel classification of human malignancies based on gene expression patterns*. J Pathol, 2001. 195(1): p. 41-52.
74. Thomas, J.G., et al., *An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles*. Genome Res, 2001. 11(7): p. 1227-36.
75. Sawiris, G.P., et al., *Development of a highly specialized cDNA array for the study and diagnosis of epithelial ovarian cancer*. Cancer Res, 2002. 62(10): p. 2923-8.
76. Alizadeh, A., et al., *The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes*. Cold Spring Harb Symp Quant Biol, 1999. 64: p. 71-8.
77. Barrans, J.D., D. Stamatiou, and C. Liew, *Construction of a human cardiovascular cDNA microarray: portrait of the failing heart*. Biochem Biophys Res Commun, 2001. 280(4): p. 964-9.
78. Takata, R., et al., *Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling*. Clin Cancer Res, 2005. 11(7): p. 2625-36.
79. Alaoui-Jamali, M.A., I. Dupre, and H. Qiang, *Prediction of drug sensitivity and drug resistance in cancer by transcriptional and proteomic profiling*. Drug Resist Updat, 2004. 7(4-5): p. 245-55.
80. Goswami, S., et al., *Breast cancer cells isolated by chemotaxis from primary tumors show increased survival and resistance to chemotherapy*. Cancer Res, 2004. 64(21): p. 7664-7.
81. Mariadason, J.M., et al., *Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin*. Cancer Res, 2003. 63(24): p. 8791-812.

82. Park, J.S., et al., *Identification of novel genes associated with the response to 5-FU treatment in gastric cancer cell lines using a cDNA microarray*. Cancer Lett, 2004. 214(1): p. 19-33.
83. Chen, C.D., et al., *Molecular determinants of resistance to antiandrogen therapy*. Nat Med, 2004. 10(1): p. 33-9.
84. Pomeroy, S.L., et al., *Prediction of central nervous system embryonal tumour outcome based on gene expression*. Nature, 2002. 415(6870): p. 436-42.
85. Hedenfalk, I.A., *Gene expression profiling of hereditary and sporadic ovarian cancers reveals unique BRCA1 and BRCA2 signatures*. J Natl Cancer Inst, 2002. 94(13): p. 960-1.
86. Nutt, C.L., et al., *Gene expression-based classification of malignant gliomas correlates better with survival than histological classification*. Cancer Res, 2003. 63(7): p. 1602-7.
87. Davis, R.E. and L.M. Staudt, *Molecular diagnosis of lymphoid malignancies by gene expression profiling*. Curr Opin Hematol, 2002. 9(4): p. 333-8.
88. Gruvberger, S., et al., *Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns*. Cancer Res, 2001. 61(16): p. 5979-84.
89. Perou, C.M., et al., *Molecular portraits of human breast tumours*. Nature, 2000. 406(6797): p. 747-52.
90. Sorlie, T., et al., *Repeated observation of breast tumor subtypes in independent gene expression data sets*. Proc Natl Acad Sci U S A, 2003. 100(14): p. 8418-23.
91. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer*. N Engl J Med, 2002. 347(25): p. 1999-2009.
92. Watanabe, H., et al., *Analysis of temporal changes in the expression of estrogen-regulated genes in the uterus*. J Mol Endocrinol, 2003. 30(3): p. 347-58.
93. Cheng, Y. and G.M. Church, *Biclustering of expression data*. Proc Int Conf Intell Syst Mol Biol, 2000. 8: p. 93-103.
94. Ben-Dor, A., et al., *Discovering local structure in gene expression data: the order-preserving submatrix problem*. J Comput Biol, 2003. 10(3-4): p. 373-84.
95. Ihmels, J., et al., *Revealing modular organization in the yeast transcriptional network*. Nat Genet, 2002. 31(4): p. 370-7.
96. Ihmels, J.H. and S. Bergmann, *Challenges and prospects in the analysis of large-scale gene expression data*. Brief Bioinform, 2004. 5(4): p. 313-27.
97. Murali, T.M. and S. Kasif, *Extracting conserved gene expression motifs from gene expression data*. Pac Symp Biocomput, 2003: p. 77-88.
98. Prelic, A., et al., *A systematic comparison and evaluation of biclustering methods for gene expression data*. Bioinformatics, 2006. 22(9): p. 1122-9.

99. Cho, R.J., et al., *Transcriptional regulation and function during the human cell cycle*. Nat Genet, 2001. 27(1): p. 48-54.
100. Cho, R.J., et al., *A genome-wide transcriptional analysis of the mitotic cell cycle*. Mol Cell, 1998. 2(1): p. 65-73.
101. Spellman, P.T., et al., *Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization*. Mol Biol Cell, 1998. 9(12): p. 3273-97.
102. Whitfield, M.L., et al., *Identification of genes periodically expressed in the human cell cycle and their expression in tumors*. Mol Biol Cell, 2002. 13(6): p. 1977-2000.
103. Zhang, B., et al., *GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies*. BMC Bioinformatics, 2004. 5: p. 16.
104. Brown, A.M., et al., *Activation of pS2 gene transcription is a primary response to estrogen in the human breast cancer cell line MCF-7*. Proc Natl Acad Sci U S A, 1984. 81(20): p. 6344-8.
105. Ghosh, A.K., R. Steele, and R.B. Ray, *Functional domains of c-myc promoter binding protein 1 involved in transcriptional repression and cell growth regulation*. Mol Cell Biol, 1999. 19(4): p. 2880-6.
106. Qin, C., P. Singh, and S. Safe, *Transcriptional activation of insulin-like growth factor-binding protein-4 by 17beta-estradiol in MCF-7 cells: role of estrogen receptor-Sp1 complexes*. Endocrinology, 1999. 140(6): p. 2501-8.
107. Augereau, P., et al., *Characterization of the proximal estrogen-responsive element of human cathepsin D gene*. Mol Endocrinol, 1994. 8(6): p. 693-703.
108. Huynh, H., X.F. Yang, and M. Pollak, *A role for insulin-like growth factor binding protein 5 in the antiproliferative action of the antiestrogen ICI 182780*. Cell Growth Differ, 1996. 7(11): p. 1501-6.
109. Charpentier, A.H., et al., *Effects of estrogen on global gene expression: identification of novel targets of estrogen action*. Cancer Res, 2000. 60(21): p. 5977-83.
110. Bamforth, S.D., et al., *Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator*. Nat Genet, 2001. 29(4): p. 469-74.
111. Eilers, M., T. Endo, and G. Schatz, *Adriamycin, a drug interacting with acidic phospholipids, blocks import of precursor proteins by isolated yeast mitochondria*. J Biol Chem, 1989. 264(5): p. 2945-50.
112. Coppola, J.A. and M.D. Cole, *Constitutive c-myc oncogene expression blocks mouse erythroleukaemia cell differentiation but not commitment*. Nature, 1986. 320(6064): p. 760-3.

113. Freytag, S.O., *Enforced expression of the c-myc oncogene inhibits cell differentiation by precluding entry into a distinct predifferentiation state in G0/G1*. Mol Cell Biol, 1988. 8(4): p. 1614-24.
114. Larsson, L.G., et al., *Phorbol ester-induced terminal differentiation is inhibited in human U-937 monoblastic cells expressing a v-myc oncogene*. Proc Natl Acad Sci U S A, 1988. 85(8): p. 2638-42.
115. Miner, J.H. and B.J. Wold, *c-myc inhibition of MyoD and myogenin-initiated myogenic differentiation*. Mol Cell Biol, 1991. 11(5): p. 2842-51.
116. Gruber, C.J., et al., *Production and actions of estrogens*. N Engl J Med, 2002. 346(5): p. 340-52.
117. Kanehisa, M., *A database for post-genome analysis*. Trends Genet, 1997. 13(9): p. 375-6.
118. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. 28(1): p. 27-30.
119. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG*. Nucleic Acids Res, 2006. 34(Database issue): p. D354-7.
120. biocarta. <http://www.biocarta.com/genes/index.asp>. 2006 [cited.
121. Tsan, M.F., *Toll-like receptors, inflammation and cancer*. Semin Cancer Biol, 2006. 16(1): p. 32-7.
122. Zhou, Y., et al., *The NFkappaB pathway and endocrine-resistant breast cancer*. Endocr Relat Cancer, 2005. 12 Suppl 1: p. S37-46.
123. Bromberg, J., *Stat proteins and oncogenesis*. J Clin Invest, 2002. 109(9): p. 1139-42.
124. Bryan, T.M. and T.R. Cech, *Telomerase and the maintenance of chromosome ends*. Curr Opin Cell Biol, 1999. 11(3): p. 318-24.
125. Bryan, T.M., et al., *Telomere elongation in immortal human cells without detectable telomerase activity*. Embo J, 1995. 14(17): p. 4240-8.
126. Tanoue, T. and E. Nishida, *Docking interactions in the mitogen-activated protein kinase cascades*. Pharmacol Ther, 2002. 93(2-3): p. 193-202.
127. Harper, S.J. and P. LoGrasso, *Signalling for survival and death in neurones: the role of stress-activated kinases, JNK and p38*. Cell Signal, 2001. 13(5): p. 299-310.
128. Vorbrueggen, G., J. Lovric, and K. Moelling, *Functional analysis of phosphorylation at serine 532 of human c-Myb by MAP kinase*. Biol Chem, 1996. 377(11): p. 721-30.
129. Mawson, A., et al., *Estrogen and insulin/IGF-1 cooperatively stimulate cell cycle progression in MCF-7 breast cancer cells through differential regulation of c-Myc and cyclin D1*. Mol Cell Endocrinol, 2005. 229(1-2): p. 161-73.
130. Kharbanda, S., et al., *Determination of cell fate by c-Abl activation in the response to DNA damage*. Oncogene, 1998. 17(25): p. 3309-18.

131. Bruemmer, D., et al., *Expression of minichromosome maintenance proteins in vascular smooth muscle cells is ERK/MAPK dependent*. Exp Cell Res, 2003. 290(1): p. 28-37.
132. Travers, M.T., et al., *Indoleamine 2,3-dioxygenase activity and L-tryptophan transport in human breast cancer cells*. Biochim Biophys Acta, 2004. 1661(1): p. 106-12.
133. Kumar, P., et al., *Estradiol-regulated expression of the immunophilins cyclophilin 40 and FKBP52 in MCF-7 breast cancer cells*. Biochem Biophys Res Commun, 2001. 284(1): p. 219-25.
134. Gougelet, A., et al., *Estrogen receptor alpha and beta subtype expression and transactivation capacity are differentially affected by receptor-, hsp90- and immunophilin-ligands in human breast cancer cells*. J Steroid Biochem Mol Biol, 2005. 94(1-3): p. 71-81.
135. Baldin, V., et al., *Cyclin D1 is a nuclear protein required for cell cycle progression in G1*. Genes Dev, 1993. 7(5): p. 812-21.
136. Li, J. and A.S. Lee, *Stress induction of GRP78/BiP and its role in cancer*. Curr Mol Med, 2006. 6(1): p. 45-54.
137. Attwooll, C., E. Lazzerini Denchi, and K. Helin, *The E2F family: specific functions and overlapping interests*. Embo J, 2004. 23(24): p. 4709-16.
138. Philippar, U., et al., *The SRF target gene Fhl2 antagonizes RhoA/MAL-dependent activation of SRF*. Mol Cell, 2004. 16(6): p. 867-80.
139. NCBI.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=full_report&list_uids=3725. 2006
140. Weinmann, A.S., et al., *Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis*. Genes Dev, 2002. 16(2): p. 235-44.
141. Takahashi, Y., J.B. Rayman, and B.D. Dynlacht, *Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression*. Genes Dev, 2000. 14(7): p. 804-16.
142. Hernando, E., et al., *Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control*. Nature, 2004. 430(7001): p. 797-802.
143. Kim, J., et al., *Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment*. Nat Methods, 2005. 2(1): p. 47-53.