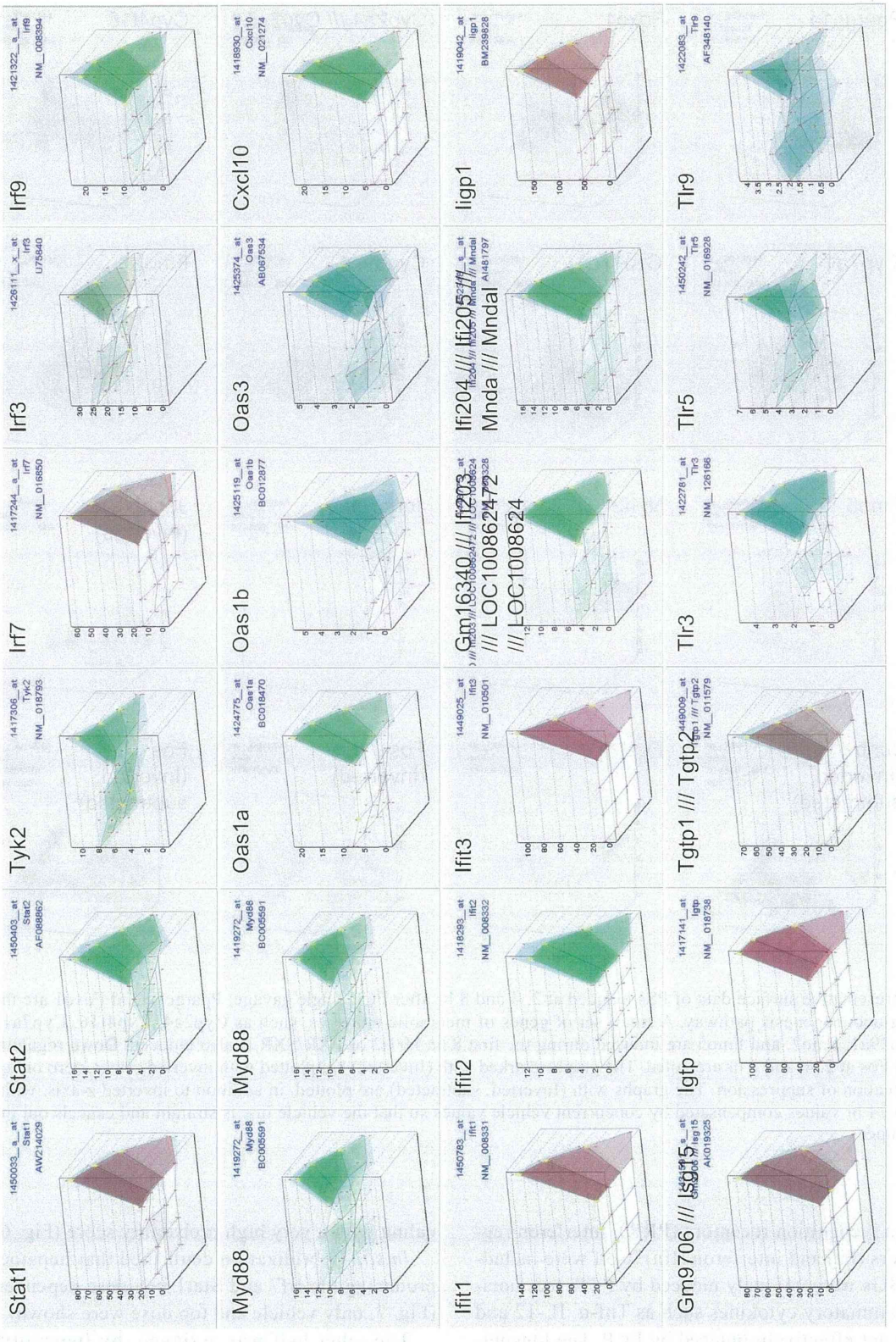

Fig. 4. Surface data of ISGs: Among about 1,200 PSs induced at 24 hr, half of them were uniquely induced by PCP and were assigned to ISG pathway from Stat1, Stat2, Tyk, to Irf7, Myd88, Oas, Ifit, Cxcl10 and other downstream targets. Some Tlrs were also uniquely up-regulated (cf. Supplementary Table 2).

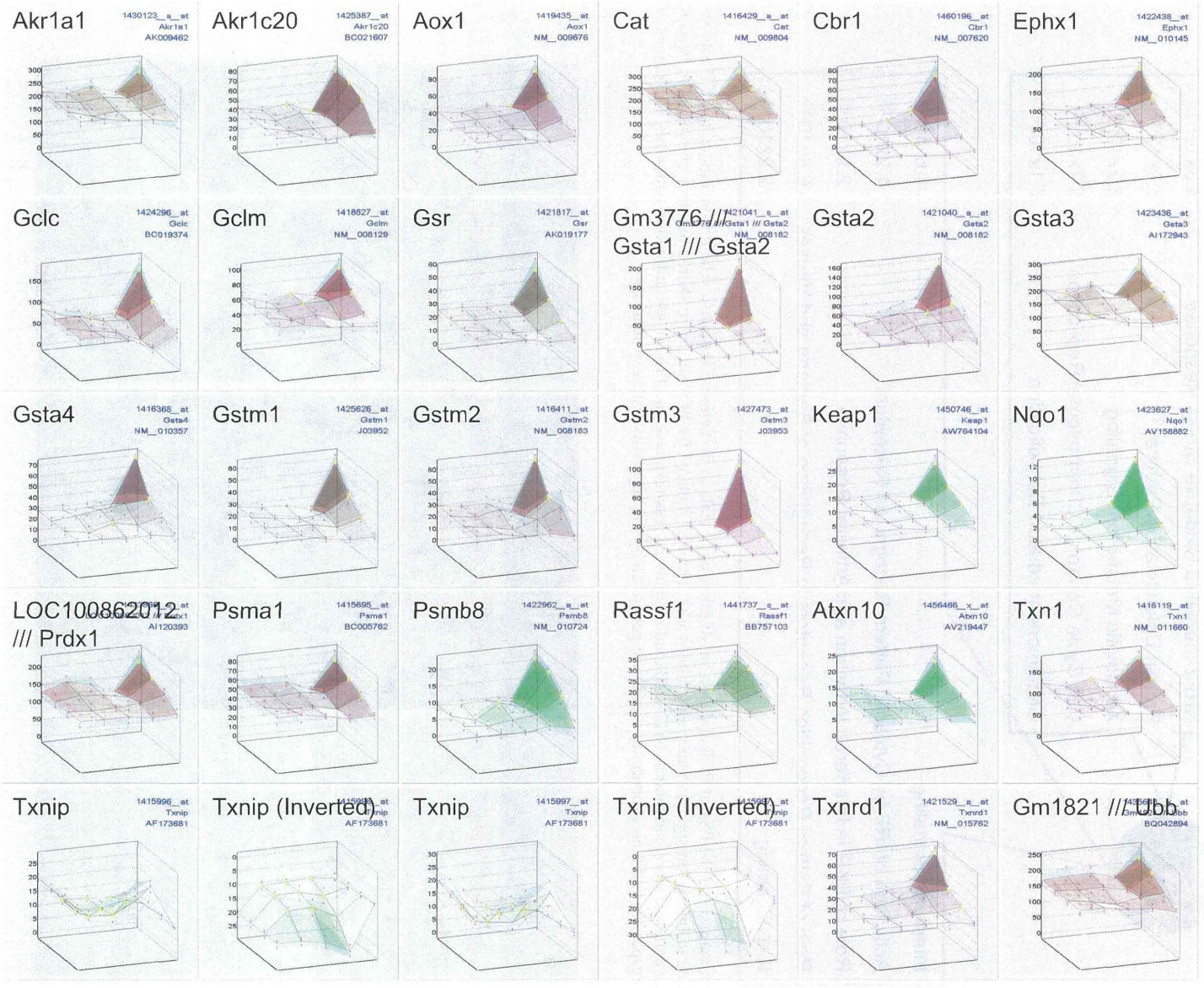

Fig. 5. Surface data of Nrf2-mediated oxidative stress response genes: Among about 1,200 PSs induced at 24 hr, another half of them were Nrf2-mediated oxidative stress response genes commonly induced by PCP and other 10 or so chemicals (cf. Supplementary Table 2). Nrf2 itself did not alter but Keap1 was clearly induced.

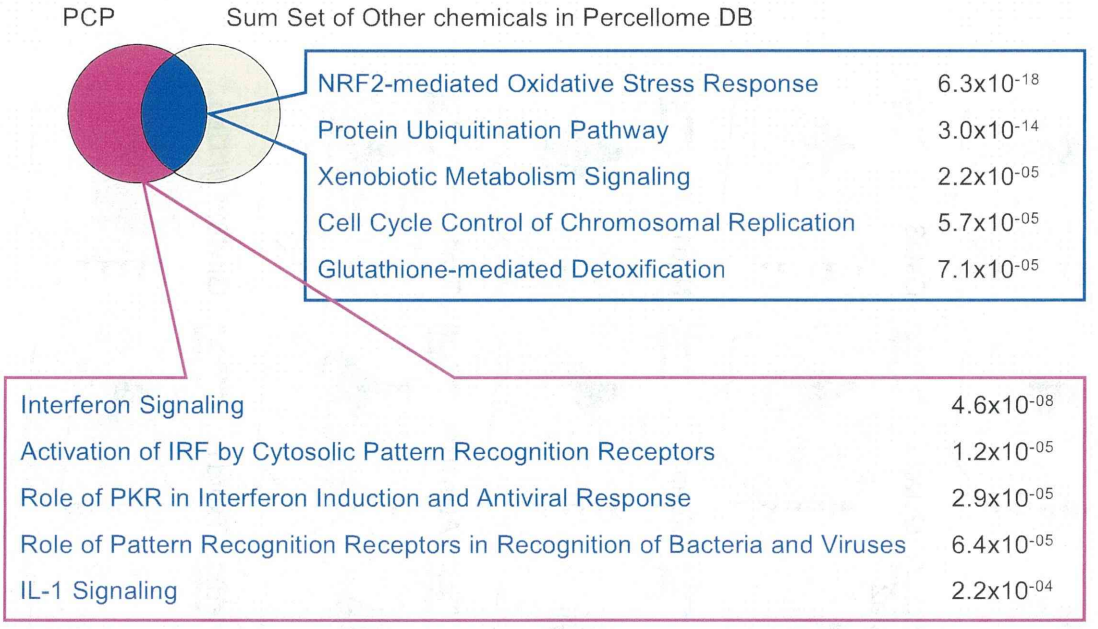

Fig. 6. Venn diagram of the PSs of PCP and sum set of other chemicals in Percellome Database. The PS list unique to PCP was assigned to Interferon signaling and related networks. The PSs induced by PCP and were shared by other chemicals in Percellome database were enriched in Nrf2-mediated oxidative stress response and Protein ubiquitination pathways. The names of the responses and their probability scores are generated by the Ingenuity Pathway analysis.

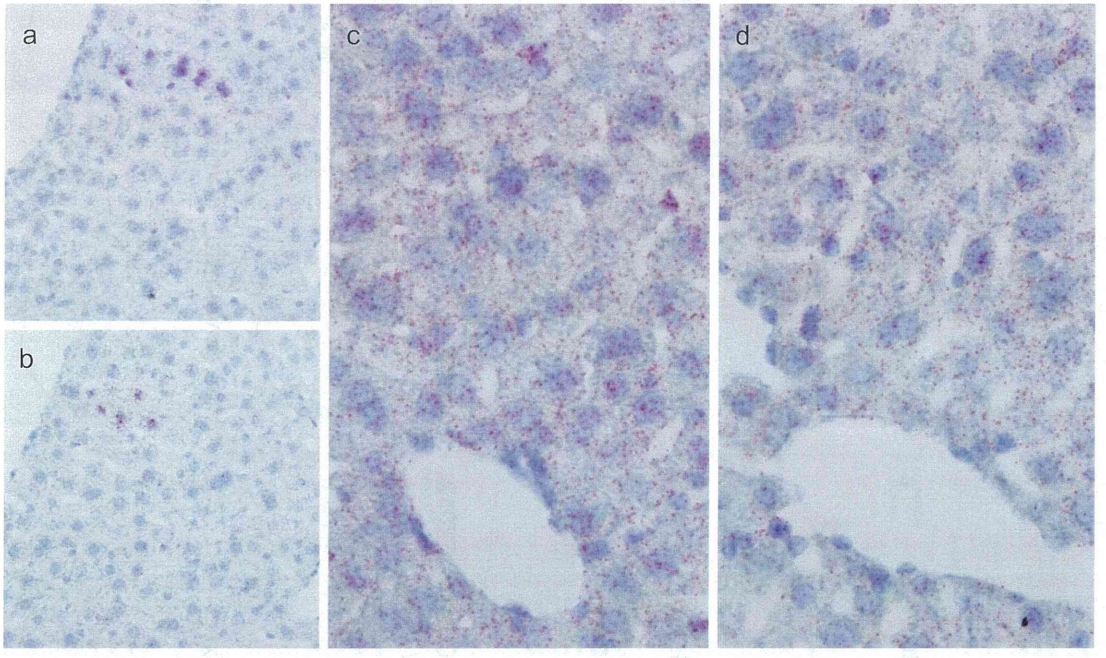

Fig. 7. *In situ* hybridization of Irf7 and Stat1. a) Vehicle control liver stained for Irf7. In a very low back ground, a small nest of hepatocytes was positively stained for Irf7. b) Vehicle control liver stained for Stat1. In a very low back ground, a small nest of hepatocytes was positively stained for Irf7. It is likely that the same hepatocyte is producing both mRNAs. c, d) High dose group stained for Irf7 and Stat1. Hepatocytes were shown to produce both mRNAs in a ubiquitous manner.

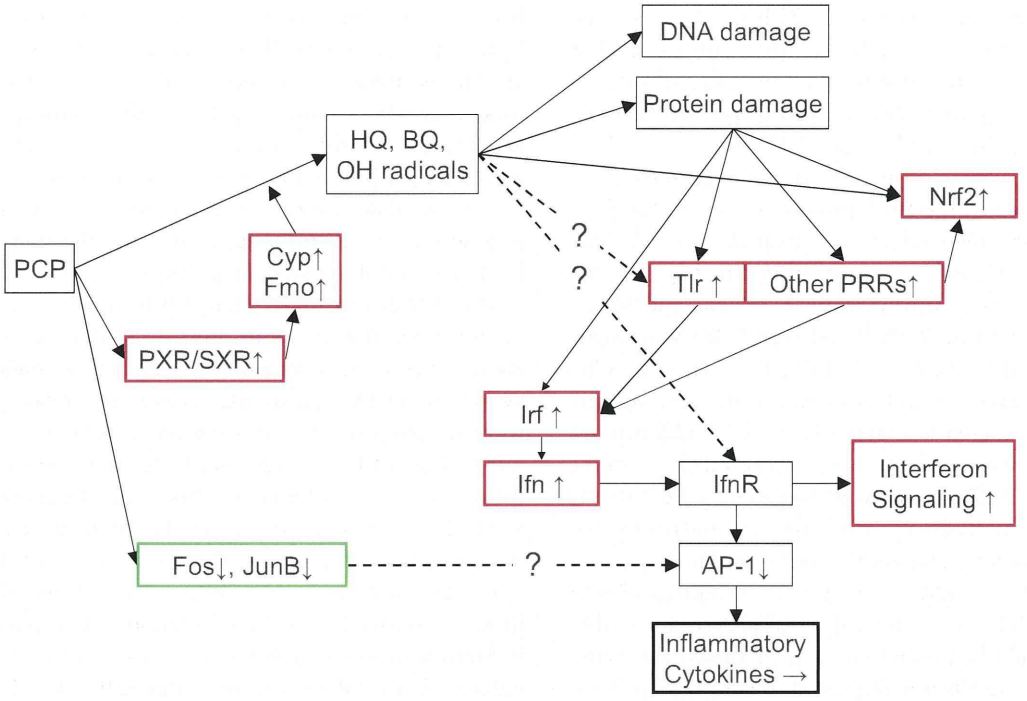

Fig. 8. Tentative summary scheme of the PCP induced networks in mouse liver. PCP or its metabolites may stimulate PXR/SXR or CAR, thereby inducing Cyp2, 4, 7, Fmo2, 5 within 8 hr to facilitate PCP metabolism, generating HQ (TCoHQ,TCpHQ), BQ (TCpBQ) and hydroxyl radical. The metabolites/radicals induce DNA damage and Protein damage. These reaction triggers Nrf2 networks and PRR system, initiating Irf mediated synthesis of Interferon alpha, which triggers the interferon signaling networks by autocrine or paracrine mechanisms. On the other hand, there remains a possibility that the metabolites may act as direct ligands to Tlr or IfnR and trigger downstream events as indicated in dotted lines with "?". Before activating interferon signaling networks, PCP suppressed Fos and JunB, which might have suppressed the inflammatory cytokine induction as shown in dotted line with "?".

accompanying distinct induction of Keap1, and other metabolic pathways (Figs. 5 and 6, Supplementary Table 1). Those networks were in the Common list mentioned above. Other networks were not effectively identified by the Ingenuity pathway analysis.

DISCUSSION

Among the chemicals tested in the Percellome project, PCP was slow to induce changes in gene expression; only around one hundred PSs were induced before 8 hr and 1,200 PSs at 24 hr. It would be plausible to hypothesize that PCP was metabolized during the first 8 hr and that the metabolite(s) then induced the 24 hr burst of ISGs and Nrf2-mediated genes. The time course of PCP action is in accord with the reported biological half-life of PCP; 6 to 27 hr in rodents (Larsen *et al.*, 1972; Braun *et al.*, 1977). A few metabolizing enzymes located downstream of PXR/SXR were induced during the first 8 hr (Fig. 3). The presence of DEHP in the top part of the common chemical list in Table 1 is also consistent with this hypothesis.

It would be of interest to ascertain whether PCP or its metabolite(s) could be PXR/SXR ligands. Metabolites known are tetrachloro-p(o)-hydroquinone (TCpHQ and TCoHQ) and tetrachloro-p-benzoquinone (chloranil, TCpBQ). Further, TCpHQ is reported to be metabolized, generating hydroxyl radicals with a help of H_2O_2 without Fenton reaction, to trichloro-hydroperoxyl-1,4-benzoquinone (TrCBQ-OOH) and trichloro-hydroxy-1,4-benzoquinone (TrCBQ-OH) (Zhu and Shan, 2009). We have no Percellome data of those metabolites and, to date, there are no reports on the interaction of PCP or its metabolites with the PXR/SXR. There are reports that PCP affects the function of estrogen receptor (Jung *et al.*, 2004) and thyroid hormone receptor (Kawaguchi *et al.*, 2008). Further study will be needed to identify the triggering event for the earliest responses to PCP.

Hepatocarcinogenic activity of PCP has been shown by rodent studies (NTP, 1999). The metabolites of PCP mentioned above were considered as the cause of oxidative stress or the hydroxyl radical insults against the liver (Zhu and Shan, 2009; Tasaki *et al.*, 2012).

It was reported that Tlr4-mediated, lipopolysaccharide-induced activation of the Ifn-β promoter was inhibited by PCP in a Myd88-independent way (Ohnishi *et al.*, 2008). On the other hand, PCP was considered to trigger Tlr4 via the induction of hydroxyl radicals (Lucas and Maes, 2013). In our experiment, PCP had significantly up-regulated Myd88, Irak1, Traf6, Tlr2, Tlr3, Tlr5, Tlr9 at 24 hr. Although the induction did not reach statistical significance, Tlr4 expression was also elevated. In addition, Irf 3, Irf7 and Irf9 were also induced. These findings might indicate that the TLR system was triggered by PCP itself, its metabolites or hydroxyl radicals via modifying the cytoplasmic proteins; abnormal proteins might be sensed by the TLRs or the pattern recognition receptors (PRR) system. Since Irf3, 7, 9 and Ifn-α1 expression are also increased, it could be possible that Irf mediated autocrine or paracrine of the Ifn-α is triggered, resulting in a burst of ISGs, as postulated by Sato *et al.* (Sato *et al.*, 2000).

Although Myd88 is mobilized, NF-κB, TNF, IL12 and CD40 were not induced, indicating that there may be a switch towards inflammatory cytokine production that was not directly induced by PCP. In relation to the switching mechanism, there is a report that isopropanol impaired AP-1 activation by removing Fos and JunB from the nuclear region of monocytes *in vitro* and suppressed Tlr4-mediated lipopolysaccharide stimulated Tnf-α production (Carignan *et al.*, 2011). In our data, Fos/JunB was down-regulated at earlier phase. As this change might correspond to AP-1 suppression (Gomard *et al.*, 2010), one possibility could be that PCP itself (as an alcohol/phenol) inhibited inflammatory cytokine production.

In contrast to the possible indirect mechanisms noted above, there are two examples that lead us to consider a possibility of direct activation of the ISGs. The first example is a low-molecular weight compound "imiquimod", a Tlr7 agonist, already on the market for treatment of skin viral infection (Hemmi *et al.*, 2002). Subcutaneous injection of imiquimod was reported to induce fever, sickness behavior and induction of ISG in rats (Damm *et al.*, 2012). The second example is a new polyfluoromethylated compound, 8-(1, 3,4-oxadiazol-2-yl)-2, 4-bis (trifluoromethyl) imidazo [1, 2-a] [1, 8] naphthyridine (RO4948191) which was reported to be an orally available low molecular weight interferon receptor agonist. RO4948191 was shown to induce a set of ISGs (Konishi *et al.*, 2012), such as Oas1, Adar, Bst1, Stat1, Ifit3, Usp18, Isg15, Herc6 and Cxcl10. These two examples of direct ligands to TLR and IFNR lead us to hypothesize that PCP and/or its metabolites may be able to directly activate these receptor systems (Fig. 8). Further investigation will be needed to clarify the molecular mechanism(s) through which PCP administration triggers the ISGs. It is tempting to speculate that this classic toxin, PCP, could be used as a new lead for orally applicable interferon-manipulating and/or cytokine switching drugs.

Hyperthermia or hyperpyrexia, profuse sweating, uncoordinated movement, muscle twitching, and coma are reported in humans and experimental animals as acute symptoms of PCP poisoning. These functional symptoms were reported to be caused by the mitochondrial uncoupling effect of PCP. Ucp2 and Ucp3 and some mitochondrial genes are induced in this study (Supplementary Table 2, Supplementary Fig 1). These genes are supportive for uncoupling effect. However, many of the Krebs cycle enzymes were not induced and those that were induced showed very small increases compared to the induction of ISGs. Some peroxidases were also mildly induced, but catalase was not induced by PCP. In the Percellome database, chemicals reported as uncouplers are aspirin, ethanol, sodium arsenite, and 2,4-dinitrophenol. Under the current computational condition, none of them was picked up as a chemical sharing PS list with PCP.

Taken together, our data may be interpreted to indicate that the functional symptoms represented by hyperthermia can be induced by PCP mainly through the activation of ISGs. This interpretation is backed up by the literature on "endogenous pyrogens" (Dinarello, 1999) and on imiquimod (Damm *et al.*, 2012).

Finally, although not yet perfected, the performance of the RSort and PE programs were demonstrated to be sufficient to sort out biologically meaningful changes for the comprehensive characterization of PCP. Manual searches employing different criteria added about 100 mildly changing PSs (data not shown), but the conclusions of our study were not affected. Nevertheless, further refinement for the better coverage is underway.

In conclusion, the RSort program-based comprehensive cross-reference of the Percellome database revealed that PCP was the only chemical among 111 orally administered chemicals that significantly induced the ISGs in hepatocytes. *In situ* hybridization confirmed that the parenchymal hepatocytes are responding to PCP. Two possible mechanisms were discussed; indirect mechanism via the PRR system, and direct stimulation of the Tlr(s) or IfnR(s). Further study is needed to clarify the possible molecular mechanisms.

ACKNOWLEDGMENTS

The author thanks all the member of the division of Cellular and Molecular Toxicology, NIHS for strong support of the project, and Dr. Bruce Blumberg for critical reading of the manuscript. The project has been supported by MHLW Health Sciences Research Grants H24-Kagaku-Shitei-006, H21-Kagaku-Ippan-001, H19-Toxico-001, H18-Kagaku-Ippan-001, H15-Kagaku-002, H14-Toxico-001, and H13-Seikatsu-012.

REFERENCES

Braun, W.H., Young, J.D., Blau, G.E. and Gehring, P.J. (1977): The pharmacokinetics and metabolism of pentachlorophenol in rats. Toxicol. Appl. Pharmacol., **41**, 395-406.

Carignan, D., Désy, O. and de Campos-Lima, P.O. (2011): The dysregulation of the monocyte/macrophage effector function induced by isopropanol is mediated by the defective activation of distinct members of the AP-1 family of transcription factors. Toxicol. Sci., **125**, 144-156.

Clayton, G.D. and Clayton, F.E. (1981): Toxicology. John Wiley Sons, New York.

Damm, J., Wiegand, F., Harden, L.M., Gerstberger, R., Rummel, C. and Roth, J. (2012): Fever, sickness behavior, and expression of inflammatory genes in the hypothalamus after systemic and localized subcutaneous stimulation of rats with the Toll-like receptor 7 agonist imiquimod. Neuroscience, **201**, 166-183.

Deichman, W., Machle, W., Kitzmiller, K.V. and Thomas, G. (1942): Acute and chronic effects of pentachlorophenol and sodium pentachlorophenate upon experimental animals. J. Pharmacol. Exp. Ther., **76**, 104-117.

Dinarello, C.A. (1999): Cytokines as endogenous pyrogens. J Infect Dis., **179 Suppl. 2**, S294-S304.

Finck, B.N., Gropler, M.C., Chen, Z., Leone, T.C., Croce, M.A., Harris, T.E., Lawrence, J.C.Jr. and Kelly, D.P. (2006): Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. Cell. Metab., **4**, 199-210.

Gomard, T., Michaud, H.A., Tempé, D., Thiolon, K., Pelegrin, M. and Piechaczyk, M. (2010): An NF-kappaB-dependent role for JunB in the induction of proinflammatory cytokines in LPS-activated bone marrow-derived dendritic cells. PLoS One, **5**, e9585.

Hemmi, H., Kaisho, T., Takeuchi, O., Sato, S., Sanjo, H., Hoshino, K., Horiuchi, T., Tomizawa, H., Takeda, K. and Akira, S. (2002): Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. Nat. Immunol., **3**, 196-200.

Jung, J., Ishida, K. and Nishihara, T. (2004): Anti-estrogenic activity of fifty chemicals evaluated by in vitro assays. Life Sci., **74**, 3065-3074.

Kanno, J., Aisaki, K., Igarashi, K., Nakatsu, N., Ono, A., Kodama, Y. and Nagao, T. (2006): "Per cell" normalization method for mRNA measurement by quantitative PCR and microarrays. BMC Genomics, **7**, 64.

Kawaguchi, M., Morohoshi, K., Saita, E., Yanagisawa, R., Watanabe, G., Takano, H., Morita, M., Imai, H., Taya, K. and Himi, T. (2008): Developmental exposure to pentachlorophenol affects the expression of thyroid hormone receptor beta1 and synapsin I in brain, resulting in thyroid function vulnerability in rats. Endocrine, **33**, 277-284.

Konishi, H., Okamoto, K., Ohmori, Y., Yoshino, H., Ohmori, H., Ashihara, M., Hirata, Y., Ohta, A., Sakamoto, H., Hada, N., Katsume, A., Kohara, M., Morikawa, K., Tsukuda, T., Shimma, N., Foster, G.R., Alazawi, W., Aoki, Y., Arisawa, M. and Sudoh, M. (2012): An orally available, small-molecule interferon inhibits viral replication. Sci. Rep., **2**, 259.

Larsen, R.V., Kirsch, L.E., Shaw, S.M., Christian, J.E. and Born, G.S. (1972): Excretion and tissue distribution of uniformly labeled 14 C-pentachlorophenol in rats. J. Pharm. Sci., **61**, 2004-2006.

Lucas, K. and Maes, M. (2013): Role of the Toll Like Receptor (TLR) Radical Cycle in Chronic Inflammation: Possible Treatments Targeting the TLR4 Pathway. Mol. Neurobiol., (in press).

NTP (1999): NTP Toxicology and Carcinogenesis Studies of Pentachlorophenol (CAS NO. 87-86-5) in F344/N Rats (Feed Studies). In: Natl Toxicol Program Tech Rep Ser), pp.1-182.

Ohnishi, T., Yoshida, T., Igarashi, A., Muroi, M. and Tanamoto, K. (2008): Effects of possible endocrine disruptors on MyD88-independent TLR4 signaling. FEMS Immunol. Med. Microbiol., **52**, 293-295.

Puigserver, P., Rhee, J., Donovan, J., Walkey, C.J., Yoon, J.C., Oriente, F., Kitamura, Y., Altomonte, J., Dong, H., Accili, D. and Spiegelman, B.M. (2003): Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1alpha interaction. Nature, **423**, 550-555.

Sato, M., Suemori, H., Hata, N., Asagiri, M., Ogasawara, K., Nakao, K., Nakaya, T., Katsuki, M., Noguchi, S., Tanaka, N. and Taniguchi, T. (2000): Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN-alpha/beta gene induction. Immunity, **13**, 539-548.

Tasaki, M., Kuroiwa, Y., Inoue, T., Hibi, D., Matsushita, K., Ishii, Y., Maruyama, S., Nohmi, T., Nishikawa, A. and Umemura, T. (2012): Oxidative DNA damage and *in vivo* mutagenicity caused by reactive oxygen species generated in the livers of p53-proficient or -deficient gpt delta mice treated with non-genotoxic hepatocarcinogens. J. Appl. Toxicol. DOI: 10.1002/jat.2807.

Zhu, B.Z. and Shan, G.Q. (2009): Potential mechanism for pentachlorophenol-induced carcinogenicity: a novel mechanism for metal-independent production of hydroxyl radicals. Chem. Res. Toxicol., **22**, 969-977.

The web site for GeneChip data

The GeneChip data of PCP are accessible at http://www.nihs.go.jp/tox/TtgPublished.htm

The Journal of Toxicological Sciences (J. Toxicol. Sci.)
Vol.38, No.2, 279-283, 2013

Letter

Neonatal exposure to 2,3,7,8-tetrachlorodibenze-p-dioxin increases the mRNA expression of prostatic proteins in C57BL mice

Nariaki Fujimoto[1], Atsuya Takagi[2] and Jun Kanno[2]

[1]Endocrine Research Group, Department of Disease Model, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan
[2]Division of Toxicology, National Institute of Health Sciences, Kamiyoga 1-18-1, Setagaya-ku, Tokyo 158-8501, Japan

(Received January 9, 2013; Accepted February 9, 2013)

ABSTRACT — The effects of neonatal exposure to low doses of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) on prostatic secretory protein expression were investigated. Male C57BL mice were treated with TCDD at 10, 100, or 1,000 ng/kg body weight at postnatal day (PND) 6. At PND42, the ventral, dorsolateral, and anterior prostatic lobes were dissected and the mRNA expression of prostatic proteins including spermine-binding protein, serine protease inhibitor Kazal type 3, prostate secretory protein 94 (PSP94), immunoglobulin binding protein-like protein (IgGBPLP), experimental autoimmune prostatitis antigen proteins, and peroxiredoxin-6 (Prdx6) was measured by quantitative PCR. There was no significant difference in the weight of the prostatic lobes between the control and TCDD-treated groups. The expression of PSP94 and Prdx6 in the ventral prostate and IgGBPLP in the dorsolateral prostate at PND42 was significantly increased by neonatal TCDD treatment in a dose-dependent manner, while no changes were noted in other prostatic secretions. These data suggest that neonatal exposure to TCDD may have effects on the neonatal differentiation of the prostate and results in the hyper-expression of some prostatic proteins later in life.

Key words: 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD), Prostatic secretion, Mouse prostate, Neonatal effects

INTRODUCTION

The developing male reproductive system of laboratory rodents is highly sensitive to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) (Mably *et al.*, 1992; Roman and Peterson, 1998; Theobald *et al.*, 2000). Its toxic effects include a decrease in the weight of the testis and accessory sex organs, degeneration of germ cells, and decreased spermatogenesis. The adverse effects of maternal exposure to TCDD on the development of the prostate gland have been studied extensively in rats and mice. In Holtzman rats, a single maternal dose of 64 ng/kg body weight (bw) of TCDD caused a significant decrease in ventral prostate (VP) weight (Mably *et al.*, 1992). More recently, it was reported that androgen receptor (AR) mRNA expression was reduced in the VP of Holtzman rats following maternal treatment with as low as 12.5 ng/kg bw TCDD (Ohsako *et al.*, 2001). In the mouse, the C57BL/6J strain appears to be sensitive to TCDD, in which the maternal administration of 5 μg/kg bw TCDD suppressed the development of the VP in the offspring, while the weight of the dorsolateral prostate (DLP) and anterior prostate (AP) decreased by approximately 50% (Lin *et al.*, 2002a). Exposure to TCDD during only the lactational period also resulted in offspring with lower prostate weights, but with less severe changes (Lin *et al.*, 2002b).

Although the previous studies have been clearly demonstrated that TCDD affect the development of the prostate morphologically, it is important to examine the effect on the prostatic function, production of prostatic proteins. We recently reported the identification of the major proteins secreted from the mouse prostate (Fujimoto *et al.*, 2006). The secreted proteins included spermine-binding protein (SBP), serine protease inhibitor Kazal type 3

Correspondence: Nariaki Fujimoto (E-mail: nfjm@hiroshima-u.ac.jp)

(SPI-KT3), prostate secretory protein 94 (PSP94), glucose-regulated protein, 78kDa (GRP78), peroxiredoxin-6 (Prdx6), probasin, experimental autoimmune prostatitis antigen protein (EAPA2), and immunoglobulin binding protein-like protein (IgGBPLP). The expression profile of these proteins would be useful for studying prostatic function and may also provide markers for evaluating the effects of environmental chemicals on the prostate. In the present study, we investigated the effects of neonatal exposure to low doses of TCDD on the mRNA expression of prostatic proteins as well as AR in the prostate.

MATERIALS AND METHODS

Animal experiments

The animal experiments were conducted under the approval of the Animal Experiment Committee of the National Institute of Health Sciences (NIHS). All experiments involving TCDD-treated animals were carried out following the rules for the use of TCDD set by NIHS. Five-day-old male C57BL mice were purchased from Charles River Japan Co. and maintained with free access to a basal diet and tap water. At postnatal day (PND) 6, the animals were divided into 4 groups (n = 6, each group): control and 3 TCDD-treated groups. TCDD (Cambridge Isotope Laboratories, Inc., Andover, MA, USA) in corn oil (50 μl) was injected intraperitoneally (ip) at doses of 0, 10, 100, or 1,000 ng/kg bw). At PND42, the animals were killed under ether anesthesia, since our previous study indicated that the mRNA expression of prostatic proteins is matured at PND42 (Fujimoto *et al.*, 2006). The prostatic lobes were dissected under a microscope, then immediately fixed in RNAlater Solution (Life technologies, Grand Island, NY, USA).

Quantification of mRNA by real-time RT-PCR

Total RNA was prepared from prostatic tissues using an RNA isolation kit (NucleoSpin RNA II; Machery-Nagel GmbH & Co. KG, Düren, Germany). An ABI Prism 7500 (Applied Biosystems/Life Technologies Co., Carlsbad, CA, USA) was employed for the RT-PCR based quantification of prostatic protein mRNAs as described previously (Fujimoto *et al.*, 2006). All mRNA levels were normalized with reference to β-actin mRNA.

Statistical analysis

Statistical comparisons were made by Dunnett's multiple comparison test.

RESULTS

Body and prostate lobe weights

There was no significant difference in body weight between the control and 3 TCDD-treated groups at PND42 (Table 1). There was no significant change in the weight of either the VP, DLP, or AP.

Expression of prostatic protein and AR mRNAs

SBP and SPI-KT3 were preferentially expressed in the VP, while probasin, EAPA2, and IgGBPLP expression was localized in the DLP and AP (Table 2). PSP94 was expressed in both the VP and DLP, while GRP78 and Prdx6 were expressed in all prostatic lobes. The effects of neonatal treatment of TCDD on mRNA expression were evident for PSP94, Prdx6, and IgGBPLP. The effects were lobe specific; that is, neonatal TCDD increased the expression of PSP94 and Prdx6 mRNA in the VP as well as IgGBPLP mRNA in the DLP in a dose-dependent manner. Neonatal TCDD exposure did not change the expression of AR mRNA in the VP or DLP, but decreased its expression in the AP.

DISCUSSION

Maternal exposure to TCDD reportedly causes irreversible changes to the reproductive systems of offspring, including reduced sperm count and reduced size of the reproductive organs. The development of the male reproductive organs in rodents, in particular the prostate gland, has been recognized as a sensitive target to

Table 1. Weight of body and prostatic lobes at PND42

Treatment	body weight (g)	VP (mg/g bw)	DLP (mg/g bw)	AP (mg/g bw)
control	16.8 ± 0.28	0.20 ± 0.03	0.21 ± 0.02	0.30 ± 0.02
TCDD 10	16.9 ± 0.37	0.19 ± 0.02	0.23 ± 0.01	0.22 ± 0.06
TCDD 100	17.7 ± 0.25	0.30 ± 0.08	0.20 ± 0.02	0.34 ± 0.03
TCDD 1000	18.9 ± 0.38	0.24 ± 0.08	0.24 ± 0.01	0.31 ± 0.03

Mean ± S.E.M. (n = 6). Male C57BL mice were treated with TCDD (10, 100, or 1,000 ng/kg bw) at postnatal day (PND) 6 and sacrificed at PND42.

Table 2. mRNA expression of prostatic proteins and AR in prostatic lobes

Treatment	SBP	SPI-KT3	PSP94	GRP78	Prdx6	Probasin	EAPA2	IgGBPLP	AR
VP									
control	315 ± 64.4	280 ± 85.7	12.4 ± 4.2	5.7 ± 0.36	1.5 ± 0.21				7.3 ± 1.24
TCDD 10	293 ± 131.2	495 ± 95.7	40.0 ± 11.9	7.6 ± 0.92	2.4 ± 0.38				8.6 ± 1.08
TCDD 100	322 ± 63.6	278 ± 44.1	65.7 ± 21.1*	8.3 ± 1.89	4.5 ± 0.79*				4.8 ± 0.27
TCDD 1000	450 ± 84.7	308 ± 46.5	110 ± 16.6**	7.1 ± 1.68	5.6 ± 1.0*				7.5 ± 0.71
DLP									
control			92.3 ± 20.2	11.2 ± 1.23	27.2 ± 4.7	7.8 ± 0.65	2.5 ± 0.44	0.80 ± 0.09	4.5 ± 0.96
TCDD 10			59.0 ± 24.3	9.5 ± 1.21	31.7 ± 1.93	9.1 ± 1.36	2.8 ± 0.44	2.2 ± 0.59	3.9 ± 0.45
TCDD 100			82.3 ± 20.7	8.9 ± 1.07	31.6 ± 3.02	8.5 ± 0.57	2.4 ± 0.29	2.7 ± 0.75*	3.5 ± 0.22
TCDD 1000			56.9 ± 15.9	7.3 ± 0.71	31.3 ± 2.46	7.9 ± 0.07	2.5 ± 0.28	3.3 ± 0.47*	3.2 ± 0.50
AP									
control				13.7 ± 3.23	38.1 ± 5.4	4.5 ± 0.66	3.4 ± 0.29	19.5 ± 1.95	3.6 ± 0.35
TCDD 10				18.0 ± 1.59	67.1 ± 4.1	5.8 ± 0.21	3.0 ± 0.33	27.8 ± 5.63	3.0 ± 0.31
TCDD 100				9.8 ± 0.91	40.9 ± 6.13	3.9 ± 0.52	3.0 ± 0.57	11.2 ± 1.90	1.9 ± 0.22
TCDD 1000				14.6 ± 1.90	73.6 ± 15.4	5.9 ± 0.75	3.3 ± 0.37	28.3 ± 6.19	2.1 ± 0.32

Mean ± S.E.M. (n = 5 or 6). Values are mRNA levels divided by beta actin mRNA levels (*$p < 0.05$ and **$p < 0.01$ vs. control).
Male C57BL mice were treated with TCDD (10, 100, or 1,000 ng/kg body weight) at postnatal day (PND) 6 and sacrificed at PND42.

TCDD, especially when it is administered maternally (Bjerke and Peterson, 1994; Lin *et al.*, 2002b; Mably *et al.*, 1992; Theobald *et al.*, 2000). In the present study, we examined the effects of neonatal TCDD exposure on the expression of prostatic proteins and demonstrated that the administration of low doses of TCDD at PND6 resulted in the abnormal hyper-expression of PSP94, Prdx6, and IgGBPLP mRNAs at PND42 in C57BL mice. Although the expression of all the prostatic proteins is regulated by androgen as we previously reported (Fujimoto *et al.*, 2006), only three of them were hyper-expressed, that may suggest the neonatal TCDD did not change androgen levels.

There is a difference in the acute lethality of TCDD among different mouse strains, with an LD50 of approximately 100 μg/kg bw in the "sensitive" C57BL/6J strain, while it is more than 3 mg/kg bw in "non-sensitive" DBA mice (Weber *et al.*, 1995). The C57BL/6J strain demonstrated a higher susceptibility to developmental disruption of the male reproductive system by maternal exposure to TCDD (Theobald *et al.*, 2000). In the rat, there are also great differences in the acute lethality of TCDD among strains, but the effects of TCDD on the development of the prostate seem to be similar among strains (Simanainen *et al.*, 2004).

The development of the prostatic gland begins with the formation of epithelial buds from the urogenital sinus at gestational day (GD) 17; these then develop into the prostatic main ducts (Cunha *et al.*, 1987). After birth, extensive branching and growth from the duct takes place to generate the mature prostate. Approximately 70-80% of ductal tips and 50-70% of branching points are formed during the first 15 days after birth, while ductal branching continues throughout adolescence (Sugimura *et al.*, 1986). Vulnerability to the effects of TCDD on the development of the prostate has been studied extensively in C57BL/6J mice, in which the oral administration of 5 μg/kg bw TCDD on GD 13 reduced VP weight by 84%. Lactational exposure alone to TCDD also significantly suppressed VP weight by 41%. For the DLP and AP, the effects were less severe, with lactational exposure alone reducing their weight by approximately 20%, while in utero exposure caused a 50% reduction (Lin *et al.*, 2002b). Our data may suggest that the prostate at PND 6 may be less susceptible for TCDD suppressing the prostatic growth but sensitive for the functional alteration. Further studies are needed to understand what timing of TCDD exposure is critical to lead to changes in expression of prostatic proteins.

Although prostatic secreted proteins are found in the seminal fluid, it is not clear what their physiological roles are. PSP94 is one of the major proteins secreted by the human prostate and is also abundantly secreted by the rodent prostate. PSP94 is known to be expressed mainly in the VP and DLP in mice. This protein may function as an immunoglobulin-binding protein and is involved in the regulation of the immune response in the female reproductive tract (Kamada *et al.*, 1998). It also functions as an inhibitor of sperm motility and of the acrosome reaction. IgGBPLP is abundantly expressed in the DLP and AP and may have a similar function to PSP94 (Kumar *et al.*, 2010). Prdx6 is an antioxidant enzyme that reduces peroxide and alkyl hydroperoxide to water and alcohol, respectively, and it may have a seminal plasma antioxidant capability (Wang *et al.*, 2004). The changes in the composition of prostatic secretions caused by the hyper-expression of these proteins might eventually affect normal fertility.

The present study demonstrated that neonatal exposure to low levels of TCDD changes the normal expression pattern of prostatic protein mRNAs later on in life, although it is not known whether these changes are physiologically detrimental. Previous studies have emphasized the suppressive effects of TCDD on the size of the prostate. However, the present study suggested that exposure to low doses of TCDD in the neonatal period may affect the expression patterns of prostate proteins.

ACKNOWLEDGMENTS

This research was partially supported by a Grant-in-Aid (H19-Kagaku-Ippan-003) from the Ministry of Health, Labour and Welfare, Japan.

REFERENCES

Bjerke, D.L. and Peterson, R.E. (1994): Reproductive toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in male rats: different effects of in utero versus lactational exposure. Toxicol. Appl. Pharmacol., **127**, 241-249.

Cunha, G.R., Donjacour, A.A., Cooke, P.S., Mee, S., Bigsby, R.M., Higgins, S.J. and Sugimura, Y. (1987): The endocrinology and developmental biology of the prostate. Endoc. Rev., **8**, 338-362.

Fujimoto, N., Akimoto, Y., Suzuki, T., Kitamura, S. and Ohta, S. (2006): Identification of prostatic-secreted proteins in mice by mass spectrometric analysis and evaluation of lobe-specific and androgen-dependent mRNA expression. J. Endo., **190**, 793-803.

Kamada, M., Mori, H., Maeda, N., Yamamoto, S., Kunimi, K., Takikawa, M., Maegawa, M., Aono, T., Futaki, S. and Koide, S.S. (1998): beta-Microseminoprotein/prostatic secretory protein is a member of immunoglobulin binding factor family. Biochem. Biophys. Acta, **1388**, 101-110.

Ko, K., Theobald, H.M. and Moore, R.W. and Peterson, R.E. (2004): Evidence that inhibited prostatic epithelial bud formation in 2,3,7,8-tetrachlorodibenzo-p-dioxin-exposed C57BL/6J fetal mice is not due to interruption of androgen signaling in the urogenital sinus. Toxic. Sci., **79**, 360-369.

Kumar, A., Jagtap, D.D., Mahale, S.D. and Kumar, M. (2010): Crystal structure of prostate secretory protein PSP94 shows an edge-to-edge association of two monomers to form a homodimer. J. Mol. Biol., **397**, 947-956.

Lin, T. m., Ko, K., Moore, R.W., Simanainen, U., Oberley, T.D. and Peterson, R.E. (2002a): Effects of aryl hydrocarbon receptor null mutation and in utero and lactational 2,3,7,8-tetrachlorodibenzo-p-dioxin exposure on prostate and seminal vesicle development in C57BL/6 mice. Toxic. Sci., **68**, 479-487.

Lin, T.M., Simanainen, U., Moore, R.W. and Peterson, R.E. (2002b): Critical windows of vulnerability for effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on prostate and seminal vesicle development in C57BL/6 mice. Toxic. Sci., **69**, 202-209.

Mably, T.A., Moore, R.W. and Peterson, R.E. (1992): In utero and lactational exposure of male rats to 2,3,7,8-tetrachlorodibenzo-p-dioxin. 1. Effects on androgenic status. Toxicol. Appl. Pharmacol., **114**, 97-107.

Ohsako, S., Miyabara, Y., Nishimura, N., Kurosawa, S., Sakaue, M., Ishimura, R., Sato, M., Takeda, K., Aoki, Y., Sone, H., Tohyama, C. and Yonemoto, J. (2001): Maternal exposure to a low dose of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) suppressed the development of reproductive organs of male rats: dose-dependent increase of mRNA levels of 5alpha-reductase type 2 in contrast to decrease of androgen receptor in the pubertal ventral prostate. Toxicol. Sci., **60**, 132-143.

Roman, B.L. and Peterson, R.E. (1998): In utero and lactational exposure of the male rat to 2,3,7,8-tetrachlorodibenzo-p-dioxin impairs prostate development. 1. Effects on gene expression. Toxicol. Appl. Pharmacol., **150**, 240-253.

Simanainen, U., Adamsson, A., Tuomisto, J.T., Miettinen, H. M., Toppari, J., Tuomisto, J. and Viluksela, M. (2004): Adult 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) exposure and effects on male reproductive organs in three differentially TCDD-susceptible rat lines. Toxic. Sci., **81**, 401-407.

Sugimura, Y., Cunha, G.R. and Donjacour, A.A. (1986): Morphogenesis of ductal networks in the mouse prostate. Biol. Reprod., **34**, 961-971.

Theobald, H.M., Roman, B.L., Lin, T.M., Ohtani, S., Chen, S.W. and Peterson, R.E. (2000): 2,3,7,8-Tetrachlorodibenzo-P-Dioxin Inhibits Luminal Cell Differentiation and Androgen Responsiveness of the Ventral Prostate Without Inhibiting Prostatic 5Alpha-Dihydrotestosterone Formation or Testicular Androgen Production in Rat Offspring. Toxic. Sci., **58**, 324-338.

Wang, M.H., Grossmann, M.E. and Young, C.Y.F. (2004): Forced expression of heat-shock protein 70 increases the secretion of Hsp70 and provides protection against tumour growth. Br. J. Cancer, **90**, 926-931.

Weber, L.W., Lebofsky, M., Stahl, B.U., Smith, S. and Rozman, K.K. (1995): Correlation between toxicity and effects on intermediary metabolism in 2,3,7,8-tetrachlorodibenzo-p-dioxin-treated male C57BL/6J and DBA/2J mice. Toxicol. Appl. Pharmacol., **131**, 155-162.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Harnessing Diversity towards the Reconstructing of Large Scale Gene Regulatory Networks

Takeshi Hase[1,2☯], **Samik Ghosh**[1,2☯*], **Ryota Yamanaka**[1,3], **Hiroaki Kitano**[1,2,4,5*]

1 The Systems Biology Institute, Shirokanedai, Minato, Tokyo, Japan, **2** Laboratory of Disease Systems Modeling, Center for Integrative Medical Sciences, RIKEN, Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, Japan, **3** Department of Genome Science, RCAST, The University of Tokyo, Komaba, Meguro, Tokyo, Japan, **4** Sony Computer Science Laboratories, Inc., Higashigotanda, Shinagawa, Tokyo, Japan, **5** Okinawa Institute of Science and Technology, Onna, Onna-son, Kunigami, Okinawa, Japan

Abstract

Elucidating gene regulatory network (GRN) from large scale experimental data remains a central challenge in systems biology. Recently, numerous techniques, particularly consensus driven approaches combining different algorithms, have become a potentially promising strategy to infer accurate GRNs. Here, we develop a novel consensus inference algorithm, Top*k*Net that can integrate multiple algorithms to infer GRNs. Comprehensive performance benchmarking on a cloud computing framework demonstrated that (i) a simple strategy to combine many algorithms does not always lead to performance improvement compared to the cost of consensus and (ii) Top*k*Net integrating only high-performance algorithms provide significant performance improvement compared to the best individual algorithms and community prediction. These results suggest that a priori determination of high-performance algorithms is a key to reconstruct an unknown regulatory network. Similarity among gene-expression datasets can be useful to determine potential optimal algorithms for reconstruction of unknown regulatory networks, i.e., if expression-data associated with known regulatory network is similar to that with unknown regulatory network, optimal algorithms determined for the known regulatory network can be repurposed to infer the unknown regulatory network. Based on this observation, we developed a quantitative measure of similarity among gene-expression datasets and demonstrated that, if similarity between the two expression datasets is high, Top*k*Net integrating algorithms that are optimal for known dataset perform well on the unknown dataset. The consensus framework, Top*k*Net, together with the similarity measure proposed in this study provides a powerful strategy towards harnessing the wisdom of the crowds in reconstruction of unknown regulatory networks.

Citation: Hase T, Ghosh S, Yamanaka R, Kitano H (2013) Harnessing Diversity towards the Reconstructing of Large Scale Gene Regulatory Networks. PLoS Comput Biol 9(11): e1003361. doi:10.1371/journal.pcbi.1003361

Editor: Andrey Rzhetsky, University of Chicago, United States of America

Received June 20, 2013; **Accepted** October 10, 2013; **Published** November 21, 2013

Funding: This work is, in part, supported by funding from the HD-Physiology Project of the Japan Society for the Promotion of Science (JSPS) to the Okinawa Institute of Science and Technology (OIST). Additional support is from a Canon Foundation Grant, the International Strategic Collaborative Research Program (BBSRC-JST) of the Japan Science and Technology Agency (JST), the Exploratory Research for Advanced Technology (ERATO) programme of JST to the Systems Biology Institute (SBI), a strategic cooperation partnership between the Luxembourg Centre for Systems Biomedicine and SBI, and from Toxicogenemics program of Ministry of Health, Labour and Welfare. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ghosh@sbi.jp (SG); kitano@sbi.jp (HK)

☯ These authors contributed equally to this work.

Introduction

Most genes do not exert their functions in isolation [1], but make their functions through regulations among them. Such regulatory interactions are in the same cell, between different cells, and even between different organs, forming large-scale gene regulatory networks (GRNs). The impact of genetic abnormality can spread through regulatory interactions in GRNs and alter the activity of other genes that do not have any genetic defects [2]. Analyses of GRNs are key to identify disease mechanisms and possible therapeutic targets for the future [1]. Therefore, reconstruction of accurate and comprehensive GRNs from genome-wide experimental data (*e.g.*, gene expression data from DNA microarray experiments) is one of the fundamental challenges in systems biology [3,4].

A plethora of algorithms have been developed to infer GRNs from gene expression data, *i.e.*, mutual-information (MI) based algorithms [5–12], correlation-based algorithms [5], Bayesian networks (BNs) [13–17], regression-based algorithms [18–22], graphical gaussian model (ggm) [23], meta predictors that combine several different methods [24,25], and several other approaches that were recently proposed [26–32], *i.e.*, random forests based algorithm [26] (GENIE3) and two-way ANOVA based algorithm [27] (ANOVA). Each network-inference algorithm generates a confidence score for a link between two genes from expression data and assumes that a predicted link with higher confidence score is more reliable. Systematic and comparative assessment of the performance of these inference algorithms remains a major challenge in network reconstruction [33].

Several studies compared performances of the network-inference algorithms [7,8,9,34]. Especially, the DREAM5 (Dialogue on Reverse Engineering Assessment and Methods) challenge evaluated performances of many and diverse network-inference algorithms (29 algorithms submitted by challenge participants and 6 commonly used "off-the-shelf" algorithms) by using benchmark dataset composed of large-scale *Escherichia coli*, *Saccharomyces cerevisiae*, and in silico regulatory networks and their corresponding expression datasets [35]. The evaluation demon-

Author Summary

Elucidating gene regulatory networks is crucial to understand disease mechanisms at the system level. A large number of algorithms have been developed to infer gene regulatory networks from gene-expression datasets. If you remember the success of IBM's Watson in "Jeopardy!" quiz show, the critical features of Watson were the use of very large numbers of heterogeneous algorithms generating various hypotheses and to select one of which as the answer. We took similar approach, "Top*k*Net", to see if "Wisdom of Crowd" approach can be applied for network reconstruction. We discovered that "Wisdom of Crowd" is a powerful approach where integration of optimal algorithms for a given dataset can achieve better results than the best individual algorithm. However, such an analysis begs the question "How to choose optimal algorithms for a given dataset?" We found that similarity among gene-expression datasets is a key to select optimal algorithms, i.e., if dataset A for which optimal algorithms are known is similar to dataset B, the optimal algorithms for dataset A may be also optimal for dataset B. Thus, our "Top*k*Net" together with similarity measure among datasets can provide a powerful strategy towards harnessing "Wisdom of Crowd" in high-quality reconstruction of gene regulatory networks.

strated that no single individual algorithm performs optimally across all the three expression-datasets, *i.e.*, GENIE3 and ANOVA perform optimally for *E. coli* dataset, while two algorithms based on regression techniques are optimal for in silico dataset. Further, algorithm-specific biases influence the recovery of different regulation patterns, *i.e.*, MI and correlation based algorithms can recover feed-forward loops most reliably, while regression and BNs can more accurately recover linear cascades than MI and correlation based algorithms [35].

Above observations suggest that different network-inference algorithms have different strengths and weaknesses [33,35]. A natural corollary to the observations is that combining multiple network-inference algorithms may be a good strategy to infer an accurate and comprehensive GRN. Recently, Marbach et al. proposed a new network-inference algorithm, "Community Prediction", by combining several network-inference algorithms that were submitted to DREAM5 challenge [35]. The Community Prediction combining 29 algorithms ("off-the-self" algorithms are not used) shows higher or at least comparable performance to the best among the 29 algorithms across all DREAM5 datasets. Further, performance of community prediction increases as the number of integrated algorithms increases. Thus, community prediction based on integration of many algorithms can be a robust approach to infer GRNs across diverse datasets and will provide a powerful framework to reconstruct unknown regulatory networks.

Analysis of DREAM5 results [35] reveal that algorithms complement each other in a context-specific manner and harnessing the combined strengths and weaknesses of diverse techniques can lead to high quality inference networks. Thus, it is important to analyze the anatomy of diversity and quantify it. This is particularly important to systematically evaluate the characteristics of individual techniques and leverage their diversity in finding an optimal combination set for a specific experimental data context. Recently, Marbach et al. showed that integration of algorithms with high-diversity outperform that with low-diversity [35]. However, their diversity analysis is qualitative and, to our knowledge, there is no measure to quantify algorithm diversity.

Analysis of small in silico datasets of the DREAM3 challenge demonstrated that integration of the best five algorithms outperforms integration of all algorithms submitted to the challenge [33]. Selection of optimal algorithms for a given expression data and integration of the selected optimal algorithms may be more powerful strategy to reconstruct accurate GRNs than using many algorithms. Development of a method to determine optimal algorithms is a key to reconstruct accurate and comprehensive GRNs, although it is difficult to identify beforehand optimal algorithms for reconstruction of an unknown regulatory network because of biological and experimental variations among expression datasets.

A measure to quantify similarity among gene-expression datasets could be a clue to determine the optimal algorithms for reconstruction of unknown regulatory networks. This is because, if expression-data associated with known regulatory network (e.g., the DREAM5 datasets) is similar to that with unknown regulatory network, optimal algorithms for data with known regulatory network could be also optimal for data with unknown regulatory network.

Motivated by the above observations and issues, this paper focuses on four strategies towards building a comprehensive network reconstruction platform –

- A computational framework to integrate diverse inference algorithms.
- Systematically assess the performance of the framework against the DREAM5 datasets composed of genome-wide transcriptional regulatory networks and their corresponding expression data from actual microarray experiments as well as in silico simulation.
- Develop a measure to quantify diversity among inference techniques towards identifying optimal combination of algorithms which elucidate accurate GRNs.
- Develop a measure to quantify similarity among expression datasets towards selecting optimal algorithms for reconstruction of unknown regulatory networks.

To investigate these possible strategies, we first develop a novel network-inference algorithm that can combine multiple network-inference algorithms. Second, to evaluate inference performances of the algorithms precisely, we used the DREAM5 datasets composed of *E. coli* and *S. cerevisiae* transcriptional regulatory networks and their corresponding expression data from large-scale microarray experiments, together with synthetic network and corresponding expression datasets (http://wiki.c2b2.columbia.edu/dream/index.php/D5c4). A cloud-based computing framework was developed on the Amazon Web Services (AWS) system to systematically benchmark the large data-sets and compute-intensive algorithms. Third, we define a mathematical function quantifying diversity between algorithm pairs to analyze the anatomy of diversity and its role in improving the performance of reverse engineering techniques. Finally, we present a similarity measure among expression-datasets and its potential to identify optimal algorithms for reconstruction of unknown regulatory networks.

Results

We developed a computational workflow for the combination of network-inference algorithms and systematic assessment of their performance. The workflow of our framework is composed of three steps (see Supplementary figure S1 and Materials and Methods for details):

I. **Inference Methods**: We obtained confidence score between gene pairs based on 29 algorithms submitted by DREAM5 participants and 6 commonly used "off the shelf" algorithms from Marbach et al [35]. Furthermore, we calculated confidence score between two genes based on other three algorithms, *i.e.*, c3net [9], ggm [23], and mrnet [8] algorithms. We used, in total, 38 network-inference algorithms for the study.

II. **Top*k*Net**: A novel algorithm to generate predicted list of regulatory links by the network-inference algorithms that can combine multiple network-inference algorithms (in this case, 38 algorithms) (see Figure 1).

III. **Performance Assessment**: Comparative evaluation of the performance of Top*k*Net with that of the 38 network-inference algorithms and Community Prediction, benchmarked using the DREAM5 network-inference challenge dataset composed of the large synthetic data (number of genes = 1,643 and sample size = 805), large-scale *E. coli* and *S. celevisiae* networks (number of gene = 4,511 and 5,950, respectively), and their corresponding real microarray gene expression data (with sample size of 805 and 536, respectively). Table 1 summarizes the different data-sets employed in the performance assessment for this study. We used a cloud-computing infrastructure built on Amazon EC2 instances to infer GRNs from the large-scale DREAM5 dataset (see Materials and Methods for details).

Top*k*Net (the maximum value of k is the number of integrated algorithms, 38 in this case) is based on leveraging the diversity of the different techniques by combining the confidence of each gene pair interaction computed by the algorithms. Top*k*Net applied a bagging method, which was introduced by Breiman L [36], to combine confidence scores between each gene pair from multiple network inference algorithms. Top1Net assumes that two genes have a regulatory links between them, if at least one network-inference algorithm assigns high confidence level to the link between them, while Top38Net assumes that two genes have a regulatory link between them, only if all the 38 algorithms assign high confidence levels to the link between them (see below and Figure 1 for details). Top*k*Net with $k = 2$–37 and Community prediction (which takes the average of ranks assigned by different algorithms), are intermediates between Top1Net and Top38Net.

Figure 1 gives an illustrative example of **Top*k*Net** (for simplicity, in this illustrative example, we used 5 individual algorithms) on a sample target network (Figure 1A). As shown in Figure 1B, a network-inference algorithm assigns a confidence level to each link and links are ranked according to their confidence levels, *i.e.*, a link with higher confidence level has higher rank value. For each link, 5 individual network-inference algorithms (represented by different colors) assigned 5 rank values to each link (Figure 1C). Among the five rank values of each link, **Top*k*Net** regards *k*th highest rank value as the rank value of the link (Figure 1D). For example, five rank values (1, 3.5, 10, 10.5, and 12) are assigned to the regulatory link between nodes 1 and 2 (Figure 1C). In this case, **Top1Net** and **Top2Net** regards 1 and 3.5 as the rank value of the link, respectively (see Figure 1D). As shown in color spectrums in Figure 1D, **Top*k*Net** algorithms reconstruct GRNs which include predicted connections from multiple algorithms.

Based on the observation that network-inference algorithms tend to assign high confidence levels to true-positive links [6,34], **Top1Net** algorithm would infer the largest number of true-positive links among all algorithms, *i.e.*, **Top*k*Net** algorithms would infer smaller number of true-positive links as the value of k increases and, at the same time, can avoid inferring false-positive links. Thus, in general, **Top1Net** would outperform other algorithms in terms of inference performances, as seen in the predicted network in Figure 1.

Comparative performance assessment

Network inference algorithms have increased following Moore's Law (doubling every two years) [33,37]. Consequently, it has become increasing important to develop comprehensive performance benchmarking platforms to compare their relative strengths and weaknesses and leverage them to improve quality of inferred network. Two key components fundamental to performance assessment are representative metrics to quantify performance and standardized data sets on which to evaluate them.

In this section, we first outline these components employed in this study on the basis of which the performance of Top*k*Net is evaluated.

Benchmarking data sets. The performance of gene network reconstruction algorithms require benchmarking against various data sets representing network dynamics (for example, gene expression profiles) for which the underlying network is known. However, the ability to generate biologically plausible networks and validate them against experimental data remains a fundamental tenet in network reconstruction. In this respect, the DREAM initiative provides a community platform for the objective assessment of inference methods. The DREAM challenges provide a common framework on which to evaluate inference techniques against well-characterized data sets. In this study, we used large scale experimental data from the DREAM5 network inference challenge.

Performance metrics. True-positive rate, false-positive rate, recall, and precision are representative metrics to evaluate performances of network inference algorithms (see Materials and Methods for details). True-positive (false-positive) rate is, among all true (false) links, how many of them are with ranks beyond a threshold. Precision is, among all links with ranks beyond a threshold, how many of them are true links, while recall is, among all true links, how many of them are with ranks beyond a threshold. For example, in Figure 1, Top2Net for threshold of 3.5 shows recall = 6/7 (one missing) and precision = 6/7 (one mistake). By changing threshold gradually, we obtained a receiver operating characteristic (ROC) and Precision/recall (PR) curves that are graphical plots of true-positive rate vs. false-positive rate and precision vs. recall, respectively. These curves are straightforward visual representation of performances of network-inference algorithms.

Further, we calculated three representative metrics, *i.e.*, AUC-PR (area under the PR curve), AUC-ROC (area under the ROC curve), and max f-score (f-score is harmonic mean of precision and recall) for these algorithms (see Materials and Methods for details). AUC-PR and AUC-ROC evaluate the average performances of network-inference algorithms, while max f-score evaluates the optimal performance of network-inference algorithms. A network-inference algorithm with higher inference performance would show higher AUC-PR, AUC-ROC, and max f-score. Moreover, DREAM5 also provides performance benchmarking package which computes an overall score (OS) across the entire dataset [35]. By using the package, we also calculated overall score to evaluate performance of Top*k*Net algorithms against community prediction and 38 algorithms (for performance of the individual 38 algorithms, see Supplementary table S1).

TopkNet performance on DREAM5 data set

To evaluate how Top*k*Net leverages diversity amongst the candidate algorithms to infer consensus network, we used the

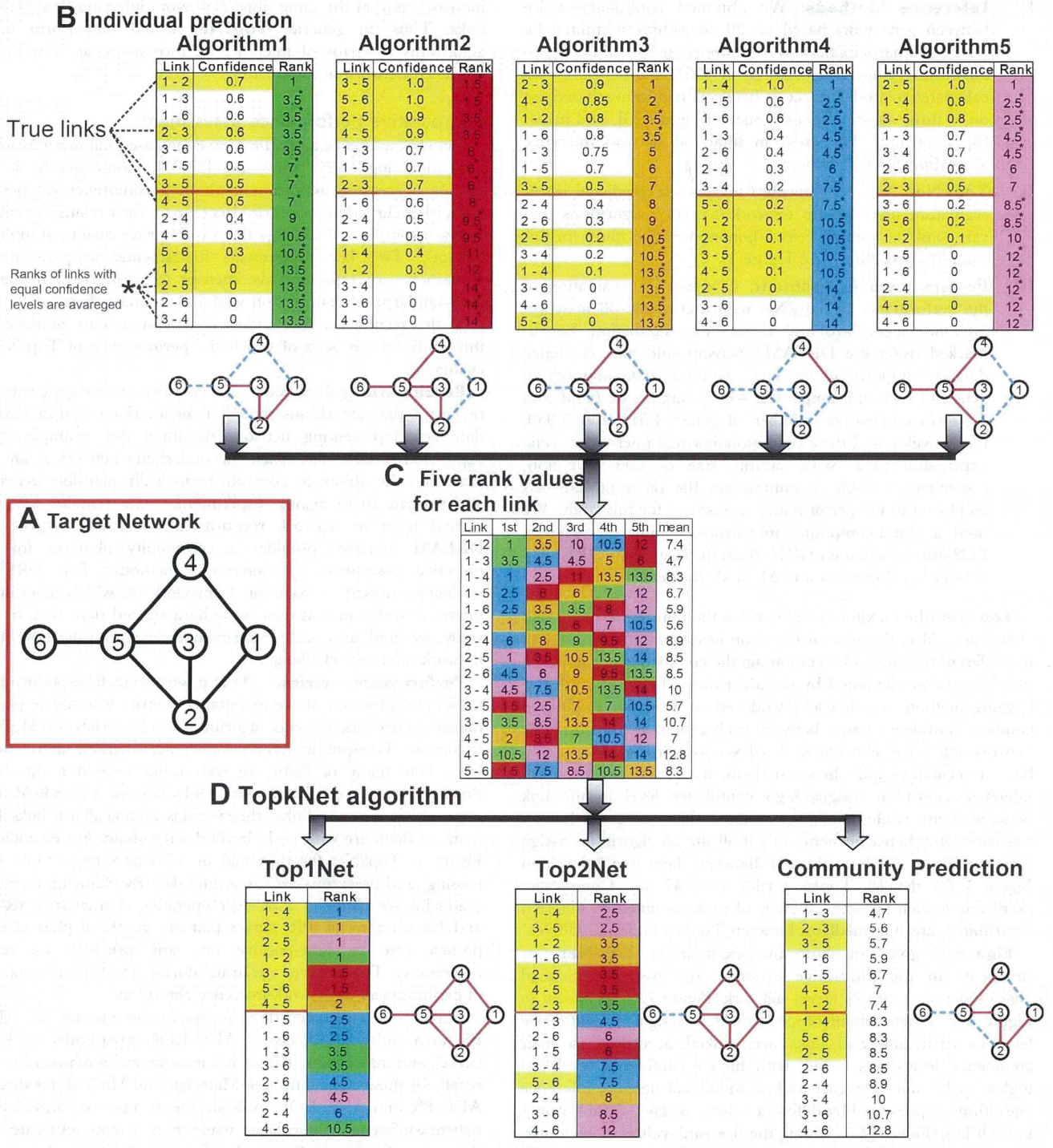

Figure 1. Example of a prediction by Top*k*Net formed from five individual network predictions. (**a**) Target Network. Circles and links are genes and regulatory links among genes, respectively. (**b**) The five lists are ranked according to the confidence levels of links, the most reliable prediction is at the top of the list and has the highest rank, *i.e.*, Algorithm1 assigns the highest confidence level and the rank value of 1 to a link between nodes 1 and 2. The true link of the target network is highlighted in yellow. We regard links with rank of 1–7 as regulatory links inferred by the algorithms because the target network composed of 7 links. Red lines and blue dashed lines represent true positive and false negative links, respectively. (**c**) Five rank values for each link and the mean value among the five values. Green, red, orange, blue, and purple represent rank values from Algorithm1, Algorithm2, Algorithm3, Algorithm4, and Algorithm5, respectively. (**d**) Rank value of a link by Top*k*Net and that by Community Prediction. Top1Net and Top2Net regards 1st and 2nd highest value among five rank values for a link as the rank value of the link, respectively. Community Prediction calculates the mean value among five rank values for a link and regards the mean as the rank of the link. For example, rank of the links between genes 1 and 2 for Community Prediction is 7.4. This example illustrates how Top1Net can be more accurate than the other algorithms.

doi:10.1371/journal.pcbi.1003361.g001

Table 1. The DREAM5 datasets used in this study.

-	In silico[1]	*E. coli*[2]	*S. celevisiae*[3]
Number of genes	1,643	4,511	5,950
Number of samples	805	805	536

[1]In silico Dream 5 dataset.
[2]Dream 5 dataset from *E.coli*.
[3]Dream5 dataset from *S.cerevisiae*.
doi:10.1371/journal.pcbi.1003361.t001

DREAM5 benchmarking data comprised of large scale synthetic and experimental gene regulatory networks for *E.coli* and *S.cerevesiae* as outlined in Table 1 and computed different performance metrics on them. As seen in the PR and ROC curves for in silico, *E. coli*, and *S. cerevisiae* datasets (Supplementary figures S2 and S3), Top6net shows constantly higher performance compared to community prediction. Other three performance metrics (AUC-PR, AUC-ROC, and Max f-score) of Top*k*Net with $k = 5$–8 are also higher than those of community prediction for all the three datasets (see Figures 2B–J). Thus, overall score of Top*k*Net with $k = 5$–8 is significantly higher than that of community prediction (see Figure 2B). However, the performance metrics of Top*k*Net is only comparable to the best individual algorithms and not significantly better. Community prediction also showed significantly lower performances than the best individual algorithm.

These results indicate that, while Top*k*Net would provide better strategy to integrate multiple algorithms than community prediction, such a strategy does not always significant increase in performance compared to the cost of integration. As seen in this section, the overall score of the best individual algorithm (40.279) is comparable to that of Top*k*Net with $k = 5$–7 (40.110–41.251) and is much higher than that of community prediction and TopkNet with $k = 1$ (30.228 and 10.432, respectively). This is because, for the DREAM5 datasets, several low-performance algorithms assign high confidence scores to many false-positive links and such false links could decrease the performance of Top*k*Net (especially, with $k = 1$) and community prediction algorithms.

Thus, by integrating only high-performance algorithms that tend to assign high confidence score to true-positive link, Top*k*Net (especially, with $k = 1$) and community prediction may show much higher performances than the best individual algorithms. To investigate this issue, we evaluate Top*k*Net (and community prediction) based on integration of 10 optimal algorithms (algorithms within top 10 highest AUC-PR) for each of the in silico, *E. coli.* and *S. cerevisiae* datasets. As seen in Supplementary figures S4 and S5, PR and ROC curves of Top1Net are constantly over those of the best individual algorithm and community prediction for in silico and *E. coli* datasets, although, for *S. cerevisiae*, PR-curve of the best individual algorithm slightly over that of Top1Net. Other three metrics (AUC-PR, AUC-ROC, and Max f-score) of Top*k*Net with low k ($k = 1$ for in silico and *E. coli* and $k = 2$ for *S. cerevisiae*) are significantly higher or at least comparable to those of the best individual algorithm and community prediction (see Figures 3B–J). Therefore, the overall score of Top*k*Net with $k = 1$ and 2 (74.935 and 73.261, respectively) are significantly higher than that of the best individual algorithm (40.279) and community prediction (56.158) (see Figure 3A). These results highlight that integration of multiple high-performance algorithms by Top1Net or Top2Net consistently reconstructs the most accurate GRNs for different datasets.

As demonstrated in this section, selection of optimal algorithms for a given expression data and Top1Net, Top2Net, and community prediction based on integration of the selected optimal algorithms could be a powerful approach to reconstruct high-quality GRNs. However, currently, to our knowledge, there is no method to determine beforehand optimal algorithms for expression data associated with an unknown regulatory network. Development of a method to determine optimal algorithms is a key to reconstruct unknown regulatory networks (We investigate this issue in the next section).

Selection of optimal algorithm pairs to infer GRNs based on algorithm diversity

Different network-inference algorithms employ different and often complementary techniques to infer gene regulatory interactions from an expression dataset. Therefore, a consensus driven approach, which leverages *diversity* in network-inference algorithms, can infer more accurate and comprehensive GRNs than a single network-inference algorithm. However, as demonstrated in this study, a simple strategy of increasing the number of algorithms may not always yield significant performance gains compared to the *cost of consensus*, *i.e.*, the computation cost (CPU time and memory usage).

It is pertinent to analyze the anatomy of diversity between different algorithms in a theoretical framework to answer the questions of -

- To what extent, then, are the algorithms different from each other?
- Does bringing diversity of the algorithms into community prediction improve the quality of inferred networks?

For the purposes, Marbach et al. conducted principal component analysis (PCA) on confidence scores from 35 network-inference algorithms [35]. They mapped 35 algorithms onto 2nd and 3rd principal components and grouped the algorithms into four clusters by visual inspection. The analysis demonstrated that integration of three algorithms from different clusters shows higher performance than that from the same cluster. It indicates that the diversity signature of the selected algorithms, and not just the number of algorithms, plays an important role in the performance of the network reconstruction techniques.

However, their algorithm diversity is qualitative and, to our knowledge, there is no quantitative measure for algorithm diversity. In order to quantify diversity among the individual algorithms employed in this study, we developed two quantitative measures of diversity which calculates distance between algorithms pairs on the basis of confidence scores of regulatory interactions inferred by the algorithms. One is based on simple Euclidean distance (EUC distance) and the other is based on EUC distance on 2nd and 3rd components from PCA analysis (PCA distance) (see Materials and Methods for details). In Figure 4, we provide a toy model to explain how diversity among network-inference algorithms is defined.

By using the diversity measures, we calculated distance among 10 optimal algorithms for each of the DREAM5 datasets to examine whether bringing quantified algorithm diversity into Top1Net (and Community prediction) improves the performances of network reconstruction. Based on the calculated distances, we defined high-diversity pairs as top 10% of algorithm pairs with highest distance, while low-diversity pairs are defined as bottom 10% of algorithm pairs with lowest distance. In this study, we have 45 algorithm pairs among 10 optimal algorithms and thus top 5 algorithm pairs with highest distance are high-diversity pairs, while

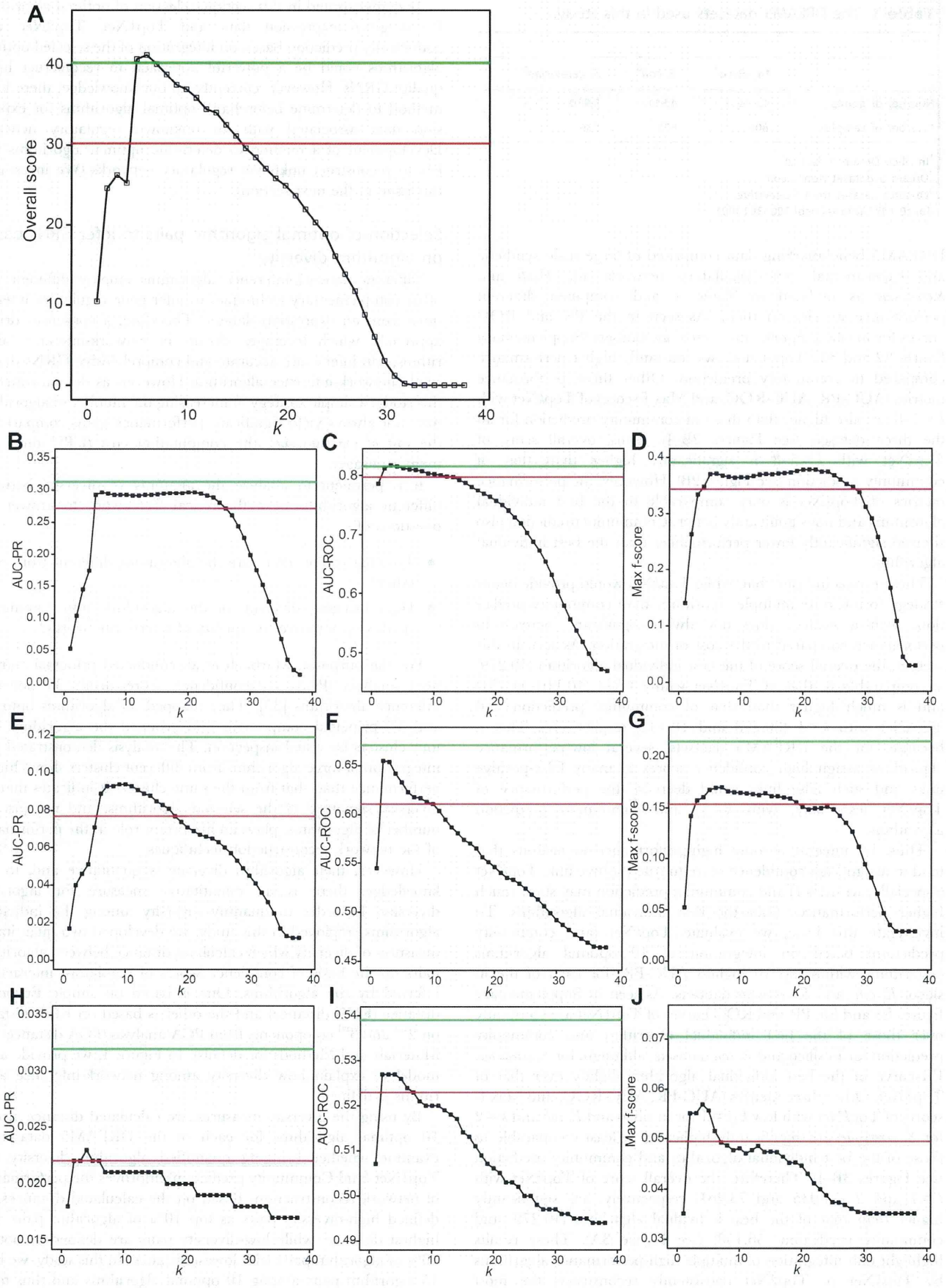
A
Overall score
k
B
AUC-PR
C
AUC-ROC
D
Max f-score
E
AUC-PR
F
AUC-ROC
G
Max f-score
H
AUC-PR
I
AUC-ROC
J
Max f-score

Figure 2. Performances of Top*k*Net and community prediction based on integration of the 38 network-inference algorithms. Black squares and lines show performances of Top*k*Net algorithm. For example, values at $k=1$ represent performances of Top1Net algorithm. Red and green lines represent performances of community prediction and those of the best algorithm, respectively. (**A**) Overall score. (**B**) AUC-PR for in silico dataset. (**C**) AUC-ROC for in silico dataset. (**D**) Max f-score for in silico dataset. (**E**) AUC-PR for *E. coli* dataset. (**F**) AUC-ROC for *E. coli* dataset. (**G**) Max f-score for *E. coli* dataset. (**H**) AUC-PR for *S. cerevisiae* dataset. (**I**) AUC-ROC for *S. cerevisiae* dataset. (**J**) Max f-score for *S. cerevisiae* dataset.
doi:10.1371/journal.pcbi.1003361.g002

bottom 5 algorithm pairs with lowest distance are low-diversity pairs.

Next, we evaluated the performances of Top1Net (or community prediction) based on integration of high-diversity pairs and those of low-diversity pair. As seen in Figures 5B–J and Supplementary figures S6B–J, AUC-PR, AUC-ROC, and max f-score of high-diversity pairs by EUC distance are higher or at least comparable to those of low-diversity pairs by EUC distance across all datasets. Especially, for *in silico* and *E. coli* datasets, AUC-PR and Max f-score of high-diversity pairs by EUC distance are significantly higher than that of low-diversity pairs by EUC distance. Thus, the overall score of high-diversity pairs is also significantly higher than that of low-diversity pair ($P<0.05$) (see Figure 5A and Supplementary figure S6A). The performances of high-diversity pairs by PCA distance are also higher or at least comparable to those of low-diversity pairs by PCA distance (see Supplementary figures S7 and S8). Furthermore, median value of the overall score of high-diversity pairs (47.725 and 50.250 by Top1Net, for EUC and PCA distances, respectively) are much higher than that by the best individual algorithms (40.279) and that by community prediction that integrates 38 network-inference algorithms (30.228). In summary, these results indicate that -,

- Even for the same number of algorithms (in this case, two algorithms are integrated), the quantitative diversity to selected pairs can improve the performance of the consensus methods (Top*k*Net and community prediction).
- Quantitative diversity-guided consensus can reduce the cost of consensus (only 2 algorithms integration instead of 38 algorithms integration in this case) without compromising the quality of the inferred network as shown in this study where the inference performance of high diversity pair is much higher than that of 38 algorithms combination.

Selection of optimal algorithms based on similarity among expression datasets towards reliable reconstruction of regulatory networks

Top1Net or Top2Net based on integration of highest-performance algorithms consistently reconstruct the most accurate GRNs, as demonstrated in the previous section (see Figure 3). However, as Marbach et al. mentioned, "*Given the biological variation among organisms and the experimental variation among gene-expression datasets, it is difficult to determine beforehand which methods will perform optimally for reconstruction an unknown regulatory network*" [35], and, to our knowledge, there is no method to select the optimal network-inference algorithms. Development of a method to select optimal network-inference algorithms for each of the expression datasets remains a major challenge in network reconstruction.

A measure to quantify similarity among expression datasets can be a key to select optimal network-inference algorithms for each of the datasets, because, if similarity between expression-data associated with known regulatory network (*e.g.*, DREAM5 datasets) and that with unknown regulatory network is high, optimal algorithms for the known dataset can be repurposed to infer regulatory network from unknown dataset. Driven by this observation, we developed a similarity measure among gene-expression datasets based on algorithm diversity proposed in previous section.

First, we briefly explain the overview of the procedure to calculate similarity among expression datasets (see Figure 6 and Materials and Methods for the details). The procedure is composed of 4 steps. (1) The expression datasets were split into a dataset for which optimal algorithms are unknown (e.g, Data1 in Figure 6A) and datasets for which optimal algorithms are known (e.g., Data2 and Data3 in Figure 6A). (2) For each of the datasets, confidence scores of links were calculated by network-inference algorithms. In the example shown in Figure 6B, each of 5 algorithms calculates 6 confidence scores for 6 links. (3) By using the confidence scores calculated in the step (2), diversity among algorithms was calculated based on a distance measure proposed in the previous section (EUC and PCA based distances, see Figure 6C and Materials and Methods for details), for each of the datasets. In the example shown in Figure 6C, we have 10 algorithm pairs among 5 algorithms and thus, as shown in matrices in the figure, we have 10 distances between two algorithms for each of the three datasets. (4) By using algorithm diversity calculated in the step (3), we calculated correlation coefficient of the algorithm distances between two datasets (see Figure 6D). In terms of algorithm diversity, the correlation coefficient is regarded as similarity measure between the two datasets. In the example shown in Figure 6D, Data1 is more similar to Data2 than Data3. Thus, optimal algorithms for Data2 are better fit than those for Data3 to infer GRN from Data1.

Next, to evaluate whether dataset similarity can be used to govern optimal selection of inference algorithms, we calculated the similarity among the DREAM5 gene-expression datasets and compared the performance of the algorithms across the datasets. As seen in Figure 7 and Table 2, correlation between *S. cerevisiae* and *E. coli* datasets (Spearman's correlation coefficient $\rho = 0.99$) is much higher than that between *E. coli* and *in silico* ($\rho = 0.87$ and 0.81 by EUC and PCA distances, respectively) and that between *S. cerevisiae* and *in silico* ($\rho = 0.83$). In terms of algorithm diversity, similarity between *E. coli* and *S. cerevisiae* datasets is much higher than that between *E. coli* and *in silico* and that between *S. cerevisiae* and *in silico*.

Further correlation of algorithm performances between dataset pair with high similarity (*e.g.*, *E. coli* and *S. cerevisiae* pair) is higher than that between dataset pair with low similarity (*e.g.*, *in silico* and *E. coli* pair and *in silico* and *S. cerevisiae* pair) (see Supplementary figure S7 and Supplementary Table S2). These results indicate that, for dataset pair with high similarity, optimal network-inference algorithms for one dataset also tend to be optimal for the other dataset.

From above observations (observations in Figure 7, Supplementary figure S9, Table 2, and Supplementary table S2), we hypothesized that, if similarity between the two expression-datasets is high, integration of algorithms that are optimal for one dataset could perform well on the other dataset. To examine this issue in more detail, we integrated algorithms that are optimal for *S. cerevisia* dataset (algorithms with 10 highest AUC-PR values on the dataset) and those for the in silico dataset and evaluated their performance of these two integrations against *E. coli* dataset.

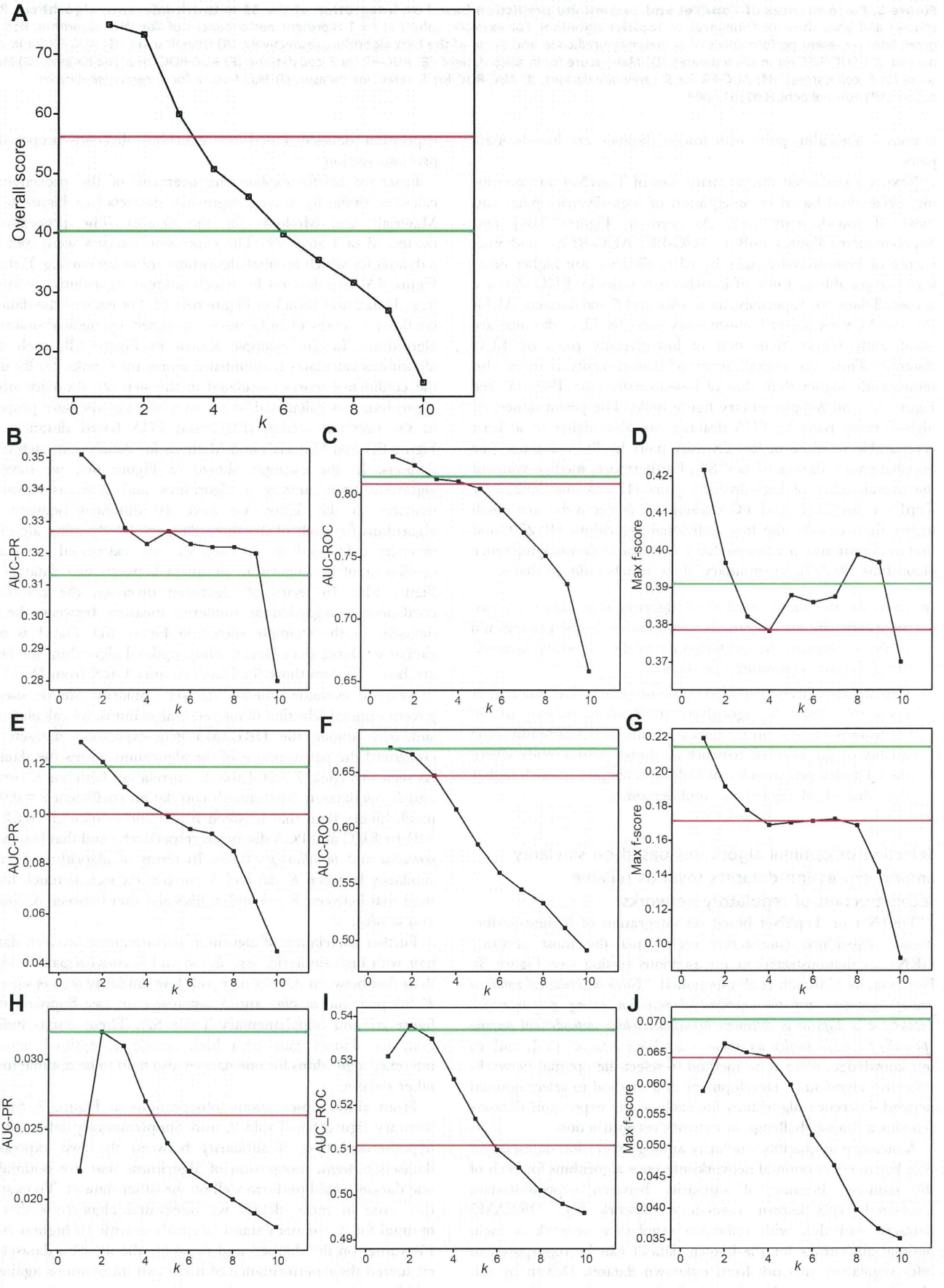
A
Overall score
k
B
AUC-PR
k
C
AUC-ROC
k
D
Max f-score
k
E
AUC-PR
k
F
AUC-ROC
k
G
Max f-score
k
H
AUC-PR
k
I
AUC_ROC
k
J
Max f-score
k

Figure 3. Performances of Top*k*Net and Community prediction based on integration of top 10 highest-performance algorithms. Black squares and lines show performances of Top*k*Net algorithm. For example, values at $k = 1$ represent performances of Top1Net algorithm. Red and green lines represent performances of community prediction and those of the best algorithm, respectively. (**A**) Overall score. (**B**) AUC-PR for in silico dataset. (**C**) AUC-ROC for in silico dataset. (**D**) Max f-score for in silico dataset. (**E**) AUC-PR for *E. coli* dataset. (**F**) AUC-ROC for *E. coli* dataset. (**G**) Max f-score for *E. coli* dataset. (**H**) AUC-PR for *S. cerevisiae* dataset. (**I**) AUC-ROC for *S. cerevisiae* dataset. (**J**) Max f-score for *S. cerevisiae* dataset.
doi:10.1371/journal.pcbi.1003361.g003

As seen in Figures 8A, B, and C, against the *E.coli* dataset, performances (AUC-PR, AUC-ROC, and max f-score) of optimal integration from *S. cerevisiae* dataset (green lines) are generally higher than those from in silico dataset (red lines). Further, against the *S. cerevisiae* dataset, we evaluate performances of optimal-algorithm integration from *E. coli* dataset and that for in silico dataset and found that optimal integration from *E. coli* dataset (green lines) generally outperform that from in silico dataset (red lines) (see Figures 8D, E, and F). Because similarity between *S. cerevisiae* and *E. coli* datasets are much higher than that between *E. coli* and in silico datasets and that between *S. cerevisiae* and in silico datasets (see Figure 7 and Table 2), these results support the above hypothesis.

Further, as shown in Figure 8, performance of Top*k*net integrating optimal algorithms from a dataset with high-similarity (green lines) is comparable to that integrating top 10 highest-performance algorithms (blue lines). Thus, data-similarity based optimal algorithm selection together with Top*k*Net (or community prediction) based integration of the selected optimal algorithms can be a powerful strategy to reconstruct unknown regulatory network.

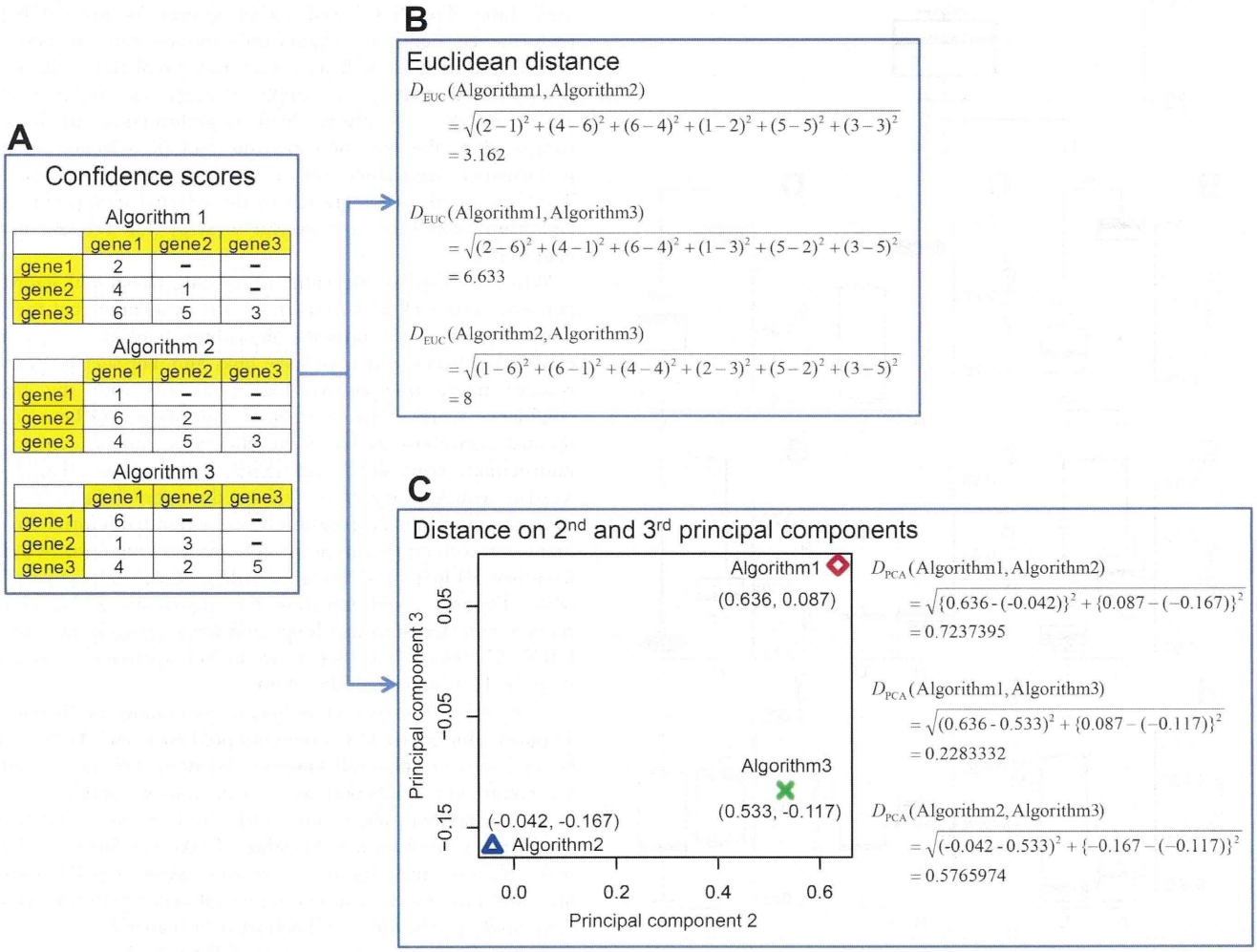

Figure 4. A toy example to calculate diversity among algorithms. (**A**) Confidence scores from algorithms. Confidence score of a link between two genes were generated by each of three algorithms. In this case, each algorithm has 6 confidence scores for 6 links. Note that the three algorithms in this example are algorithms to infer non-directional algorithms and make symmetrical matrices of confidence scores, i.e., confidence score of link from gene1 to gene 2 is same as that from gene2 to gene1. Thus, for simplicity, upper triangles of confidence score matrices are not shown in the figure. (**B**) Diversity among algorithms based on Euclidean distances. In this example, each of three algorithms has a vector of 6 confidence scores for 6 links between two genes. Euclidean distance between two vectors of confidence scores from two algorithms is calculated and is defined as diversity between the two algorithms. (**C**) Diversity among algorithms based on 2nd and 3rd components of PCA analysis. In this example, PCA analysis is conducted on three vectors of 6 confidence scores from three network-inference algorithms and the three algorithms are mapped on to 2nd and 3rd principal components (see left panel of C). Euclidean distance between two algorithms is calculated by using the 2nd and 3rd principal components and is defined as diversity between the two algorithms.
doi:10.1371/journal.pcbi.1003361.g004