Coupling Regulatory Networks and Microarays: Revealing Molecular Regulations of Breast Cancer Treatment Responses

Lefteris Koumakis[1], Vassilis Moustakis[1,2], Michalis Zervakis[3], Dimitris Kafetzopoulos[4], and George Potamias[1,*]

[1] Institute of Computer Science, FORTH
{koumakis,potamias}@ics.forth.gr
[2] Department of Production Engineering, Technical Univsrsity of Chania
moustakis@dpem.tuc.gr
[3] Department of Electronic and Computer Engineering, Technical University of Chania
michalis@display.tuc.gr
[4] Institute of Molecular Biology & Biotechnology, FORTH
kafetzo@imbb.forth.gr

Abstract. Moving towards the realization of genomic data in clinical practice, and following an individualized healthcare approach, the function and regulation of genes has to be deciphered and manifested. Two of the most significant forms of molecular data come form microarray gene expression sources, and gene interactions sources – as encoded in Gene Regulatory Networks (GRNs). The usual computational task is the gene selection procedure with the GRNs to be mainly utilized for annotation and enrichment purposes. In this study we present a novel perception of these resources. Initially we locate all functional **path-modules** encoded in GRNs and we try to assess which of them are compatible and match the gene-expression profiles of samples that belong to different phenotypes. The differential power of the selected path-modules is computed and their biological relevance is assessed. The whole approach was applied on a set of microarray studies with the target of revealing putative regulatory mechanisms that govern and putatively guide the treatment responses of BRCA patients. The results were quite satisfactory according to their biological and clinical relevance.

1 Introduction

Advances in highthroughput technologies (e.g., microarrays, SNP mapping and copy-number variations etc.) have put the foundation stones for the vision of contemporary *personalized medicine*. On the other hand, *systems biology* follows a 'holistic' approach in order to explore and study the behavior of biological components, trying to uncover and model cell interaction events and, in a way, reproduce the function of organisms. In such a context, we need computational methods that not only combine information and data from dispersed and heterogeneous data sources but also distil the knowledge and provide a systematic, genome-scale view of biology [1]. The

I. Maglogiannis, V. Plagianakos, and I. Vlahavas (Eds.): SETN 2012, LNAI 7297, pp. 239–246, 2012.

advantage of this approach is that it can identify emergent properties of the underlying molecular system as a 'whole' – an endeavor of limited success if targeted genes, reactions or even molecular pathways are studied in isolation. Genes and proteins do not function independently, but participate in complex, interconnected pathways and *gene regulatory networks* (GRN) that govern the function of cells, tissues, organs and organisms seen as functional biological systems and not just as a 'bag of molecules' [2]. At the same time, most of the known and established GNRs are based on laborious wet-lab experiments that make their generation and validation a rather difficult as well as time- and cost-consuming task. A major challenge is to accelerate our understanding of the *molecular mechanisms* of these variations and to produce targeted individualized therapies. Faced with such a challenge we devised and present an integrated methodology that 'amalgamates' knowledge and data from both GRNs and MA gene-expression sources. A preliminary realization of the methodology is implemented in a system called *MinePath*. MinePath aims to uncover potential gene-regulatory 'fingerprints' and mechanisms that underlie and govern the molecular profiles of diseases.

2 Microarrays and GRNs: Techniques and Limitations

MA experiments involve more variables (genes) than samples (patients). This fact, leads to results with poor biological significance. To remedy this there is an open debate whether we should concentrate on gathering more data or on building new algorithms. Simon et al., [3], published a very strict criticism on the common pitfalls of microarray data mining, while in [4] the authors comment about the bias in the gene selection procedure.

In the light of the these observations, and in order to overcome the posted limitations we have to view MA based gene-expression profiles *just as an instance* of biological information, strongly connected - rather than isolated, from other sources of related biological knowledge, e.g., GRNs. GRNs are network structures that depict the interaction of DNA segments during the transcription of the genes into mRNA. From a computational point of view, GRNs can be conceived as analogue biochemical computers that regulate the level of expression of target genes [5]. The network by itself acts as a mechanism that determines cellular behavior where the nodes are genes and edges are functions that represent the molecular reactions between the nodes. These functions can be perceived as Boolean functions, where nodes have only two possible states ("on" and "off"), and the whole network represented as a simple *directed graph* [6]. It is indicative that most of the relations in known and established GRNs have been derived from laborious and extensive laboratory experiments and careful study of the existing biochemical literature. Thus GRNs are far from complete.

A number of different methodologies have been proposed to help overcome this, and help to identify useful biological knowledge from GRNs with very few of them to be considered superior to the others - mainly because of the intrinsically noisy property of the data, 'the curse of dimensionality', and the unknown 'true' underlying networks. In this paper we present a novel methodology that couples microarray gene-expression profiles wit GRNs. The methodology aims towards the identification of differentially expressed functional GRN ***path-modules***.

3 MinePath: Revealing Phenotype-Specific Regulation

Online public repositories contain a variety of information that includes not only the GRNs per se but links and rich annotations for the respective nodes (genes) and edges (reactions). In the current study we utilize the KEGG pathways repository[1]. KEGG provides a format representation standardized by its own markup description language (KGML[2]). A preliminary implementation of our methodology is implemented in a system called *MinePath*, and it unfolds into four phases. **(i) Pathway decomposition.** MinePath relies on a novel approach for GRN processing that takes into account all possible functional interactions of the network, i.e., the network's functional sub-paths. Different GRNs are downloaded from the KEGG repository. With an XML parser (operated on KEGG's KGML representation scheme of GRNs) the network is **decomposed** into its all-possible sub-paths (see Figure 1.a for an exemplification). After parsing a set of (targeted) GRNs, the decomposed sub-paths are stored in a database that acts as a repository for future reference. As the database repository could contain sub-paths from a variety of different GRNs we may combine different molecular pathways and networks – a major need for molecular biology and a big challenge for systems biology and contemporary bioinformatics research.

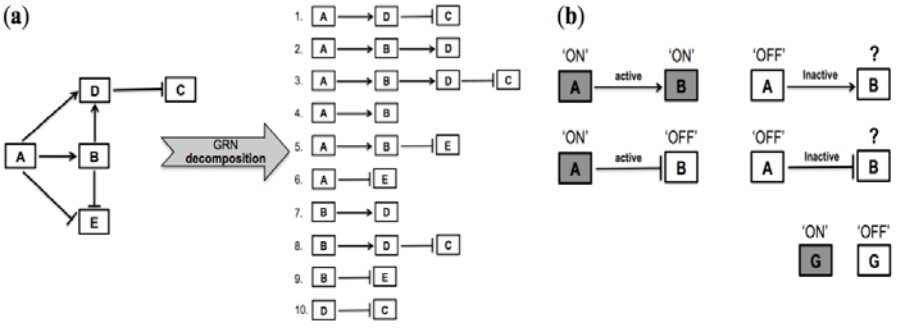

Fig. 1. (a) GRN decomposition: the artificial GRN (left) is decomposed into its all-possible sub-paths (10; right). (**b**) Functional path-modules: a reaction is considered as 'active' only-and-only-if its starting gene is active (e.g., 'ON' → 'ON' or, 'ON' —| 'OFF' for active activation/expression or, inhibition reactions, respectively; otherwise is considered as inactive, 'OFF' → ? or, 'OFF' —| ?, with the state of the regulated gene on the right of the reaction being undetermined).

(**ii**) Inference **of functional path-modules.** Each GRN sub-path is interpreted according to Kauffman's principles and semantics [6]: ① the network is a directed graph with genes (inputs and outputs) being the graph nodes and the edges between them representing the *causal* links between them, i.e., the *regulatory* reactions; ② each node can be in one of the two states, 'ON', the gene is expressed or up-regulated (i.e., the respective substance being present) or, 'OFF', the gene is not-expressed or

[1] KEGG: Kyoto Encyclopedia of Genes and Genomes; `http://www.genome.jp/kegg/`
[2] KGML (KEGG Markup Language); `http://www.genome.jp/kegg/xml/`

down-regulated; and ③ time is viewed as proceeding in discrete steps - at each step the new state of a node is a Boolean function of the prior states of the nodes with arrows pointing towards it. In order to cope with and reveal functional regulatory mechanisms we impose over the formed sub-paths the following requirement: for a sub-path to be considered as functional it should be 'active' during the GRN regulation process - in other words we assume that all genes in a sub-path are functional. For example consider the reaction A → B (see Figure 1.b), if A is 'ON' then the activation/expression ('→') regulatory reaction is active, resulting into the activation/expression of gene B ('ON') – the same holds for an inhibition (—|) reaction. In the case that gene A is 'OFF' then the reaction is considered as inactive with the state of the regulated gene B to remain undetermined ('?'). Under this assumption, a **path-module** is just a sub-path (atomic or more complex) for which all its reactions are considered as active. So, the state of all genes engaged in a path-module that forms an *ordered regulation pattern*, e.g., the pattern of the complex regulatory mechanism A → D —| C is <'ON', 'ON', 'OFF'>.

(iii) Matching gene-expression profiles and path-modules. The next step is to locate microarray experiments and respective gene-expression data for which we expect (suspect) the targeted GRNs play an important role - for example, the cell-cycle and apoptosis GRNs play an important role in tumorgenesis and cancer progression. The samples of a binary transformed (discretized) gene-expression matrix are matched against functional path-modules of target GRNs. (retrieved form the described repository). We follow an information-theoretic gene-expression discretization process (detailed in [7]).

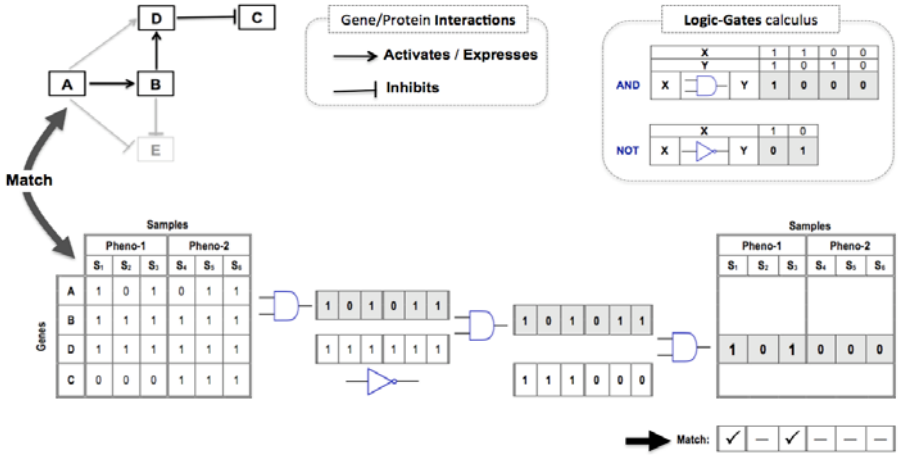

Fig. 2. Matching gene-expression sample profiles with GRN functional path-modules: a logic-gates approach

As an example, assume the gene-expression binary profiles of six artificial samples for genes A, B, D and C - with '1' to denote 'ON' and '0' to denote 'OFF' - three of them are assigned to phenotype-1 (S_1, S_2, and S_3) and the other three to phenotype-2 (S_4, S_5, and S_6) – refer to Figure 2. Furthermore, assume the artificial GRN shown in

the lower left part of Figure 2, and its sub-path A → B → D —| C (in bold). We follow a **logic-gates** process that aims to match the path-module instance of the sub-path with the respective samples' binary instances. The process results into the formation of an ordered pattern that indicate the samples for which the target sub-path is consistent with ('1's) or not ('0's), i.e., the respective path-module A='ON' → B='ON' → D='ON' —| C='OFF' is active.

(iv) The differential-power of path-modules. Note that for the finally inferred pattern of Figure 2, <1,0,1,0,0,0>, value '1' occurs in positions one and three which means that the examined path-module is active for samples one and three; in all other samples it is inactive ('0'). As samples one and three belong to phenotype-1, the target path-module matches 2 out of 3 phenotype-1 samples, and zero phenotype-2 samples. In general, assume that there are S_1 and S_2 samples that belong to phenotype-1 and phenotyp2, respectively, and that path-module P_i matches $S_{i;1}$ and $S_{i;2}$ samples form phenotype-1 and phenotype-2, respectively. Formula 1, computes the **differential power** of a path-module with respect to the two phenotypic classes;

$$\frac{S_{i;1}}{S_1} - \frac{S_{i;2}}{S_2} \tag{1}$$

The formula posses a *polarity* characteristic according the class phenotype: positive for class S_1 and negative for class S_2; e.g., for the above example, the differential power of path-module A='ON' → B='ON' → D='ON' —| C='OFF' is (2/3) – 0 = 0.67, and as it positive it is interpreted and considered as a regulation mechanism that governs phenotype-1.

4 The Regulation of Breast Cancer Treatment Response

Most of breast cancer (BRCA) cases are estrogen responsive, a series of growth-promoting pathways are activated, for example, the estrogen receptor (ER) related ErbB signaling GRN. In an effort to reveal the underlying regulatory mechanisms that govern BRCA patients' treatment responses we applied the presented MinePath methodology on a set of four independent gene-expression studies targeting the ER phenotypic status of the respective patients, i.e., ER+ (ER positive) vs. ER- (ER negative). The details of the gene-expression data from the four studies are: GSE2034 (the GEO-Gene Expression Omnibus[3] study code), 286 patients [8]; GSE2990, 183 patients [9]; GSE3494, 247 patients [10]; and GSE7390, 198 patients [11]. We targeted 14 pathways, all of which are engaged within the 'Pathways in Cancer' integrated pathway (KEGG code: hsa05200), e.g., ErbB (hsa04012), MAPK (hsa04010), mTOR (hsa04150) etc. After applying the aforementioned matching process we selected the 100 path-modules with the highest differential power – 50 for ER+ and 50 for the ER- phenotypes, respectively. Inspecting the results we observed that the pathway that engage a significantly larger, with respect to all other targeted pathways, number of the selected sub-paths is the ErbB signaling pathway. Figure 3 shows the ErbB signaling pathway as colored with the help of the KEGG Mapper/Search&Color[4] tool (symbols and coloring scheme are shown at the bottom

[3] http://www.ncbi.nlm.nih.gov/geo/
[4] http://www.genome.jp/kegg/mapper.html

of the figure). Note the two different functional path *cascades* for the ER+ (black arrows) and ER- (grey arrows) phenotypes, respectively. Both have extra-cellular origins:

❶ The ER- path originates from TGFα (transforming growth factor, alpha), AR (amphiregulin), BTC (betacellulin), and EPR (epiregulin) epidermal growth factors that activate both **ErbB-1** and **ErbB-2** EGF-receptors; then, the two receptors initiate the path GRB2 → GAB1 → **PI3K → PKB/Akt** that guides to the activation of mTOR that activates p70S6K which signals "protein synthesis", and inhibits BAD which signals "cell survival";

❷ The ER+ path originates from the extra-cellular NRG1, NRG2 (neuregulin1,2) growth factors that activate **ErbB-3** and **ErbB-4** viral oncogenes followed by the **PI3K → PKB/Akt** activation reaction which is also part of the ER- path. But now, PKB/Akt acts just as an inhibitor of GSK-3 and blocking of "Metabolism". Moreover, PKB/Akt activates mTOR, which now acts as an inhibitor of EIF-4EBP with the result of blocking "protein synthesis". According to the recent biomedical literature the aforementioned results are quite relevant to the estrogen-receptor status - we focused our exploration on the mechanisms underlying the resistance to pure estrogen antagonists (e.g., fulvestrant5).

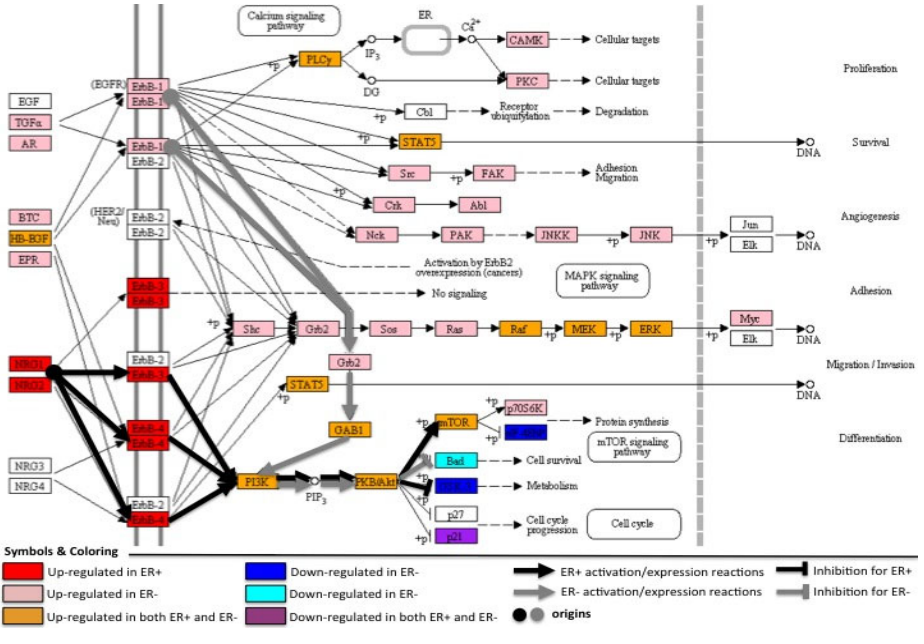

Fig. 3. Regulation of ER+ and ER- phenotypes in the ErbB signaling GRN

[5] Fulvestrant (Faslodex, AstraZeneca) is a drug treatment of hormone receptor-positive metastatic BRCA in postmenopausal women with disease progression following anti-estrogen therapy.

Recent studies show the significant role of both ErbB3 and ErbB4 as alternative targets for the treatment of BRCA patients; as Sutherland notes in [12]: "… *recent studies now implicates the other two ErbB family members, ErbB-3 and -4. Exposure of ER+ breast cancer cells to the pure antiestrogen, fulvestrant, increased levels of ErbB-3 or ErbB-4 and sensitivity to the growth-stimulatory effects of heregulin β1, a potent ligand for these receptors. Thus, the initial growth inhibitory effects of fulvestrant appear compromised by cellular plasticity that allows rapid compensatory growth stimulation via ErbB-3/4 …*"; In addition, Hutcheson et al., [13], investigated whether induction of ErbB3 and/or ErbB4 may provide an alternative resistance mechanism to antihormonal action - their conclusion is that fulvestrant treatment is sensitive to the actions of the ErbB3/4 ligand HRGb1 (NRG1) with enhanced ErbB3/4-driven signaling activity, and significant increases in cell proliferation; the same results are also reported in other relevant studies related to the treatment of BRCA patients [14, 15].

5 Conclusions

We have presented an integrated methodology for the coupling of both GRNs and MA gene expression profiles. In the heart of the methodology are the decomposition of GRNs into functional sub-paths, and the matching of these sub-paths with samples' gene expression profiles, in order to compute their differential power with target phenotypic classes. The whole methodology is preliminary implemented in a system called MinePath. MinePath was applied on a set of four gene-expression studies with the target of identifying putative mechanisms that underlie and govern the treatment response of BRCA patients according to their ER-status profiles. Results were quite indicative and strongly supported by the relevant biomedical literature. Our on-going work and future R&D plans include: (a) further experimentation with various real-world microarray studies and different GRNs; (c) elaboration on more sophisticated path/gene-expression profile matching formulas and operations; (d) incorporation of different gene coding schemes in order to cope with microarray experiments from different platforms and nomenclatures; (e) incorporation of a GRN visualization component, and (e) porting of the whole methodology in a scientific workflow environment enabled by the development of respective Web-Services.

References

1. Ideker, T., Galitski, T., Hood, L.: A new approach to decoding life: systems biology. Annual Review of Genomics and Human Genetics 2, 343–372 (2001)
2. Collins, F.S., Eric, D., Green, E.D., Guttmacher, A.E., Mark, S., Guyer, M.S.: A vision for the future of genomics research. Nature 422, 835–847 (2003)
3. Simon, R., Radmacher, M.D., Dobbin, K., McShane, L.M.: Pitfalls in the Use of DNA Microarray Data for Diagnostic Classification. Journal of the National Cancer Institute 95(1), 14–18 (2003)
4. Ambroise, C., McLachlan, G.J.: Selection bias in gene extraction on the basis of microarray gene-expression data. PNAS 99(10), 6562–6566 (2002)

5. Arkin, A., Ross, J.: Computational functions in biochemical reaction networks. Biophys. J. 67(2), 560–578 (1994)
6. Kauffman, S.A.: The Origins of Order: Self-Organization and Selection in Evolution. Oxford Univ. Press, New York (1993)
7. Potamias, G., Koumakis, L., Moustakis, V.: Gene Selection via Discretized Gene-Expression Profiles and Greedy Feature-Elimination. In: Vouros, G.A., Panayiotopoulos, T. (eds.) SETN 2004. LNCS (LNAI), vol. 3025, pp. 256–266. Springer, Heidelberg (2004)
8. Wang, Y., et al.: Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet. 365(9460), 671–679 (2005)
9. Sotiriou, C., et al.: Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. 98(4), 262–272 (2006)
10. Miller, L.D., et al.: An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. PNAS 102(38), 13550–13555 (2005)
11. Desmedt, C., et al.: Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin. Cancer Res. 13(11), 3207–3214 (2007)
12. Sutherland, R.L.: Endocrine resistance in breast cancer: new roles for ErbB3 and ErbB4. Breast Cancer Research 13(3), 106 (2011)
13. Hutcheson, I.R., et al.: Heregulin beta1 drives gefitinib-resistant growth and invasion in tamoxifen-resistant MCF-7 breast cancer cells. Breast Cancer Research 9(4), R50 (2007)
14. Zhu, Y., Sullivan, L.L., Nair, S.S., Williams, C.C., Pandey, A.K., Marrero, L., Vadlamudi, R.K., Jones, F.E., et al.: Coregulation of estrogen receptor by ERBB4/HER4 establishes a growth-promoting autocrine signal in breast tumor cells. Cancer Research 66(16), 7991–7998 (2006)
15. Sonne-Hansen, K., et al.: Breast cancer cells can switch between estrogen receptor alpha and ErbB signaling and combined treatment against both signaling pathways postpones development of resistance. Breast Cancer Research and Treatment 121(3), 601–613 (2010)