Figure 8. Structures of neutral, monosialyl, and disialyl PA-oligosaccharides in iPSCs, iPSC-CM, and heart cells. Glucose units (GU) were calculated from the peak elution times for the ODS column in Figure 5, 6 and 7, and the amide column (data not shown). Average mass (Mass) calculated from the *m/z* values of $[M+Na]^+$ or $[M+H]^+$ ion for neutral, $[M-H]^-$ ion for monosialyl, and $[M-H]^-$ & $[M+Na-2H]^-$ ions for disialyl PA-oligosaccharides.
doi:10.1371/journal.pone.0111064.g008

iPSC-CMs (959A2-1 CM: 77.4%, 959C1-1 CM: 60.0% and 956F-1 CM: 65.1%), and lowest in the Heart (46.9%). The quantity of monofucosylated, difucosylated, and other types of *N*-glycans were greater in the iPSC-CMs and Heart (Figure 8, 9).

Sialyl *N*-glycans increased with cardiomyogenic differentiation

The quantity of monosialyl *N*-glycans (MS) calculated from the total volume of M1–M23 increased in iPSC-CMs (959A2-1 CM: 6.4%, 959C1-1 CM: 15.7% and 956F-1 CM: 10.5%) and Heart (19%) and were low in iPSCs (959A2-1: 0.5%, 959C1-1: 0.7% and 956F-1: 1.1%). The disialyl *N*-glycans (DS; D1–D12) yielded a similar pattern. The quantity of asialyl *N*-glycans (AS; N1–N17) decreased in iPSC-CMs (959A2-1 CM: 89.2%, 959C1-1 CM: 79.4% and 956F-1 CM: 81.7%) and Heart (55.3%) in comparison to the iPSCs (959A2-1: 96.9%, 959C1-1: 98.1% and 956F-1: 95.8%) (Figure 9, 10).

Rarely expressed *N*-glycans

The sialic acids identified in this study were either *N*-acetyl neuraminic acid (NeuAc) or *N*-glycolyl neuraminic acid (NeuGc). The quantity of monosialyl and disialyl *N*-glycans containing only NeuAc (A, A/A) was lowest in iPSCs (959A2-1: 2.5%, 959C1-1: 1.7% and 956F-1: 3.7%) and similar in iPSC-CMs (959A2-1 CM: 10.6%, 959C1-1 CM: 21% and 956F-1 CM: 18%) and the Heart (8%). The quantity of monosialyl and disialyl *N*-glycans containing only NeuGc (G, G/G) was markedly higher in the Heart (32.8%) than in iPSCs (959A2-1: 0.6%, 959C1-1: 0.1% and 956F-1: 0.5%) or iPSC-CMs (959A2-1 CM: 0%, 959C1-1 CM: 0% and 956F-1 CM: 0%) (Figure 10a).

Expression of glycosyl transferase, ST3Gal-III, ST3Gal-IV, ST6Gal-I, and CMAH in the iPSCs, iPSC-CMs, and Heart was assessed by RT-PCR to explore the glycan structures responsible for the differences between groups. The Heart expressed high levels of CMAH (0.91±0.13/GAPDH); levels in the iPSCs and iPSC-CMs were markedly lower (iPSCs: 959A2-1 0.011±0.0065/GAPDH, 959C1-1 0.013±0.0070/GAPDH, 956F-1 0.0045±0.0042/GAPDH, $P<0.05$; iPSC-CM: 959A2-1 CM 0.21±0.16/GAPDH, 959C1-1 CM 0.19±0.04, 956F-1 CM 0.45±0.31, $P<0.05$). Expression of ST3Gal-III was significantly higher in the Heart (0.98±0.13/GAPDH) than in iPSCs (959A2-1: 0.21±0.05/GAPDH, 959C1-1: 0.18±0.07/GAPDH, 956F-1: 0.27±0.05/GAPDH) and iPSC-CMs (959A2-1 CM: 0.40±0.10/GAPDH, 959C1-1 CM: 0.35±0.09/GAPDH, 956F-1 CM: 0.66±0.18); expression of ST3Gal-IV did not differ between groups. ST6Gal-I expression was significantly higher in iPSC-CMs (959A2-1 CM: 1.87±0.41/GAPDH, 959C1-1 CM: 1.95±0.22/GAPDH, 956F-1 CM: 3.08±1.27/GAPDH) than in iPSCs (959A2-1: 0.51±0.18/GAPDH, 959C1-1: 0.40±0.09/GAPDH, 956F-1: 0.62±0.29/GAPDH) and the Heart (1.04±0.13/GAPDH) (Figure 10b).

Discussion

Sixty-eight different *N*-glycans were isolated from iPSCs, iPSC-CMs, and the Heart. The structures of 60 *N*-glycans were identified, based on their HPLC elution peaks (Figure 8, Table S1-S5). Each preparation contained a combination of neutral, monosialyl, and disialyl *N*-glycans.

The molar ratios of high-mannose, monofucosylated, and difucosylated *N*-glycans were substantially different between groups (Figure 9), although no clear differences in the abundance of these glycans were found. The decrease in high-mannose *N*-glycans and increase of fucosylated *N*-glycans in iPSC-CMs versus iPSCs is consistent with a previous report on a comparison of ESC derived cardiomyocytes to undifferentiated ESCs [18]. Generally, all *N*-glycans are synthesized from the high-mannose type by a large array of sequentially and competitively acting biosynthetic enzymes located throughout the endoplasmic reticulum and Golgi apparatus [26], indicating that the high-mannose type of *N*-glycans could be categorized as a marker of immaturity. In this study, the high-mannose *N*-glycans were highest in the immature iPSC and lowest in the Heart, or mature tissue; thus, the quantity of high-mannose-type *N*-glycans might be an indicator of maturity in iPSC-derivatives and the iPSC-CMs in our protocol may still be immature in comparison to cardiac tissue.

Clear differences in glycan abundance were observed, such as hybrid and complex types represented by N9-1, N9-3, N15, N16, M1, M2-1, M2-2, M7, M8, M10, M12, M13, M14-1, M14-2, M17, M18, M20-2, D6 and D9 in iPSC-CMs, M2-3, M3, M4, M9, M11-1, M11-2, M20-1, M21, D1, D2, D3, D5-1, D5-2, D10-2 and D11 in Heart and N14 and M15 in iPSCs; these may also be indicators of maturation stage. In addition, expression of monosialyl and disialyl *N*-glycans in iPSC-CMs fell between the levels observed in the iPSCs and Heart, as were the molar ratios, indicating that the iPSC-CMs may still be immature stage. While many *N*-glycolyl neuraminic acid (NeuGc) structures were detected in the Heart, iPSCs and iPSC-CMs did not contain NeuGc in their sialyl structures, except for D8. Moreover, the molar ratio of NeuAc was low in iPSCs and iPSC-CMs. This finding is one of the clearest differences between iPSCs or iPSC-CMs and Heart cells.

The proposed spectra-based composition of the D8 glycans in iPSCs was [(Hexose)5(HexNAc)5(NeuGc)2(PA)1], indicating that it contains NeuGc. However, D8 might be quite a rare exception because transcript levels of CMAH, which catalyzes the conversion of NeuAc to NeuGc, was quite low in iPSCs in comparison to the Heart. This data suggests that during the process of reprograming, iPSCs suppress or eliminate CMAH activity. We conclude that iPSCs contain less sialic acid (especially NeuGc) and high-mannose structures are abundant in the *N*-glycans. In contrast, heart cells produce numerous sialyl-*N*-glycans, especially NeuGc. Transcript levels of CMAH tended to increase in iPSC-CMs relative to iPSCs, suggesting cardiomyogenic differentiation may induce expression of CMAH. If the iPSC-CMs could be matured more closely to the Heart by some additional methods of culture, the quantity of high mannose type of *N*-glycans might decrease more closely to the Heart, and might produce *N*-glycans containing NeuGc, followed by the expression of CMAH.

A terminal NeuGc, the Hanganutziu-Deicher (H-D) epitope [27], is widely distributed in the animal kingdom with the exception of humans and chickens. Expression of NeuGc is controlled by CMAH activity. Irie et al. [28] and Chou et al. [29] cloned the cDNA for human CMAH and reported that the *N*-terminal truncation of human CMAH is caused by deletion of Exon 6, a 92-base pair segment in the genomic DNA. Expression of this truncation in the heart eliminates NeuGc in sialyl

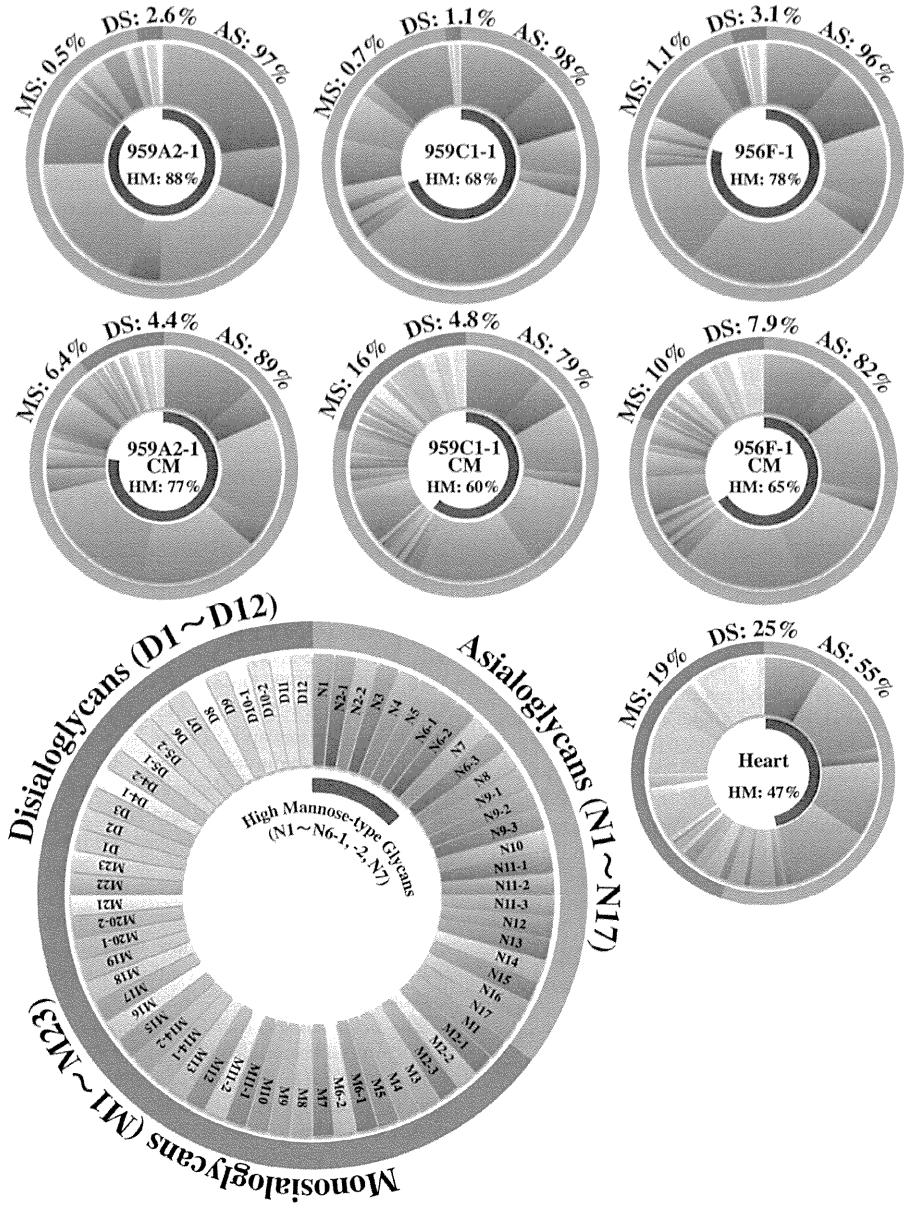


Figure 9. Relative quantities of neutral, monosialyl, and disialyl PA-oligosaccharides in iPSCs, iPSC-CM, and heart cells. Relative quantities of each glycan, calculated from the peak area in Figure 5, 6 and 7 vs. total *N*-glycan content in each cell, were expressed in the doughnut charts. Relative quantities of the asialoglycans, the monosialoglycans and the disialoglycans were showed outside of the charts, and relative quantities of the high mannose type glycans were showed inside of the charts. Asialoglycan (AS): the total volume of N1-N17; Monosialoglycan (MS): the total volume of M1-M23; Disialoglycan (DS): the total volume of D1-D12, High mannose-type glycan (HM): the total volume of N1–N6-1, N6-2, N7.
doi:10.1371/journal.pone.0111064.g009

structures. If human iPSCs or iPSC-CMs do not express CMAH in the same way as murine iPSCs or iPSC-CMs, there may be no difference between human iPSCs, iPSC-CMs, and the human Heart. Further study on human iPSC-CM will be needed to completely understand the features of the sialyl acid of *N*-glycans.

It was reported that human iPSCs produced α2,6sialyl glycans but did not contain α2,3sialyl structures, in contrast to human fibroblast, the origin of iPSCs, which produced α2,3sialyl but not α2,6sialyl structures [30,31]. The murine iPSCs in this study contained α2,3sialyl structures in NeuAc, M5, M23, D4-1, D10-1 and D12, and the iPSC-CMs produced α2,3 and α2,6sialyl structures in NeuAc. These differences may be due to variations between species, because mouse Heart cells also contained α2,3 and α2,6sialyl structures in NeuGc. Further studies are needed to characterize the glycome shift in the production and differentiation of iPSCs.

Type I Lactose structures were not detected, although over 98% of glycans in each cell were accounted for in this study. The *N*-glycans of N9-3, M8, M12, M17, and M23, which were identified after α-galactosidase digestion, contained Galα1-6Gal, not only in

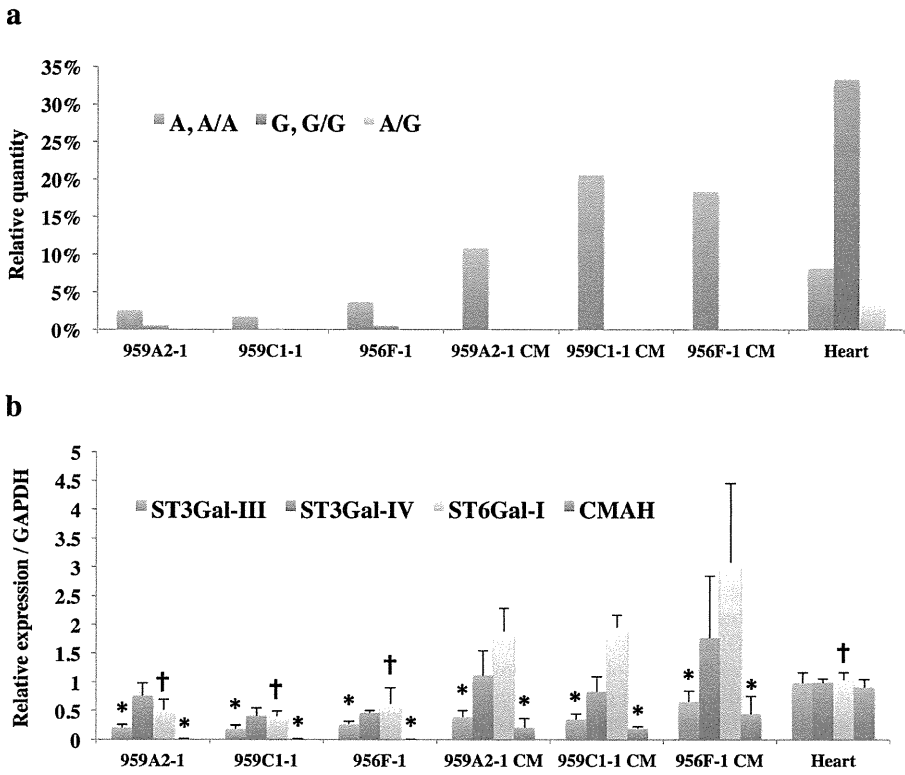


Figure 10. Rarely expressed NeuGc-containing glycans in iPSCs and iPSC-CMs. (a) Relative quantities of NeuAc- and NeuGc-containing glycans; Monosialoglycans containing NeuAc and Disialoglycans containing two NeuAc (A, A/A): the total volume of M1, M2-1, M2-2, M5-M8, M10-M14, M16-M19, M20-2, M21-M23, D4-1, D4-2, D6, D7, D9, D10-1, D12, Disialoglycan containing NeuAc and NeuGc (A/G): D11, Monosialoglycan containing NeuGc and Disialoglycan containing two NeuGc (G, G/G): the total volume of M2-3, M3, M4, M9, M15, M20-1, D1-D3, D5-1, D5-2, D8, D10-2. (b) Transcript expression of ST3Gal-III, ST3Gal-IV, ST6Gal-I, and CMAH; Transcript expression of glycosyltransferases in iPSCs, iPSC-CM, and heart cells was analyzed by real-time PCR. Results are expressed as the mean ± standard deviation. *P<0.05 vs. Heart, †P<0.05 vs. iPSC-CM (all of the 959A2-1 CM, 959C1-1 CM and 956F-1 CM).
doi:10.1371/journal.pone.0111064.g010

the neutral glycans but also in the monosialyl *N*-glycans of the iPSC-CM preparation. The same structure was not found in iPSCs, but only one structure, M23, was present in Heart cells. Therefore, in iPSC-CMs, Galα1-6Gal enzyme activity appears to be up-regulated in comparison to wild-type myocardium, although enzyme activity was not assessed by RT-PCR because of the limited availability of genetic sequence data.

The D8 was identified in all of three iPSC lines and not in the iPSC-CMs and Heart. This structure, unfortunately not identified in this study, may be useful as markers of undifferentiated iPSCs in the same way as well-known pluripotency biomarkers such as stage-specific embryonic antigens (SSEA)-3, SSEA-4 (glycosphingolipids) [32].

Previous MALDI-TOF/MS and MS/MS studies concluded that many kinds of *N*-glycans are found in organs and cells. The number of detected *N*-glycans is attributed to the sensitivity of the MS and HPLC methods employed. That is, MS data are sensitive and can be rapidly obtained, but a glycan structure is identified based only on the calculated molecular weight. Therefore, discriminating between isomeric structures is difficult. On the other hand, it thus appears that the accuracy of the data presented here using HPLC mapping in conjunction with a MALDI-TOF technique provides much more detailed information. Our data were used to identify the representative features of each *N*-glycan in these three cell types.

There may be a concern that the heart tissue used in this study contains connective tissues, vessels or nerves other than cardiomyocytes. Therefore, some of the *N*-glycans detected from the Heart sample might be derived from the tissues other than cardiomyocytes. However, heart is majority composed by cardiomyocytes, and furthermore, even if a small amount of *N*-glycans derived from connective tissues were contaminated in the Heart sample, the main evidences in this study, such as the proportion of the high-mannose type *N*-glycans, the ratio of the active sialyltransferase genes, the existence of NeuGc, and the uncommonness of Galα1-6 Gal, are essentially not affected.

In summary, murine iPSCs were rich in high-mannose type *N*-glycans but very poor in sialyl type *N*-glycans. Murine heart tissue contained a relatively low volume of high-mannose glycans, but was very rich in neuraminic acid, especially NeuGc type sialyl structures. Under these conditions, the volume of each type of glycan was similar for iPSC-CMs and iPSCs. That is, they were rich in high-mannose and relatively poor in sialyl type *N*-glycans by volume. In addition, most of the sialyl structures of the iPSC-CMs were different from those of the Heart, and the iPSC-CMs expressed no NeuGc. Moreover, the iPSC-CMs produced several unique glycans with the Galα1-6Gal structure. These results

provide important data that can be useful in future clinical iPSC studies.

It is quite important to investigate the meaning of *N*-glycans transitions during the cardiomyogenic differentiation presented in this study, for deeply understanding the relationship between the *N*-glycan expression and cardiomyogenic differentiation. Knock-out or knock-down of the genes related to cardiomyogenic differentiation or glycosylation may be useful for such purpose. However, the *N*-glycan signature in the cell surface is determined by a variety of the genes. Knock-out or knock-down of a single gene related to cardiomyogenic differentiation would alter an array of gene expressions, such as sarcomere proteins, transcriptional factors, or cell surface proteins, all of which would affect the signature of *N*-glycans in the cell surface. Therefore, the data interpretation for relationship between expression of a single gene and *N*-glycan signature would be difficult. Some different experimental approach may be needed to investigate the meaning of change in *N*-glycan expression during cardiomyogenic differentiation.

Supporting Information

Table S1 Structures and relative quantities of neutral (Table S1, S2) PA-oligosaccharides derived from iPSC, iPSC-CM, and heart cells. a. Glucose units (GU) were calculated from the peak elution times of the peaks obtained from the ODS column in Figure 5, 6, 7 and the Amide column (data not shown). b. Average mass calculated from the *mlz* values of $[M+Na]^+$ or $[M+H]^+$ ion for neutral, $[M-H]^-$ ion for mono-sialyl, and $[M-H]^-$ & $[M+Na-2H]^-$ ions for di-sialyl PA-oligosaccharides. c. PA-oligosaccharide structures. d. mol% was calculated from the peak area versus total *N*-glycan content in each cell
(TIFF)

Table S2 Structures and relative quantities of neutral (Table S1, S2) PA-oligosaccharides derived from iPSC, iPSC-CM, and heart cells.
(TIFF)

Table S3 Structures and relative quantities of mono-sialyl (Table S3, S4) PA-oligosaccharides derived from iPSC, iPSC-CM, and heart cells.
(TIFF)

Table S4 Structures and relative quantities of mono-sialyl (Table S3, S4) PA-oligosaccharides derived from iPSC, iPSC-CM, and heart cells.
(TIFF)

Table S5 Structures and relative quantities of disialyl PA-oligosaccharides derived from iPSC, iPSC-CM, and heart cells.
(TIFF)

Video S1
(MP4)

Acknowledgments

Our deepest appreciation goes to Professor Shinya Yamanaka and Keisuke Okita of the Center for iPS Cell Research and Application, Kyoto University, who kindly provided the murine iPSCs. We also thank Sachiko Kondo and Uichiro Yabe of MBL, Nagoya, Japan, who gave invaluable comments regarding *N*-glycan analysis.

Author Contributions

Conceived and designed the experiments: TK S. Miyagawa S. Miyagawa JL YS. Performed the experiments: TK AY NK AK EI AM HE KT. Analyzed the data: TK S. Miyagawa YS. Contributed reagents/materials/analysis tools: TK AY JL. Wrote the paper: TK S. Miyagawa YS SF. Obtained permission for use of cell line: S. Miyagawa AS YS.

References

1. Gonzales C, Pedrazzini T (2009) Progenitor cell therapy for heart disease. Exp Cell Res 315: 3077–3085.
2. Shah AM, Mann DL (2011) In search of new therapeutic targets and strategies for heart failure: recent advances in basic science. Lancet 378: 704–712.
3. Yoshida Y, Yamanaka S (2010) Recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration. Circulation 122: 80–87.
4. Yoshida Y, Yamanaka S (2011) iPS cells: A source of cardiac regeneration. Journal of Molecular and Cellular Cardiology 50: 327–332.
5. Kawamura M, Miyagawa S, Miki K, Saito A, Fukushima S, et al. (2012) Feasibility, safety, and therapeutic efficacy of human induced pluripotent stem cell-derived cardiomyocyte sheets in a porcine ischemic cardiomyopathy model. Circulation 126: S29–37.
6. Mercola M, Colas A, Willems E (2013) Induced pluripotent stem cells in cardiovascular drug discovery. Circ Res 112: 534–548.
7. Sinnecker D, Goedel A, Laugwitz KL, Moretti A (2013) Induced pluripotent stem cell-derived cardiomyocytes: a versatile tool for arrhythmia research. Circ Res 112: 961–968.
8. Kamakura T, Makiyama T, Sasaki K, Yoshida Y, Wuriyanghai Y, et al. (2013) Ultrastructural maturation of human-induced pluripotent stem cell-derived cardiomyocytes in a long-term culture. Circ J 77: 1307–1314.
9. Kuzmenkin A, Liang H, Xu G, Pfannkuche K, Eichhorn H, et al. (2009) Functional characterization of cardiomyocytes derived from murine induced pluripotent stem cells in vitro. FASEB J 23: 4168–4180.
10. Varki A (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3: 97–130.
11. Haltiwanger RS, Lowe JB (2004) Role of glycosylation in development. Annu Rev Biochem 73: 491–537.
12. Ohtsubo K, Marth JD (2006) Glycosylation in cellular mechanisms of health and disease. Cell 126: 855–867.
13. Surani MA (1979) Glycoprotein synthesis and inhibition of glycosylation by tunicamycin in preimplantation mouse embryos: compaction and trophoblast adhesion. Cell 18: 217–227.
14. Akama TO, Nakagawa H, Sugihara K, Narisawa S, Ohyama C, et al. (2002) Germ cell survival through carbohydrate-mediated interaction with Sertoli cells. Science 295: 124–127.
15. Hato M, Nakagawa H, Kurogochi M, Akama TO, Marth JD, et al. (2006) Unusual N-glycan structures in alpha-mannosidase II/IIx double null embryos identified by a systematic glycomics approach based on two-dimensional LC mapping and matrix-dependent selective fragmentation method in MALDI-TOF/TOF mass spectrometry. Mol Cell Proteomics 5: 2146–2157.
16. Lau KS, Partridge EA, Grigorian A, Silvescu CI, Reinhold VN, et al. (2007) Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. Cell 129: 123–134.
17. Kraushaar DC, Rai S, Condac E, Nairn A, Zhang S, et al. (2012) Heparan sulfate facilitates FGF and BMP signaling to drive mesoderm differentiation of mouse embryonic stem cells. J Biol Chem 287: 22691–22700.
18. Amano M, Yamaguchi M, Takegawa Y, Yamashita T, Terashima M, et al. (2010) Threshold in stage-specific embryonic glycotypes uncovered by a full portrait of dynamic N-glycan expression during cell differentiation. Mol Cell Proteomics 9: 523–537.
19. Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S (2008) Generation of mouse induced pluripotent stem cells without viral vectors. Science 322: 949–953.
20. Miki K, Uenaka H, Saito A, Miyagawa S, Sakaguchi T, et al. (2012) Bioengineered myocardium derived from induced pluripotent stem cells improves cardiac function and attenuates cardiac remodeling following chronic myocardial infarction in rats. Stem Cells Transl Med 1: 430–437.
21. Yu T, Miyagawa S, Miki K, Saito A, Fukushima S, et al. (2013) In vivo differentiation of induced pluripotent stem cell-derived cardiomyocytes. Circ J 77: 1297–1306.
22. Tohyama S, Hattori F, Sano M, Hishiki T, Nagahata Y, et al. (2013) Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell 12: 127–137.
23. Takahashi N, Khoo KH, Suzuki N, Johnson JR, Lee YC (2001) N-glycan structures from the major glycoproteins of pigeon egg white: predominance of terminal Galalpha(1)Gal. J Biol Chem 276: 23230–23239.

24. Takahashi N, Kato K (2003) GALAXY(Glycoanalysis by the Three Axes of MS and Chromatography): a Web Application that Assists Structural Analyses of N-Glycans. Trends in Glycoscience and Glycotechnology 15 No.84: 235–251.
25. Yagi H, Takahashi N, Yamaguchi Y, Kimura N, Uchimura K, et al. (2005) Development of structural analysis of sulfated N-glycans by multidimensional high performance liquid chromatography mapping methods. Glycobiology 15: 1051–1060.
26. Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA (2014) Emerging principles for the therapeutic exploitation of glycosylation. Science 343: 1235681.
27. Varki A (2009) Multiple changes in sialic acid biology during human evolution. Glycoconj J 26: 231–245.
28. Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A (1998) The molecular basis for the absence of N-glycolylneuraminic acid in humans. J Biol Chem 273: 15866–15871.
29. Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, et al. (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. Proc Natl Acad Sci U S A 95: 11751–11756.
30. Hasehira K, Tateno H, Onuma Y, Ito Y, Asashima M, et al. (2012) Structural and quantitative evidence for dynamic glycome shift on production of induced pluripotent stem cells. Mol Cell Proteomics 11: 1913–1923.
31. Tateno H, Toyota M, Saito S, Onuma Y, Ito Y, et al. (2011) Glycome diagnosis of human induced pluripotent stem cells using lectin microarray. J Biol Chem 286: 20345–20353.
32. Fujitani N, Furukawa J, Araki K, Fujioka T, Takegawa Y, et al. (2013) Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. Proc Natl Acad Sci U S A 110: 2105–2110.

Biochemical and Biophysical Research Communications 456 (2015) 860–865

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Therapeutic potential of human adipose tissue-derived multi-lineage progenitor cells in liver fibrosis

Hanayuki Okura [a,b], Mayumi Soeda [a], Mitsuko Morita [a], Maiko Fujita [a], Kyoko Naba [a], Chiyoko Ito [a], Akihiro Ichinose [c], Akifumi Matsuyama [a,b,*]

[a] Platform of Therapeutics for Rare Disease, National Institute of Biomedical Innovation, 5-5-2-602 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
[b] The Center for Medical Engineering and Informatics, Osaka University, 2-2 Yamada-oka, Suita, Osaka 565-0879, Japan
[c] Department of Plastic Surgery, Kobe University Hospital, 7-5-2 Kusunoki-cho, Chuo-ku, Kobe, Hyogo, Japan

ARTICLE INFO

Article history:
Received 22 November 2014
Available online 6 December 2014

Keywords:
hADMPCs
Chronic hepatitis
Fibrosis
MMPs
Mouse

ABSTRACT

Introduction: Liver fibrosis is characterized by excessive accumulation of extracellular matrix. In a mouse model of liver fibrosis, systemic injection of bone marrow mesenchymal stem cells (BM-MSCs) was considered to rescue the diseased phenotype. The aim of this study was to assess the effectiveness of human adipose tissue-derived multi-lineage progenitor cells (hADMPCs) in improving liver fibrosis.

Methods and results: hADMPCs were isolated from subcutaneous adipose tissues of healthy volunteers and expanded. Six week-old male nude mice were treated with carbon tetra-chloride (CCl_4) by intraperitoneal injection twice a week for 6 weeks, followed by a tail vein injection of hADMPCs or placebo control. After 6 more weeks of CCl_4 injection (12 weeks in all), nude mice with hADMPCs transplants exhibited a significant reduction in liver fibrosis, as evidenced by Sirius Red staining, compared with nude mice treated with CCl_4 for 12 weeks without hADMPCs transplants. Moreover, serum glutamic pyruvate transaminase and total bilirubin levels in hADMPCs-treated nude mice were lower levels than those in placebo controls. Production of fibrinolytic enzyme MMPs from hADMPCs were examined by ELISA and compared to that from BM-MSCs. MMP-2 levels in the culture media were not significantly different, whereas those of MMP-3 and -9 of hADMPCs were higher than those by BM-MSCs.

Conclusion: These results showed the mode of action and proof of concept of systemic injection of hADMPCs, which is a promising therapeutic intervention for the treatment of patients with liver fibrosis.

1. Introduction

Various conditions such as viral hepatitis, chronic alcohol abuse, metabolic diseases, autoimmune diseases and bile duct epithelial injury can cause liver fibrosis [1,2]. Liver fibrosis is reversible, whereas cirrhosis, the end-stage result of fibrosis, is in general irreversible [3]. Liver fibrosis is characterized by excessive accumulation of extracellular matrix, with the formation of scar tissue encapsulating the area of injury [4]. The prognosis of patients with liver fibrosis is poor, but liver transplantation seems to improve the prognosis [5,6]. However, limited numbers of donor livers are available for the millions of patients who need them worldwide [7]. Thus, there is a need for novel therapeutic approaches.

Recently, cell therapy has been proposed as an attractive tool for treatment of patients with severe liver disease [8–13]. Stem/progenitor cells, which possess certain characteristics including self-renewal, proliferation, longevity, and differentiation, are valuable in cell therapy [14]. Several groups have demonstrated the effectiveness of bone marrow-derived mesenchymal stem cells (BM-MSCs) in animal models of liver fibrosis and cirrhosis [15–18]. However, others have reported the lack of any changes in the extent of liver fibrosis or liver function tests following the use of BM-MSCs in a rat model of severe chronic liver injury [19]. Thus, the therapeutic efficacy of BM-MSCs transplantation remains controversial at present [19].

Adipose tissue-derived progenitor/stem cells are an attractive cell source for cell therapy of liver fibrosis, based on several properties of these cells; (1) ample production of fibrinolytic enzymes and cytokines [20], (2) ease of obtaining stem cells compared to other tissue-specific stem cells including BM-MSCs, [21]. The use of human adipose tissue-derived multi-lineage progenitor cells (hADMPCs) supports the view that cytokine production could mediate the therapeutic actions of hADMPCs in liver fibrosis.

* Corresponding author at: Platform of Therapeutics for Rare Diseases, National Institute of Biomedical Innovation, 5-5-2-602 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. Fax: +81 78 304 6176.
E-mail address: akifumi-matsuyama@umin.ac.jp (A. Matsuyama).

http://dx.doi.org/10.1016/j.bbrc.2014.11.122
0006-291X/

In the present study, we investigated the efficacy of treatment using hADMPCs in nude mice with CCl_4-induced chronic liver dysfunction and the mechanism of their action in improvement of liver fibrosis.

2. Materials and methods

2.1. Adipose tissue

Adipose tissue samples were resected from 7 human subjects during plastic surgery (all females, age, 20–60 years) as excess discards. About 10–50 g subcutaneous adipose tissue was collected from the sample of each subject. All subjects provided informed consent. The protocol was approved by the Review Board for Human Research of Kobe University, Graduate School of Medicine, Osaka University, Graduate School of Medicine and National Institute of Biomedical Innovation, Japan.

2.2. Isolation and expansion of hADMPCs

hADMPCs were prepared as described previously [8–10]. Briefly, the resected excess adipose tissue was minced and then digested at 37 °C for 1 h in Hank's balanced salt solution (HBSS, GIBCO Invitrogen, Grand Island, NY) with Liberase (Roche Diagnostics, Germany) as indicated by the manufacturer. Digests were filtered through a cell strainer (BD Bioscience, San Jose, CA) and centrifuged at 800×*g* for 10 min. Red blood cells were excluded using density gradient centrifugation with Lymphoprep ($d = 1.077$; Nycomed, Oslo, Norway), and the remaining cells were cultured in Dulbecco's modified Eagle's medium (DMEM, GIBCO Invitrogen) with 10% defined fetal bovine serum (FBS, Biological Industries, Israel) for 24 h at 37 °C. Following incubation, the adherent cells were washed extensively and then treated with 0.2 g/l ethylenediaminetetraacetate (EDTA) solution (Nacalai Tesque, Kyoto, Japan). The resulting suspended cells were replated on retronectin (RN)-coated dishes (Takara, Kyoto, Japan) in SteMedis (Nipro, Osaka, Japan), 1× insulin-transferring selenium (Nipro, Osaka), 1 nM dexamethasone (MSD, Tokyo, Japan), 100 μM ascorbic acid 2-phosphate (Sawai Pharmaceuticals Co., Osaka), 10 ng/ml epidermal growth factor (EGF, PeproTec, Rocky Hill, NJ), and 5% FBS (FBS, Biological Industries, Israel). The culture medium was changed twice a week and then the cells were applied for the experiments after 5–6 passages.

2.3. Flow cytometric analysis of hADMPCs

hADMPCs were characterized by flow cytometry. Cells were detached and stained with anti-human CD31, CD34, CD44, CD45, CD56, CD73, CD90, CD105 or CD166 antibodies (BD Lyoplate™ Screening Panels, BD Bioscience, San Jose, CA). Isotype-identical antibodies served as controls. After washing with Dulbecco's phosphate-buffered saline (PBS, Nacalai Tesque), cells were incubated with PE-labeled goat anti-mouse Ig antibody (BD PharMingen) for 30 min at 4 °C. After three washes, the cells were resuspended in PBS and analyzed by flow cytometry using a guava easyCyte flow cytometry systems (Merck Millipore, Darmstadt, Germany).

2.4. Adipogenic, osteogenic and chondrogenic differentiation procedure

Tri-lineage differentiation was examined as described previously [22]. Briefly, for adipogenic differentiation, the cells were cultured in Differentiation Medium (Zen-Bio, Inc.). After three days, half of the medium was replaced with Adipocyte Medium (Zen-Bio, Inc.) every two days. Five days after differentiation, characterization of adipocytes was confirmed by microscopic observation of intracellular lipid droplets after Oil Red O staining. Osteogenic differentiation was induced by culturing the cells in DMEM containing 10 nM dexamethasone, 50 mg/dl ascorbic acid 2-phosphate, 10 mM β-glycerophosphate (Sigma), and 10% FBS. Differentiation was examined by Alizarin red staining. For chondrogenic differentiation, 2×10^5 cells of the hADMPCs were centrifuged at 400×*g* for 10 min. The resulting pellets were cultured in chondrogenic medium (α-MEM supplemented with 10 ng/ml TGF-β, 10 nM dexamethasone, 100 M ascorbate, and 10 μl/ml 100× ITS Solution) for 14 days. For Alcian Blue staining, nuclear counter-staining with Weigert's hematoxylin was followed by 0.5% Alcian Blue 8GX for proteoglycan-rich cartilage matrix.

2.5. Animal model of liver fibrosis and cell administration

Chronic liver fibrosis was induced in nude mice using the procedure described previously [23,24] with some modification. Briefly, 6-week-old male nude mice (body weight of 20–30 g purchased from CLEA, Tokyo) were treated with a mixture of CCl_4 (Wako Pure Chemicals, Osaka) (0.3 ml/kg) and olive oil (Wako Pure Chemicals) (1:1 vol/vol) by intra-peritoneal injection twice a week for 6 weeks, and this was followed by a tail vein injection of hADMPCs (1.0×10^6 cells/kg body weight, $n = 4$) or placebo control ($n = 5$), and followed by 6 more weeks of CCl_4 treatment.

2.6. Liver function tests and histological analysis

Blood specimens were collected by cardiac puncture at the end of the experiment. Measurement of serum albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and total-bilirubin levels by routine laboratory methods was outsourced to Oriental Yeast Co. (Shiga, Japan).

Hematoxylin and eosin (H&E) staining and Sirius Red (SR) staining were performed to determine the extent of liver inflammation and fibrosis. The stained slides were viewed on a BioZero laser scanning microscope (Keyence, Osaka). The area of liver fibrosis was quantified with SR staining. Briefly, the fibrotic area (red staining) was assessed at 40× magnification using computer-assisted image analysis with All-in-One analysis software (Keyence, Osaka). Sixteen fields were randomly selected for each group.

2.7. Measurement of MMP-2, -3 and -9 production by hADMPCs

One million cells of hADMPCs and BM-MSCs (DS Pharma Biomedical, Osaka) were seeded onto 6 well plates and then cultured for 24 h. The supernatants were harvested, centrifuged, and frozen at −80 °C until analysis. MMP-2, MMP-3 and MMP-9 were measured by enzyme-linked immunosorbent assay (ELISA) kits from R&D Systems (Minneapolis, MN) using the instructions supplied by the manufacturer.

2.8. Statistical analysis

Serum parameters and fibrotic area are presented as mean ± SD. Differences between groups were assessed for statistical significance by the Student's *t*-test, with $p < 0.05$ considered statistically significant.

3. Results

3.1. Characterization of hADMPCs

Flow cytometry was used to assess markers expressed by hADMPCs (Fig. 1A). The cells were negative for markers of

hematopoietic lineage (CD45) and hematopoietic stem cells, CD34 and CD133. They were also negative for CD31, an endothelial cell-associated marker, and c-Kit (CD117), a cell surface antigen. However, they stained positively for several surface markers characteristic of mesenchymal stem cells, but not embryonic stem (ES) cells, such as CD29, CD44 (hyaluronan receptor), CD73 and CD105 (endoglin).

Next, we examined the adipogenic, osteogenic and chondrogenic differentiation potentials of hADMPCs. Adipogenic differentiation was confirmed by accumulation of intracellular lipid droplets stained with Oil Red O (Fig. 1B). Differentiation and induction of hADMPCs was associated with increase in the amount of Oil Red O-stained lipid droplets, indicating that hADMPCs can differentiate into adipocytes. Osteogenic induction was examined by Alizarin red S staining (Fig. 1B). Induction of hADMPCs for osteogenesis was associated with Alizarin red S staining and appearance of mineralized nodules. The chondrogenic potential of hADMPCs is shown in Fig. 1B. Induction of chondrogenesis by pellet culture resulted in staining of extracellular matrices of hADMPCs-derived pellet-cultured chondrocytes for Alcian Blue, indicating the chondrogenic differentiation potential of hADMPCs. These results confirmed the tri-lineage differentiation potential of hADMPCs and the mesenchymal stem cell properties of hADMPCs.

3.2. Effects of hADMPC on CCl_4-induced chronic liver dysfunction in nude mice

We adopted the CCl_4-induced chronic mouse fibrosis model in this study rather than the CCl_4-induced acute model because CCl_4-induced acute liver fibrosis resolves spontaneously [25]. For this purpose, 9 male nude mice were injected intraperitoneally with CCl_4 twice weekly for 6 weeks, and then divided into two groups, 4 animals received hADMPCs transplantation via the tail vein and the other 5 vehicle control received Ringer's solution with 1/30 volume of heparin. All animals were followed for 6 weeks after the last injection (a).

H&E staining of liver sections showed reduced hepatocyte vacuolar degeneration in hADMPC-transplanted CCl_4-injured mice compared with the control (Fig. 2B). The peri-lobular regions were the main areas affected by CCl_4 hepatotoxicity while the centrilobular regions seemed to be the least affected. These findings suggest intact albumin secretion, which was confirmed by Sirius Red (SR) staining of control liver sections. SR staining of sections from hADMPC-transplanted mice showed mild liver fibrosis, while that of sections from control group mice showed moderate fibrosis (Fig. 2B). Quantitative image analysis of the fibrotic area in SR-stained sections confirmed the efficacy of hADMPC-transplantation on liver fibrosis. The mean fibrotic area was significant lower in hADMPC-transplanted CCl_4-injured mice ($1.8 \pm 1.1\%$ of fibrotic areas) than control mice ($10.9 \pm 3.9\%$ of fibrotic areas) ($p < 0.05$), indicating that hADMPC-transplantation ameliorated liver fibrosis and increased the area containing hepatocytes (c).

3.3. Functional recovery of liver damage following transplantation of hADMPCs

We next evaluated the effects of cell transplantation on the extent of liver injury and liver function. Serum transaminase levels (AST and ALT) were significantly higher in mice with liver damage (control), but the increase was attenuated by hADMPCs transplantation (Fig. 3A and B). These results confirmed the effectiveness of hADMPCs in the treatment of liver damage associated with fibrosis.

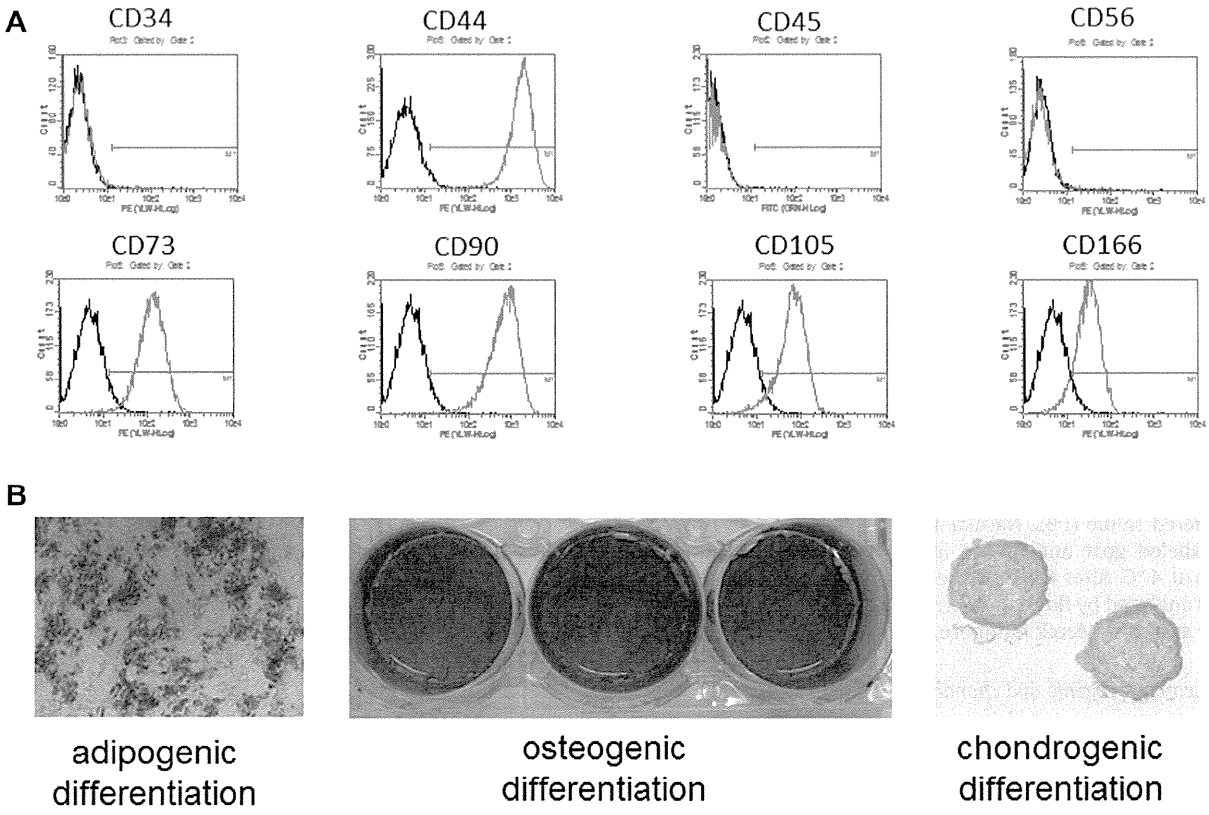


Fig. 1. Characterization of hADMPCs. (A) Flow cytometric characterization of hADMPCs. (B) An isotype-matched negative control indicated as red curve. (C) Adipogenic, osteogenic and chondrogenic differentiation potentials of hADMPCs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

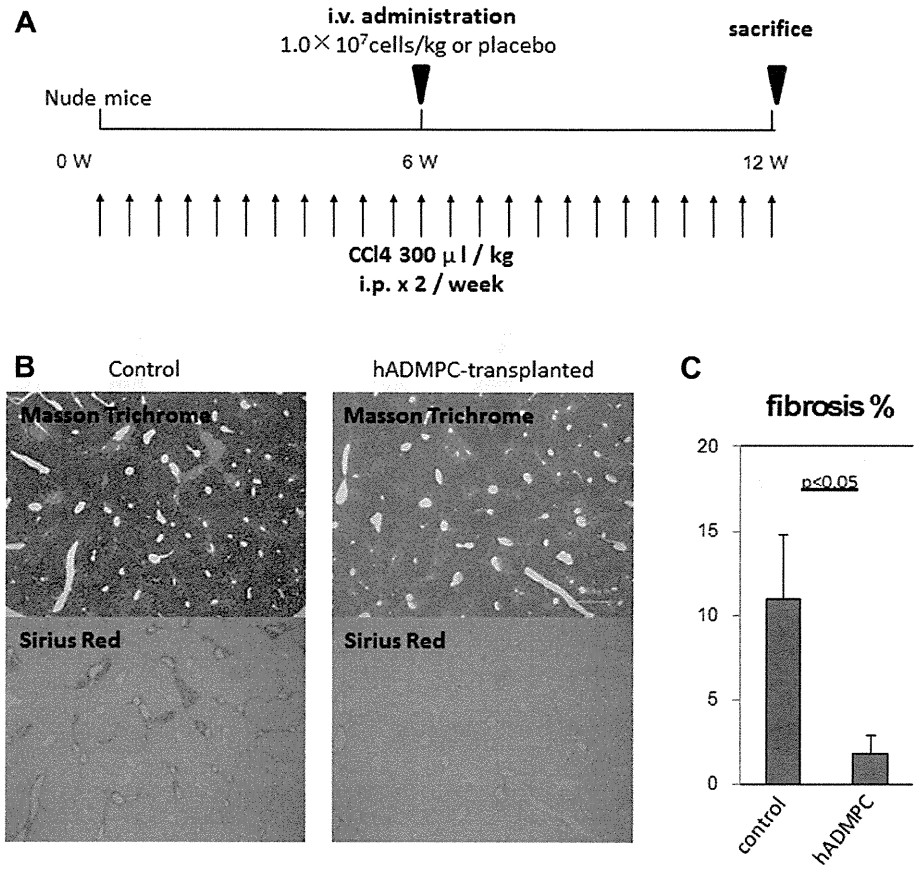


Fig. 2. Assessment of liver fibrosis in hADMPC-transplanted nude mice and controls. (A) Diagram of the treatment protocol. (B) Extracellular deposition of collagen fibers stained with Sirius Red. (C) Quantification of collagen by image analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Interestingly, serum albumin level remained high after hADMPCs, similar to the control (Fig. 3C). These results could be explained by damage of the centrilobular region, the main site of albumin production. Considered together, the results suggest the beneficial effects of hADMPCs in attenuating liver damage and recovery of liver function.

3.4. hADMPCs-induced functional recovery is mediated by MMP release

Finally, we analyzed the mechanism of the hepatoprotective effect of hADMPCs. For this purpose, we measured the amount of the fibrinolytic enzymes, MMP-2, MMP-3 and MMP-9, secreted by hADMPCs by ELISA (Fig. 4). After 3-day culture, the amounts of enzymes production by hADMPCs and BM-MSCs were measured. There was no significant difference in MMP-2 production by hADMPCs and BM-MSCs (59.7 ± 2.3 vs 58.3 ± 0.0 ng/ml from 1.0×10^4 cells cultured for 3 days). On the other hand, MMP-3 and MMP-9 production levels were significantly higher in hADMPCs than BM-MSCs (6.84 ± 2.3 vs 0.03 ± 0.0 ng/ml, $p < 0.05$, 0.462 ± 0.015 vs 0.003 ± 0.008 ng/ml, $p < 0.05$, from 1.0×10^4 cells cultured for 3 days, respectively).

4. Discussion

The major finding of the present study was improvement of liver fibrosis in CCl_4-induced mice after systemic administration of hADMPCs, and that this effect was mediated, at least in part, through the production of fibrinolytic MMP-2, -3, and -9, from hADMPCs, suggesting that these cells could be particularly effective in resolving liver fibrosis.

Liver transplantation is an established treatment for severe liver cirrhosis, although the number of patients who could benefit from such treatment is small due to the limited number of donors [5]. Cell therapy has been proposed as an alternative and attractive tool for treating patients with severe liver disease [8–13]. Among the cell therapy tested so far, hepatocyte replacement therapy had been examined. Isolated hepatocytes from human liver [13], regenerated hepatocyte-like or –progenitor cells from embryonic, induce pluripotent [26,27], or hepatic progenitor cells [11], and *in situ* reprogrammable cells (9, 10) have been tested for their efficacy in animal models. The strategy has also been successful in clinical trials involving patients with certain inherited diseases [28]. Although large numbers of hepatocytes or hepatocyte-like cells are needed for meaningful cure and there should be no room for the cells in fibrotic hepatic parenchyma to engraft, such replacement therapies, however, do not seem to be clinically fruitful for liver fibrosis. We hypothesized that fibrolytic enzymes produced by hADMPCs could be useful for treatment of liver fibrosis, and therefore shifted the treatment strategy to improvement of liver fibrosis with cell-based fibrinolytic enzymes delivery. In this strategy, hADMPCs derived-MMPs should produce lysis of excess extracellular matrices and make room for the patient's own proliferative hepatocytes.

To establish the cell-based fibrinolytic enzyme delivery therapy as first line next to liver transplantation, some challenging issues

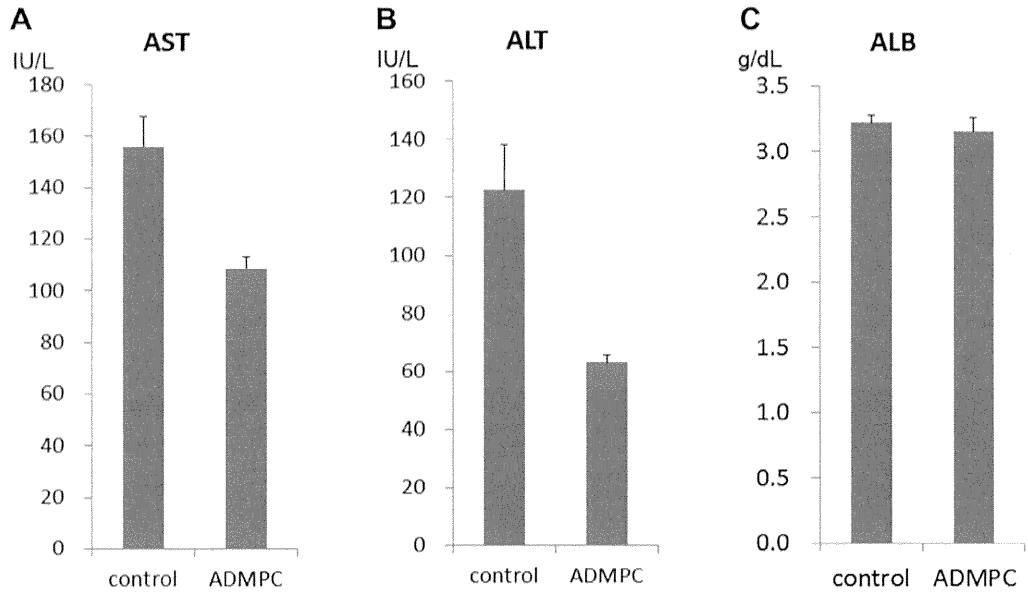


Fig. 3. Examination of serum parameters. (A, B) Transaminase (AST and ALT), (C) Serum albumin. Data are mean ± SD. C: control mice; T: mice transplanted mice with hADMPCs.

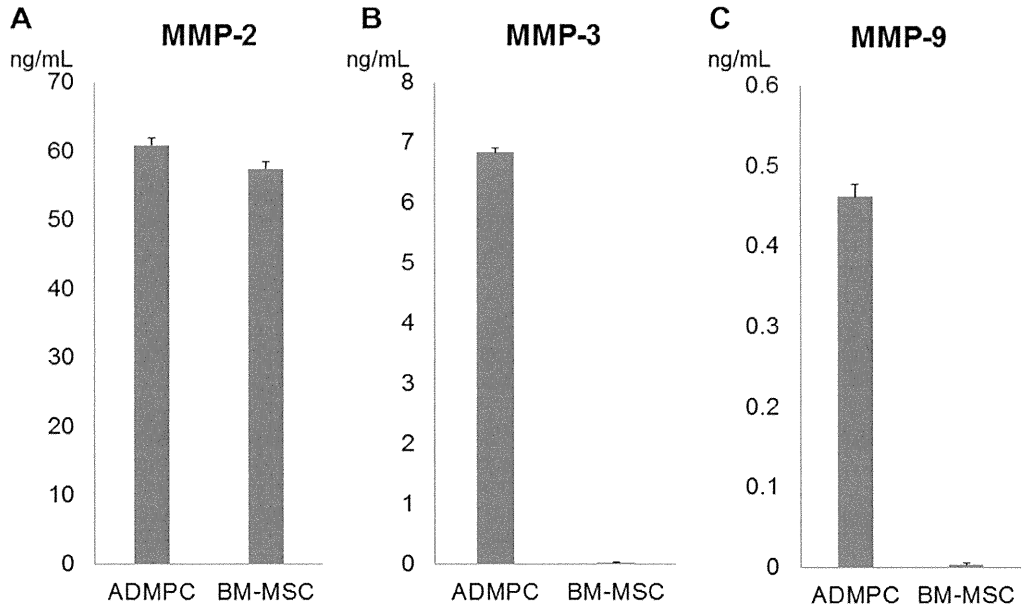


Fig. 4. Quantitative analysis of MMP-2, MMP-3 and MMP-9 level produced by hADMPCs and BM-MSCs. The amount of MMP-2 (A), MMP-3 (B) and MMP-9 (C) after 3 days of culture.

should be dealt with; (1) the cells should be obtained easily and ethically in large quantities, (2) the cells should improve liver fibrosis and liver panel, and (3) the cells act as vehicle for the delivery of MMPs.

The first issue is whether the cells could be obtained easily and ethically in large quantities. hADMPCs is favorable for the therapy because adipose tissue, from which ADMPCs are obtained, is easily and safely accessible and large quantities of the tissues can be obtained without serious ethical issues, since liposuction surgery yields from 100 ml to >3 L of lipoaspirate tissue [8–10]. Therefore, hADMPCs can potentially be applied not only for autologous but also allogenic cell-based enzyme delivery in the future. Based on the above advantages, hADMPCs represent a potentially promising source of cells for the therapy.

Second, we need to show that hADMPCs-administration results in improvement of liver fibrosis and liver panel, as proof-of-concept of therapy. As shown in Fig. 2, hADMPC significantly improved liver fibrosis in CCl_4-treated nude mice (a model of chronic liver cirrhosis). The treatment also resulted in improvement of serum transaminase levels. In this model, massive fibrosis was mainly noted in the peri-hepatic lobular region but not in the centrilobular regions surrounding the central veins. Albumin is known to be mainly produced by hepatocytes in the centrilobular region. This is the most likely reason for the lack of difference in serum albumin

levels between hADMPC-transplanted animals and controls. These results indicate that hADMPCs transplantation showed the proof-of-concept to liver dysfunction associated with fibrosis.

Finally, an important issue in this kind of therapy is whether the cells secrete sufficient amount of MMPs. One study reported that matrix metalloproteinase gene delivery could decrease collagen fibers and reduce liver fibrosis [29]. The mode of action was considered to be the strong expression MMPs on the transplanted cells, indicating that MMPs-producing cells other than BM-MSCs [30] are suitable for use for the cell-based enzyme delivery. The present study showed that hADMPCs expressed MMP-2, -3 and -9 (Fig. 2). There was no significant difference in MMP-2 production between hADMPC and BM-MSCs. However, the production of MMP-3 and MMP-9 from hADMPCs was superior to that from BM-MSCs. MMP-3 and MMP-9 are known to lyse collagen types III and I, which are major compartment of liver fibrotic lesion [29]. These data highlight the potential effectiveness of hADMPCs in the treatment of liver fibrosis and the superiority of hADMPCs compared to other therapies.

In conclusion, the present study demonstrated that systemic administration of hADMPCs significantly attenuated liver fibrosis and improved liver function, and that the therapeutic effect of hADMPCs was in part due the secretion of fibrinolytic enzymes, MMPs. These proofs of concept and mode of action prompted us to choose hADMPCs for cell therapy of liver fibrosis. hADMPCs therapy, as cell-based enzyme delivery therapy, has the potential to be an effective source of inducers that support liver regeneration.

Declaration

The authors declare no conflict of interest.

Acknowledgments

This study was partly supported by a grant-in-aid for A.M. from the Ministry of Health, Labor and Welfare, Japan.

References

[1] A. Kasahara, H. Tanaka, T. Okanoue, et al., Interferon treatment improves survival in chronic hepatitis C patients showing biochemical as well as virological responses by preventing liver-related death, J. Viral Hepat. 11 (2004) 148–156.
[2] T. Saito, K. Misawa, S. Kawata, Intern. Med. 6 (2007) 101–103.
[3] V. Manne, E. Akhtar, S. Saab, J. Clin. Gastroenterol. 8 (2014) e76–e84.
[4] A. Mallat, S. Lotersztajn, Am. J. Physiol. Cell Physiol. 305 (2013) C789–C799.
[5] C.L. Chen, S.T. Fan, S.G. Lee, et al., Living-donor liver transplantation: 12 years of experience in Asia, Transplantation 75 (2003) S6–S11.
[6] Y. Takada, M. Ueda, T. Ito, Living donor liver transplantation as a second-line therapeutic strategy for patients with hepatocellular carcinoma, Liver Transpl. 12 (2006) 912–919.
[7] R. Rai, Liver transplantation – an overview, Indian J. Surg. 75 (2013) 185–191.
[8] H. Okura, H. Komoda, A. Saga, Tissue Eng. Part C Methods 16 (2010) 761–770.
[9] H. Okura, A. Saga, Y. Fumimoto, Tissue Eng. Part C Methods 17 (2011) 145–154.
[10] A. Saga, H. Okura, M. Soeda, HMG-CoA reductase inhibitor augments the serum total cholesterol-lowering effect of human adipose tissue-derived multilineage progenitor cells in hyperlipidemic homozygous Watanabe rabbits, Biochem. Biophys. Res. Commun. 412 (2011) 50–54.
[11] R. Semeraro, V. Cardinale, G. Carpino, et al., The fetal livers as cell source for the regenerative medicine of liver and pancreas, Ann. Transl. Med. 1 (2013) 13.
[12] K. Sun, X. Xie, J. Xi, et al., Cell-based therapy for acute and chronic liver failures: distinct diseases, different choices, Sci. Rep. 4 (2014) 6494.
[13] C. Jorns, E.C. Ellis, G. Nowak, et al., Hepatocyte transplantation for inherited metabolic diseases of the liver, J. Intern. Med. 272 (2012) 201–223.
[14] M.F. Pittenger, A.M. Mackay, S.C. Beck, et al., Multilineage potential of adult human mesenchymal stem cells, Science 284 (1999) 143–147.
[15] T. Saito, K. Okumoto, H. Haga, et al., Potential therapeutic application of intravenous autologous bone marrow infusion in patients with alcoholic liver cirrhosis, Stem Cells Dev. 20 (2011) 1503–1510.
[16] I. Sakaida, Autologous bone marrow cell infusion therapy for liver cirrhosis, J. Gastroenterol. Hepatol. 3 (2008) 1349–1353.
[17] M. Abdel Aziz, H. Atta, S. Mahfouz, Therapeutic potential of bone marrow-derived mesenchymal stem cells on experimental liver fibrosis, 2007, Clin. Biochem. 40 (2007) (2007) 893–899.
[18] M. Hardjo, M. Miyazaki, M. Sakaguchi, et al., Suppression of carbon tetrachloride-induced liver fibrosis by transplantation of a clonal mesenchymal stem cell line derived from rat bone marrow, Cell Transplant. 18 (2009) 89–99.
[19] A.B. Carvalho, L.F. Quintanilha, J.V. Dias, et al., Bone marrow multipotent mesenchymal stromal cells do not reduce fibrosis or improve function in a rat model of severe chronic liver injury, Stem Cells 26 (2008) 1307–1314.
[20] Y. Ding, D. Xu, G. Feng, et al., Mesenchymal stem cells prevent the rejection of fully allogenic islet grafts by the immunosuppressive activity of matrix metalloproteinase-2 and -9, Diabetes 58 (2009) 1797–1806.
[21] J.M. Gimble, B.A. Bunnell, T. Frazier T, et al., Adipose-derived stromal/stem cells: a primer, Organogenesis 9 (2013) 3–10.
[22] H. Komoda, H. Okura, C.-M. Lee, et al., Reduction of N-glycolylneuraminic acid xenoantigen on human adipose tissue-derived stromal cells/mesenchymal stem cells leads to safer and more useful cell sources for various stem cell therapies, Tissue Eng. Part A 16 (2010) 1143–1155.
[23] I. Sakaida, S. Terai, N. Yamamoto, et al., Transplantation of bone marrow cells reduces CCl_4-induced liver fibrosis in mice, Hepatology 40 (2004) 1304–1311.
[24] T. Takami, S. Terai, I. Sakaida, Advanced therapies using autologous bone marrow cells for chronic liver disease, Discov. Med. 14 (2012) 7–12.
[25] L.W. Weber, M. Boll, A. Stampfl, Hepatotoxicity and mechanism of action of haloalkanes: carbon tetrachloride as a toxicological model, Crit Rev Toxicol. 33 (2003) 105–136.
[26] D.A. Chistiakov, P.A. Chistiakov, Strategies to produce hepatocytes and hepatocyte-like cells from pluripotent stem cells, Hepatol. Res. 42 (2012) 111–119.
[27] M. Imamura, T. Aoi, A. Tokumasu, et al., Induction of primordial germ cells from mouse induced pluripotent stem cells derived from adult hepatocytes, Mol. Reprod. Dev. 77 (2010) 802–811.
[28] J. Meyburg, G.F. Hoffmann, Liver, liver cell and stem cell transplantation for the treatment of urea cycle defects, Mol. Genet. Metab. 100 (Suppl. 1) (2010) S77–S83.
[29] Y. Iimuro, D.A. Brenner, Matrix metalloproteinase gene delivery for liver fibrosis, Pharm. Res. 25 (2008) 249–258.
[30] B. Usunier, M. Benderitter, R. Tamarat, A. Chapel, Management of fibrosis: the mesenchymal stromal cells breakthrough, Stem Cells Int. 2014 (2014) 340257.

OPEN ACCESS Freely available online

PLOS | ONE

Computational Promoter Modeling Identifies the Modes of Transcriptional Regulation in Hematopoietic Stem Cells

Sung-Joon Park[1], Terumasa Umemoto[2], Mihoko Saito-Adachi[1], Yoshiko Shiratsuchi[2], Masayuki Yamato[2], Kenta Nakai[1]*

1 Human Genome Center, the Institute of Medical Science, the University of Tokyo, Tokyo, Japan, **2** Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, Tokyo, Japan

Abstract

Extrinsic and intrinsic regulators are responsible for the tight control of hematopoietic stem cells (HSCs), which differentiate into all blood cell lineages. To understand the fundamental basis of HSC biology, we focused on differentially expressed genes (DEGs) in long-term and short-term HSCs, which are closely related in terms of cell development but substantially differ in their stem cell capacity. To analyze the transcriptional regulation of the DEGs identified in the novel transcriptome profiles obtained by our RNA-seq analysis, we developed a computational method to model the linear relationship between gene expression and the features of putative regulatory elements. The transcriptional regulation modes characterized here suggest the importance of transcription factors (TFs) that are expressed at steady state or at low levels. Remarkably, we found that 24 differentially expressed TFs targeting 21 putative TF-binding sites contributed significantly to transcriptional regulation. These TFs tended to be modulated by other nondifferentially expressed TFs, suggesting that HSCs can achieve flexible and rapid responses via the control of nondifferentially expressed TFs through a highly complex regulatory network. Our novel transcriptome profiles and new method are powerful tools for studying the mechanistic basis of cell fate decisions.

Citation: Park S-J, Umemoto T, Saito-Adachi M, Shiratsuchi Y, Yamato M, et al. (2014) Computational Promoter Modeling Identifies the Modes of Transcriptional Regulation in Hematopoietic Stem Cells. PLoS ONE 9(4): e93853. doi:10.1371/journal.pone.0093853

Editor: Connie J. Eaves, B.C. Cancer Agency, Canada

Received December 18, 2013; **Accepted** March 7, 2014; **Published** April 7, 2014

Funding: This work was supported by Research on Applying Health Technology, Health and Labour Sciences by the Ministry of Health, Labour and Welfare, Japan. 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: knakai@ims.u-tokyo.ac.jp

Introduction

Hematopoiesis is a complex and dynamic process, which generates mature blood cells throughout the life of organisms. In the adult bone marrow, long-term hematopoietic stem cells (LT-HSCs) maintain a balanced pool of stem cells, which also differentiates into more mature short-term hematopoietic stem cells (ST-HSCs), multipotent progenitors with a lower self-renewal capacity. It is believed that the blood lineage choice of HSCs is governed by a stepwise cell fate decision [1,2]. However, recent studies have raised questions about the hierarchical hematopoietic system [3,4]. Many studies based on genome-wide gene expression profiling [5–9] have demonstrated that specific extrinsic and intrinsic regulators play key roles in hematopoiesis [10–12]. Recently, high-throughput sequencing techniques have been applied widely [13–15], which have provided new insights into *in vivo* transcription factor (TF) binding and epigenetic modifications [16–18]. Systems biology approaches are also enhancing our understanding of the regulatory dynamics of hematopoiesis [19].

Despite the biological importance of the formation of all blood cells via a transition from LT-HSC to ST-HSC, little is known about the mechanism that underlies this early differentiation. A major explanation for this deficiency is a lack of comprehensive genome-wide identification studies and characterizations of the regulatory elements that govern gene expression in HSCs. The profiling of potential key regulators [8,17,20] and the large-scale integration of datasets [21,22] have improved our understanding greatly. However, these studies are limited to a small number of factors that function in heterogeneous HSCs, which were isolated using different combinations of monoclonal antibodies. Therefore, unconsidered key regulators may exist at this early stage of hematopoiesis. Indeed, novel key factors [23,24] and new multipotent progenitors [3,4,25] have been identified recently.

To address these deficiencies, we developed a computational method on the basis of novel transcriptome data from adult mouse bone marrow HSCs; $CD34^{-}KSL$ ($c\text{-}kit^{+}Sca1^{+}Lin^{-}$) LT-HSCs and $CD34^{+}KSL$ ST-HSCs, a widely used strategy to isolate HSCs at high purity [26,27]. Our method uses a regression-based approach [28–30] to model the linear relationships between gene expression and the characteristics of regulatory elements compiled from a database. In the present study, we extended this regression modeling-based approach using large-scale log-linear modeling (LLM) [31], which considered the combinatorial nature of TFs. Thus, our method can systematically infer the regulation modes exerted by TFs that are probably necessary for gene expression, as well as suggesting synergistic TF modules. Using our transcriptome profiles and this novel method, we characterized transcriptional regulatory modes related to HSCs, which suggested the

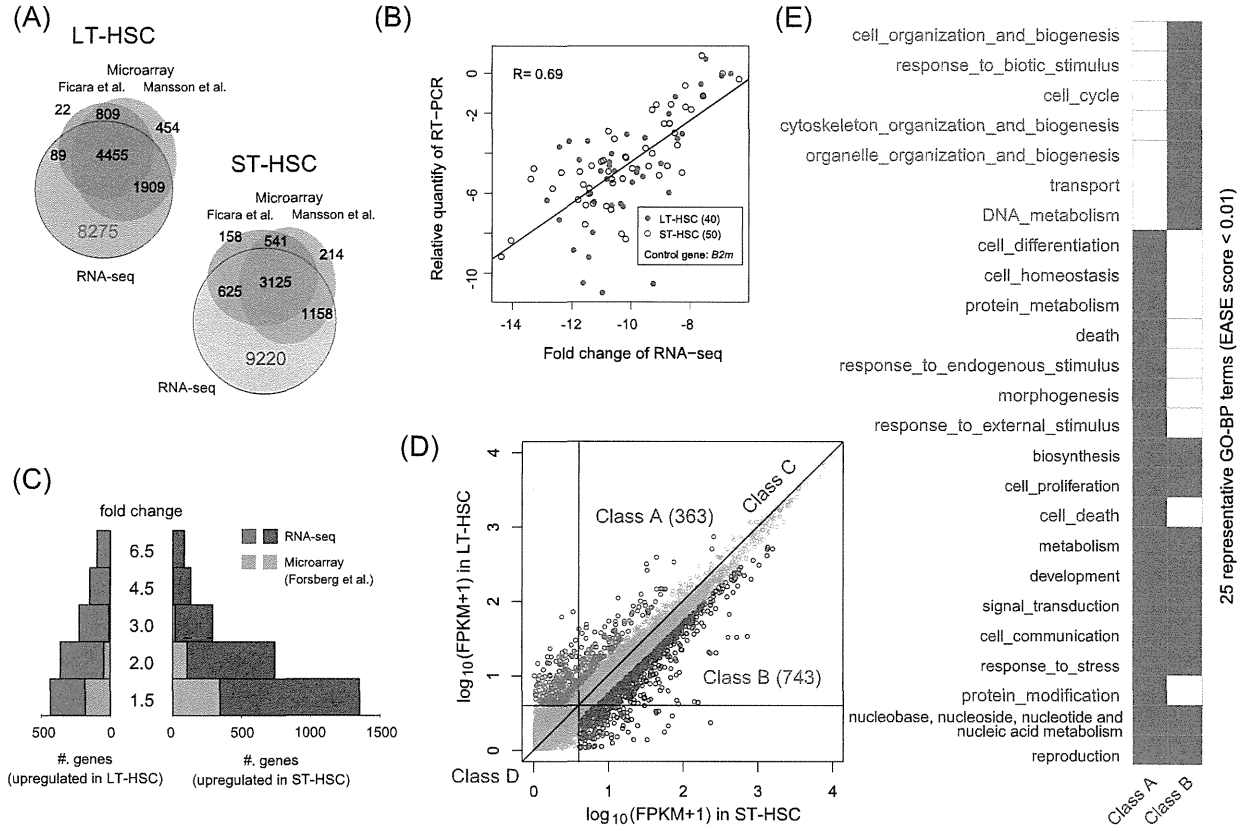


Figure 1. Extensive transcriptome discovery based on the RNA-seq assay. (A) Our RNA-seq assay discovered over 8200 mRNAs that were not detected in microarray-based studies. (B) RNA quantities relative to those of the housekeeping gene beta-2 microglobulin (*B2m*) were correlated in qRT-PCR and RNA-seq assays, but variations were also observed in genes that were expressed at low levels. (C) Analysis of gene expression changes detected a transcriptionally active state in ST-HSCs with a larger number of genes than those considered previously. (D) We categorized genes into 4 classes; Class A and Class B, in which FC >2 and FPKM >3, Class C (6332 genes), in which FC ≤ 2 and FPKM >3, and Class D (6006 genes), in which FPKM ≤ 3. Class A and Class B represented DEGs, Class C represented steady-state transcription genes, and Class D represented genes with noisy expression and/or functional low-expression genes. (E) Enriched GO biological process (GO-BP) terms detected by DAVID (EASE score, <0.01, complete lists in Tables S10 and S11).
doi:10.1371/journal.pone.0093853.g001

functional importance of TFs expressed at steady-state or low levels. Remarkably, we identified 24 differentially expressed TFs that targeted 21 putative TF-binding sites (TFBSs) in LT-HSCs. These TFs might be essential for maintaining the HSC capacity during the early stage of hematopoiesis.

Results

Extensive transcriptome discovery

RNA-seq analysis of HSCs. To establish transcriptional profiles, we extracted total RNA from mouse LT-HSCs ($CD34^-KSL$) and ST-HSCs ($CD34^+KSL$), and performed SOLiD RNA-seq assays in triplicate. We generated 44–70 million 50 bp short reads, among which 44%–63% were mapped uniquely to the mouse genome (mm9) via our recursive mapping strategy [32]. These uniquely mapped reads (uni-reads) were used for further analysis (Table S1). We used the TopHat/Cufflinks pipeline [33] to quantify the RNA abundance of RefSeq genes as fragments per kilobase of exon per million mapped reads (FPKM). This analysis confirmed the high reproducibility among replicates (Figure S1A). We also assessed the overlap between our profile and public expression profiles [8,9]. This comparison showed that our RNA-seq assay uniquely identified 8275 and 9220 genes from LT- and ST-HSCs, respectively (Figure 1A). This indicates that our study successfully identified a more detailed transcriptome landscape than previous studies.

The application of different monoclonal antibodies to purify HSC populations may have diverse effects on the resulting expression profiles [2], which are related to issues regarding the functional purification of HSCs [10,26] and the heterogeneous expression in single cells [4,10,34]. In fact, a comparison between our findings and the results of an RNA-seq analysis of HSCs isolated using distinct markers [15] demonstrated that there were great differences, particularly among genes that were expressed at low levels (Figure S1B). In addition, we performed qRT-PCR using 90 genes that were randomly selected from our samples, and confirmed that RNA quantities relative to the housekeeping gene *B2m* were in overall agreement (Figure 1B). However, genes that were expressed at low levels were substantially different. These results suggest the difficulty in detecting and quantifying rare transcripts in HSCs.

Identification of differentially expressed genes (DEGs). We identified genes with high expression levels (FPKM, >3) and calculated the fold change (FC) in gene

Table 1. Top ten differentially expressed transcription factors.

Class	Gene	FC*	Microarray†	Description
A	Rorc	6.4252		RAR-related orphan receptor gamma
	Hoxb5	5.1317		homeobox B5
	Rarb	3.8601		retinoic acid receptor, beta
	Cdkn1c	3.8479	M,Fo	cyclin-dependent kinase inhibitor 1C (P57)
	Fosb	3.0942	Fi,M	FBJ osteosarcoma oncogene B
	Car1	2.9839	M	carbonic anhydrase 1
	Id1	2.9708		inhibitor of DNA binding 1
	Klf1	2.8796	M	Kruppel-like factor 1 (erythroid)
	Nr4a1	2.7957	Fi,M	nuclear receptor subfamily 4, group A, member 1
	Foxa3	2.7509		forkhead box A3
B	Satb1	3.7749	Fi,M,Fo	special AT-rich sequence binding protein 1
	Hnf4a	3.1733		hepatic nuclear factor 4, alpha
	Trf	2.5921		transferrin
	Hmgb2	2.0842	M	high mobility group box 2
	Runx3	1.9827		runt related transcription factor 3
	Irf8	1.8349		interferon regulatory factor 8
	Arid5a	1.7884		AT rich interactive domain 5A (MRF1-like)
	Uhrf1	1.4536		ubiquitin-like, containing PHD and RING finger domains, 1
	Zfp422	1.4477	Fi,M	zinc finger protein 422
	Notch1	1.3403		notch 1

*$\log_2$ fold change.
†M: Mansson et al. [8], Fo: Forsberg et al. [6], Fi: Ficara et al. [9].
doi:10.1371/journal.pone.0093853.t001

expression. This analysis detected the transcriptionally active state of ST-HSCs (Figure 1C), which supported the results of previous studies [6,7,15]. Our RNA-seq assay detected a higher number of DEGs than those reported previously, which may have been related to our more comprehensive transcriptome discovery method. We categorized the genes into 4 classes using a change of 2-fold as the threshold [15] (Figure 1D): Class A, 363 genes upregulated in LT-HSC; Class B, 743 genes downregulated in LT-HSC; Class C, 6332 genes with FC ≤ 2 and FPKM >3; and Class D, 6006 genes with low expression (FPKM, ≤ 3). Thus, Class A and Class B represented DEGs, Class C represented steady-state transcription genes, and Class D represented genes with noisy expression and/or functional low expression genes.

We searched for any gene ontology (GO) terms enriched in DEGs using the DAVID Bioinformatics Resources [35]. Figure 1E shows the representative GO terms (Tables S10 and S11 for complete lists). This analysis showed that DEGs were involved in the immune response, cell–cell communication, and signal transduction. This was not surprising because extrinsic and intrinsic signals and molecules contribute to the biology of HSCs in the bone marrow microenvironment [1,10,11,36]. In addition to these common biological processes, Class A genes were involved particularly in cell death, cell differentiation, and homeostasis, whereas Class B genes were involved in DNA repair, cell cycle progression, and cell organization. These results were consistent with those of previous studies that showed that apoptosis and cell-cycle regulators play critical roles in maintaining a balanced pool of HSCs and in the expansion of progenitor populations [5,37,38].

Differentially expressed cell-surface molecules and TFs. DEGs included 77 cell-surface molecules with the "cell surface" (GO:0009986) GO term (Table S2), some of which are known to be associated with hematopoiesis: in Class A, *Vwf*, *Lhcgr*, *Cxcl12*, and *Tgfbr3*; in Class B, *CD244*, *CD33*, and *Clec12a*. *CD34*, which was used to isolate HSCs in this study, exhibited an upregulation of over 12-fold in ST-HSCs compared with LT-HSCs. To obtain high HSC purities, these cell-surface molecules will be useful as alternative or additional markers.

DEGs also included 57 TFs that were annotated in TRANSFAC [39], i.e., 31 in Class A and 26 in Class B (Tables 1 and S3). These differentially expressed TFs included known hematopoietic regulators (e.g., *Gata2*, *Tal1*, and *Satb1*) and previously unconsidered TFs, such as the hepatocyte nuclear factor *Foxa3*, the BTB-domain zinc finger *Zbtb20*, the DNA-binding domain *Arid5a*, and the epigenetic regulator *Uhrf1*. It was noteworthy that a large number of TFs belonged to Class C (303 TFs) and Class D (341 TFs) (Tables S4 and S5). In particular, TFs with synergistic functions in HSCs [17] and that belonged to TF families, such as Fox, Lmo, and Sox (which are required by HSCs), were present in Class C and/or Class D. These results may suggest that, in addition to differentially expressed TFs, TFs with coding genes that are expressed at stable or low levels are functionally important molecules.

Computational modeling of DEG promoters

Workflow overview of promoter modeling. To determine the upstream regulatory elements that are essential for DEG transcription, we used a linear regression model that was used widely for this purpose in previous studies [28,30]. The underlying assumption of this model is that the expression levels of genes are controlled by the sum of the independent activities of regulators, such as DNA-binding factors or epigenetic marks. These activities

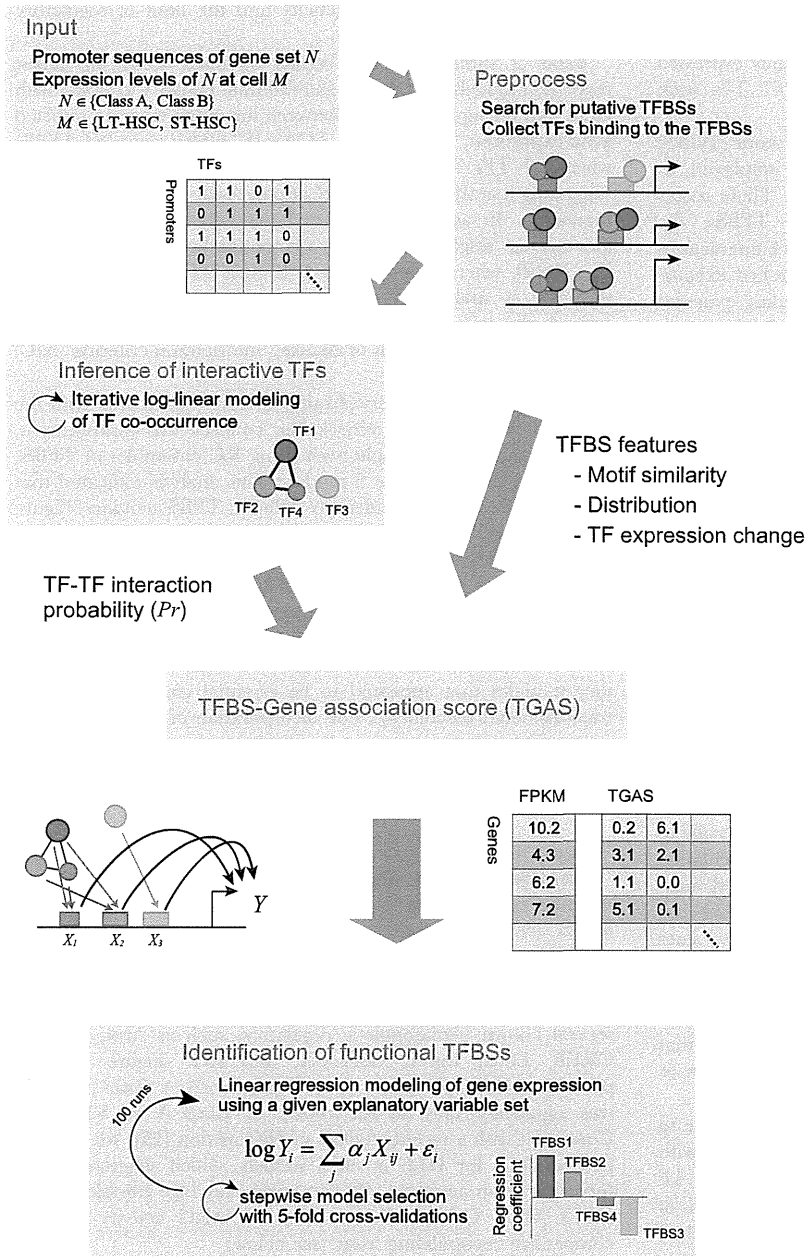


Figure 2. Overview of computational promoter modeling. We searched putative TFBSs and mouse TFs from ±5kb DNA sequences of TSSs, and used these for inferring TF–TF interaction probability and calculating TGASs. We searched the best combination of TFBSs represented by TGASs to predict FPKMs of a gene class in LT- or ST-HSCs by performing 5-fold CVs iteratively.
doi:10.1371/journal.pone.0093853.g002

can be approximated using high-throughput *in vivo* experiments [40,41] or knowledge-based computational approaches [25,30]. As a preliminary test, we applied the linear regression model described in our previous study [29] using ChIP-seq data for 10 major TFs [17]. In this approach, we used genome-wide TF-binding instances that occurred within ±5kb regions from transcription start sites (TSSs), and predicted the FPKMs of DEGs by using a simple linear regression model with rigorous statistical tests. However, we were unable to detect any significant effects, and the correlation between the observed and predicted FPKMs was <0.3. This failure may, in part, reflect the possibility that these TFs exert regulatory functions as distal enhancers, rather than through proximal promoters [17,42].

To identify regulators from proximal promoter regions comprehensively, we used TRANSFAC [39], which is a database that curates >1.5 million ChIP-seq sites, and designed a workflow coupled with intensive computations (Figure 2). First, we prepared the promoter sequences of DEGs and searched for putative TFBSs and mouse TFs that are known to bind to the TFBSs in TRANSFAC using the MATCH tool [43]. This procedure

identified 140 and 141 TFBSs for Class A and Class B promoters, respectively. Among these, 70 TFBSs in Class A and 69 TFBSs in Class B were targeted by at least one TF with a highly expressed coding gene (FPKM, >3). In total, 265 and 267 TFs were involved in Class A and Class B, respectively.

Next, we calculated the TFBS–gene association score (TGAS) using 5 distinct scoring schemes, which were employed as explanatory variables in a linear regression model. These scores considered matrix similarity, positional bias of TFBSs, the expression levels of TFs, and the probability of TF–TF interactions (Materials and Methods). Given a TGAS, we searched exhaustively for the best combination of TFBSs, including pairwise interactions between TFBSs. We performed a 5-fold cross-validation (CV) to avoid the risk of over-fitting. This procedure was repeated 100 times with different random seeds. An ensemble of 100 regression coefficients (RCs) for a TFBS provides statistical information of the estimated regulatory activity of the TFBS. We conducted statistical tests using these ensembles. We applied this workflow to 4 regression models to predict the expression levels of each of the Class A and Class B genes in LT- or ST-HSCs.

We attempted to characterize promoter architectures by testing the different TGASs mentioned above, rather than by comparing our approach with other modeling methods. This was because of the difficulty of implementing existing methods using our inputs and analyzing their results. We also aimed to determine regulatory activities by analyzing 4 models. We characterized the context-dependent function of regulators that activated and repressed the transcription of distinct genes depending on the cellular context [44,45]. Thus, our approach provided a detailed picture of the regulatory modes involved in context-dependent gene expression.

Inference of higher-order TF interactions. The co-occupancy of a promoter by multiple TFs contributes synergistically to transcriptional regulation. We considered this when calculating TGAS by performing probabilistic LLM [31] coupled with iterative random sampling. The input matrix used for LLM, i.e., n promoters in rows $\times m$ TFs in columns, comprised binary values that represent the existence of TFBSs for m TFs in n promoters. Using this matrix, LLM was employed to infer the conditional (in)dependency of TF occurrences, i.e., TF–TF interactions in higher-order conditional distributions. It should be noted that LLM cannot determine whether an interaction is competitive or cooperative.

The huge number of TFs means that LLM is not adequate to compute them all; therefore, we performed random sampling with 10 arbitrary selected TFs, which means that an inferred TF–TF interaction was observed constantly in the 2^8 state combinations of 8 TFs. This sampling procedure was terminated if an outcome had no effect during 10^5 runs. We calculated the interaction probability Pr for all possible TF pairs using this iterative sampling procedure (Materials and Methods). After repeating the sampling procedure 1,367,639 times for Class A and 1,406,837 times for Class B, we retrieved 50 and 77 interactions ($Pr=1.0$) from Class A and Class B, respectively (Tables S12 and S13).

Performance of regression models. Overall, Pearson's correlation coefficient Rs in learning and testing of 5-fold CVs showed a slight over-fitting in the range of <0.2 (Figure S5), which was acceptable in our sense. One of the reasons for this over-fitting was the unbalanced numbers between testing genes and TFBSs; e.g., 72 Class A genes (a subset of 5-fold CV) were tested by a model with over 100 predictors that were trained by the remaining Class A genes. This implies that the constructing of a model to generalize the gene regulation for an HSC population is a highly difficult challenge that is associated with the degrees of functional purity and heterogeneity and the limit of regulatory features used in the modeling.

Table 2 summarizes the results obtained from the linear regression models. The results showed that TGAS V coupled with LLM had the highest mean R between the observed and predicted gene expression. Interestingly, TGAS IV, which removed TFBSs where all TFs had FPKM ≤ 3, yielded poor-quality models, suggesting that these TFBSs were also necessary for modeling gene regulation. In addition, compared with the main effect terms (denoted as "single" in Table 2), a large number of pairwise terms, i.e., $A \times B$, where A and B are 2 distinct TFBSs that were not included as main effect terms, contributed to the modeling. Indeed, the initial models that comprised only the main effect terms selected on the basis of Akaike's information criterion (AIC) showed an $R<0.6$.

The improvement observed using TGAS V compared with the use of TGAS III was not remarkable. To assess this improvement, we performed a two-sample t-test using RC ensembles of TFBSs that were common in the 2 models. This analysis indicated that these models yielded considerably different TFBS activities (Figure S2A). In most cases, TF interaction scores (Equation 9) were ineffectively small. However, specific TFBSs, such as AP-1, Ets, and Ebox, had high scores (Figure S2B) because of the relatively larger number of TFs that interacted to occupy these TFBSs ($Pr>0$). This apparently affected the different estimations.

Overall, pairwise interactions between TFBSs reflected regulatory modules that appeared to be essential components of the transcriptional machinery. The incorporation of cooperative and competitive interactions among TFs into quantitative models is also essential for determining the transcriptional network based on a fine-tuned explanation of gene expression.

Propensity of inferred TFBS activities

Identification of significant TFBSs and changes of regulatory activities. To assess the statistical significance of TFBS activities, we performed single-sample t-tests using RC ensembles on the basis of TGAS V (Figure S3). This analysis identified 142 TFBSs that rejected the null hypothesis that the mean value of RCs was equal to zero ($p<0.05$). This included several known hematopoietic regulators, such as Arnt, C/EBP, CREB, Ebox, Egr-1, GATA-X, and IRF (Figure 3A). In particular, GATA-X targeted by *Gata1*, *Gata2*, or *Gata3* (Class A) was significant only in the model of Class A in ST-HSCs. Consistent with a recent analysis of *Hlf* function [25], we inferred positive RCs for HLF in all 4 models, which suggests that it functions as an activator. *Hlf* was upregulated by 1.9-fold in LT-HSCs (Class C). We also validated PPARG activity using a competitive repopulating assay (see below).

Overall, 83 of the 142 TFBSs were detected by all 4 models, among which 14 TFBSs were unique in Class A or Class B (Table S6). Furthermore, 73% (61/83) of the common TFBSs appeared to have same effects on the activities in LT- and ST-HSCs, e.g., a positive RC in LT-HSC was also positive in ST-HSC. Interestingly, this effect was the opposite in Class A and Class B, for which typical examples are marked by rectangles in Figure 3A. There were also exceptional cases, including IRF-2, HOXA7, and DMRT3. The results obtained using TGAS III had similar properties.

Gain and loss of activities during HSC progression. To analyze the change of TFBS activities between LT- and ST-HSCs, we tested 2 RC ensembles of a TFBS using a two-sample t-test under the null hypothesis that the mean values were equal. This analysis found that the null hypothesis was rejected for 71 TFBSs (Class A) and 58 TFBSs (Class B) ($p<0.001$) (Figure 3B). The

Table 2. Result obtained using the linear regression models.

			Linear regression		TFBS contents	
Class	**Cell**	**TGAS***	**TFBSs**	**R†**	**Single**	**Pairwise**
A	LT-HSC	I	83.91 (5.7238)§	0.8016 (0.0205)	18.98 (2.1070)	64.93 (5.7745)
		II	98.69 (5.5492)	0.8482 (0.0197)	30.67 (2.6685)	68.02 (5.8378)
		III	103.73 (4.5296)	0.8722 (0.0134)	29.14 (2.5220)	74.59 (5.0062)
		IV	47.28 (2.9260)	0.6771 (0.0165)	11.15 (1.2835)	36.13 (2.9888)
		V	108.12 (4.9138)	0.8850 (0.0124)	31.82 (2.6921)	76.30 (5.1449)
		V-1	84.42 (4.0748)	0.8334 (0.0154)	18.59 (6.3594)	65.83 (7.0129)
		V-2	51.90 (2.9648)	0.7164 (0.0155)	11.96 (1.1128)	39.94 (3.0588)
		V-3	91.38 (4.3053)	0.8284 (0.0146)	28.01 (2.5120)	63.37 (4.8634)
	ST-HSC	I	83.02 (5.3907)	0.8087 (0.0204)	20.66 (1.9709)	62.36 (5.9389)
		II	101.65 (4.7188)	0.8463 (0.0180)	37.47 (2.8088)	64.18 (5.6416)
		III	106.77 (4.0394)	0.8730 (0.0114)	36.29 (2.7579)	70.48 (4.5902)
		IV	50.34 (3.1376)	0.6786 (0.0215)	17.63 (2.0768)	32.71 (3.6064)
		V	108.49 (4.5618)	0.8777 (0.0132)	37.62 (2.7378)	70.87 (5.2548)
		V-1	85.01 (4.2883)	0.8289 (0.0160)	22.75 (2.1372)	62.26 (4.3327)
		V-2	53.32 (3.1012)	0.6867 (0.0191)	21.53 (2.2470)	31.79 (3.5222)
		V-3	86.71 (4.8853)	0.8126 (0.0196)	26.03 (2.6399)	60.68 (5.4934)
B	LT-HSC	I	77.82 (5.6451)	0.6177 (0.0183)	21.98 (2.1400)	55.84 (6.2749)
		II	100.86 (4.3244)	0.7016 (0.0147)	30.33 (2.8002)	70.53 (4.9748)
		III	105.78 (3.8251)	0.7311 (0.0125)	27.96 (2.4614)	77.82 (4.1434)
		IV	49.50 (2.8231)	0.5410 (0.0143)	15.29 (1.7164)	34.21 (3.1058)
		V	108.45 (4.2270)	0.7466 (0.0111)	27.20 (2.4819)	81.25 (4.3183)
		V-1	87.86 (3.7895)	0.6736 (0.0159)	28.59 (3.1051)	59.27 (4.9272)
		V-2	53.74 (2.7879)	0.5548 (0.0145)	15.54 (1.5324)	38.20 (3.1969)
		V-3	84.45 (3.8350)	0.6662 (0.0149)	24.95 (2.5744)	59.50 (4.6573)
	ST-HSC	I	77.65 (4.7924)	0.6077 (0.0175)	21.42 (2.0745)	56.23 (5.3514)
		II	100.69 (5.3846)	0.6980 (0.0169)	25.74 (2.4602)	74.95 (6.0056)
		III	105.87 (4.1633)	0.7262 (0.0140)	24.73 (2.4448)	81.14 (4.8477)
		IV	50.07 (2.8679)	0.5160 (0.0161)	14.36 (1.7235)	35.71 (3.3920)
		V	107.32 (4.4763)	0.7325 (0.0135)	24.77 (2.6338)	82.55 (4.6720)
		V-1	86.98 (3.8781)	0.6716 (0.0166)	22.62 (2.4891)	64.36 (4.6206)
		V-2	54.09 (2.9397)	0.5354 (0.0164)	15.65 (1.8993)	38.44 (3.5080)
		V-3	84.96 (4.1639)	0.6544 (0.0165)	21.47 (2.4185)	63.49 (4.9830)

E.g., at the top line, the final regression model predicted Class A FPKMs in LT-HSCs using TGAS I, resulting in the correlation coefficient R = 0.8016. This model included 83.91 TFBSs consisting of 18.98 single TFBSs and 64.93 pairwise TFBSs.

*TFBS-Gene association scores; (I) MATCH score only, (II) including distribution of TFBSs, (III) including expression changes in TFs, (IV) same as (III) but only including TFBSs targeted by highly expressed TFs, and (V) including the TF–TF interactions in the log-linear model. (V) was modified to remove TFs: coded by undetectable transcripts (V-1), those that belonged to Class D (V-2), or by removing the 21 TFBSs in Figure 4B (V-3).

†Pearson's correlation coefficient; once the final regression model was found, *R* reflecting the model quality is calculated to measure the degree of correlation between the observed and predicted FPKMs.

§Data are presented as the means (and standard deviation in parentheses).

doi:10.1371/journal.pone.0093853.t002

multiple-testing correction reduced these numbers to 49 and 42 in Class A and Class B, respectively ($p < 0.001$) (Tables S8 and S9). Interestingly, although these TFBSs had different mean values, the effects of the activities were mostly unchanged; a positive (negative) activity in LT-HSC was still positive (negative) in ST-HSC, i.e., 75% (53/71) in Class A and 84% (49/58) in Class B. In most cases, the strengths of these activities increased markedly in ST-HSC, i.e., 85% (45/53) in Class A and 76% (37/49) in Class B. These results suggest that the maintenance of self-renewal and the differentiation competence in ST-HSCs require a vigorous transcriptional program.

As an intuitive insight into the gain and loss of activities during HSC progression, we found that downregulation of Class A in ST-HSC relative to LT-HSC was accompanied by a gain of negative RCs in ST-HSC (e.g., CKROX, GABP, C/EBPdelta, and myogenin/NF-1) and by a loss of positive RCs in LT-HSC (e.g., IRF-2, HEN1, POU6F1, and RBP-jkappa). Similarly, upregulation of Class B in ST-HSC relative to LT-HSC was

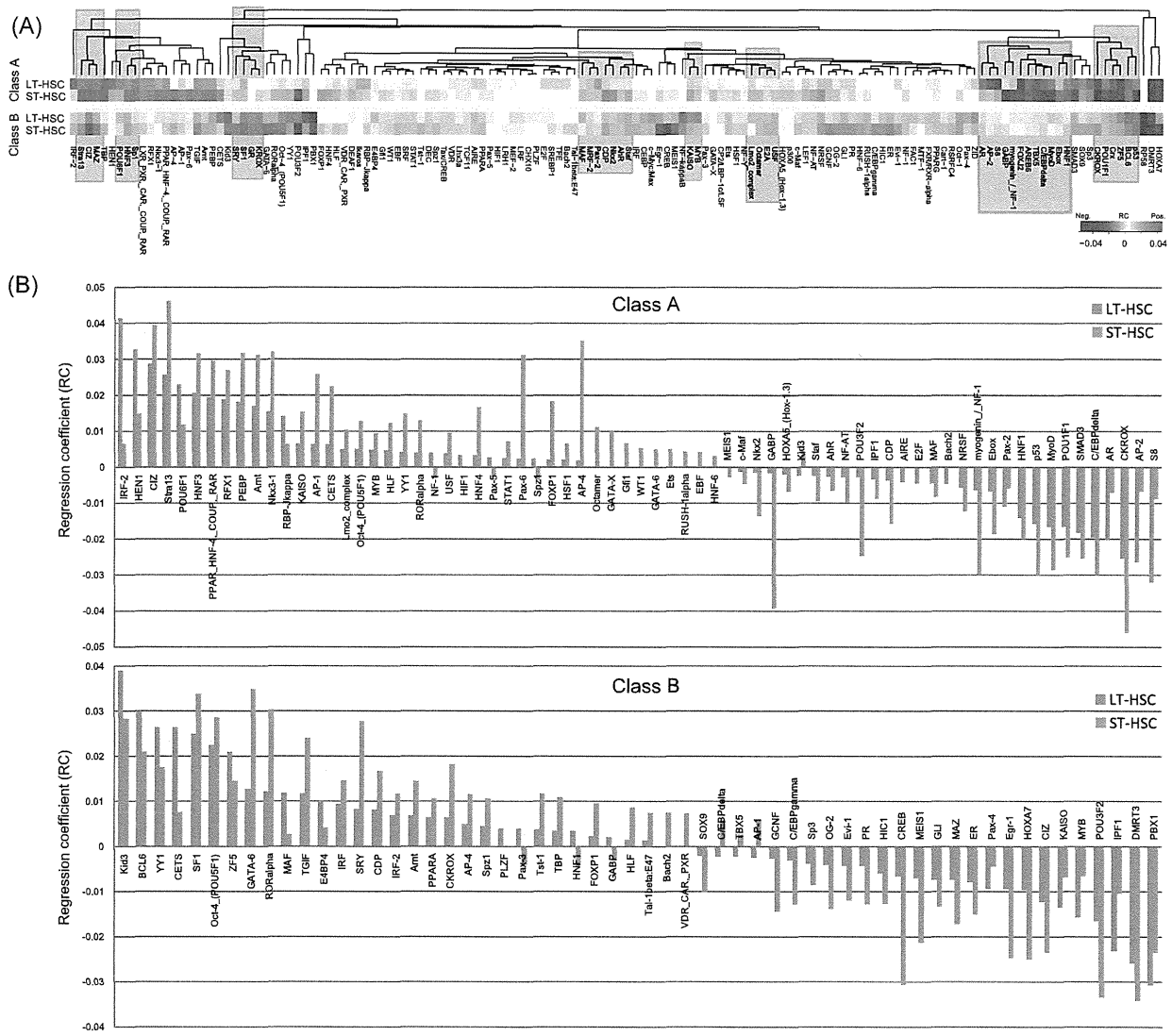


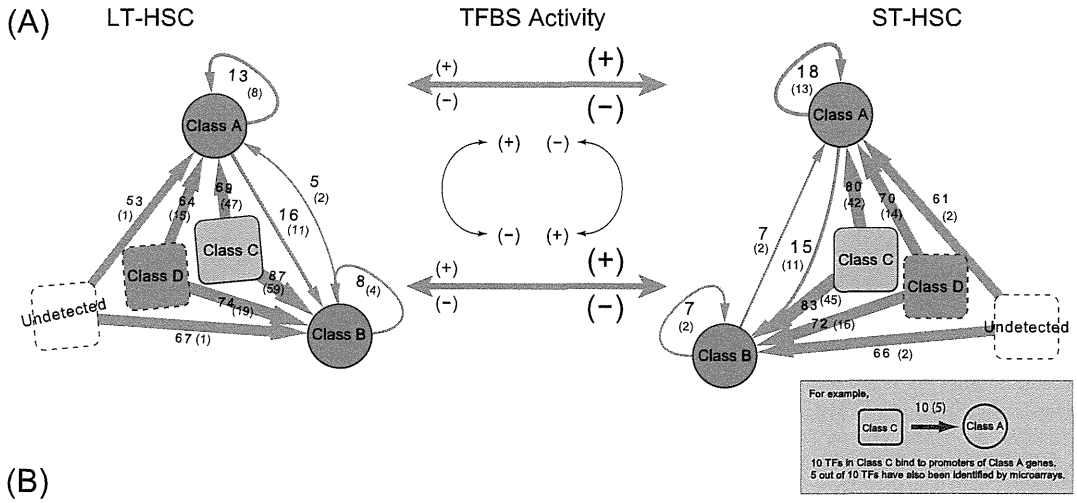
Figure 3. Propensity of significant TFBS activities. (A) Heat map showing regression coefficients (RCs) of 142 potentially important TFBSs ($p<0.05$). Rectangles indicate typical cases of opposing RCs in Class A and Class B. (B) We found that 71 and 58 TFBSs from Class A and Class B promoters, respectively, had significantly different RCs in LT- and ST-HSCs ($p<0.001$).
doi:10.1371/journal.pone.0093853.g003

followed by a gain of positive RCs in ST-HSC (e.g., SF1, Oct-4, GATA-6, and RORalpha) and by a loss of negative RCs in LT-HSC (e.g., PBX1, IPF1, MYB, KAISO, and Pax-4). However, many of the TFBSs in each class exhibited activity changes that differed greatly from our intuitive expectations, which suggests that the high level of complexity in the transcriptional circuit is related to context-dependent gene expression.

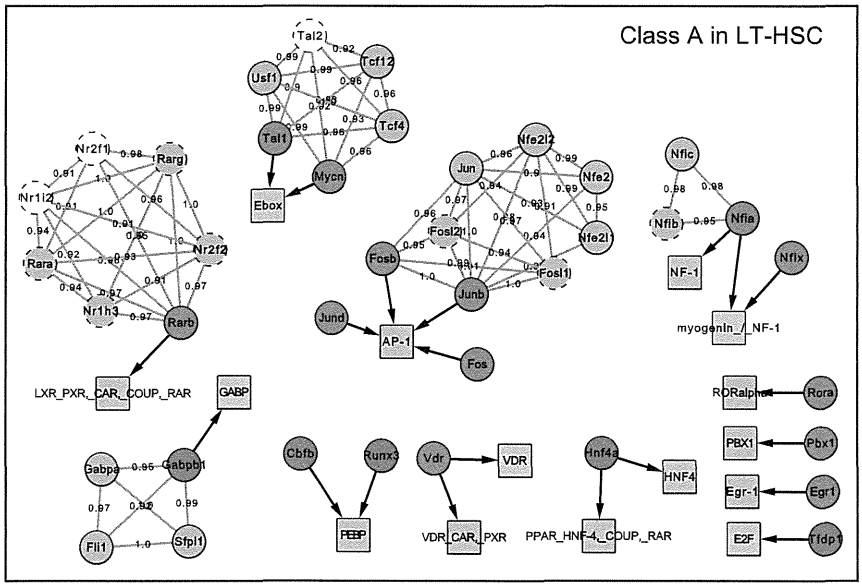
Functional importance of TFs coded by rare transcripts

Regulatory effects of TFs from gene classes that were expressed at low and undetectable levels. We constructed TF–gene networks on the basis of the links between the 142 TFBSs and their downstream target genes. The networks had vast numbers of edges: 40,896 edges among 204 TFs that targeted 114 TFBSs of Class A in LT-HSC; 45,882 edges among 237 TFs that targeted 114 TFBSs of Class A in ST-HSC; 97,946 edges among 253 TFs that targeted 134 TFBSs of Class B in LT-HSC; and 96,975 edges among 243 TFs that targeted 125 TFBSs of Class B in ST-HSC.

The majority of TFs involved in these networks belonged to Class D and transcripts that were not detected in our RNA-seq assay (Figure 4A). Only a small portion of these genes were detected by microarray analyses [8,9], i.e., the numbers in parentheses in Figure 4A. Our qRT-PCR assay detected only 1 or 2 of these genes, suggesting that they originated from rare transcripts, i.e., TF-coding genes expressed at low or undetectable levels in HSCs. To assess the importance of these TFs, we modified TGAS V to remove the regulatory effects from the TFs; by setting $F=0$ (Equation 8) for unexpressed TF-coding genes (TGAS V-1) and for TF-coding genes in Class D (TGAS V-2). As a result, Rs were lower than TGAS V when the model removed these effects (Table 2), which suggests their important contribution


(A)
LT-HSC
TFBS Activity
ST-HSC
Class A
Class B
Class C
Class D
Undetected
13 (8)
5 (2)
53 (1)
64 (15)
69 (47)
16 (11)
87 (59)
8 (4)
74 (19)
67 (1)
18 (13)
80 (42)
70 (14)
61 (2)
7 (2)
15 (11)
83 (45)
72 (16)
7 (2)
66 (2)
For example,
10 (5)
10 TFs in Class C bind to promoters of Class A genes.
5 out of 10 TFs have also been identified by microarrays.

(B)


Class A in LT-HSC
Tal2
Tcf12
Usf1
Tcf4
Tal1
Mycn
Ebox
Nr2f1
Rarg
Nr1i2
Nr2f2
Rara
Nr1h3
Rarb
LXR_PXR_CAR_COUP_RAR
GABP
Gabpa
Gabpbp1
Fli1
Sfpi1
Nfe2l2
Jun
Nfe2
Fosl2
Fosl1
Fosb
Junb
Jund
AP-1
Fos
Nfic
Nfib
Nfia
NF-1
Nfix
myogenin_/_NF-1
Cbfb
Runx3
PEBP
Vdr
VDR
VDR_CAR_PXR
Hnf4a
HNF4
PPAR_HNF4_COUP_RAR
RORalpha
Rora
PBX1
Pbx1
Egr-1
Egr1
E2F
Tfdp1
TFBS
positive
negative
binding
TF
Class A
Class B
Class C
Class D
interaction
undetected

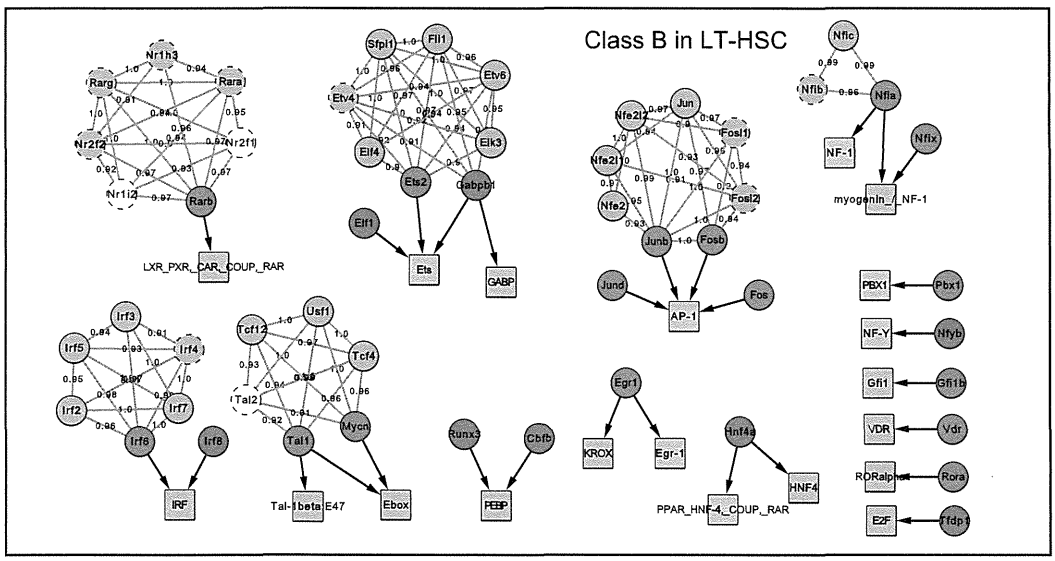

Class B in LT-HSC
Nr1h3
Rarg
Rara
Nr2f2
Nr2f1
Nr1i2
Rarb
LXR_PXR_CAR_COUP_RAR
Sfpi1
Fli1
Etv6
Etv4
Elf4
Elk3
Ets2
Gabpbp1
Elf1
Ets
GABP
Nfe2l2
Jun
Fosl1
Fosl2
Nfe2
Junb
Fosb
Jund
Fos
AP-1
Nfic
Nfib
Nfia
NF-1
Nfix
myogenin_/_NF-1
Irf3
Irf5
Irf4
Irf2
Irf7
Irf6
Irf8
IRF
Usf1
Tcf12
Tcf4
Tal2
Tal1
Mycn
Tal1_beta_E47
Ebox
Runx3
Cbfb
PEBP
Egr1
KROX
Egr-1
Hnf4a
HNF4
PPAR_HNF4_COUP_RAR
PBX1
Pbx1
NF-Y
Nfyb
Gfi1
Gfi1b
VDR
Vdr
RORalpha
Rora
E2F
Tfdp1

Figure 4. Inference of transcriptional regulatory networks. (A) Systematic representation of TF–gene networks and the change of TFBS activities between LT-HSCs and ST-HSCs. Genes that produce TFs that putatively bind to important TFBSs (Figure 3A) existed in each class. Some of them were not detected in the RNA-seq assay, and were categorized as "Undetected". The numbers on the gray-colored arrows denote the number of TFs in the corresponding class that bind to Class A or B gene promoters, suggesting that the majority of TFs belonged to nondifferentially expressed gene classes. The numbers in parentheses indicate TFs that were detected in microarray-based studies, suggesting the extensive discovery of our assay. As shown in the middle panel, we inferred that the positive or negative activities of TFBSs are mostly unchanged between cells, but are inverted between Class A and Class B. (B) Subnetworks of (A) in LT-HSCs. The majority of TF-coding genes were not differentially expressed, whereas 24 TFs binding to 21 TFBSs were present among DEGs (Class A and Class B) and interacted strongly with nondifferentially expressed TFs (Figure S4 shows the subnetworks in ST-HSCs).
doi:10.1371/journal.pone.0093853.g004

to the modeling. Indeed, many known factors [17,46] were present in these categories.

Competitive repopulation assay with activated *Pparg*. The suggestion that TF-coding genes expressed at low levels are important contributors to transcriptional regulation prompted us to investigate the function of *Pparg*, which remains controversial in HSC biology [47]. *Pparg* was categorized into Class D (0.3747 FPKM in LT-HSC and 0.2616 FPKM in ST-HSC), and its binding site PPARG had negative RCs in all 4 models (Figure 3A). To confirm this PPARG activity, we treated LT-HSCs with GW1929, a high agonist of *Pparg* [48,49].

As shown in Figure 5A, we performed a transplantation assay using LT-HSCs that were cultured for 5 days with or without GW1929. GW1929-treated HSCs exhibited decreased chimerism at 20 weeks after the transplantation compared with the controls (Figure 5B). The contribution of T-cell, B-cell, and myeloid lineages to the total donor-derived cells was not highly different (Figure 5C). These results suggest the possibility that the excessive activity of PPARG influences negatively the long-term repopulating activity of HSCs, which supports the capacity of our approach to infer the activities of regulatory elements in HSCs.

Identification of potential key regulators

Differentially expressed TFs and their target sites. The regulatory networks (Figure 4A) involved differentially expressed TFs, i.e., 18 TFs regulated Class A in LT-HSC (13 from Class A and 5 from Class B) and 24 TFs regulated Class B in LT-HSC (16 from Class A and 8 from Class B). These TFs targeted 21 TFBSs that are well-studied hematopoietic regulators, including the Fos/Jun complex [50], Ebox-binding bHLH TFs [51], the GABP complex [52], and retinoic acid receptors [53]. In particular, AP-1 and Egr-1 appeared in all of the models and were targeted by immediate early response genes that are important for apoptosis and differentiation [50] and that are downregulated in ST-HSCs [54]. Interestingly, our model showed that some of these TFs are highly modulated by other TFs that were not differentially expressed (Figures 4B and S4). This may explain the observation that the models with TGAS V-1 and TGAS V-2 reduced the predictive performance.

Putative function of the differentially expressed TFs. Many recent studies have reported that epigenetic effects are important factors in hematopoiesis [16,18,55]. What would happen if the 21 TFBSs targeted by differentially expressed TFs were turned off by DNA methylation, for example? This question was suggested by the recent finding that CpG-methylated regions colocalize with TFBSs in HSCs [56]. To answer this question, we removed each set of TFBSs that appeared in Figures 4B and S4, and performed regression modeling in this condition. The results showed slightly lower Rs (TGAS V-3 in Table 2), however, the

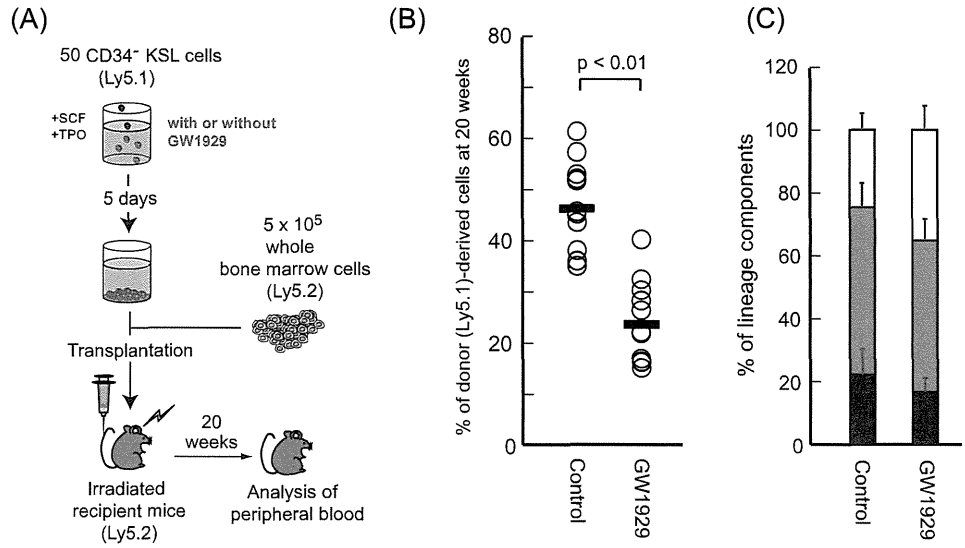


Figure 5. Long-term competitive reconstitution assay. (A) Scheme of the competitive repopulation assay using GW1929, a high agonist of *Pparg*. (B) Analysis of the proportion of donor-derived $CD34^-$KSL HSCs obtained from untreated (Control) and treated (GW1929) recipient mice at 20 weeks posttransplant. Each dot represents a single mouse. (C) Relative contributions of $CD4^+$ or $CD8^+$ (T-cell lineage), $B220/CD45R^+$ (B-cell lineage) and Mac-1^+ or Gr-1^+ (myeloid lineage) cells in donor-derived Ly-5.1^+ cells of recipient mice at 20 weeks posttransplant. Black, T-cell lineage; gray, B-cell lineage; white, myeloid lineage. Data are presented as the mean $\pm$SD.
doi:10.1371/journal.pone.0093853.g005