


# SCIENTIFIC REPORTS



OPEN

## Combining membrane proteomics and computational three-way pathway analysis revealed signalling pathways preferentially regulated in human iPSCs and human ESCs

Wei-Sheng Tien<sup>1,2</sup>, Pei-Mien Chen<sup>4</sup>, Ching-Yu Chuang<sup>5</sup>, Shook-Mun Lui<sup>1</sup>, Hung-Chih Kuo<sup>3</sup>, Yu-Ju Chen<sup>4</sup> & Kun-Pin Wu<sup>1</sup> 

Owing to the clinical potential of human induced pluripotent stem cells (hiPSCs) in regenerative medicine, a thorough examination of the similarities and differences between hiPSCs and human embryonic stem cells (hESCs) has become indispensable. Moreover, as the important roles of membrane proteins in biological signalling, functional analyses of membrane proteome are therefore promising. In this study, a pathway analysis by the bioinformatics tool GSEA was first performed to identify significant pathways associated with the three comparative membrane proteomics experiments: hiPSCs versus precursor human foreskin fibroblasts (HFF), hESCs versus precursor HFF, and hiPSCs versus hESCs. A following three-way pathway comparison was conducted to identify the differentially regulated pathways that may contribute to the differences between hiPSCs and hESCs. Our results revealed that pathways related to oxidative phosphorylation and focal adhesion may undergo incomplete regulations during the reprogramming process. This hypothesis was supported by another public proteomics dataset to a certain degree. The identified pathways and their core enriched proteins could serve as the starting point to explore the possible ways to make hiPSCs closer to hESCs.

The breakthroughs in induced pluripotent stem cell (iPSC) technology shed light on the regenerative medicine; by introducing a panel of transcription factors, somatic cells can be converted into iPSCs that exhibit pluripotency and self-renewal capacity similar to embryonic stem cells (ESCs)<sup>1</sup>. The use of human embryo in academic research or clinical practice is quite limited due to the ethical controversies, iPSCs obtained by reprogramming from somatic cells therefore display superior potential for therapeutic applications<sup>2</sup>. How close the iPSCs to their target ESCs, however, has not been thoroughly elucidated. Although many iPSCs were reported remarkably similar to ESCs, iPSCs and ESCs with different gene expression profiles were constantly found in different studies<sup>3,4</sup>. Thus, prior to the clinical application of iPSCs, identifying the similarities and differences between the two types of pluripotent cells and the biological events related to these subtle differences will be valuable.

Comparative proteomics analyses based on mass spectrometry (MS) have been applied to gain insights into the similarities and differences between iPSCs and ESCs<sup>5,6</sup>. Meanwhile, the behaviour of pluripotent cells has been shown to be tightly controlled by extrinsic and intrinsic factors<sup>7</sup>, the plasma membrane-associated proteins that serve as interfaces between the cell and the surrounding environment are of importance in cellular signalling process and therefore are regarded as biomarker candidates<sup>8,9</sup>. Thus, the comparison of membrane proteome may

<sup>1</sup>Institute of Biomedical Informatics, National Yang Ming University, Taipei, 112, Taiwan. <sup>2</sup>Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, 115, Taiwan. <sup>3</sup>Institute of Cellular and Organismic Biology, Academia Sinica, Taipei, 115, Taiwan. <sup>4</sup>Institute of Chemistry, Academia Sinica, Taipei, 115, Taiwan. <sup>5</sup>Genomics Research Center, Academia Sinica, Taipei, 115, Taiwan. Correspondence and requests for materials should be addressed to K.-P.W. (email: [kpwu@ym.edu.tw](mailto:kpwu@ym.edu.tw))

Received: 24 May 2017  
Accepted: 25 October 2017  
Published online: 08 November 2017

help us identify pathways that translate external factors into internal signals for self-renewal and differentiation induction<sup>10</sup>. Although membrane proteins are known to serve as surface specific markers and of clinical potential<sup>11</sup>, the analyses of membrane proteins from whole cell lysates are technically difficult due to the relatively low abundance and hydrophobic properties of membrane proteins<sup>12</sup>. The studies of iPSCs and ESCs exclusively on membrane proteome therefore became promising but challenging. Nevertheless, currently, no pairwise comparisons were specifically performed on human membrane proteome of ESCs, iPSCs, and their precursor somatic cells to the best of our knowledge. In this context, we aimed at conducting a comparative proteomics analysis of the membrane protein profiles of human ESCs (hESCs), human iPSCs (hiPSCs), and human precursor foreskin fibroblast cell lines (HFF).

There are usually functional analyses following comparative proteomics to gain insights into the underlying biological mechanism behind the proteomic data. Over the past decade, several pathway analysis methods have been developed and applied to numerous comparative omics studies to perform functional analyses<sup>13</sup>. The biomolecules involved in the underlying mechanism behind the long lists of omics profiles are expected to be identified by these pathway analysis tools or databases such as KEGG<sup>14,15</sup>. Pathway analysis approaches have also been used to identify pathways significantly regulated between hiPSCs and their precursor somatic cells, or to assess the differences between hiPSCs and hESCs at the pathway level<sup>5,16,17</sup>. Each of the above two applications of pathway analysis, however, can only identify the differences between hiPSCs and precursor somatic cells, or the differences between hiPSCs and hESCs; we did not know the relations between the two sorts of differences. In other words, we did not know how the reprogramming of hiPSCs affects the differences between hiPSCs and hESCs. Given that hiPSCs are reprogrammed from somatic cells and designated to function as hESCs in an undifferentiated state, a three-way pathway comparison of pathways regulated between hiPSCs and hESCs, hiPSCs and precursor somatic cells, and hESCs and precursor somatic cells might identify pathways that contribute to the subtle differences between hiPSCs and hESCs. Nevertheless, none of the current research performed such a three-way comparison of the pathways regulated among hiPSCs, hESCs, and somatic precursor cells.

In this study, a computational three-way pathway analysis was conducted using the bioinformatics tool Gene Set Enrichment Analysis (GSEA)<sup>18,19</sup>. Three membrane protein profiles of expression ratio of hiPSCs to HFF, hESCs to HFF, and hiPSCs to hESCs were subjected to GSEA to find significantly regulated pathways. On the basis of these identified pathways, our hiPSCs and hESCs samples were found very similar at protein and pathway levels. Moreover, by matching the significant pathways of hiPSCs to hESCs against those of hiPSCs to HFF and hESCs to HFF, certain pathways were hypothesized to undergo an incomplete regulation during the reprogramming process of hiPSCs, which may contribute to the subtle differences between hiPSCs and hESCs. Our results showed that during the reprogramming process of hiPSCs, pathway *focal adhesion* may undergo an incomplete repression and *oxidative phosphorylation* and *TCA cycle and respiratory electron transport* may undergo an incomplete activation.

## Results and Discussion

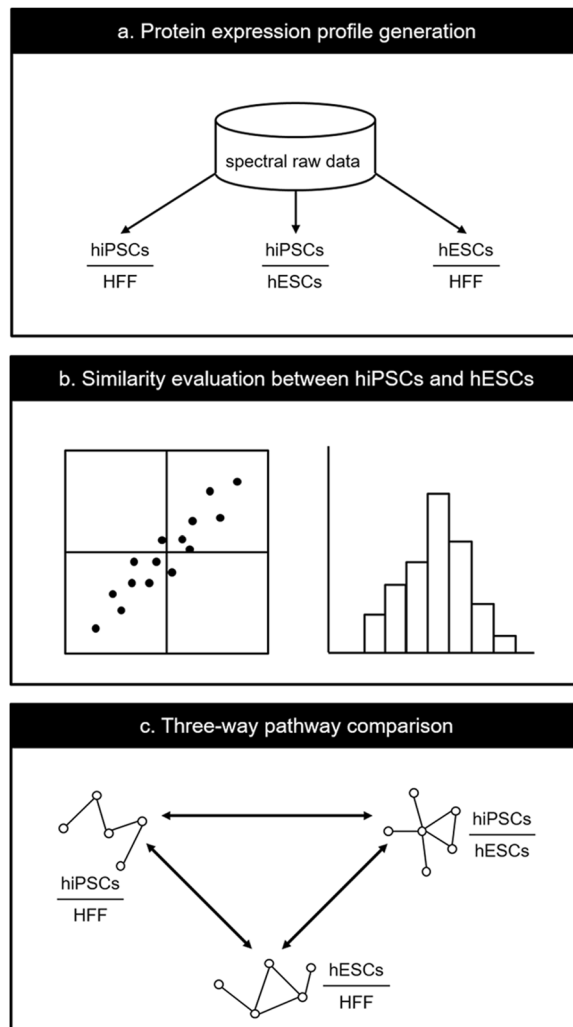
The proteomics analysis in this study was composed of three steps as depicted in Fig. 1. Since our goal aimed at capturing differences that were consistently revealed between hiPSCs and hESCs, the first step of the analysis was to compile our mass spectrometry-based membrane proteomic data into three comparative protein expression profiles: one for hiPSCs versus somatic precursor HFF (hiPSCs/HFF), one for hESCs versus somatic precursor HFF (hESCs/HFF), and one for hiPSCs versus hESCs (hiPSCs/hESCs) (Fig. 1a).

The second step of our analysis was to estimate how close our hiPSC cell lines to hESC cell lines at protein expression level (Fig. 1b); the hiPSC cell lines were expected to be very similar to the hESC cell lines. Since the output of IDEAL-Q, the software we used to quantify proteins, is not absolute expression level but relative expression ratio, we can only compare protein profiles using ratios with respect to the same base (i.e., same denominator). So the similarity between the hiPSC and hESC cell lines was assessed by the following two measures: 1) the Spearman correlation coefficient between the hESCs/HFF and hiPSCs/HFF profiles, and 2) a histogram to show the distribution of the protein expression  $\log_2$  ratios of the hiPSCs/hESCs profile. If most proteins in hiPSCs and hESCs had similar expression levels, there would be a high accumulation of proteins with zero  $\log_2$  ratio in the histogram.

The final step of our analysis was to estimate how close our hiPSC cell lines approach to hESC cell lines at functional level by identifying significant pathways related to the differences among hiPSCs, hESCs, and somatic HFF (Fig. 1c). On the basis of pathways reported by the GSEA, a three-way pathway comparison was performed to answer the following three questions:

1. The similarity between the hiPSCs and hESCs.
2. The pathways related to the differences between the hiPSCs and hESCs.
3. The biological processes during the reprogramming of hiPSCs which may contribute to the differences between hiPSCs and hESCs.

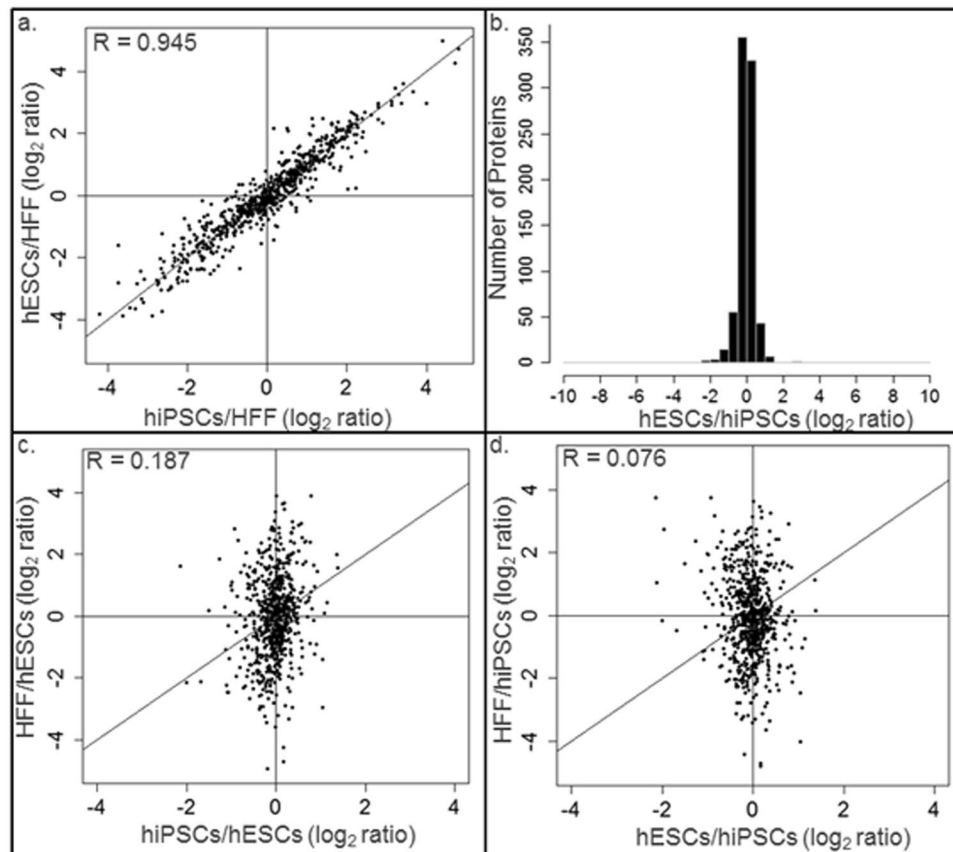
**Expression profiles of hiPSCs/HFF, hESCs/HFF, and hiPSCs/hESCs.** The protein identification was achieved by shotgun-based mass spectrometry analysis. To ensure statistical confidence of protein identification, we performed a decoy database search, obtained by searching against a Mascot-created randomized protein sequence database with identical validation criteria and search parameters, to evaluate the false discovery rate (FDR) of protein identification. It is noted that all the protein identification results were obtained at false discovery rate < 1% at protein level with at least one unique peptide. An hiPSCs/HFF profile of 792 proteins was obtained by taking the intersection of the two protein identification results for the hiPSC CFB46 cell line versus somatic precursor HFF (1100 proteins) and the hiPSC CFB50 cell line versus somatic precursor HFF



**Figure 1.** The workflow of our three-way pathway comparison on the comparative membrane proteomics of hiPSCs, hESCs, and HFF. **(a)** On the basis of spectral raw data, three profiles of protein expression  $\log_2$  ratio, hiPSCs/HFF, hESCs/HFF, and hiPSCs/hESCs, were generated for hiPSCs versus HFF, hESCs versus HFF, and hiPSCs versus hESCs. **(b)** The similarity between hiPSCs and hESCs at protein expression level were evaluated by the Spearman correlation coefficient between hiPSCs/HFF and hESCs/HFF and a histogram describing the distribution of protein  $\log_2$  ratios of hiPSCs/hESCs. **(c)** A three-way pathway comparison was performed on the significant pathways (obtained from GSEA) of hiPSCs/HFF, hESCs/HFF, and hiPSCs/hESCs to identify important pathways that may contribute to the differences between hiPSCs and hESCs during the reprogramming process of hiPSCs.

(1111 proteins). Similarly, an hESCs/HFF profile of 752 proteins was obtained by taking the intersection of the two profiles for the hESC NTU1 cell line versus somatic precursor HFF (1110 proteins) and the hESC H9 cell line versus somatic precursor HFF (1105 proteins). To generate the hiPSCs/hESCs profile, we first generated two profiles: one of 864 proteins for hiPSCs versus hESC NTU1 by taking the intersection of the two profiles for hiPSC CFB46 versus hESC NTU1 (1125 proteins) and hiPSC CFB50 versus hESC NTU1 (1128 proteins), and the other of 870 proteins for hiPSCs versus hESC H9 by taking the intersection of the two profiles for hiPSC CFB46 versus hESC H9 (1115 proteins) and hiPSC CFB50 versus hESC H9 (1122 proteins). The hiPSCs/hESCs profile of 810 proteins was thereafter obtained by taking the intersection of the two generated profiles. The raw protein profiles from mass spectrometry and the above three integrated protein profiles were listed in Supplementary Tables S1 and S2, respectively.

**Similarity measure of hiPSCs and hESCs at protein expression level.** The protein  $\log_2$  ratios of the hESCs/HFF profile against that of the hiPSCs/HFF profile were depicted in Fig. 2a. A Spearman correlation coefficient of 0.945 was received to show the high similarity between the proteomic profiles of hESCs and hiPSCs. Moreover, a histogram showing the distribution of the  $\log_2$  ratios of the hiPSCs/hESCs profile was depicted in Fig. 2b. The narrow and high spike near the zero of the x-axis also revealed the high similarity between our hiPSC and hESC samples.



**Figure 2.** Similarity evaluation of the membrane proteomes among hiPSCs, hESCs, and HFF. **(a)** The protein  $\log_2$  ratios of the profile hiPSCs/HFF are plotted against those of hESCs/HFF. A Spearman correlation coefficient of 0.945 was obtained. **(b)** The histogram depicted the distribution of protein  $\log_2$  ratios of the profile hiPSCs/hESCs. **(c)** The protein  $\log_2$  ratios of the profile hiPSCs/hESCs are plotted against those of HFF/hESCs. A Spearman correlation coefficient of 0.187 was obtained. **(d)** The protein  $\log_2$  ratios of the profile hESCs/hiPSCs are plotted against those of HFF/hiPSCs. A Spearman correlation coefficient of 0.076 was obtained.

The protein  $\log_2$  ratios of the HFF/hESCs against that of the hiPSCs/hESCs profile were depicted in Fig. 2c. A Spearman correlation coefficient of 0.187 revealed that our hiPSC samples were very different from HFF. The protein  $\log_2$  ratios of the HFF/hiPSCs profile against that of the hESCs/hiPSCs profile were depicted in Fig. 2d. A Spearman correlation coefficient of 0.076 reveal that our hESC samples were also very different from HFF.

**Three-way pathway comparison of hiPSCs/HFF, hESCs/HFF, and hiPSCs/hESCs.** *Pathway comparison between hiPSCs/HFF and hESCs/HFF.* To evaluate the results of the reprogramming of hiPSCs, we assessed the similarity between hiPSCs and hESCs at the functional level; we identified significant pathways associated with the profiles hiPSCs/HFF and hESCs/HFF. As shown in Table 1, the GSEA reported 13 significant pathways (4 up-regulated and 9 down-regulated) common to both hESCs/HFF and hiPSCs/HFF while 6 significant pathways (5 up-regulated and 1 down-regulated) exclusive to hiPSCs/HFF. The differentially expressed proteins that contributed to the significance of these pathways were summarized in Supplementary Table S3. Our results revealed that the two pluripotent cells did share some regulation patterns when compared to the somatic HFF cell. The four common down-regulated pathways *platelet activation signaling and aggregation*, *hemostasis*, *adaptive immune system*, and *immune system* are related to hemostasis and immune system. To the best of our knowledge, currently, no evidence was reported to show the regulation of immune-related signalling during reprogramming processes. The GSEA results in this study, however, showed that numerous down-regulated proteins were found enriched in the immune-related pathway between the pluripotent cells and their differentiated counterpart (Supplementary Table S3), which suggests the possible correlation between the reprogramming processes and immune signalling. For example, our results showed that TAP-1 (Antigen peptide transporter 1) and HLA-C (Major Histocompatibility Complex, Class I, C) were down-regulated in hESCs and hiPSCs; the two proteins were reported involved in an immune evasion mechanism of pluripotent cells<sup>20</sup>.

Another three common down-regulated pathways *regulation of actin cytoskeleton*, *focal adhesion*, and *leukocyte transendothelial migration* are related to cell motility and cell adhesion. Previous studies have revealed that pluripotent stem cells adopt tight colony morphology with low adhesive strength, in contrast to the parental somatic fibroblast cells. The differentiation of hiPSCs significantly increases the adhesion strength to the ECM and the differences in adhesive strength were found correlated to more focal adhesions in parental cells than in

Pathway	hiPSCs/HFF (p-val/FDR)	hESCs/HFF (p-val/FDR)
REACTOME: platelet activation signaling and aggregation	DOWN 0.000/0.000	DOWN 0.000/0.000
REACTOME: hemostasis	DOWN 0.000/0.002	DOWN 0.000/0.007
REACTOME: adaptive immune system	DOWN 0.002/0.003	DOWN 0.002/0.002
REACTOME: immune system	DOWN 0.000/0.003	DOWN 0.002/0.006
KEGG: regulation of actin cytoskeleton	DOWN 0.000/0.000	DOWN 0.000/0.000
KEGG: focal adhesion	DOWN 0.000/0.000	DOWN 0.000/0.001
KEGG: leukocyte transendothelial migration	DOWN 0.101/0.132	DOWN 0.060/0.088
REACTOME: axon guidance	DOWN 0.000/0.004	DOWN 0.000/0.010
REACTOME: developmental biology	DOWN 0.002/0.003	DOWN 0.002/0.009
REACTOME: TCA cycle and respiratory electron transport	UP 0.000/0.004	UP 0.000/0.001
KEGG: oxidative phosphorylation	UP 0.000/0.001	UP 0.000/0.000
KEGG: Alzheimer's disease	UP 0.159/0.196	UP 0.084/0.074
KEGG: Huntington's disease	UP 0.000/0.001	UP 0.000/0.000
KEGG: tight junction	DOWN 0.222/0.214	N/A—
REACTOME: respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins	UP 0.000/0.003	N/A—
KEGG: Parkinson's disease	UP 0.000/0.001	N/A—
REACTOME: post-translational protein modification	UP 0.004/0.020	N/A 0.224/0.601
REACTOME: metabolism of proteins	UP 0.002/0.039	N/A 0.088/0.694
REACTOME: asparagine N-linked glycosylation	UP 0.105/0.204	N/A 0.534/0.626

**Table 1.** Significant pathways of hiPSCs/HFF and hESCs/HFF. UP: up-regulation; DOWN: down-regulation; N/A: not applicable; hiPSCs: human induced pluripotent stem cells; hESCs: human embryonic stem cells; HFF: human foreskin fibroblast; p-val: p-value; FDR: false discovery rate; —: not in GSEA.

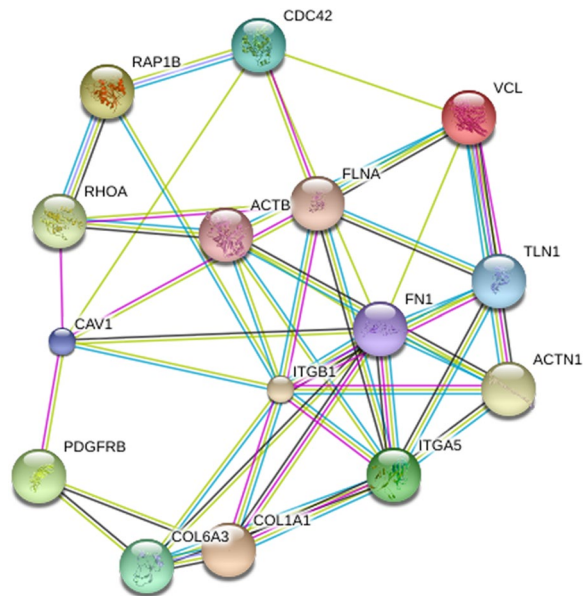
hiPSCs<sup>21</sup>. Protein signatures related to focal adhesion such as vinculin, talin, actin stress fibre, and  $\alpha_5\beta_1$ -integrin were found enriched in human fibroblasts when compared to hiPSCs, and both hESCs and hiPSCs possess lower adhesion strength than fibroblasts<sup>21</sup>. Moreover, to support stem cell self-renewal and prevent differentiation, the pluripotent stem cells were reported to promote the inactivation of focal adhesion kinase (FAK) signalling<sup>22</sup>. The FAK signalling was mentioned related to the focal adhesion pathway<sup>23,24</sup>. Based on our results along with the previous findings, the proteins related to focal adhesion pathway might be modulated to be down-regulated in stem cells to maintain the undifferentiated states.

In our proteomic data, the enriched protein PDGFR $\beta$  (Platelet-derived growth factor receptor-beta) was down-regulated in the *focal adhesion* pathway of both hESCs/HFF and hiPSCs/HFF. The PDGFR $\beta$  was reported up-regulated in human fibroblasts when compared to pluripotent cells and known to be involved in cell migration that is suggested related to focal adhesion<sup>25,26</sup>. Our results were basically in agreement with previous findings regarding the shift of adhesive properties during reprogramming<sup>21</sup>. All the enriched proteins in our study were further subjected to the STRING database for protein-protein interaction analysis. A tightly connected protein interaction module was formed by 15 enriched proteins (in both hESCs/HFF and hiPSCs/HFF) including small GTPase proteins RHOA (ras homolog gene family, member A), RAP1B (RAP1B, a member of RAS oncogene family), and CDC42 (cell division cycle 42). Remarkably, all the members of this module were found within the *focal adhesion* pathway, which implies that focal adhesion may function between pluripotent cells and differentiated fibroblast cells (Fig. 3 and Table 2).

As to the remaining two common down-regulated pathways *axon guidance* and *development biology*, more than half of enriched proteins also appeared in the cell motility and cell adhesion pathways. Thus the down-regulation of these proteins in cell motility and cell adhesion may also contribute to the statistical significance of the two pathways.

On the other hand, the two pathways *TCA cycle and respiratory electron transport* and *oxidative phosphorylation* were found up-regulated in both pluripotent cells when compared to somatic fibroblast cells. The two pathways are mitochondrial-mediated metabolic pathways. Previous studies concluded that pluripotent cells favour glycolysis as a source of energy, whereas the differentiated somatic cells rely on mitochondrial oxidative phosphorylation for energy production<sup>27</sup>. Nevertheless, proper mitochondrial network integrity and functions were reported to be important for the maintenance of pluripotency<sup>28</sup>. Our results were in agreement with the finding of a recent membrane proteomic study on murine iPSCs; the up-regulated proteins in either murine ESCs or iPSCs were found enriched in oxidative phosphorylation and electron transport chain (relative to fibroblast cells)<sup>29</sup>. Another recent proteomic study on human iPSCs also obtained similar results that proteins involved in oxidative phosphorylation were up-regulated<sup>30</sup>. In addition, a study focusing on the energy metabolism of hiPSCs revealed that despite the lower mitochondrial activity, the protein expression level of mitochondrial complexes II, III, and V was found up-regulated in pluripotent cells when compared to differentiated cells possibly due to a higher level of c-myc<sup>31</sup>. All the aforementioned studies support our results (Table 1 and Supplementary Table S3) to the up-regulation of certain mitochondrial complex components in pluripotent cells at protein expression level. The elevated expression of numerous proteins in the electron transport was suggested to be involved in the





**Figure 3.** The protein interaction module in the *focal adhesion* pathway regulated between pluripotent cells and somatic fibroblast cells. This module was composed of 15 proteins that were enriched in the *focal adhesion* pathway. The interactions between these proteins were suggested by the STRING database.

Gene Symbol	Protein Description
ACTB	Actin, cytoplasmic 1
ACTN1	Alpha-actinin-1
CAV1	Isoform Alpha of Caveolin-1
CDC42	Isoform 2 of Cell division control protein 42 homolog
COL1A1	Collagen alpha-1(I) chain
COL6A3	alpha 3 type VI collagen isoform 4 precursor
FLNA	Isoform 2 of Filamin-A
FN1	Isoform 1 of Fibronectin
ITGA5	Integrin alpha-5
ITGB1	Isoform Beta-1A of Integrin beta-1
PDGFRB	Beta-type platelet-derived growth factor receptor
RAP1B	Ras-related protein Rap-1b
RHOA	Transforming protein RhoA
TLN1	Talin-1
VCL	Isoform 1 of Vinculin

**Table 2.** Fifteen enriched proteins within a tightly connected protein interaction module regulated between pluripotent cells and differentiated fibroblast cells in *focal adhesion* pathway. The protein interaction modules were revealed based on the STRING database.

stoichiometric change of the electron transport chain, which may affect the composition of respiratory supercomplexes and lead to a decrease in efficiency of oxidative phosphorylation<sup>32,33</sup>.

The remaining two common up-regulated pathways *Alzheimer's disease* and *Huntington's disease* are related to the biological signalling in diseases. As depicted in the KEGG database, oxidative phosphorylation and mitochondrial-related functions are involved in these disease pathways, the up-regulation of the mitochondrial-related proteins may also contribute to the statistical significance of the two disease pathways. During the pathogenesis of the two diseases, mitochondrial abnormalities and impaired mitochondrial functions have been revealed<sup>34,35</sup>. There is also evidence that the mitochondrial and related respiratory functions are modulated during the reprogramming process in favour of pluripotent glycolytic metabolism<sup>27</sup>. Alzheimer's and Huntington's diseases are known to be caused by the accumulation of disordered protein  $\beta$ -amyloid precursor protein (APP) and huntingtin (HTT), respectively<sup>34,36</sup>. Although the proteins APP was not found enriched in our membrane proteomic data, the level of APP was reported to be highly expressed in undifferentiated human transformed pluripotent stem cells<sup>37</sup>. Huntingtin was absent from our data because it is not a membrane protein. However, HTT is required for normal hematopoiesis and has been reported to be expressed in embryonic cells<sup>38</sup>.

Whether or not the proteins APP and HTT are involved in the regulation of mitochondrial function during the reprogramming process may require further investigation.

We also investigated those pathways that were differentially regulated between hiPSCs/HFF and hESCs/HFF. As shown in Table 1, the three pathways *tight junction*, *respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins*, and *Parkinson's disease* were found significant only in hiPSCs/HFF. As to the hESCs/HFF profile, the three pathways were not identified as significant by the GSEA due to the parameter setting of the GSEA; only those pathways containing no less than 20 mapped proteins can be served as the candidates of significant pathways. Both *tight junction* and *Parkinson's disease* contained 19 proteins in the hESCs/HFF profile, and *respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins* contained 18 proteins in the hESCs/HFF. Since the number of mapped proteins of the three pathways was close to the threshold 20, the three pathways were suggested to be significant to both hiPSCs/HFF and hESCs/HFF.

While further dug into the *tight junction* pathway, we found that many down-regulated enriched proteins in both hiPSCs/HFF and hESCs/HFF were also enriched in cell motility and cell adhesion-related pathways. Since *tight junction* is known to be involved in cell adhesion signalling<sup>39</sup>, the *tight junction* pathway seems to possess similar down-regulation pattern as well as other cell motility and cell adhesion-related pathways. Likewise, many enriched proteins in *respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins* and *Parkinson's disease* pathways were mostly the mitochondrial-related proteins that were up-regulated in *oxidative phosphorylation* and *TCA cycle and respiratory electron transport* pathways. As the aforementioned pathways *Alzheimer's disease* and *Huntington's disease*, the statistical significance of *respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins* and *Parkinson's disease* may also come from the up-regulation of the mitochondrial-related pathways.

The three pathways *post-translational protein modification*, *metabolism of protein*, and *asparagine N-linked glycosylation* were also reported significant only in hiPSCs/HFF; the three pathways are related to protein metabolism and post-translational modification. We found that proteins UGGT1 (UDP-glucose glycoprotein glucosyltransferase 1), SLC25A4 (solute carrier family 25, member 4), PIGK (phosphatidylinositol glycan anchor biosynthesis, class K), and WBSR17 (Williams-Beuren syndrome chromosome region 17) were up-regulated and enriched in hiPSCs/HFF but missed in hESCs/HFF. In addition, proteins DOLPP1 (dolichyl pyrophosphate phosphatase 1), FAU (ubiquitin-like protein FUBI), GALNT2 (polypeptide N-acetylgalactosaminyltransferase 2), GALNT7 (polypeptide N-acetylgalactosaminyltransferase 7), RPS19 (ribosomal protein S19), RPS20 (ribosomal protein S20), and PIGT (phosphatidylinositol glycan anchor biosynthesis) were up-regulated at a higher expression level in hiPSCs/HFF than in hESCs/HFF. Pathways related to protein metabolism and post-translational modification may possess different expression patterns among the two pluripotent cells while compared to somatic HFF.

**Functional analysis of the differences between hiPSCs and hESCs.** To reveal the difference in biological functions between hiPSCs and hESCs, we examined the significant pathways of the hiPSCs/hESCs profile. There were 11 pathways (4 up-regulated and 7 down-regulated) differentially regulated between the two pluripotent cells (Table 3). The 11 pathways can be characterized into three functional categories after a cross-reference to the significant pathways of hiPSCs/HFF and hESCs/HFF. As listed in Table 3, the three protein modification-related pathways *N-glycan biosynthesis*, *post-translational modification*, and *asparagine N-linked glycosylation* were up-regulated in hiPSCs/hESCs. The *post-translational modification* and *asparagine N-linked glycosylation* were also up-regulated in hiPSCs/HFF. None of the three was up-regulated in hESCs/HFF. The result indicates that the pathways related to protein modification might be specifically regulated in hiPSCs when compared to either pluripotent hESCs or somatic HFF.

The pathway *focal adhesion* was found to be significant in all the three profiles hiPSCs/hESCs, hiPSCs/HFF, and hESCs/HFF. The pathway was up-regulated in hiPSCs/hESCs and down-regulated in both hiPSCs/HFF and hESCs/HFF. The down-regulation of *focal adhesion* in both hiPSCs/HFF and hESCs/HFF suggested that the pathway may be less activated in both pluripotent cells than in somatic HFF cells. The up-regulation of *focal adhesion* in hiPSCs/hESCs further revealed that the pathway may be more activated in hiPSCs than in hESCs. On the basis of the regulation direction of *focal adhesion* in the three profiles, we hypothesized that *focal adhesion* may undergo an incomplete repression during the reprogramming process of hiPSCs (Fig. 4a).

There were 7 down-regulated significant pathways *TCA cycle and respiratory electron transport*, *respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins*, *mitochondrial protein import*, *oxidative phosphorylation*, *Alzheimer's disease*, *Huntington's disease*, and *Parkinson's disease* in hiPSCs/hESCs. When cross-referencing to the significant pathways of hiPSCs/HFF and hESCs/HFF, pathways related to mitochondria functions (*oxidative phosphorylation* and *TCA cycle and respiratory electron transport*) and diseases (*Alzheimer's disease* and *Huntington's disease*) were up-regulated in both hiPSCs/HFF and hESCs/HFF. The 4 pathways are either mitochondrial-mediated metabolic pathways or selected as significant pathways because most of their members are involved in mitochondrial function. We therefore paid our attention only to those mitochondrial-related pathways. According to the regulation direction of the mitochondrial-related pathways in the three profiles, *oxidative phosphorylation* and *TCA cycle and respiratory electron transport* were hypothesized to undergo an incomplete activation during the reprogramming process of hiPSCs (Fig. 4b).

The reprogramming of hiPSCs has been shown to be a multi-stage process that involves large-scale changes in the transcriptional and epigenetic states of somatic cells. Several expressed genes and pathways in somatic cells were suggested to act as antagonizing barriers during the reprogramming<sup>40</sup>. The barriers may jeopardize the changes in the adhesive signature in reprogramming and lead to the incomplete regulation of the *focal adhesion* pathways. Besides, metabolic change may serve as a vital progress during the reprogramming of hiPSCs<sup>41</sup>. The genes involved in glycolysis and oxidative phosphorylation undergo epigenetic and gene expression changes for

Pathway	hiPSCs/hESCs (p-val/ FDR)	hiPSCs/HFF (p-val/ FDR)	hESCs/HFF (p-val/ FDR)
KEGG: N-glycan biosynthesis	UP 0.006/0.013	N/A—	N/A—
REACTOME: post translational protein modification	UP 0.000/0.006	UP 0.004/0.020	N/A 0.224/0.601
REACTOME: asparagine N-linked glycosylation	UP 0.008/0.037	UP 0.105/0.204	N/A 0.534/0.626
KEGG: focal adhesion	UP 0.062/0.111	DOWN 0.000/0.000	DOWN 0.000/0.001
KEGG: oxidative phosphorylation	DOWN 0.033/0.232	UP 0.000/0.001	UP 0.000/0.000
REACTOME: TCA cycle and respiratory electron transport	DOWN 0.056/0.187	UP 0.000/0.004	UP 0.000/0.001
KEGG: Alzheimer's disease	DOWN 0.066/0.088	UP 0.159/0.196	UP 0.084/0.074
KEGG: Huntington's disease	DOWN 0.057/0.113	UP 0.000/0.001	UP 0.000/0.000
KEGG: Parkinson's disease	DOWN 0.063/0.150	UP 0.000/0.001	N/A—
REACTOME: respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins	DOWN 0.032/0.223	UP 0.000/0.003	N/A —
REACTOME: mitochondrial protein import	DOWN 0.061/0.144	N/A —	N/A —

**Table 3.** Three-way pathway comparison of hiPSCs/hESCs, hiPSCs/HFF, and hESCs/HFF. UP: up-regulation; DOWN: down-regulation; N/A: not applicable; hiPSCs: human induced pluripotent stem cells; hESCs: human embryonic stem cells; HFF: human foreskin fibroblast; p-val: p-value; FDR: false discovery rate; —: not in GSEA.

progressive resetting of metabolite levels<sup>42</sup>. The reprogrammed iPSCs, however, may still retain specific chromatin modifications related to cellular respiration and metabolism. Similar to the epigenetic memory, a notion of “metabolic memory” retained in iPSCs was proposed to detain the metabolic reprogramming process and provide a metabolic barrier for reprogramming<sup>43</sup>. The reprogramming barriers and metabolic memory may lead to the incomplete regulation of hiPSCs.

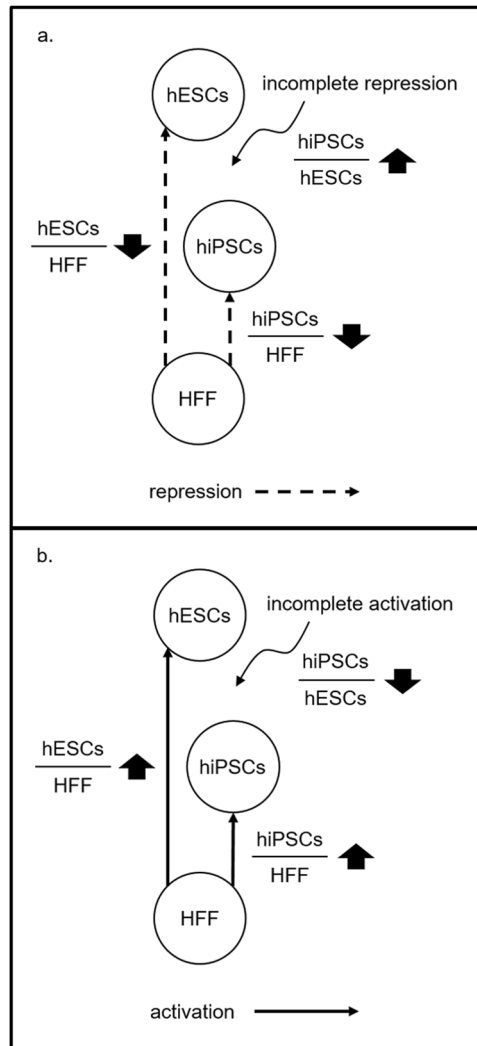
There were several membrane proteins that were enriched in the aforementioned significant pathways and whose regulation directions concurred with the regulation directions of these pathways. These proteins may serve as the starting points of the research to make hiPSCs closer to hESCs since membrane proteins usually play an important role in biological signalling. For the hypothetical incomplete repressed pathway *focal adhesion*, several proteins such as ACTN1 (actinin, alpha 1), VCL, and Rap1B were down-regulated in both hiPSCs/HFF and hESCs/HFF but up-regulated in hiPSCs/hESCs (Table 4A). For the hypothetical incomplete activated pathways *oxidative phosphorylation* and *TCA cycle and respiratory electron transport*, several proteins such as UQCRQ (ubiquinol-cytochrome c reductase, complex III subunit VII), CYC1 (cytochrome c 1), COX4I1 (cytochrome c oxidase subunit IV isoform 1), and ATP5H (ATP synthase, H<sup>+</sup> transporting, mitochondrial F<sub>0</sub> complex, subunit d) were found up-regulated in hiPSCs/HFF and hESCs/HFF and down-regulated in hiPSC/hESCs (Table 4B). These membrane proteins might all be used for the further study of dissimilarity between hiPSCs and hESCs.

Even though a comprehensive three-way pathway comparison has been performed on the three profiles hiPSCs/hESCs, hiPSCs/HFF, and hESCs/HFF, we are not sure if the obtained hypothetical incomplete repressed and activated pathways are actually involved in the biological events behind the reprogramming process of hiPSCs. To address this issue, we surveyed related proteomic studies and found a proteomics research that supports our results to certain degree<sup>5</sup>. There were two human fibroblast (HF) cell lines IMR90 fetal fibroblasts and 4Skin fetal fibroblasts used in the study, and each of them had two biological replicas. The study therefore obtained 12 profiles of protein relative expression ratio, in which 4 profiles recorded log<sub>2</sub> ratios of hiPSCs to somatic cells, 4 recorded log<sub>2</sub> ratios of hESCs to somatic cells, and 4 recorded log<sub>2</sub> ratios of hiPSCs to hESCs. To be compared to our results, the 12 obtained profiles were subjected to our data integration procedure to produce three integrated profiles, one for hiPSCs versus somatic HFs (hiPSCs/HFs), one for hESCs versus somatic HFs (hESCs/HFs), and one for hiPSCs versus hESCs (hiPSCs/hESCs). The three integrated profiles are shown in Table S4. The three profiles were subjected to GSEA to perform pathway analyses; the identified significant pathways were summarized in Supplementary Table S5 and the differentially expressed proteins that contributed to the significance of these pathways were summarized in Supplementary Table S6. We adopted the same strategy of our pathway analysis. The identified significant pathways of hiPSCs/hESCs were cross-referenced to those of hiPSCs/HFs and hESCs/HFs, and the results are listed in Table 5. The results showed that the pathway *focal adhesion* was up-regulated in hiPSCs/hESCs and down-regulated in both hiPSCs/HFs and hESCs/HFs, which is in agreement with our result: *focal adhesion* may undergo an incomplete repression process during the reprogramming of hiPSCs. On the other hand, the pathway *TCA cycle and respiratory electron transport* was found down-regulated in hiPSCs/hESCs but missed in both hESCs/HFs and hiPSCs/HFs. Although the mitochondrial-related pathway did not totally possess the same regulation patterns as ours, hiPSCs in the study also showed the same trend towards up-regulation when compared to hESCs as revealed by our study. The conclusion drawn from this study is similar to ours, which greatly increase the confidence of our results because there are usually significant differences among pluripotent cells obtained from different laboratories. The *focal adhesion* and mitochondrial-related pathways are therefore suggested to be incomplete events during the reprogramming process of hiPSCs.

## Methods

**Culture of hiPSCs and hESCs.** Two hESC cell lines NTU1 and H9<sup>44,45</sup>, two hiPSC cell lines CFB46 and CFB50 (retroviral reprogrammed and passed from our previous studies)<sup>46,47</sup>, and their somatic precursor HFF





**Figure 4.** The hypothetical incomplete regulation of pathways during the reprogramming of hiPSCs. **(a)** Down-regulation in hiPSCs/HFF and hESCs/HFF and up-regulation in hiPSCs/hESCs suggest an incomplete repression of pathways. **(b)** Up-regulation in hiPSCs/HFF and hESCs/HFF and down-regulation in hiPSCs/hESCs suggest an incomplete activation of pathways.

cells (human foreskin fibroblast cells) were prepared for this study. Both hiPSC cell lines and the two hESC cell lines were maintained on MEF (mouse embryonic fibroblast feeder) with the serum-free medium as previously described<sup>47,48</sup>, and the somatic precursor HFF cells were obtained and cultured in DMEM with 20% FBS (fetal bovine serum) as described in a previous study as well<sup>47</sup>.

**Membrane fractionation and gel-assisted digestion of membrane proteins.** The isolation of membrane proteins and the following gel-assisted digestion were performed using the protocol we conducted in our membrane proteomic study<sup>49</sup>. In brief, after washed with PBS (phosphate buffered saline), the cells were collected and homogenized in hypotonic buffer (1.5 mM MgCl<sub>2</sub>, 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, 10 mM KCl, and 1 × protease inhibitor mixture (Calbiochem)). We first removed nuclei by a centrifugation at 3,000 × g for 10 min at 4 °C. The post-nuclear supernatant was then mixed with 1.8 M sucrose and subjected to a second centrifugation at 13,000 × g for 1 h at 4 °C to obtain membrane pellet. After washed with ice-cold 0.1 M Na<sub>2</sub>CO<sub>3</sub> (pH 11.5), the membrane pellet was dissolved in 90% (v/v) formic acid and vacuum-dried.

To prepare samples for following MS analyses, the membrane pellet was subjected to gel-assisted digestion<sup>49</sup>. The membrane protein pellet was first resuspended and added with 5 mM TCEP (Tris(2-carboxyethyl) phosphine hydrochloride) and 2 mM MMTS (methyl methanethiosulfonate) for protein reduction and protein alkylation, respectively. Next, the protein sample was applied with 10% (w/v) APS (Ammonium persulfate), Acrylamide/bisacrylamide (40%, 29:1, v/v), and 100% TEMED (N,N,N',N'-tetramethylethylenediamine) to polymerize as a gel. After the gel was cut into small pieces and washed, the tryptic digestion (protein: trypsin 10:1, w/w) was performed in 25 mM TEABC (triethylammonium bicarbonate) at 37 °C overnight. A sequential extraction with

Gene Symbol	hiPSCs/hESCs	hiPSCs/HFF	hESCs/HFF
<b>A.</b>			
ACTN1	0.355	-2.251	-2.696
CAV1	0.195	-3.113	-3.461
FN1	0.350	-1.622	-1.992
RAP1B	0.394	-1.632	-2.044
RHOA	0.233	-1.010	-1.265
VCL	0.647	-1.647	-2.371
<b>B.</b>			
ATP5H	-0.361	2.195	2.603
ATP5J2	-0.283	1.804	2.089
ATP5L	-0.329	1.498	1.917
COX4I1	-0.429	1.299	1.794
CYC1	-0.578	0.813	1.576
NDUFA9	-0.427	0.918	1.216
NDUFB10	-0.227	1.319	1.403
NDUFB6	-0.221	2.024	2.223
UQCRRS1	-0.501	1.793	2.349
UQCRRQ	-0.577	2.115	2.672

**Table 4.** The expression  $\log_2$  ratio of the discriminated proteins in the hypothetical incomplete regulated pathways. **(A)** Pathway *focal adhesion* **(B)** Pathways *oxidative phosphorylation* and *TCA cycle and respiratory electron transport*.

25 mM TEABC, 0.1% (v/v) TFA (trifluoroacetic acid) in water, 0.1% (v/v) TFA in ACN (acetonitrile), and 100% ACN was performed to extract peptides from the gel. The peptides were then concentrated in a SpeedVac and subjected to MS analysis.

**Mass spectrometry analysis for protein identification and quantification.** In this study, the liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses were performed for protein identification and quantitation using Waters Q-TOF Premier mass spectrometer (Waters Corp., Milford, MA). Follow our previous study<sup>50</sup>, the samples were first injected into a trap column (20-mm  $\times$  180- $\mu$ m), separated by a capillary column (200-mm  $\times$  75  $\mu$ m), and eluted with a linear gradient of 0–80% of 0.1% (v/v) FA in ACN at 300 nl/min for 120 min. Data acquisition was operated under the duty cycle of full MS scan ( $m/z$  400–1600, 0.6 s) followed by three MS/MS scans ( $m/z$  100–1990, 1.2 s for each scan) on the three highest peaks present in the full MS scan.

The MS/MS data were processed and searched against International Protein Index (IPI) human database<sup>51</sup> (v3.29, 68161 sequences) using Mascot v2.2 (Matrix science, London, United Kingdom, <http://www.matrixscience.com/>). The search parameters were set as follows: 0.3-Da and 0.1-Da mass tolerances for MS and MS/MS fragment ions, respectively; oxidation (Met) and methylation (Cys) as variable modifications; up to 2 missed cleavages. Only unique peptides with scores higher than 35 ( $p < 0.05$ ) were confidently assigned. A decoy database search was also performed to evaluate the false discovery rate of protein identification by searching against a Mascot-created randomize protein sequence database with identical validation criteria and search parameters. The Mascot search results were exported in extensive markup language (XML) data format.

In the current study, protein quantitation was performed by a label-free approach using the IDEAL-Q software<sup>52</sup>. The spectral raw data were first converted into mzXML format using a computational tool, massWolf v4.0 (<http://sourceforge.net/projects/sashimi/files/massWolf> (MassLynx converter)). The obtained mzXML data coupled with Mascot search results in XML format were then submitted to IDEAL-Q for quantitation analysis. The peptide abundance was determined by extracted ion chromatography (XIC) and further normalized to internal peptide standard. The protein ratio was calculated by the weighted ratio of normalized peptide abundance in different samples.

**Data processing and integration.** For a comparative proteomics analysis, a protein profile of expression ratio is basically expected for the following pathway analyses. The scale of spectral proteomic data, however, is usually much smaller than that of genomic data. To quantify as many peptides as possible, for each sample we produced two spectral datasets by repeated MS experiments. The two datasets were subjected to the IDEAL-Q to perform protein identification and generate two protein profiles of expression ratio for the sample; the two profiles were further integrated into a single one for pathway analyses. The integration of the two profiles was done as follows. First, all the protein expression ratios were transformed into  $\log_2$  ratios. Next, for each protein that had a different  $\log_2$  ratio in the two profiles, we either removed the protein from our profiles or determined a unique  $\log_2$  ratio for the protein based on the following two criteria. 1) If the expression of the redundant protein was in opposite up- and down-regulation directions in the two profiles and the inconsistency in expression is significant (numerical difference between the two fold changes  $> 5$ ), the protein was removed. 2) Otherwise, the  $\log_2$  ratio of the protein was set to be the average of the two  $\log_2$  ratios of the protein. Finally, all the left proteins were pooled together to form a new profile. We totally obtained eight integrated membrane protein profiles of expression ratio

Pathways	hiPSCs/hESCs (p-val/ FDR)	hiPSCs/HFs (p-val/ FDR)	hESCs/HFs (p-val/ FDR)
KEGG: DNA replication	UP 0.096/0.245	UP 0.000/0.000	UP 0.000/0.000
KEGG: ECM-receptor interaction	UP 0.003/0.025	DOWN 0.000/0.000	DOWN 0.000/0.000
KEGG: arrhythmogenic right ventricular cardiomyopathy (ARVC)	UP 0.011/0.145	DOWN 0.051/0.060	DOWN 0.011/0.013
KEGG: focal adhesion*	UP 0.003/0.136	DOWN 0.000/0.000	DOWN 0.000/0.000
KEGG: regulation of actin cytoskeleton	UP 0.002/0.107	DOWN 0.000/0.000	DOWN 0.000/0.000
KEGG: epithelial cell signaling in helicobacter pylori infection	UP 0.083/0.247	DOWN 0.076/0.118	DOWN 0.017/0.021
REACTOME: cell surface interactions at the vascular wall	UP 0.002/0.185	DOWN 0.031/0.031	DOWN 0.004/0.006
REACTOME: integrin cell surface interactions	UP 0.003/0.123	DOWN 0.005/0.007	DOWN 0.004/0.005
REACTOME: glucose transport	DOWN 0.000/0.000	UP 0.000/0.000	UP 0.000/0.000
REACTOME: regulation of glucokinase by glucokinase regulatory protein	DOWN 0.000/0.000	UP 0.000/0.000	UP 0.000/0.000
REACTOME: transport of ribonucleoproteins into the host nucleus	DOWN 0.000/0.000	UP 0.000/0.000	UP 0.000/0.000
REACTOME: processing of capped intron-containing pre-mRNA	DOWN 0.000/0.000	UP 0.000/0.000	UP 0.000/0.000
REACTOME: NEP/NS2 interacts with the cellular export machinery	DOWN 0.027/0.030	UP 0.000/0.000	UP 0.000/0.000
KEGG: arginine and proline metabolism	DOWN 0.055/0.058	N/A 0.615 /0.684	UP 0.053/0.105
KEGG: fatty acid metabolism	DOWN 0.101/0.198	N/A 0.946/1.000	UP 0.025/0.078
KEGG: peroxisome	DOWN 0.000/0.001	N/A 0.589/0.623	N/A 0.388/0.553
REACTOME: TCA cycle and respiratory electron transport*	DOWN 0.000/0.042	N/A 0.999/1.000	N/A 0.327/0.442
REACTOME: respiratory electron transport ATP synthesis by chemiosmotic coupling and heat production by uncoupling proteins*	DOWN 0.086/0.224	N/A 0.728/0.778	N/A 0.558/0.641
REACTOME: pyruvate metabolism and citric acid (TCA) cycle	DOWN 0.000/0.005	N/A 0.998/1.000	N/A 0.539/0.639

**Table 5.** Three-way pathway comparison of hiPSCs/hESCs, hiPSCs/HFs, and hESCs/HFs on a public dataset. \*Significant pathways found in both public dataset and our dataset; UP: up-regulation; DOWN: down-regulation; N/A: not applicable; hiPSCs: human induced pluripotent stem cells; hESCs: human embryonic stem cells; HF: human fibroblast; p-val: p-value; FDR: false discovery rate.

for the following eight experimental samples: hiPSC CFB46 versus somatic precursor HFF, hiPSC CFB50 versus somatic precursor HFF, hESC NTU1 versus somatic precursor HFF, hESC H9 versus somatic precursor HFF, hiPSC CFB46 versus hESC NTU1, hiPSC CFB46 versus hESC H9, hiPSC CFB50 versus hESC NTU1, and hiPSC CFB50 versus hESC H9. The eight profiles respectively contained 1100, 1111, 1110, 1105, 1125, 1115, 1128, and 1122 protein expression ratios.

The eight profiles were further integrated into three profiles: one for hiPSCs versus somatic precursor HFF (hiPSCs/HFF), one for hESCs versus somatic precursor HFF (hESCs/HFF), and one for hiPSCs versus hESCs (hiPSCs/hESCs). We generated the hiPSCs/HFF profile by merging two hiPSC biological replicas, hiPSC CFB46 and hiPSC CFB50. Similarly, the hESCs/HFF profile was obtained by merging two hESC cell lines, hESC NTU1 and hESC H9. To generate the hiPSCs/hESCs profile, we first generated a profile for hiPSCs versus hESC NTU1 and a profile for hiPSCs versus hESC H9 by merging hiPSC biological replicas, and further integrated the two profiles into the hiPSCs/hESCs profile by merging the two hESC cell lines. All the aforementioned merges of two profiles were done by selecting those proteins quantified in both profiles, and the  $\log_2$  ratio of each selected protein was set to be the average of the  $\log_2$  ratios of the protein in the two source profiles.

**Statistical correlation measure.** Since the distributions of the protein expression ratios in our datasets were not characterized by any parameter, the Spearman correlation coefficients were therefore used to evaluate the similarity between our profiles. Given two ranked protein expression profiles  $\mathbf{x}$  and  $\mathbf{y}$ , the Spearman correlation coefficient  $r_s$  of the two profiles is defined as follows:

$$r_s = 1 - \frac{6 \times \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)},$$

where  $n$  is the number of protein expression ratios in a profile, and  $x_i$  and  $y_i$  are protein expression ratios belong to  $\mathbf{x}$  and  $\mathbf{y}$ , respectively.  $r_s$  receives a value of +1 or -1 if  $\mathbf{x}$  and  $\mathbf{y}$  are perfectly monotonic positive or negative correlated, respectively; it receives a value of zero if  $\mathbf{x}$  and  $\mathbf{y}$  are not correlated with each other. In this study, the calculation of the Spearman correlation coefficients was carried out by the R language (<https://www.r-project.org/>).

**Enrichment of significant pathway.** The bioinformatics tool Gene Set Enrichment Analysis (GSEA)<sup>18,19</sup> was used in this study to enrich the target pathways with statistically significant difference between two cell types (e.g., hiPSC versus hESC). Given a protein profile sorted by the expression ratio between the two cell types, the target pathway is regarded as significant if most of its protein members are enriched in the top (up-regulated)

or bottom (down-regulated) region of the profile. On the basis of the Kolmogorov-Smirnov test, the GSEA first calculates an ES score (enrichment score) for the target pathway to measure the degree to which the pathway is enriched in the top-ranked or bottom-ranked region of the profile. Next, a normalized ES (NES) is calculated from the ES to take into account the pathway size. Finally, a false discovery rate (FDR) is determined based on the NES to represent the significance of the target pathway. The GSEA takes into consideration all the proteins from the experimental results without any filtration procedure. Thus, all the proteins are able to contribute to the enrichment analysis even if some of them have a very small fold change in expression<sup>13</sup>. In this study, the parameters of the GSEA were empirically set as follows:

1. The “difference of classes” metric was selected to rank proteins according to the  $\log_2$  ratio between two phenotypic classes.
2. The canonical pathways from KEGG database (MSigDB, v5.0)<sup>14,15</sup> and Reactome (MSigDB, v5.0)<sup>53,54</sup> databases were selected as our target pathways.
3. The size of a target pathway ranged from 20 to 200 proteins.
4. The permutation test of type “gene\_set” was performed 1000 times for the target pathway.
5. The pathways with an FDR <25% were reported as significant pathways.

**Protein-protein interaction database.** To identify protein functions behind our experimental data, the protein-protein interaction database STRING (Search Tool for the Retrieval of Interacting Genes)<sup>55</sup> was used to capture the functional modularity and interconnectivity among input proteins. Protein interaction relationship is encoded into networks in the STRING database. Proteins are represented by nodes in the networks and known or predicted interactions are represented by edges. Each edge in the networks is labelled a score in the range between 0 and 1 to represent the confidence level of the associated interaction. In this study, *Homo sapiens* was used as our model organism; those enriched proteins revealed by the GSEA pathway analyses were subjected to STRING v10 to identify interactions among these proteins with a confidence score higher than 0.4.

## 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. Reiland, S., Salekdeh, G. H. & Krijgsveld, J. Defining pluripotent stem cells through quantitative proteomic analysis. *Expert Rev. Proteomics* **8**, 29–42, <https://doi.org/10.1586/epr.10.100> (2011).
3. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123, <https://doi.org/10.1016/j.stem.2009.06.008> (2009).
4. Liu, Y., Cheng, D., Li, Z., Gao, X. & Wang, H. The gene expression profiles of induced pluripotent stem cells (iPSCs) generated by a non-integrating method are more similar to embryonic stem cells than those of iPSCs generated by an integrating method. *Genet. Mol. Biol.* **35**, 693–700, <https://doi.org/10.1590/S1415-47572012005000050> (2012).
5. Munoz, J. *et al.* The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol. Syst. Biol.* **7**, 550, <https://doi.org/10.1038/msb.2011.84> (2011).
6. Kim, S. Y. *et al.* Comparative proteomic analysis of human somatic cells, induced pluripotent stem cells, and embryonic stem cells. *Stem Cells Dev.* **21**, 1272–1286, <https://doi.org/10.1089/scd.2011.0243> (2012).
7. Gudjonsson, T. & Magnusson, M. K. Stem cell biology and the cellular pathways of carcinogenesis. *APMIS* **113**, 922–929, [https://doi.org/10.1111/j.1600-0463.2005.apm\\_371.x](https://doi.org/10.1111/j.1600-0463.2005.apm_371.x) (2005).
8. Volonté, C. & D’Ambrosi, N. Membrane compartments and purinergic signalling: the purinome, a complex interplay among ligands, degrading enzymes, receptors and transporters. *FEBS J.* **276**, 318–329, <https://doi.org/10.1111/j.1742-4658.2008.06793.x> (2009).
9. Ludwig, J. A. & Weinstein, J. N. Biomarkers in cancer staging, prognosis and treatment selection. *Nat. Rev. Cancer* **5**, 845–856, <https://doi.org/10.1038/nrc1739> (2005).
10. Harkness, L. *et al.* Identification of a membrane proteomic signature for human embryonic stem cells independent of culture conditions. *Stem Cell Research* **1**, 219–227, <https://doi.org/10.1016/j.scr.2008.06.001> (2008).
11. Bonardi, F. *et al.* A Proteomics and Transcriptomics Approach to Identify Leukemic Stem Cell (LSC) Markers. *Molecular & Cellular Proteomics* **12**, 626–637, <https://doi.org/10.1074/mcp.M112.021931> (2013).
12. Helbig, A. O., Heck, A. J. R. & Slijper, M. Exploring the membrane proteome—challenges and analytical strategies. *J. Proteomics* **73**, 868–878, <https://doi.org/10.1016/j.jprot.2010.01.005> (2010).
13. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13, <https://doi.org/10.1093/nar/gkn923> (2009).
14. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
15. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–114, <https://doi.org/10.1093/nar/gkr988> (2012).
16. Phanstiel, D. H. *et al.* Proteomic and phosphoproteomic comparison of human ES and iPSC cells. *Nat. Methods* **8**, 821–827, <https://doi.org/10.1038/nmeth.1699> (2011).
17. Yamana, R. *et al.* Rapid and deep profiling of human induced pluripotent stem cell proteome by one-shot NanoLC-MS/MS analysis with meter-scale monolithic silica columns. *J. Proteome Res.* **12**, 214–221, <https://doi.org/10.1021/pr300837u> (2013).
18. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550, <https://doi.org/10.1073/pnas.0506580102> (2005).
19. Mootha, V. K. *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273, <https://doi.org/10.1038/ng1180> (2003).
20. Suárez-Alvarez, B. *et al.* Epigenetic mechanisms regulate MHC and antigen processing molecules in human embryonic and induced pluripotent stem cells. *PLoS One* **5**, e10192, <https://doi.org/10.1371/journal.pone.0010192> (2010).
21. Singh, A. *et al.* Adhesion strength-based, label-free isolation of human pluripotent stem cells. *Nat. Methods* **10**, 438–444, <https://doi.org/10.1038/nmeth.2437> (2013).
22. Villa-Diaz, L. G., Kim, J. K., Laperle, A., Palecek, S. P. & Krebsbach, P. H. Inhibition of Focal Adhesion Kinase Signaling by Integrin  $\alpha 6 \beta 1$  Supports Human Pluripotent Stem Cell Self-Renewal. *STEM CELLS* **34**, 1753–1764, <https://doi.org/10.1002/stem.2349> (2016).
23. Parsons, J. T., Slack-Davis, J., Tilghman, R. & Roberts, W. G. Focal Adhesion Kinase: Targeting Adhesion Signaling Pathways for Therapeutic Intervention. *Clinical Cancer Research* **14**, 627–632, <https://doi.org/10.1158/1078-0432.ccr-07-2220> (2008).
24. Wozniak, M. A., Modzelewska, K., Kwong, L. & Keely, P. J. Focal adhesion regulation of cell behavior. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1692**, 103–119, <https://doi.org/10.1016/j.bbamcr.2004.04.007> (2004).

25. Hewitt, K. J. *et al.* PDGFR $\beta$  expression and function in fibroblasts derived from pluripotent cells is linked to DNA demethylation. *J. Cell Sci.* **125**, 2276–2287, <https://doi.org/10.1242/jcs.099192> (2012).
26. Gupton, S. L., Eisenmann, K., Alberts, A. S. & Waterman-Storer, C. M. mDia2 regulates actin and focal adhesion dynamics and organization in the lamella for efficient epithelial cell migration. *J. Cell Sci.* **120**, 3475–3487, <https://doi.org/10.1242/jcs.006049> (2007).
27. Folmes, C. D. L. *et al.* Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab.* **14**, 264–271, <https://doi.org/10.1016/j.cmet.2011.06.011> (2011).
28. Xu, X. *et al.* Mitochondrial regulation in pluripotent stem cells. *Cell Metab.* **18**, 325–332, <https://doi.org/10.1016/j.cmet.2013.06.005> (2013).
29. Hao, J. *et al.* Reprogramming- and pluripotency-associated membrane proteins in mouse stem cells revealed by label-free quantitative proteomics. *J. Proteomics* **86**, 70–84, <https://doi.org/10.1016/j.jprot.2013.04.018> (2013).
30. Pripuzova, N. S. *et al.* Development of a protein marker panel for characterization of human induced pluripotent stem cells (hiPSCs) using global quantitative proteome analysis. *Stem Cell Res.* **14**, 323–338, <https://doi.org/10.1016/j.scr.2015.01.009> (2015).
31. Varum, S. *et al.* Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One* **6**, e20914, <https://doi.org/10.1371/journal.pone.0020914> (2011).
32. Qin, H. *et al.* Systematic Identification of Barriers to Human iPSC Generation. *Cell* **158**, 449–461, <https://doi.org/10.1016/j.cell.2014.05.040>.
33. Hansson, J. *et al.* Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency. *Cell Reports* **2**, 1579–1592, <https://doi.org/10.1016/j.celrep.2012.10.014>.
34. Yao, J. & Brinton, R. D. Estrogen Regulation of Mitochondrial Bioenergetics. *Advances in Pharmacology* **64**, 327–371, <https://doi.org/10.1016/B978-0-12-394816-8.00010-6> (2012).
35. Spuch, C., Ortolano, S. & Navarro, C. New Insights in the Amyloid-Beta Interaction with Mitochondria. *Journal of Aging Research* **2012**, <https://doi.org/10.1155/2012/324968> (2012).
36. Jin, Y. N. & Johnson, G. V. W. The interrelationship between mitochondrial dysfunction and transcriptional dysregulation in Huntington disease. *Journal of Bioenergetics and Biomembranes* **42**, 199–205, <https://doi.org/10.1007/s10863-010-9286-7> (2010).
37. Venkataramani, V. *et al.* Amyloid Precursor Protein Is a Biomarker for Transformed Human Pluripotent Stem Cells. *The American Journal of Pathology* **180**, 1636–1652, <https://doi.org/10.1016/j.ajpath.2011.12.015> (2012).
38. Metzler, M. *et al.* Huntingtin is required for normal hematopoiesis. *Human Molecular Genetics* **9**, 387–394, <https://doi.org/10.1093/hmg/9.3.387> (2000).
39. Matter, K. & Balda, M. S. Signalling to and from tight junctions. *Nat. Rev. Mol. Cell Biol.* **4**, 225–236, <https://doi.org/10.1038/nrm1055> (2003).
40. Boekema, E. J. & Braun, H.-P. Supramolecular Structure of the Mitochondrial Oxidative Phosphorylation System. *Journal of Biological Chemistry* **282**, 1–4, <https://doi.org/10.1074/jbc.R600031200> (2007).
41. Panopoulos, A. D. *et al.* The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res* **22**, 168–177, <http://www.nature.com/cr/journal/v22/n1/supinfo/cr2011177s1.html> (2012).
42. Zhang, J. *et al.* Michael A. Metabolic Regulation in Pluripotent Stem Cells during Reprogramming and Self-Renewal. *Cell Stem Cell* **11**, 589–595, <https://doi.org/10.1016/j.stem.2012.10.005> (2012).
43. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290, <http://www.nature.com/nature/journal/v467/n7313/abs/nature09342.html#supplementary-information> (2010).
44. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
45. Chen, H. F. *et al.* Derivation, characterization and differentiation of human embryonic stem cells: comparing serum-containing versus serum-free media and evidence of germ cell differentiation. *Hum. Reprod.* **22**, 567–577, <https://doi.org/10.1093/humrep/del412> (2007).
46. Huang, H.-P. *et al.* Factors from human embryonic stem cell-derived fibroblast-like cells promote topology-dependent hepatic differentiation in primate embryonic and induced pluripotent stem cells. *J. Biol. Chem.* **285**, 33510–33519, <https://doi.org/10.1074/jbc.M110.122093> (2010).
47. Huang, H.-P. *et al.* Epithelial cell adhesion molecule (EpcAM) complex proteins promote transcription factor-mediated pluripotency reprogramming. *J. Biol. Chem.* **286**, 33520–33532, <https://doi.org/10.1074/jbc.M111.256164> (2011).
48. Chuang, C.-Y. *et al.* Meiotic competent human germ cell-like cells derived from human embryonic stem cells induced by BMP4/WNT3A signaling and OCT4/EpcAM (epithelial cell adhesion molecule) selection. *J. Biol. Chem.* **287**, 14389–14401, <https://doi.org/10.1074/jbc.M111.338434> (2012).
49. Han, C.-L. *et al.* A multiplexed quantitative strategy for membrane proteomics: opportunities for mining therapeutic targets for autosomal dominant polycystic kidney disease. *Mol. Cell. Proteomics* **7**, 1983–1997, <https://doi.org/10.1074/mcp.M800068-MCP200> (2008).
50. Han, C.-L. *et al.* An informatics-assisted label-free approach for personalized tissue membrane proteomics: case study on colorectal cancer. *Mol. Cell. Proteomics* **10**, M110.003087, <https://doi.org/10.1074/mcp.M110.003087> (2011).
51. Kersey, P. J. *et al.* The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4**, 1985–1988, <https://doi.org/10.1002/pmic.200300721> (2004).
52. Tsou, C.-C. *et al.* IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Mol. Cell. Proteomics* **9**, 131–144, <https://doi.org/10.1074/mcp.M900177-MCP200> (2010).
53. Milacic, M. *et al.* Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers* **4**, 1180–1211, <https://doi.org/10.3390/cancers4041180> (2012).
54. Croft, D. Building models using Reactome pathways as templates. *Methods Mol. Biol.* **1021**, 273–283, [https://doi.org/10.1007/978-1-62703-450-0\\_14](https://doi.org/10.1007/978-1-62703-450-0_14) (2013).
55. Snel, B., Lehmann, G., Bork, P. & Huynen, M. A. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442–3444 (2000).

## Acknowledgements

We would like to thank Hsuan-Cheng Huang for his discussion on this work. This work was financially supported by the Ministry of Science and Technology, Taiwan.

## Author Contributions

W.S.T., S.M.L. and K.P.W. conceived the bioinformatics study. P.M.C., C.Y.C., H.C.K., and Y.J.C. performed the biological experiments. W.S.T. and K.P.W. wrote the main manuscript text and prepared all figures. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-15347-z>.

**Competing Interests:** The authors declare that they have no competing interests.



**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017