A Coupled Three-Step Network-Based Approach to Identify Genes Associated with Breast Cancer

Michael Netzer*, Xiaocong Fang†, Michael Handler*, Christian Baumgartner*
*Research Group for Clinical Bioinformatics,
Institute of Electrical, Electronic and Bioengineering, UMIT, 6060 Hall in Tirol, Austria
Email: michael.netzer|michael.handler|christian.baumgartner@umit.at
†Zhongshan Hospital, Fudan University
200032 Xuhui District, Shanghai,China
fangxiaocong@gmail.com

***Abstract*—New biomarker candidates for breast cancer recurrence are urgently needed to provide patients an optimal treatment and avoid "overtreatment" where patients needlessly suffer from the toxic side effects of chemotherapy. In this work, we present a new network-based approach to identify biomarker candidates for breast cancer recurrence. Our coupled three-step strategy first selects relevant genes by using a filter approach. In the second step, we infer a new type of network where the number of edges of each vertex (i.e., degree) represents the predictive value of the underlying gene. In the third step, we conduct a database search for biomedical interpretation of our findings. Using a breast cancer microarray dataset we could verify top ranked genes associated with breast cancer and other pathophysiological processes.**

***Index Terms*—biomarker discovery; microarray; feature selection; networks; breast cancer**

I. Introduction

Recent biotechnological advances in the "omics" sciences such as microarrays have led to high amounts of data. In particular, this data is characterized by a high number of features (e.g., genes) and a small number of samples (large p, small n problem) [1]. To identify highly predictive biomarker candidates, sophisticated feature selection approaches, including filters and wrappers, are required [2]. Wrapper approaches [3] use learning algorithms (i.e., a classifier if the dependent variable is categorical) and a search strategy to identify sets of highly discriminating features. Filter approaches that calculate a measure for every feature representing its predictive ability are independent of a classifier and generally have less extensive computational costs [4].

Advances of computer systems have also led to extensive biological network analysis [5]–[7], increasing understanding of interactions between genes, proteins and metabolites.

In this work, we introduce a new three-step network-based approach to identify genes related to breast cancer recurrence. New biomarker candidates demonstrating recurrence are urgently needed for diagnosis and to provide the patients with an optimal treatment plan. According to Weigelt et al. [8], currently, more than 80% of female patients with breast cancer receive adjuvant chemotherapy, though only approximately 40% relapse and ultimately die of metastatic breast cancer. Consequently, many patients are over-treated needlessly, suffering from toxic side effects of chemotherapy [8].

Section "Materials and Methods" describes our new three-step network-based approach. In the "Results and Discussion" section, we show our results for an example dataset. Finally, we conclude and discuss our method and findings.

II. Materials and Methods

A. Data

We used 59 preprocessed microarray spectra (platform GPL1390) from the study of Hoadley et al. [9] available at the GEO database [10] ($n_s = 59$). The data includes 29 controls and 30 cases (breast cancer recurrence within 36 months). In our experiments we only use genes where the ensembl gene id is available (n_f = 11638).

B. Computational approach

Given is a set of tuples (dataset) $T = \{x_i, y_i\}_{i=1}^{n_s}$ with $x_i \in \mathbb{R}^{n_f}$ and $y_i \in \{case, control\}$, where n_f is the number of features, n_s is the number of samples and y is the set of classlabels. Recently, we developed a new supervised approach to infer networks based on the ratios between metabolites [11]. Therefore, we first calculated all ratios R between features F, where $r_{ij} = |\log_2\left(\frac{f_i}{f_j}\right)|$ with $i > j$, and $f \in F, r \in R$.

Note that the logarithm induces symmetry of the ratios and their reciprocals, respectively. We then created a graph G with:

$$G_{ij} = \begin{cases} 1 & \text{if } |s_{ij}| > \tau \\ 0 & \text{else,} \end{cases} \quad (1)$$

for $i, j \in 1, \ldots, n_f$, where τ is a defined threshold. The score s representing the discriminatory ability was calculated using the BI filtering method [12]. Briefly, BI combines a discriminance measure (DA^*), the coefficient of variation CV and Δ_{change} representing the strength of the change. For the unpaired BI, we define $\Delta = \frac{\overline{x}}{\overline{x}_{ref}}$, where $\overline{x}$ is the mean of the comparison group (e.g., case) and $\overline{x}_{ref}$ is the mean of the reference group (e.g., control).

In our recent work, we could show that this new type of network, based on the ratios, outperforms other networks as

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-190-8

correlation networks in terms of accuracy (see also Netzer et al., 2011). However, this approach has several limitations: i) Creating the network for n features results in $\frac{n \cdot (n-1)}{2}$ comparisons and BI calculations ($\mathcal{O}(n^2)$); ii) by definition the values for f must be positive to calculate the logarithms and $\overline{x}_{ref}$ must be $\neq 0$ to calculate the BI scores.

In particular, when dealing with standard normal distributed microarray data ($N(0,1)$) we have a high number of features ($n > 10,000$) including negative values for f and $\overline{x}_{ref}$ is close to zero.

Therefore, in this work, we propose a new generic three-step strategy to overcome the afore mentioned restrictions for standard normal distributed datasets with a high number of features (e.g., preprocessed microarray data):

Step 1: In order to reduce the number of features we first perform a feature selection. Therefore, we use a filter method and calculate the score s representing the discriminatory ability of each feature. We remove all features f with a score less than a defined threshold ($s_f < \tau$). Finally, we obtain our reduced dataset T_r.

Step 2: Given a standard normal distributed dataset we calculate all differences D between features F in T_r, where $d_{ij} = |f_i - f_j|$ with $i > j$, and $f \in F, d \in D$. Similar to equation 1 we finally construct the graph with

$$G_{ij} = \begin{cases} 1 & \text{if } s_{ij} > \tau \\ 0 & \text{else,} \end{cases} \tag{2}$$

In our experiments, we used the information gain [13] to calculate the score s on the distances d for step 1 and 2. The information gain IG of a feature f is given by [14]

$$IG = H(Y) - H(Y|X), \text{ where} \tag{3}$$

$$H(Y) = -\sum_{y \in Y} p(y) \log_2(p(y)), \text{ and} \tag{4}$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x)), \tag{5}$$

where $H(Y)$ denotes the entropy for Y (class variable) and $H(Y|X)$ is the entropy of Y after observing X.

The information gain easily allows to identify features with no or less discriminatory ability. Therefore, we set $\tau = 0$ (filtering threshold) to remove features with no information regarding the class attribute. In addition, we also used the information gain because it can also deal with values ≤ 0.

Step 3: After the network is inferred the genes are ranked according to the topological descriptor of the vertices. In this study we ranked the genes according to their degree (i.e., number of edges).

To verify and interpret our findings a database search to multiple repositories such as the Database for Annotation, Visualization and Integrated Discovery (DAVID) [15] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was conducted [16].

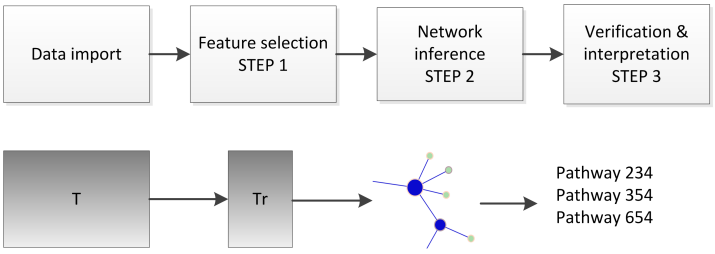

Fig. 1. The proposed workflow. The second row outlines the result of each process from row one. T denotes the entire dataset and T_r denotes the reduced dataset after feature selection.

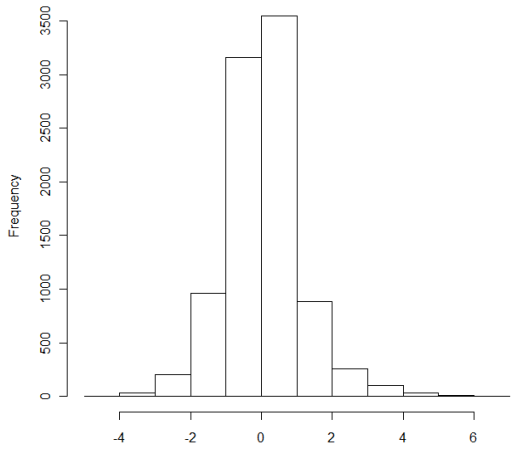

Fig. 2. Standard normal distribution of the dataset.

The feature selection was performed using Weka [17]. We used R [18] to implement the network approach methods. There exist several R packages for handling and analyzing graphs such as `graph` [19], `igraph` [20], `QuACN` [21] and `BioNet` [22]. The overall workflow including data import, feature selection, network inference and pathway analysis is depicted in Fig. 1.

The used hardware platform was an Intel Centrino 2x-1.83 GHz PC with 2048 MB RAM.

III. Results and Discussion

This works aims at identifying new genes related to breast cancer recurrence. The distribution of the used dataset is shown in Fig. 2. After feature selection of step 1 a total of 156 genes yielded an information gain > 0. The resulting network after applying step 2 to the reduced dataset is shown in Fig. 3. The ten top ranked genes using the degree (i.e., number of edges) are listed in Table I.

The related pathophysiological processes of the identified gene set are shown in Table II.

Kallilreins gene family are a group of serine proteases and known to be involved in several endocrine related malignancies

[23]. Kallikrein gene 7 (KLK7) was first reported in 1991 in the desquamation of stratum corneum [24] and later reported to be highly upregulated in ovarian carinomas [25]. Recent studies have reported that KLK7 can be up regulated primarily by estrogens and glucocorticoids [26], [27]. The expression of the KLK7 gene was supposed to be the most independent prognostic marker for the survival of patients with breast cancer [27].

Salivary amylase alpha 1 (AMY1) is involved in the starch and sucrose metabolism. It was reported to be highly expressed in individuals with high starch diets [28] and highly activated under psychosocial stress [29]. There is no existing evidence which supports the association between AMY1 and breast cancer, however, recent study reported AMY1 is an important modulator of cAMP-dependent protein kinase (PKA) which has versatile functions in cells [30].

The PH domain of PHLDA2 can compete with the PH domain of some other proteins, thereby interfering with their binding to phosphatidylinositol 4,5-bisphosphate (PIP2) and phosphatidylinositol 3,4,5-triphosphate (PIP3) in membrane lipids thus involved in various biological processes [15].

SnrpC belongs to the U1 small nuclear ribonucleoprotein C family and was reported to be involved in the splicing of mRNA [31]. The peptidylarginine deiminases 2 (PAD2) is a member of PADs which catalyze citrullination by converting arginine residues of proteins [32], [33]. It was also reported to play a role in inflammatory response, cell apoptosis [34], [35] as well as the gene regulation in the mammary glands [36].

Wee1 is another gene which was previously reported to be involved in tumourigenesis [37], [38] as well as the signaling regulation in breast cancer stem cells [39]. It was suggested to act as a tumor suppressor via regulating Cyclin and cyclin-dependent kinase complexes [37].

SOCS family proteins are part of a classical negative feedback system that regulates cytokine signal transduction of which SOCS2 appears to regulate the growth hormone/IGF1 signaling pathway [15], [40]. It is also suggested to play a role in the oncogenesis of ovary and mammary gland [41].

SOX5 is one of the high-mobility group (HMG) which has been recognized as a key player in the regulation of embryonic development [42] and in the determination of cell fate [43] reported to be involved in the progression of glioma and prostate cancer [44], [45].

KIAA0494 was supposed to play a role in the splicing of eukaryotic pre-mRNAs [46] but there are no more studies provided on its effect on tumorgenesis. CTP synthase (CTPS) plays a predominant role in CTP synthesis by converting UTP to CTP thus controling cell proliferation, differentiation and apoptosis [47]. Previous studies have demonstrated that CTPS depletion resulted in stabilization of wild-type p53 and showed antitumor effects in breast cancer cells [48].

Among the top 10 most related genes, we found that half of them had been previously reported to play roles in tumorogenesis, such as KLK7, Wee1, SOCS, SOX5 and CTPS. Furthermore, some had been studied in breast cancer (KL7, Wee1,SOCS, CTPS). However, the others were first suspected to be involved in the development or progression of breast cancer. Our results revealed that it is possible that these genes provide new targets for the control of breast cancer. However, further studies are warranted and essential to verify and validate these promising findings.

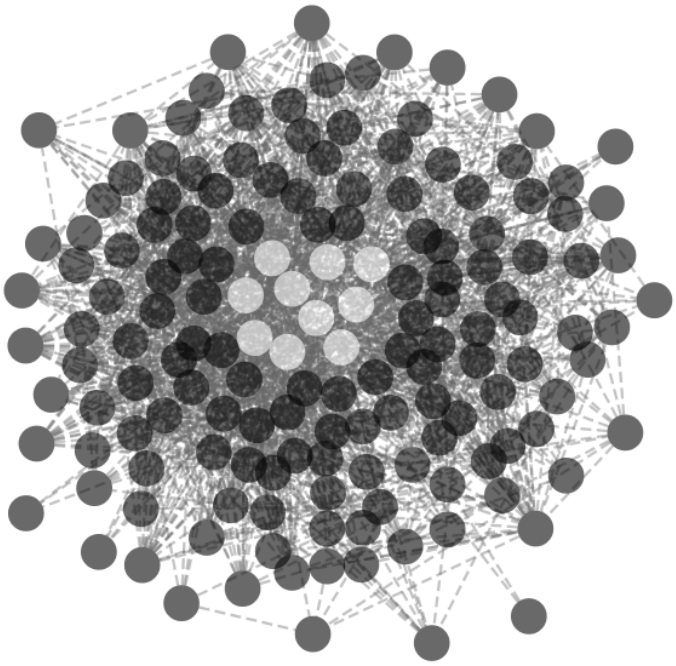

Fig. 3. The resulting network in which the ten white vertices represent the genes with highest degree. The network is plotted using Cytoscape [49].

TABLE I
THE GENE ID, DEGREE AND NAME OF THE TOP 10 RANKED FEATURES.

Rank	Gene ID	Degree	Gene name
1	ENSG00000169035	86	kallikrein-related peptidase 7 (KLK7)
2	ENSG00000051415	68	amylase, alpha 1A, 1B, 1C
3	ENSG00000181649	66	pleckstrin homology-like domain family A, member 2 (phlda2)
4	ENSG00000124562	54	small nuclear ribonucleoprotein polypeptide C
5	ENSG00000117115	52	peptidyl arginine deiminase, type II
6	ENSG00000166483	51	WEE1 homolog (S. pombe)
7	ENSG00000120833	49	suppressor of cytokine signaling 2
8	ENSG00000134532	47	SRY (sex determining region Y)-box 5
9	ENSG00000159658	45	KIAA0494
10	ENSG00000171793	43	CTP synthase

IV. CONCLUSION

In this work, we introduced a new network-based approach to identify biomarker candidates for breast cancer recurrence. Our main contribution is to propose a new workflow that

TABLE II
THE RELATED PATHOPHYSIOLOGICAL PROCESSES OF THE TOP 10 RANKED GENES. THE ASSOCIATED GENE NAMES CAN BE FOUND IN TABLE I.

Rank	Related pathophysiological processes
1	Play a role in the desquamation of the skin, the stratum corneum. Up-regulated in by estrogens and glucocorticoids.
2	involved in starch and sucrose metabolism, dalivary secretion, carbohydrate digestion and absorption.
3	compete to bind phosphoinositides in the membrane lipids with a broad specificity in various biological processes
4	associated with snRNP U1. and involved in the splicesome pathway
5	catalyzes the deimination of arginine residues of proteins
6	participate in the cell cycle with an increased synthesis during S and G2 phases
7	a negative regulator in the growth hormone/IGF1 signaling pathway
8	binds specifically to the DNA sequence 5'-AACAAT-3', overexpressed in glioma and prostate tumor
9	involved in the splicing of eukaryotic pre-mRNAs
10	catalyzes the ATP-dependent amination of UTP to CTP, play a role in pyrimidine metabolism

includes the filtering of relevant genes and inferring the network based on an information entropy measure. Our three-step strategy selected in the first step relevant features using the information gain. In the second step we inferred a new type of a network where the number of edges of each vertex (i.e., degree) represents the predictive ability of the underlying feature. Finally we used the DAVID and KEGG databases to verify and interpret top ranked genes.

Using our breast cancer microarray dataset from the GEO database we could identify a set of known and unexpected genes associated with breast cancer and other pathophysiological processes. The proposed generic method can also be applied to other biomedical questions (e.g., other diseases) or types of data such as metabolic datasets.

ACKNOWLEDGMENT

This work was supported by the Austrian Genome Research Program GEN-AU (Bioinformatics Integration Network, BIN III) and partly by the Tiroler Wissenschaftsfond.

REFERENCES

[1] A. A. Amini and M. J. Wainwright, "High-dimensional analysis of semidefinite relaxations for sparse," *The Annals of Statistics*, vol. 37, p. 28772921, 2009.

[2] M. Netzer, G. Millonig, M. Osl, B. Pfeifer, S. Praun, J. Villinger, W. Vogel, and C. Baumgartner, "A new ensemble-based algorithm for identifying breath gas marker candidates in liver disease using ion molecule reaction mass spectrometry." *Bioinformatics*, vol. 25, no. 7, pp. 941–947, Apr 2009.

[3] I. Inza, P. Larraaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in dna microarray domains." *Artif Intell Med*, vol. 31, pp. 91–103, Jun 2004.

[4] C. Baumgartner and A. Graber, *Successes and new directions in data mining*. Idea Group Inc., 2007, ch. 7, Data mining and knowledge discovery in metabolomics, pp. 141–166.

[5] F. Emmert-Streib and M. Dehmer, "Networks for systems biology: conceptual connection of data and function." *IET Syst Biol*, vol. 5, no. 3, pp. 185–207, May 2011.

[6] M. Dehmer, N. Barbarini, K. Varmuza, and A. Graber, "Novel topological descriptors for analyzing biological networks." *BMC Struct Biol*, vol. 10, p. 18, 2010.

[7] S. Bergmann, J. Ihmels, and N. Barkai, "Similarities and differences in genome-wide expression data of six organisms." *PLoS Biol*, vol. 2, no. 1, p. E9, Jan 2004.

[8] B. Weigelt, J. L. Peterse, and L. J. van 't Veer, "Breast cancer metastasis: markers and models." *Nat Rev Cancer*, vol. 5, no. 8, pp. 591–602, Aug 2005.

[9] K. A. Hoadley, V. J. Weigman, C. Fan, L. R. Sawyer, X. He, M. A. Troester, C. I. Sartor, T. Rieger-House, P. S. Bernard, L. A. Carey, and C. M. Perou, "Egfr associated expression profiles vary with breast tumor subtype." *BMC Genomics*, vol. 8, p. 258, 2007.

[10] T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva, "Ncbi geo: archive for functional genomics data sets–10 years on." *Nucleic Acids Res*, vol. 39, no. Database issue, pp. D1005–D1010, Jan 2011.

[11] M. Netzer, K. M. Weinberger, M. Handler, M. Seger, X. Fang, K. G. Kugler, A. Graber, and C. Baumgartner, "A computational strategy for the identification and kinetic analysis of metabolic biomarkers: A pilot study for profiling human response to physical exercise," *J Clin Bioinforma*, vol. 1, no. 34, 2011.

[12] C. Baumgartner, G. D. Lewis, M. Netzer, B. Pfeifer, and R. E. Gerszten, "A new data mining approach for profiling and categorizing kinetic patterns of metabolic biomarkers after myocardial injury." *Bioinformatics*, vol. 26, no. 14, pp. 1745–1751, Jul 2010.

[13] J. R. Quinlan, "Induction of decision trees," *Mach Learn*, vol. 1, no. 1, pp. 81–106, 1986.

[14] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, University of Waikato, 1999.

[15] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, "David: Database for annotation, visualization, and integrated discovery." *Genome Biol*, vol. 4, no. 5, p. P3, 2003.

[16] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes and genomes." *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan 2000.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, 2009.

[18] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[19] R. Gentleman, E. Whalen, W. Huber, and S. Falcon, *graph: graph: A package to handle graph data structures*, 2009, r package version 1.26.0.

[20] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006. [Online]. Available: http://igraph.sf.net

[21] L. A. J. Mueller, K. G. Kugler, A. Dander, A. Graber, and M. Dehmer, "QuACN: an R package for analyzing complex biological networks quantitatively," *Bioinformatics*, vol. 27, no. 1, pp. 140–141, Jan 2011.

[22] D. Beisser, G. Klau, T. Dandekar, T. Mueller, and M. Dittrich, "Bionet an r-package for the functional analysis of biological networks," *Bioinformatics*, vol. 26, pp. 1129–1130, 2009.

[23] F. Fritzsche, T. Gansukh, C. A. Borgoo, M. Burkhardt, S. Pahl, E. Mayordomo, K.-J. Winzer, W. Weichert, C. Denkert, K. Jung, C. Stephan, M. Dietel, E. P. Diamandis, E. Dahl, and G. Kristiansen, "Expression of human kallikrein 14 (klk14) in breast cancer is associated with higher tumour grades and positive nodal status." *Br J Cancer*, vol. 94, no. 4, pp. 540–547, Feb 2006.

[24] A. Lundstrm and T. Egelrud, "Stratum corneum chymotryptic enzyme: a proteinase which may be generally present in the stratum corneum and with a possible involvement in desquamation." *Acta Derm Venereol*, vol. 71, no. 6, pp. 471–474, 1991.

[25] H. Tanimoto, L. J. Underwood, K. Shigemasa, M. S. Y. Yan, J. Clarke, T. H. Parmley, and T. J. O'Brien, "The stratum corneum chymotryptic enzyme that mediates shedding and desquamation of skin cells is highly overexpressed in ovarian tumor cells." *Cancer*, vol. 86, no. 10, pp. 2074–2082, Nov 1999.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-190-8

[26] G. M. Yousef, A. Scorilas, A. Magklara, A. Soosaipillai, and E. P. Diamandis, "The klk7 (prss6) gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family - genomic characterization, mapping, tissue expression and hormonal regulation." *Gene*, vol. 254, no. 1-2, pp. 119–128, Aug 2000.

[27] M. Talieri, E. P. Diamandis, D. Gourgiotis, K. Mathioudaki, and A. Scorilas, "Expression analysis of the human kallikrein 7 (klk7) in breast tumors: a new potential biomarker for prognosis of breast carcinoma." *Thromb Haemost*, vol. 91, no. 1, pp. 180–186, Jan 2004.

[28] G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, and A. C. Stone, "Diet and the evolution of human amylase gene copy number variation." *Nat Genet*, vol. 39, no. 10, pp. 1256–1260, Oct 2007.

[29] U. M. Nater, N. Rohleder, J. Gaab, S. Berger, A. Jud, C. Kirschbaum, and U. Ehlert, "Human salivary alpha-amylase reactivity in a psychosocial stress paradigm." *Int J Psychophysiol*, vol. 55, no. 3, pp. 333–342, Mar 2005.

[30] M. Furusawa, T. Taira, S. M. M. Iguchi-Ariga, and H. Ariga, "Amy-1 interacts with s-akap84 and akap95 in the cytoplasm and the nucleus, respectively, and inhibits camp-dependent protein kinase activity by preventing binding of its catalytic subunit to a-kinase-anchoring protein (akap) complex." *J Biol Chem*, vol. 277, no. 52, pp. 50 885–50 892, Dec 2002.

[31] Y. Muto, D. P. Krummel, C. Oubridge, H. Hernandez, C. V. Robinson, D. Neuhaus, and K. Nagai, "The structure and biochemical properties of the human spliceosomal protein u1c." *J Mol Biol*, vol. 341, no. 1, pp. 185–198, Jul 2004.

[32] R. Raijmakers, A. J. W. Zendman, W. V. Egberts, E. R. Vossenaar, J. Raats, C. Soede-Huijbregts, F. P. J. T. Rutjes, P. A. van Veelen, J. W. Drijfhout, and G. J. M. Pruijn, "Methylation of arginine residues interferes with citrullination by peptidylarginine deiminases in vitro." *J Mol Biol*, vol. 367, no. 4, pp. 1118–1129, Apr 2007. [Online]. Available: http://dx.doi.org/10.1016/j.jmb.2007.01.054

[33] E. R. Vossenaar, A. J. W. Zendman, W. J. van Venrooij, and G. J. M. Pruijn, "Pad, a growing family of citrullinating enzymes: genes, features and involvement in disease." *Bioessays*, vol. 25, no. 11, pp. 1106–1118, Nov 2003.

[34] E. R. Vossenaar, T. R. D. Radstake, A. van der Heijden, M. A. M. van Mansum, C. Dieteren, D.-J. de Rooij, P. Barrera, A. J. W. Zendman, and W. J. van Venrooij, "Expression and activity of citrullinating peptidylarginine deiminase enzymes in monocytes and macrophages." *Ann Rheum Dis*, vol. 63, no. 4, pp. 373–381, Apr 2004.

[35] H. J. Lee, M. Joo, R. Abdolrasulnia, D. G. Young, I. Choi, L. B. Ware, T. S. Blackwell, and B. W. Christman, "Peptidylarginine deiminase 2 suppresses inhibitory kappab kinase activity in lipopolysaccharide-stimulated raw 264.7 macrophages." *J Biol Chem*, vol. 285, no. 51, pp. 39 655–39 662, Dec 2010.

[36] B. D. Cherrington, E. Morency, A. M. Struble, S. A. Coonrod, and J. J. Wakshlag, "Potential role for peptidylarginine deiminase 2 (pad2) in citrullination of canine mammary epithelial cell histones." *PLoS One*, vol. 5, no. 7, p. e11768, 2010.

[37] T. Yoshida, S. Tanaka, A. Mogi, Y. Shitara, and H. Kuwano, "The clinical significance of cyclin b1 and wee1 expression in non-small-cell lung cancer." *Ann Oncol*, vol. 15, no. 2, pp. 252–256, Feb 2004.

[38] S. Backert, M. Gelos, U. Kobalz, M. L. Hanski, C. Bhm, B. Mann, N. Lvin, A. Gratchev, U. Mansmann, M. P. Moyer, E. O. Riecken, and C. Hanski, "Differential gene expression in colon carcinoma cells and tissues detected with a cdna array." *Int J Cancer*, vol. 82, no. 6, pp. 868–874, Sep 1999.

[39] H. Wang, M. Huang, D. Y. Zhang, and F. Zhang, "Global profiling of signaling networks: study of breast cancer stem cells and potential regulation." *Oncologist*, vol. 16, no. 7, pp. 966–979, 2011.

[40] E. Rico-Bautista, A. Flores-Morales, and L. Fernndez-Prez, "Suppressor of cytokine signaling (socs) 2, a protein with multiple functions." *Cytokine Growth Factor Rev*, vol. 17, no. 6, pp. 431–439, Dec 2006.

[41] K. D. Sutherland, G. J. Lindeman, D. Y. H. Choong, S. Wittlin, L. Brentzell, W. Phillips, I. G. Campbell, and J. E. Visvader, "Differential hypermethylation of socs genes in ovarian and breast carcinomas." *Oncogene*, vol. 23, no. 46, pp. 7726–7733, Oct 2004.

[42] Y. Kamachi, M. Uchikawa, and H. Kondoh, "Pairing sox off: with partners in the regulation of embryonic development." *Trends Genet*, vol. 16, no. 4, pp. 182–187, Apr 2000.

[43] L. H. Pevny and R. Lovell-Badge, "Sox genes find their feet." *Curr Opin Genet Dev*, vol. 7, no. 3, pp. 338–344, Jun 1997.

[44] S. Ma, Y. P. Chan, B. Woolcock, L. Hu, K. Y. Wong, M. T. Ling, T. Bainbridge, D. Webber, T. H. M. Chan, X.-Y. Guan, W. Lam, J. Vielkind, and K. W. Chan, "Dna fingerprinting tags novel altered chromosomal regions and identifies the involvement of sox5 in the progression of prostate cancer." *Int J Cancer*, vol. 124, no. 10, pp. 2323–2332, May 2009.

[45] E. Tchougounova, Y. Jiang, D. Brster, N. Lindberg, M. Kastemar, A. Asplund, B. Westermark, and L. Uhrbom, "Sox5 can suppress platelet-derived growth factor b-induced glioma development in ink4a-deficient mice through induction of acute cellular senescence." *Oncogene*, vol. 28, no. 12, pp. 1537–1548, Mar 2009.

[46] K. Szafranski, S. Schindler, S. Taudien, M. Hiller, K. Huse, N. Jahn, S. Schreiber, R. Backofen, and M. Platzer, "Violating the splicing rules: Tg dinucleotides function as alternative 3' splice sites in u2-dependent introns." *Genome Biol*, vol. 8, no. 8, p. R154, 2007.

[47] K. J. M. Schimmel, H. Gelderblom, and H. J. Guchelaar, "Cyclopentenyl cytosine (cpec): an overview of its in vitro and in vivo activity." *Curr Cancer Drug Targets*, vol. 7, no. 5, pp. 504–509, Aug 2007.

[48] M. Huang, P. Whang, P. Lewicki, and B. S. Mitchell, "Cyclopentenyl cytosine induces senescence in breast cancer cells through the nucleolar stress response and activation of p53." *Mol Pharmacol*, vol. 80, no. 1, pp. 40–48, Jul 2011.

[49] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader, "Integration of biological networks and gene expression data using cytoscape." *Nat Protoc*, vol. 2, no. 10, pp. 2366–2382, 2007.

Copyright (c) IARIA, 2012. **ISBN: 978-1-61208-190-8**