OUKA
The University of Osaka
Institutional Knowledge Archive

Title	CXCL4/PF4 is a predictive biomarker of cardiac differentiation potential of human induced pluripotent stem cells
Author(s)	大橋, 文哉
Citation	大阪大学, 2019, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/73508
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

博　士　論　文

題目：CXCL4/PF4 is a predictive biomarker of cardiac differentiation potential of human induced pluripotent stem cells

（CXCL4/PF4 がヒト iPS 細胞由来心筋細胞の分化能を予測できるバイオマーカーである）

大阪大学大学院薬学研究科

創成薬学専攻　遺伝子細胞医薬学

大橋　文哉

目次

1. Abstract....3

2. Introduction....5

3. Results....7

3.1. Outline of the workflow for selecting predictive biomarkers for cardiac differentiation....7

3.2. Differences in cardiac differentiation abilities of hiPSC lines.....11

3.3. Expression of germ layer-related genes in EBs derived from hiPSC lines.....20

3.4. Comprehensive gene expression analysis of undifferentiated hiPSCs using miRNA and mRNA arrays....25

3.5. Profiling of TSS expression in undifferentiated hiPSCs using CAGE....29

3.6. Validation of 22 candidate genes as predictors of cardiac differentiation potential....35

4. Discussion....43

5. Conclusion....48

6. Materials and Methods....48

7. References....59

8. Acknowledgements....69

Abstract

【背景】

iPS 細胞は様々な細胞へ分化しうる多能性を有しており [1]、再生医療や薬剤スクリーニングへの応用が進みつつある [2-8]。近年では、創薬研究用に疾患を持つ患者の体細胞から樹立した疾患特異的 iPS 細胞株や再生医療用にいくつかの HLA 型の iPS 細胞株のストックも行われ、様々な種類の iPS 細胞株が作成されている [9,10]。 一方、ES 細胞や iPS 細胞は株の種類により、各組織への分化指向性が異なることが分かってきた [11-13]。その原因としてはエピジェネティックな変化の影響を受ける可能性が報告されている [14-16] 。iPS 細胞は由来となる体細胞の遺伝子的背景が大きく異なるので [17]、ES 細胞と比べて分化誘導効率がばらつく可能性が考えられる。また、目的とする組織に応じて最適な iPS 細胞を選ぶための指標が求められるが、未分化マーカーではその分化指向性を判別することができない。そこで、我々は心筋細胞に分化しやすい iPS 細胞株をスクリーニングするためのバイオマーカーを探索することを目的とした。これまでバイオマーカーは単一の遺伝子解析プラットフォームを用いて探索されてきたが [25-27]、数多くの候補遺伝子から偽陽性を最小限にして、最適なバイオマーカーを選ぶことは困難な作業である。そこで、本研究では CAGE 解析、mRNA アレイ解析、microRNA アレイ解析を組み合わせることで、バイオマーカーを特定可能であることを仮説とした。

【方法および結果】

(1)6 種類の iPS 細胞株で同一の分化誘導法を用いて、心筋細胞へ分化誘導を行った。分化誘導後の心筋細胞をフローサイトメトリーや定量 PCR 等で解析して、心筋細胞への分化指向性が高い iPS 細胞株(High)と低い iPS 細胞株(Low)に群分けをした。その結果、High 群と Low 群の 2 群間のトロポニン陽性率に有意な差があった(High v.s. Low: 81.1%±3.3% vs 13.6±1.9%; $p<0.01$, $n=6$)。また、High 群では分化誘導後の未分化細胞の残存率が Low 群と比べて有意に低い結果であった。

(2)未分化細胞の遺伝子発現を網羅的に解析するために mRNA アレイ解析、microRNA アレイ解析、

CAGE 解析を用いた。分化指向性の High 群と Low 群の 2 群間で未分化細胞の遺伝子発現量を比較し、有意な変化があった遺伝子群を抽出した。その結果、High 群では WNT シグナルやミトコンドリア機能に関連する遺伝子群が Low 群と比べて有意に高かった。また、遺伝子解析手法を組み合わせることで分化指向性を予測できる可能性があるバイオマーカー候補遺伝子を 22 個に絞り込むことができた。

(3)テストセットとして 13 種類の iPS 細胞を用いて心筋細胞の分化能をバイオマーカー候補遺伝子で判別可能かをバリデーションした。High 株と Low 株の 2 つの群に分けて、バイオマーカー候補遺伝子の発現量を比較した結果、22 個の遺伝子のうち 20 個の遺伝子で 2 群間の有意な差は見られなかったが、2 つの候補遺伝子はテストセットでも有意な差が見られた。正に相関した遺伝子として *CXCL4/PF4* があり、負に相関した遺伝子として *TMEM64* があった。

(4)心筋細胞への分化誘導に関連するシグナル伝達の調節実験を行い、同一 iPS 細胞株で心筋細胞への分化誘導の変化とバイオマーカーが相関するかを検証した。WNT シグナル伝達の阻害剤、刺激剤、ミトコンドリア機能の阻害剤、対照群として DMSO を分化誘導前の iPS 細胞にそれぞれ添加して、心筋細胞へ分化した。その結果、WNT シグナルの刺激やミトコンドリア機能の阻害により、心筋細胞への分化誘導効率が対照群と比べて低下した。その心筋分化誘導効率の変化と相関して *CXCL4/PF4* 発現量が低下していたが、*TMEM64* は相関性が見られなかった。

(5)心筋細胞への分化誘導における PF4 の機能を調べるために、分化誘導前の iPS 細胞に PF4 を 2 日間前処理してから心筋細胞へ分化誘導を行った。その結果、PF4 処理群では対照群と比べて、分化誘導後の心筋細胞関連遺伝子(*MYH7*, *MYL2* および *TNT2*)の発現が有意に高くなり、強く拍動する胚葉体数が増加した。

【結論】

CXCL4/PF4 の発現量が高い iPS 細胞株が心筋細胞への高い分化指向性を有する可能性あり、心筋細胞を製造する上で最適な iPS 細胞株を選ぶための新たなバイオマーカーになりうることが示唆された。

Introduction

Human induced pluripotent stem cells (hiPSCs) are capable of differentiating into various tissues[1], thereby acting as a source of cells for regenerative medicine and drug discovery[2-8]. Technological advancements in the development of disease-specific hiPSCs from somatic cells of patients have enabled the study of the pathology of rare diseases[9,10]. Several studies have suggested that the direction of differentiation of tissues derived from the endoderm, mesoderm, and ectoderm varies depending on the line of human embryonic stem cells (hESCs) and hiPSCs[11-13]. Variation in the direction of differentiation among hiPSC lines is the result of differences in somatic tissue of origin and epigenetic changes[14-16]. As the genetic backgrounds of the somatic cells used to derive hiPSCs differ significantly, the epigenetic variation between hiPSCs and hESCs is large[17]. Biomarkers are required for selecting suitable hiPSC lines with high differentiation potential for specific tissues. Several studies have previously investigated biomarkers associated with differentiation potential of hiPSCs[18-24]. However, current pluripotency markers such as *OCT-4*, *LIN28*, and *NANOG* cannot be used to distinguish the direction of differentiation.

The purpose of the present study was to identify a biomarker for predicting efficient cardiac differentiation that can be used for selecting individual hiPSC lines by comparing the gene expression profiles of undifferentiated hiPSC lines with varying cardiac differentiation potential.

Biomarkers have been searched using single genome-wide analyses[25-27]. However, selection of appropriate genes from among the many candidate genes while minimizing the occurrence of false positives using this approach is challenging.

In this study, we hypothesized that biomarkers can be selected using three different platforms of genetic analyses. We comprehensively analysed the gene expression of hiPSCs using cap analysis of gene expression (CAGE), mRNA array, and microRNA (miRNA) array to screen for biomarkers of cardiac differentiation potential. CAGE has been used to analyse transcription start sites and can measure the activity of alternative promoters via absolute quantitation. In contrast, microarray analysis has been used to quantify transcript expression in samples based on the intensity ratio of the hybridisation signal. Our proposed method of using three gene analysis platforms for identifying novel predictive biomarkers of hiPSCs with high cardiac differentiation potential will reveal useful genes that can be important for selecting desired hiPSC lines.

We showed the schematic illustration of the experimental design to identify biomarkers for cardiac differentiation (Fig. 1).

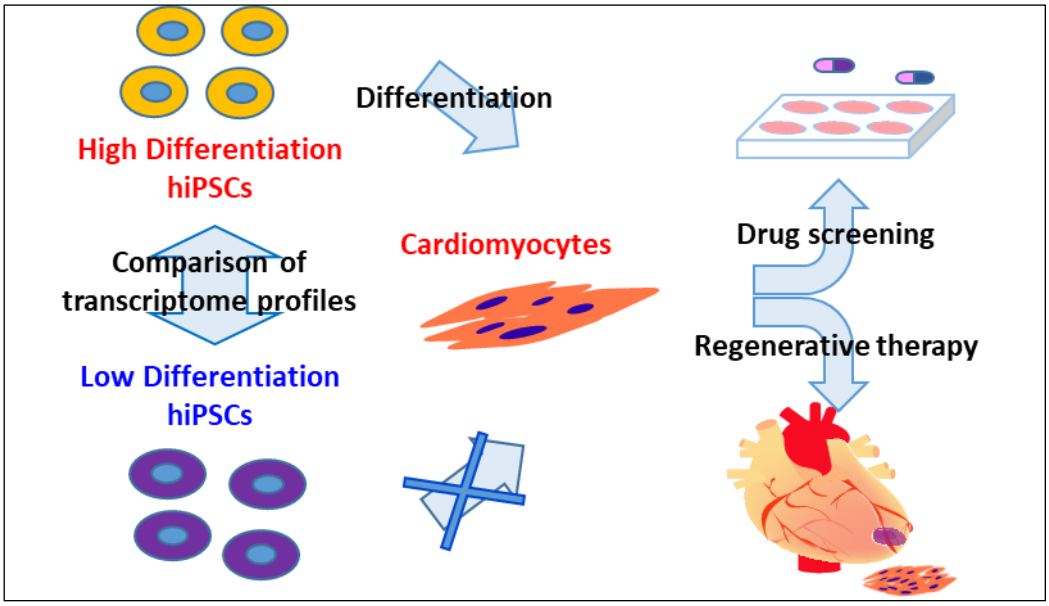

Figure 1 Strategy for identification of biomarkers for cardiac differentiation.

Results

Outline of the workflow for selecting predictive biomarkers for cardiac differentiation.

To compare the *in vitro* cardiac differentiation efficiency of hiPSC lines, six hiPSC lines were cultured and differentiated into cardiomyocytes under identical conditions as a training set (Table 1). Two types of human somatic tissues were used to establish hiPSCs, namely, dermal fibroblasts and umbilical cord fibroblasts. Five hiPSC lines were generated using retroviral vectors and one hiPSC line using episomal vectors.

Table 1 Information about six hiPSC lines of the training set.

Cell Line	Vectors	Reprogramming factors	Cell source	Donor Racial	Donor Age	Donor Sex	Culture Method	Reference
201B7	Retrovirus	Oct3/4, Sox2, Klf4, c-Myc	Human dermal fibroblast	Caucasian	36	Female	MEF	Takahashi K, et al. Cell. 2007 Nov 30;131(5):861-72.
253G1	Retrovirus	Oct3/4, Sox2, Klf4	Human dermal fibroblast	Caucasian	36	Female	MEF	Nakagawa M, et al. Nat Biotechnol. 2008 Jan;26(1):101-6. Epub 2007 Nov 30.
409B2	Episomal	Oct3/4, Sox2, Klf4, L-Myc, Lin28+ shRNA for p53	Human dermal fibroblast	Caucasian	36	Female	MEF	Okita K, et al. Nat Methods. 2011 May;8(5):409-12.
RIKEN-1A (R-1A)	Retroviral	Oct3/4, Sox2, Klf4, c-Myc	Umbilical cord fibroblast	Japanese	-	Female	MEF	Fujioka T, et al. Hum Cell. 2010 Aug;23(3):113-8
RIKEN-2A (R-2A)	Retroviral	Oct3/4, Sox2, Klf4, c-Myc	Umbilical cord fibroblast	Japanese	-	Male	MEF	Fujioka T, et al. Hum Cell. 2010 Aug;23(3):113-8
RIKEN-12A (R-12A)	Retroviral	Oct3/4, Sox2, Klf4	Umbilical cord fibroblast	Japanese	-	Male	MEF	Fujioka T, et al. Hum Cell. 2010 Aug;23(3):113-8

We performed miRNA array, mRNA array, and CAGE on the undifferentiated hiPSCs to develop comprehensive transcript expression profiles of the undifferentiated hiPSCs. Next, we analysed the cardiomyocytes derived from hiPSCs using flow cytometry, quantitative reverse transcription-polymerase chain reaction (qRT-PCR), immunostaining, and beating analysis, and then determined the cardiac differentiation efficiency ranking. Based on the ranking, the hiPSCs lines were divided into high and low purity groups. To select candidate genes for predictive biomarkers, we compared the mRNA and miRNA expression and the transcription start sites (TSS) in undifferentiated hiPSCs to those of the high and low differentiation groups. Finally, using 13 hiPSC lines as a test set, we examined whether hiPSC lines with high capability of cardiomyogenic differentiation could be selected using the biomarker candidates (Fig. 2).

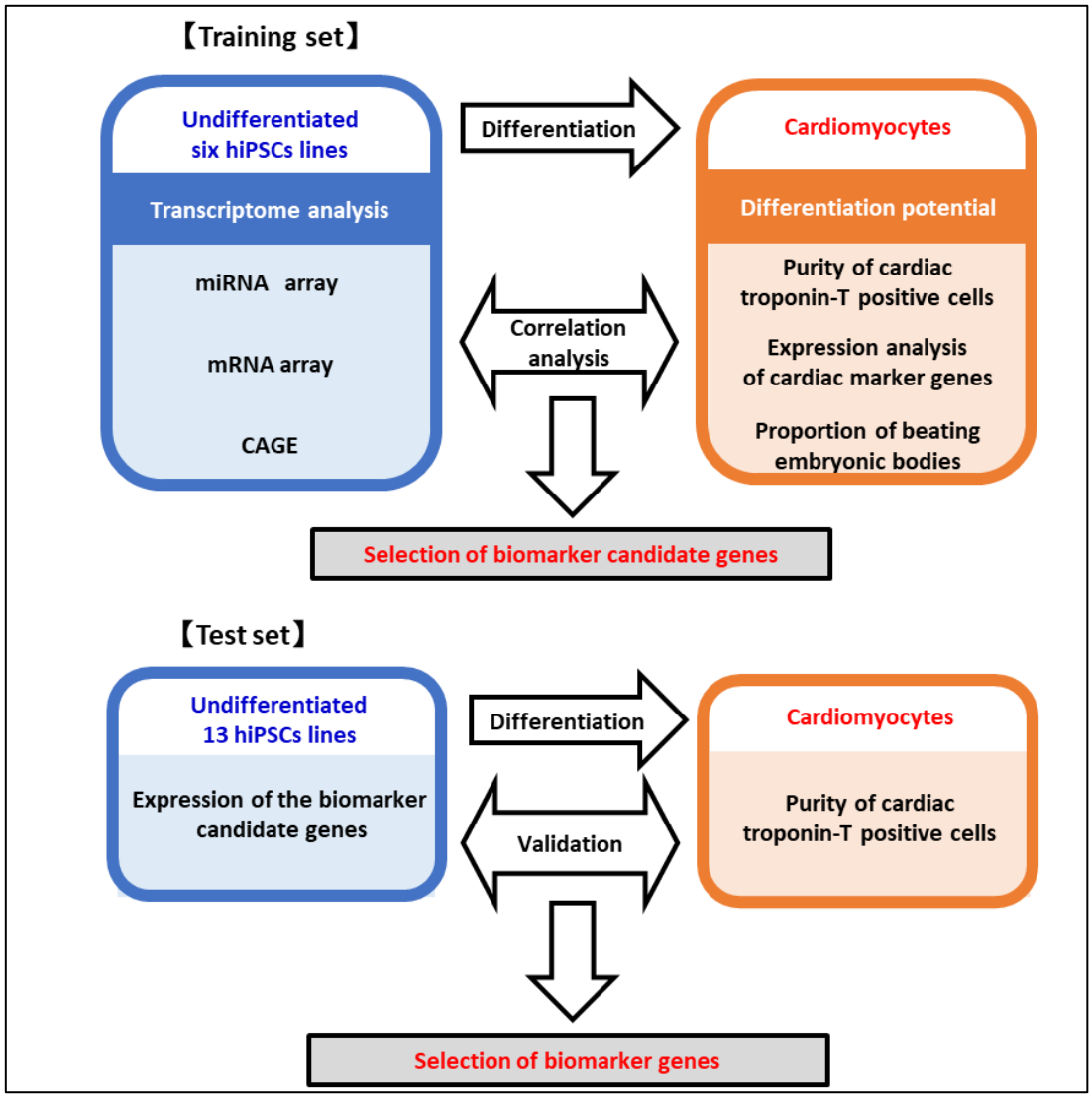

Figure 2 Workflow for selecting biomarkers to predict the cardiac differentiation potential of hiPSCs.

For cardiac differentiation of the hiPSCs, we used an embryoid body (EB) differentiation method (Fig. 3) based on a previous protocol with some modifications[28-30], as EBs formed using three-dimensional (3D) differentiation methods can be easily scaled up for clinical application of hiPSC-

derived cardiomyocytes. The formation of EBs is a critical process in hiPSCs differentiation[31]. Therefore, we compared the cardiac differentiation efficiency at the EB formation step of the cardiac differentiation process between two methods, the small cell clump method and the single cell method. The small cell clump method enzymatically disrupted hiPSC colonies into small cell clumps, whereas the single cell method enzymatically dissociated them into single cells. After evaluating cardiomyocytes derived from 253G1 hiPSCs for cardiac troponin T (cTnT) expression using flow cytometry, we concluded that EB formation using the single cell method was more efficient and robust with respect to cardiomyocyte differentiation than the small cell-clump method (single cells, 83.4 ± 1.2%, n = 110 vs. small clumps, 37.2 ± 2.2%, n = 46) (Fig. 4). In subsequent experiments, we determined the protocol for the use of the single cell method for cardiac differentiation as shown in Fig. 3.

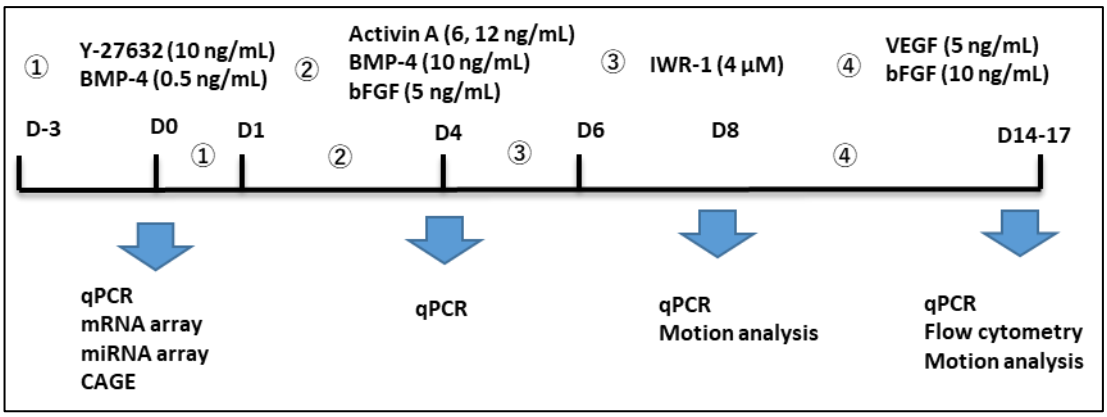

Figure 3 Schematic of the culturing process for cardiac differentiation in EB suspension cultures.

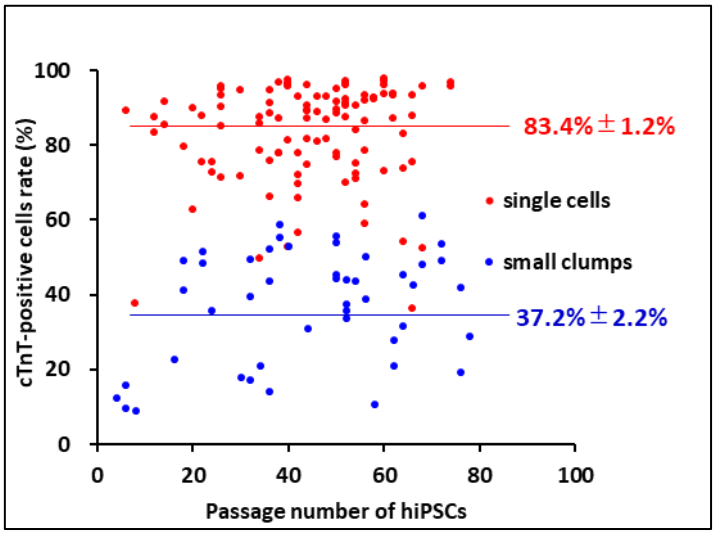

Figure 4 Effect of passage number and differentiation protocols on cardiac differentiation efficiency. Quantification of cTnT expression measured using flow cytometry in 253G1 hiPSCs prepared using two different protocols. Red dots: single cells (83.4 ± 1.2%, n = 110); blue dots: small clumps (37.2 ± 2.2%, n = 46). Data are expressed as mean ± SEM.

Differences in cardiac differentiation abilities of hiPSC lines.

First, we analysed the expression of undifferentiated cell markers of each hiPSC line using qRT-PCR. In addition to the similarities in cell morphology with pluripotent stem cells, the expression levels of the undifferentiated cell markers showed no significant difference among the six hiPSC lines (Fig. 5).

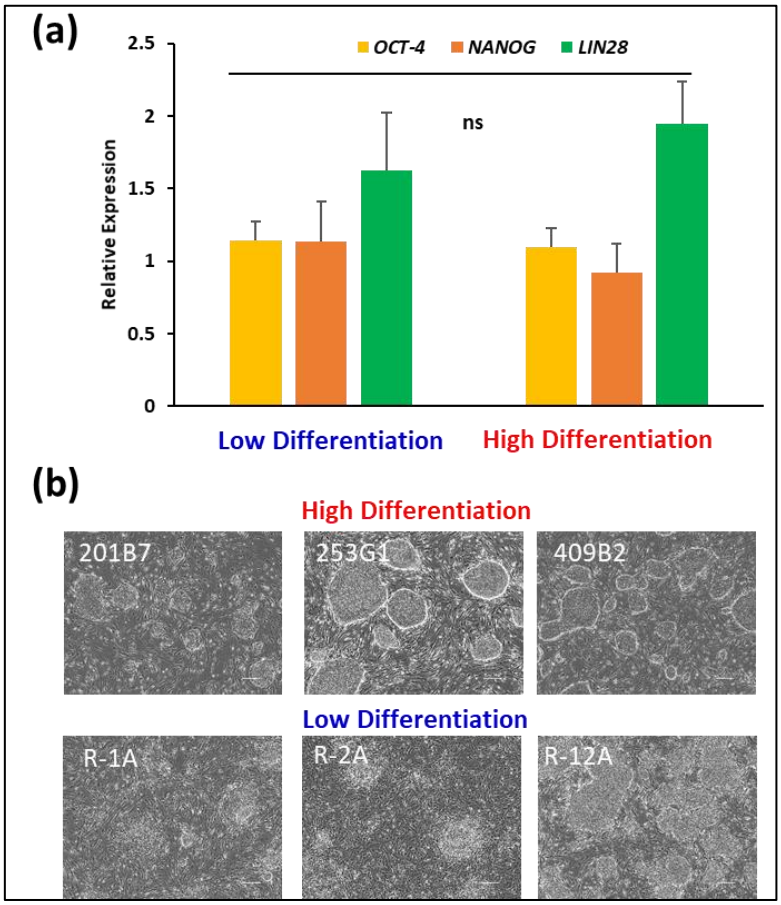

Figure 5 Undifferentiated gene expression profiles in a training set of hiPSCs lines. (a) Comparison of undifferentiated-related genes of hiPSCs lines (ns, not significant). Data are expressed as mean ± SEM (n = 3). All mRNA values are shown as fold change relative to the expression of R-2A in Low differentiation group. (b) Images of hiPSCs from experiments on training set of hiPSC lines. Scale bars = 300 μm.

Next, we subjected all the hiPSC lines to cardiac differentiation. A previous study reported that the highest number of cardiomyocytes was obtained from the mesoderm induced with 10 ng/mL BMP-4

and 6 ng/mL of Activin A, while higher or lower levels of Activin A and BMP-4 showed lower cardiomyocyte differentiation efficiencies[32]. Based on these findings, we applied two different concentrations of Activin A (6 ng/mL and 12 ng/mL) to the hiPSC lines with 10 ng/mL BMP-4, as shown in the Matrials and Methods. Flow cytometry analysis revealed differences in the prevalence of cTnT-positive cells among hiPSC lines. The percentage of cTnT-positive cells ranged from 7.7 ± 2.6% in RIKEN-12A cells to 92.3 ± 1.2% in 409 B2 cells (Fig. 6a). We calculated the percentage of beating EBs in differentiated hiPSCs at 8 d and 17 d post-differentiation. In the high differentiation group, approximately 50% and 90% of the EBs demonstrated rhythmical and synchronous beating at 8 d and 17 d post-differentiation, respectively. In contrast, the proportion of beating EBs was low in the low differentiation group, even after 17 d of differentiation (Fig. 6b).

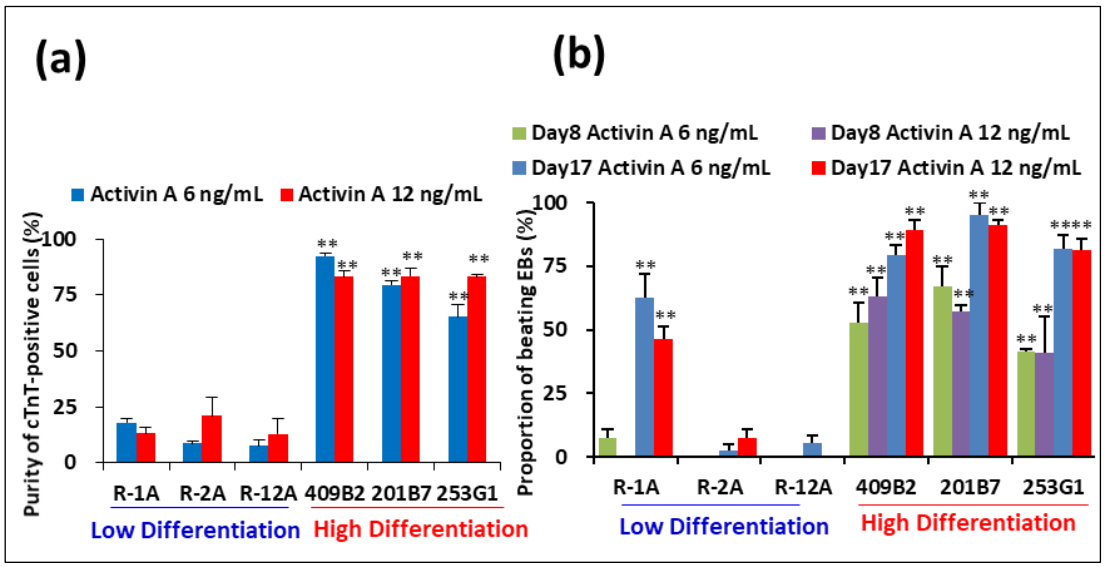

Figure 6 (a) Comparison of the cardiac differentiation ability of hiPSC lines using flow cytometric analysis 17 d post-induction of cardiac differentiation with different concentrations of Activin A; 6 ng/mL and 12 ng/mL. Data are expressed as mean ± SEM (n = 3). $^{**}p < 0.01$ vs. R-2A, ANOVA and Dunnett's test. (b) Beating analysis of rhythmic and synchronous beating of EBs at 8 d and 17 d post-differentiation with different concentrations of Activin A; 6 ng/mL and 12 ng/mL. Data are expressed as mean ± SEM (n = 3). $^{**}p < 0.01$ vs. R-2A, ANOVA and Dunnett's test.

With respect to cardiac differentiation capability, hiPSC lines could be categorised into two distinct groups, the low (R-1A, R-2A, and R-12A) and high differentiation groups (409B2, 201B7, and 253G1). Similarly, cardiomyocytes derived from the hiPSC lines in the high differentiation group expressed *TNT2* and *NKX2.5* at > 20-fold higher levels than those in the low differentiation group (Fig. 7). Differences in the concentrations of Activin A did not affect the percentage of cTnT-positive cells in the hiPSC lines with low differentiation efficiency.

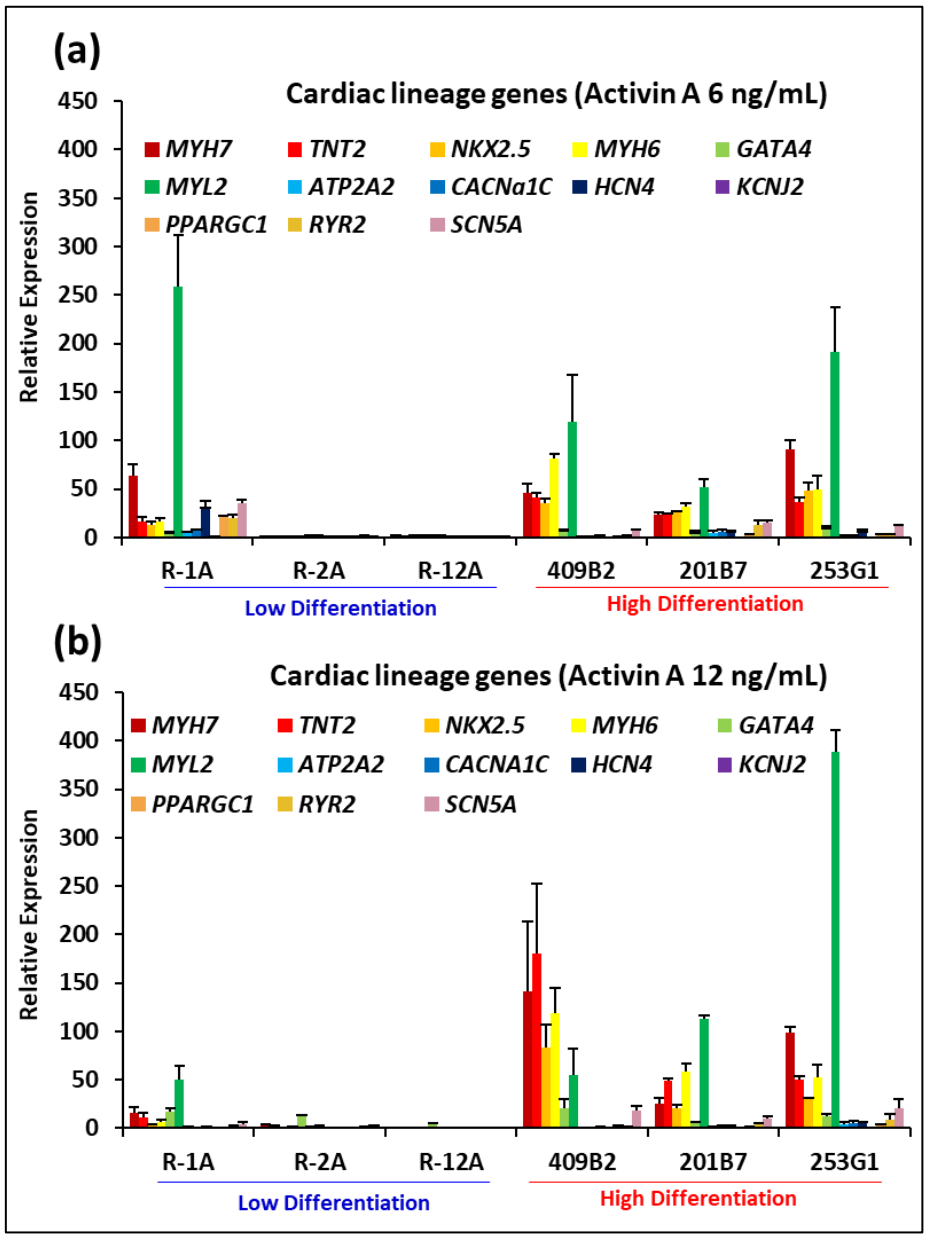

Figure 7 Cardiac lineage gene expression in EBs at 17 d post-differentiation. Differences in cardiac lineage differentiation potential among hiPSC lines induced with different concentrations of Activin A; 6 ng/mL (a) and 12 ng/mL (b). All mRNA values are shown as fold change relative to the expression of R-2A at 6 ng/mL Activin A. Data are expressed as mean ± SEM (n = 3).

Immunostaining also revealed that the number of cTnT-positive and NKX2.5-positive cells in the cardiac population was higher in the high differentiation group than in the low differentiation group (Fig. 8). Furthermore, in the low differentiation group, we detected many α-SMA and vimentin-positive smooth muscle cells and fibroblasts (Fig. 9), which was consistent with the results of previous reports[33,34].

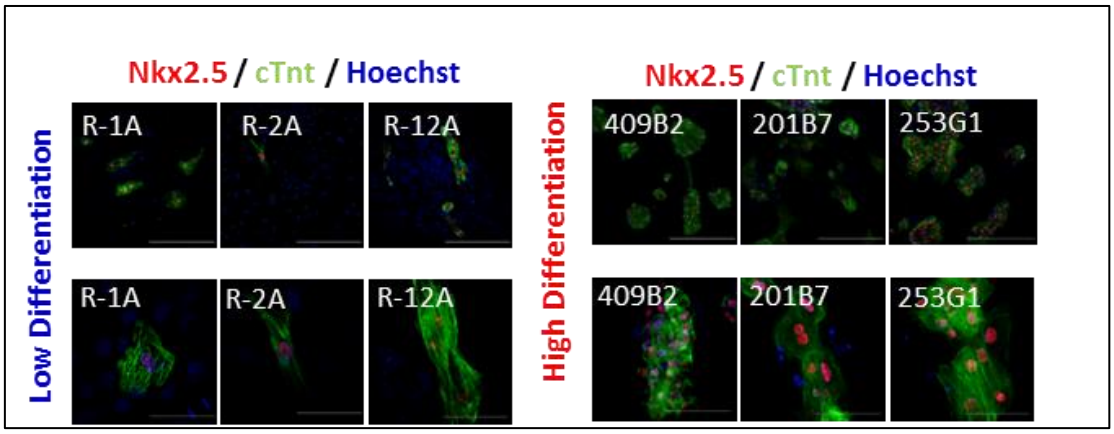

Figure 8 Immunofluorescence of hiPSC-derived cardiomyocytes for cardiac-specific markers. Micrographs show cTnT (green), Nkx2.5 (red), and Hoechst (blue) staining. Upper panels show low magnification and lower panels high magnification. Scale bars in upper panels = 300 μm, in lower panels = 100 μm.

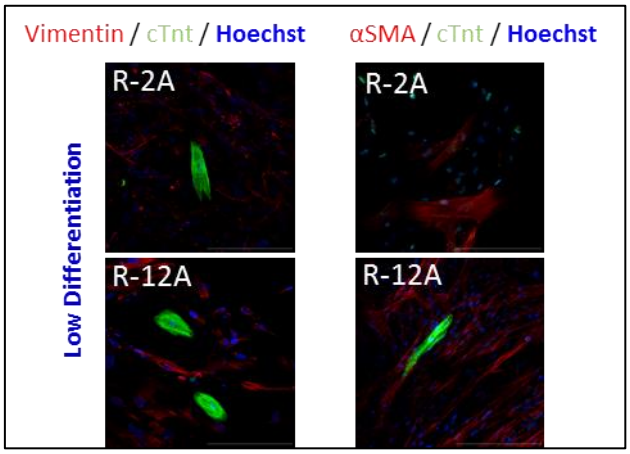

Figure 9 Immunofluorescence of cardiomyocytes derived from the low differentiation group (R-2A and R-12A) for cardiac-specific markers and smooth muscle cell and fibroblast markers. Left micrographs show cTnT (green), vimentin (red), and Hoechst (blue) staining. Right micrographs show cTnT (green), αSMA (red), and Hoechst (blue) staining. Scale bars = 100 μm.

The presence of residual undifferentiated cells after differentiation is a critical issue regarding clinical application of hiPSC-derived cardiomyocytes[4,35-38]. Therefore, we compared the expression of an hiPSC marker gene (*LIN28*) between the high and low differentiation groups on day 17 after the induction of differentiation. The expression of the undifferentiated hiPSC marker *LIN28* [39] was significantly higher in the low differentiation group than in the high differentiation group, suggesting that the lower efficiency in cardiac differentiation correlated inversely with a higher proportion of residual undifferentiated hiPSCs (Fig. 10). This underscored the importance of selecting hiPSC lines

with high differentiation potential for clinical application of hiPSC-derived cardiomyocytes.

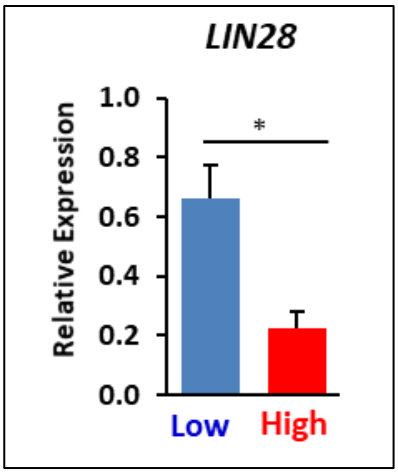

Figure 10 Expression of an undifferentiated hiPSC marker gene (*LIN28*) among hiPSC lines at 17 d post-cardiac differentiation. All mRNA values are shown as fold change relative to the expression of R-2A in Low differentiation group. Data are expressed as mean ± SEM (n = 6). *$p < 0.05$, *t*-test.

To examine the expression of cardiomyocyte-related genes in EBs, we extracted RNA from hiPSC-derived cardiomyocytes on day 17 after the induction of differentiation and performed qRT-PCR analysis. The expression of cardiomyocyte-related genes in EBs formed after induction with 12 ng/mL Activin A was higher in the high differentiation group than in the low differentiation group. Heat maps demonstrated the relative expression levels of cardiomyocyte-related genes and cardiac maturation-related genes in EBs derived from the hiPSC lines (Fig. 11).

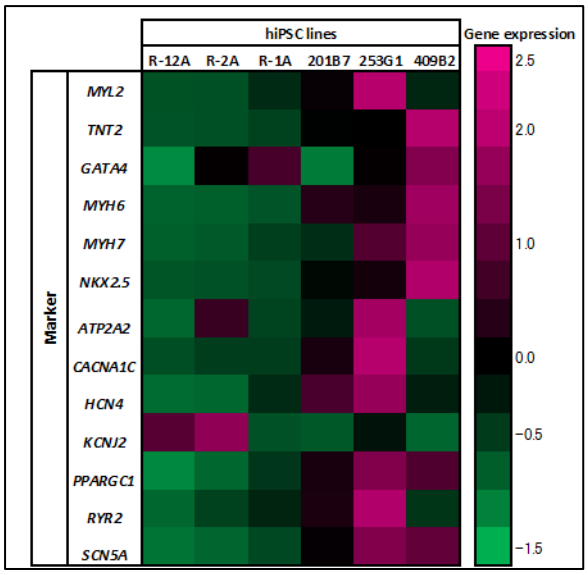

Figure 11 Heat map of cardiomyocyte-related genes and maturation-related genes among six hiPSC lines.

The cardiomyogenic ranking in EBs at 17 d post-differentiation was calculated based on the expression levels of six cardiomyocyte-related genes that were used as markers. We performed principal component analysis (PCA) to determine the rank of cardiac differentiation capacity among the six hiPSC lines. The ranking of hiPSC lines according to their cardiac differentiation potential (from highest to lowest) was 409B2, 253G1, 201B7, R-1A, R-2A, and R-12A (Fig. 12). The expression of cardiomyocyte-related genes was also high in the high differentiation group of EBs at 9 d post-differentiation. In addition, PCA based on the expression of cardiomyocyte-related genes in EBs demonstrated that the ranking order on day 9 post-differentiation was identical to that on day 17. We recalculated cTnT-positive rates of EBs at day 17 in the high and low differentiation groups and observed a significant difference in the cTnt-positive rate between the two groups; low = 13.6% ± 1.9 vs. high = 81.1% ± 3.3% ($p < 0.01$) (Fig. 13).

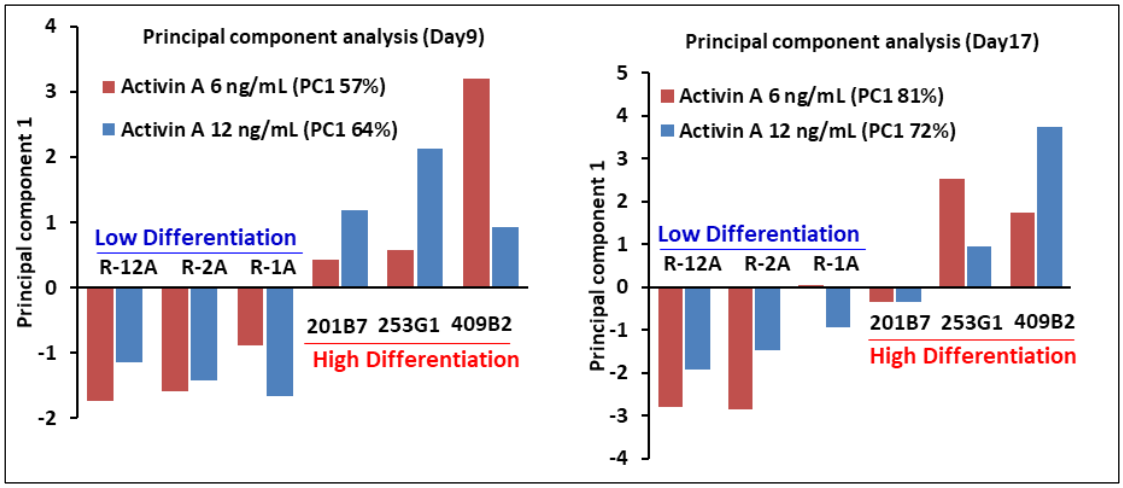

Figure 12 Principal component analysis of cardiac differentiation ability among six hiPSC lines at 17 d and 9 d post-cardiac differentiation with different concentrations of Activin A; 6 ng/mL and 12 ng/mL (PC1, first principal component scores).

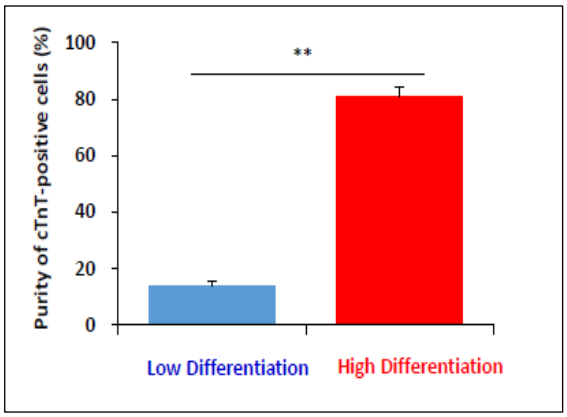

Figure 13 Percentage of cTnT-positive cells generated from hiPSCs in high and low differentiation groups at 17 d post-differentiation in a training set of hiPSC lines. Data are expressed as mean ± SEM (n = 6). $^{**}p < 0.01$, *t*-test.

Expression of germ layer-related genes in EBs derived from hiPSC lines.

To determine the differentiation direction of hiPSC lines in the early stage of differentiation, we

analysed gene expression of the three embryonic germ layers in EBs on day 4 of cardiac differentiation using qRT-PCR. The radar charts were based on the expression of genes specific to the mesendoderm and mesoderm (*GOOSECOID*, *PDGFR-A*, *FLK1*, and *BRACHYURY*), endoderm (*HNF3*, *SOX7*, *SOX17*, and *AMN*), and ectoderm (*ZIC1*, *SOX1*, and *PAX6*), which enabled assessment of the differentiation direction of the hiPSC lines with high and low differentiation potential. The hiPSC lines in the high differentiation group showed high expression of the mesodermal gene *FLK1* during the early stage of differentiation, whereas the hiPSC lines in the low differentiation group showed high expression of the endodermal gene *AMN* and the ectodermal gene *PAX6* (Fig. 14).

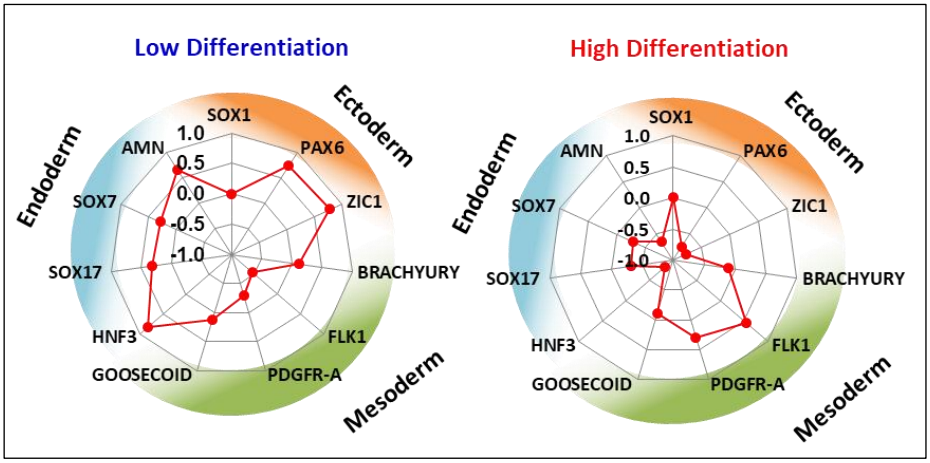

Figure 14 Comparison of differentiation direction of hiPSC lines at 4 d post-cardiac differentiation. Radar chart based on expression levels of endoderm, ectoderm, and mesoderm-related genes at 4 d post-cardiac differentiation.

We compared the ranking order of the hiPSC lines for cardiac differentiation potential at 17 d post-differentiation (Fig. 12) and the expression of germ-layer related genes at 4 d using Spearman's rank correlation analysis. The expression levels of *FLK1* correlated positively with the cardiac differentiation potential of hiPSC lines (Spearman correlation coefficient (rs) = 0.75, $p = 0.05$). In contrast, the expression of *ZIC1* and *AMN* showed negative correlation (*ZIC1*, rs = −0.57, $p = 0.18$; *AMN*, rs = −0.89, $p < 0.05$) (Table 2). Considering that cardiomyocytes are generated from the mesoderm[40], we inferred that mesodermal genes were upregulated in hiPSCs with high cardiac differentiation potential during the early stage of differentiation. The expression of germ layer-related genes at 4 d post-differentiation reflected a lineage-specific direction of hiPSC lines, suggesting that the differentiation process of EBs can be confirmed by measuring the expression levels of *FLK1*, *ZIC1*, and *AMN*.

Lineage	Marker	rs	*p*-value
Mesoderm	*FLK1*	0.75	0.05
	PDGFR-A	0.29	0.53
	GOOSECOID	0.21	0.64
	BRACHYURY	0.11	0.82
Endoderm	*SOX17*	0.39	0.38
	AMN	-0.89	0.01
	SOX7	-0.11	0.82
	HNF3	-0.36	0.43
Ectoderm	*SOX1*	-0.29	0.53
	PAX6	-0.79	0.04
	ZIC1	-0.57	0.18

Spearman correlation coefficient (rs) and *p*-value were calculated for statistical analysis.

Table 2 Spearman's rank correlation analysis between cardiac differentiation capacity at 17 d post-differentiation and the expression of EB-related genes at 4 d post-differentiation.

In addition, we observed that the diameter and area of EBs at 6 d post-differentiation were significantly different between the high and low differentiation groups of hiPSC lines. The mean diameters of the high and low differentiation groups were 286 ± 15 µm and 202 ± 9 µm, respectively ($p < 0.01$), whereas the mean areas of the high and low differentiation groups were 67329 ± 7145 µm^2 and 33676 ± 2892 µm^2 ($p < 0.01$), respectively (Fig. 15).

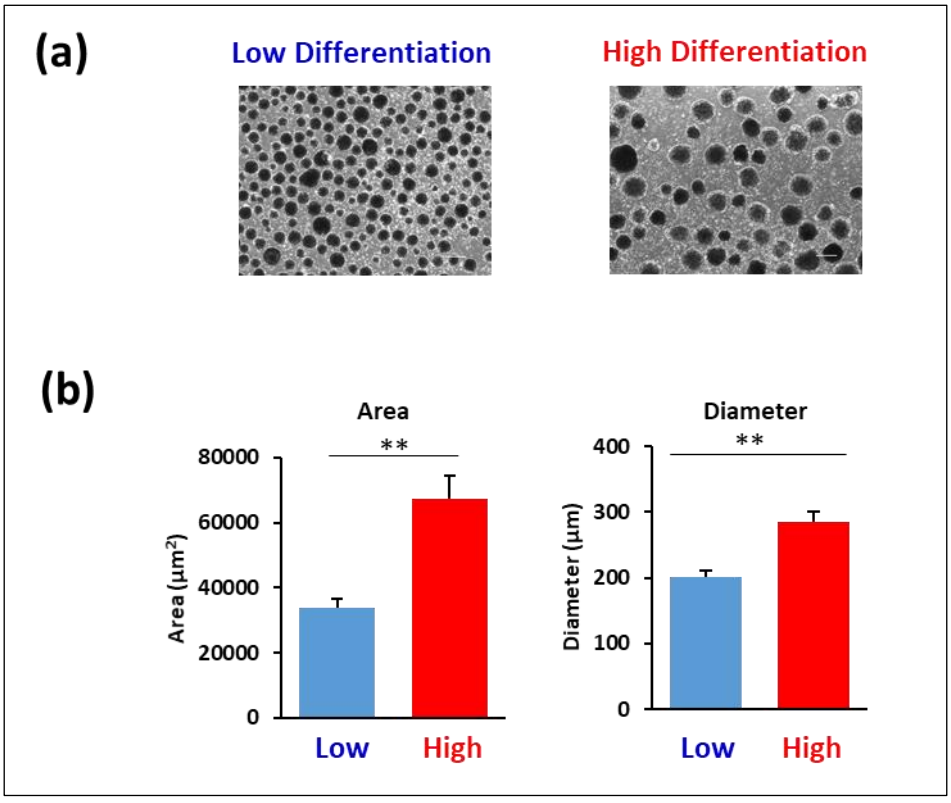

Figure 15 (a) Images of EBs 6 d post-cardiac differentiation of a training set of hiPSC lines. Scale bars = 300 µm. (b) Cross-sectional area and diameter of EBs at 6 d post-cardiac differentiation. Data are expressed as mean ± SEM (n = 6). **$p < 0.01$, *t*-test.

During the time course of cardiac differentiation, the diameter of EBs in the high differentiation group was significantly larger than that in the low differentiation group (Fig. 16a). We also demonstrated that the number of cells per EB was significantly different between the high and low differentiation groups (Fig. 16b). Previous studies also reported that the size of EBs was associated with cardiac differentiation in ES cells[31,41]. Therefore, the differences in cell proliferation may have a significant impact on cell differentiation in EBs, suggesting that the non-invasive measurement of EB size could be useful for process control.

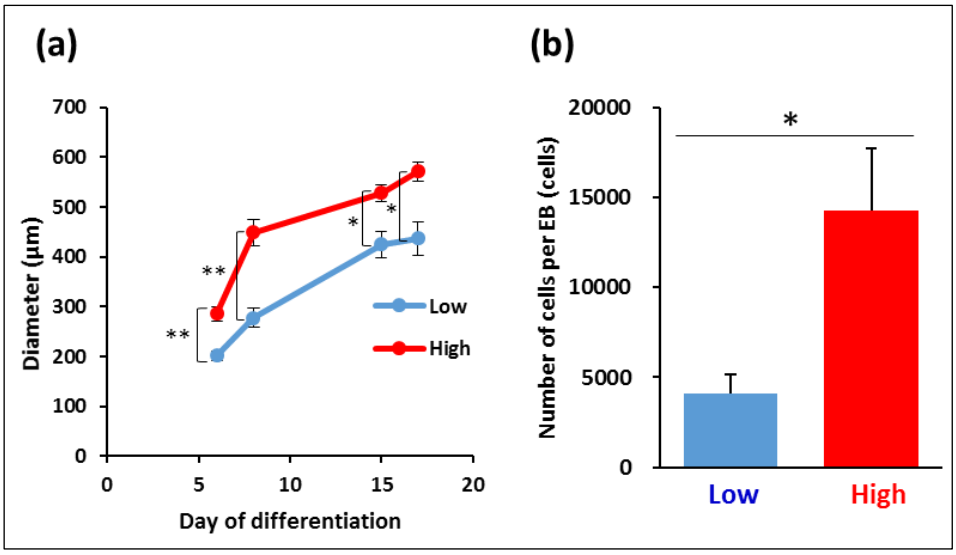

Figure 16 (a) The diameter of EBs in the high and low differentiation groups during the cardiac differentiation. Data are expressed as mean ± SEM (n = 6). $^{**}p < 0.01$; $^{*}p < 0.05$, *t*-test. (b) Number of cells per EB at 17 d post-cardiac differentiation. Data are expressed as mean ± SEM (n = 6). $^{*}p < 0.05$, *t*-test.

Comprehensive gene expression analysis of undifferentiated hiPSCs using miRNA and mRNA arrays.

Short non-coding RNAs, such as miRNAs, play important roles in silencing targeted genes by regulating post-transcriptional events[42]. Over 500 human miRNAs and small nucleolar RNAs (snoRNAs) have been described and each can regulate hundreds of different mRNAs[43]. Previous reports showed that miRNAs can control cell lineage determination and maturation of hiPSCs, possibly by regulating the transcriptome of undifferentiated hiPSCs[44-46]. Therefore, we performed miRNA array analysis and investigated miRNA expression profiles in the hiPSC lines (Fig. 17a). Differential analysis based on miRNA array comparing the high and low differentiation groups identified three miRNAs and two snoRNAs that were statistically different ($p < 0.05$, fold change > 2; Fig. 17b and Table 3). The miRNAs were expressed at significantly higher levels in the low differentiation group than in the high differentiation group, suggesting that these miRNAs may be involved in inhibiting the cardiac differentiation of hiPSCs and in maintaining cells in a state of self-renewal. To characterise the miRNA-mediated regulation of cardiac differentiation, miRNA target prediction was performed using the Ingenuity software (Qiagen). In total, 1924 genes were identified as being regulated by has-miR-139, mml-miR-204, and hsa-miR-629.

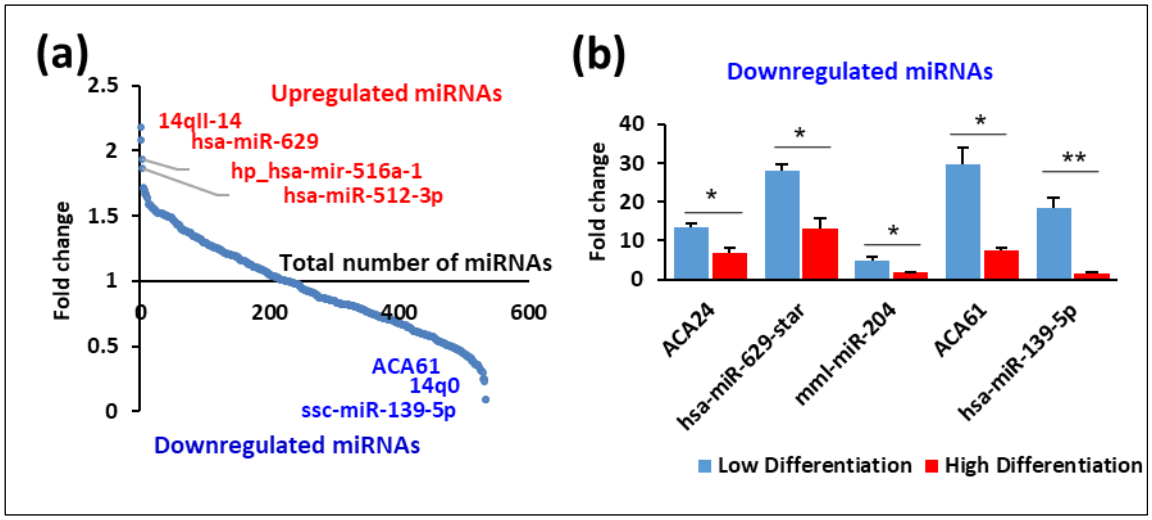

Figure 17 Transcriptome analysis of undifferentiated hiPSCs using miRNA array. (a) Plot showing expression of relative fold change expression between high and low differentiation groups. The x-axis indicates miRNA ranks for relative fold change and the y-axis shows the expression ratio (high/low) based on the differential profiles of 535 miRNAs in hiPSCs. (b) The differential expression of miRNAs in the high differentiation and low differentiation groups at a $p < 0.05$ and fold change > 2. Data are expressed as mean ± SEM (n = 3). $^{**}p < 0.01$; $^{*}p < 0.05$, *t*-test.

Probe Set ID	Fold Change	Type	Subcellular Location	Role in cell
ACA24	↓ − 2.0	small nucleolar RNA	-	-
hsa-miR-629-star	↓ − 2.1	mature microRNA	Cytoplasm	-
mml-miR-204	↓ − 2.6	mature microRNA	Cytoplasm	expression in, differentiation, antiapoptosis, activation in, proliferation, maturation
ACA61	↓ − 4.0	small nucleolar RNA	-	-
hsa-miR-139	↓ − 11.4	mature microRNA	Cytoplasm	growth

Three miRNAs and 2 snoRNAs showed significant difference in their levels in the high and low purity groups ($p < 0.05$, fold change > 2).

Table 3 Differentially expressed miRNAs and snoRNAs in the high and low purity groups.

Next, gene expression of the high and low differentiation groups was compared quantitatively to identify genes that were expressed in undifferentiated hiPSCs and may regulate cardiac differentiation. To perform a comprehensive analysis of the gene expression of undifferentiated hiPSCs, the mRNA array was used to identify differentially expressed transcripts in the high differentiation group compared to that in the low differentiation group. We identified 20 upregulated genes and 11 downregulated genes related to cardiac differentiation capability (fold change > 2 and $p < 0.01$; Fig. 18a and 18b). Ingenuity canonical pathway analysis showed that the genes differentially expressed between the high and low differentiation groups were involved in chemokine signaling and the Wnt/Ca+ signaling pathway (Fig. 18c).

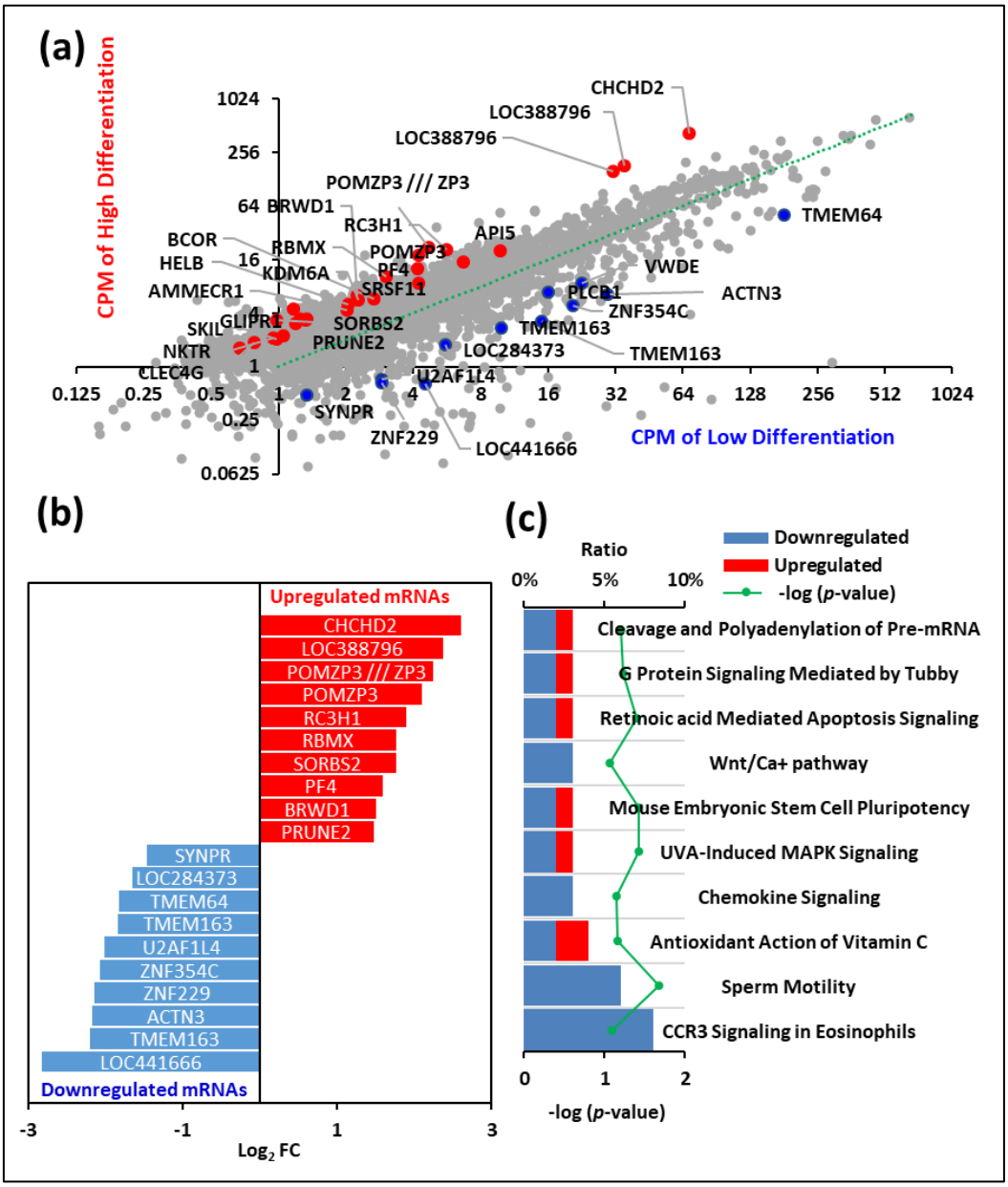

Figure 18 Transcriptome analysis of undifferentiated hiPSCs using GeneChip array.

(a) Scatter plot showing counts per million (CPM) of high differentiation (y-axis) vs. CPM of low differentiation (x-axis) from the mRNA array analysis. Red and blue coloured points and gene names indicate mRNAs that were significantly changed (fold change > 2 and $p < 0.01$, *t*-test). (b) Graph

showing the fold changes of the top and bottom differentially expressed genes from a GeneChip analysis comparing the high differentiation group with the low differentiation group. Selected genes have been colour-coded and labelled. Red, top 10 expression in the high differentiation group ($p <$ 0.01, *t*-test); blue, top 10 expression in the low differentiation group ($p < 0.01$, *t*-test). (c) Pathway analysis of the collective expression levels of interacting genes involved in specific pathways.

Profiling of TSS expression in undifferentiated hiPSCs using CAGE.

To further investigate the correlation between the genetic state of hiPSC lines and cardiac differentiation potential, we comprehensively compared TSS expression between the high and low differentiation groups using CAGE[47-49], which allowed identification of specific promoters for hiPSC differentiation. We identified 159 upregulated-transcripts and 707 downregulated-transcripts with more than twofold change in expression and FDR (false discovery rate) less than 0.01 in the high differentiation group compared to that in the low differentiation group (Fig. 19a). The genes identified using CAGE with the highest differentially expressed fold-changes in either direction are shown in Fig. 19b. Pathway analysis of these differentially expressed genes demonstrated that genes related to BMP receptors and human embryonic stem cell pluripotency were possibly involved in cardiac differentiation (Fig. 19c).

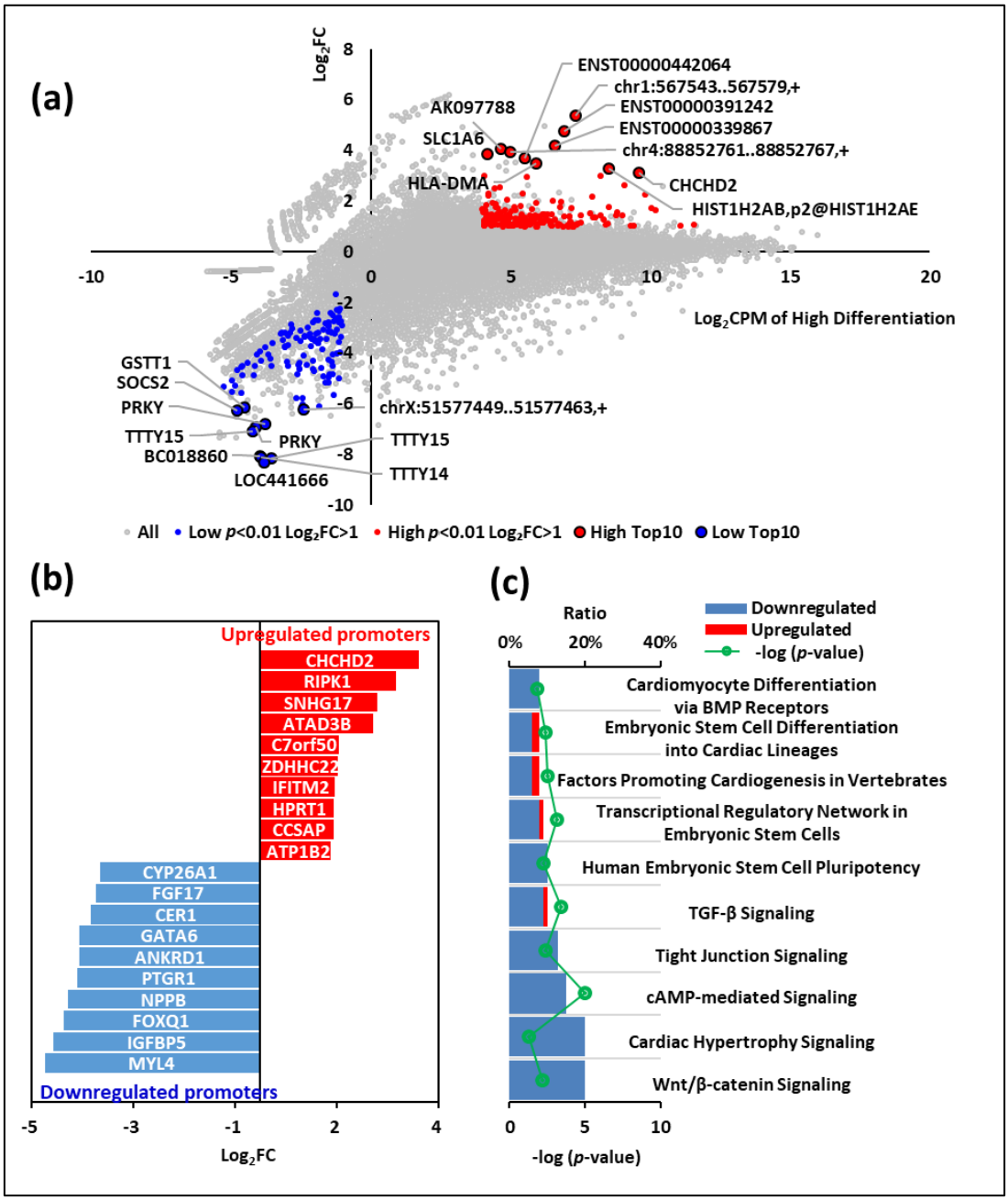

Figure 19 Undifferentiated hiPSC transcriptome profiling using CAGE. (a) Scatter plot shows log fold changes of individual genes (x-axis) vs. log CMP of high differentiation (y-axis) from CAGE. (b) Graph showing the fold changes of the top and bottom differentially expressed promoters

according to CAGE comparing the high differentiation group with the low differentiation group. Selected genes have been color-coded and labelled. Red, top 10 expression in the high differentiation group; blue, top 10 expression in the low differentiation group (FDR < 0.01, fold change > 2). (c) Pathway analysis of the collective expression levels of interacting genes involved in specific pathways.

In addition, Venn diagram analysis indicated that five genes (*CHCHD2*, *PF4*, *ZNF229*, *ZNF354C*, and *LOC441666*) were differentially expressed in both mRNA array and CAGE, 130 genes were differentially expressed in CAGE and were targets of differentially expressed miRNAs, and one gene (*TMEM64*) that was differentially expressed in the mRNA array was the target of differentially expressed miRNAs. Genes were categorised as positive predictors (upregulated in the high differentiation group) and negative predictors (downregulated in the high differentiation group) of cardiac differentiation potential (Fig. 20). In addition to these eight genes (*CHCHD2*, *PF4*, *ZNF229*, *TMEM64*, *FGF17*, *GATA6*, *ANKRD1*, and *IGFBP5*) identified using multiple analysis platforms, we selected 14 genes related to cell differentiation that were differentially expressed in at least one platform. In total, we selected 22 genes as biomarker candidates and listed their cellular roles (Table 4).

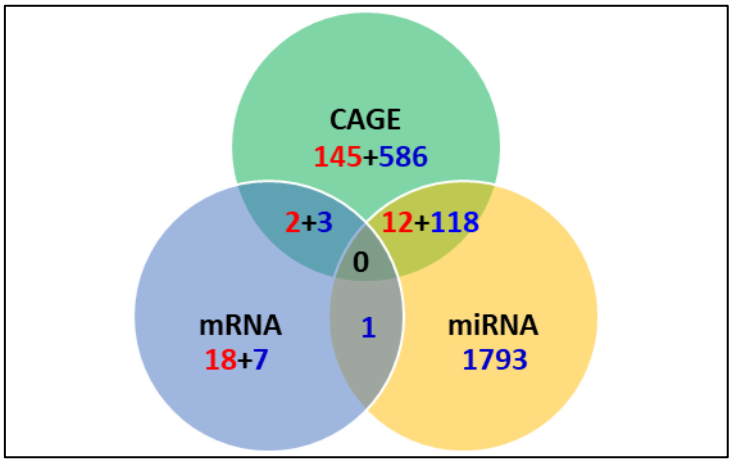

Figure 20 List of predictive biomarkers for cardiac differentiation. Venn diagram analysis visualising the overlap among the candidate genes identified using CAGE, mRNA array, and miRNA array analyses. The number of upregulated and downregulated mRNAs in hiPSC lines of the high differentiation group compared to that in the low differentiation group is indicated by red (up) and blue (down) colours, respectively.

Table 4 Predictive biomarker candidates for cardiac differentiation of hiPSC lines.

Positive predictors:

Gene name	Analysis	Entrez gene name	Role in cell
CHCHD2	CAGE & mRNA	coiled-coil-helix-coiled-coil-helix domain containing 2	phosphorylation in, differentiation, expression in, migration by, signaling in, formation in, formation
PF4	CAGE & mRNA	platelet factor 4	proliferation, chemotaxis, activation, binding, growth, chemoattraction, differentiation, migration, survival, adhesion
KDM6A	mRNA	lysine demethylase 6A	differentiation, identity, expression in, remodelling, cell viability, proliferation
BCOR	mRNA	BCL6 corepressor	differentiation, formation
POMZP3	mRNA	POM121 and ZP3 fusion	formation
RBMX	mRNA	RNA binding motif protein, X-linked	alternative splicing by, homologous recombination in, expression in
RC3H1	mRNA	ring finger and CCCH-type domains 1	proliferation, homeostasis, quantity, number, expression in, abnormal morphology, degradation in, differentiation
GLIPR1	mRNA	GLI pathogenesis related 1	apoptosis, sensitivity, destruction in, cell cycle progression, transactivation in, degradation in, binding in, ubiquitination in
RIPK1	CAGE	receptor interacting serine/threonine kinase 1	apoptosis, activation in, cell death, necroptosis, necrosis, expression in, production in, survival, proliferation, formation in
C7orf50	CAGE	chromosome 7 open reading frame 50	unknown

Negative predictors:

Gene name	Analysis	Entrez gene name	Role in cell
ZNF229	CAGE & mRNA	zinc finger protein 229	unknown
PLCB1	mRNA	phospholipase C beta 1	differentiation, expression in, activation in, cell death, G2/M phase transition, loss, binding, size, hypertrophy, fusion
TMEM64	mRNA & miRNA	transmembrane protein 64	differentiation
PTGR1	CAGE	prostaglandin reductase 1	survival
FOXQ1	CAGE	forkhead box Q1	formation by, expression in, quantity, migration, proliferation
MYL4	CAGE	myosin light chain 4	unknown
FGF17	CAGE & miRNA	fibroblast growth factor 17	proliferation, abnormal morphology, phosphorylation in, survival
GATA6	CAGE & miRNA	GATA binding protein 6	differentiation, expression in, proliferation, apoptosis, transcription in, transactivation in, specification, growth, abnormal morphology
ANKRD1	CAGE & miRNA	ankyrin repeat domain 1	apoptosis, response, differentiation, cell viability, expression in, colony formation
IGFBP5	CAGE & miRNA	insulin like growth factor binding protein 5	growth, migration, apoptosis, proliferation, survival, translation in, differentiation, cell spreading, expression in, quantity
WNT3	CAGE	Wnt family member 3	expression in, binding in, accumulation in, signaling in, transcription in, phosphorylation in, proliferation, differentiation, stabilization in
IGF2	CAGE	insulin like growth factor 2	proliferation, differentiation, growth, migration, phosphorylation in, apoptosis, expression in, activation in, survival, quantity

Validation of 22 candidate genes as predictors of cardiac differentiation potential.

Differential gene expression between the high and low differentiation groups of undifferentiated hiPSCs revealed potential candidate genes related to cardiac differentiation potential. The expression of 22 biomarker candidate genes that predicted the cardiac differentiation capacity of hiPSC lines was confirmed in 13 additional hiPSC lines. A previous report showed that culture conditions affect the pluripotency and cardiac differentiation of hiPSCs[50]. Therefore, hiPSC lines of the test set were evaluated under on-feeder and feeder-free conditions to determine the effects of extracellular matrix on cardiac differentiation (Table 5).

Cell line name	Vectors	Reprogramming factors	Source				Culture method	cardiac troponin T positive cells rate (%)	Reference
			Origin	Racial	Age	Sex			
454 E2	Episomal	Oct3/4, Sox2, Klf4, L-Myc, Lin28+ shRNA for p53	Dental pulp cell	Japanese	16	F	MEF	5%	Kajiwara M, et al. Proc. Natl. Acad. Sci. USA. 2012 Jul 31;109(31):12538-43.
606A1	Episomal	Oct3/4, Sox2, Klf4, L-Myc, Lin28	Cord blood cell	Japanese	-	F	MEF	16%	Kajiwara M, et al. Proc. Natl. Acad. Sci. USA. 2012 Jul 31;109(31):12538-43.
648A1	Episomal	Oct3/4, Sox2, Klf4, L-Myc, Lin28+ shRNA for p53	Peripheral blood mononuclear cell	Japanese	30s	M	MEF	17%	Okita K, et al. Stem Cells. 2013 Mar;31(3):458-66.
Tic	Retroviral	Oct3/4, Sox2, Klf4, c-Myc	Human fetus lung cell	no data	14 week fetus	M	MEF	4%	Fujioka T, et al. Hum Cell. 2010 Aug;23(3):113-8
ATCC-HYR0103	Retroviral	Oct4, Sox2, Klf4, Myc	Primary hepatic fibroblast	Hispanic	31	M	MEF	3%	-
ChiPSC18	Retroviral	Oct4, Sox2, Klf4, Myc	Human dermal fibroblast	North african	32	M	iMatrix511	69%	Asplund A, et al. Stem Cell Rev. 2016 Feb;12(1):90-104.
ChiPSC21	Retroviral	Oct4, Sox2, Klf4, Myc	Human dermal fibroblast	North african	26	M	iMatrix511	8%	Asplund A, et al. Stem Cell Rev. 2016 Feb;12(1):90-104.
ChiPSC22	Retroviral	Oct4, Sox2, Klf4, Myc	Human dermal fibroblast	North african	32	M	iMatrix511	33%	Asplund A, et al. Stem Cell Rev. 2016 Feb;12(1):90-104.
PCi-1432	Episomal	Sox2, Oct4, Klf, cMyc	Peripheral blood mononuclear cell	South european	30-50	M	iMatrix511	3%	-
PCi-1503	Episomal	Oct4, Sox2, Lin28, Klf4, L-Myc	Peripheral blood mononuclear cell	Asian	no data	F	iMatrix511	1%	-
PCi-1533	Episomal	Oct4, Sox2, Lin28, Klf4, L-Myc	Peripheral blood mononuclear cell	African	26	F	iMatrix511	11%	-
Ff-201B7	Retrovirus	Oct3/4, Sox2, Klf4, c-Myc	Human dermal fibroblast	Caucasian	36	F	iMatrix511	84%	Takahashi K, et al. Cell. 2007 Nov 30;131(5):861-72.
Ff-253G1	Retrovirus	Oct3/4, Sox2, Klf4	Human dermal fibroblast	Caucasian	36	F	iMatrix511	79%	Nakagawa M, et al. Nat Biotechnol. 2008 Jan;26(1):101-6. Epub 2007 Nov 30.

Table 5 Information about the 13 hiPSC lines of the test set.

To validate the biomarker candidate genes using a method other than CAGE, and mRNA and miRNA arrays, we measured the expression of each biomarker candidate gene in these undifferentiated hiPSC lines using qRT-PCR. The cells were then differentiated into cardiomyocytes using the protocol shown in Fig. 3. The differentiated cells of each hiPSC line were divided into high

and low differentiation groups based on whether the cTnT-positive rate was 50% or higher, and the cTnt-positive rates were recalculated for the differentiated hiPSC lines of each group (Table 5 and Fig. 21).

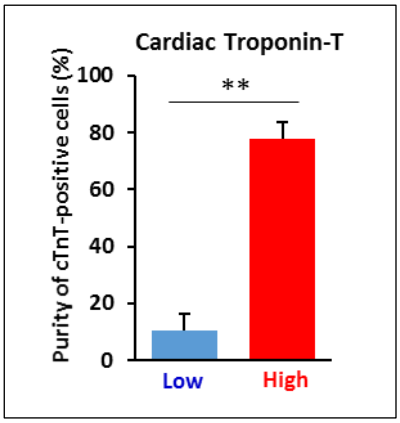

Figure 21 Validation results for expression of 22 genes in a test set of 13 hiPSC lines. (a) Percentage of cTnT-positive cells generated from hiPSCs in high and low differentiation groups at 17 d post-differentiation in a test set of hiPSC lines. Data are expressed as mean ± SEM (n = 3-10). $^{**}p < 0.01$, *t*-test.

Next, we compared the expression levels of the biomarker candidate genes between the two groups. Twenty of the 22 genes exhibited no significant difference in expression between the two groups (Fig. 22). These observations suggested that those 20 candidate genes were false positives. Two of the candidate genes in the test set correlated either positively or negatively with cardiac differentiation potential. The gene that positively correlated with cardiac differentiation potential was *PF4* (Fig. 23a, and 23c), which is known to be involved in cellular functions, including proliferation, chemotaxis, activation, binding, growth, chemoattraction, differentiation, migration, survival, and adhesion. PF4 has been reported as one of the most potent antiangiogenic chemokines influencing angiogenesis[51], suggesting that *PF4*-high expressing hiPSC lines may be efficiently induced to differentiate into cardiomyocytes. In addition, *TMEM64* was found to negatively correlate with cardiac differentiation potential (Fig. 23b and 23d). Reports show that *TMEM64* is involved in WNT signaling[52], possibly affecting the differentiation of cells into cardiomyocytes. Although WNT3[20],

IGF2[18], and CHCHD2[19] have been reported as differentiation markers, results from the present study revealed only low correlation of these genes as a predictive marker for cardiac differentiation.

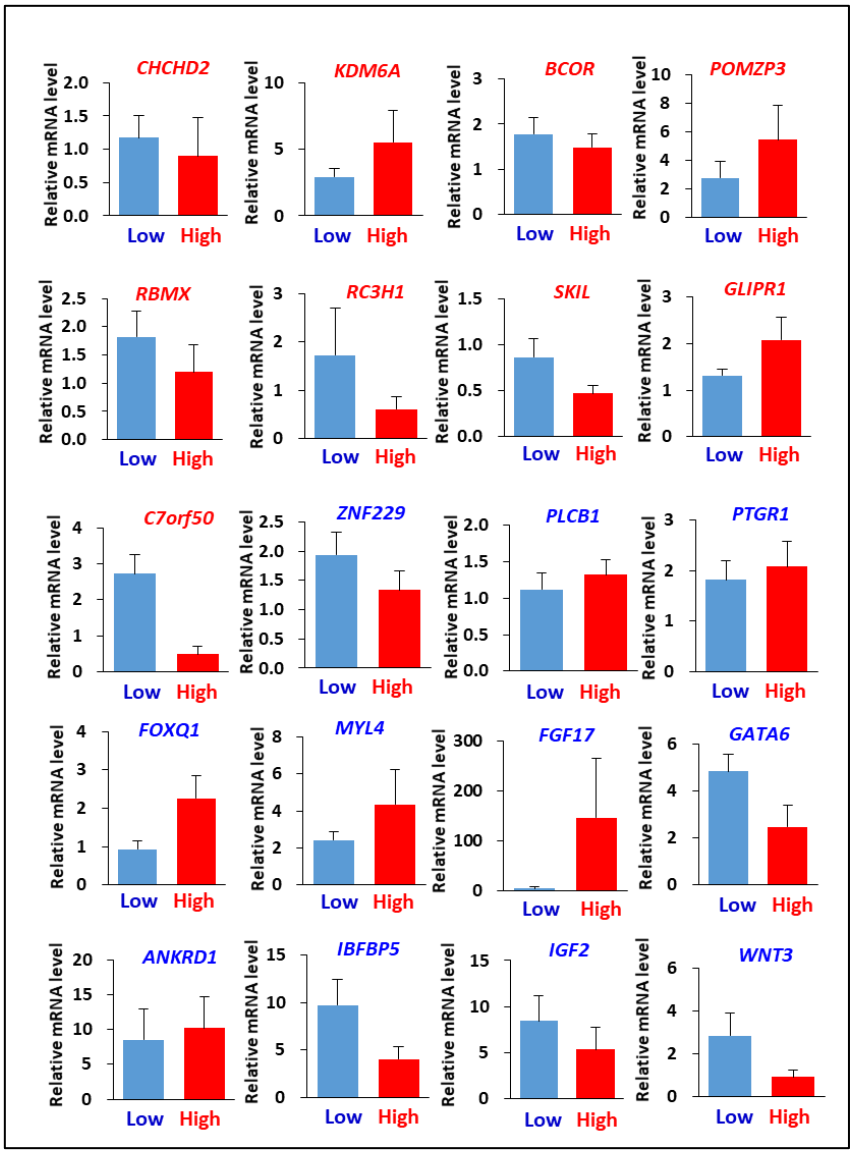

Figure 22 Expression profile of 20 candidate genes based on qRT-PCR for positive predictors (red) and negative predictors (blue). All mRNA values are shown as fold change relative to the expression of PCi-1533.

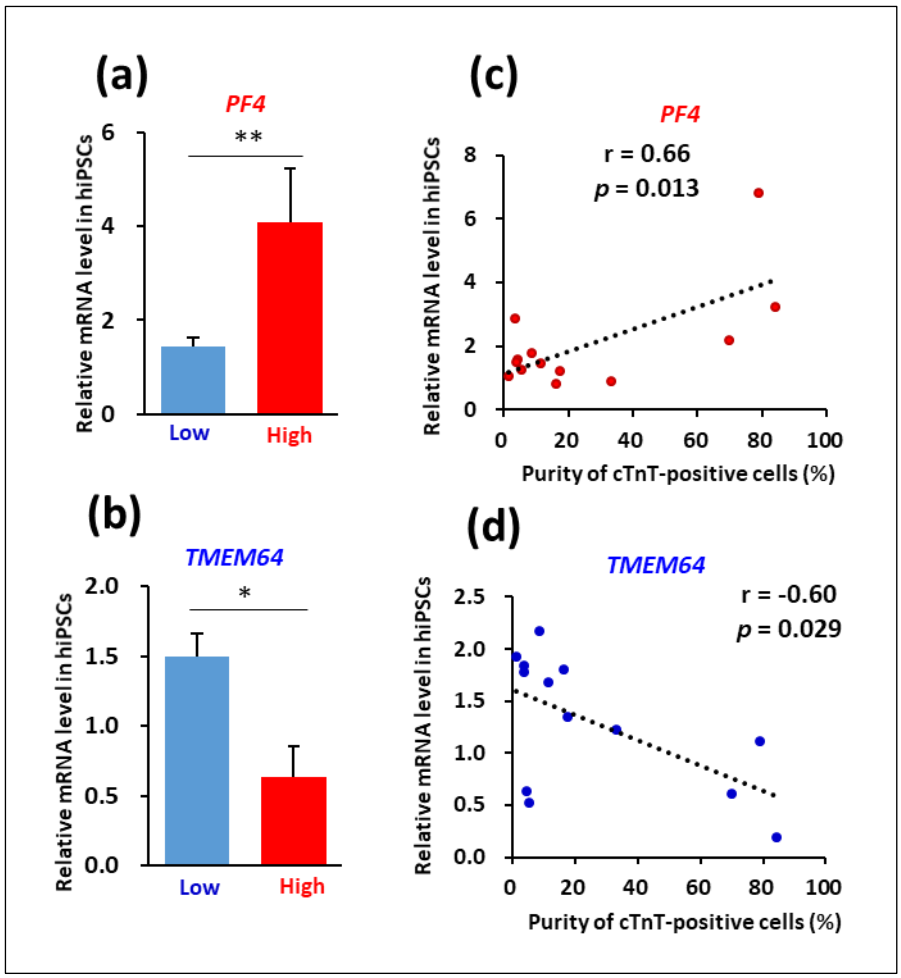

Figure 23 Validation results for expression of 22 genes in a test set of 13 hiPSC lines. (a,b) *PF4* and *TMEM64* mRNA expression levels in the high and low differentiation groups were quantified using qRT-PCR. Data are expressed as mean ± SEM (n = 3-10). $^{**}p < 0.01$; $^{*}p < 0.05$, *t*-test. (c,d) *PF4* and *TMEM64* mRNA levels in undifferentiated hiPSCs correlated with their cardiac differentiation efficiency along with r and *p* values. All mRNA values are shown as fold change relative to the expression of PCi-1533. Each dot indicates the expression level in each hiPSC line.

As WNT signaling and mitochondrial function have been reported to play critical roles in the cardiac differentiation of hiPSCs[20,53], we attempted to differentiate hiPSCs after treatment with WNT signaling inhibitors IWR-1 and IWP-2, a WNT signaling activator CHIR99021, and a mitochondrial function inhibitor MitoBlock-6[54]. We observed that CHIR99021 and Mitoblock-6 significantly decreased the cardiac differentiation efficiency compared to that of vehicle (Fig. 24a). In addition, treatment with CHIR99021 suppressed *PF4* expression in hiPSC lines, indicating that the activity of WNT signaling was associated not only with the cardiomyocyte differentiation efficiency, but also with *PF4* expression (Fig. 24b). Furthermore, the 2D plots of the raw data points of Fig. 24 indicated a significant correlation between *PF4* gene expression in hiPSCs with a common genetic background and purity of cTnT-positive cells after their differentiation (Fig. 24d). In contrast, *TMEM64* expression did not correlate with changes in cardiac differentiation potential (Fig. 24c and 24e).

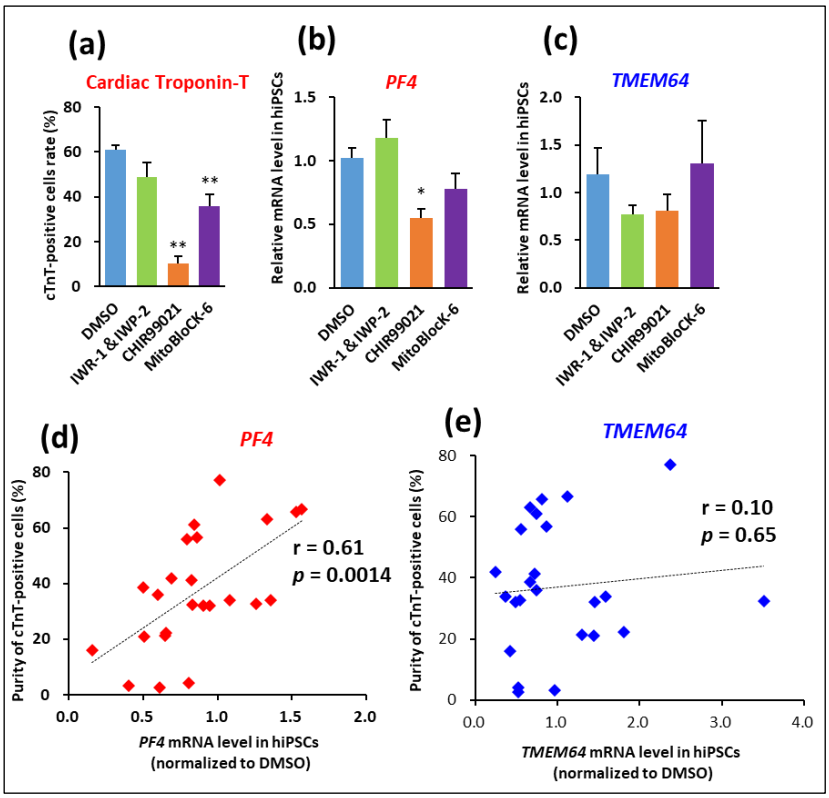

Figure 24 (a) Percentage of cTnT-positive cells generated from hiPSC lines with a common genetic background (253G1) incubated with IWR-1 & IWP-2, CHIR99021, and Mitoblock-6 at 17 d post-differentiation. Data are expressed as mean ± SEM (n = 6). $^{**}p < 0.01$ vs. DMSO, ANOVA and Dunnett's test. (b, c) *PF4* and *TMEM64* mRNA expression levels of the hiPSC lines incubated with IWR-1 & IWP-2, CHIR99021, and Mitoblock-6. All mRNA values are shown as fold change relative to the expression of mRNA in DMSO-treated control cells. Data are expressed as mean ± SEM (n = 6). $^{*}p < 0.05$ vs. DMSO, ANOVA and Dunnett's test. (d, e) Pearson correlation analysis between *PF4* and *TMEM64* mRNA levels in undifferentiated hiPSCs and their cardiac differentiation efficiency along with r and *p* values.

Finally, we performed experiments to examine the functional significance of PF4 in cardiomyocyte differentiation of hiPSCs. In brief, cardiomyocyte differentiation of hiPSCs, which were cultured in the presence of PF4 (1 μM)[55] for 2 days prior to differentiation induction, led to 1.3- to 10.0-fold higher expression levels of cardiomyocyte-specific genes (*MYH7*, *MYL2,* and *TNT2*) at 14 d post-differentiation, compared with the controls (Fig. 25). Similarly, PF4 treatment of hiPSCs resulted in an increased number of strong beating EBs, compared to the controls. The cross-sectional area of the EBs, which was associated with the cardiomyogenic potential of hiPSC lines (Fig. 15), was significantly larger in the PF4-treated group, compared with that in the control group. In addition, the proportion of beating EBs in the PF4-treated group was significantly higher than that in the control group (Fig. 26).

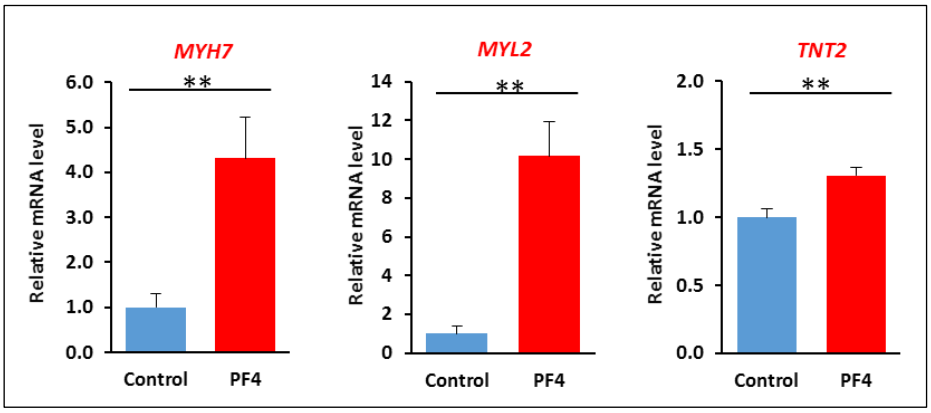

Figure 25 Expression of cardiomyocyte-specific genes *(MYH7*, *MYL2*, and *TNT2*) at 14 d post-differentiation in EBs derived from hiPSCs treated with PF4 (1 μM) or DMSO (1%) as a control for

2 days prior to differentiation induction. The mRNA values are shown as fold changes relative to the expression in the controls. Data are expressed as mean ± SEM (n = 9). **$p < 0.01$, *t*-test.

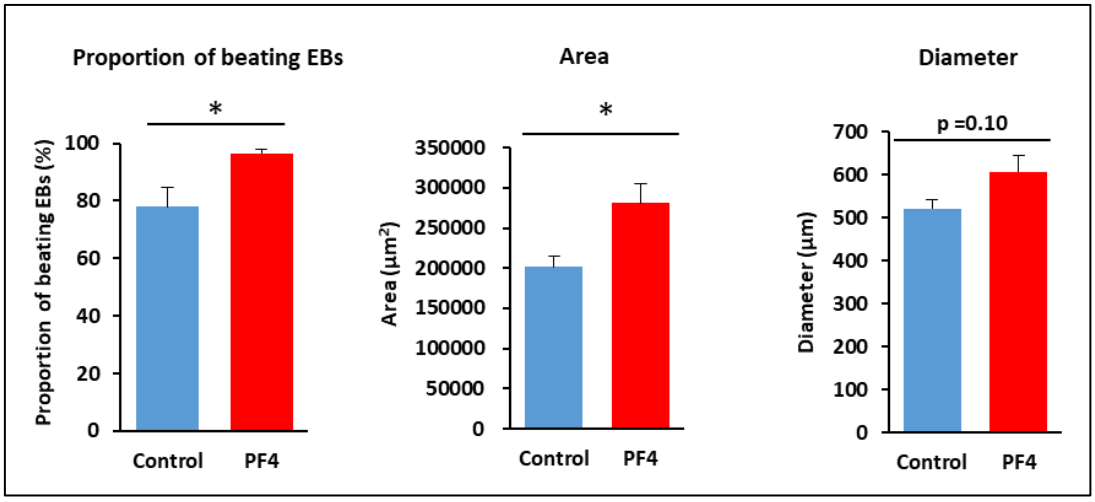

Figure 26 Proportion of beating EBs, and cross-sectional area and diameter of EBs derived from 253G1 hiPSCs treated with PF4 (1 µM) or DMSO (1 %) as a control for 2 days prior to differentiation induction. Data were obtained from EBs at 14 d post-differentiation and are expressed as mean ± SEM (n = 6). *$p < 0.05$, *t*-test.

Discussion

Transcriptome comparison of hiPSC lines between the high and low differentiation groups revealed that *PF4* was a novel biomarker that may be used to distinguish between high and low cardiac differentiation potential of hiPSC lines. For clinical application of hiPSC-derived cardiomyocytes, suitable hiPSC lines must be selected, from which highly purified cardiomyocytes containing minimal undifferentiated hiPSCs can be generated. Our results suggest that the cardiac differentiation potential of individual hiPSC lines can be predicted by assessing *PF4* expression in hiPSCs. In addition, it was possible to reduce the potential risk of tumorigenicity from residual undifferentiated hiPSCs in the end product. In our study, we identified *PF4* as a marker for selection of hiPSCs as a raw material for cardiomyocyte production. The data from the test set of 13 hiPSC lines suggest the robustness of this biomarker, though the usefulness of *PF4* as a biomarker could be changed by further optimization of the differentiation protocol.

We showed variability in cardiac differentiation potential of hiPSCs lines by assessing cardiomyocytes differentiated from hiPSCs. We focused on the gene expression profiles of hiPSC lines to identify predictive biomarkers for hiPSCs with cardiac differentiation potential. As factors specific to the original cells and reprogramming methods are both known to influence the differentiation efficiency of hiPSC lines[56], we used hiPSC lines that were derived from different

somatic origins and reprogramming methods. Our results showed that hiPSC lines derived from skin fibroblasts tended to differentiate more efficiently into cardiomyocyte than hiPSC lines derived from umbilical cord fibroblasts. Furthermore, we used 253G1, 201B7, and 409B2, which were generated from the same individual and dermal fibroblasts, and different reprogramming methods as shown in Table 1. Our results demonstrate that the hiPSC lines expressing high levels of PF4 had a higher capacity to differentiate into cardiomyocytes. The levels of PF4 and the purity of cardiomyocytes were not different among the three hiPSC lines (data not shown). We also found that the cardiac differentiation capacity of hiPSCs was not affected by the MEF and feeder-free culture condition (data not shown). Taken together, these findings suggest that their efficiencies for cardiomyocyte differentiation do not depend on the reprogramming methods. Considering the epigenetic influence of somatic cells on differentiation, analysis of DNA methylation profiles of hiPSC lines and large-scale differentiation experiments might be necessary in the future, as has been reported previously[13,14].

In the present study, we evaluated hiPSCs in an undifferentiated state using three comprehensive gene expression analysis approaches that included miRNA array, mRNA array, and CAGE, as multiple strategies for identifying biomarkers that are clearly required to reduce the number of false positives. In the process of identifying biomarkers for differentiation potential, the three platforms

for quantitating transcript expression allowed us to better understand the mechanisms underlying hiPSC differentiation into cardiogenic lineages. Furthermore, the three independent genetic analyses identified common genes by comparing the transcript expression patterns between hiPSC lines with high and low differentiation potential. Validation using qRT-PCR analysis of candidate genes of the test set of hiPSC lines further decreased the number of false positives. Indeed, *PF4* expression overlapped as a candidate gene between the CAGE and GeneChip analyses, and 20 false positive markers were eliminated using the qRT-PCR-based validation. Using our comprehensive gene analysis data of commercially available undifferentiated hiPSCs, researchers can search for new differentiation biomarkers in desired cell types such as blood, liver, and neural cells.

Moreover, the marker for cardiac progenitors was used to identify the cardiomyogenic potential of the hiPSC lines at the early stage of differentiation[32,57]. However, our results demonstrated that the cardiac differentiation capacity could be predicted by measuring the expression of *PF4* in the undifferentiated hiPSCs. PF4, a known heparin neutralising factor released from platelets, plays a key role in the activation and differentiation of monocytes and macrophages and is associated with systemic sclerosis and cancer[58]. In addition, PF4 levels in the blood have been proposed as biomarkers for determining cancer types[59]. We demonstrated that *PF4* expression decreased in hiPSC lines with low cardiac differentiation potential and that higher percentage of residual

undifferentiated cells remained post-differentiation in the low differentiation group of hiPSC lines than in the high differentiation group. Taken together, these results suggested that *PF4* can be used to distinguish among hiPSC lines associated with tumorigenicity after induction of differentiation. In addition, we found that the pretreatment of hiPSCs with PF4 enhanced the cardiac differentiation potential of hiPSCs, suggesting that PF4 has a cause-and-effect relationship with their cardiac differentiation. These observations suggest that PF4 gene expression, which could vary not only between hiPSC lines but also between pharmacological conditions, reflect the potential of hiPSCs to differentiate into cardiomyocytes and can thus be used as a quality control marker of hiPSCs. On the other hand, further experiments will be required to improve hiPSCs in the low differentiation group or to confirm the loss-of-function experiments using inhibitors.

Although our observations revealed that *PF4* was responsible for cardiac differentiation in hiPSCs, the mechanism underlying the regulation of cardiac differentiation remains unknown. As PF4 is a chemokine that suppresses FGF2-dependent ERK phosphorylation[60], we presumed that it may possibly modulate FGF signals in hiPSCs. FGF2 and BMP signals are important for inducing differentiation of cells into cardiomyocytes[61,62], and FGF2/ERK signals suppress BMP signaling-induced Smad1 phosphorylation in ES cells. Indeed, the expression of pluripotency markers and proliferative ability are not altered in FGF knockout stem cells, which rather show difficulty in

differentiating into neural cells[63,64]. Therefore, suppression of the FGF2 signal via PF4 may play a key role in directing cardiac differentiation in hiPSCs.

In addition, we observed that WNT activation reduced *PF4* expression in hiPSCs, which suggested a role of WNT signaling in the regulation of cardiac differentiation via *PF4*. These results suggest that *PF4* is a novel biomarker for selecting hiPSC lines that are likely to differentiate into cardiomyocytes, which can be used in regenerative therapy and drug screening. In the present study, we focused on the differentiation process of hiPSC lines. However, the gene expression patterns of hESC lines and the mechanism for the maintenance of their pluripotency are known to be similar to those of hiPSC lines. Thus, it seems likely that our biomarker could be applicable to hESC lines.

In summary, hiPSCs with high cardiomyogenic potential showed high expression levels of *PF4*, suggesting that this gene may be used as a biomarker for the selection of hiPSCs that are suitable for generating cardiomyocytes. In the current study, we demonstrated the feasibility of using our approach to efficiently select critical biomarkers in hiPSC lines, which can then be used for screening the cardiac differentiation potential of hiPSCs. In addition to the identification of *PF4*, this new approach will facilitate the identification of novel biomarkers for the differentiation potential of hiPSCs. This new strategy may be beneficial in selecting novel biomarkers and in eliminating false-positives.

Conclusion

Differential global gene expression of hiPSC lines with high and low differentiation capacity showed that cardiac differentiation potential may be predicted by measuring the expression of *PF4*. The identification of *PF4* as a cardiomyogenic differentiation potential marker of hiPSC lines may be helpful in selecting hiPSCs lines for regenerative therapy and drug screening.

Materials and Methods

Cell culture and differentiation of hiPSCs.

We used commercially available hiPSC lines as listed in Table 1 and Table 5. To differentiate hiPSCs into cardiomyocytes, we modified a previously described protocol[28-30]. Undifferentiated hiPSCs in the training set were cultured on mouse embryo fibroblast (MEF) feeders (ReproCell, Tokyo, Japan) treated with mitomycin C in primate embryonic stem (ES) medium (ReproCell) supplemented with 5 ng/mL basic fibroblast growth factor (bFGF; ReproCell). The hiPSC lines were passaged twice a week on MEF feeders. Feeder-free hiPSCs in the test set were cultured on iMatrix-511 (Nippi, Inc., Tokyo, Japan) using StemFit (Ajinomoto Co., Inc., Tokyo, Japan).

When the hiPSCs were differentiated into cardiomyocytes, the MEF feeders were removed using

human ES/iPS dissociation solution (ReproCell). Cell colonies were then dissociated into single cells using Accumax (Innovation Cell Technologies, San Diego, CA, USA). Feeder-free hiPSCs were dissociated into single cells using Accutase (Innovation Cell Technologies). Embryoid bodies (EBs) were generated in EZSPHERE (AGC Techno Glass, Shizuoka, Japan) with StemPro-34 medium (Gibco, Grand Island, NY, USA) containing 50 μg/mL ascorbic acid (Wako, Osaka, Japan), 2 mM L-glutamine (Gibco), and 400 μM L-thioglycerol (Sigma-Aldrich, Saint Louis, MO, USA) with 0.5 ng/mL bone morphogenic protein 4 (BMP-4; R&D Systems) and 10 μM Y-27632 (Wako).

The next day, 5 mL StemPro-34 medium containing 10 ng/mL BMP-4 (R&D Systems), 6 ng/mL bFGF, and 6 ng/mL or 12 ng/mL Activin A (R & D Systems) was added into the EZSPHERE. On day 4, the EBs were transferred to a low-attachment plate (Corning, NY, USA), and the medium was changed to StemPro-34 medium containing 4 μM IWR-1 (Sigma-Aldrich). After day 6, the medium was changed to StemPro-34 medium containing vascular endothelial growth factor (VEGF; R & D Systems; 5 ng/mL) and bFGF (10 ng/mL) after every 2 days. The EBs were differentiated under hypoxic conditions (5% O_2 and 5% CO_2) at 37°C in a hypoxic incubator (Thermo Fisher Scientific, Waltham, MA, USA). The differentiated EBs were collected on days 14–18.

To study the effect of cardiac differentiation, hiPSCs were preincubated in a medium containing 1% dimethyl sulphoxide (DMSO; Wako), 10 μM IWR-1, 5 μM IWP-2 (Sigma-Aldrich), 6 μM

CHIR99021 (StemCell Technologies, Vancouver, Canada) or 10 μM Mitoblock-6 (Focus Biomolecules, Plymouth Meeting, PA, USA) for overnight, or in a medium containing 1 μM Recombinant Human PF4 (Peprotech, New Jersey, USA) or 1% DMSO for 2 days. Following preincubation, the hiPSCs were differentiated into cardiomyocytes as described above.

Affymetrix miRNA labelling, array hybridisation, and data pre-processing.

Undifferentiated hiPSCs were maintained on Matrigel (Corning, NY, USA) -coated dishes in mTeSR1 medium (StemCell Technologies, Vancouver, Canada). Total RNA was isolated from hiPSC lines using a miRNeasy mini kit (Qiagen, Hilden, Germany) and treated with DNase I according to the manufacturer's instructions. Total RNA containing low molecular weight RNA (from six hiPSC lines of the training set, n = 6 for each line) were labelled using the FlashTag Biotin HSR RNA labelling kit (Affymetrix, Sunnyvale, CA, USA), according to the manufacturer's instructions. Labelled RNA was processed for microarray hybridisation to miRNA 3.0 array (Affymetrix). An Affymetrix GeneChip fluidics station was used to perform streptavidin/phycoerythrin staining. The hybridisation signals on the microarray were scanned using a GeneChip Scanner 3000 (Affymetrix), and normalisation was performed using the miRNA array RMA+DABG analysis and the Expression Console software (Affymetrix). The National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) accession number for the miRNA array data is GSE117739.

GeneChip profiling and biostatistical analysis.

Undifferentiated hiPSCs were maintained on Matrigel (Corning)-coated dishes in mTeSR1 medium (StemCell Technologies). Total RNA was isolated from hiPSC lines using an RNeasy mini kit (Qiagen) and treated with DNase I according to the manufacturer's instructions. RNA samples (from six hiPSCs lines of the training set, n = 6 for each line) were converted into biotinylated cRNA using GeneChip 3' IVT PLUS reagent kit (Affymetrix). Labelled RNA was processed for microarray hybridisation to Human Genome U133 Plus 2.0 GeneChips (Affymetrix). An Affymetrix GeneChip fluidics station was used to perform streptavidin/phycoerythrin staining. The hybridisation signals on the microarray were scanned using a GeneChip Scanner 3000 (Affymetrix) and analysed using Expression console software (Affymetrix). Normalisation was performed after global scaling, with the arrays scaled to a trimmed average intensity of 500 after excluding the 2% probe sets with the highest and lowest values. The hybridisation experiments were performed with six samples of each hiPSC line. The NCBI GEO accession number for the microarray data is GSE88963. To identify the probe sets related to cardiac differentiation of hiPSCs, paired sample *t*-tests were conducted to compare high and low differentiation groups based on PCA ranking. Statistical significance was defined as $p < 0.01$ as determined by the Student's *t*-test with a fold change > 2.

CAGE profiling.

Undifferentiated hiPSCs were maintained on MEF feeder (ReproCell)-coated dishes in primate ES medium (ReproCell). Eleven hiPSC lines of a training set, including four hiPSCs with different passages for 253G1 and two for 201B7 and R-12A were used for total RNA extraction and purification using the TRIzol tissue kit (Invitrogen) according to the manufacturer's protocol. RNA quality was assessed using a NanoDrop spectrophotometer and the Agilent total RNA nano kit. The standard CAGE protocol was adapted for sequencing on an Illumina platform[49,65], the details of which are provided in the supplemental experimental procedures (CAGE profiling and data processing).

Data analysis.

Data are expressed as mean ± standard error of mean (SEM). Statistical significance was determined using a two-tailed Student's *t* test or analysis of variance (ANOVA), as appropriate. Differences between groups were considered statistically significant at p-values < 0.05.

Flow cytometry.

After differentiation into cardiomyocytes, the EBs were dissociated with TrypLE Select (Gibco) for 5–10 min and neutralised with DMEM (Nacalai Tesque, Kyoto, Japan) containing 10% foetal bovine serum (FBS; Sigma-Aldrich). Thereafter, the cells were fixed with CytoFix fixation buffer (Becton Dickinson, East Rutherford, NJ, USA) for 30 min at 4°C. Then, the fixed cells were stained with the

cardiac troponin T (cTnt) antibody CT3 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) or normal mouse IgG (Santa Cruz Biotechnology, Inc.) in perm/wash buffer (Becton Dickinson). Alexa Fluor 488-conjugated goat anti-mouse IgG (Thermo Fisher Scientific) was used as a secondary antibody. The stained cells were analysed using a flow cytometer (Becton Dickinson). Data were analysed using a FACS Canto II system (Becton Dickinson).

Beating analysis.

Differentiated EBs were monitored using a motion analysis system (SI8000 View; Sony, Tokyo, Japan). The proportion of beating EBs and the cross-sectional area and diameter of EBs in the video images were calculated, using SI8000C analyser (Sony) and a fluorescence microscope (BZX-710; Keyence, Osaka, Japan).

Immunofluorescence staining.

For immunofluorescence staining, cells were fixed with 4% paraformaldehyde for 10 min at room temperature and washed with phosphate-buffered saline (PBS). After incubation with blocking solution containing 10% normal bovine serum albumin (Sigma-Aldrich) for 1 h at room temperature, the cells were permeabilised by incubation with 0.25% Triton X-100 for 10 min at room temperature. The cells were then incubated overnight with primary antibodies for cTnt (Santa Cruz Biotechnology), Nkx2.5 (Abcam, Cambridge, United Kingdom), alpha-smooth muscle actin (αSMA,

Abcam), and vimentin (Abcam) at 4°C. Anti-rabbit or anti-mouse IgG secondary antibodies conjugated with fluorescein such as Alexa Fluor 488 or Alexa Fluor 555 (Thermo Fisher Scientific) were used for visualisation of the antigens of interest. Nuclei were counterstained with Hoechst 33258 (Dojindo, Kumamoto, Japan). Immunofluorescent images were examined under a confocal microscope (FV1200 or SD-OSR, Olympus, Tokyo, Japan).

Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis.

Undifferentiated hiPSCs were lysed using QIAzol lysis reagent (Qiagen, Hilden, Germany). RNA was extracted using a miRNeasy mini kit (Qiagen) according to the manufacturer's protocol. cDNA was synthesised using a SuperScript VILO cDNA synthesis kit (Thermo Fisher Scientific). qPCR was performed using a ViiA 7 real-time PCR system (Applied Biosystems, Carlsbad, CA, USA), PCR primers, SYBR Green PCR master mix (Applied Biosystems) or Taqman probe, and Taqman Gene Expression master mix (Applied Biosystems) as described in Table 6. Relative expression analysis was performed using the expression level of *GAPDH*, a housekeeping gene, as a reference.

Heat map and principal component analysis (PCA).

A heat map was generated based on the standardised data using Excel 2016 (Microsoft, Redmond, WA, USA), representing the cardiac marker genes *TNT2*, *NKX2.5*, *GATA4*, *MYL2*, *MYH6*, and *MYH7*, and cardiac maturation marker genes *SCN5A*, *RYR2*, *PPARGC1*, *KCNJ2, HCN4*, *CACNA1C*,

and *ATP2A2* expressed in the differentiated hiPSC lines. PCA was also performed, and the first principal component scores were calculated for the cardiac marker genes on days 9 and 17 for six hiPSC lines from the training set using the JMP 12 software (SAS Institute, Cary, NC, USA).

Table 6 List of primers used in the current study.

Oligonucleotide primers:

Gene	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')
NKX2.5	ACCTCAACAGCTCCCTGACTC	ATAATCGCCGCCACAAACTCTCC
TNNT2	TTCACCAAAGATCTGCTCCTCGCT	TTATTACTGGTGTGGAGTGGGTGTGG
MYL2	TGTCCCTACCTTGTCTGTTAGCCA	ATTGGAACATGGCCTCTGGATGGA
MYH6	TCAGCTGGAGGCCAAAAGTAAAGGA	TTCTTGAGCTCTGAGCACTCGTCT
MYH7	TCGTGCCTGATGACAAACAGGAGT	ATACTCGGTCTCGGCAGTGACTTT
GATA4	AGGCCTCTTGCAATGCGGA	CTGGTGGTGGCGTTGCTGG
LIN28	CACGGTGCGGGCATCTG	CCTTCCATGTGCAGCTTACTC
GAPDH	CAATGACCCCTTCATTGACC	TTGATTTTGGAGGGATCTCG
NANOG	AAAGAATCTTCACCTATGCC	GAAGGAAGAGGAGAGACAGT
SOX1	CACAACTCGGAGATCAGCAA	GGTACTTGTAATCCGGGTGC
PAX6	GTCCATCTTTGCTTGGGAAA	TAGCCAGGTTGCGAAGAACT
ZIC1	CTGGCTGTGGCAAGGTCTTC	CAGCCCTCAAACTCGCACTT
FLK1	TGATCGGAAATGACACTGGA	CACGACTCCATGTTGGTCAC
PDGFR-A	ACAGGTTGGTGTGGGTTCAT	CTGCATCTTCCAAAGCATCA
BRACHYURY	AATTGGTCCAGCCTTGGAAT	CGTTGCTCACAGACCACA
GOOSECOID	GAGGAGAAAGTGGAGGTCTG	CTCTGATGAGGACCGCTTCTG
SOX17	CTCTGCCTCCTCCACGAA	CAGAATCCAGACCTGCACAA
HNF3	GGAGCGGTGAAGATGGAA	TACGTGTTCATGCCGTTCAT
AMN	GACTCTGACCGCTTCTCCTG	CACTAGGCGGAAAGAAGACG
SOX7	ACGCCGAGCTCAGCAAGAT	TCCACGTACGGCCTCTTCTG
OCT-4	CAATTTGCCAAGCTCCTGAAG	AAAGCGGCAGATGGTCGTT

Taqman primers:

Gene	Assay reference
ANKRD1	Hs00173317_m1
ATP2A2	Hs00544877_m1
BCOR	Hs00372378_m1
C7orf50	Hs00260594_m1
CACNA1C	Hs00167681_m1
CHCHD2	Hs00853326_g1
FGF17	Hs00182599_m1
FOXQ1	Hs00536425_s1
GAPDH	Hs02758991_g1
GATA6	Hs00232018_m1
GLIPR1	Hs01564146_g1
HCN4	Hs00975492_m1
IGFBP5	Hs00181213_m1
KCNJ2	Hs01876357_s1
KDM6A	Hs00253500_m1
MYL4	Hs04187281_m1
PF4	Hs00427220_g1
PLCB1	Hs01001930_m1
POMZP3	Hs02341150_m1
PPARGC1	Hs01016719_m1
PTGR1	Hs00400932_m1
RBMX	Hs00953944_g1
RC3H1	Hs01367694_m1
RIPK1	Hs01041869_m1
RYR2	Hs00181461_m1
SCN5A	Hs00165693_m1
SKIL	Hs01045418_m1
TMEM64	Hs01595139_m1
WNT3	Hs00902257_m1
ZNF229	Hs00970996_g1

Pathway analyses.

We determined the intracellular location and biological function of a group of genes with significantly changed expression using Qiagen's Ingenuity Pathway Analysis (IPA) software. Enriched human homologues were classified into functional categories based on either biological function or canonical pathways. A right-tailed Fisher's exact test was used to calculate *p*-values to determine the probability of each enriched human homologue associated with the dataset based on chance alone. We selected human homologues that had been functionally annotated as "differentiation" by IPA and further confirmed their roles based on functional descriptions in the NCBI and GeneCard databases.

CAGE profiling and data processing.

cDNA was synthesised from total RNA using random primers in the presence of trehalose and sorbitol under high temperature. The ribose diols in the cap structure and the 3' end of the RNA were oxidised and then biotinylated by reacting the generated aldehyde groups with biotin hydrazide (Vector Laboratories, Burlingame, CA, USA). Single-stranded RNAs were digested with RNase ONE ribonuclease (Promega, Fitchburg, WI, USA). The biotinylated RNA/cDNA was selected using Dynabeads M-270 streptavidin (Thermo Fisher Scientific). After capturing on the beads, the cDNA was released into the supernatant by heat denaturation. The supernatant was subjected to RNase

ONE/H digestion, followed by purification using AMPure XP (Bio Rad, Hercules, CA, USA) and adaptor ligation to both ends of the cDNA. A double-stranded cDNA library was created using DeepVent (exo-) DNA polymerase. The CAGE cDNA libraries were sequenced using an Illumina HiSeq 2500 sequencer (Illumina). The data have been submitted to DDBJ Read Archive (DRA) under accession number DRA007185. The obtained reads were aligned with the reference human genome (GRCh37) using Burrows-Wheeler Alignment tool (BWA)[66], and low quality alignments with a mapping quality of 20 or less were discarded. The 5′-ends of the remaining alignments were counted based on the robust set of CAGE peaks defined in a previous study[67] and submitted to the Functional Annotation of Mammalian Genome 5 (FANTOM5) web resource[68]. The read counts were normalised as counts per million based on the normalisation (size) factor calculated using the RLE (relative log ratio) method[69]. Expression analysis was conducted using edgeR[70].

References

1 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, https://doi.org/10.1016/j.cell.2006.07.024 (2006).

2 Kawamura, M. *et al.* Feasibility, safety, and therapeutic efficacy of human induced pluripotent stem cell-derived cardiomyocyte sheets in a porcine ischemic cardiomyopathy model. *Circulation* **126**, S29-S37, https://doi.org/10.1161/circulationaha.111.084343 (2012).

3 Miki, K. *et al.* 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, https://doi.org/10.5966/sctm.2011-0038 (2012).

4 Miyagawa, S. & Sawa, Y. From bench to bedside, work in cell-based myocardial regeneration therapy. *J. Biomed. Sci. Eng.* **7**, 86, https://doi.org/10.4236/jbise.2014.72012 (2014).

5 Takeda, M. *et al.* Development of in vitro drug-induced cardiotoxicity assay by using three-dimensional cardiac tissues derived from human induced pluripotent stem cells. *Tissue Eng. Part C: Methods* **24**, 56-67, https://doi.org/10.1089/ten.tec.2017.0247 (2018).

6 Liang, P. *et al.* Drug screening using a library of human induced pluripotent stem cell-derived cardiomyocytes reveals disease specific patterns of cardiotoxicity. *Circulation* **127**, 1677-1691, https://doi.org/10.1161/circulationaha.113.001883 (2013).

7 Ishida, M. *et al.* Transplantation of human induced pluripotent stem cell-derived cardiomyocytes is superior to somatic stem cell therapy for restoring cardiac function and oxygen consumption in a porcine model of myocardial infarction. *Transplantation*, https://doi.org/10.1097/tp.0000000000002384 (2018).

8 Yoshida, S. *et al.* Maturation of human induced pluripotent stem cell-derived cardiomyocytes by soluble factors from human mesenchymal stem cells. *Mol Ther*. https://doi.org/10.1016/j.ymthe.2018.08.012 (2018).

9 Sun, N. *et al.* Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Sci. Transl. Med.* **4**, 130ra147, https://doi.org/10.1126/scitranslmed.3003552 (2012).

10 Wu, H. *et al.* Epigenetic regulation of phosphodiesterases 2A and 3A underlies compromised beta-adrenergic signaling in an iPSC model of dilated cardiomyopathy. *Cell Stem Cell* **17**, 89-100, https://doi.org/10.1016/j.stem.2015.04.020 (2015).

11 Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* **26**, 313-315, https://doi.org/10.1038/nbt1383 (2008).

12 Bock, C. *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439-452, https://doi.org/10.1016/j.cell.2010.12.032 (2011).

13 Lee, J. -H. *et al.* Somatic transcriptome priming gates lineage-specific differentiation potential of human-induced pluripotent stem cell states. *Nat. Commun.* **5**, 5605, https://doi.org/10.1038/ncomms6605 (2014).

14 Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285-290, https://doi.org/10.1038/nature09342 (2010).

15 Hartjes, K. A. *et al.* Selection via pluripotency-related transcriptional screen minimizes the influence of somatic origin on iPSC differentiation propensity. *Stem Cells* **32**, 2350-2359, https://doi.org/10.1002/stem.1734 (2014).

16 Hu, S. *et al.* Effects of cellular origin on differentiation of human induced pluripotent stem cell-derived endothelial cells. *JCI Insight* **1**, https://doi.org/10.1172/jci.insight.85558 (2016).

17 Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68-73, https://doi.org/10.1038/nature09798 (2011).

18 Nishizawa, M. *et al.* Epigenetic variation between human induced pluripotent stem cell lines is an indicator of differentiation capacity. *Cell Stem Cell* **19**, 341-354, https://doi.org/10.1016/j.stem.2016.06.019 (2016).

19 Zhu, L. *et al.* The mitochondrial protein CHCHD2 primes the differentiation potential of human induced pluripotent stem cells to neuroectodermal lineages. *J. Cell Biol.* **215**, 187-202, https://doi.org/10.1083/jcb.201601061 (2016).

20 Jiang, W., Zhang, D., Bursac, N. & Zhang, Y. WNT3 is a biomarker capable of predicting the definitive endoderm differentiation potential of hESCs. *Stem Cell Rep.* **1**, 46-52, https://doi.org/10.1016/j.stemcr.2013.03.003 (2013).

21 Koyanagi-Aoi, M. *et al.* Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc. Natl. Acad. Sci.***110**, 20569-20574,

https://doi.org/10.1073/pnas.1319061110 (2013).

22 Iwashita, H. *et al.* Secreted cerberus1 as a marker for quantification of definitive endoderm differentiation of the pluripotent stem cells. *PloS One* **8**, e64291, https://doi.org/10.1371/journal.pone.0064291 (2013).

23 Lee, J.-H. *et al.* Lineage-specific differentiation is influenced by state of human pluripotency. *Cell Rep.* **19**, 20-35, https://doi.org/10.1016/j.celrep.2017.03.036 (2017).

24 Yamamoto, T. *et al.* Differentiation potential of pluripotent stem cells correlates to the level of CHD7. *Sci. Rep.* **8**, 241, https://doi.org/10.1038/s41598-017-18439-y (2018).

25 Yasuda, S. *et al.* AW551984: a novel regulator of cardiomyogenesis in pluripotent embryonic cells. *Biochem. J.* **437**, 345-355, https://doi.org/10.1042/bj20110520 (2011).

26 Kuroda, T. *et al.* Identification of a gene encoding slow skeletal muscle troponin T as a novel marker for immortalization of retinal pigment epithelial cells. *Sci. Rep.* **7**, 8163, https://doi.org/10.1038/s41598-017-08014-w (2017).

27 Martinez-Fernandez, A., Li, X., Hartjes, K. A., Terzic, A. & Nelson, T. J. Natural cardiogenesis-based template predicts cardiogenic potential of induced pluripotent stem cell lines. *Circ. Cardiovasc. Genet.* **6**, 462-471, https://doi.org/10.1161/circgenetics.113.000045 (2013).

28 Matsuura, K. *et al.* Fabrication of mouse embryonic stem cell-derived layered cardiac cell sheets using a bioreactor culture system. *PLoS One* **7**, e52176, https://doi.org/10.1371/journal.pone.0052176 (2012).

29 Kawamura, T. *et al.* Cardiomyocytes derived from MHC-homozygous induced pluripotent

stem cells exhibit reduced allogeneic immunogenicity in MHC-matched non-human primates. *Stem Cell Rep.* **6**, 312-320, https://doi.org/10.1016/j.stemcr.2016.01.012 (2016).

30 Miki, K. *et al.* Efficient detection and purification of cell populations using synthetic microRNA switches. *Cell Stem Cell* **16**, 699-711, https://doi.org/10.1016/j.stem.2015.04.005 (2015).

31 Kurosawa, H. Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. *J. Biosci. Bioeng.* **103**, 389-398, https://doi.org/10.1263/jbb.103.389 (2007).

32 Kattman, S. J. *et al.* Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* **8**, 228-240, https://doi.org/10.1016/j.stem.2010.12.008 (2011).

33 Matsuura, K. *et al.* Creation of human cardiac cell sheets using pluripotent stem cells. *Biochem. Biophys. Res. Commun.* **425**, 321-327, https://doi.org/10.1016/j.bbrc.2012.07.089 (2012).

34 Iseoka, H. *et al.* Pivotal role of non-cardiomyocytes in electromechanical and therapeutic potential of induced pluripotent stem cell-derived engineered cardiac tissue. *Tissue Eng. Part A* **24**, 287-300, https://doi.org/10.1089/ten.tea.2016.0535 (2018).

35 Kawamura, A. *et al.* Teratocarcinomas arising from allogeneic induced pluripotent stem cell-derived cardiac tissue constructs provoked host immune rejection in mice. *Sci. Rep.* **6**, 19464, https://doi.org/10.1038/srep19464 (2016).

36 Matsuura, K. *et al.* TRPV-1-mediated elimination of residual iPS cells in bioengineered

cardiac cell sheet tissues. *Sci. Rep.* **6**, 21747, https://doi.org/10.1038/srep21747 (2016).

37 Sougawa, N. *et al.* Immunologic targeting of CD30 eliminates tumourigenic human pluripotent stem cells, allowing safer clinical application of hiPSC-based cell therapy. *Sci. Rep.* **8**, 3726, https://doi.org/10.1038/s41598-018-21923-8 (2018).

38 Yasuda, S. *et al.* Tumorigenicity-associated characteristics of human iPS cell lines. PLoS ONE 13(10): e0205022. https://doi.org/10.1371/journal.pone.0205022 (2018)

39 Kuroda, T. *et al.* Highly sensitive in vitro methods for detection of residual undifferentiated cells in retinal pigment epithelial cells derived from human iPS cells. *PLoS One* **7**, e37342, https://doi.org/10.1371/journal.pone.0037342 (2012).

40 Masumoto, H. *et al.* Pluripotent stem cell-engineered cell sheets reassembled with defined cardiovascular populations ameliorate reduction in infarct heart function through cardiomyocyte-mediated neovascularization. *Stem Cells* **30**, 1196-1205, https://doi.org/10.1002/stem.1089 (2012).

41 Bauwens, C. L. *et al.* Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories. *Stem Cells* **26**, 2300-2310, https://doi.org/10.1634/stemcells.2008-0183 (2008).

42 Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457**, 413, https://doi.org/10.1038/nature07756 (2009).

43 Kaikkonen, M. U., Lam, M. T. & Glass, C. K. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc. Res.* **90**, 430-440, https://doi.org/10.1093/cvr/cvr097 (2011).

44 Kim, H. et al. miR-371-3 expression predicts neural differentiation propensity in human pluripotent stem cells. Cell stem cell 8, 695-706, https://doi.org/10.1016/j.stem.2011.04.002 (2011).

45 Kuppusamy, K. T. *et al.* Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2785-2794, https://doi.org/10.1073/pnas.1424042112 (2015).

46 Ivey, K. N. *et al.* MicroRNA regulation of cell lineages in mouse and human embryonic stem cells. *Cell Stem Cell* **2**, 219-229, https://doi.org/10.1016/j.stem.2008.01.016 (2008).

47 Yoshida, E. *et al.* Promoter-level transcriptome in primary lesions of endometrial cancer identified biomarkers associated with lymph node metastasis. *Sci. Rep.* **7**, 14160, https://doi.org/10.1038/s41598-017-14418-5 (2017).

48 Yoshihara, M. *et al.* Discovery of molecular markers to discriminate corneal endothelial cells in the human body. *PloS One* **10**, e0117581, https://doi.org/10.1371/journal.pone.0117581 (2015).

49 Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci.* **100**, 15776-15781, https://doi.org/10.1073/pnas.2136655100 (2003).

50 Ojala, M. *et al.* Culture conditions affect cardiac differentiation potential of human pluripotent stem cells. *PLoS One* **7**, e48659, https://doi.org/10.1371/journal.pone.0048659 (2012).

51 Vandercappellen, J., Van Damme, J. & Struyf, S. The role of the CXC chemokines platelet

factor-4 (CXCL4/PF-4) and its variant (CXCL4L1/PF-4var) in inflammation, angiogenesis and cancer. *Cytokine Growth Factor Rev.* **22**, 1-18, https://doi.org/10.1016/j.cytogfr.2010.10.011 (2011).

52 Kim, H. *et al.* Tmem64 modulates calcium signaling during RANKL-mediated osteoclast differentiation. *Cell Metab.* **17**, 249-260, https://doi.org/10.1016/j.cmet.2013.01.002 (2013).

53 Minami, I. *et al.* A small molecule that promotes cardiac differentiation of human pluripotent stem cells under defined, cytokine- and xeno-free conditions. *Cell Rep.* **2**, 1448-1460, https://doi.org/10.1016/j.celrep.2012.09.015 (2012).

54 Sugimoto, M. *et al.* A simple and robust method for establishing homogeneous mouse epiblast stem cell lines by wnt inhibition. *Stem Cell Rep.* **4**, 744-757, https://doi.org/10.1016/j.stemcr.2015.02.014 (2015).

55 Fox, J. M. et al. CXCL4/Platelet Factor 4 is an agonist of CCR1 and drives human monocyte migration. Sci Rep 8, 9466, http://dx.doi.org/10.1038/s41598-018-27710-9 (2018).

56 Sanchez-Freire, V. *et al.* Effect of human donor cell source on differentiation and function of cardiac induced pluripotent stem cells. *J. Am. Coll. Cardiol.* **64**, 436-448, https://doi.org/10.1016/j.jacc.2014.04.056 (2014).

57 Takeda, M. *et al.* Identification of Cardiomyocyte-Fated Progenitors from Human-Induced Pluripotent Stem Cells Marked with CD82. *Cell Rep* **22**, 546-556, https://doi.org/10.1016/j.celrep.2017.12.057 (2018).

58 van Bon, L. *et al.* Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis.

New Engl. J. Med. **370**, 433-443, https://doi.org/10.1056/nejmc1402401 (2014).

59 Cheng, S. H. *et al.* 4q loss is potentially an important genetic event in MM tumorigenesis: identification of a tumor suppressor gene regulated by promoter methylation at 4q13. 3, platelet factor 4. *Blood* **109**, 2089-2099, https://doi.org/10.1182/blood-2006-04-018770 (2007).

60 Lord, M. S., Cheng, B., Farrugia, B. L., McCarthy, S. & Whitelock, J. M. Platelet factor 4 binds to vascular proteoglycans and controls both growth factor activities and platelet activation. *J. Biol. Chem.* **292**, 4054-4063, https://doi.org/10.1074/jbc.m116.760660 (2017).

61 Kawai, T. *et al.* Efficient cardiomyogenic differentiation of embryonic stem cell by fibroblast growth factor 2 and bone morphogenetic protein 2. *Circulation J.* **68**, 691-702, https://doi.org/10.1253/circj.68.691 (2004).

62 Rosenblatt-Velin, N., Lepore, M. G., Cartoni, C., Beermann, F. & Pedrazzini, T. FGF-2 controls the differentiation of resident cardiac precursors into functional cardiomyocytes. *J. Clin. Investig.* **115**, 1724-1733, https://doi.org/10.1172/jci23418 (2005).

63 Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* **137**, 3351-3360, https://doi.org/10.1242/dev.050146 (2010).

64 Stavridis, M. P., Collins, B. J. & Storey, K. G. Retinoic acid orchestrates fibroblast growth factor signalling to drive embryonic stem cell differentiation. *Development* **137**, 881-890, https://doi.org/10.1242/dev.043117 (2010).

65 Murata, M. *et al.* Detecting expressed genes using CAGE. Methods Mol Biol. **1164**, 67-85, https://doi.org/10.1007/978-1-4939-0805-9_7 (2014).

66 Li, H. & Durbin, R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).

67 Forrest, A.R., *et al*. A promoter-level mammalian expression atlas. *Nature* **507**, 462, https://doi.org/10.1038/nature13182 (2014).

68 Lizio, M., *et al*. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22, https://doi.org/10.1186/s13059-014-0560-6 (2015).

69 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106, https://doi.org/10.1186/gb-2010-11-10-r106 (2010).

70 Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, https://doi.org/10.1093/bioinformatics/btp616 (2010).

Acknowledgements

本研究の遂行ならびに本論文の執筆にあたり、終始適切な御指導、御鞭撻を賜りました大阪大学大学院薬学研究科　佐藤　陽治　招聘教授　（国立医薬品食品衛生研究所　再生・細胞医療製品部　部長をご兼任）に心より御礼を申し上げます。再生医療の実用化に向けてレギュラトリーサイエンスを推進する佐藤　陽治先生の姿に魅了され、大阪大学大学院薬学研究科　博士課程に進学しました。社会人として仕事を継続しながらも連携大学院生として研究することを快く受け入れて頂き、さらに様々な分野の専門家の方々と知り合う機会を与えて頂き、専門知識を学ぶことができたことが学位取得に繋がったものと深く感謝しております。御指導いただいた経験を糧に今後も精進して参ります。

本研究の遂行ならびに本論文の執筆にあたり、有益な御助言と御指導、ならびに、研究の立ち上げから終始絶え間ない議論を賜りました国立医薬品食品衛生研究所　再生・細胞医療製品部　安田　智　室長、三浦　巧　室長、黒田　拓也　主任研究官に深く感謝の意を表します。研究の進め方に悩んだときの国立医薬品食品衛生研究所の皆様からの適切なアドバイスに都度助けられて、論文を完成させることができました。博士後期課程でこのような素晴らしい研究指導を賜りましたことに喜びを実感すると共に、皆様に心より御礼申し上げます。また、村岡ひとみ　技術研究員には、研究を遂行するにあたり甚大なるご協力を頂き、厚く御礼申し上げます。

本研究を行うにあたり、博士後期課程への進学を後押していただき、学会発表や論文執筆のご指導を頂きました大阪大学大学院医学系研究科の澤　芳樹　教授、宮川　繁　特任教授に深く感謝致します。ご多忙にもかかわらず、心臓血管外科　研究室に出向していた私に懇切丁寧にご指導いただき、研究することの意義を学び、博士後期課程への進学を志すことができました。本研究に対して多大なご協力や御助言を賜りました齋藤　充弘　特任准教授、吉田　昇平　先生、伊東　絵望子　氏、寒川　延子　氏、武田　真希　氏、原田　明希摩　氏、吉用　賢治　氏、石川　烈　氏、井出　由利　氏、永楽　星子　氏ならびに日頃の研究活動において、共に励まし、支えてくれた大阪大学大学院医学系研究科　心臓血管外科の皆様に深く感謝致します。

本論文の審査にあたり、様々な御助言を賜りました大阪大学大学院薬学研究科　水口　裕之　教授、櫻井　文教　准教授、藤尾　慈　教授、堤　康央　教授に深く感謝いたします。

本研究を遂行するにあたり、共同研究として iPS 細胞株の CAGE 解析を行い、数多くの御助言、御協力を賜りました理化学研究所　林崎　良英　先生、河合　純　先生、川路　英哉　先生、伊藤　昌可　先生、鈴木　直子氏、山本　由美子氏ならびに RIKEN GeNAS チームの皆様に感謝致します。

会社の業務を継続しながら博士後期課程への進学や本研究を行うことに快くご承認頂き、全面的なサポートをして頂いたテルモ株式会社　鮫島　正　執行役員、大河原　順一　主任研究員、ならびに業務を通じて技術的な御指導や御協力を頂いた伊勢岡　弘子　氏、大山　賢二　氏、松田　勇　氏に深く感謝致します。

最後に、研究に深い理解をしてくれ、常日頃より献身的に支えて頂いた家族に心より感謝致します。