A System Biology Approach for the Steady-State Analysis of Gene Signaling Networks

Purvesh Khatri[1], Sorin Draghici[1,⋆], Adi L. Tarca[1,2], Sonia S. Hassan[2], and Roberto Romero[2]

[1] Department of Computer Science, Wayne State University
[2] Perinatology Research Branch, NIH/NICHD, Detroit, MI 48201
purvesh@cs.wayne.edu, sorin@wayne.edu,
atarca@med.wayne.edu, shassan@med.wayne.edu,
prbchiefstaff@med.wayne.edu

Abstract. The existing approaches used to identify the relevant pathways in a given condition do not consider a number of important biological factors such as magnitude of each gene's expression change, their position and interactions in the given pathways, etc. Recently, an impact analysis approach was proposed that considers these crucial biological factors to analyze regulatory pathways at systems biology level. This approach calculates perturbations induced by each gene in a pathway, and propagates them through the entire pathway to compute an impact factor for the given pathway. Here we propose an alternative approach that uses a linear system to compute the impact factor. Our proposed approach eliminates the possible stability problems when the perturbations are propagated through a pathway that contains positive feedback loops. Additionally, the proposed approach is able to consider the type of genes when calculating the impact factors.

1 Introduction

While high-throughput life science technologies have enabled the collection of large amount of data, they have also posed challenges related to the extraction of knowledge from these data. For instance, the typical result of a microarray experiment is a list of differentially expressed (DE) genes that quantitatively reflect the changes in gene activity in response to a particular treatment, or in a given condition. The challenge common to all experiments is to translate such lists of DE genes into a better understanding of the underlying phenomenon. An automated Gene Ontology (GO) based approach has been proposed in order to help in this process [1,2]. This approach uses an over-representation analysis (ORA) of the list of DE genes in order to identify the GO categories that are significantly over- or under-represented in a given condition. This type of analysis has been very successful to the point of becoming a *de facto* standard in the analysis of microarray data [3]. A more recent approach considers the distribution

[⋆] To whom the correspondence should be addressed.

L. Rueda, D. Mery, and J. Kittler (Eds.): CIARP 2007, LNCS 4756, pp. 32–41, 2007.

of the pathway genes in the entire list of genes and performs a functional class scoring (FCS) which also allows adjustments for gene correlations [4,5,6,7,8].

Both ORA and FCS techniques currently used are limited by the fact that each functional category is analyzed independently without a unifying analysis at a pathway or system level [8]. This approach is not well suited for a systems biology approach that aims to account for system level dependencies and interactions, as well as identify perturbations and modifications at the pathway or organism level [9]. In particular, all existing ORA and FCS approaches ignore a number of important biological factors including the amount of change in gene expression, the interactions between genes and their positions on a given pathway [10].

Recently, an impact analysis method was proposed that combines these important biological factors with the classical statistical analysis in order to identify the most perturbed signaling pathways in a given condition [10]. An *impact factor* (IF) is calculated for each pathway incorporating parameters such as the normalized fold change of the differentially expressed genes, the statistical significance of the set of pathway genes, and the topology of the pathway.

In this paper, we propose using a different approach to calculate the impact factors. Rather than propagating the perturbation through the pathway in a neural network-like fashion, here we propose to calculate the stable-state values of the perturbations by using a system of simultaneous equations. The main differences occur when pathways includes loops, which is true for most of the known gene signaling pathways. In such cases, in the previously described impact analysis the computation of the gene perturbation factors (PFs) was done through an iterative process. Problems are created by the fact that in the graph that describes the given pathway, multiple paths of different length are usually available to propagate the signal from any one source node to any one destination node. In order to address this, the previous version of the impact analysis approximates the PFs by going around each loop only once. No such approximation is necessary when the pathways are described by a system of simultaneous equations in which the PF of each gene is a function of the PFs of all other genes on the pathway. The previous approach of approximating the PFs by propagating the perturbations from node to node is still used when the system does not have an exact algebraic solution.

2 Impact Analysis

The aim of this approach is to establish a model that accounts for both the statistical significance of the set of genes and the perturbations of the individual genes on each pathway. A variety of models can be proposed here, but Occam's razor suggests to start with the simplest possible model and increase its complexity only if this model fails to capture the complexity of the phenomenon studied. One of the simplest possible models is a linear additive model in which the *impact factor* (IF) of a pathway P_i can be calculated as the sum between a probabilistic term and a perturbation term:

$$IF(P_i) = log\left(\frac{1}{p_i}\right) + \frac{\left|\sum_{g \in P_i} PF(g)\right|}{\overline{|\Delta E|} \cdot N_{de}(P_i)} \tag{1}$$

The first term captures the significance of the given pathway P_i as provided by the classical statistical approaches, where p_i corresponds to the probability of obtaining a value of the statistic used at least as extreme as the one observed when the null hypothesis is true. We would like the IF to be large for severely impacted pathways (small p-values) so the first term uses $1/p_i$ rather than p_i. The log function is necessary to map the exponential scale of the p-values to a linear scale compatible with our intended linear model. The p_i value can be calculated using either an ORA (e.g., z-test [11], contingency tables [12,13], etc.), a FCS approach (e.g., GSEA [6,7]) or other more recent approaches [8,14,15].

The second term in (1) is a functional term that depends on the specific genes that are differentially expressed as well as on the interactions described by the pathway (i.e., its topology). In essence, this term sums up the values of the *perturbation factors* (PF) for all genes g on the given pathway P_i. The perturbation factor of a gene g_i is calculated as follows:

$$PF(g_i) = \alpha(g_i) \cdot \Delta E(g_i) + \sum_{j=1}^{n} \beta_{ji} \cdot \frac{PF(g_j)}{N_{ds}(g_j)} \tag{2}$$

In (2), the first term captures the quantitative information from the gene expression experiment. The factor $\Delta E(g_i)$ represents the signed normalized measured expression change of the gene g_i. The factor $\alpha(g_i)$ is a weight that captures the potential for systemic changes associated with the type of gene g_i. For most genes, α will be 1. However, if the gene is a transcription factor or similar, α can take a larger value set by the user. Thus, the user can divide the genes into various categories and associate different weights to various categories depending on the target organism.

The second term is a sum of the perturbation factors of all the genes g_j on the pathway P_i, normalized by the number of downstream genes of each such gene $N_{ds}(g_j)$, and weighted by a factor β_{ji}, whose absolute value quantifies the strength of the interaction between g_j and g_i. The sign of β reflects the type of interaction: +1 for induction, -1 for repression. Note that β will have non-zero value only for the genes that directly interact with the gene g_i. The second term here is similar to the PageRank index used by Google [16,17,18] only that we weight the downstream instead of the upstream connections (a web page is important if other pages point to it whereas a gene is important if it influences other genes).

Under the null hypothesis which assumes that the list of differentially expressed genes only contains random genes, the likelihood that a pathway has a large impact factor is proportional to the number of such “differentially expressed” genes that fall on the pathway, which in turn is proportional to the size of the pathway. Thus, we need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, $N_{de}(P_i)$. Furthermore, various technologies can yield systematically different estimates of the fold changes. For instance, the fold changes reported by microarrays tend to be compressed with respect to those reported by RT-PCR [19,20]. In order to make the impact factors as independent as possible from the technology, and also comparable between problems, we also divide the second term in (1) by the mean absolute fold change $\overline{|\Delta E|}$, calculated across all differentially expressed genes. Assuming that there are at least some differentially expressed genes anywhere in the data set[1], both $\overline{|\Delta E|}$ and $N_{de}(P_i)$ are different from zero so the second term is properly defined.

Note that (2) essentially describes the perturbation factor PF for a gene g_i as a linear function of the perturbation factors of all genes in a given pathway. In the stable state of the system, all relationships must hold, so the set of all equations defining the impact factors for all genes form a system of simultaneous equations. Equation 2 can be re-written as:

$$PF(g_i) = \alpha(g_i) \cdot \Delta E(g_i) + \beta_{1i} \cdot \frac{PF(g_1)}{N_{ds}(g_1)} + \beta_{2i} \cdot \frac{PF(g_2)}{N_{ds}(g_2)} + \cdots + \beta_{ni} \cdot \frac{PF(g_n)}{N_{ds}(g_n)} \quad (3)$$

Rearranging (3) gives

$$PF(g_i) - \beta_{1i} \cdot \frac{PF(g_1)}{N_{ds}(g_1)} - \beta_{2i} \cdot \frac{PF(g_2)}{N_{ds}(g_2)} - \cdots - \beta_{ni} \cdot \frac{PF(g_n)}{N_{ds}(g_n)} = \alpha(g_i) \cdot \Delta E(g_i) \quad (4)$$

Using (4), a pathway P_i composed of n genes can be described as follows:

$$\begin{pmatrix} 1 - \frac{\beta_{11}}{N_{ds(g_1)}} & -\frac{\beta_{21}}{N_{ds(g_2)}} & \cdots & -\frac{\beta_{n1}}{N_{ds(g_n)}} \\ -\frac{\beta_{12}}{N_{ds(g_1)}} & 1 - \frac{\beta_{22}}{N_{ds(g_2)}} & \cdots & -\frac{\beta_{n2}}{N_{ds(g_n)}} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{\beta_{1n}}{N_{ds(g_1)}} & -\frac{\beta_{2n}}{N_{ds(g_2)}} & \cdots & 1 - \frac{\beta_{nn}}{N_{ds(g_n)}} \end{pmatrix} \begin{pmatrix} PF(g_1) \\ PF(g_2) \\ \cdots \\ PF(g_n) \end{pmatrix} = \begin{pmatrix} \alpha(g_1) \cdot \Delta E(g_1) \\ \alpha(g_2) \cdot \Delta E(g_2) \\ \cdots \\ \alpha(g_n) \cdot \Delta E(g_n) \end{pmatrix}$$

$$\begin{pmatrix} PF(g_1) \\ PF(g_2) \\ \cdots \\ PF(g_n) \end{pmatrix} = \begin{pmatrix} 1 - \frac{\beta_{11}}{N_{ds(g_1)}} & -\frac{\beta_{21}}{N_{ds(g_2)}} & \cdots & -\frac{\beta_{n1}}{N_{ds(g_n)}} \\ -\frac{\beta_{12}}{N_{ds(g_1)}} & 1 - \frac{\beta_{22}}{N_{ds(g_2)}} & \cdots & -\frac{\beta_{n2}}{N_{ds(g_n)}} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{\beta_{1n}}{N_{ds(g_1)}} & -\frac{\beta_{2n}}{N_{ds(g_2)}} & \cdots & 1 - \frac{\beta_{nn}}{N_{ds(g_n)}} \end{pmatrix}^{-1} \begin{pmatrix} \alpha(g_1) \cdot \Delta E(g_1) \\ \alpha(g_2) \cdot \Delta E(g_2) \\ \cdots \\ \alpha(g_n) \cdot \Delta E(g_n) \end{pmatrix}$$

Since the perturbations of the genes are obtained as the solution of a linear system, this approach aims to characterize the steady state of the system rather than rapidly transient states before an equilibrium has been established. Once the perturbation factors of all genes in a given pathway are calculated, (1) is used to calculate the impact factor of each pathway. The impact factor of each pathway is then used as a score to assess the impact of a given gene expression data set on all pathways (the higher the impact factor the more significant the pathway).

For some pathways, the matrix describing the interactions between the genes may be singular. In such cases, the perturbation factors as approximating by propagating the perturbations as previously described [10].

[1] If there are no differentially expressed genes anywhere, the problem of finding the impact on various pathways is meaningless.

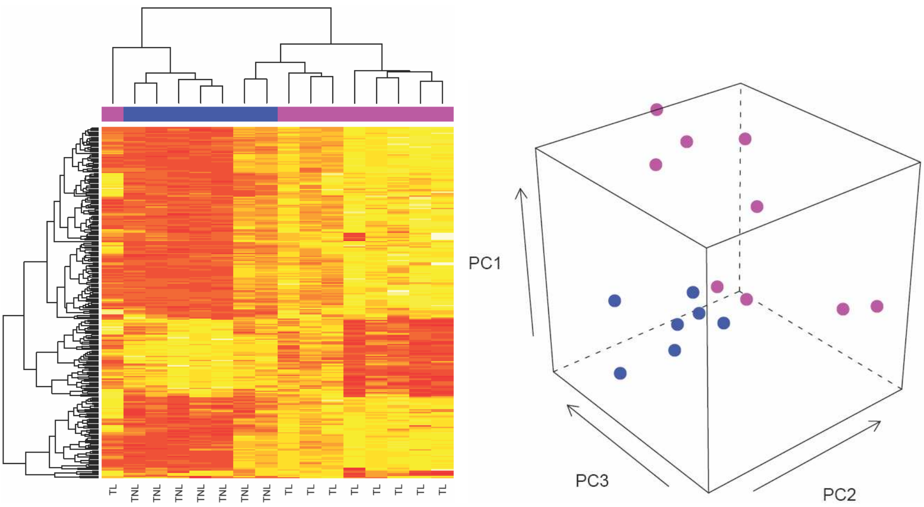
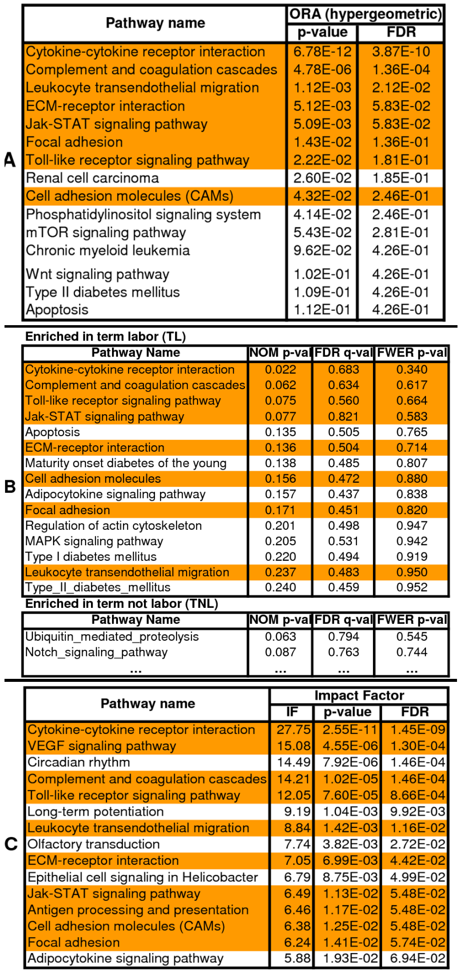


Fig. 1. Cervix data quality assessment. Unsupervised bi-clustering (left panel) of the cervix data using the 263 genes with the largest variability irrespective of the group identifies the two groups, term labor (TL) and term not labor (TNL), in the dataset. Visualization of the 16 samples using PCA (right panel) also shows that the samples are linearly separable using the first 3 principal components.

3 Results and Discussion

We used the proposed pathway impact analysis approach to analyze the differences between cervix tissue in women after term labor ($n = 9$) and those who reached the term without the on-set of labor ($n = 7$). The results obtained from the impact analysis were compared with the results obtained using ORA (hypergeometric p-value) and GSEA. The cervical transcriptome was profiled with Affymetrix HG-U133 Plus 2.0 microarrays. The details of this study and its biological significance are described elsewhere [21,22].

The microarray data was pre-processed using Robust Multi-array Average (RMA) [23]. In order to assess the quality of the microarray data, we used two unsupervised methods. First, we used a bi-clustering procedure [24] that hierarchically partitions the genes and the samples simultaneously. We used 263 genes for clustering that exhibit the largest variability among all 16 samples irrespective of the group they belong to. This approach is unsupervised since it does not use group information. The result of the bi-clustering is shown in Fig.1. As shown in the Fig. 1, the clustering retrieves the two groups of the samples. Next, we applied the principal component analysis (PCA) [24] using all probesets on the HG-U133 plus 2.0 microarray. The results of PCA are shown in Fig. 1. As shown in Fig. 1, the two groups of samples can be separated in the space of the first 3 principal components with a hyperplane. Both types of results indicate that the data is meaningful in terms of differences between classes.



A

Pathway name	ORA (hypergeometric)	
	p-value	FDR
Cytokine-cytokine receptor interaction	6.78E-12	3.87E-10
Complement and coagulation cascades	4.78E-06	1.36E-04
Leukocyte transendothelial migration	1.12E-03	2.12E-02
ECM-receptor interaction	5.12E-03	5.83E-02
Jak-STAT signaling pathway	5.09E-03	5.83E-02
Focal adhesion	1.43E-02	1.36E-01
Toll-like receptor signaling pathway	2.22E-02	1.81E-01
Renal cell carcinoma	2.60E-02	1.85E-01
Cell adhesion molecules (CAMs)	4.32E-02	2.46E-01
Phosphatidylinositol signaling system	4.14E-02	2.46E-01
mTOR signaling pathway	5.43E-02	2.81E-01
Chronic myeloid leukemia	9.62E-02	4.26E-01
Wnt signaling pathway	1.02E-01	4.26E-01
Type II diabetes mellitus	1.09E-01	4.26E-01
Apoptosis	1.12E-01	4.26E-01

B

Enriched in term labor (TL)

Pathway Name	NOM p-val	FDR q-val	FWER p-val
Cytokine-cytokine receptor interaction	0.022	0.683	0.340
Complement and coagulation cascades	0.062	0.634	0.617
Toll-like receptor signaling pathway	0.075	0.560	0.664
Jak-STAT signaling pathway	0.077	0.821	0.583
Apoptosis	0.135	0.505	0.765
ECM-receptor interaction	0.136	0.504	0.714
Maturity onset diabetes of the young	0.138	0.485	0.807
Cell adhesion molecules	0.156	0.472	0.880
Adipocytokine signaling pathway	0.157	0.437	0.838
Focal adhesion	0.171	0.451	0.820
Regulation of actin cytoskeleton	0.201	0.498	0.947
MAPK signaling pathway	0.205	0.531	0.942
Type I diabetes mellitus	0.220	0.494	0.919
Leukocyte transendothelial migration	0.237	0.483	0.950
Type_II_diabetes_mellitus	0.240	0.459	0.952

Enriched in term not labor (TNL)

Pathway Name	NOM p-val	FDR q-val	FWER p-val
Ubiquitin_mediated_proteolysis	0.063	0.794	0.545
Notch_signaling_pathway	0.087	0.763	0.744
...	...	...	...

C

Pathway name	Impact Factor		
	IF	p-value	FDR
Cytokine-cytokine receptor interaction	27.75	2.55E-11	1.45E-09
VEGF signaling pathway	15.08	4.55E-06	1.30E-04
Circadian rhythm	14.49	7.92E-06	1.46E-04
Complement and coagulation cascades	14.21	1.02E-05	1.46E-04
Toll-like receptor signaling pathway	12.05	7.60E-05	8.66E-04
Long-term potentiation	9.19	1.04E-03	9.92E-03
Leukocyte transendothelial migration	8.84	1.42E-03	1.16E-02
Olfactory transduction	7.74	3.82E-03	2.72E-02
ECM-receptor interaction	7.05	6.99E-03	4.42E-02
Epithelial cell signaling in Helicobacter	6.79	8.75E-03	4.99E-02
Jak-STAT signaling pathway	6.49	1.13E-02	5.48E-02
Antigen processing and presentation	6.46	1.17E-02	5.48E-02
Cell adhesion molecules (CAMs)	6.38	1.25E-02	5.48E-02
Focal adhesion	6.24	1.41E-02	5.74E-02
Adipocytokine signaling pathway	5.88	1.93E-02	6.94E-02

Fig. 2. A comparison between the results of the classical approaches (A - hypergeometric, B - GSEA) and the results of the pathway impact analysis (C) for a set of differentially expressed genes in term labor. The pathways marked in red are well supported by the existing literature. After correcting for multiple comparisons, GSEA does not identify any pathway as significantly impacted in this condition at any of the usual significance levels (1%, 5% or 10%). The hypergeometric model identifies *cytokine-cytokine receptor interaction*, *complement and coagulation cascades* and *leukocyte transendothelial migration* as significantly impacted pathways at 5%, and *ECM-receptor interaction* and *Jak-STAT signaling* at 10% after correction for multiple comparisons. In contrast, in addition to the 3 pathways identified by the hypergeometric at 5% significance, the impact analysis also identifies *VEGF signaling*, *toll-like receptor signaling* and *ECM-receptor interaction*. Furthermore, at 10%, the impact analysis identifies *Jak-STAT signaling*, *antigen processing and presentation*, *cell adhesion molecules* and *focal adhesion* as significantly impacted pathways.

Next, we applied a moderated t-test [25] to select a list of DE genes. The p-values obtained from the moderated t-test were corrected using the False Discovery Rate method [23]. We selected 960 genes with corrected p-value less than 0.05 and fold change greater than 2 as DE genes that are both meaningful and verifiable. These 960 genes were used as the input to the ORA analysis using hypergeometric distribution and the impact analysis. GSEA was applied on the normalized expression matrix of all 19,886 unique genes on the array.

Figure 2 shows the comparison between the two classical approaches (hypergeometric and GSEA) and the pathway impact analysis. Note that the figure only shows the top 15 pathways as identified by each approach. For the rest of this section we will discuss the significance of a pathway as indicated by the FDR corrected p-values unless noted otherwise.

When considering the nominal p-value, GSEA finds the *cytokine-cytokine receptor interaction* pathway significant at 5%. However, when the correction for multiple comparisons is applied, GSEA does not find any significantly impacted pathways at any of the usual (1%, 5% or 10%) significance levels.

The hypergeometric model yields 3 pathways significant at the 5% significance level: *cytokine-cytokine receptor interaction*, *complement and coagulation cascades* and *leukocyte transendotheial migration*. These pathways are compatible with our current understanding of the phenomena involved in labor. The *cytokine-cytokine receptor interaction* and *leukocyte transendothelial migration* pathways are both associated with the innate immune system. The involvement of the innate immune system in cervical dilation and remodeling is well established in the literature [26,27]. Also, the *complement and coagulation cascades* include a part of the PLAU signaling and plasmin signaling pathways. There are several studies suggesting the involvement of plasminogen in cervical dilation and remodeling after term labor [28,29]. In particular, the plasminogen activation cascade plays an integral role in the remodeling of extracellular matrices during pregnancy and parturition [28]. In essence, the top 3 pathways identified by the classical ORA approach appear to be relevant.

At the same significance level, the impact analysis agrees on these pathways, but also identifies 7 additional pathways. Among these, *VEGF signaling*, *toll-like receptor signaling* and *ECM-receptor interaction* also appear to be very relevant. In fact, 2 of these 3 pathways point in the same direction: *toll-like receptor signaling* is another pathway associated with the innate immune system while *ECM-receptor interaction* describes the interactions between trans-membrane proteins and the extra-cellular matrix, already known to be heavily remodeled during pregnancy [30,31]. The remaining pathway, VEGF contains a number of genes previously shown to be differentially expressed between labor and non-labor (see Fig. 3) [21]. Finally, if the significance level were to be relaxed to 10%, the impact analysis also identifies *antigen processing and presentation* pathway, which is again part of the immune system.

It is important to point out that neither the hypergeometric model nor GSEA manage to identify any adhesion-related pathway at the usual 1% or 5% levels. Similarly, in spite of the differential expression of a number of genes related to

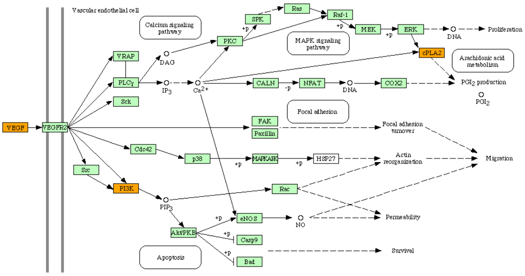


Fig. 3. The VEGF signaling pathway is one of the pathways identified by impact analysis. The genes found to be differentially expressed between labor and non-labor are highlighted in red. A more complete discussion about this pathway and its role in parturition is available elsewhere [21].

the VEGF-signaling, neither GSEA nor the classical ORA approach indicate that this pathway may be meaningful.

4 Conclusion

The classical statistical approaches used to identify significantly impacted pathways in a given condition only consider the number of differentially expressed genes and completely ignore other important biological factors. The impact analysis method uses a systems biology approach to extend the classical approach by incorporating important biological factors such as the magnitude of the expression changes, the topology and the type of signaling interactions between the genes on the pathway, and position of the differentially expressed genes on the pathway. The previously described impact analysis approach first computes the perturbations introduced by the differentially expressed genes in a pathway, and then propagates these perturbations throughout the pathway in order to calculate its impact factor. The perturbation propagation approach yields only an approximation of the gene perturbations when the pathways include loops. Here, we describe a modified impact analysis approach that addresses these stability issues. The results obtained on a human uterine cervix data set suggest that: i) the modified impact analysis approach has a higher statistical power and ii) it can identify several additional pathways that are likely to be involved in cervical dilation and re-modeling.

Acknowledgements

This material is based upon work supported by the following grants: NSF DBI 0234806, CCF 0438970, 1R01HG003491, 1U01CA117478, 1R21CA100740, 1R01 NS045207, 5R21EB000990, 2P30 CA022453. Onto-Tools currently runs on equipment provided by Sun Microsystems EDU 7824-02344-U and by NIH(NCRR) 1S10 RR017857-01. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, NIH, DOD or any other of the funding agencies.

References

1. Khatri, P., Drăghici, S., Ostermeier, G.C., Krawetz, S.A.: Profiling gene expression using Onto-Express. Genomics 79(2), 266–270 (2002)
2. Drăghici, S., Khatri, P., Martins, R.P., Ostermeier, G.C., Krawetz, S.A.: Global functional profiling of gene expression. Genomics 81(2), 98–104 (2003)
3. Khatri, P., Draghici, S.: Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 21(18), 3587–3595 (2005)
4. Pavlidis, P., Qin, J., Arango, V., Mann, J.J., Sibille, E.: Using the gene ontology for microarray data mining: A comparison of methods and application to age effects in human prefrontal cortex. Neurochemical Research 29(6), 1213–1222 (2004)
5. Goeman, J.J., van de Geer, S.A., de Kort, F., van Houwelingen, H.C.: A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 20(1), 93–99 (2004)
6. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Altshuler, D., Groop, L.C.: Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics 34(3), 267–273 (2003)
7. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P.: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceeding of The National Academy of Sciences of the USA 102(43), 15545–15550 (2005)
8. Tian, L., Greenberg, S.A., Kong, S.W., Altschuler, J., Kohane, I.S., Park, P.J.: Discovering statistically significant pathways in expression profiling studies. Proceeding of The National Academy of Sciences of the USA 102(38), 13544–13549 (2005)
9. Stelling, J.: Mathematical models in microbial systems biology. Current opinion in microbiology 7(5), 513–518 (2004)
10. Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R.: A systems biology approach for pathway level analysis. Genome Research 17 (2007)
11. Doniger, S.W., Salomonis, N., Dahlquist, K.D., Vranizan, K., Lawlor, S.C., Conklin, B.R.: MAPPFinder: using Gene Ontology and GenMAPP to create a global gene expression profile from microarray data. Genome biology 4(1), R7 (2003)
12. Pan, D., Sun, N., Cheung, K.H., Guan, Z., Ma, L., Holford, M., Deng, X., Zhao, H.: PathMAPA: a tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for Arbidopsis. BMC Bioinformatics 4(1), 56 (2003)
13. Pandey, R., Guru, R.K., Mount, D.W.: Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. Bioinformatics 20(13), 2156–2158 (2004)
14. Breslin, T., Krogh, M., Peterson, C., Troein, C.: Signal transduction pathway profiling of individual tumor samples. BMC Bioinformatics 6, 1471–2105 (2005)

15. Robinson, P.N., Wollstein, A., Bohme, U., Beattie, B.: Ontologizing gene-expression microarray data: characterizing clusters with gene ontology. Bioinformatics 20(6), 979–981 (2004)
16. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30(1–7), 107–117 (1998)
17. Haveliwala, T.: Efficient computation of PageRank. Technical Report 1999-31, Database Group, Computer Science Department, Stanford University (February 1999)
18. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report (1998)
19. Canales, R.D., Luo, Y., Willey, J.C., Austermiller, B., Barbacioru, C.C., Boysen, C., Hunkapiller, K., Jensen, R.V., Knight, C.R., Lee, K.Y., Ma, Y., Maqsodi, B., Papallo, A., Peters, E.H., Poulter, K., Ruppel, P.L., Samaha, R.R., Shi, L., Yang, W., Zhang, L., Goodsaid, F.M.: Evaluation of dna microarray results with quantitative gene expression platforms. Nat. Biotechnol. 24(9), 1115–1122 (2006)
20. Draghici, S., Khatri, P., Eklund, A.C., Szallasi, Z.: Reliability and reproducibility issues in DNA microarray measurements. Trends Genet. 22(2), 101–109 (2006)
21. Hassan, S.S., Romero, R., Haddad, R., Hendler, I., Khalek, N., Tromp, G., Diamond, M.P., Sorokin, Y., Malone, J.J.: The transcriptome of the uterine cervix before and after spontaneous term parturition. Am. J. Obstet. Gynecol. 195(3), 778–786 (2006)
22. Hassan, S.S., Romero, R., Tarca, A.L., et al.: Signature pathways identified from gene expression profiles in the human uterine cervix before and after spontaneous term parturition. Am. J. Obstet. Gynecol. 197(3), 250.e1–250.e7 (2007)
23. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P.: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2), 249–264 (2003)
24. Tarca, A.L., Carey, V.J., Chen, X.W., Romero, R., Draghici, S.: Machine learning and its applications to biology. PLoS Comput. Biol. 3(6), e116 (2007)
25. Smyth, G.K.: In: Limma: linear models for microarray data, pp. 397–420. Springer, New York (2005)
26. Saito, S., Shima, T., Nakashima, A., Shiozaki, A., Ito, M., Sasaki, Y.: What is the role of regulatory t cells in the success of implantation and early pregnancy? J. Assist Reprod. Genet. Epub. ahead of print (August 2007)
27. King, A., Kelly, R., Sallenave, J., Bocking, A., Challis, J.: Innate immune defences in the human uterus during pregnancy. Placenta Epub ahead of print (July 2007)
28. Tsatas, D., Baker, M.S., Rice, G.E.: Differential expression of proteases in human gestational tissues before, during and after spontaneous-onset labour at term. J. Reprod. Fertil. 116(1), 43–49 (1999)
29. Koelbl, H., Kirchheimer, J., Tatra, G.: Influence of delivery on plasminogen activator inhibitor activity. J. Perinat. Med. 17(2), 107–111 (1989)
30. Turpeenniemi-Hujanen, T., Feinberg, R.F., Kauppila, A., Puistola, U.: Extracellular matrix interactions in early human embryos: implications for normal implantation events. Fertil Steril 64(1), 132–138 (1995)
31. Xu, P., Wang, Y.L., Zhu, S.J., Luo, S.Y., Piao, Y.S., Zhuang, L.Z.: Expression of matrix metalloproteinase-2, -9, and -14, tissue inhibitors of metalloproteinase-1, and matrix proteins in human placenta during the first trimester. Biol. Reprod. 62(4), 988–994 (2000)