UNRAVELING DYNAMIC ACTIVITIES OF AUTOCRINE PATHWAYS THAT CONTROL DRUG-RESPONSE TRANSCRIPTOME NETWORKS

YOSHINORI TAMADA[1][*][†], HIROMITSU ARAKI[2][*], SEIYA IMOTO[1][*], MASAO NAGASAKI[1], ATSUSHI DOI[2], YUKIKO NAKANISHI[2], YUKI TOMIYASU[2], KAORI YASUDA[2], BEN DUNMORE[3], DEBORAH SANDERS[3], SALLY HUMPHREYS[3], CRISTIN PRINT[4], D. STEPHEN CHARNOCK-JONES[3], KOUSUKE TASHIRO[5], SATORU KUHARA[5], SATORU MIYANO[1]

[1] *Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku Tokyo, 108-8639, Japan*

[2] *Systems Pharmacology Research Institute, GNI Ltd., 3-8-33 Momochihama, Sawaraku-ku, Fukuoka, 814-0011, Japan*

[3] *Department of Obstetrics & Gynaecology, University of Cambridge, The Rosie Hospital, Robinson Way, Cambridge CB2 2SW, United Kingdom*

[4] *Department of Molecular Medicine & Pathology, School of Medical Sciences, The University of Auckland, Private Bag 92019, Auckland, New Zealand*

[5] *Graduate School of Genetic Resources Technology, Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan*

Some drugs affect secretion of secreted proteins (e.g. cytokines) released from target cells, but it remains unclear whether these proteins act in an autocrine manner and directly effect the cells on which the drugs act. In this study, we propose a computational method for testing a biological hypothesis: there exist autocrine signaling pathways that are dynamically regulated by drug response transcriptome networks and control them simultaneously. If such pathways are identified, they could be useful for revealing drug mode-of-action and identifying novel drug targets. By the node-set separation method proposed, dynamic structural changes can be embedded in transcriptome networks that enable us to find master-regulator genes or critical paths at each observed time. We then combine the protein-protein interaction network with the estimated dynamic transcriptome network to discover drug-affected autocrine pathways if they exist. The statistical significance (p-values) of the pathways are evaluated by the meta-analysis technique. The dynamics of the interactions between the transcriptome networks and the signaling pathways will be shown in this framework. We illustrate our strategy by an application using anti-hyperlipidemia drug, Fenofibrate. From over one million protein-protein interaction pathways, we extracted significant 23 autocrine-like pathways with the Bonferroni correction, including VEGF−NRP1−GIPC1−PRKCA−PPARα, that is one of the most significant ones and contains PPARα, a target of Fenofibrate.

---

[*]These authors contributed equally to this work.

[†]Corresponding author: tamada@ims.u-tokyo.ac.jp

1. Introduction

Understanding mode-of-action of drugs has received considerable attention in pharmacogenomics. Drug-response pathways at a transcriptome level are successfully predicted by cutting-edge computational techniques.[1,2] On the other hand, some drugs affect the pathways at protein level. For example, drugs affect secretion of secreted proteins (e.g. cytokines and growth factors) which are released from target cells. There is a possibility that these proteins have effects on target cells through drug-effected autocrine pathways. From the drug development viewpoint, these pathways could be useful for revealing drug mechanism of action, potentiation of drug effects and avoidance of side effects.

To validate the existence of such drug-affected autocrine pathways, we propose a novel computational method for finding signaling pathways that have the potential to regulate transcriptome networks. The method combines transcriptome networks estimated as drug-response pathways from mRNA expression data with proteome networks represented by protein-protein interactions to extract such pathways. First, we estimate a dynamic transcriptome network from drug-response time-course microarray data by dynamic Bayesian networks with nonparametric regression.[5] For this, we propose the node-set separation method that enables us to find sub-networks significantly activated at observed time points, master-regulator genes and critical paths in the drug-response pathways. We then combine protein-protein interaction (PPI) network with the estimated dynamic transcriptome network. The candidate signaling pathways that connect a ligand or a receptor to the key genes in the transcriptome network are extracted and evaluated based on statistical hypothesis testing at each observed time. Based on the computed p-values, the candidate drug-affected autocrine pathways are selected by multiplicity corrected significance level.

Comparing with the existing computational methods to find mode-of-action of drugs, the novelty of the proposed method is to consider the dynamical interactions between transcriptome and proteome networks. The methods that consider the mode-of-action of drugs only in transcriptome networks cannot be used for our purpose. Moreover, dynamic structural changes of transcriptome networks obtained by the proposed node-set separation method is a key for our purpose, but it cannot be obtained by ordinary dynamic Bayesian networks. For computational autocrine pathway identification, correlation of microarray data between known ligand-receptor pairs was used in cancer cells.[6] The proposed method can be

considered as an extension of this research direction to extract dynamics of autocrine pathways from drug-response data. We would like to emphasize again that the existence of drug-affected autocrine pathways is a hypothesis and their existence has not been demonstrated. Therefore, to our knowledge, this is the first genome-wide investigation to test this hypothesis.

We illustrate the whole process of the proposed method by an application using anti-hyperlipidemia drug Fenofibrate, which is known as an agonist of PPARα. The drug effect of Fenofibrate is not only lipid lowering, but also anti-inflammatory in vascular cells.[3,4] The molecular mechanism of lipid lowering by Fenofibrate is well known. The mechanism depends on ligand dependent transcription-activity in which PPARα binds to a specific DNA element termed the PPAR response element (PPRE) of beta-oxidation enzymes and proteins implicated in the reverse cholesterol transport pathways. However, the molecular mechanism mediating transrepression of the anti-inflammatory effects of Fenofibrate which may include post-transcriptional modification of PPARα has not been established.[3,4] Our method extracted candidate autocrine pathways including PPARα with highly statistical significance. These pathways may be involved in the trans-repression properties of PPARα. We also extracted 23 candidate drug-affected autocrine pathways from over one million protein-protein interaction pathways. Therefore, we observed candidate drug-affected autocrine pathways that support the hypothesis.

2. Methods

2.1. *Dynamic Transcriptome Network*

Figure 1 represents the overview of the proposed method. Based on drug response time-course microarray data, we estimate a dynamic transcriptome network by the dynamic Bayesian network (DBN) model with non-parametric regression.[5] However, ordinary dynamic Bayesian networks can estimate a network from time-course data, while at each observed time, different sub-networks have high activity and transmit information of external signals to other sub-networks. Therefore, we need to extend dynamic Bayesian networks to capture this feature.

The key idea of our DBN estimation, called node-set separation method, is to define the *active gene set* for each time point. That is, a gene in an active gene set is determined as a differentially expressed gene comparing with the controls. Let $\mathcal{A}_t = \{g_i : \mathrm{pv}(g_i, t) \leq \theta_t\}$ be the active gene set at time t for $t = 1, \ldots, T$, where g_i represents the ith gene, $\mathrm{pv}(g_i, t)$ is the
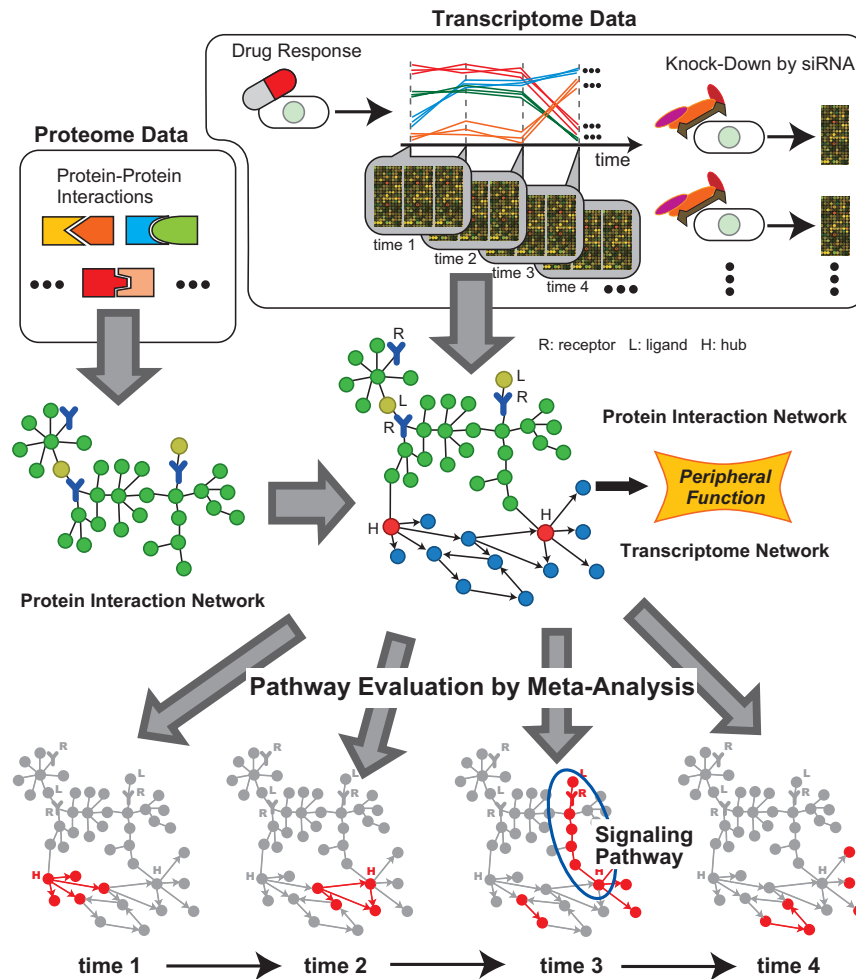
Figure 1.   Overview of the proposed method.

p-value of g_i at time t, and θ_t is the threshold for time t that could be determined by using false discovery rate for example. In our case, the p-value of each gene is computed by comparing triplet expression values of the gene at a time with control four replicate expression values, i.e., expression data of non-treated cells. We then define the *node set* $\mathcal{N}_t = \mathcal{A}_{t-1} \cup \mathcal{A}_t$ for $t = 1, ..., T$, where $\mathcal{A}_0$ is the empty set.

The definition of the node set has the basis on the Markov process

of the dynamic Bayesian networks, i.e., the DBN assumes the first order Markov process among time-course data; this yields the decomposition $\Pr(\boldsymbol{X}_1, ..., \boldsymbol{X}_T) = \prod_{t=1}^{T} \Pr(\boldsymbol{X}_t | \boldsymbol{X}_{t-1})$, where $\boldsymbol{X}_t$ is the expression data vector at time t and $\boldsymbol{X}_0$ is the empty set. The transcriptome network at time t, we denote G_t, is estimated for the node set $\mathcal{N}_t$ by the DBN and nonparametric regression with whole expression data $\boldsymbol{X}_1, ..., \boldsymbol{X}_T$.[5] Finally the dynamic transcriptome network is obtained by $G = G_1 \cup \cdots \cup G_T$. The advantage of this estimation procedure, i.e., using node set $\mathcal{N}_t$ separately, by comparing with other algorithms that use $\mathcal{N} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_T$ as the node set is not only finding dynamics of transcriptome networks, but also possibility to reduce false positive edges in the network, because we can reduce the size of the gene set for each observed time efficiently; this can increase the accuracy of the structure learning.

2.2. *Signaling Pathway Extraction*

2.2.1. *PPI Paths for Candidate of Signaling Pathways*

First, we define master-regulator genes in each node set $\mathcal{N}_t$, based on the estimated G_t for $t = 1, ..., T$. In this paper, the hub genes in $\mathcal{N}_t$ are defined as the top 5% genes; the genes in $\mathcal{N}_t$ are ranked according to the numbers of their direct child-genes in G_t. We denote the set of hub genes of $\mathcal{N}_t$ as $\mathcal{H}_t$. We also focus on the direct parents of the hub genes and represent the set of parent genes of the hub genes in $\mathcal{H}_t$ as $\mathcal{P}_t$. Since the hub genes and their direct parents could control the transcription levels of many genes in $\mathcal{N}_t$, we thus define the set of master-regulator genes at time t by $\mathcal{M}_t = \mathcal{H}_t \cup \mathcal{P}_t$.

We then focus on the PPI network for exploring candidates of signaling pathways affecting master-regulator genes. On the PPI network, for $g_i \in \mathcal{M}_t$, we search receptors and ligands, denoted by r_j, that connect g_i by l or less edges, i.e., g_i connects with r_j by $l-2$ or less intermediate proteins. We denote the kth PPI path for the genes in $\mathcal{M}_t$ ending at $g_i \in \mathcal{M}_t$ as $s_{tk} = r_j - p_1 - p_2 - \cdots - g_i$ where p_1 and p_2 represent the intermediate proteins in the PPI network.

2.2.2. *P-Values for PPI Paths by Meta-Analysis*

Let $[p_i]$ represent the gene for the ith protein in the PPI network, i.e., if p_i is a protein translated from the i'th gene, we have $[p_i] = g_{i'}$. We also define $[r_j]$ in the same way. We assess the significance of s_{tk} using the p-values, $\mathrm{pv}([p'], t)$ for $p' \in s_{tk} \backslash \{g_i\}$, by statistical meta-analysis.[7] That

is, we regard the p-value of each genes in s_{tk} as an evidence whether the PPI pathway s_{tk} is activated or not. We use the statistical meta-analysis method for integrating p-values of genes in s_{tk} into the p-value of s_{tk}.

The integrated p-value for s_{tk} will be computed under the null hypothesis: all p-values $\mathrm{pv}([p'], t)$ are not significant, and the alternative hypothesis: at least one or more p-values $\mathrm{pv}([p'], t)$ are significant. That is, if the null hypothesis is not rejected, s_{tk} seems to be not functional; otherwise if we observe the small p-value, s_{tk} is activated and is functional. For the meta-analysis, we use Fisher's inversion method to integrate p-values. We remove the p-value of g_i for the meta-analysis, because g_i was selected as a significant genes in $\mathcal{N}_t$. Therefore, it is obvious that s_{tk} is decided as significant if g_i is included in the meta-analysis calculation, and is meaningless.

Since the node set $\mathcal{N}_t$ is constructed by the active gene sets of time t and $t-1$, there are two ways to assess the significance of s_{tk} by using either p-value at time t or $t-1$. In the real data analysis, we test both cases and assess the significance of each PPI path. We determine s_{tk} is significant if and only if either $\mathrm{pv}(s_{tk}, t) < \xi_t$ or $\mathrm{pv}(s_{tk}, t-1) < \xi_{t-1}$ holds, where $\mathrm{pv}(s_{tk}, t)$ is the integrated p-value of the PPI path s_{tk} with p-values at time t and ξ_t is the threshold determined by considering multiplicity of the testings. In the real data analysis, we use 1% significant level with the Bonferroni correction. Obviously, other methods for controlling multiplicity of testing, such as family-wise error rate, false discovery rate and so on, can be used for reducing false negatives. The reason why we choose the Bonferroni method is that since we use the results of statistical tests for mRNA expression data for finding the significance of protein levels, some changes of protein levels are not measured normally. Therefore, we choose the most strict correction method to achieve a conservative method.

Finally, we should indicate that if high-throughput protein expression data such as time course protein array data are available, we can replace the results of statistical testings for mRNA transcriptional data by the results from protein expression data. Our method can be applied directly to high-throughtput protein expression data immediately.

3. Discovering Fenofibrate-Affected Pathways

In order to demonstrate the capability of the proposed approach, we analyzed gene expression data of human umbilical vein endothelial cells (HUVECs) treated with the anti-hyperlipidemia drug, Fenofibrate. Fenofibrate is an agonist of the peroxisome proliferator-activated receptor α

(PPARα), which is known as a transcription factor that induces genes related to the lipid metabolism. Also, recent studies revealed that Fenofibrate has anti-inflammatory effects.[3,4] Indeed, PPARα activators positively regulate the secretion of secreted proteins.[8] Therefore, we think a PPARα activator is an optimal target (example) to evaluate our approach presented. This application aims at identifying Fenofibrate-affected autocrine pathways related to its anti-inflammatory effects, and at elucidation of unknown modes-of-action.

3.1. *Data Set*

3.1.1. *HUVEC Gene Expression Data and Transcriptome Networks*

We used CodeLink$^{\mathrm{TM}}$ Human UniSet I 20K arrays for measuring drug-response time course and knockdown expression data. For the time course data, we observed 6 time points including the control, 2, 4, 6, 8 and 18 hours after treated with 25 μM Fenofibrate in 3 or 4 replicates. For the knockdown gene expression data we knocked down 400 transcripts, which are mainly transcription factors, by siRNAs. We excluded probes which have less than 90% G flags for all 400 arrays from the knockdown expression data. Missing values were imputed by LSimput.[9]

We selected genes whose SAM[10] (Significance Analysis of Microarrays) q value $\leq 5\%$ and fold change ≥ 1.5 at time t for $\mathcal{A}_t$ where $t = 1, \ldots, 5$ corresponding to 2hr, 4hr, 6hr, 8hr and 18hr, respectively. If a gene has more than one probe in the microarrays, we selected the one that has the smallest average of SAM p values for all the time points. This is required for the later steps to match each probe in the microarrays to a protein in the PPI network. Finally, the numbers of genes in $\mathcal{A}_t$ are 14, 5, 144, 129 and 370, respectively. The numbers of genes in $\mathcal{N}_t$ are 14, 19, 144, 200, and 454, respectively. The total number of unique genes in the network is 527. The online supplement[30] gives the complete list of these genes.

The transcriptome networks G_t ($t = 1, \ldots, 5$) were estimated with the prior networks which were estimated from knockdown gene expression data to incorporate transcriptome level changes which can be observed by knocking-down genes by siRNAs.[11] The reliability of the edges in the estimated networks is calculated by the bootstrap method[1] with 1 000 iterations. Edges whose bootstrap probability is less than threshold 0.05 were removed from the final transcriptome networks.

Table 1.   Summary of the dynamic transcriptome networks and the extracted pathways on the PPI network ($l = 4$).   "$|\mathcal{H}_t|$" is the number of hub genes, "$|\mathcal{M}_t|$" the number of hubs and their parents in G_t, and "# s_{tk}" the number of all possible pathways.   Columns "eval $t-1$" and "eval t" are the numbers of pathways that are statistically significant if evaluated by $\mathrm{pv}(s_{tk}, t-1)$ and $\mathrm{pv}(s_{tk}, t)$ respectively.

G_t	nodes	edges	$|\mathcal{H}_t|$	$|\mathcal{M}_t|$	# s_{tk}	eval $t-1$	eval t
1 (2hr)	14	59	1	9	590 644	−	125
2 (2hr/4hr)	19	91	1	2	28 384	23	271
3 (4hr/6hr)	144	625	7	31	1 016 831	673	1 870
4 (6hr/8hr)	200	874	10	42	1 861 999	1 744	587
5 (8hr/18hr)	454	1 982	22	51	1 194 215	436	150

3.1.2. *Protein-Protein Interaction Data*

We used the integrated PPI data set publicly available in Genome Network Platform[24] (GNP) (released on May 27 2008).   GNP collected the PPI data from the public PPI databases of BIND[25] (released on Jun. 25 2006), BioGRID[26] (ver. 2.0.37), HPRD[27] (rel. 7), IntAct[28] (released on Jan. 25 2008), and MINT[29] (released on Dec. 21 2007).   In addition to the literature based data, it also contains their own experimental data by yeast two hybrid experiments.   In total, GNP PPI data consists of 49 950 non redundant PPIs for 10 103 unique Entrez Gene IDs.   By removing proteins which does not have the corresponding probes in the time course data, the final PPI network contains 42 570 edges for 9 016 proteins.   We extracted 308 receptor and 149 ligand (in total 457) proteins from the PPI network, which are used as starting nodes r_j of the pathway extraction.

3.2. **Results**

Table 1 summarizes the number of nodes ($|\mathcal{N}_t|$), edges, hubs ($|\mathcal{H}_t|$), and hubs and their parents ($|\mathcal{M}_t|$) in the dynamic transcriptome network G_t. Here hub genes are defined as top 5 % genes ranked according to the number of direct child-genes in G_t.   These hub and their parent genes were used as target nodes of the pathway extraction in the later step.   The hub genes include cell growth related genes (SESN2), inflammatory response transcription factors (CEBPB, ANKRD1, PPARα), as well as many genes of unknown function.   The complete lists of hub genes and their parents are available on the online supplement.[30]

Next we checked the number of possible pathways to determine the appropriate l (maximum distance).   Table 2 shows the number of all possible
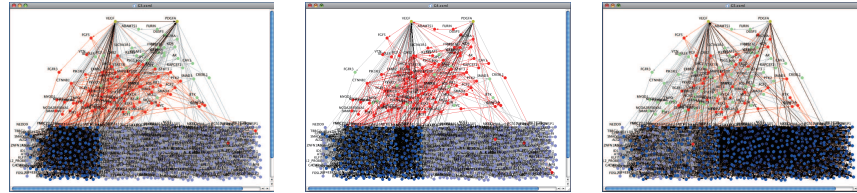
Figure 2.   Extracted autocrine ligand pathways for G_3 (left), G_4 (center), and G_5 (right). The top two nodes VEGF and PDGFA are ligand genes in the network. Nodes and edges in the bottom are the transcriptome network. The middle part contains the proteins and the extracted significant PPI pathways. The large version of this figure is available on the online supplement.[30]

Table 2.   The number of possible pathways s_{5k} and the final significant pathways with $\mathrm{pv}(s_{5k}, 5) < \xi_t$ with respect to the maximum distance l.

l	all	final
1	13	3
2	651	3
3	27 373	43
4	1 194 215	150
5	51 078 582	806

Table 3.   The numbers of all the ALPs from ligands in $G_{t'}$ to $\mathcal{M}_t$. Column "total" represents the number of all possible pathways ($l = 4$). Columns "eval $t-1$" and "eval t" are the numbers of ALPs that are statistically significant if evaluated by $\mathrm{pv}(s_{tk}, t-1)$ and $\mathrm{pv}(s_{tk}, t)$ respectively.

$G_{t'}$ ligands	$\mathcal{M}_t$	total	eval $t-1$	eval t
2 (2hr/4hr)	3 (4hr/6hr)	437	35	126
2 (2hr/4hr)	4 (6hr/8hr)	894	160	27
3 (4hr/6hr)	4 (6hr/8hr)	1 448	177	28
2 (2hr/4hr)	5 (8hr/18hr)	533	30	27
3 (4hr/6hr)	5 (8hr/18hr)	873	23	23
4 (6hr/8hr)	5 (8hr/18hr)	873	23	23

pathways from ligands or receptors to $\mathcal{M}_5$ (the hubs and their parents in 8hr/18hr transcriptome network G_5) evaluated by p-values of 18hr Fenofibrate time course gene expression data ($\mathrm{pv}(s_{5k}, 5)$). According to this table, we decided to use $l = 4$ since it seems to be the most realistic and appropriate for the later analysis.

The number of the extracted pathways are shown in Table 1. For example, there are 1 194 215 possible pathways within distace $l = 4$ from receptors or ligands to hub and their parents in G_5. Out of these huge number of pathways, there are only 150 statistically significant pathways if their p-values were evaluated by the gene expression data observed at time $t = 5$ (18hr). The complete list of the significant pathways are avaiable on the online supplement.[30]

In order to confirm that the method can capture known pathways related to Fenofibrate, we focused on PPI paths related to PPARα, since PPARα is a target of Fenofibrate. In the dynamic transcriptome network analysis,
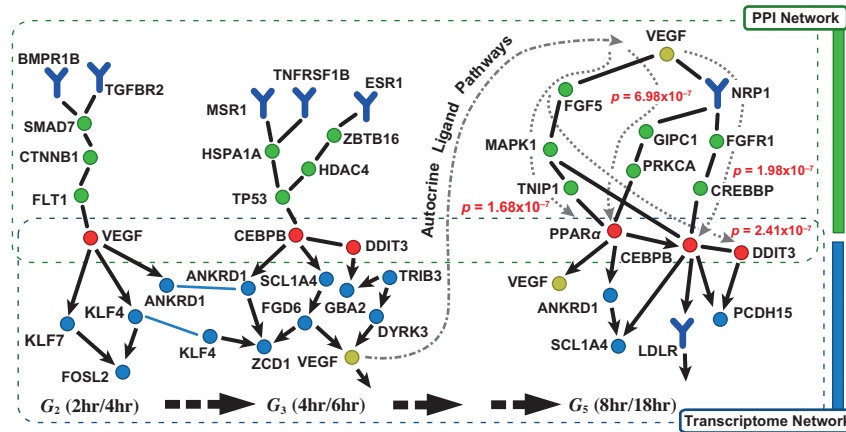
Figure 3.   The top four autocrine pathways and their p-values, which are connected to G_5 (8hr/18hr transcriptome network) hub and their parent genes. Some parts of transcriptome networks and the significant pathways at the previous times are also presented to illustrate dynamical changes extracted by the proposed method.

PPARα is included in the node sets $\mathcal{N}_4$ and $\mathcal{N}_5$, i.e., PPARα was overexpressed at 8 and 18 hours. In both times, PPARα was selected as a hub gene, in G_4 PPARα has 21 children and 31 in G_5. Since we would like to investigate drug-affected autocrine pathways, we first limited the candidate PPI paths by autocrine ligand pathways (ALPs) that connect ligands included in earlier time transcriptome networks, i.e., active ligands in earlier times, to hub genes and their parent genes.

By the Bonferroni correction with 1% significance level, only 23 pathways from ligands in G_3 or G_4 to $\mathcal{M}_5$ evaluated by 8hr expression data remained as significant ALPs (Table 3). Among them, we found that the pathway including PPARα as a hub gene of the gene network has high statistical significance (the fourth highest significance). This ALP is VEGF$-$NRP1$-$GIPC1$-$PRKCA$-$PPARα. PRKCA, protein kinase C alpha, is located on the upstream of PPARα. PRKCA is one of the members of serine- and threonine-specific protein kinases and is related to phospholyration of many genes including PPARα. Protein kinase C inhibitor inactivates the phosphorylation of PPARα and induces the trans-repression activity of PPARα in hepatocytes.[14, 15] Our method was able to extract this known relationship, which is related to PPARα's trans-repression, with high statistical significance. VEGF, vascular endothelial growth factor A, is also included in this pathway. VEGF is a member of the PDGF/VEGF

growth factor family and is the predominant regulator of angiogeneis. It has been reported that Fenofibrate induces VEGF mRNA and prevents cell apoptotic cell death in human retinal endothelial cells (HRECs).[9] VEGF is also significantly up regulated in our microarray experiment. From this, our method suggests that the trans-repression property of Fenofibrate might be caused by PRKCA mediated thorough VEGF signaling.

We also found out that CEBPB (CCAAT/enhancer binding protein beta) is the ending node of the ALP and a hub gene of the gene network with the highest significance level. This ALP is VEGF−NRP1−FGFR1−CREBBP−CEBPB. CEBPB is a transcription factor which can bind as a homodimer or heterodimiers with the related proteins to certain DNA regulatory regions. CEBPB is important in the regulation of genes involved in immune and inflammatory responses and has been shown to bind to the IL-1 response element in the IL-6 gene, as well as to regulatory regions of several acute-phase and cytokine genes. Fenofibrate decreases CRP (C-reactive Protein) and fibrinogen which are risk factors of vascular disorders, through CEBPB.[16,17] Our method extracted the ALP including CEBPB that is a key gene for anti-inflammatory action and suggests that this ALP might also mediate the anti-inflammatory effects of Fenofibrate.

DDIT3 (DNA-damage-inducible transcript 3), which is a transcription factor of C/EBP family, is the ending node of the third highest ALP. This gene is involved in atherogenes induced by oxLDL (oxidized LDL)[18] and increases promoter activity of MCP-1.[19] PPARα increases uptake of oxLDL in bovine aortic endothelial cells (BAEC)[20] and induces atherogenesis in response to oxidized phospholipids, constituents of oxLDL, through the up-regulation of MCP-1 and IL-8 in human aortic endothelial cells (HAEC).[21] From these facts and our results, DDIT3 may partly play a role in oxLDL-induced atherogenesis through PPARα.

4. Discussion

In this paper, based on the assumption of the existence of drug-affected autocrine pathways, we presented a novel computational method capable of finding them. The results suggest that the autocrine pathways exist and have an important role in the regulation of transcriptome level networks affected by drugs. Previously, identification of autocrine signaling loops in cancer cells by microarray data was proposed,[6] but only co-expression of known ligand-receptor pairs was considered. The method proposed in

this paper can be viewed as an extension of their work by adding PPI networks and dynamics of transcriptome network that are essential to extract drug-affected autocrine pathways. Moreover, unlike the transcriptome level analyses,[1, 2, 6] the proposed method can capture proteomic and metabolomic level signaling pathways by exploiting the PPI networks and drug-response gene expression data. By combining the dynamic transcriptome network estimated by the node-set separation method and the meta-analysis technique, the method can extract statistically significant pathways from a huge number of possible candidates. Compared to the existing methods for analyzing signaling pathways,[22] the proposed method can extract genome-wide dynamical changes of the signaling pathways that are statistically significant. Although the proposed method currently uses only gene expression data, it can be easily applied to the genome-wide proteomic time course data when they are available.

In the real data analysis, we applied the method to transcription expression data in Fenofibrate treated HUVECs. Our proposed method extracted statistically significant pathways from over a million of the possible PPI pathways. Among them, the pathway from VEGF to PPARα through PRKCA has high statistical significance. PRKCA regulates the trans-repression activity of PPAR by the phosphorylation of PPARα. Thus, we speculate that this pathway might involve the transrepression effects of Fenofibrate. Other pathways (e.g. VEGF to CEBPB) which might be related to the drug effect of Fenofibrate are also extracted by this method. The method proposed in this study has potential advantages for drug discovery and development. First, our approach can outline drug mode of action not only at the level of mRNA regulatory relationships, but also protein-protein interactions. Pathways extracted in this study can not be detected by using gene networks based only on transcriptome data. Secondly, potentiation of existing drugs has recently been recognized as an important area of anti-tumor research.[23] We think our approach will be able to detect candidate targets in silico. For example, NRP1 (Neuropilin-1) is a co-receptor for VEGF and is present in both the pathways from VEGF to PPARα and from VEGF to CEBPB. Small molecules targeting NRP1 have the potential for potentiation of the anti-inflammatory effects of Fenofibrate.

Acknowledgments

References

1. S. Imoto *et al.*, *Pac. Symp. Biocomput.* **11**, 559–571 (2006)
2. D. di Bernardo *et al.*, *Nat. Biotechnol.* **23**(3), 377–383 (2005).
3. P. Lefebvre *et al.*, *J. Clin. Invest.* **116**(3), 571–580 (2006).
4. D. S. Straus and C. K. Glass, *Trends Immunol.* **28**(12), 551–558 (2007).
5. S. Kim *et al.*, *Biosystems* **75**(1-3), 57–65 (2004).
6. T. G. Graeber and D. Eisenberg, *Nat. Genet.* **29**(3), 295–300 (2001).
7. P. K. Gupta *et al.*, *Proc. 3rd ISBRA, LNBI* **4463**, 146–157 (2007).
8. L. Nilsson *et al.*, *Arterioscler. Thromb. Vasc. Biol.* **19**(6), 1577–1581 (1999).
9. T. H. Bø *et al.*, *Nucleic Acids Res.* **32**(3), e34 (2004).
10. V. G. Tusher *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **98**(9), 5116–5121 (2001).
11. M. Affara *et al.*, *Philos. Trans. R. Soc. Lond. B* **362**(1484), 1469–1487 (2007).
12. J. Kim *et al.*, *Exp. Eye Res.* **84**(5), 886–893 (2007).
13. K. A. Burns and J. P. Vanden Heuvel, *Biochim. Biophys. Acta.* **1771**(8), 952–960 (2007).
14. C. Blanquart *et al.*, *Mol. Endocrinol.* **18**(8), 1906–1918 (2004).
15. J. P. Gray *et al.*, *Biochemistry* **44**(30), 10313–10321 (2005).
16. R. Kleemann *et al.*, *Blood* **101**(2), 545–551 (2003).
17. P. Gervois *et al.*, *J. Biol. Chem.* **276**(36), 33471–33477 (2001).
18. T. Thum and J Borlak, *J. Biol. Chem.* **283**(28), 19456–19464 (2008).
19. K. Kodama *et al.*, *Am. J. Physiol. Cell Physiol.* **289**(3), C582–C590 (2005).
20. K. Hayashida *et al.*, *Biochem. Biophys. Res. Commun.* **323**(3), 1116–1123 (2004).
21. H. Lee *et al.*, *Circ. Res.* **87**(6), 516–521 (2000).
22. K. A. Janes and M. B. Yaffe, *Nat. Rev. Mol. Cell Biol.* **7**(11), 820–828 (2006).
23. S. Morgan-Lappe *et al.*, *Oncogene* **25**(9), 1340–1348 (2006).
24. GNP: `http://genomenetwork.nig.ac.jp/index_e.html`
25. BIND: `http://bind.ca/`
26. BioGRID: `http://www.thebiogrid.org/`
27. HPRD: `http://www.hprd.org/`
28. IntAct: `http://www.ebi.ac.uk/intact/`
29. MINT: `http://mint.bio.uniroma2.it/mint/`
30. Online Supplementary Information: `http://bonsai.ims.u-tokyo.ac.jp/~tamada/suppl/PSB2009/`.