mesenchymal marker) is known to be representative of breast CSCs (Al-Hajj et al. 2003). Interestingly, activation of the EMT program by the induction of Snail or Twist genes or addition of recombinant TGF-β resulted in the enrichment of a CD24-low CD44-high population that had a high capability to form spheroids in vitro and subcutaneous tumors in vivo (Mani et al. 2008). Induction of oncogenic Ras also induced EMT and enriched the CSC population in breast cancer cells (Morel et al. 2008). In the liver, TGF-β signaling appears to induce the differentiation of hepatic stem/progenitor cells and suppress the development of HCC (Mishra et al. 2009), suggesting that it may not work in the same manner observed in breast cancers. Regardless, the association between the liver CSC phenotypes and the induction of EMT/MET programs is completely unclear and should be pursued in future studies (Fig. 16.2).

5 Conclusions

There is accumulating evidence that liver CSCs play a key role in the development and perpetuation of HCC, and the relevance of targeting CSCs has also become clear. Yet, experimental models for the treatment of HCC are still in the preliminary stages. Identification of useful CSC markers and exploration of their roles in maintaining stem-like traits are critical steps toward the clinical application of the CSC hypothesis for the improved diagnosis and the treatment of HCC.

References

Al-Hajj M, Wicha MS, Benito-Hernandez A et al (2003) Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 100:3983–3988

Balzar M, Winter MJ, de Boer CJ et al (1999) The biology of the 17-1A antigen (Ep-CAM). J Mol Med 77:699–712

Boman BM, Huang E. (2008) Human colon cancer stem cells: a new paradigm in gastrointestinal oncology. J Clin Oncol 26:2828–2838

Bonnet D, Dick JE. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 3:730–737

Calmont A, Wandzioch E, Tremblay KD et al (2006) An FGF response pathway that mediates hepatic gene induction in embryonic endoderm cells. Dev Cell 11:339–348

Calvisi DF, Thorgeirsson SS. (2005) Molecular mechanisms of hepatocarcinogenesis in transgenic mouse models of liver cancer. Toxicol Pathol 33:181–184

Challen GA, Little MH. (2006) A side order of stem cells: the SP phenotype. Stem Cells 24:3–12

Chiba T, Kita K, Zheng YW et al (2006) Side population purified from hepatocellular carcinoma cells harbors cancer stem cell-like properties. Hepatology 44:240–251

Chiba T, Zheng YW, Kita K et al (2007) Enhanced self-renewal capability in hepatic stem/progenitor cells drives cancer initiation. Gastroenterology 133:937–950

Clarke MF, Dick JE, Dirks PB et al (2006) Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells. Cancer Res 66:9339–9344

Crosby HA, Kelly DA, Strain AJ. (2001) Human hepatic stem-like cells isolated using c-kit or CD34 can differentiate into biliary epithelium. Gastroenterology 120:534–544

Dabeva MD, Shafritz DA. (2003) Hepatic stem cells and liver repopulation. Semin Liver Dis 23:349–362

Dan YY, Riehle KJ, Lazaro C et al (2006) Isolation of multipotent progenitor cells from human fetal liver capable of differentiating into liver and mesenchymal lineages. Proc Natl Acad Sci U S A 103:9912–9917

Dean M, Fojo T, Bates S. (2005) Tumour stem cells and drug resistance. Nat Rev Cancer 5:275–284

Dome B, Timar J, Ladanyi A et al (2009) Circulating endothelial cells, bone marrow-derived endothelial progenitor cells and proangiogenic hematopoietic cells in cancer: from biology to therapy. Crit Rev Oncol Hematol 69:108–124

Dudas J, Mansuroglu T, Batusic D et al (2009) Thy-1 is expressed in myofibroblasts but not found in hepatic stellate cells following liver injury. Histochem Cell Biol 131:115–127

Dunsford HA, Sell S. (1989) Production of monoclonal antibodies to preneoplastic liver cell populations induced by chemical carcinogens in rats and to transplantable Morris hepatomas. Cancer Res 49:4887–4893

Escribano L, Ocqueteau M, Almeida J et al (1998) Expression of the c-kit (CD117) molecule in normal and malignant hematopoiesis. Leuk Lymphoma 30:459–466

Fan J, Li R, Zhang R et al (2007) Effect of Bcl-2 and Bax on survival of side population cells from hepatocellular carcinoma cells. World J Gastroenterol 13:6053–6059

Fausto N. (2004) Liver regeneration and repair: hepatocytes, progenitor cells, and stem cells. Hepatology 39:1477–1487

Fialkow PJ. (1976) Clonal origin of human tumors. Biochim Biophys Acta 458:283–321

Gangenahalli GU, Singh VK, Verma YK et al (2006) Hematopoietic stem cell antigen CD34: role in adhesion or homing. Stem Cells Dev 15:305–313

Gilyarov A.V. (2008) Nestin in central nervous system cells. Neurosci Behav Physiol 38:165–169

Goodell MA, Brose K, Paradis G et al (1996) Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J Exp Med 183:1797–1806

Graf T, Stadtfeld, M. (2008) Heterogeneity of embryonic and adult stem cells. Cell Stem Cell 3:480–483

Hanahan D, Weinberg, RA. (2000) The hallmarks of cancer. Cell 100:57–70

Herlyn M, Steplewski Z, Herlyn D et al (1979) Colorectal carcinoma-specific antigen: detection by means of monoclonal antibodies. Proc Natl Acad Sci U S A 76:1438–1442

Hu C, Li H, Li J et al (2008) Analysis of ABCG2 expression and side population identifies intrinsic drug efflux in the HCC cell line MHCC-97L and its modulation by Akt signaling. Carcinogenesis 29:2289–2297

Hugo H, Ackland ML, Blick T et al (2007) Epithelial – mesenchymal and mesenchymal – epithelial transitions in carcinoma progression. J Cell Physiol 213:374–383

Inada M, Follenzi A, Cheng K et al (2008) Phenotype reversion in fetal human liver epithelial cells identifies the role of an intermediate meso-endodermal stage before hepatic maturation. J Cell Sci 121:1002–1013

Jelnes P, Santoni-Rugiu E, Rasmussen M et al (2007) Remarkable heterogeneity displayed by oval cells in rat and mouse models of stem cell-mediated liver regeneration. Hepatology 45: 1462–1470

Jensen CH, Jauho EI, Santoni-Rugiu E et al (2004) Transit-amplifying ductular (oval) cells and their hepatocytic progeny are characterized by a novel and distinctive expression of delta-like protein/preadipocyte factor 1/fetal antigen 1. Am J Pathol 164:1347–1359

Jordan CT, Guzman ML, Noble, M. (2006) Cancer stem cells. N Engl J Med 355:1253–1261

Kamiya A, Kinoshita T, Ito Y et al (1999) Fetal liver development requires a paracrine action of oncostatin M through the gp130 signal transducer. EMBO J 18:2127–2136

Karnoub AE, Dash AB, Vo AP et al (2007) Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. Nature 449:557–563

Khurana S, Mukhopadhyay, A. (2008) Hematopoietic progenitors from early murine fetal liver possess hepatic differentiation potential. Am J Pathol 173:1818–1827

Kinoshita T, Miyajima, A. (2002) Cytokine regulation of liver development. Biochim Biophys Acta 1592:303–312

Koenig S, Probst I, Becker H et al (2006) Zonal hierarchy of differentiation markers and nestin expression during oval cell mediated rat liver regeneration. Histochem Cell Biol 126:723–734

Kordes C, Sawitza I, Muller-Marbach A et al (2007) CD133+ hepatic stellate cells are progenitor cells. Biochem Biophys Res Commun 352:410–417

Kuwahara R, Kofman AV, Landis CS et al (2008) The hepatic stem cell niche: identification by label-retaining cell assay. Hepatology 47:1994–2002

Lessard J, Sauvageau, G. (2003) Bmi-1 determines the proliferative capacity of normal and lèukaemic stem cells. Nature 423:255–260

Li L, Krantz ID, Deng Y et al (1997) Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nat Genet 16:243–251

Limaye PB, Bowen WC, Orr AV et al (2008) Mechanisms of hepatocyte growth factor-mediated and epidermal growth factor-mediated signaling in transdifferentiation of rat hepatocytes to biliary epithelium. Hepatology 47:1702–1713

Lozier J, McCright B, Gridley, T. (2008) Notch signaling regulates bile duct morphogenesis in mice. PLoS ONE 3:e1851

Ma S, Chan KW, Hu L et al (2007) Identification and characterization of tumorigenic liver cancer stem/progenitor cells. Gastroenterology 132:2542–2556

Ma S, Chan KW, Lee TK et al (2008a) Aldehyde dehydrogenase discriminates the CD133 liver cancer stem cell populations. Mol Cancer Res 6:1146–1153

Ma S, Lee TK, Zheng BJ et al (2008b) CD133+ HCC cancer stem cells confer chemoresistance by preferential expression of the Akt/PKB survival pathway. Oncogene 27:1749–1758

Maetzel D, Denzel S, Mack B et al (2009) Nuclear signalling by tumour-associated antigen EpCAM. Nat Cell Biol 11:162–171

Mani SA, Guo W, Liao MJ et al (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 133:704–715

Masson NM, Currie IS, Terrace JD et al (2006) Hepatic progenitor cells in human fetal liver express the oval cell marker Thy-1. Am J Physiol Gastrointest Liver Physiol 291:G45–54

Michalopoulos GK, Barua L, Bowen WC. (2005) Transdifferentiation of rat hepatocytes into biliary cells after bile duct ligation and toxic biliary injury. Hepatology 41:535–544

Mishra L, Banker T, Murray J et al (2009) Liver stem cells and hepatocellular carcinoma. Hepatology 49:318–329

Mishra PJ, Humeniuk R, Medina DJ et al (2008) Carcinoma-associated fibroblast-like differentiation of human mesenchymal stem cells. Cancer Res 68:4331–4339

Mizrak D, Brittan M, Alison M.R. (2008) CD133: molecule of the moment. J Pathol 214:3–9

Morel AP, Lievre M, Thomas C et al (2008) Generation of breast cancer stem cells through epithelial-mesenchymal transition. PLoS ONE 3:e2888

Murry CE, Keller, G. (2008) Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell 132:661–680

Niki T, Pekny M, Hellemans K et al (1999) Class VI intermediate filament protein nestin is induced during activation of rat hepatic stellate cells. Hepatology 29:520–527

Nitou M, Sugiyama Y, Ishikawa K et al (2002) Purification of fetal mouse hepatoblasts by magnetic beads coated with monoclonal anti-e-cadherin antibodies and their in vitro culture. Exp Cell Res 279:330–343

O'Brien CA, Pollett A, Gallinger S et al (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 445:106–110

Ober EA, Verkade H, Field HA et al (2006) Mesodermal Wnt2b signalling positively regulates liver specification. Nature 442:688–691

Oda T, Elkahloun AG, Pike BL et al (1997) Mutations in the human Jagged1 gene are responsible for Alagille syndrome. Nat Genet 16:235–242

Oertel M, Menthena A, Chen YQ et al (2007) Comparison of hepatic properties and transplantation of Thy-1(+) and Thy-1(-) cells isolated from embryonic day 14 rat fetal liver. Hepatology 46:1236–1245

Oertel M, Menthena A, Chen YQ et al (2008) Purification of fetal liver stem/progenitor cells containing all the repopulation potential for normal adult rat liver. Gastroenterology 134:823–832

Oertel M, Shafritz DA. (2008) Stem cells, cell transplantation and liver repopulation. Biochim Biophys Acta 1782:61–74

Overturf K, al-Dhalimy M, Ou CN et al (1997) Serial transplantation reveals the stem-cell-like regenerative potential of adult mouse hepatocytes. Am J Pathol 151:1273–1280

Potter VR. (1978) Phenotypic diversity in experimental hepatomas: the concept of partially blocked ontogeny. The 10th Walter Hubert Lecture. Br J Cancer 38:1–23

Rege TA, Hagood JS. (2006) Thy-1, a versatile modulator of signaling affecting cellular adhesion, proliferation, survival, and cytokine/growth factor responses. Biochim Biophys Acta 1763: 991–999

Ricci-Vitiani L, Lombardi DG, Pilozzi E et al (2007) Identification and expansion of human colon-cancer-initiating cells. Nature 445:111–115

Roskams T, Cassiman D, De Vos R et al (2004a) Neuroregulation of the neuroendocrine compartment of the liver. Anat Rec A Discov Mol Cell Evol Biol 280:910–923

Roskams TA, Libbrecht L, Desmet VJ. (2003) Progenitor cells in diseased human liver. Semin Liver Dis 23:385–396

Roskams TA, Theise ND, Balabaud C et al (2004b) Nomenclature of the finer branches of the biliary tree: canals, ductules, and ductular reactions in human livers. Hepatology 39:1739–1745

Rossi JM, Dunn NR, Hogan BL et al (2001) Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. Genes Dev 15:1998–2009

Schmelzer E, Wauthier E, Reid LM. (2006) The phenotypes of pluripotent human hepatic progenitors. Stem Cells 24:1852–1858

Schmelzer E, Zhang L, Bruce A et al (2007) Human hepatic stem cells from fetal and postnatal donors. J Exp Med 204:1973–1987

Sell S. (1993) Cellular origin of cancer: dedifferentiation or stem cell maturation arrest? Environ Health Perspect 101 (Suppl 5):15–26

Sell S. (2002) Cellular origin of hepatocellular carcinomas. Semin Cell Dev Biol 13:419–424

Sell S. (2003) The hepatocyte: heterogeneity and plasticity of liver cells. Int J Biochem Cell Biol 35:267–271

Shi GM, Xu Y, Fan J et al (2008) Identification of side population cells in human hepatocellular carcinoma cell lines with stepwise metastatic potentials. J Cancer Res Clin Oncol 134: 1155–1163

Singh SK, Hawkins C, Clarke ID et al (2004) Identification of human brain tumour initiating cells. Nature 432:396–401

Sjoblom T, Jones S, Wood LD et al (2006) The consensus coding sequences of human breast and colorectal cancers. Science 314:268–274

Slack JM. (2008) Origin of stem cells in organogenesis. Science 322:1498–1501

Strain AJ, Crosby HA, Nijjar S et al (2003) Human liver-derived stem cells. Semin Liver Dis 23:373–384

Suetsugu A, Nagaki M, Aoki H et al (2006) Characterization of CD133+ hepatocellular carcinoma cells as cancer stem/progenitor cells. Biochem Biophys Res Commun 351:820–824

Suzuki A, Sekiya S, Onishi M et al (2008) Flow cytometric isolation and clonal identification of self-renewing bipotent hepatic progenitor cells in adult mouse liver. Hepatology 48:1964–1978

Takahashi K, Okita K, Nakagawa M et al (2007a) Induction of pluripotent stem cells from fibroblast cultures. Nat Protoc 2:3081–3089

Takahashi K, Tanabe K, Ohnuki M et al (2007b) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861–872

Takaishi S, Okumura T, Wang TC. (2008) Gastric cancer stem cells. J Clin Oncol 26:2876–2882

Tanimizu N, Miyajima, A. (2004) Notch signaling controls hepatoblast differentiation by altering the expression of liver-enriched transcription factors. J Cell Sci 117:3165–3174

Tanimizu N, Nishikawa M, Saito H et al (2003) Isolation of hepatoblasts based on the expression of Dlk/Pref-1. J Cell Sci 116:1775–1786

Turley EA, Veiseh M, Radisky DC et al (2008) Mechanisms of disease: epithelial-mesenchymal transition--does cellular plasticity fuel neoplastic progression? Nat Clin Pract Oncol 5: 280–290

Van Den Heuvel MC, Slooff MJ, Visser L et al (2001) Expression of anti-OV6 antibody and anti-N-CAM antibody along the biliary line of normal and diseased human livers. Hepatology 33:1387–1393
Visvader JE, Lindeman GJ. (2008) Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat Rev Cancer 8:755–768
Vogelstein B, Kinzler KW. (2004) Cancer genes and the pathways they control. Nat Med 10: 789–799
Wauthier E, Schmelzer E, Turner W et al (2008) Hepatic stem cells and hepatoblasts: identification, isolation, and ex vivo maintenance. Methods Cell Biol 86:137–225
Weiss TS, Lichtenauer M, Kirchner S et al (2008) Hepatic progenitor cells from adult human livers for cell transplantation. Gut 57:1129–1138
Went P, Vasei M, Bubendorf L et al (2006) Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers. Br J Cancer 94:128–135
Wicha MS, Liu S, Dontu, G. (2006) Cancer stem cells: an old idea--a paradigm shift. Cancer Res 66:1883–1890; discussion 1895–1886
Wood LD, Parsons DW, Jones S et al (2007) The genomic landscapes of human breast and colorectal cancers. Science 318:1108–1113
Wu C, Alman BA. (2008) Side population cells in human cancers. Cancer Lett 268:1–9
Yamashita T, Budhu A, Forgues M et al (2007) Activation of hepatic stem cell marker EpCAM by Wnt-beta-catenin signaling in hepatocellular carcinoma. Cancer Res 67:10831–10839
Yamashita T, Forgues M, Wang W et al (2008) EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. Cancer Res 68:1451–1461
Yamashita T, Ji J, Budhu A et al (2009) EpCAM-Positive Hepatocellular Carcinoma Cells Are Tumor-Initiating Cells With Stem/Progenitor Cell Features. Gastroenterology 136:1012–1024
Yang W, Yan HX, Chen L et al (2008a) Wnt/beta-catenin signaling contributes to activation of normal and tumorigenic liver progenitor cells. Cancer Res 68:4287–4295
Yang ZF, Ho DW, Ng MN et al (2008b) Significance of CD90+ cancer stem cells in human liver cancer. Cancer Cell 13:153–166
Yang ZF, Ngai P, Ho DW et al (2008c) Identification of local and circulating cancer stem cells in human liver cancer. Hepatology 47:919–928
Yin AH, Miraglia S, Zanjani ED et al (1997) AC133, a novel marker for human hematopoietic stem and progenitor cells. Blood 90:5002–5012
Yin S, Li J, Hu C et al (2007) CD133 positive hepatocellular carcinoma cells possess high capacity for tumorigenicity. Int J Cancer 120:1444–1450
Yovchev MI, Grozdanov PN, Zhou H et al (2008) Identification of adult hepatic progenitor cells capable of repopulating injured rat liver. Hepatology 47:636–647
Zaret KS, Grompe, M. (2008) Generation and regeneration of cells of the liver and pancreas. Science 322:1490–1494
Zen Y, Fujii T, Yoshikawa S et al (2007) Histological and culture studies with respect to ABCG2 expression support the existence of a cancer cell hierarchy in human hepatocellular carcinoma. Am J Pathol 170:1750–1762
Zender L, Spector MS, Xue W et al (2006) Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell 125:1253–1267
Zhou S, Schuetz JD, Bunting KD et al (2001) The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nat Med 7:1028–1034
Zou G.M. (2008) Cancer initiating cells or cancer stem cells in the gastrointestinal tract and liver. J Cell Physiol 217:598–604

The Fourth International Conference on Computational Systems Biology (ISB2010)
Suzhou, China, September 9–11, 2010
Copyright © 2010 ORSC & APORC, pp. 196–203

Significant Regulatory Networks from Goto-Kakizaki Rat Liver Microarray Data during Diabetic Progression

Huarong Zhou[1] Shigeru Saito[2,3] Guanying Piao[4] Rongkuan Hu[4]
Jiguang Wang[1] Zhi-Ping Liu[1] Katsuhisa Horimoto[2,5*]

1 Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China
2 Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan
3 INFOCOM Corporation, Tokyo 150-0001, Japan
4 Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei 230027, China
5 Institute of Systems Biology, Shanghai University, Shanghai 200444, China
*corresponding author

Abstract In the aim of identifying significant transcriptional regulatory networks in the liver contributing to diabetes, we have performed comprehensive active regulatory network survey by network screening in 4weeks (w), 8-12w, and 18-20w Goto-Kakizaki (GK) rat liver microarray data. The comprehensive survey of the consistency between the networks and the measured data by the network screening approach in the case of non-insulin dependent diabetes in the GK rat reveals: 1. More pathways are active during inter-middle stage diabetes; 2. Inflammation, hypoxia, increased apoptosis, decreased proliferation, and altered metabolism are characteristics and display as early as 4weeks in GK strain; 3. Diabetes progression accompanies insults and compensations;

Keywords GK rat; Regulatory Network; Network Screening; Active Pathways; Diabetic progression

1 Introduction

Type 2 diabetes mellitus (T2DM) is a complex systemic disease, with significant disorders of metabolism [1]. The liver, a central energy metabolic organ, plays a critical role in the development of diabetes [2]. Although gene expression levels are able to be measured via microarray since 1996, it is difficult to evaluate the contributions of one altered gene expression to a specific disease. One of the reasons is that a whole network picture responsible for a specific phase of diabetes is missing, while a single gene has to be put into a network picture to evaluate its

importance.

In the aim of identifying significant transcriptional regulatory networks in the liver contributing to diabetes, we have performed comprehensive active regulatory network survey by network screening in 4weeks (w), 8-12w, and 18-20w Goto-Kakizaki (GK) rat liver microarray data. We identify active regulatory networks in GK rat by network screening in the following procedure. First, the regulatory networks are compiled by using the known binary relationships between the transcriptional factors and their regulated genes and the biological classification scheme, and second, the consistency of each regulatory network with the microarray data measured in GK rat is estimated to detect the active networks under the corresponding conditions.

2 Materials and Methods

2.1 Network Screening

The candidates of active regulatory networks are detected by network screening in the following manner. First, the regulatory network sets are generated by combining the binary relationships between transcriptional factors (TFs) and their regulating genes, which are compiled in TRANSFAC database [3], and the functional gene sets defined in the Molecular Signatures Database (MSigDB) [4]. Then, we calculate the graph consistency probability (GCP) [5], which expresses the consistency of a given network structure with the monitored expression data of the constituent genes in this study, for each of the network structures obtained at the first step. In addition, in each reference network, the enrichment probability of the genes with the significant differences between GK and WKY rats (expression signature) is further tested. For this purpose, the expression signature is determined using the Student's t-test (for a false discovery rate [FDR] $< 5\%$ in expression between GK and WKY rats). The number of genes included in the expression signature is tested for each network, based on the hyper-geometric probability. Thus, we refine the candidates of active regulatory networks, in terms of both the network structure by GCP and the extent of gene expression by enrichment probability. The significance of both thresholds is set to 0.05.

2.2 Microarray data

Microarray dataset is cited from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/projects/geo/) database (GSE 13271). The data are composed of 31,099 probes measured by using Affymetrix Microarray Suite 5.0 (Affymetrix), which are reduced into 14,506 genes, for 5 samples of male Goto-Kakizaki (GK) spontaneously diabetic rats and WKY rats at each of 5 time points (4, 8, 12, 16, and 20 weeks of age). Hyperglycemia begins to show at 4 weeks of age and stabilize after 16 weeks in GK, thus we divided data into three functional groups: 4w, 8-12w, and 16-20w.

3 Results and Discussion

3.1 Activated pathways revealed by network screening and their functions

We identify a total of 20 and 19 differentially activating transcriptional regulatory networks in GK and WKY rats, respectively. There are fewer pathways activating at 4w and 16-20w in GK rats which are at the beginning and the steady state of diabetes. While during 8-12w, more pathways are significantly activated, which indicates a dynamic process involving dysfunctions and compensations in the development of diabetes, as showed outside blood glucose fluctuations. There are more active pathways in the 4w and 8-12w than those in the 16-20w in WKY, which may be due to body growth and development. It is worth pointing out that many activating pathways in WKY are diminished in GK rats at 4w, suggesting that those pathways in the liver important to keep glucose metabolism homeostasis are dysfunction at very early stages of diseases.

Table 1: Active regulatory networks classification according to their functions.

	GK	WKY
Metabolism	HSC_LATEPROGENITORS_ADULT ATRIA_UP GLYCEROPHOSPHOLIPID_METABOLISM GOLUB_ALL_VS_AML_UP HOHENKIRK_MONOCYTE_DEND_UP HSC_LATEPROGENITORS_ADULT LONGEVITYPATHWAY VHL_NORMAL_UP	HASLINGER_B_CLL_MUTATED VEGF_HUVEC_30MIN_UP YAGI_AML_PROG_ASSOC ZHAN_MM_CD138_CD1_VS_REST
Immune	HSC_LATEPROGENITORS_ADULT LINDSTEDT_DEND_8H_VS_48H_DN LEI_HOXC8_DN TESTIS_EXPRESSED_GENES TSADAC_RKOEXP_UP	NGUYEN_KERATO_UP ICF_UP
Transcription	HSC_LATEPROGENITORS_ADULT ATRIA_UP GOLUB_ALL_VS_AML_UP HOHENKIRK_MONOCYTE_DEND_UP HSC_LATEPROGENITORS_ADULT MEF2DPATHWAY P35ALZHEIMERSPATHWAY	VEGFPATHWAY HCC_SURVIVAL_GOOD_VS_POOR_UP HSC_LATEPROGENITORS_SHARED SCHURINGA_STAT5A_UP NUCLEAR_RECEPTORS CELL_DEATH NI2_LUNG_DN PARK_RARALPHA_MOD NUCLEAR_RECEPTORS TGFBPATHWAY
Signaling transduction	INTEGRINPATHWAY INTEGRIN_MEDIATED_CELL_ADHESION_KEGG MEF2DPATHWAY P35ALZHEIMERSPATHWAY RCC_NL_UP VHL_NORMAL_UP ASTON_OLIGODENDROGLIA_MYELINATION_SUBSET BRCA_BRCA1_NEG LEI_HOXC8_DN TESTIS_EXPRESSED_GENES TSADAC_RKOEXP_UP VEGFPATHWAY HSC_LATEPROGENITORS_ADULT	P21_P53_MIDDLE_DN UVB_NHEK1_C2 ALKPATHWAY BRENTANI_PROTEIN_MODIFICATION CELL_DEATH NI2_LUNG_DN PARK_RARALPHA_MOD TGFBPATHWAY NUCLEAR_RECEPTORS

3.2 Classification of activated pathways revealed in terms of their functions

Apart from the view of differentially activated networks along the time points, the networks in the GK and WKY strains can be classified into 4 functional categories in Table 1, which are metabolism, immune, transcription, and signal transduction. Note that some activated pathways share their functions. In that case, they are listed under several functional groups as long as the condition met. Then, we combine the activated networks belonging to the same functional category, if any constituent genes of transcriptional factor (TF) and its regulated gene share each other in the networks. Thus TF-gene expression networks for each functional category are created (Figs 1A-E). Interestingly, significantly activated networks in GK and WKY strains are very different even in the same functional category. We will describe the details of the activated networks in 4 functional categories, below.

3.2.1 Metabolism

Metabolic TF regulatory network in WKY rats reveals increased expression of several genes are important to keep metabolic homeostasis, e.g. bone gamma-carboxyglutamic acid-containing protein (BGLAP), Hepatocyte nuclear factor 4 alpha (HNF4A) and Lipoprotein lipase (LPL) (Fig.1A). In addition to its role in bone-building, BGLAP stimulates pancreatic beta cells releasing more insulin and increases insulin sensitivity via enhancing adipocytes adiponectin secretion [6]. HNF4A plays a key role in liver development. Mutations in this gene have been associated with maturity-onset non-insulin-dependent diabetes of the young (MODY) [7]. Our analysis indicates that reduced HNF4A expression may also favor T2DM development in GK rats. LPL is an enzyme that hydrolyzes triglyceride in lipoproteins such as very low-density lipoproteins (VLDL) and reforms high-density lipoproteins (HDL). Lipoprotein lipase deficiency leads to elevated levels of triglycerides in the bloodstream [8]. Interestingly, like HNF4A, LPL is also suggested to be a diabetes susceptibility gene by human studies [9].

Metabolic networks in GK rats are more complicated than those in WKY rats (Fig.1B). Besides the reduced expression of three genes described in the previous paragraph in diabetic GK rats, some pathways identified by network screening further contribute to metabolism disorders. Cytokines induce activation of the JAK-STAT pathway results in expression of various suppressors of cytokine signaling (SOCS). Expression of SOCS2 and STAT5 but not SOCS3 is decreased in GK rats. Decreased expression of SOCS2 leads to enlarged internal organs, which consists with the description in the original paper that liver weight as a percentage of total body weight is significantly larger in GK [10]. Insulin directly stimulates SOCS2 and STAT5 expression, and the decreased SOCS2 and STAT5 levels are due to insulin deficiency or resistance. IGF-1 (insulin-like growth factor-1) has functions similar to insulin, and it can also improve blood sugar profiles in type 2 diabetics [11]. IGF-1 levels are increased at 4w, but significantly decreased, thereafter may partially explain the insulin resistance after 8 weeks of age in GK rats.

We also observed some compensative pathways activation in GK to fight against

insulin resistance. For instance, insulin receptor substrate 2 (IRS2) is up-regulated and SOCS1 is down-regulated at 8-12w. Cytokine-induced SOCS-1 interacts with the phosphorylated insulin receptor and promotes ubiquitination (Ub) and degradation of IR-IRS complex, thereby preventing insulin signaling pathways [12]. Decreased SOCS-1 is correlated to insulin sensitivity. However, compensations fail to stop development of diabetes.

3.2.2 Immune

Many proinflammatory pathways are activating in GK compared to WKY rats (Fig 1C). From the TF-regulatory gene expression networks in GK rats, two hubs which play important role in immune damages are displayed.

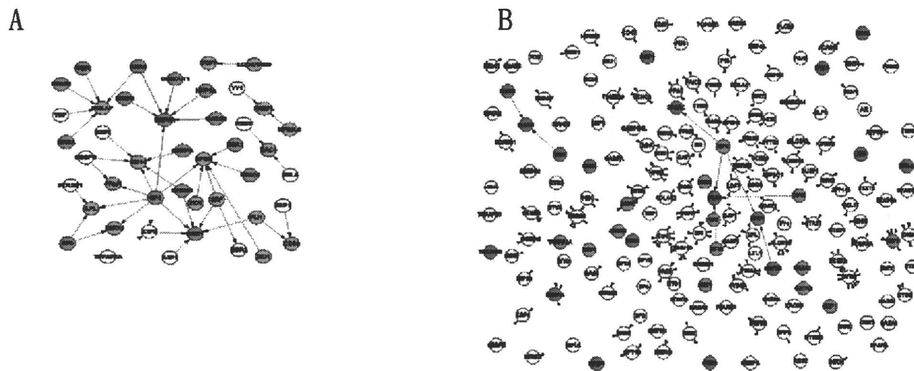
Cytochrome b-245, beta polypeptide (CYBB) is a gene encoding gp91(phox) protein, a phagocyte NADPH oxidase. The protein is also known as P91-PHOX and NOX2. Reactive oxygen species (ROS) produced by NOX2 are able to kill phagocytized bacteria. Because of its highly reactive nature, CYBB has been considered harmful mediators of inflammation [13]. NF-KB and interferon-gamma further increase CYBB expression. Prolonged highly CYBB expression enhanced production of reactive oxygen species, which are critical sources mediating neurovascular damage. Significantly overexpressed CYBB in GK stain is a critical contributor to the microvascular complications associated with diabetes.

Activating transcription factor 3 (ATF3) is a stress-inducible gene and encodes ATF3 transcription factors. ATF3 expression has been reported up-regulated in insulitis and type 1 or type 2 diabetics. Induction of ATF3 is mediated by proinflammatory factors, such as nitric oxide and NF-κB. Importantly, the induction of ATF3 leads cell apoptosis, while signals without ATF3 up-regulation do not cause cell damage [14]. Increased gene expression of ATF3 in GK rats are related to increased immune response and apoptosis.

Besides these two hubs, about 20 immune related genes are changed in GK strain. In sum, inflammation is significantly increased in diabetic Gk rats.

3.2.3 Transcription

Pathways analysis reveals that WKY transcriptional network is a balanced and well-controlled system.



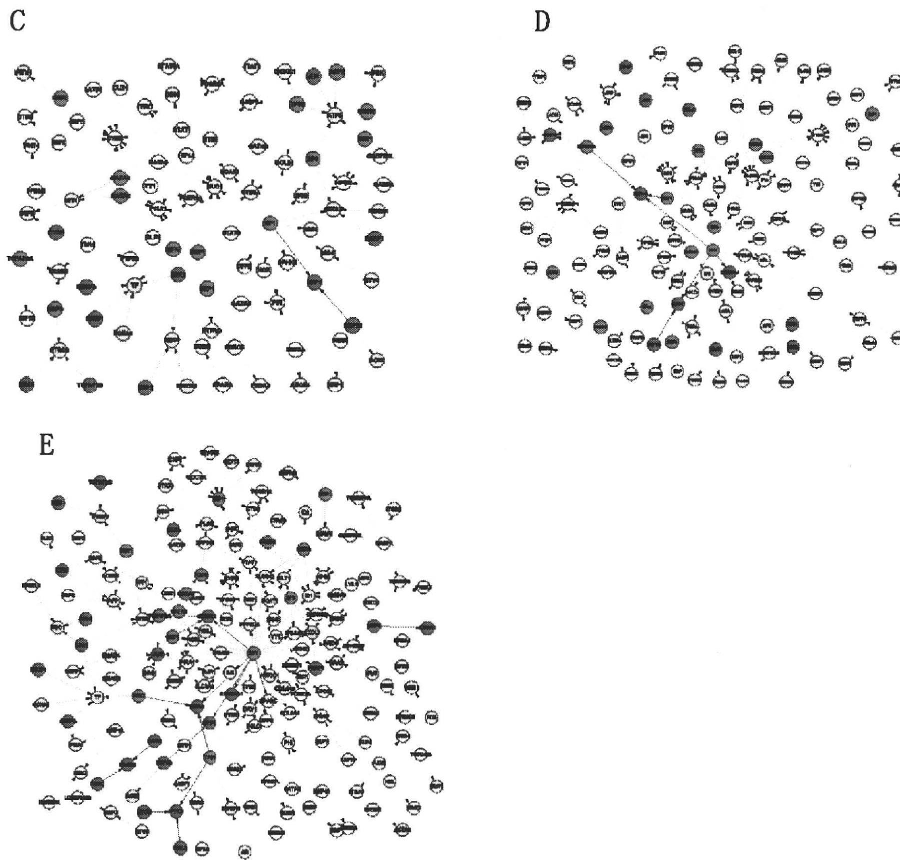


Fig 1. Combined networks in the four functional categories. A: metabolism in WKY, B: metabolism in GK, C: immune in GK, D: transcription in GK, and E: signal transduction in GK (see details in the text).

Some involves in cell replication, good survival and self renewal. Others, including P21-P53_Middle_DN, UBV_NHEK1_C2, and TGFBPATHWAY, emphasize anticancer and cell cycle checkpoints regulation (Table 1).

In GK rats, two out of 7 pathways are related to apoptosis (Table 1 and Fig 1D). Caspase 1 (CASP1), which has been shown to induce cell apoptosis, is overexpressed. Transforming growth factor alpha (TGFA), which stimulates neural cell proliferation, is inhibited. Interestingly, diabetes activates several genes involving in neurodegenerative disorders. Alzheimer's disease shares many commons with T2DM, so that some scientists proposed to call Alzheimer's disease "type 3 diabetes" or "diabetes of the brain." Calpain small subunit 1 (CAPNS1), a highly-conserved cysteine protease, which have been implicated in neurodegenerative processes after oxidative stress stimulation, is more active in GK. Casein kinase I isoform alpha (CSNK1A1), also called CK1α, is associated with phosphorylate tau and amyloid formation. The expression of CK1α gene is much higher in GK.

3.2.4 Signal Transduction

The key difference in signal transduction category is activation of hypoxia and coagulation related pathways in GK rats (Table 1 and Fig 1E). Coagulation factor XIII A chain (F13A1) is the last zymogen activating in the blood coagulation cascade, which stabilize clots. In GK rats, F13A1 gene expression levels are significantly elevated which enhance thrombosis. Macrophages expressing high affinity immunoglobulin gamma Fc receptor I (FcgRIa) also display coagulation function via binding platelets and initiate thrombosis. Tissue plasminogen activator (PLAT) breakdowns blood clots. GK rats present significantly higher PLAT expression levels, which may explain hemolysis and thrombosis co-existing in diabetics. Dr. Auwerx reported in diabetics, PLAT and plasminogen activator (PA) inhibitor are both activated [16]. The elevated levels of PA-inhibitor activity abolish PLAT activity inducing a reduced fibrinolytic capacity.

3.3 Further remarks

In order to understand the dynamical changes of regulatory networks in the development of diabetes, the active networks can be also interpreted in terms of each time segments. The characteristic features of the active networks, especially relationship between active networks and diabetic progression, will be reported in near future.

This study is the first time to use network screening to explain the role of liver in development of diabetes and the underline mechanism. The results provide many important rational information and insights into guiding experiments design. It is worth pointing out that the molecular relationships change dynamically, depending on the conditions in a living cell, which suggests implicitly that all of the relationships in the knowledge-based network do not always exist.

Acknowledges

We are grateful to Dr. Jiarui Wu, Dr. Jacob Sten Petersen, and Dr. Trine Ryberg Clausen for their comments and support. Work was supported by grants from NN-CAS Research Foundation under NO. NNCAS-2009-1 (H.Z.), and Development of Analysis Technology for Induced Pluripotent Stem (iPS) Cell', from The New Energy and Industrial Technology Development Organization, Japan (S.S. and K.H.).

References

[1] Smyth S, Heron A: Diabetes and obesity: the twin epidemics. *Nat Med* 2006, **12:**75-80.

[2] Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007, **445:**881-885.

[3] Wingender, E: TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief. Bioinformatics 2008, 9:326-332.

[4] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment

analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 2005, 102:15545-15550.

[5] Saito S, Aburatani S, Horimoto K: Network evaluation from the consistency of the graph structure with the measured data. BMC Sys Biol 2008, 2:84.

[6] Lee NK, Sowa H, Hinoi E, Ferron M, Ahn JD, Confavreux C, Dacquin R, Mee PJ, McKee MD, Jung DY, Zhang Z, Kim JK, Mauvais-Jarvis F, Ducy P, Karsenty G: Endocrine regulation of energy metabolism by the skeleton. Cell 2007, 130:456-469.

[7] Beijers HJ, Losekoot M, Odink RJ, Bravenboer B: Hepatocyte nuclear factor (HNF)1A and HNF4A substitution occurring simultaneously in a family with maturity-onset diabetes of the young. Diabet Med 2009, 26:1172-1174.

[8] Nikkilä EA, Huttunen JK, Ehnholm C: Postheparin plasma lipoprotein lipase and hepatic lipase in diabetes mellitus. Relationship to plasma triglyceride metabolism. Diabetes 1977, 26: 11-21.

[9] Das B, Pawar N, Saini D, Seshadri M: Genetic association study of selected candidate genes (ApoB, LPL, Leptin) and telomere length in obese and hypertensive individuals. BMC Med Genet 2009, 10:99.

[10] Greenhalgh CJ, Rico-Bautista E, Lorentzon M, Thaus AL, Morgan PO, Willson TA, Zervoudakis P, Metcalf D, Street I, Nicola NA, Nash AD, Fabri LJ, Norstedt G, Ohlsson C, Flores-Morales A, Alexander WS, Hilton DJ: SOCS2 negatively regulates growth hormone action in vitro and in vivo. J Clin Invest 2005, 115:397–406.

[11] Haluzik M, Yakar S, Gavrilova O, Setser J, Boisclair Y, LeRoith D: Insulin Resistance in the Liver-Specific IGF-1 Gene-Deleted Mouse Is Abrogated by Deletion of the Acid-Labile Subunit of the IGF-Binding Protein-3 Complex Relative Roles of Growth Hormone and IGF-1 in Insulin Resistance. Diabetes 2003, 52:2483-2489.

[12] Rui L, Yuan M, Frantz D, Shoelson S, White MF: SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. J Biol Chem 2002, 277:42394-42398.

[13] Bolscher BG, de Boer M, de Klein A, Weening RS, Roos D: Point mutations in the beta-subunit of cytochrome b558 leading to X-linked chronic granulomatous disease. Blood 1991, 77: 2482–2487.

[14] Fan F, Jin S, Amundson SA, Tong T, Fan W, Zhao H, Zhu X, Mazzacurati L, Li X, Petrik KL, Fornace AJ Jr, Rajasekaran B, Zhan Q: ATF3 induction following DNA damage is regulated by distinct signaling pathways and over-expression of ATF3 protein suppresses cells growth. Oncogene 2002, 1:7488-7496.

[15] Auwerx J, Bouillon R, Collen D, Geboers J: Tissue-type plasminogen activator antigen and plasminogen activator inhibitor in diabetes mellitus. Arteriosclerosis 1988, 8:68-72.

The Fourth International Conference on Computational Systems Biology (ISB2010)
Suzhou, China, September 9–11, 2010
Copyright © 2010 ORSC & APORC, pp. 381–388

Potential Linkages Between the Inner and Outer Cellular States of Human Induced Pluripotent Stem Cells

Shigeru Saito[1,2,*] Yasuko Onuma[3,*] Yuzuru Ito[3,*] Hiroaki Tateno[4,*]
Masashi Toyoda[5,*] Hidenori Akutsu[5] Koichiro Nishino[5]
Emi Chikazawa[5] Yoshihiro Fukawatase[5] Yoshitaka Miyagawa[6]
Hajime Okita[6] Nobutaka Kiyokawa[6] Yohichi Shimma[4]
Akihiro Umezawa[5] Jun Hirabayashi[4] Katsuhisa Horimoto[1,7,§]
Makoto Asashima[3,8,§]

1 Computational Biology Research Center, National Institute of Advanced Industrial Science Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
2 INFOCOM CORPORATION, Sumitomo Fudosan Harajuku Building, 2-34-17, Jingumae, Shibuya-ku, Tokyo, 150-0001, Japan
3 Research Center for Stem Cell Engineering, National Institute of Advanced Industrial Science Technology (AIST), Tsukuba Central 4, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8562, Japan.
4 Research Center for Medical Glycoscience, National Institute of Advanced Industrial Science Technology (AIST), Tsukuba Central 2, 1-1-1 Umezono, Ibaraki 305-8568, Japan.
5 Department of Reproductive Biology, National Research Institute for Child Health and Development, 2-10-1 Ookura, Setagaya-ku, Tokyo 157-8535, Japan
6 Department of Developmental Biology and Pathology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan
7 Institute for Systems Biology, Shanghai University, Shangda Road 99, Shanghai 200444, China
8 Department of Life Sciences (Biology), Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
* These authors contributed equally to this work.
§ Correspondence author

Abstract We analyzed both RNA profile to reveal gene expression changes and glycan profile to identify structural changes in glycans between four parental somatic cell (SC) lines and nine hiPSC lines that were originally established. The sum of these information by a combination of standard statistical techniques and a network approach showed significant differences in expression between the iPSCs and SCs. Subsequent network analysis of the gene expression and glycan signatures revealed glycan transfer network associated with known epitopes for differentiation. The present study is the first to uncover the relationships between gene expression patterns and cell surface changes in hiPSCs, and reinforces the importance of the cell surface to identify established iPSCs from SCs.

Keywords Induced pluripotent stem cell; DNA microarray; Lectin microarray; Network Analysis

1 Introduction

Human iPS cells (hiPSCs) attract great attention for the application to drug screening and analysis of the mechanisms of diseases, and even as the next generation materials for regenerative medicine. Genetic reprogramming to a pluripotent state of human somatic cells was first achieved by ectopic expression of four factors (Sox2, Oct4, Klf4 and c-Myc) using retrovirus [1]. Subsequently, this method was applied to various human cells using different combination of defined factors [2, 3]. However, transcription factor-induced acquisition of replication competence and pluripotency raises the question as to how exogenous factors induce changes in the inner and outer cellular states [4-6].

Here, we analyzed both RNA profile to reveal gene expression changes and glycan profile to identify structural changes in glycans between four parental somatic cell (SC) lines and nine hiPSC lines that were originally established. The sum of these information by a combination of standard statistical techniques and a network approach showed significant differences in expression between the iPSCs and SCs. Subsequent network analysis of the gene expression and glycan signatures revealed glycan transfer network associated with known epitopes for differentiation. These results shed new light on the potential linkages between the inner and outer cellular states for acquisition of replication competence and pluripotency in hiPSCs.

2 Materials and Methods

2.1 Cell experiments

Cell experiment was performed as described [7, 8]. Somatic cell pellets were harvested by scraping. The hiPSCs were incubated at 37°C in a solution that contained 1 mg/ml collagenase IV (Invitrogen, Carlsbad, CA), 1 mM $CaCl_2$, 20% KNOCKOUT™ Serum Replacement (KSR), and 10% ACCUMAX (Innovative Cell Technologies, Inc., San Diego, CA). When the edges of the colonies started to dissociate from the bottom of the dish, the collagenase solution was removed and the cells were washed with medium. Colonies were then picked up and collected.

2.2 Gene expression analysis

Changes in mRNA levels were monitored using an Agilent Whole Human Genome Microarray chip (G4112F). This array covers 41,000 well-characterized human genes and transcripts. After background correction using a Normal plus Exponential convolution model, which adjusts the foreground to the background, we used an offset to dampen the variation of the log-ratios for intensities close to zero.

2.3 Network Analysis

We calculated the graph consistency probability [9], which reflects the

consistency of a given network structure with the expression data of the constituent genes monitored in the present study, for each of the 146 network structures constructed using the previous ChIP-on-Chip data of four factors [10] and the gene classification scheme of the Molecular Signatures Database (MSigDB) [11]. The statistical significance level of the thresholds was set at 0.05.

2.4 Glycan analysis

We analyzed cell surface glycans using a lectin microarray [12]. Briefly, 43 lectins were spotted at a concentration of 0.5 mg/ml in triplicate onto epoxysilane-coated glass slides. Cell membrane fractions labeled with Cy3 NHS ester were then incubated with the lectin microarray, and fluorescence images were acquired using an evanescent-field activated fluorescence scanner SC profiler. The fluorescence signal for each spot was quantified with background value correction, and the lectin signals were defined by the average of triplicate spots.

3 Results and Discussion

3.1 Gene expression signature of hiPSCs descended from different parent SCs

To determine the gene expression signatures of hiPSCs, a detailed genome-wide expression analysis was performed to compare iPSCs and their parental SCs from amniotic mesodermal (AM), placental artery endotherial (PAE), uterine endometrium (UtE), and MRC-5 (MRC) cell sources. In total, 51 cell samples of 13 cell lines (39 hiPSCs samples of 9 hiPSC lines [7, 8]) were studied in the present study, so as to statically compare the hiPSCs and parental SCs. Unsupervised hierarchical clustering of the gene expression data across the four hiPSC lines (AM, PAE, UtE, MRC) and their corresponding parental SCs revealed interesting patterns of gene expression heatmap (Fig. 1A). First, the hiPSCs were clearly distinguishable from their respective parental SCs. Second, the four hiPSC lines had gene expression profiles that were linked to those of their parental SCs, while the hiPSCs from different passages clustered more closely with each other than with the hiPSCs from the corresponding parental SCs (Fig. 1A). In support of these findings, a Pearson correlation analysis demonstrated that the gene expression profiles of hiPSCs from different passages were more closely related to each other than to the hiPSCs from the same parental SCs, (Fisher's z-transformation comparison of correlations) (data not shown).

Analyses of the differences in gene expression between the four hiPSC lines and the parental SC lines revealed that 8,287 (out of 16,483) genes in the AM cells, 7,249 genes in the MRC cells, 7,465 genes in the PAE cells, and 6,314 genes in the UtE cells showed significant differences between the hiPSC lines and the corresponding parental lines, as determined using the Student's *t*-test (for a false discovery rate [FDR] $< 5\%$ and requiring a$\geq$2.0 -fold change in expression between the cells). In total, 2,502 genes were categorized into a gene expression signature common to the above four gene sets with expression differences (Fig. 1B). In this expression signature, 62% of the genes (1549 genes) were upregulated and 38%

(953 genes) downregulated in the hiPSCs compared to the parental SCs.

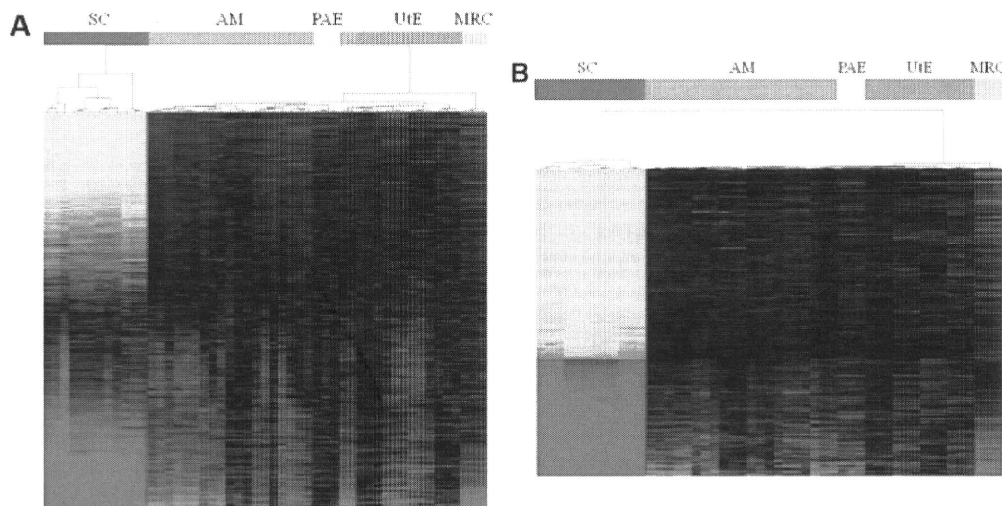


Figure 1 Gene Expression Signature. A: Heat map and hierarchical clustering for all cells and genes. Cell types are indicated by the bars, and the following abbreviations are used for the source of cell types of the human somatic cells (SCs) and induced pluripotent stem cells (hiPSCs): AM, amniotic membrane cell; PAE, placental artery cell; UtE, uterine endometrium cell; MRC, MRC-5 cell. **B**: Heat map of 2,502 genes in the hiPSCs and parental cells. Cell types are indicated by the bars.

3.2 Significant correlation between reprogramming and glycan biosynthesis based on network analysis

To elucidate the nature of the expression signature of the hiPSCs, we incorporated information on gene binding and function into a network analysis approach. To prepare the network analysis, we identified in the expression signature 146 regulatory networks of 313 genes, which were classified with their functions using the gene sets defined previously [11] (see also Methods), among 519 genes that were identified as being bound by four factors in ChIP-on-chip experiments [10]. We then analyzed the 146 reference networks, which were regarded as being directly induced by the four factors (OCT3/4, SOX2, KLF4, c-MYC), to define the network signature of the hiPSC, according to the two following thresholds: 1) the enrichment probability of the genes in the expression signature for each network; and 2) the consistency of the network structure in relation to the gene expression profile [9]. Thus, we defined as the network signature 28 networks of 76 genes that fulfilled these conditions (Fig. 2A).

As expected, the network signature almost completely covered the pathways that were previously implicated in the reprogramming of hiPSC pluripotency (Figs. 2A and B). For example, the relationship between reprogramming for pluripotency and signal transduction was emphasized for the TGF-β [13], Wnt [14], and MAPK

pathways [15]. In addition, pathways related to cell-cell interactions were implicated. Although the molecular mechanisms underlying the cell-cell interactions in the inner cellular states are less understood, several studies have reported the importance of cellular communication through the extracellular matrix with respect to changes in the cellular states, such as those that occur during development and differentiation [16]. Furthermore, relationships to cancer-related pathways were identified, which is consistent with the fact that the four factors induce various cancer cells [17]; this finding may be useful in the prevention of cancer induction by hiPSCs. Although several pathways in the network signature remain to be characterized, the network signature provides clues as to the molecular mechanisms underlying reprogramming for hiPSC pluripotency and self-renewal.

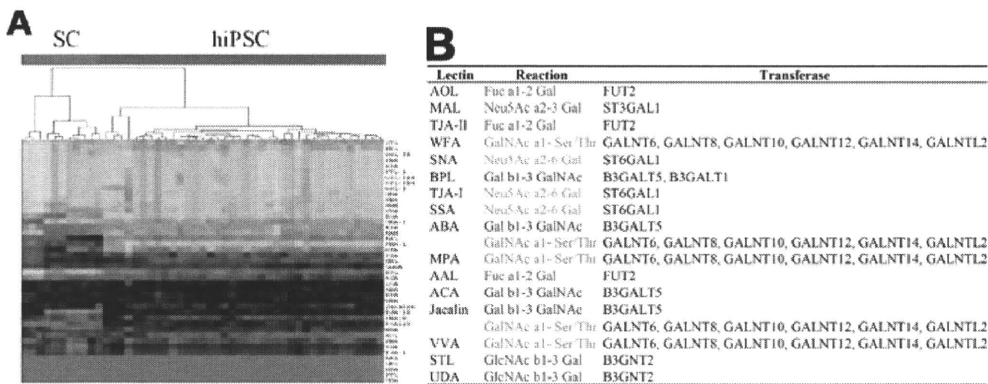
Figure 2 Network signature. **A**: List of network signatures. The pathways with significant probabilities are classified into the following categories: 2nd-7th row, signal transduction; 8th-12th, cell-cell interactions; 13th and 14th, glycan biosynthesis; 15th-21st, cancer; and otherwise, unclassified pathways. **B**: Schematic presentation of networks. The four induced factors are described in the center, and the binding genes, which are grouped according to the classification scheme described in **A**, are connected by thin lines.

Interestingly, two regulatory networks related to the glycome for linkage of the inner and outer cellular states appeared in the network signature (Fig. 2A). In general, glycan biosynthesis is a multi-step process that requires a variety of enzymes, i.e., glycosyltransferases and enzymes involved in cytosolic sugar metabolism, and in many cases, glycan biosynthesis follows a glycan-specific linear pathway. As most glycosyltransferases are regulated at the transcription level, an assessment of the transcriptional profile of glycan biosynthesis genes is warranted. In the two pathways, we found three genes (*ST6GAL1*, *B3GNT3*, and *GCNT2*) related to glycan transfer and two genes (*EXT1* and *HS6ST2*) related to heparan sulfate biosynthesis that were included in the expression signature. These findings are consistent with recent studies that have revealed associations between *N*-glycan and the maintenance of embryonic stem cell (ESC) pluripotency [18] and between heparan sulfate and the reprogramming of ESCs [19]. Therefore, the identified genes in the above two pathways are candidates for the maintenance of the outer

cellular state of iPSCs.

3.3 Unique glycan signatures of hiPSCs distinct from parent SCs

In addition to the expression and network signatures of the inner cell state, we examined the differences in the outer cellular states of the hiPSCs and parental SCs using a lectin array, which detects glycan structures on cell surface proteins based on glycan-lectin interactions [12]. In this analysis, the hiPSCs were clearly distinct from their parental SCs, and the dendrogram of the lectin microarray generated by unsupervised hierarchical clustering showed a clear separation between the hiPSCs and the parental SCs (Fig. 3A). Although the binding relationships between lectins and glycans and the relationships between the changes in glycan structures and the corresponding glycosyltransferases are redundant [20], we summarized the lectin-glycan-glycosyltransferase relationships using KEGG GLYCAN [21] and manual curation of previous papers. We found strong correlations between the gene expression profiles of the glycosyltransferases and the corresponding lectin fluorescence intensities (data not shown). This result indicates that glycosyltransferases are coordinately expressed with reprogramming, which the result that the hiPSCs bear glycan structures that are distinct from those of their parental SCs, reflecting reprogramming of the inner cellular state.



Lectin	Reaction	Transferase
AOL	Fuc a1-2 Gal	FUT2
MAL	Neu5Ac a2-3 Gal	ST3GAL1
TJA-II	Fuc a1-2 Gal	FUT2
WFA	GalNAc a1-Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
SNA	Neu5Ac a2-6 Gal	ST6GAL1
BPL	Gal b1-3 GalNAc	B3GALT5, B3GALT1
TJA-I	Neu5Ac a2-6 Gal	ST6GAL1
SSA	Neu5Ac a2-6 Gal	ST6GAL1
ABA	Gal b1-3 GalNAc	B3GALT5
	GalNAc a1-Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
MPA	GalNAc a1-Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
AAL	Fuc a1-2 Gal	FUT2
ACA	Gal b1-3 GalNAc	B3GALT5
Jacalin	Gal b1-3 GalNAc	B3GALT5
	GalNAc a1-Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
VVA	GalNAc a1-Ser/Thr	GALNT6, GALNT8, GALNT10, GALNT12, GALNT14, GALNTL2
STL	GlcNAc b1-3 Gal	B3GNT2
UDA	GlcNAc b1-3 Gal	B3GNT2

Figure 3 Glycan signature. **A**:Heat map and hierarchical clustering of lectins. The hiPSCs and parental SCs are discriminated. **B**: Correspondence between lectin gene expression patterns and glycan signatures. The lectin-glycosyltransferase relationships are described, together with their reactions. The lectins were selected under the condition that the corresponding glycosyltransferases were found in the expression signature.

Based on the Student's *t*-test (FDR <0.01) analysis, 28 of the 43 lectins in the lectin microarray showed significant differences between the hiPSCs and parental SCs (data not shown). We assigned 16 lectins to the glycan signature, which interacted with the 12 glycosyltransferases, which were related to the six patterns of glycan reactions, based on correspondence with the expression signature (Fig. 3B).

3.4 Molecular candidates for linkages between the inner and outer cellular states

Based on the correspondences between the expression and network signatures and between the expression and glycan signatures, we identified a total of 14 glycosyltransferases, owing to the appearance of *ST6GAL1* in both sets of correspondences. These glycosyltransferases are potential candidates for the linkage between the inner and outer cellular states in hiPSCs. Interestingly, these glycosyltransferases may be related to the biosynthesis of a glycolipid that is characteristic of hiPSCs. Indeed, allocation of the above glycosyltransferases to the pathways of "Glycan Biosynthesis and Metabolism" in KEGG GLYCAN revealed that the glycosyltransferases identified in the present study are important in glycolipid biosynthetic pathway. We identified *B3GALT5* in the biosynthetic pathway for the carbohydrate chains of the globo-series glycosphingolipids that carry the well-known SSEA-3 and SSEA -4 epitopes for ESCs and iPSCs, [22 ,23], and although *FUT2* is not directly involved in the synthesis of these glycans, it was found in the neighboring pathway that leads to the type IV H antigen. Furthermore, *B3GALT1* and *GCNT2*, in addition to *B3GALT5* and *FUT2*, were found in the extensive biosynthetic pathway of the carbohydrate chains of the lacto- and neolacto-series glycosphingolipids that carry SSEA-1, which is intensively expressed in ESCs but is absent in cells that have differentiated from ESCs [24]. In addition, members of the *GALNT* family, responsible for the *O*-glycan biosynthetic pathway of sialyl-T antigen, which is the most abundant glycan in several carcinoma cell lines, and *ST6GAL1* were only found in the *N*-glycan biosynthetic pathway, which is involved in the generation of cell-surface carbohydrate determinants and the differentiation antigens HB-6, CDw75, and CD76 [25]. These analyses identify the glycosyltransferases that are directly and indirectly related to known glycan epitopes, thereby indicating the key molecules and the marker epitopes involved in reprogramming.

Acknowledges

We thank to H. Abe, M. Yamazaki - Inoue and M. Machida, H. Sakaguchi for providing expert technical assistance. This work was supported by the project grant titled 'Development of Analysis Technology for Induced Pluripotent Stem (iPS) Cell', from The New Energy and Industrial Technology Development Organization (NEDO).

References

[1] Takahashi K, et al.: Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007, 131:861–872.

[2] Lowry WE, et al.: Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci USA* 2008, 105:2883–2888.

[3] Huangfu D, et al.: Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol* 2008, 26:1269–1275.

[4] Kim J, Chu J, Shen X, Wang J, Orkin SH: An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 2008, 132:1049–1061.

[5] Chen X, et al.: Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008, 133:1106–1117.

[6] Mikkelsen TS, et al.: Dissecting direct reprogramming through integrative genomic