

# Integrative Self-Organizing Map—A Mean Pattern Model

Zihua Yang<sup>1</sup>, Zhengrong Yang<sup>2</sup>

<sup>1</sup>Wolfson Institute of Preventive Medicine, University of Queen Mary, London, UK

<sup>2</sup>School of Biosciences, University of Exeter, Devon, UK

Email: z.h.yang@qmul.ac.uk, z.r.yang@ex.ac.uk

Received June 2013

## ABSTRACT

We propose an integrative self-organizing map (iSOM) for exploring differential expression patterns across multiple microarray experiments. The algorithm is based on the assumption that observed differential expressions are random samples of a mean pattern model which is unknown *a priori*. The learning mechanism of iSOM is similar to the conventional SOM. The mean pattern model which underlies the proposed iSOM models mean differential expressions using a one-dimension of mean differential expressions for the mean differential expressions. The feature map of an iSOM model can be used to reveal correlation between multiple medically/biologically related disease types or multiple platform experiments for one disease. We illustrate applications of iSOM using simulated data and real data.

**Keywords:** Integrative Study of Microarrays; Pattern Discovery; Differential Expressions; Self-Organizing Map

## 1. Introduction

The self-organizing map (SOM) is a popular unsupervised artificial neural network algorithm [1] used for topological pattern recognition. It explores hidden patterns in data and visualizes it in a two-dimensional array. In this array, each grid or neuron preserves or demonstrates a local pattern of the whole pattern hidden in data. The local patterns smoothly changes across the grids of the array. Neighboring neurons therefore show similar local patterns. SOM and its many variants have been widely used in data analysis/mining.

However, SOM and its variants are designed for discovering topological structure hidden in one data set or experiment. Thus they cannot be directly used for an integrative study across multiple data sets for exploring common patterns.

In cancer research, we may often wish to search for common cancer signatures [2-17], based on the understanding that diseases may have some common gene expression pattern in spite of diseases heterogeneity. For instance, it is believed that common gene signature may exist among various cancers [12] as well as among inflammatory diseases [18]. The commonly used method for detecting a common gene signature is to integrate multiple microarray data sets into one study. From this, we can detect a subset of genes whose expressions can be highly correlated with experimental design for as many microarray data sets as possible. Various classification algorithms have been employed to train a classifier to maximize the prediction power of signatures [19].

In addition to these classification models, it is also important to determine signatures in terms of their common or distinct expression differentiation without classification labels. A simple way is to pool all the data sets into one data set and then use a clustering algorithm to partition the pooled data. However this is not possible when data sets have different number of samples (dimensions). On the other hand, separately analyzing each data set individually and then integrating the separate results may be inaccurate and inefficient. One popularly approach is to use non-negative matrix factorization (NMF) [20,21]. However it has been noted that NMF is a linear algorithm which may not be able to explore complex pattern across multiple data sets [22].

We propose an extension to SOM, integrative SOM (iSOM), for exploring structural relationships in integrative studies. Based on modern high-resolution microarray technology, we assume that the differential expression variance of a gene is relatively low and that most microarray expression data demonstrate a high positive correlation among replicates. We therefore propose a mean pattern model, *i.e.* all differential expressions are assumed to be random samples drawn from a mean pattern model. The model can be considered as a library of all possible mean differential expressions with variances. An integrative study of multiple microarray data sets then aims to reveal how the hidden and unknown mean pattern model shapes the observed differential expressions from multiple data sets. We can thereafter discover how differentially expressed genes are common or distinct in multiple biologically related disease types or in different

platforms of the same study.

The proposed iSOM is composed of two major steps. In the first step, the vector of all differential expression matrices across multiple data sets is analyzed using SOM leading to an array in which each neuron represents one local pattern, *i.e.*, one mean differential expression. In the second step, we assume this mean pattern model underlies the observed differential expressions. The standard SOM learning rule is thus altered. We show in this paper how iSOM can be useful for correlated pattern discovery using both simulated data and real data.

## 2. Methods

### 2.1. Data Pre-Processing and Filtering

Each microarray expression data set is first normalized to the logarithm scale. After normalization, a significance analysis is carried out using eBayes [23]. Only those genes which show sufficient significant differential expressions are selected for further analysis. This is because the major aim of integrative study such as common gene signature discovery will not be interested in genes, which do not show significant differential expression. Afterwards, we save the differential expression matrix for significantly differentially expressed genes for each data.

### 2.2. Self-Organizing Map Algorithm

The conventional self-organizing map (SOM) is a two layer neural network in which the first layer is composed of input neurons for input variables ( $\mathbf{x}_n$ ,  $n$  stands for the  $n^{\text{th}}$  input vector) while the second layer is composed of an array of output neurons. Each output neuron has a weight vector acting as a parameter vector,  $\mathbf{w}_k$ , where  $k$  represents the  $k^{\text{th}}$  neuron. Note that  $\mathbf{w}_k$  and  $\mathbf{x}_n$  have the same dimensionality. The competitive learning of SOM is to update  $\mathbf{w}_k$  based on the distance between  $\mathbf{w}_k$  and  $\mathbf{x}_n$

$$\Delta \mathbf{w}_k^{t+1} = \eta^t \alpha_n (\mathbf{x}_n - \mathbf{w}_k^t)$$

where  $0 < \eta^t < 1$  is called a learning rate at time  $t$  and  $\alpha_n$  is the neighborhood constrain associated with  $\mathbf{x}_n$ . Note that SOM employs an online learning strategy, *i.e.* model parameters are undated once one input vector is fed.

### 2.3. iSOM

We assume that each output neuron has a scalar weight functioning as a mean differential expression. This means that the feature map of iSOM is an array of mean differential expressions. We denote  $\mu_k$  as the weight (mean differential expression) of the  $k^{\text{th}}$  output neuron. Suppose we have two data sets, X and Y (it is easy to

generalize the analysis of two data sets to multiple data sets), we denote the  $n^{\text{th}}$  input of X as  $\mathbf{x}_n$  and the  $m^{\text{th}}$  input of Y as  $\mathbf{y}_m$ . We assume that each vector of differential expression samples is a random sample drawn from a hidden signal, *i.e.* mean differential expression expressed by  $\mu_k$ . Using the standard SOM learning rule, the learning rules for iSOM are

$$\Delta \mu_k = \eta^t \alpha_n (\mathbf{x}_n - \mu_k \mathbf{i})^t \mathbf{i}$$

$$\Delta \mu_k = \eta^t \alpha_m (\mathbf{y}_m - \mu_k \mathbf{i})^t \mathbf{i}$$

where  $\mathbf{i}$  is a vector of ones. These learning rules are used in building an iSOM model.

### 2.4. Random Selection of Expressions in Training

We wish to train an iSOM in such a way as to ensure the successful discovery of differential expression pattern between multiple data sets. Suppose we have two data sets, X and Y, their relationship may be one of the following: 1)  $X \supset Y$ ; 2)  $Y \supset X$ ; 3)  $X \equiv Y$ ; 4)  $(X \setminus Y) \neq \phi$  and  $(Y \setminus X) \neq \phi$ . When  $X \equiv Y$ , any training procedure will do well. For the remaining cases, the order of data selection of a training process is crucial. For instance, if  $X \supset Y$  and we use Y to train an iSOM first, the iSOM will have fully learnt differential expression patterns from Y and leave no space for extra differential expression pattern in X. Consequently, the data structure learned from Y will then be lost during the learning process using X. This will then lead to biased pattern discovery. The same problem occurs also in the other two scenarios. To avoid this, we propose a random sampling strategy for iSOM training which randomly selects one microarray data set and one differential expression vector (corresponding to a gene) at every step of the training process. This ensures unbiased pattern discovery across multiple microarray data sets.

## 3. Results

### 3.1. Simulated Scenarios

We design three simulated scenarios. The first satisfies the condition  $X \equiv Y$ . Here we design a mean pattern model with ten differential expression means ( $\mathcal{G}$ ) as  $-5, -4, -3, -2, -1, 1, 2, 3, 4, \text{ and } 5$ . The X space is of two dimensions and the Y space is of three dimensions. Both X and Y spaces are composed of random vectors drawn from the mean pattern model. Each mean differential expression value is used to draw 100 vectors randomly for both X and Y spaces. In total, there are 2000 data points. Each vector is drawn from the mean pattern model,  $\mathbf{z}_n \sim \mathcal{G}(\mu_k, \sigma^2)$ , where H means the number of output neurons,  $\mu_k \in \mathcal{G}$ ,  $\mathbf{z}_n$  is either  $\mathbf{x}_n$  or  $\mathbf{y}_n$  and  $\sigma \in (0.1, 0.2, 0.3, 0.4, 0.5)$ . Each vector is labeled indi-

cating which mean differential expression value the vector is drawn from. We then measure whether this topological structure is maintained during iSOM modeling. For two data sets, iSOM generates two output maps named as  $\Phi_x = (\phi_1^x, \phi_2^x, \dots, \phi_H^x)$  and  $\Phi_y = (\phi_1^y, \phi_2^y, \dots, \phi_H^y)$  for two (or more) data sets, where  $\phi_h^x$  and  $\phi_h^y$  are assumed to be multivariate mean differential expression patterns drawn from the same mean differential expression

$$E(\mathbf{x}_n | \forall \mathbf{x}_n \in \phi_h^x) = E(\mathbf{y}_n | \forall \mathbf{y}_n \in \phi_h^y) \in \{\mu_k\} \quad (1)$$

We therefore compare  $\Phi_x$  against  $\Phi_y$  for each neuron to examine whether Equation (1) is satisfied. **Table 1** shows this evaluation for five variance values. It can be seen that the relationship (the designed topological structure) is well maintained during iSOM modeling. The maximum error rate is about 5.6% (=113/2000) for simulation with a relatively large s.d. of  $\sigma = 0.5$ . This data set revealed no distinct genes. Therefore there is no measurement for checking if distinct genes can be well revealed.

The second simulated scenario satisfies the condition  $X \supset Y$ . The mean differential expression structure is the same as above. The X space is composed of all ten differential expression patterns while the Y space is composed of eight of them (-3, -2, -1, 1, 2, 3, 4, 5). This means that the X space has 200 genes which are distinct - only occurring in the X space not the Y space. There are therefore 1800 data points. In **Table 1**, we also found the error to be small. The maximum error rate is 8.6%. In addition to error, we also examined how the distinct genes from the X data set can be revealed. Suppose the X data set contains some distinct differential expression patterns, which are not found from the Y data set. This means that some neurons  $\phi_h^x$  contain vectors drawn from the mean pattern model, which cannot be found

**Table 1.** Evaluation on three simulated data sets. “Sigma” means standard deviation. “Error” means the number of times that Equation (1) is violated. “Distinct” is the percentage of distinct genes of one mean differential expression from one data set are identified. In the second simulated scenario, only one data set has distinct genes, therefore two measurements are used. In the third simulated scenario, both data sets have distinct genes, therefore we use four measurements.

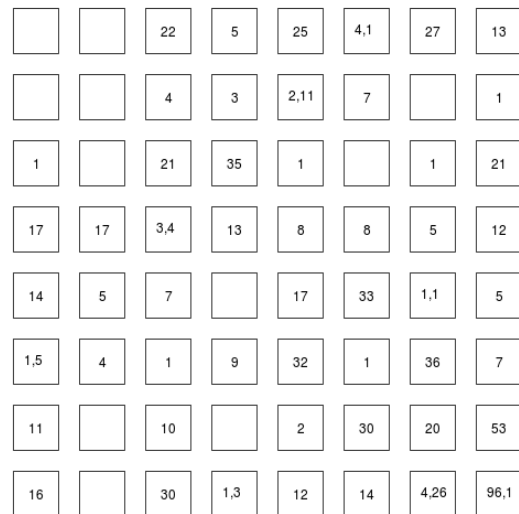
Sigma	Error			Distinct	
	Toy 1	Toy 2	Toy 3	Toy 2	Toy 3
0.1	0	0	0	100/100	100/100/100/100
0.2	0	0	0	100/100	100/100/100/100
0.3	0	24	2	100/100	100/100/100/99
0.4	23	61	29	100/99	100/98/99/100
0.5	113	155	35	100/94	100/98/98/100

from the corresponding neuron from  $\Phi_y$ , i.e.  $\phi_h^y$ . **Table 1** shows the percentages of distinct differential expression patterns which were uniquely preserved during iSOM learning. It can be seen that iSOM is well adapted for this kind of patterns.

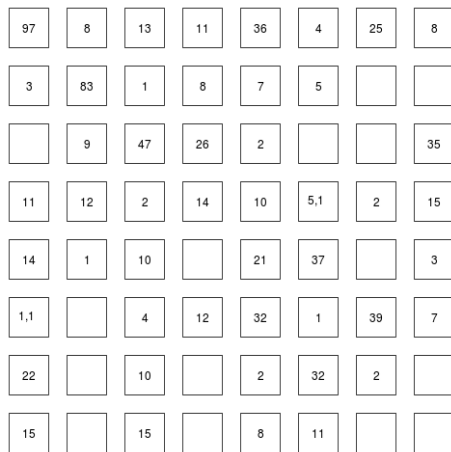
The third simulated scenario satisfies the conditions  $(X \setminus Y) \neq \emptyset$  and  $(Y \setminus X) \neq \emptyset$ . The mean differential expression pattern structure is the same as above. The X space is composed of eight differential expression patterns centered at (-5, -4, -3, -2, -1, 1, 2, 3) while for Y the patterns are centered around (-3, -2, -1, 1, 2, 3, 4, 5). In this case, we have distinct differential expression patterns from both data spaces. In this case, iSOM can still discover this kind of dual distinct differential expression patterns—**Table 1**. In addition, the error is still very small. **Figure 1** shows the distribution of data points of the first simulated scenario mapped to the iSOM model. A single number in a grid representing the number of data points falling into the grid in a grid indicates that the neuron is composed of data points as random samples of a single mean differential expression value. When there are more than one numbers, it means that data points of multiple mean differential expression values are mapped to the same neuron. **Figure 2** shows the distribution of data points of the second simulated scenario mapped to the iSOM model. We can see that some neurons are uniquely occupied by a single data set.

### 3.2. Real Data

We use two data sets downloaded from Gene Expression Omnibus (GEO), GSE12630, of breast cancer metastasis and liver cancer metastasis. Both data sets contain four replicates. Using eBayes with a critical  $p$  value of 0.05, 2359 differentially expressed genes were found for the



**Figure 1.** The distribution of data points of the first data set of the third simulated scenario.



**Figure 2.** The distribution of data points of the second data set of the third simulated scenario.

breast cancer data set and 19029 differentially expressed genes for the liver cancer data set. The two data sets are then used for the integrative analysis using iSOM. The neuron array is set to be ten by ten (100 neurons).

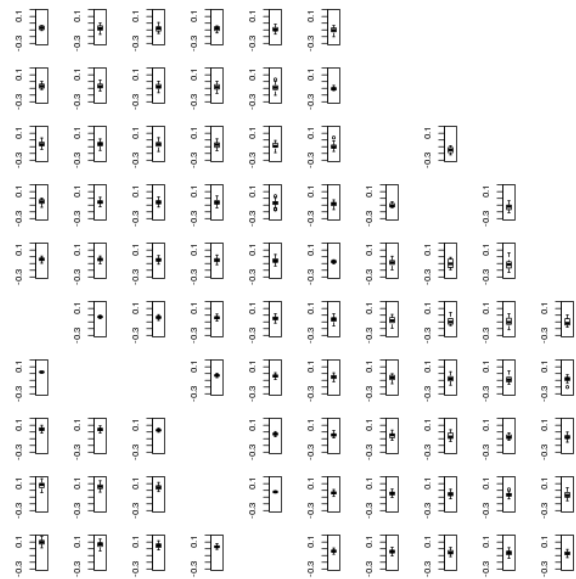
Using iSOM, we found seven common differentially expressed genes between the breast cancer metastasis and liver cancer metastasis datasets. Among them, SRSF1 (probe set ID 201741\_x\_at) was an up-regulated gene while the six others were down-regulated—**Table 2**. The gene SRSF1 has been studied in relation to breast cancer [24] and liver cancer [25]. Some of the down-regulated genes have been studied in the context of both breast and liver cancer metastasis—**Table 2**. The number of unique differentially expressed genes for breast cancer metastasis is 773 while the number of unique differentially expressed genes for liver cancer metastasis is 297. **Figures 3** and **4** show the differential expression patterns (derived using iSOM) for the breast cancer metastasis data and liver cancer metastasis data respectively. Comparing these two maps, it can be seen that both the number of common differentially expressed genes and the number of distinct differentially expressed genes are quite small.

### 4. Conclusion

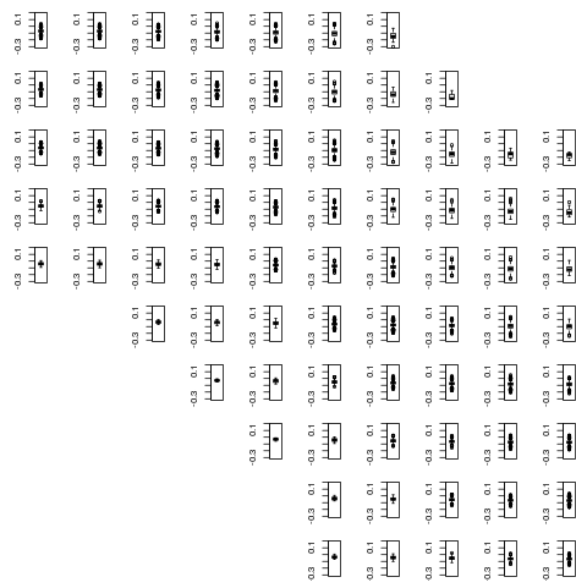
We have presented a novel extension to SOM for integrative studies of microarray expression data sets. The proposed integrative SOM (iSOM) is based on the assumption that the microarray expression data under consideration has small variance across replicates. This assumption is reasonable considering recent technology improvement in the microarray experimental precision. We assume that differential expressions across multiple microarray expression data sets with medical or biological relevance are random samples from a mean pattern model. This mean pattern model is a one-dimension structure of

**Table 2.** Six common down-regulated genes between the breast and liver cancer metastasis datasets. Relevant references indices are given under ‘breast’ and ‘liver’.

Probe set ID	Gene symbol	Breast	Liver
200736_s_at	GPX1	[26]	[27]
200872_at	S100A10	[28]	[29]
201868_s_at	TBL1X	[30]	n.a.
202069_s_at	IDH3A	[31]	n.a.
201150_s_at	TIMP3	[32]	[33]
202430_s_at	PLSCR1	n.a.	[34]



**Figure 3.** Differential expression pattern for breast cancer metastasis.



**Figure 4.** Differential expression pattern for liver cancer metastasis.

mean differential expressions. It is this linear structure that shapes the observed differential expressions from multiple data sets in a multivariate model. The iSOM approach can also be used in cross-omics and cross-species studies.

## REFERENCES

- [1] T. Kohonen, "Self-Organizing Maps," Springer, Berlin, 2001. <http://dx.doi.org/10.1007/978-3-642-56927-2>
- [2] J. Yao, Q. Zhao, Y. Yuan, L. Zhang, X. Liu, W. K. Yung and J. N. Weinstein, "Identification of Common Prognostic Gene Expression Signatures with Biological Meanings From Microarray Gene Expression Datasets," *PLoS One*, Vol. 7, 2012, Article ID: e45894. <http://dx.doi.org/10.1371/journal.pone.0045894>
- [3] Y. Park, E. S. Park, S. B. Kim, S. C. Kim, B. H. Sohn, I. S. Chu, W. Jeong, G. B. Mills, L. A. Byers and J. S. Lee, "Development and Validation of a Prognostic Gene-Expression Signature for Lung Adenocarcinoma," *PLoS One*, Vol. 7, 2012, Article ID: e44225. <http://dx.doi.org/10.1371/journal.pone.0044225>
- [4] A. Zaravinos, G. I. Lambrou, D. Volanis, D. Delakas and D. A. Spandidos, "Spotlight on Differentially Expressed Genes in Urinary Bladder Cancer," *PLoS One*, No. 6, 2011, Article ID: e18255. <http://dx.doi.org/10.1371/journal.pone.0018255>
- [5] M. LaBonte, P. M. Wilson, W. Fazzone, S. Groshen, H. J. Lenz and R. D. Ladner, "DNA Microarray Profiling of Genes Differentially Regulated by the Histone Deacetylase Inhibitors Vorinostat and LBH589 in Colon Cancer Cell Lines," *BMC Med Genomics*, Vol. 2, 2009, p. 67. <http://dx.doi.org/10.1186/1755-8794-2-67>
- [6] M. Howlett, A. S. Giraud, H. Lescesen, C. B. Jackson, A. Kalantzis, I. R. Van Driel, L. Robb, M. Van der Hoek, M. Ernst, T. Minamoto, A. Boussioutas, H. Oshima, M. Oshima and L. M. Judd, "The Interleukin-6 Family Cytokine Interleukin-11 Regulates Homeostatic Epithelial Cell Turnover and Promotes Gastric Tumor Development," *Gastroenterology*, Vol. 136, 2009, pp. 967-977. <http://dx.doi.org/10.1053/j.gastro.2008.12.003>
- [7] A. Alaiya, M. Al-Mohanna, M. Aslam, Z. Shinwari, L. Al-Mansouri, M. Al-Rodayan, M. Al-Eid, I. Ahmad, K. Hanash, A. Tulbah, A. Mahfooz and C. Adra, "Proteomics-Based Signature for Human Benign Prostate Hyperplasia and Prostate Adenocarcinoma," *International Journal of Oncology*, Vol. 38, 2011, pp. 1047-1057. <http://dx.doi.org/10.3892/ijo.2011.937>
- [8] L. Cheng, W. Lu, B. Kulkarni, T. Pejovic, X. Yan, J. H. Chiang, L. Hood, K. Odunsi and B. Lin, "Analysis of Chemotherapy Response Programs in Ovarian Cancers by the Next-Generation Sequencing Technologies," *Gynecologic Oncology*, Vol. 117, 2010, pp. 159-169. <http://dx.doi.org/10.1016/j.ygyno.2010.01.041>
- [9] A. Planche, M. Bacac, P. Provero, C. Fusco, M. Delorenzi, J. C. Stehle and I. Stamenkovic, "Identification of Prognostic Molecular Features in the Reactive Stroma of Human Breast and Prostate Cancer," *PLoS One*, Vol. 6, 2011, Article ID: e18640. <http://dx.doi.org/10.1371/journal.pone.0018640>
- [10] A. Collura, L. Marisa, D. Trojan, O. Buhard, A. Lagrange, A. Saget, M. Bombled, P. Méchighel, M. Ayadi, M. Mulleris, A. de Reynies, M. Svrcek, J. F. Fléjou, J. C. Florent, F. Mahuteau-Betzer, A. M. Faussat and A. Duval, "Extensive Characterization of Sphere Models Established From Colorectal Cancer Cell Lines," *Cellular and Molecular Life Sciences*, 2012, in Press.
- [11] M. Stevenson, W. Mostertz, C. Acharya, W. Kim, K. Walters, W. Barry, K. Higgins, S. A. Tuchman, J. Crawford, G. Vlahovic, N. Ready, M. Onaitis and A. Potti, "Characterizing the Clinical Relevance of an Embryonic Stem Cell Phenotype in Lung Adenocarcinoma," *Clinical Cancer Research*, Vol. 15, 2009, pp. 7553-7561. <http://dx.doi.org/10.1158/1078-0432.CCR-09-1939>
- [12] E. Markert, A. J. Levine and A. Vazquez, "Proliferation and Tissue Remodeling in Cancer: The Hallmarks Revisited," *Cell Death & Disease*, Vol. 3, 2012, p. 3397. <http://dx.doi.org/10.1038/cddis.2012.140>
- [13] N. Fankhauser, I. Cima, P. Wild and W. Krek, "Identification of a Gene Expression Signature Common to Distinct Cancer Pathways," *Cancer Information*, Vol. 11, 2012, pp. 139-146. <http://dx.doi.org/10.4137/CIN.S9542>
- [14] M. Daves, S. G. Hilsenbeck, C. C. Lau and T. K. Man, "Meta-Analysis of Multiple Microarray Datasets Reveals a Common Gene Signature of Metastasis in Solid Tumors," *BMC Med Genomics*, Vol. 4, 2011, p. 56. <http://dx.doi.org/10.1186/1755-8794-4-56>
- [15] S. De and F. Michor, "DNA Secondary Structures and Epigenetic Determinants of Cancer Genome Evolution," *Nature Structural & Molecular Biology*, Vol. 18, 2011, pp. 950-955. <http://dx.doi.org/10.1038/nsmb.2089>
- [16] F. Buffa, A. L. Harris, C. M. West and C. J. Miller, "Large Meta-Analysis of Multiple Cancers Reveals a Common, Compact and Highly Prognostic Hypoxia Metagene," *British Journal of Cancer*, Vol. 102, 2010, pp. 428-435. <http://dx.doi.org/10.1038/sj.bjc.6605450>
- [17] N. Slavov and K. A. Dawson, "Correlation Signature of the Macroscopic States of the Gene Regulatory Network in Cancer," *Proceedings of the National Academy of Sciences*, Vol. 106, 2009, pp. 4079-4084. <http://dx.doi.org/10.1073/pnas.0810803106>
- [18] I. Wang, B. Zhang, X. Yang, J. Zhu, S. Stepaniants, C. Zhang, Q. Meng, M. Peters, Y. He, C. Ni, D. Slipetz, M. A. Crackower, H. Houshyar, C. M. Tan, E. Asante-Appiah, G. O'Neill, M. J. Luo, R. Thieringer, J. Yuan, C. S. Chiu, P. Y. Lum, J. Lamb, Y. Boie, H. A. Wilkinson, E. E. Schadt, H. Dai and C. Roberts, "Systems Analysis of Eleven Rodent Disease Models Reveals an Inflammation Signature and Key Drivers," *Molecular Systems Biology*, Vol. 8, 2012, p. 594. <http://dx.doi.org/10.1038/msb.2012.24>
- [19] S. Michiels, S. Koscielny and C. Hill, "Prediction of cancer Outcome with Microarrays: A Multiple Random Validation Strategy," *Lancet*, Vol. 365, 2005, pp. 488-492, 2005. [http://dx.doi.org/10.1016/S0140-6736\(05\)17866-0](http://dx.doi.org/10.1016/S0140-6736(05)17866-0)
- [20] E. Segal, N. Friedman, D. Koller and A. Regev, "A Module Map Showing Conditional Activity of Expression Modules in Cancer," *Nature Genetics*, Vol. 36, 2004, pp.

- 1090-1098. <http://dx.doi.org/10.1038/ng1434>
- [21] P. Tamayo, D. Scanfeld, B. L. Ebert, M. A. Gillette, C. W. Roberts and J. P. Mesirov, "Metagene Projection for Cross-Platform, Cross-Species Characterization of Global Transcriptional States," *Proceedings of the National Academy of Sciences*, Vol. 104, 2007, pp. 5959-5964. <http://dx.doi.org/10.1073/pnas.0701068104>
- [22] K. Devarajan, "Nonnegative Matrix Factorization: An Analytic and Interpretive Tool in Computational Biology," *PLoS Computational Biology*, Vol. 4, 2008, Article ID: e1000029. <http://dx.doi.org/10.1371/journal.pcbi.1000029>
- [23] B. Efron, R. Tibshirani, J. D. Storey and V. Tusher, "Empirical Bayes Analysis of a Microarray Experiment," *Journal of American Statistical Association*, Vol. 96, 2001, pp. 1151-1160. <http://dx.doi.org/10.1371/journal.pcbi.1000029>
- [24] H. Gautrey and A. J. Tyson-Capper, "Regulation of Mcl-1 by SRSF1 and SRSF5 in Cancer Cells," *PLoS One*, Vol. 7, 2012, Article ID: e51497. <http://dx.doi.org/10.1371/journal.pone.0051497>
- [25] Ú. Muñoz, J. E. Puche, R. Hannivoort, U. E. Lang, M. Cohen-Naftaly and S. L. Friedman, "Hepatocyte Growth Factor Enhances Alternative Splicing of the Kruppel-Like Factor 6 (KLF6) Tumor Suppressor to Promote Growth through SRSF1," *Molecular Cancer Research*, Vol. 10, 2012, pp. 1216-1227. <http://dx.doi.org/10.1158/1541-7786.MCR-12-0213>
- [26] M. Kulak, A. R. Cyr, G. W. Woodfield, M. Bogachek, P. M. Spanheimer, T. Li, D. H. Price, F. E. Domann and R. J. Weigel, "Transcriptional Regulation of the GPX1 Gene by TFAP2C and Aberrant CpG Methylation in Human Breast Cancer," *Oncogene*, 2013, in Press.
- [27] N. Blum, K. Mueller, D. Lippmann, C. C. Metges, T. Linn, J. Pallauf and A. S. Mueller, "Feeding of Selenium Alone or in Combination with Glucoraphanin Differentially Affects Intestinal and Hepatic Antioxidant and Phase II Enzymes in Growing rats," *Biological Trace Element Research*, Vol. 151, 2013, pp. 384-399. <http://dx.doi.org/10.1007/s12011-012-9567-6>
- [28] T. Reddy, C. Li, X. Guo, H. K. Myrvang, P. M. Fischer and L. V. Dekker, "Design, Synthesis, and Structure-Activity Relationship Exploration of 1-Substituted 4-Aroyl-3-hydroxy-5-phenyl-1H-pyrrol-2(5H)-One Analogues as Inhibitors of the Annexin A2-S100A10 Protein Interaction," *Journal of Medicinal Chemistry*, Vol. 54, 2011, pp. 2080-2094. <http://dx.doi.org/10.1021/jm101212e>
- [29] Y. Tan, S. Y. Ma, F. Q. Wang, H. P. Meng, C. Mei, A. Liu and H. R. Wu, "Proteomic-Based Analysis for Identification of Potential Serum Biomarkers in Gallbladder Cancer," *Oncology Reports*, Vol. 26, 2011, pp. 853-859.
- [30] V. Perissi, C. Scafoglio, J. Zhang, K. A. Ohgi, D. W. Rose, C. K. Glass and M. G. Rosenfeld, "TBL1 and TBLR1 Phosphorylation on Regulated Gene Promoters Overcomes Dual CtBP and NCoR/SMRT Transcriptional Repression Checkpoints," *Molecular Cell*, Vol. 29, 2008, pp. 755-766. <http://dx.doi.org/10.1016/j.molcel.2008.01.020>
- [31] A. Qattan, M. Radulovic, M. Crawford and J. Godovac-Zimmermann, "Spatial Distribution of Cellular Function: The Partitioning of Proteins between Mitochondria and the Nucleus in MCF7 Breast Cancer Cells," *Journal of Proteome Research*, Vol. 11, 2012, pp. 6080-6101.
- [32] T. Neill, H. Painter, S. Buraschi, R. T. Owens, M. P. Lisanti, L. Schaefer and R. V. Iozzo, "Decorin Antagonizes the Angiogenic Network: Concurrent Inhibition of Met, Hypoxia Inducible Factor 1 $\alpha$ , Vascular Endothelial Growth Factor A, and Induction of Thrombospondin-1 and TIMP3," *The Journal of Biological Chemistry*, Vol. 287, 2012, pp. 5492-5506. <http://dx.doi.org/10.1074/jbc.M111.283499>
- [33] S. Turner, D. Mangnall, N. C. Bird, R. A. Bunning and M. E. Blair-Zajdel, "Expression of ADAMTS-1, ADAMTS-4, ADAMTS-5 and TIMP3 by Hepatocellular Carcinoma Cell Lines," *International Journal of Oncology*, Vol. 41, 2012, pp. 1043-1049.
- [34] W. Cui, S. Y. Li, J. F. Du, Z. M. Zhu and P. An, "Silencing Phospholipid Scramblase 1 Expression by RNA Interference in Colorectal Cancer and Metastatic Liver Cancer," *Hepatobiliary & Pancreatic Diseases International*, Vol. 11, 2012, pp. 393-400. [http://dx.doi.org/10.1016/S1499-3872\(12\)60197-0](http://dx.doi.org/10.1016/S1499-3872(12)60197-0)