

## RESEARCH ARTICLE

# *In vitro* methods to ensure absence of residual undifferentiated human induced pluripotent stem cells intermingled in induced nephron progenitor cells

Hiraku Tsujimoto<sup>1,2\*</sup>, Naoko Katagiri<sup>1,2</sup>, Yoshihiro Ijiri<sup>1,2</sup>, Ben Sasaki<sup>1</sup>, Yoshifumi Kobayashi<sup>2</sup>, Akira Mima<sup>2</sup>, Makoto Ryosaka<sup>1,2</sup>, Kenichiro Furuyama<sup>1</sup>, Yoshiya Kawaguchi<sup>1</sup>, Kenji Osafune<sup>1\*</sup>

**1** Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan, **2** Rege Nephro Co., Ltd., Med-Pharm Collaboration Building, Kyoto University, Kyoto, Japan

\* [osafu@cira.kyoto-u.ac.jp](mailto:osafu@cira.kyoto-u.ac.jp) (KO); [hiraku.tsujimoto@cira.kyoto-u.ac.jp](mailto:hiraku.tsujimoto@cira.kyoto-u.ac.jp) (HT)



## OPEN ACCESS

**Citation:** Tsujimoto H, Katagiri N, Ijiri Y, Sasaki B, Kobayashi Y, Mima A, et al. (2022) *In vitro* methods to ensure absence of residual undifferentiated human induced pluripotent stem cells intermingled in induced nephron progenitor cells. PLoS ONE 17(11): e0275600. <https://doi.org/10.1371/journal.pone.0275600>

**Editor:** Li-Ping Liu, Affiliated Hospital of Jiangsu University, CHINA

**Received:** June 8, 2022

**Accepted:** September 20, 2022

**Published:** November 15, 2022

**Copyright:** © 2022 Tsujimoto et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Kenji Osafune (email: [osafu@cira.kyoto-u.ac.jp](mailto:osafu@cira.kyoto-u.ac.jp)), or the Project Administrator, Hiraku Tsujimoto (email: [hiraku.tsujimoto@cira.kyoto-u.ac.jp](mailto:hiraku.tsujimoto@cira.kyoto-u.ac.jp)). Data for the statistical analysis of Figs 1A, 1D, 1E and 2A and 2B, qPCR following MACS analysis (Figs 3B and 3C and S12) and the ddPCR analysis with different sample concentrations (Fig 4B and 4D) were uploaded to the Center For Open Science (OSF)

## Abstract

Cell therapies using human induced pluripotent stem cell (hiPSC)-derived nephron progenitor cells (NPCs) are expected to ameliorate acute kidney injury (AKI). However, using hiPSC-derived NPCs clinically is a challenge because hiPSCs themselves are tumorigenic. *LIN28A*, *ESRG*, *CNMD* and *SFRP2* transcripts have been used as a marker of residual hiPSCs for a variety of cell types undergoing clinical trials. In this study, by reanalyzing public databases, we found a baseline expression of *LIN28A*, *ESRG*, *CNMD* and *SFRP2* in hiPSC-derived NPCs and several other cell types, suggesting *LIN28A*, *ESRG*, *CNMD* and *SFRP2* are not always reliable markers for iPSC detection. As an alternative, we discovered a lncRNA marker gene, *MIR302CHG*, among many known and unknown iPSC markers, as highly differentially expressed between hiPSCs and NPCs, by RNA sequencing and quantitative RT-PCR (qRT-PCR) analyses. Using *MIR302CHG* as an hiPSC marker, we constructed two assay methods, a combination of magnetic bead-based enrichment and qRT-PCR and digital droplet PCR alone, to detect a small number of residual hiPSCs in NPC populations. The use of these *in vitro* assays could contribute to patient safety in treatments using hiPSC-derived cells.

## Introduction

Currently, kidney transplantation is the only curative treatment for patients with end-stage renal failure by restoring kidney function. Due to the shortage of donor kidneys, however, regenerative medicine is eagerly awaited. A series of recent discoveries reported that human fetal nephron progenitor cells (NPCs) are effective for the treatment of chronic kidney disease (CKD) [1] and that human induced pluripotent stem cell (hiPSC)-derived NPCs can ameliorate acute kidney injury (AKI) in mice [2–4]. These findings have stimulated interest in cell therapies for CKD, in which AKI plays an important role [5]. NPCs are the source of nephron-

repository ([https://osf.io/2dtzr/?view\\_only=eae62786eb84405ad7d2e9913972da1](https://osf.io/2dtzr/?view_only=eae62786eb84405ad7d2e9913972da1)). Since HLA homozygous individuals are good donors for allogeneic transplantation, the ethics committee of our institute has decided that the RNA nucleotide sequence data from such individuals should be treated as personal information; therefore, we could not upload the data to the public database. Instead, the transcript count data are attached as [S1 Dataset](#). The locations of the other RNA sequencing data used in this study are listed in [S4 Table](#). The sequences of recombinant DNA used in this study are listed in [S6 Table](#).

**Funding:** This research is supported by RegeNephro Co., Ltd. (150204400001), by the Japan Agency for Medical Research and Development (AMED) through its research grant 'Research Project for Practical Applications of Regenerative Medicine (22bk0104113h0003)' to K. O. and by iPS Cell Research Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** KO is a founder and member of the scientific advisory boards of iPS Portal Inc. and a founder and chief scientific advisor of RegeNephro Co., Ltd. HT, NK, YK, AM, MR, and YI are employees of RegeNephro Co., Ltd. KO is one of the inventors of the NPC differentiation method from hiPSCs (PCT/JP2018/019886, WO 2018/216743). HT and KO have a patent pending on a method for detecting undifferentiated PSCs. These interests do not alter our adherence to PLOS ONE policies on sharing data and materials. BS, KF, and YK declare no competing interests.

constituent epithelia and form glomeruli and renal tubules by reciprocal interactions with the ureteric bud, a collecting duct progenitor tissue [6, 7]. These NPCs are considered to deliver a variety of nutritional factors that energize failing kidneys and produce direct and local paracrine effects [2].

However, using hiPSC-derived cell products clinically is a challenge because hiPSCs themselves are tumorigenic. Indeed, tumor formation has been associated with many stem cell therapies [8, 9]. Avoiding hiPSC contamination or residual hiPSCs requires a thoroughly controlled cell manufacturing process and assays that ensure the absence of hiPSCs. Several *in vitro* tests for this purpose exist [10–12], including flow cytometry [13], quantitative RT-PCR (qRT-PCR) [13], droplet digital PCR (ddPCR) [14, 15], culture methods for efficient hiPSC growth [15, 16], the detection of marker molecules released into the culture medium [17, 18], and enrichment using magnetic beads [15, 19].

The clinical application of NPCs has another challenge in terms of the absolute number of cells used. If the findings in mice are translated to human dosages, cells on the order of  $10^9$  are needed [2]. In this case, even if a small proportion of residual hiPSCs are present, the absolute number would still risk tumor formation. Therefore, highly sensitive tests are required.

Accordingly, in this study, we discovered a suitable marker, *MIR302CHG*, among many known and unknown iPSC markers, for the detection of residual hiPSCs and developed assays to confirm the absence of residual undifferentiated hiPSCs intermingled in hiPSC-derived NPC populations.

## Results

### Expression of *LIN28A*, *ESRG*, *CNMD* and *SFRP2* in various cell types

Since highly sensitive hiPSC detection methods using *LIN28A*, *ESRG*, *CNMD* and *SFRP2* as hiPSC markers have been reported for retinal pigment epithelial cells [13], dopaminergic progenitor cells [11], cardiomyocytes [14], T cells [15] and liver bud [20], we analyzed the expression of these four genes in public RNA-seq datasets of various human fetal tissues [21] using GREIN [22]. We found similarly low expressions of *LIN28A* in the fetal kidney, liver, and pancreas as in the heart and similarly low expressions of *CNMD* and *SFRP2* in the kidney and pancreas as in the liver ([S1A Fig](#)). Next, we obtained publicly available RNA-seq datasets of these cell types but derived from hiPSCs. We found the expressions of *LIN28A*, *ESRG*, *CNMD* and *SFRP2* varied at different differentiation stages, with some stages expressing relatively high *LIN28A* in hiPSC-derived liver cells [23], pancreatic progenitor cells [24], liver organoids [25], and NPCs [26] ([S1B–S1E Fig](#)) and others expressing high *ESRG*, *CNMD* and *SFRP2* in hiPSC-derived kidney organoids [26] ([S1E Fig](#)). These findings suggested that *LIN28A*, *ESRG*, *CNMD* and *SFRP2* are not suitable markers for detecting residual hiPSCs among several cell types. However, since those datasets are based on tissues produced from hiPSCs in laboratories and the manufacturing process is not as strictly controlled as in a cell conditioning facility, the possibility of hiPSC contamination during the sample preparation cannot be completely ruled out.

### Expression of *LIN28A*, *CNMD* and *SFRP2* in purified NPCs

We next examined the expressions of *Lin28A*, *Cnmd* and *Sfrp2*, which encode proteins in mouse developing kidneys using GUDMAP, a publicly available molecular atlas of gene expressions for developing organs of the genitourinary tract [27], and confirmed relatively high *Lin28A* and *Sfrp2* and *Cnmd* expression in E11.5 metanephric mesenchyme and podocytes, respectively ([S2 Fig](#)). To confirm the expression of *LIN28A*, *CNMD* and *SFRP2* in our induced NPCs, we performed a differentially expressed gene (DEG) analysis of deposited bulk

RNA-seq data for purified hiPSC-derived OSR1-GFP(+)/SIX2-tdTomato(+) NPCs [2, 28] and 201B7 hiPSCs [29, 30] using Salmon [31], which quantifies transcript abundance using cDNA sequences. We used Homo sapiens GRCh38 cDNA gene annotation (Ensembl), which covers all transcripts of Ensembl genes excluding ncRNA. An enrichment analysis using the gene ontology (GO) biological process for DEGs showed that gene sets related to kidney development (cluster B) were upregulated in NPCs, while gene sets related to stem cell maintenance (cluster A) were upregulated in hiPSCs (S3A Fig). These findings suggest that although the sequence datasets of hiPSCs and NPCs were from different origins, the analysis correctly reflected DEGs. We then confirmed that the difference in average scaled TPM values for *LIN28A*, *CNMD* and *SFRP2* between purified hiPSC-derived OSR1-GFP(+)/SIX2-tdTomato(+) NPCs and hiPSCs was not large, although the difference in *LIN28A* and *CNMD1* was statistically significant (Fig 1A).

We also re-analyzed our previously reported single-cell RNA-seq (scRNA-seq) data of purified hiPSC-derived OSR1-GFP(+)/SIX2-tdTomato(+) NPCs [29] and found that the cells expressing these genes are distributed unbiasedly compared to the distribution of the cells expressing NPC marker genes in the dataset (Figs 1B and S4). An analysis of correlation between the expression of these genes and NPC markers (*SIX2* and *PAX2*) showed weak or almost no correlation (S4F Fig). Thus, these data suggest that no specific NPC populations expresses of *LIN28A*, *CNMD* and *SFRP2*. The qRT-PCR analysis also showed non-negligible *LIN28A* expression when we induced NPCs using an HLA homozygous hiPSC line, Ff-I 14s04 (S3B–S3D Fig). Therefore, to detect hiPSCs intermingled in NPCs populations with high sensitivity, we sought new genetic markers that are highly expressed in hiPSCs compared to NPCs.

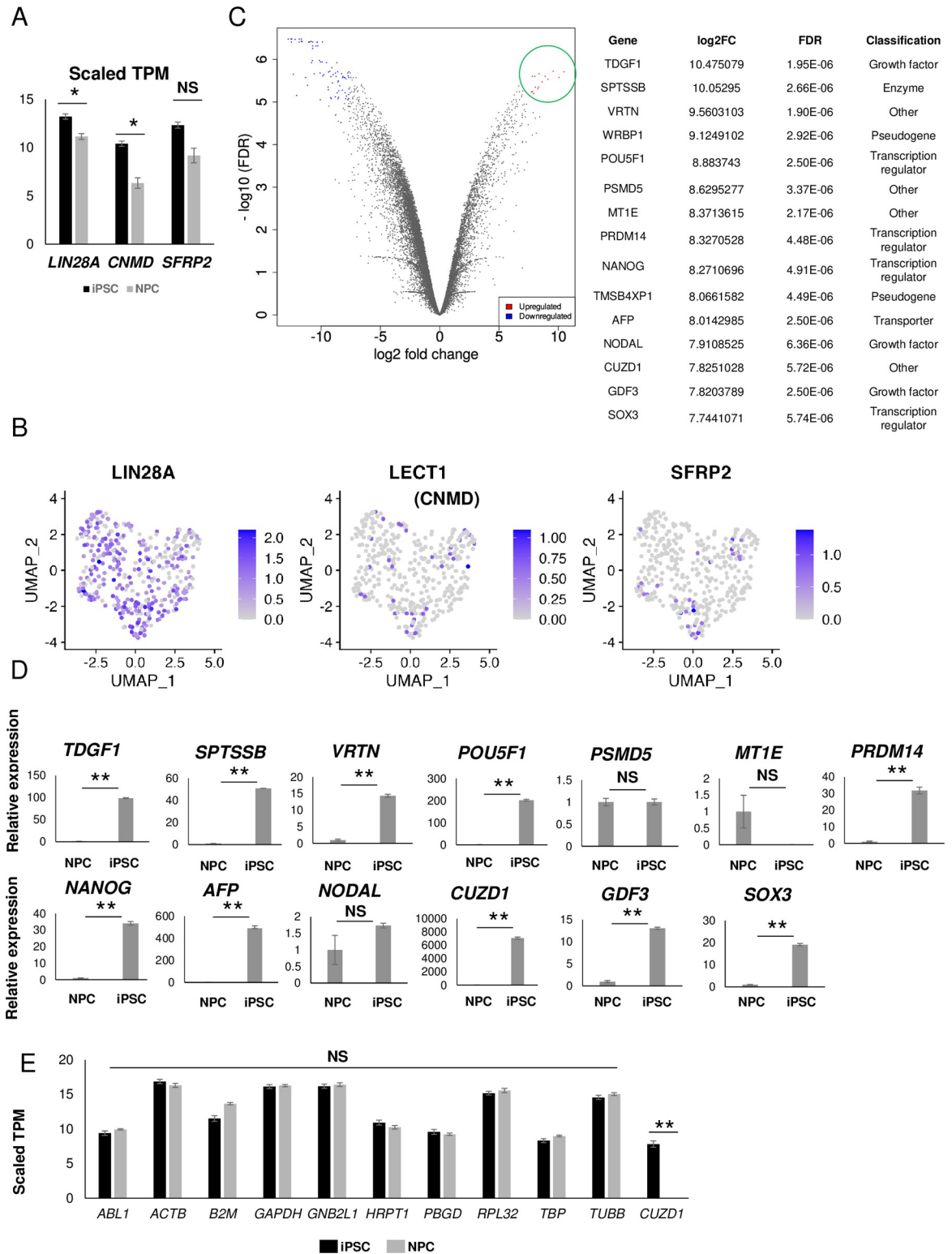
### Identification of genes highly expressed in hiPSCs compared to NPCs

We then analyzed the aforementioned bulk RNA-seq datasets of purified hiPSC-derived OSR1-GFP(+)/SIX2-tdTomato(+) NPCs and hiPSCs and extracted 15 genes with particularly high expression values ( $FC > 200$  &  $FDR < 10^{-5}$ ; Fig 1C) in hiPSCs. We selected 13 protein-coding genes and performed a qRT-PCR analysis to compare the expressions between NPCs and hiPSCs using the QHJI 14s04 iPSC line, which was derived from a clinical-grade iPSC bank [32]. Among them, *CUZD1* was markedly and differentially expressed in hiPSCs (Fig 1D). For a housekeeping gene, we selected *TBP*, which is stably expressed in NPCs and hiPSCs and similar to *CUZD1* in scaled TPM expression in hiPSCs (Fig 1E).

### Detection of undifferentiated hiPSCs in NPCs using *CUZD1* copy numbers determined by ddPCR

Next, to create a hypothetical situation where NPCs are contaminated with hiPSCs, we manually contaminated the NPCs with hiPSCs. By generating hiPSCs constitutively expressing a red fluorescent protein, tdTomato, we confirmed that the percentages calculated from the cell count and the actual percentage measured by flow cytometry were almost identical (S5A Fig).

Next, we performed one-step ddPCR [14] on hiPSCs mixed at a ratio of  $1:10^4$  with NPCs. However, we could not achieve stable and sensitive detection (S5B Fig). We considered RNA degradation during the droplet preparation as the reason. Therefore, we switched to a two-step ddPCR, in which cDNA is used during the droplet preparation and PCR reaction and further normalization using a housekeeping gene is performed. For the housekeeping gene, we selected *TBP*, which shows relatively low expression, to reduce possible artifacts of the multiplex PCR reaction (Fig 1D). As a result, while the  $1:10^4$  ratio might be sufficient to discriminate hiPSCs according to the estimated copy number ratio (estimated *CUZD1* copy number/estimated *TBP* copy number) evaluated by ddPCR, we expected slight experimental errors would



**Fig 1. Identification of markers highly expressed in human induced pluripotent stem cells (hiPSCs) compared to purified OSR1(+) SIX2(+) induced nephron progenitor cells (NPCs).** (A) Scaled TPM values of *LIN28A*, *CNMD* and *SFRP2* in hiPSCs and NPCs.

(B) Uniform Manifold Approximation and Projection (UMAP) plots for *LIN28A*, *LECT1 (CNMD)*, *SFRP2* expression in the scRNA-seq data of purified OSR1(+)SIX2(+) NPCs reported in Tsujimoto et al. (2020). (C) A volcano plot (left panel) of differentially regulated genes at  $\log_{10}(\text{FDR}) > 5$  and  $\log_2(\text{fold change}) < -7$  (blue dots) or  $> 7$  (red dots) and a table (right panel) of 15 candidate iPSC markers

as genes with  $\log_{10}(\text{FDR}) > 5$  and  $\log_2(\text{fold change}) > 7$  (green circle). The gene classification by IPA is shown. (D) qRT-PCR analysis of the 13 candidate protein-coding iPSC markers. Each value was normalized to that of NPCs. (E) Scaled TPM values of 10 housekeeping genes and *CUZD1* in hiPSCs (black bars) and NPCs (grey bars). Data are represented as the mean  $\pm$  SEM ( $n = 3$ ) in (A), (D) and (E). \* $p < 0.05$  and \*\* $p < 0.01$  by paired Student's t-tests against iPSCs with the Bonferroni correction. NS, not significant.

<https://doi.org/10.1371/journal.pone.0275600.g001>

make them indistinguishable (S5C Fig). To further clarify the sensitivity of the ddPCR assay, we performed a ddPCR analysis using cDNAs derived from mixtures of hiPSCs and NPCs at various ratios ( $10^{-6}$ :1 to 1:1). The *TBP*-normalized *CUZD1* expression pattern suggested limited assay performance when the iPSC proportion was low (S5D Fig). Therefore, we sought the differentiation stage when *CUZD1* expression was lowest, finding it was in day 4 cells (S6 Fig). Therefore, we also performed the assay using day 4 cells, but found only limited improvement (S5E Fig).

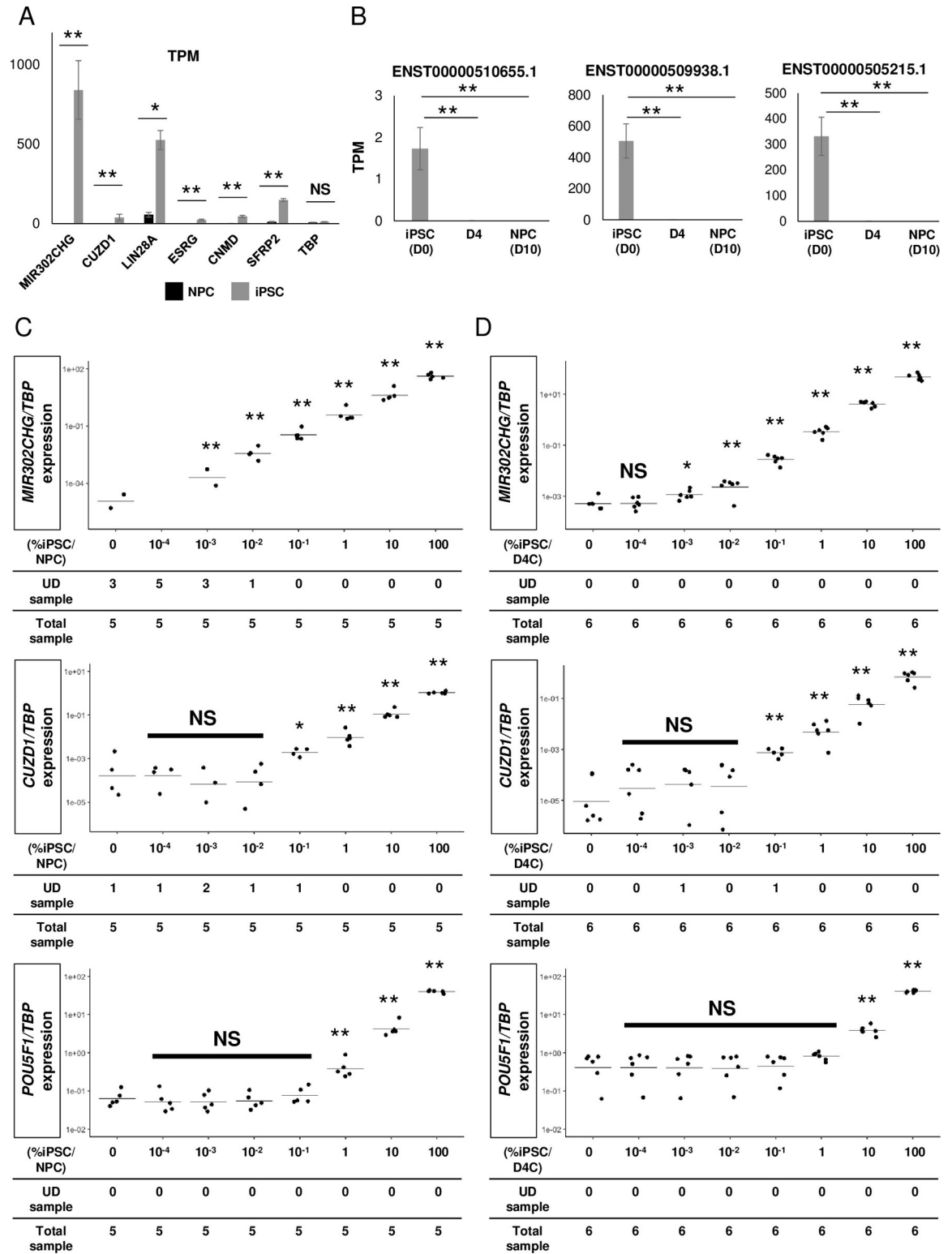
### Identification of lncRNA that has super-high differential expression in hiPSCs compared to NPCs or day 4 cells

Based on the results of the ddPCR, we suspected that NPCs had a trace expression of *CUZD1*. Publicly available RNA-seq datasets of 27 different human tissues also showed a trace expression of *CUZD1* in several tissues including the kidney (S7A Fig) [33]. Thus, to identify markers that have a high expression in hiPSCs, but not in NPCs, we performed RNA-seq with increased sequencing depth. In addition, we applied a genomic mapping method using STAR [34] to extend our analysis to markers that do not encode proteins. The average leads per sample was 70M. We identified 26 extremely highly expressed genetic markers ( $\log_2\text{FC} > 10$  &  $\text{FDR} < 10^{-5}$ ) (S7B Fig) and found a pluripotency-associated lncRNA, *MIR302CHG* [35], exhibited higher TPM values in hiPSCs than either *LIN28A* or *CUZD1* (Fig 2A). Unexpectedly, we found *MIR302CHG* was highly differentially expressed in hiPSCs compared to both NPCs (S7B Fig) and day 4 cells from the NPC differentiation protocol ( $\log_2\text{FC}$ : 15.9 &  $\text{FDR}$ :  $8.9 \times 10^{-30}$ ; S1 Table). *MIR302CHG* is highly expressed not only in the QHJI 14s04 line but also in several hiPSC lines, including 1231A3 [36], 1383D6 [36], 201B7 [37] and 317-12 [38] (S7C Fig). An isoform analysis revealed that hiPSCs, human embryonic stem cells (hESCs), anterior primitive streak (PS) and definitive endoderm (DE) highly expressed splice variants 2 (ENST00000509938.1) and 3 (ENST00000505215.1) of *MIR302CHG* compared to differentiated cells such as pancreatic progenitors, foregut, hepatocytes, liver organoids, neural progenitors, neurons, cardiomyocytes, lung progenitors, lung organoids and endothelial cells (Figs 2B and S7D and S8–S10). Fortunately, since this lncRNA has a poly(A) 3' tail (RNA central, release 19), we could perform a reverse transcription (RT) reaction using oligo deoxythymine (dT). Next, we performed qRT-PCR using a mixture of cDNAs derived from hiPSCs and NPCs or day 4 cells at various percentages (100% to  $10^{-4}\%$ ). The *TBP*-normalized *MIR302CHG* expression patterns showed better performance than the expression of *CUZD1* or *POU5F1* when the hiPSC percentages were low (Fig 2C and 2D and S1 Data). However, we sporadically obtained undermined CT values at low hiPSC percentages, suggesting a need for further improvement.

### Improvement of the qRT-PCR assay by magnetic bead-based enrichment

To overcome the difficulty in detecting a very small amount of hiPSCs, we tried iPSC enrichment using the hiPSC culture assay [15]. However, the assay was not suitable for NPCs because NPCs can grow in hiPSC culture media (Essential 8 or AK03N) (S11 Fig).

Alternatively, we tested hiPSC enrichment using magnetic beads conjugated with antibodies against two well-known iPSC surface antigens, TRA-1-60 and SSEA-4 [15]. Using cell



**Fig 2. Identification of markers highly expressed in hiPSCs compared to NPCs derived from a clinical-grade iPSC stock line.** (A) TPM values of *MIR302CHG*, *CUZD1*, *LIN28A*, *ESRG*, *CNMD*, *SFRP2* and *TBP* in hiPSCs and NPCs show the high absolute expression of *MIR302CHG*. Data are represented as the mean  $\pm$  SEM (n = 3). (B) TPM values of three *MIR302CHG* splice variants, *ENST00000510655.1*, *ENST00000509938.1* and *ENST00000505215.1*, in hiPSCs and NPCs. Data are represented as the mean  $\pm$  SEM (n = 3). (C, D) Scatter plots of the *TBP*-normalized gene expression of *MIR302CHG*, *CUZD1* and *POU5F1* by qRT-PCR for mixtures of

hiPSCs and NPCs (C) or day 4 cells (D). The dots and lines in the center of the scatter plots indicate the experimental data and mean values of the data, respectively. UD: undetermined; D4C: day 4 cell. \* $p < 0.05$  and \*\* $p < 0.01$  by paired Student's t-tests against iPSCs with the Bonferroni correction (A, B) and by Tukey-Kramer post hoc tests against the sample with 0% iPSCs (C, D). NS, not significant.

<https://doi.org/10.1371/journal.pone.0275600.g002>

mixtures containing hiPSCs constitutively expressing EGFP (317–12 cells) [38] and NPCs derived from a non-fluorescent hiPSC line (QHJI 14s04), we evaluated the degree of magnetic separation of the positive and negative fractions by FACS using GFP as an indicator. As a result, we successfully enriched hiPSCs in NPC populations using beads conjugated with antibodies against TRA-1-60 but not SSEA-4 (Fig 3A).

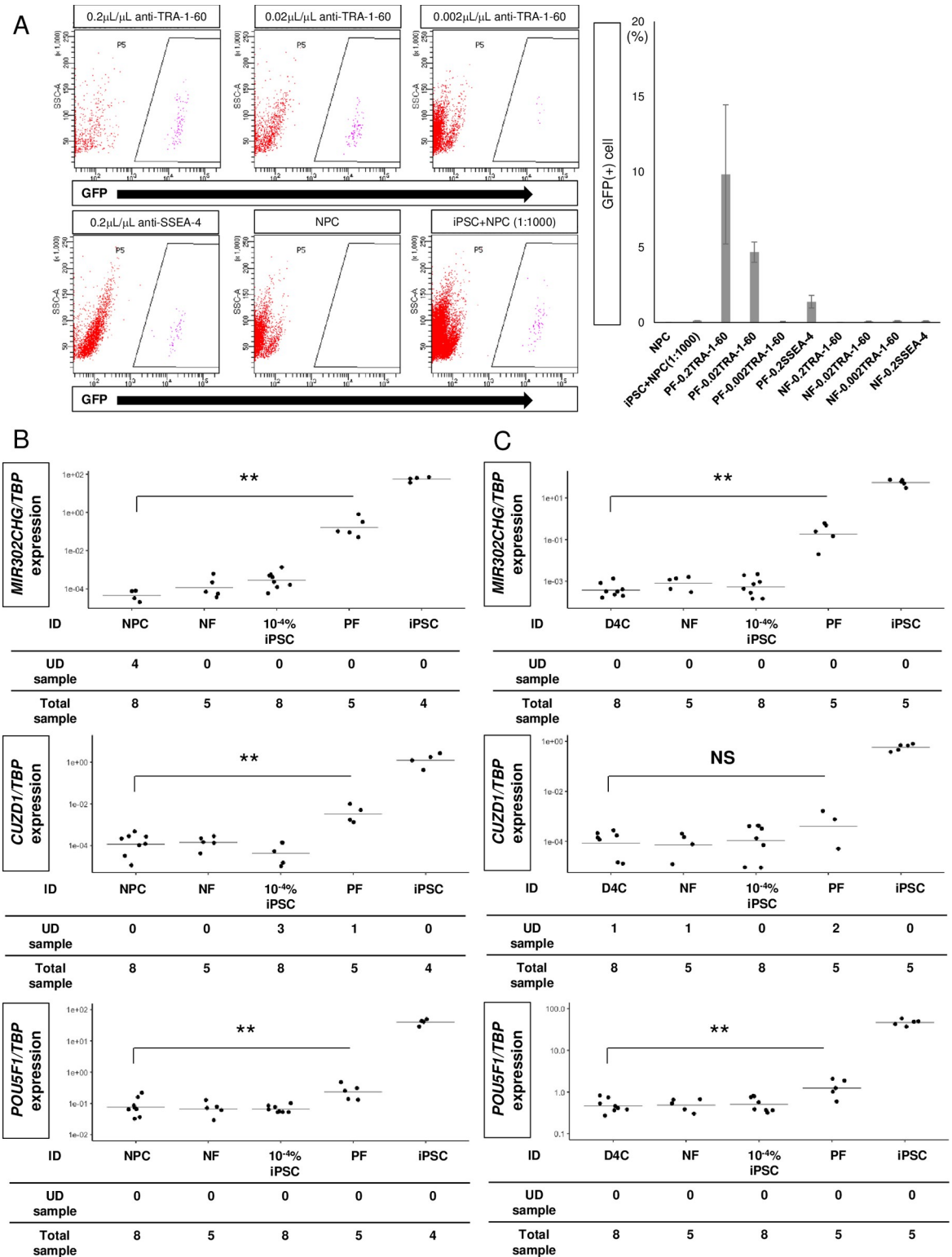
Next, we prepared cell mixtures containing  $1-3 \times 10^2$  hiPSCs in  $1-3 \times 10^8$  NPCs or day 4 cells (0.0001%) and enriched hiPSCs with anti-TRA-1-60 antibody-conjugated magnetic beads and confirmed a marked increase in the *MIR302CHG/TBP* ratio upon qRT-PCR analysis (Fig 3B and 3C and S1 Data). We also confirmed that the expression values of *MIR302CHG/TBP* were below detection level in cDNAs derived from the positive fractions when NPCs alone were treated with the magnetic beads (S12 Fig). These results confirmed high sensitivity for hiPSC detection (0.0001% hiPSCs in NPCs) by qRT-PCR for *MIR302CHG* following MACS enrichment using anti-TRA-1-60 antibody-conjugated beads.

### Improvement of the qRT-PCR assay by ddPCR

Although magnetic bead-based enrichment improves the sensitivity, the enrichment needs to be performed on the same day of the cell harvesting. Therefore, we also developed a method that can be performed only with cDNAs that can be stably cryopreserved. Using two-step ddPCR with the PCR parameters suggested by the manufacturer, we confirmed the *MIR302CHG/TBP* ratios evaluated by the ddPCR assay were generally proportional to the cDNA concentration ratios of hiPSCs in the NPC populations (Fig 4A). By optimizing the temperature (Fig 4B) and the loading amount of RT products (Fig 4C–4E and S1 Data), further improvement of the hiPSC detection was achieved, such that we could detect hiPSCs making up 0.001% of NPC populations without the magnetic bead-based enrichment (Fig 4D).

### Discussion

In this study, we examined protein-coding genes and comprehensively compared their gene expressions in hiPSCs and NPCs. By experimentally examining their usefulness as hiPSC markers, we showed that *CUZD1* did not perform as expected, probably due to its low absolute expression in hiPSCs. However, we confirmed experimentally that the lncRNA *MIR302CHG*, a non-protein encoding gene and known undifferentiated hESC marker in the *MIR302/367* cluster host gene [39, 40], is a good marker. Previous reports have shown that the transduction of *MIR302/367* using lentivirus vectors more efficiently reprograms human fibroblasts to the pluripotent state than the transduction of OCT3/4 (POU5F1), SOX2, KLF4 and MYC [35, 37] and that the direct transfection of a combination of *MIR200c* plus *MIR302s* and *MIR369s* family miRNAs reprograms human adipose stromal cells and fibroblasts to pluripotency [41]. Even for lncRNAs that do not encode proteins including *MIR302CHG* transcripts, a cDNA library can be synthesized and purified by RNA degradation after a RT reaction using poly dT sequences by taking advantage of the poly(A) 3' tail. Accordingly, we discovered *MIR302CHG* is a suitable marker for the detection of hiPSCs intermingled in hiPSC-derived NPC products. Using *MIR302CHG*, we developed highly sensitive assays to detect residual hiPSCs using two approaches. Both of these approaches ensured the absence of residual undifferentiated hiPSCs intermingled in hiPSC-derived NPC populations.



**Fig 3. Combination assay using magnetic bead-based cell isolation and qRT-PCR analysis to detect hiPSCs intermingled in NPCs derived from a clinical-grade iPSC stock line.** (A) Flow cytometric analysis of the MACS positive fractions of 1:1,000 mixtures of GFP (+) hiPSCs and GFP (-) NPCs at three different concentrations of anti-TRA-1-60 antibody-conjugated beads (upper panels) and one concentration of anti-SSEA-4 antibody-conjugated beads (lower left panel), GFP (-) NPCs (without MACS selection; lower center panel), and a 1:1,000 mixture of GFP (+) hiPSCs and GFP (-) NPCs (without MACS selection; lower right panel). The bar graph in the right panel



shows the mean  $\pm$  SEM of GFP (+) cells from the flow cytometric analysis at various bead concentrations in staining reagent. PF: positive fraction; NF: negative fraction; 0.2TRA-1-60: 0.2  $\mu$ L TRA-1-60 beads / 1  $\mu$ L staining reagent; 0.02TRA-1-60: 0.02  $\mu$ L TRA-1-60 beads / 1  $\mu$ L staining reagent; 0.002TRA-1-60: 0.002  $\mu$ L TRA-1-60 beads / 1  $\mu$ L staining reagent; 0.2SSEA-4: 0.2  $\mu$ L SSEA-4 beads / 1  $\mu$ L staining reagent. (B) Scatter plots of the *TBP*-normalized gene expressions of *MIR302CHG*, *CUZD1* and *POU5F1* in NPCs (NPC), the MACS negative fraction of the mixture (NF), 0.0001% hiPSC/NPC mixtures ( $10^{-4}$ % iPSC), the MACS positive fraction of the mixture (PF), and hiPSCs (iPSC) by qRT-PCR analysis. (C) Scatter plots of the *TBP*-normalized gene expressions of *MIR302CHG*, *CUZD1*, and *POU5F1* in day 4 cells (D4C), NF,  $10^{-4}$ % iPSC, PF and iPSC by qRT-PCR. The dots and lines in the center of the scatter plots in (B) and (C) indicate the experimental data and mean values of the data, respectively. UD: undetermined. \* $p < 0.05$  and \*\* $p < 0.01$  by Tukey-Kramer post hoc tests against the sample with NPCs. NS, not significant.

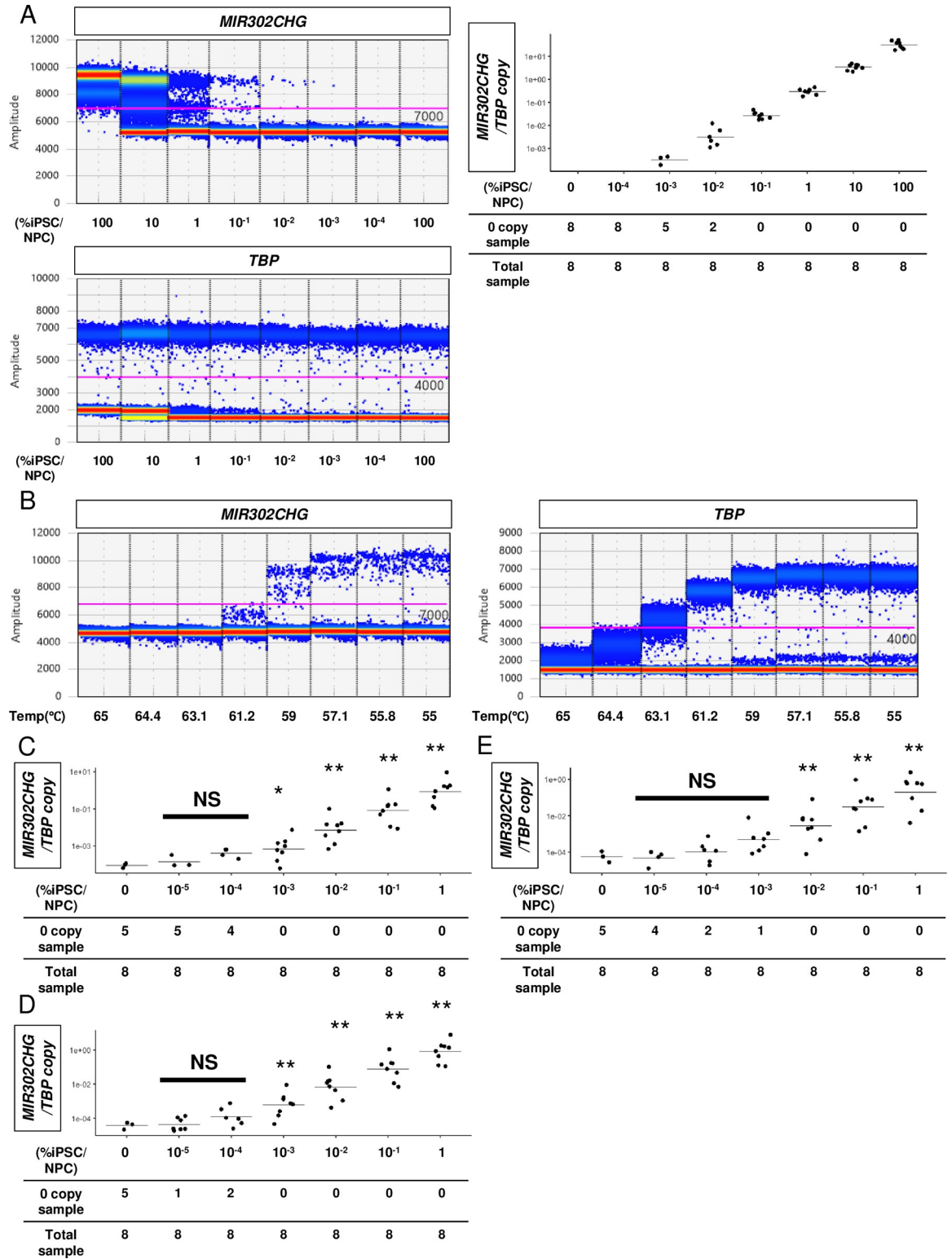
<https://doi.org/10.1371/journal.pone.0275600.g003>

TRA-1-60 is commonly used to assess the pluripotency of hiPSCs and hESCs. The TRA-1-60 antibody recognizes the type 1 lactosamine structure of podocalyxin [42, 43]. This glycan structure is different from that of podocalyxin expressed in the podocytes of kidneys [44]. Around 90% of the hiPSCs we used express the TRA-1-60 antigen (QHJI 14s04 Cell information). Although the discovery of novel hiPSC glycoproteins is in progress [45], the direct relationship between these antigens and tumorigenicity is unknown. Therefore, a detection method by ddPCR without using surface antigen enrichment is currently preferred.

In a previous report, a combination of magnetic bead-based enrichment and efficient culture assay could detect 10 (0.00002%) hiPSCs spiked into  $5 \times 10^7$  T cells [15]. Although the hiPSC culture assay is not suitable for NPCs because NPCs can grow in hiPSC culture media (Essential 8 or AK03N) (S11 Fig), by substituting the culture assay with a qRT-PCR assay, we achieved a comparably high sensitivity (0.0001% hiPSCs).

However, our method has several limitations. First, it is an *in vitro* assay and therefore should be combined with an *in vivo* tumorigenesis test. Ideally, future studies should investigate the minimum permissible level of the *MIR302CHG/TBP* ratio and the probability of tumorigenesis *in vivo* in large animals at doses similar to those in humans. Second, our study experimentally tested only hiPSC-derived NPCs and not absolute negative control cells, such as human embryonic NPCs. However, we are confident that our tested hiPSC-derived NPCs did not contain residual hiPSCs at levels higher than the desired sensitivity based on the results of the artificial contamination tests (Figs 3 and 4 and S12). The possible reason why we sporadically observed positive droplets even in samples without hiPSC addition (Figs 4 and S12) is that DNA fragments in the environment or residual DNA fragments after the DNase reaction could have been contaminated. Further improvement could be achieved by oligo dT bead purification and standardization of the loading amount for ddPCR using cDNA concentrations only derived from RNA with poly(A) 3' tail. Third, we did not examine a method to remove the detected hiPSCs. Previous reports showed ways to eliminate hiPSCs intermingled in cardiomyocytes, neurons and hepatocytes using Orlistat [46] or Atorvastatin [47]. Future studies should develop such a method for NPCs. Fourth, our method cannot detect malignantly transformed cells associated with cultures, which are tumorigenic cells that form in cell preparations. Fifth, although we standardized and described the assay procedures in detail, the ddPCR and qRT-PCR results presented here are from a single institution (CiRA) and external validation at other institutions is also necessary. Finally, the sample size we used was small. The statistical analyses in Figs 3B and 3C and 4C and 4D tested whether the samples without hiPSCs show different values compared to the samples of interest. In future work, a cutoff value for the diagnosis of hiPSC contaminants should be established, and the sensitivity and specificity of the test at the cutoff value should be considered to guarantee the safety of the cell product. If these tests are to be used industrially, further validation will be necessary due to fluctuations in the manufacturing at cell conditioning facilities and the need to standardize the assays.

Despite the above, the use of the described *in vitro* assays will contribute to patient safety in treatments using hiPSC-derived cells.



**Fig 4. Development of a digital droplet PCR assay for hiPSCs intermingled in NPCs derived from a clinical-grade iPSC stock line.** (A) Scatter plots of positive and negative droplets for digital droplet PCR (ddPCR) in a hiPSC/NPC cDNA dilution series sample. The purple lines indicate the cutoff values (left panels). A ddPCR analysis of *MIR302CHG* using an hiPSC cDNA dilution series sample in NPC cDNA shows a concentration-dependent change in the *TBP*-normalized copy number of *MIR302CHG* (right panel). (B) A representative result of ddPCR at annealing temperatures of 55°C to 65°C using 0.1% hiPSC/NPC cDNA samples shows a better

separation of the positive and negative droplets of *MIR302CHG* and that *TBP* varies at an annealing temperature below 57.1 °C. (C-E) Scatter plots of the ddPCR-estimated copy number ratios of *TBP*-normalized *MIR302CHG* using 5% (C), 15% (D) and 30% (E) of the RT products in the ddPCR reaction mix. \* $p < 0.05$  and \*\* $p < 0.01$  by the Tukey-Kramer post hoc tests against the sample with 0% hiPSCs. NS, not significant.

<https://doi.org/10.1371/journal.pone.0275600.g004>

## Materials and methods

### Ethics statement

Experiments using hiPSCs were approved by the ethics committee of the Center for iPS Cell Research and Application (CiRA), Kyoto University and performed according to the guidelines of the Declaration of Helsinki. No donors were from a vulnerable population, and all donors provided written informed consent that was freely given. We used 9 hiPSC lines including 5 parental hiPSCs (201B7, QHJI 14s04, 585A1, 1231A3 and 1383D6) and their offspring hiPSCs (4A6, 317–12, Ff-I 14s04 and 585A1-tdTomato). The lines 201B7, 1231A3 and 1383D6 were established from human primary cells purchased from commercial cell distributors (201B7: Cell Applications, Inc., USA; 1231A3 and 1383D6: Cellular Technology Ltd, USA) in CiRA and widely distributed for biomedical research (e.g., RIKEN BioResource Research Center, <https://web.brc.riken.jp/en/>) [36, 37, 48]. 201B7 and the other 4 parental hiPSCs were established from dermal fibroblasts and peripheral blood mononuclear cells (PBMCs), respectively.

The PBMC donor recruitment was conducted from June 2008 to present including the donor of 585A1 and from January 2013 to July 2019 including the QHJI donor. Written informed consent was obtained from the QHJI donor (older than 20 years) from whom HLA homozygous hiPSCs were derived as part of a large initiative named ‘Research on an HLA-Homozygous Donor-Derived iPS Cell Stock from Apheresis Donors for Regenerative Medicine’ (S1 File) [32, 49] and from an adult healthy donor (30–40 years old) from whom 585A1 (RIKEN BRC: HPS0354) and 585A1-tdTomato (this study) were derived as a part of another large initiative named ‘Genetic Analysis Study Using Human Disease-Specific iPS Cells’ and ‘The Generation of Human Disease-Specific iPS Cells and the Use of Such iPS Cells for Disease Analysis’ (S1 File) [50].

The donors of 585A1 and QHJI and their families did not receive monetary benefits. However, necessary research-related expenses were covered by the research fund of the initiatives. For the donors of human primary cells, which we purchased from commercial cell distributors (201B7, 1231A3 and 1383D6), no information about monetary reward was provided.

### Cell culture

hiPSCs were maintained with feeder-free cultures using Stem Fit AK02N medium (Ajinomoto) or Stem Fit AK03N medium (Ajinomoto) on cell culture plates coated with iMatrix [51]. The cells were passaged using 0.5 mM EDTA/PBS (Thermo Fisher Scientific) every four or five days and routinely tested for mycoplasma contamination. Details about the differentiation of hiPSCs towards NPCs are described elsewhere [29].

For the NPC growth assay using iPSC culture media, we used the OSR1-GFP/SIX2-tdTomato reporter iPSC-derived NPCs. NPCs were seeded into 96-well low cell-binding U-bottom plates (Thermo Fisher Scientific) at a density of  $4.0 \times 10^4$  NPCs/well with serum-free differentiation medium consisting of DMEM/F12 Glutamax (Thermo Fisher Scientific), B27 supplement minus vitamin A (Thermo Fisher Scientific) and 0.5×Penicillin/streptomycin (hereafter called basal medium) containing 1  $\mu$ M CHIR99021 (Wako), 100 ng/mL FGF9 (Peprotech) and 10  $\mu$ M Y-27632 to form aggregates. After spinning down the plate for 3 min at 200g, the cells

were incubated for 24 h at 37°C. After the 24 h, the aggregates were subjected to iPSC growth media, Essential 8 medium (Thermo Fisher Scientific) or Stem Fit AK03N medium. The medium was changed every 1–2 days. We harvested NPC aggregates on days 1, 3, 5, 8 and 10. The aggregates were washed with PBS (-), incubated with Accumax (Innovative cell technologies) for 5–10 min at 37°C and dissociated into single cells by pipetting. The number of cells in the aggregates was counted using a TC20 Automated Cell Counter (Bio-Rad). The fluorescent images of days 0, 1 and 10 aggregates were captured using a BZ-X700 (KEYENCE).

### RT-PCR and real-time quantitative RT-PCR (qRT-PCR)

Total RNA was isolated using the RNeasy Mini kit (QIAGEN) according to the manufacturer's recommendations followed by cDNA synthesis using standard protocols for ReverTra Ace (TOYOBO). qRT-PCR was performed using the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific) and TB Green Premix Ex Taq II (Tli RNaseH Plus), ROX plus (Takara). Denaturation was performed at 95°C for 30 s, followed by 45 cycles at 95°C for 5 s and at 60°C for 34 s. Comparative Ct method quantification was used to analyze the data for the gene expression levels, and the values were normalized to those of *TBP* (housekeeping gene). The PCR reactions were performed in at least triplicate for each sample. The primer sequences are listed in [S2 Table](#). Student's t-tests against iPSCs with the Bonferroni correction based on the number of comparisons (genes) and zero imputation to the undetermined CT value were performed in the statistical analysis in [Fig 1D](#). Statistical analyses in [Figs 2C and 2D](#) and [3B and 3C](#) were performed using log<sub>10</sub> transformed target expression values normalized to *TBP* expression values. Samples for which the CT values were not determined were removed from the analyses. One-way ANOVA and Tukey-Kramer post hoc tests were performed for multiple group comparisons.

### Immunostaining

NPC aggregates and kidney organoids were fixed with 4% PFA (Nacalai tesque)/PBS(-) overnight at 4°C. The fixed aggregates and organoids were washed with PBS(-) twice, treated with 30% sucrose (Nacalai tesque)/PBS(-) overnight at 4°C, and then frozen with OCT compound (Sakura Finetek). The frozen sections were incubated with Blocker Casein in PBS (Thermo Fisher Scientific) for 30 min at room temperature. Primary and secondary antibodies were incubated with the samples overnight at 4°C. The antibodies and lectins used in this study are listed in [S3 Table](#).

### RNA sequencing

Total RNA was extracted using the RNeasy Mini Kit. We prepared sequencing libraries using TruSeq Stranded Total RNA (Illumina). The library was sequenced using the NovaSeq SP Reagent Kit v1.5 (200 cycles) with 101-8-8-101 cycles. Gene expression quantification was done using the analysis pipeline (2.3.4) used in the ENCODE project (<https://www.encodeproject.org/pipelines/ENCPL002LPE/>). GRCh38 ENSEMBL release 104 was used for the reference sequence. Gene definitions were based on GRCh38 GENCODE release 38. We performed a comparative analysis of hiPSCs and NPCs using DESeq2 [52] with the expected count calculated by the above pipeline (STAR-RSEM). Previously reported RNA-seq data were also re-analyzed ([S4 Table](#)). For NPCs [29] and hiPSCs (201B7) [30], transcript quantification was performed using Salmon [31] with Homo sapiens GRCh38 cDNA gene annotation (Ensembl). The output from Salmon was then processed using the R/Bioconductor package tximport to get gene expression values. DEG analysis was performed using iDEP (version 0.92) [53]. Enrichment (GO Biological Process) analysis for the k-means-clustered top 1,000

DEGs was performed using iDEP. We used GREIN [22] to obtain gene-level normalized counts per million (CPM) or (transcript-level) transcripts per million (TPM) from the RNA-seq data of the human fetal transcriptional atlas [21], hepatocyte-like cells [23], pancreatic progenitors [24], liver organoids [25], kidney organoids [26], endoderm lineages [54], three germ layer lineage cells [55], neuron lineages [56], cardiomyocytes [57, 58], lung lineages [59, 60], endothelial cells [58], hiPSCs [23, 24, 56–58] and hESCs [55, 58, 60]. We obtained the *CUZD1* read per million (RPKM) values of 27 types of organs and tissues from 95 human individuals [33] in the NCBI Gene database (Gene ID: 50624). The sum of the Scaled TPM (Fig 1A and 1D) or TPM (Fig 2A and 2B) plus 0.01 was subjected to log<sub>10</sub> transformation and then used for the statistical analyses in Figs 1A and 1D and 2A and 2B. Student's t-tests against hiPSCs with the Bonferroni correction based on the number of comparisons (genes or transcripts) were performed in the statistical analyses in Figs 1A and 1D and 2A and 2B.

### Single cell RNA sequencing

We obtained processed scRNA-seq data (count matrix) of our NPCs induced with or without activin A treatment at stage 4 from a public database (GSE146119), and alignment to the reference genome hg19, filtering, debarcoding and UMI counting was conducted using the Cell Ranger v2.1.0 pipeline (10X Genomics) [29].

We performed Seurat-based (Seurat v4.1.1) filtering using three criteria: number of detected features (nFeature\_RNA) per cell, number of UMIs expressed per cell (nCount\_RNA) and percent of mitochondrial gene count (percent.mt) using the following threshold parameters: nFeature\_RNA (500 to 5000) and percentage of mitochondrial genes expressed (<5%) to remove multiplets (S4A Fig) [61].

We normalized our dataset using the SCTransform [62] framework with the percent of mitochondrial gene count to regress out in a second non-regularized linear regression. Then, we performed principal component (PC) analysis and determined the K-nearest neighbor graph using the first 30 PCs. The origin of the NPC dataset was visualized on a UMAP (RunUMAP, dims = 1:30). To visualize the expression of known marker genes for each cell on UMAP plots, we used the FeaturePlot function in Seurat. To visualize the expression of known markers in each NPC dataset, we used the VlnPlot function in Seurat. An analysis of correlation between the expression of hiPSC markers (*LIN28A*, *CNMD*, and *SFRP2*) and NPC markers (*SIX2* and *PAX2*) was performed using FeatureScatter function in Seurat.

### One-step RT-ddPCR

Total RNA was extracted using the RNeasy Mini Kit. One-Step RT-ddPCR was performed using a One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad), 900 nM forward and reverse primers, 250 nM probes labeled with FAM dye and ZEN/Iowa Black fluorescent quencher, and 1 mM manganese acetate solution (Bio-Rad), as previously described [14] with slight modifications. A total RNA sample (1 ng) was added to the mixture. The available sequences of primers and probes used in the present study are listed in S5 Table. Droplets were generated in 8-well cartridges using the QX200 droplet generator (Bio-Rad) or Automated Droplet Generator (Bio-Rad) according to the manufacturer's instructions. The water-in-oil emulsions were transferred to a 96-well plate, and RT-PCR was performed using a C1000 Touch Thermal Cycler (Bio-Rad). The thermal cycling conditions were as follows: 30 min reverse transcription at 60°C, followed by 5 min enzyme activation at 95°C, and 40 cycles of a thermal profile comprising 30 s denaturation at 94°C and 60 s annealing/extension at 63°C. After the PCR amplification, the products were denatured for 10 min at 98°C and cooled at 4°C until the fluorescence intensity was measured. Fluorescence intensities of each droplet from the samples

were measured using the QX200 droplet reader. Positive droplets containing amplification products were distinguished from negative droplets and counted by applying a fluorescence amplitude threshold in QuantaSoft software. The threshold was manually determined based on the distribution of the droplet and set at 4000. QuantaSoft software provides concentration results in copies of target per microliter (copies/ $\mu$ L). The number of copies of target per template RNA was calculated as a concentration (copies/ $\mu$ L) multiplied by the reaction volume (20  $\mu$ L).

### RT-ddPCR

Total RNA was isolated using the RNeasy Mini Kit according to the manufacturer's recommendations followed by cDNA synthesis using standard protocols for ReverTra Ace (TOYOBO). RT-ddPCR was performed using ddPCR™ Supermix for Probes (No dUTP, BioRad), 900 nM forward and reverse primers, 250 nM probes labeled with FAM or HEX dye and ZEN-Iowa Black Fluorescent quencher according to the manufacturer's recommendations. A cDNA sample from a total RNA sample (300–1,000 ng) was added to the mixture. The available sequences of primers and probes used in the present study are listed in S5 Table. Droplet generation, water-in-oil emulsions transfer, and RT-PCR were done following the one-step RT-ddPCR procedure [14]. The thermal cycling conditions were as follows: 10 min enzyme activation at 95°C and 40 cycles of a thermal profile comprising 30 s denaturation at 94°C and 60 s annealing/extension at 57.1°C. After the PCR amplification, the products were denatured for 10 min at 98°C and cooled at 4°C until the fluorescence intensity was measured. The fluorescence intensities of each droplet from the samples were measured using the QX200 droplet reader. Positive droplets containing amplification products were distinguished from negative droplets and counted by applying a fluorescence amplitude threshold in QuantaSoft software. The threshold was manually determined based on the distribution of the droplet and set at 4000 for *TBP* and 7000 for *MIR302CHG*. QuantaSoft software provides the estimated concentration results in copies of target per microliter (copies/ $\mu$ L). The ratio of the copy number of target genes per copy number of *TBP* was calculated.

The sum of the *TBP*-normalized copy number plus the minimum *TBP*-normalized copy number of the dataset were subjected to log10 transformation and then used for the statistical analyses in Fig 4C–4E. One-way ANOVA and Tukey-Kramer post hoc tests were performed for multiple group comparisons.

### Magnetic bead-based enrichment

Magnetic bead-based enrichment was performed by MACS (Miltenyi Biotec). hiPSCs and NPCs were dissociated using Accumax (Innovative Cell Technologies, Inc.), and the cell suspension was spun down at 200g for 7 min at 4°C. After centrifugation, the supernatant was discarded, and the cell pellets were reconstituted in MACS rinsing solution with 0.5% BSA (Wako or Miltenyi Biotec) and 10  $\mu$ M Y-27632. Cell suspensions containing hiPSCs and NPCs ( $1:10^6$ ) were prepared and incubated with 1  $\mu$ L anti-human TRA-1-60 microbeads (Miltenyi Biotec) per  $10^6$  cells for 15 min at 4°C. Tapping was performed every five minutes. The suspensions were then filtered using a 100- $\mu$ m pore cell strainer (Falcon) and applied to an LS column attached to a MidiMACS separator (Miltenyi Biotec). The column was then washed twice using 3 mL of MACS rinsing solution with 0.5% BSA. In the process of applying the labeled cells and washing, the syringe was gently pressed to let the drops fall at about one drop per second. The positive fractions were flushed out and collected by gently pressing with a syringe at a rate of 2–3 drops per second. After centrifugation at 200g for 5 min, the cells from both the positive and negative fractions were dissolved using buffer RLT (QIAGEN) with  $\beta$ -mercaptoethanol and

stored at  $-80^{\circ}\text{C}$  until the analysis. Total RNA was isolated using the RNeasy Mini Kit with DNase I treatment according to the manufacturer's instructions. The RNA quantity was determined using NanoDrop (Thermo Fisher Scientific). The RNA samples of the TRA-1-60 positive fraction (12.9  $\mu\text{L}$ ) and flow-through fractions (up to 1,000 ng of RNA) were used in a total of 20  $\mu\text{L}$  RT reaction mix, and cDNA synthesis was performed according to standard protocols for ReverTra Ace (TOYOBO).

### Flow cytometry and cell sorting

We prepared cell samples by incubation with Accumax (Innovative Cell Technologies) for 7 min at  $37^{\circ}\text{C}$  and dissociation by pipetting or using each fraction of the MACS separation. Dead cells stained with 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI; 0.1 ng/mL; Thermo Fisher Scientific) were excluded from the analyses. The cells were analyzed and sorted using a FACS Aria II cell sorter (BD), and the data were analyzed using the FACS Diva (BD) software program.

In [S4 Fig](#), we manually prepared and analyzed a 1:10,000 mixture of hiPSCs constitutively expressing tdTomato and induced NPCs from a non-fluorescent hiPSC line (ff-I 14s04) by flow cytometry and examined whether the manual mixing was done accurately. The tdTomato (+) fractions sorted by flow cytometry were analyzed again to test whether the positive fractions were not noise such as bubbles.

In [Fig 3](#), we used the mixtures of hiPSCs constitutively expressing EGFP (317–12 cells) and NPCs derived from a non-fluorescent hiPSC line (QHJI 14s04) and evaluated the degree of magnetic bead-based enrichment in the positive and negative fractions by flow cytometry using GFP as an indicator.

### Generation of an hiPSC line constitutively expressing tdTomato

The hiPSC line constitutively expressing tdTomato was generated as described previously with some modifications [63, 64]. Briefly, the EGFP sequence between the NcoI and EcoRI sites was replaced with a PCR-amplified tdTomato fragment (see the primer sequences in [S6 Table](#)) in the *piggyBac* transposon vector pPV-EF1a-EiP-A (provided by Prof. Akitsu Hotta) to create pPV-EF1a-tdTomato-A, allowing us to ubiquitously express tdTomato under the human EF1a promoter. pPV-EF1a-tdTomato-A plasmids were electroporated into 585A1 hiPSCs [50] with the *piggyBac* transposase-expressing vector pHL-EF1a-hcPBase-A [63] using a NEPA21 electroporator (Nepa Gene). Highly tdTomato-expressing hiPSCs were FACS-sorted after a 5-day culture and isolated for single cell expansion. The isolated colony with the highest tdTomato expression was used in this study.

### Supporting information

**S1 Fig. Analysis of *LIN28A*, *ESRG*, *CNMD* and *SFRP2* expression in various human fetal tissues and hiPSC-derived progenitor cells and organoids using publicly available RNA sequencing datasets.** (A) Scatter plots of the CPM values of *LIN28A*, *ESRG*, *CNMD* and *SFRP2* in the various human fetal tissues from first or second trimesters reported in Roost et al. (2015). (B-E) Scatter plots of the CPM values of *LIN28A*, *ESRG*, *CNMD* and *SFRP2* in hiPSCs, day 12 hiPSC-derived hepatoblasts, and day 20 hiPSC-derived hepatocytes with or without methoxamine treatment reported in Kotaka et al. (2017) (B), hiPSCs or hiPSC-derived pancreatic progenitors reported in Kimura et al. (2020) (C), day 6 hiPSC-derived foregut spheroids, days 20 and 25 hiPSC-derived liver organoids reported in Ouchi et al. (2019) (D), and hiPSC-derived kidney progenitors (day 0–3 pellets) and organoids (day 11–18 pellets)

reported in Takasato et al. (2020) (E).  
(TIF)

**S2 Fig. Microarray analysis of *Lin28a*, *Cnmd* and *Sfrp2* in mouse metanephric mesenchyme.** (A-C) Microarray analysis of *Lin28a* (A), *Cnmd* (B) and *Sfrp2* (C) in mouse developing kidneys using GUDMAP.

(TIF)

**S3 Fig. Differential gene expression analysis of hiPSCs and hiPSC-derived OSR1(+)SIX2(+)NPCs and induction of NPCs and kidney organoids from a clinical-grade iPSC line.** (A) Heatmap of the top 1,000 DEGs between hiPSCs and hiPSC-derived OSR1(+)SIX2(+)NPCs and the results of the enrichment analysis of three gene clusters defined by the gene expression patterns. The blue, yellow and purple bands on the left of the heatmap and the table correspond to each of the three clusters. (B) Representative immunofluorescence images of an induced NPC aggregate for NPC markers, SIX2 and PAX2. Scale bar, 500  $\mu$ m. (C) Representative bright field and immunofluorescence images of kidney organoids derived from induced NPCs for markers of renal lineage cells (PAX8), glomeruli (PODX) and renal tubules (LTL and CDH1). Scale bars, 500  $\mu$ m. (D) q RT-PCR analysis of the expression of *LIN28A* and markers for hiPSCs (*POU5F1* and *NANOG*) and NPCs (*OSR1* and *SIX2*). Each value was normalized to that of hiPSCs.

(TIF)

**S4 Fig. scRNA-seq analysis of purified OSR1(+)SIX2(+) induced NPCs for iPSC and NPC marker genes.** (A) Violin plots of the cells in hiPSC-derived metanephric and mesonephric NPC populations reported in Tsujimoto et al. (2020) with standard quality control parameters after filtering. (B) UMAP plots for hiPSC-derived metanephric and mesonephric NPCs. The number of cells in each NPC population: metanephric NPCs, 190; and mesonephric NPCs, 199. (C, D) Violin plots of representative NPC (C) and iPSC (D) markers of hiPSC-derived metanephric and mesonephric NPCs. (E) UMAP plots of representative NPC markers (*SIX2*, *PAX2*, *PAX8* and *WT1*) in hiPSC-derived metanephric and mesonephric NPCs. (F) Scatter plots of iPSC markers (*LIN28A*, *CNMD* and *SFRP2*) and NPC markers (*SIX2* and *PAX2*). Numbers above the plots are Pearson correlation coefficients. Mesonephric NPCs were induced from hiPSCs by the same protocol as the metanephric NPC induction except for removing activin A at Stage 4. MESO, mesonephric NPC; META, metanephric NPC.

(TIF)

**S5 Fig. Digital droplet PCR assays of hiPSCs intermingled in induced NPCs using *CUZD1*.** (A) Flow cytometric analysis of a 1:10,000 mixture of hiPSCs constitutively expressing tdTomato and induced NPCs from a clinical-grade hiPSC line (left and center panels) indicates that the manual mixing was done accurately. A histogram (right panel) of the tdTomato(+) fractions that were sorted and FACS analyzed again suggests that the positive fractions were not noise such as bubbles. (B) A box plot for the estimated number of *CUZD1* copy/total RNA (1 ng) of hiPSCs (iPSC), 1:10,000 mixture of hiPSCs and induced NPCs (0.01%iPSC), and NPCs (NPC) using a one-step ddPCR assay suggests that the estimation is biased for some technical reasons. (C) Representative scatter plots of positive and negative droplets for *CUZD1* and *TBP* using a RT-ddPCR of iPSC, 0.01%iPSC, and NPC (left panels) and a box plot for the *TBP*-normalized estimated number of *CUZD1* copies according to an RT-ddPCR (right panel). (D, E) A ddPCR analysis of *CUZD1* using an hiPSC cDNA dilution series sample diluted by the cDNAs of NPCs (D) or day 4 cells (E) shows some concentration-dependent changes in the *TBP*-normalized copy number of *CUZD1*.

(TIF)



**S6 Fig. Marker gene expressions in hiPSCs, primitive streak and NPCs at each stage of the NPC induction protocol.** Scatter plots of *TBP*-normalized expression values by qRT-PCR for marker genes of hiPSCs (*MIR302CHG*, *CUZD1*, and *POU5F1*), primitive streak (*TBX6*) and NPCs (*OSR1* and *SIX2*). UD: number of samples with undetermined CT values. (TIF)

**S7 Fig. Identification of a lncRNA, *MIR302CHG*, as a novel detection marker for hiPSCs intermingled in induced NPCs.** (A) A scatter plot of the RPKM values of *CUZD1* in 27 tissue samples from 95 human individuals reported by Fagerberg et al. (2014). (B) A volcano plot of the DEGs at  $\log_{10}(\text{FDR}) > 5$  and  $\log_2(\text{fold change}) < -10$  (blue dots) or  $> 10$  (red dots) and a list of 26 candidate hiPSC markers. A gene classification by IPA is shown. (C) Scatter plots of *TBP*-normalized expression values of several hiPSC lines normalized to the QHJI 14s04 hiPSC line by qRT-PCR for marker genes of hiPSCs (*MIR302CHG*, *CUZD1* and *POU5F1*). (D) A scatter plot of *TBP*-normalized qRT-PCR dCT values of hiPSCs (iPSC) and induced NPCs (NPC) for 15 *MIR302CHG* primers. Primers #1-#5 are specific for both ENST00000509938.1 and ENST00000505215.1, while primers #6-#15 are specific for ENST00000509938.1. Primer #2 was used for the other qRT-PCR assays. UD: undetermined CT values. (TIF)

**S8 Fig. *MIR302CHG* expression in hiPSC-derived pancreatic progenitor cells, hESC-derived endoderm lineages, and hiPSC-derived hepatocyte and liver organoids from publicly available RNA-seq datasets.** (A-D) Scatter plots of the TPM values of each transcript variant of *MIR302CHG* in hiPSCs and hiPSC-derived pancreatic progenitors reported in Kimura et al. (2020) (A), hESC-derived endoderm lineages reported in Loh et al. (2013) (B), hiPSCs, hiPSC-derived D12 hepatoblasts and D20 hepatocytes reported in Kotaka et al. (2017) (C), and hiPSC-derived day 6 foregut spheroids and days 20 and 25 hiPSC-derived liver organoids reported in Ouchi et al. (2019) (D). (TIF)

**S9 Fig. *MIR302CHG* expression in hESC-derived three germ layer lineages, and hiPSC-derived neuron, cardiomyocyte, and lung lineages from publicly available RNA-seq datasets.** (A-D) Scatter plots of the TPM values of each transcript variant of *MIR302CHG* in hESC-derived definitive endoderm, splanchnic mesoderm, neural progenitor cells and pre-neural crest cells reported in Cliff et al. (2017) (A), hiPSCs and hiPSC-derived neural progenitor cells and neurons reported in Chen et al. (2013) (B), hiPSCs from 58 Yoruba individuals and hiPSC-derived cardiomyocytes reported in Banovich et al. (2018) (C), hiPSC-derived D5 definitive endoderm, D10 anterior foregut, D15 and D20 lung progenitors and D52 and D66 lung epithelial cells reported in Kerschner et al. (2020) (D), and hESCs, hESC-derived D15 lung progenitors, D35 SFTPC(+) and (-) cells of lung organoids, and GW21 human fetal lung cells (E). (TIF)

**S10 Fig. *MIR302CHG* expression in hiPSC- and hESC-derived cardiomyocytes and endothelial cells from publicly available RNA-seq datasets.** Scatter plots of the TPM values of each transcript variant of *MIR302CHG* in hiPSCs, in vitro fertilization embryo-derived hESCs, somatic cell nuclear transfer-derived hESCs, and their derivative cardiomyocytes and endothelial cells reported in Zhao et al. (2017). (TIF)

**S11 Fig. NPC growth assay using iPSC media.** (A) The cell number of each cell aggregate treated with Essential 8 or AK03N media at 1, 3, 5, 8 and 10 days after seeding. (B, C)

Representative brightfield or fluorescent images of OSR1 (GFP) SIX2 (tdTomato) from days 0 (B), 1 and 10 (C) cell aggregates treated with Essential 8 or AK03N media.  $**p < 0.01$  by paired Student's t-tests comparing the cell number in day 1 (D1) and day 10 (D10) cell aggregates. Data are represented as the mean  $\pm$  SEM ( $n = 12$ ). Scale bars, 300  $\mu\text{m}$ . (TIF)

**S12 Fig. The combination assay of magnetic bead-based cell isolation and qRT-PCR analysis for NPCs.** Scatter plots of the *TBP*-normalized gene expressions by qRT-PCR for *MIR302CHG*, *CUZD1*, and *POU5F1* in NPCs, and MACS positive or negative fraction flow-throughs of the NPCs (PF or NF). The dots and lines in the center of the scatter plots indicate the experimental data and mean values of the data, respectively. UD: number of samples with undetermined CT values. (TIF)

**S1 Table. Highly expressed genetic markers ( $\log_2\text{FC} > 10$  &  $\text{FDR} < 10^{-5}$ ) in hiPSCs compared to day 4 cells from the NPC differentiation protocol.** (DOCX)

**S2 Table. Primer sequences for qRT-PCR used in this study.** (DOCX)

**S3 Table. Antibodies and lectin used in this study.** (DOCX)

**S4 Table. RNA sequencing data used in this study.** (DOCX)

**S5 Table. Primer and probe sequences for ddPCR used in this study.** (DOCX)

**S6 Table. Recombinant DNA used in this study.** (DOCX)

**S1 Data. Adjusted  $p$ -values by Tukey-Kramer post hoc tests.** (XLSX)

**S1 Dataset. Transcript count data.** (XLSX)

**S1 File. Informed consent forms.** (ZIP)

## Acknowledgments

The authors would like to thank Dr. Hotta for providing the synthetic DNA constructs used to generate the hiPSC line expressing tdTomato, Dr. Peter Karagiannis for critically reading and revising the manuscript, Drs. Masaki Nomura, Toshikazu Araoka, Shin-Ichi Mae, Tatsuya Shimizu, Kosuke Toyohara, Maki Kotaka, Kanae Mitsunaga, Wang Changting, Hiroaki Magotani and Shunjiro Sugimoto and Ms. Yumi Tsubokura for helpful discussion and technical support, and the CiRA Foundation (<https://www.cira-foundation.or.jp/e/>) for providing iPSCs from their iPSC stock and for the RNA sequencing analysis of the cells on a PC isolated from the network. The authors would also like to thank Drs. Mika Suzuki and Michihiro Tanaka for helpful discussions on ethical issues and information security advice for the gene expression analysis of the iPSCs from the iPSC stock project, respectively.

## Author Contributions

**Conceptualization:** Hiraku Tsujimoto, Kenji Osafune.

**Formal analysis:** Hiraku Tsujimoto, Naoko Katagiri, Ben Sasaki, Yoshifumi Kobayashi, Akira Mima, Makoto Ryosaka, Kenichiro Furuyama, Yoshiya Kawaguchi, Kenji Osafune.

**Investigation:** Hiraku Tsujimoto, Naoko Katagiri, Yoshihiro Ijiri, Ben Sasaki.

**Methodology:** Hiraku Tsujimoto, Kenji Osafune.

**Project administration:** Hiraku Tsujimoto.

**Supervision:** Yoshiya Kawaguchi, Kenji Osafune.

**Writing – original draft:** Hiraku Tsujimoto.

**Writing – review & editing:** Hiraku Tsujimoto, Naoko Katagiri, Yoshihiro Ijiri, Ben Sasaki, Yoshifumi Kobayashi, Akira Mima, Makoto Ryosaka, Kenichiro Furuyama, Yoshiya Kawaguchi, Kenji Osafune.

## References

1. Harari-Steinberg O, Metsuyanin S, Omer D, Gnatek Y, Gershon R, Pri-Chen S, et al. Identification of human nephron progenitors capable of generation of kidney structures and functional repair of chronic renal disease. *EMBO Mol Med.* 2013; 5: 1556–1568. <https://doi.org/10.1002/emmm.201201584> PMID: 23996934
2. Toyohara T, Mae S-I, Sueta S-I, Inoue T, Yamagishi Y, Kawamoto T, et al. Cell Therapy Using Human Induced Pluripotent Stem Cell-Derived Renal Progenitors Ameliorates Acute Kidney Injury in Mice. *Stem Cells Transl Med.* 2015; 4: 980–992. <https://doi.org/10.5966/sctm.2014-0219> PMID: 26198166
3. Hoshina A, Kawamoto T, Sueta S, Mae S, Araoka T, Tanaka H, et al. Development of new method to enrich human iPSC-derived renal progenitors using cell surface markers. *Sci Rep.* 2018; 8: 6375. <https://doi.org/10.1038/s41598-018-24714-3> PMID: 29686294
4. Imberti B, Tomasoni S, Ciampi O, Pezzotta A, Derosas M, Xinaris C, et al. Renal progenitors derived from human iPSCs engraft and restore function in a mouse model of acute kidney injury. *Sci Rep.* 2015; 5: 8826. <https://doi.org/10.1038/srep08826> PMID: 25744951
5. Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V. Chronic kidney disease. *Lancet (London, England).* 2021. [https://doi.org/10.1016/S0140-6736\(21\)00519-5](https://doi.org/10.1016/S0140-6736(21)00519-5)
6. Lu BC, Cebrian C, Chi X, Kuure S, Kuo R, Bates CM, et al. Etv4 and Etv5 are required downstream of GDNF and Ret for kidney branching morphogenesis. *Nat Genet.* 2009; 41: 1295–1302. <https://doi.org/10.1038/ng.476> PMID: 19898483
7. Osafune K, Takasato M, Kispert A, Asashima M, Nishinakamura R. Identification of multipotent progenitors in the embryonic mouse kidney by a novel colony-forming assay. *Development.* 2006; 133: 151–61. <https://doi.org/10.1242/dev.02174> PMID: 16319116
8. Thirabanjasak D, Tantiwongse K, Thorner PS. Angiomyeloproliferative lesions following autologous stem cell therapy. *J Am Soc Nephrol.* 2010; 21: 1218–1222. <https://doi.org/10.1681/ASN.2009111156> PMID: 20558536
9. Berkowitz AL, Miller MB, Mir SA, Cagney D, Chavakula V, Guleria I, et al. Glioproliferative Lesion of the Spinal Cord as a Complication of “Stem-Cell Tourism.” *N Engl J Med.* 2016; 375: 196–198. <https://doi.org/10.1056/NEJMc1600188> PMID: 27331440
10. Sato Y, Bando H, Di Piazza M, Gowing G, Herberts C, Jackman S, et al. Tumorigenicity assessment of cell therapy products: The need for global consensus and points to consider. *Cytotherapy.* 2019; 21: 1095–1111. <https://doi.org/10.1016/j.jcyt.2019.10.001> PMID: 31711733
11. Doi D, Magotani H, Kikuchi T, Ikeda M, Hiramatsu S, Yoshida K, et al. Pre-clinical study of induced pluripotent stem cell-derived dopaminergic progenitor cells for Parkinson’s disease. *Nat Commun.* 2020; 11: 3369. <https://doi.org/10.1038/s41467-020-17165-w> PMID: 32632153
12. Tsujimoto H, Osafune K. Current status and future directions of clinical applications using iPSC cells—focus on Japan. *FEBS J.* 2021; febs.16162. <https://doi.org/10.1111/febs.16162> PMID: 34407307
13. Kuroda T, Yasuda S, Kusakawa S, Hirata N, Kanda Y, Suzuki K, et al. Highly sensitive in vitro methods for detection of residual undifferentiated cells in retinal pigment epithelial cells derived from human iPSC cells. *PLoS One.* 2012;7. <https://doi.org/10.1371/journal.pone.0037342> PMID: 22615985

14. Kuroda T, Yasuda S, Matsuyama S, Tano K, Kusakawa S, Sawa Y, et al. Highly sensitive droplet digital PCR method for detection of residual undifferentiated cells in cardiomyocytes derived from human pluripotent stem cells. *Regen Ther.* 2015; 2: 17–23. <https://doi.org/10.1016/j.reth.2015.08.001> PMID: 31245455
15. Watanabe T, Yasuda S, Kusakawa S, Kuroda T, Futamura M, Ogawa M, et al. Multisite studies for validation and improvement of a highly efficient culture assay for detection of undifferentiated human pluripotent stem cells intermingled in cell therapy products. *Cytotherapy.* 2021; 23: 176–183. <https://doi.org/10.1016/j.jcyt.2020.07.009> PMID: 32978066
16. Tano K, Yasuda S, Kuroda T, Saito H, Umezawa A, Sato Y. A novel in vitro method for detecting undifferentiated human pluripotent stem cells as impurities in cell therapy products using a highly efficient culture system. *PLoS One.* 2014;9. <https://doi.org/10.1371/journal.pone.0110496> PMID: 25347300
17. Zhang Y, Feng G-H, Xu K, Wang L, Cui P, Li Y, et al. A non-invasive method to determine the pluripotent status of stem cells by culture medium microRNA expression detection. *Sci Rep.* 2016; 6: 22380. <https://doi.org/10.1038/srep22380> PMID: 26927897
18. Tateno H, Onuma Y, Ito Y, Hiemori K, Aiki Y, Shimizu M, et al. A medium hyperglycosylated podocalyxin enables noninvasive and quantitative detection of tumorigenic human pluripotent stem cells. *Sci Rep.* 2014; 4: 4069. <https://doi.org/10.1038/srep04069> PMID: 24518842
19. Wang Z, Gagliardi M, Mohamadi RM, Ahmed SU, Labib M, Zhang L, et al. Ultrasensitive and rapid quantification of rare tumorigenic stem cells in hPSC-derived cardiomyocyte populations. *Sci Adv.* 2020; 6: eaay7629. <https://doi.org/10.1126/sciadv.aay7629> PMID: 32440533
20. Sekine K, Tsuzuki S, Yasui R, Kobayashi T, Ikeda K, Hamada Y, et al. Robust detection of undifferentiated iPSC among differentiated cells. *Sci Rep.* 2020; 10: 10293. <https://doi.org/10.1038/s41598-020-66845-6> PMID: 32581272
21. Roost MS, Van Iperen L, Ariyurek Y, Buermans HP, Arindrarto W, Devalla HD, et al. KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. *Stem Cell Reports.* 2015; 4: 1112–1124. <https://doi.org/10.1016/j.stemcr.2015.05.002> PMID: 26028532
22. Mahi N Al, Najafabadi MF, Pilarczyk M, Kouril M, Medvedovic M. GREIN: An Interactive Web Platform for Re-analyzing GEO RNA-seq Data. *Sci Rep.* 2019;9. <https://doi.org/10.1038/s41598-019-43935-8> PMID: 31110304
23. Kotaka M, Toyoda T, Yasuda K, Kitano Y, Okada C, Ohta A, et al. Adrenergic receptor agonists induce the differentiation of pluripotent stem cell-derived hepatoblasts into hepatocyte-like cells. *Sci Rep.* 2017; 7: 1–13. <https://doi.org/10.1038/s41598-017-16858-5> PMID: 29196668
24. Kimura A, Toyoda T, Iwasaki M, Hiramura R, Osafune K. Combined Omics Approaches Reveal the Roles of Non-canonical WNT7B Signaling and YY1 in the Proliferation of Human Pancreatic Progenitor Cells. *Cell Chem Biol.* 2020; 27: 1561-1572.e7. <https://doi.org/10.1016/j.chembiol.2020.08.018> PMID: 33125912
25. Ouchi R, Togo S, Kimura M, Shinozawa T, Koido M, Koike H, et al. Modeling Steatohepatitis in Humans with Pluripotent Stem Cell-Derived Organoids. *Cell Metab.* 2019; 1–11. <https://doi.org/10.1016/j.cmet.2019.05.007> PMID: 31155493
26. Takasato M, Er PX, Chiu HS, Maier B, Baillie GJ, Ferguson C, et al. Kidney organoids from human iPSCs contain multiple lineages and model human nephrogenesis. *Nature.* 2015; 526: 564–568. <https://doi.org/10.1038/nature15695> PMID: 26444236
27. Harding SD, Armit C, Armstrong J, Brennan J, Cheng Y, Haggarty B, et al. The GUDMAP database—an online resource for genitourinary research. *Development.* 2011; 138: 2845–2853. <https://doi.org/10.1242/dev.063594> PMID: 21652655
28. Mae S-I, Shono A, Shiota F, Yasuno T, Kajiwara M, Gotoda-Nishimura N, et al. Monitoring and robust induction of nephrogenic intermediate mesoderm from human pluripotent stem cells. *Nat Commun.* 2013; 4: 1367. <https://doi.org/10.1038/ncomms2378> PMID: 23340407
29. Tsujimoto H, Kasahara T, Sueta S ichi, Araoka T, Sakamoto S, Okada C, et al. A Modular Differentiation System Maps Multiple Human Kidney Lineages from Pluripotent Stem Cells. *Cell Rep.* 2020; 31: 107476. <https://doi.org/10.1016/j.celrep.2020.03.040> PMID: 32268094
30. Matsuda M, Yamanaka Y, Uemura M, Osawa M, Saito MK, Nagahashi A, et al. Recapitulating the human segmentation clock with pluripotent stem cells. *Nature.* 2020; 580: 124–129. <https://doi.org/10.1038/s41586-020-2144-9> PMID: 32238941
31. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017; 14: 417–419. <https://doi.org/10.1038/nmeth.4197> PMID: 28263959
32. Umekage M, Sato Y, Takasu N. Overview: an iPSC cell stock at CiRA. *Inflamm Regen.* 2019; 39: 17. <https://doi.org/10.1186/s41232-019-0106-0> PMID: 31497180

33. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014; 13: 397–406. <https://doi.org/10.1074/mcp.M113.035600> PMID: 24309898
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29: 15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
35. Anokye-Danso F, Trivedi CM, Jühr D, Gupta M, Cui Z, Tian Y, et al. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*. 2011; 8: 376–88. <https://doi.org/10.1016/j.stem.2011.03.001> PMID: 21474102
36. Nakagawa M, Taniguchi Y, Senda S, Takizawa N, Ichisaka T, Asano K, et al. A novel efficient feeder-free culture system for the derivation of human induced pluripotent stem cells. *Sci Rep*. 2014; 4: 3594. <https://doi.org/10.1038/srep03594> PMID: 24399248
37. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*. 2007; 131: 861–872. <https://doi.org/10.1016/j.cell.2007.11.019> PMID: 18035408
38. Ocegüera-Yanez F, Kim S II, Matsumoto T, Tan GW, Xiang L, Hatani T, et al. Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. *Methods*. 2016; 101: 43–55. <https://doi.org/10.1016/j.ymeth.2015.12.012> PMID: 26707206
39. Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, et al. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol*. 2004; 270: 488–498. <https://doi.org/10.1016/j.ydbio.2004.02.019> PMID: 15183728
40. Houbaviiy HB, Murray MF, Sharp PA. Embryonic stem cell-specific microRNAs. *Dev Cell*. 2003; 5: 351–358. [https://doi.org/10.1016/s1534-5807\(03\)00227-2](https://doi.org/10.1016/s1534-5807(03)00227-2) PMID: 12919684
41. Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, et al. Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs. *Cell Stem Cell*. 2011; 8: 633–638. <https://doi.org/10.1016/j.stem.2011.05.001> PMID: 21620789
42. Natunen S, Satomaa T, Pitkänen V, Salo H, Mikkola M, Natunen J, et al. The binding specificity of the marker antibodies Tra-1-60 and Tra-1-81 reveals a novel pluripotency-associated type 1 lactosamine epitope. *Glycobiology*. 2011; 21: 1125–1130. <https://doi.org/10.1093/glycob/cwq209> PMID: 21159783
43. Fujitani N, Furukawa J ichi, Araki K, Fujioka T, Takegawa Y, Piao J, et al. Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. *Proc Natl Acad Sci U S A*. 2013; 110: 2105–2110. <https://doi.org/10.1073/pnas.1214233110> PMID: 23345451
44. Tateno H. Stem cell glycomics: Structure, function, and application. *Seikagaku*. Japanese Biochemical Society; 2016. pp. 761–765. <https://doi.org/10.14952/SEIKAGAKU.2016.880761>
45. Tateno H, Toyota M, Saito S, Onuma Y, Ito Y, Hiemori K, et al. Glycome diagnosis of human induced pluripotent stem cells using lectin microarray. *J Biol Chem*. 2011; 286: 20345–20353. <https://doi.org/10.1074/jbc.M111.231274> PMID: 21471226
46. Tanosaki S, Tohyama S, Fujita J, Someya S, Hishiki T, Matsuura T, et al. Fatty Acid Synthesis Is Indispensable for Survival of Human Pluripotent Stem Cells. *iScience*. 2020;23. <https://doi.org/10.1016/j.isci.2020.101535> PMID: 33083764
47. Nakashima Y, Miyagi-Shiohira C, Noguchi H, Omasa T. Atorvastatin Inhibits the HIF1 $\alpha$ -PPAR Axis, Which Is Essential for Maintaining the Function of Human Induced Pluripotent Stem Cells. *Mol Ther*. 2018; 26: 1715–1734. <https://doi.org/10.1016/j.ymthe.2018.06.005> PMID: 29929789
48. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, et al. A more efficient method to generate integration-free human iPS cells. *Nat Methods*. 2011; 8: 409–412. <https://doi.org/10.1038/nmeth.1591> PMID: 21460823
49. Hanatani T, Takasu N. CiRA iPSC seed stocks (CiRA's iPSC Stock Project). *Stem Cell Res*. 2021;50. <https://doi.org/10.1016/j.scr.2020.102033> PMID: 33445065
50. Okita K, Yamakawa T, Matsumura Y, Sato Y, Amano N, Watanabe A, et al. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells*. 2013; 31: 458–466. <https://doi.org/10.1002/stem.1293> PMID: 23193063
51. Miyazaki T, Isobe T, Nakatsuji N, Suemori H. Efficient Adhesion Culture of Human Pluripotent Stem Cells Using Laminin Fragments in an Uncoated Manner. *Sci Rep*. 2017; 7: 41165. <https://doi.org/10.1038/srep41165> PMID: 28134277
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
53. Ge SX, Son EW, Yao R. iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 2018;19. <https://doi.org/10.1186/s12859-018-2486-6> PMID: 30567491

54. Loh KM, Ang LT, Zhang J, Kumar V, Ang J, Auyeong JQ, et al. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell*. 2014; 14: 237–252. <https://doi.org/10.1016/j.stem.2013.12.007> PMID: 24412311
55. Cliff TS, Wu T, Boward BR, Yin A, Yin H, Glushka JN, et al. MYC Controls Human Pluripotent Stem Cell Fate Decisions through Regulation of Metabolic Flux. *Cell Stem Cell*. 2017; 21: 502–516.e9. <https://doi.org/10.1016/j.stem.2017.08.018> PMID: 28965765
56. Chen J, Lin M, Foxe JJ, Pedrosa E, Hrabovsky A, Carroll R, et al. Transcriptome Comparison of Human Neurons Generated Using Induced Pluripotent Stem Cells Derived from Dental Pulp and Skin Fibroblasts. Kerkis I, editor. *PLoS One*. 2013; 8: e75682. <https://doi.org/10.1371/journal.pone.0075682> PMID: 24098394
57. Banovich NE, Li YI, Raj A, Ward MC, Greenside P, Calderon D, et al. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res*. 2018; 28: 122–131. <https://doi.org/10.1101/gr.224436.117> PMID: 29208628
58. Zhao M-T, Chen H, Liu Q, Shao N-Y, Sayed N, Wo H-T, et al. Molecular and functional resemblance of differentiated cells derived from isogenic human iPSCs and SCNT-derived ESCs. *Proc Natl Acad Sci U S A*. 2017; 114: E11111–E11120. <https://doi.org/10.1073/pnas.1708991114> PMID: 29203658
59. Kerschner JL, Paranjapye A, Yin S, Skander DL, Bebek G, Leir S, et al. A functional genomics approach to investigate the differentiation of iPSCs into lung epithelium at air-liquid interface. *J Cell Mol Med*. 2020; 24: 9853–9870. <https://doi.org/10.1111/jcmm.15568> PMID: 32692488
60. Jacob A, Morley M, Hawkins F, McCauley KB, Jean JC, Heins H, et al. Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell*. 2017; 21: 472–488.e10. <https://doi.org/10.1016/j.stem.2017.08.014> PMID: 28965766
61. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multi-modal single-cell data. *Cell*. 2021; 184: 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048> PMID: 34062119
62. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019; 20: 296. <https://doi.org/10.1186/s13059-019-1874-1> PMID: 31870423
63. Ishida K, Xu H, Sasakawa N, Lung MSY, Kudryashev JA, Gee P, et al. Site-specific randomization of the endogenous genome by a regulatable CRISPR-Cas9 piggyBac system in human cells. *Sci Rep*. 2018; 8: 1–12. <https://doi.org/10.1038/s41598-017-18568-4> PMID: 29321585
64. Gee P, Lung MSY, Okuzaki Y, Sasakawa N, Iguchi T, Makita Y, et al. Extracellular nanovesicles for packaging of CRISPR-Cas9 protein and sgRNA to induce therapeutic exon skipping. *Nat Commun*. 2020; 11. <https://doi.org/10.1038/s41467-020-14957-y> PMID: 32170079