


RESEARCH

Open Access



# Long noncoding RNA *CCDC144NL-AS1* knockdown induces naïve-like state conversion of human pluripotent stem cells

Yingying Wang<sup>1</sup>, Baosen Guo<sup>2</sup>, Zengrong Xiao<sup>2</sup>, Haijun Lin<sup>2</sup>, Xi Zhang<sup>1</sup>, Yueqiang Song<sup>1</sup>, Yalei Li<sup>2</sup>, Xuehu Gao<sup>2</sup>, Jinjun Yu<sup>2</sup>, Zhihua Shao<sup>1</sup>, Xuekun Li<sup>3</sup>, Yuping Luo<sup>1,4\*</sup> and Siguang Li<sup>1,5\*</sup> 

## Abstract

**Background:** Human naïve pluripotency state cells can be derived from direct isolation of inner cell mass or primed-to-naïve resetting of human embryonic stem cells (hESCs) through different combinations of transcription factors, small molecular inhibitors, and growth factors. Long noncoding RNAs (lncRNAs) have been identified to be crucial in diverse biological processes, including pluripotency regulatory circuit of mouse pluripotent stem cells (PSCs), but few are involved in human PSCs' regulation of pluripotency and naïve pluripotency derivation. This study initially planned to discover more lncRNAs possibly playing significant roles in the regulation of human PSCs' pluripotency, but accidentally identified a lncRNA whose knockdown in human PSCs induced naïve-like pluripotency conversion.

**Methods:** Candidate lncRNAs tightly correlated with human pluripotency were screened from 55 RNA-seq data containing human ESC, human induced pluripotent stem cell (iPSC), and somatic tissue samples. Then loss-of-function experiments in human PSCs were performed to investigate the function of these candidate lncRNAs. The naïve-like pluripotency conversion caused by *CCDC144NL-AS1* knockdown (KD) was characterized by quantitative real-time PCR, immunofluorescence staining, western blotting, differentiation of hESCs in vitro and in vivo, RNA-seq, and chromatin immunoprecipitation. Finally, the signaling pathways in *CCDC144NL-AS1*-KD human PSCs were examined through western blotting and analysis of RNA-seq data.

**Results:** The results indicated that knockdown of *CCDC144NL-AS1* induces naïve-like state conversion of human PSCs in the absence of additional transcription factors or small molecular inhibitors. *CCDC144NL-AS1*-KD human PSCs reveal naïve-like pluripotency features, such as elevated expression of naïve pluripotency-associated genes, increased developmental capacity, analogous transcriptional profiles to human naïve PSCs, and global reduction of repressive chromatin modification marks. Furthermore, *CCDC144NL-AS1*-KD human PSCs display inhibition of MAPK (ERK), accumulation of active  $\beta$ -catenin, and upregulation of some LIF/STAT3 target genes, and all of these are concordant with previously reported traits of human naïve PSCs.

**Conclusions:** Our study unveils an unexpected role of a lncRNA, *CCDC144NL-AS1*, in the naïve-like state conversion of human PSCs, providing a new perspective to further understand the regulation process of human early pluripotency states conversion. It is suggested that *CCDC144NL-AS1* can be potentially valuable for future research on deriving higher quality naïve state human PSCs and promoting their therapeutic applications.

**Keywords:** lncRNA, Human PSCs, Human naïve pluripotency, MAPK/ERK signaling pathway, Wnt signaling pathway

\* Correspondence: [luoyuping@163.com](mailto:luoyuping@163.com); [siguangli@163.com](mailto:siguangli@163.com)

<sup>1</sup>Stem Cell Translational Research Center, Tongji Hospital, Tongji University School of Medicine, Shanghai 200065, China

Full list of author information is available at the end of the article



## Background

Human pluripotent stem cells (PSCs), including human embryonic stem cells (ESCs) and human induced pluripotent stem cells (iPSCs), have the ability to self-renew indefinitely and possess the potential to differentiate into derivatives of all three embryonic germ layers [1–3]. Human PSCs (hPSCs) provide a unique model to study the early process of human embryogenesis, supply an investigational tool to disease modeling and drug discovery, and also could be a probable source of cells for regenerative medicine [4, 5]. Whereas, unlike mouse ESCs [6], which are derived from pre-implantation embryos and can be maintained in a naïve, ground state in vitro [7, 8], human PSCs exhibit developmentally more advanced or primed pluripotency akin to that of mouse post-implantation epiblast-derived stem cells (EpiSCs) [9–12].

Recent years' studies demonstrate that human ESCs can be converted to the naïve state by transgene dependent or independent ways. Continuous exogenous expression of different combinations of pluripotency-associated factors, such as octamer-binding protein 4 (*OCT4*, also known as *POU5F1*) and Kruppel-like factor 4 (*KLF4*), or *KLF2* and *KLF4*, or *NANOG* and *KLF2*, allowed the derivation and expansion of human cells encompassing ground pluripotency attributes of mouse ESCs in “2i/LIF” conditions supplemented with MEK inhibitor (PD0325901), GSK3 inhibitor (CHIR99021), and human LIF [13, 14].

Short-term overexpression of *NANOG* and *KLF2* in 2i/LIF followed by 2iL/Gö medium containing titrated 2i with LIF and protein kinase C inhibitor (Gö6983) also allowed the derivation and maintenance of ground-state human PSCs [15]. In addition, the temporary expression of STAT3 in 2i/LIF could reprogramme human ESCs to naïve-like pluripotency as well [16]. By contrast, transgene-independent human naïve pluripotency induction methods implicate in multiple small molecular inhibitors and growth factors. For instance, culture conditions containing 2i/LIF in company with inhibitors of Jun N-terminal kinase (JNK), P38, aPKC, Rho-associated protein kinase (ROCK), and growth factors FGF2 and TGF $\beta$  were described for inducing and maintaining human naïve PSCs [17]. Moreover, alternative conditions for inducing human naïve PSCs were reported, such as “3iL” condition which contained MEK inhibitor, GSK3 inhibitor, BMP inhibitor, and human LIF in mTeSR1 medium, and “5i/L/A” condition compromised of inhibitors for MEK, GSK3, ROCK, BRAF, and SRC and growth factors human LIF and activin [14, 18]. Additionally, human naïve PSCs can also be derived through directly culturing isolated cells of human inner cell mass in “2iL/GöY” medium which contained inhibitors for GSK3, MEK, PKC, and ROCK; growth factor human LIF; and ascorbic acid [19]. All artificial human naïve PSCs raise the feasibility and practical avenues to acquire an earlier pluripotency state than conventional primed human PSCs in vitro.

Previous studies of the underlying molecular mechanisms of pluripotency maintenance and lineage specification demonstrate a complex network involving transcription factors [20–24], chromatin regulators [25–28], and noncoding RNAs [29–35]. Long noncoding RNAs (lncRNAs) have been identified to be extensively transcribed in the mammalian genome [36, 37]. They are currently defined as a heterogeneous class of RNA polymerase II transcripts longer than 200 nucleotides with no apparent protein-coding potential [38]. lncRNAs show precisely regulated temporal and spatial expression patterns [39], and a large number of studies have corroborated lncRNAs' essential roles in diverse biological processes, such as immune response, cell cycle regulation, T cell development, and nervous system development [40–43]. Likewise, lncRNAs have been validated as an integral part of mouse PSCs' pluripotency regulatory circuit [32, 33, 35, 44]. Nevertheless, only a tiny fraction of lncRNAs were functionally characterized in the intricate network of human PSCs in contrast to hundreds of them identified in mouse PSCs [34, 45, 46]. So, we initially planned to discover more lncRNAs possibly playing significant roles in the regulation of human PSCs' pluripotency.

Intriguingly, in our subsequent research, we identified a lncRNA, *CCDC144NL-AS1*, the downregulation of which in human PSCs induced naïve-like pluripotency conversion without exogenous factors. *CCDC144NL-AS1*-knockdown (KD) human PSCs exhibited a series of naïve pluripotency specific features, such as naïve-like dome-shaped clones, increased expression of naïve pluripotency-associated genes, elevated developmental capacity, analogous transcriptional profiles to human naïve PSCs, and global reduction of repressive chromatin-modifying marks. Furthermore, *CCDC144NL-AS1* knockdown in human PSCs resulted in MAPK (ERK) inhibition,  $\beta$ -catenin activation, and elevated expression of some LIF/STAT3 target genes, and all of these were required for naïve pluripotency maintenance. Our study unveils a lncRNA's unexpected role in the complex conversion process of human early pluripotency states, and it can be potentially valuable for future research on deriving higher quality naïve state human PSCs.

## Methods

### Cell culture

The hESC lines H9 and H14 and hiPSC line HDF-iPS were used in this study. The cells were cultured feeder free on Matrigel (BD Biosciences) in TeSR-E8 medium (Stem Cell Technologies), and their medium was changed daily. The hPSCs were passaged with Versene (Gibco) every 5–7 days. Culture conditions of stable CCDC144NL-AS1-KD and NC hPSCs were the same as that of hPSCs. Rock inhibitor Y27632 (Merck Millipore) was suggested to be added into the culture medium for 1 day at the final concentration of 10  $\mu$ M during the first

few passages of *CCDC144NL-AS1*-KD hPSCs. Single cells of these cells were obtained through Accutase (Gibco) digestion, and Y27632 (10  $\mu$ M) was added to the medium for 1 day after plating cells.

#### Plasmid construction, lentivirus production, and infection

For *CCDC144NL-AS1* knockdown, shRNA sequences targeting *CCDC144NL-AS1* were cloned into the lentivirus vector GV248 (from GeneChem company) downstream of an hU6 promoter (hU6-sh-*CCDC144NL-AS1*-Ubiquitin-EGFP-IRES-puromycin). The effective target sequence of *CCDC144NL-AS1* used in this study is GTGTGGGAAGCTATAAGCATT. For the overexpression of *CCDC144NL-AS1*, *CCDC144NL-AS1* cDNA was cloned into the lentivirus vector CV084 (from GeneChem company) downstream of a Ubi promoter (Ubi-*CCDC144NL-AS1*-SV40-Neomycin). For the production of lentiviruses, each lentivirus vector was co-transfected with packing vectors into 293T cells, then virus supernatant was collected and concentrated through ultracentrifuge (Beckman, Avanti J-301). Single cells of hPSCs were infected with viruses in TeSR-E8 medium supplemented with 2  $\mu$ g/ml polybrene (Sigma-Aldrich), and the medium was renewed 6 h after virus infection. Positively infected cells were isolated by adding 2  $\mu$ g/ml puromycin (*CCDC144NL-AS1* knockdown experiments) or 400  $\mu$ g/ml G418 (*CCDC144NL-AS1* overexpression experiments) into the culture medium 48 h after virus infection. The concentrations and time of antibiotics utilized should be adjusted according to different cell viability.

#### RNA extraction, reverse transcription, and quantitative real-time PCR

For the expression analysis, total RNA of cells was extracted by RNAiso Plus (Takara), according to the manufacturers' instructions. The reverse transcription was performed by PrimeScript RT Master Mix (Takara). Quantitative real-time PCR (qPCR) reactions were carried out using FastStart Universal SRBR Green Master (Roche). Primer sequences are listed in Additional file 1: Table S1. Relative expression levels of genes were calculated by  $\Delta\Delta$ Ct method normalized to *GAPDH* compared with control samples.

#### Immunofluorescence staining

Cultured cells or teratoma sections were fixed with 4% paraformaldehyde in PBS for 15 min, washed with PBS for three times, and then permeabilized in 0.1% Triton X-100 (1% Triton X-100 for teratoma sections) for 30 min. After blocking in 3% donkey or goat serum for an hour, primary antibodies were incubated overnight at 4 °C. Then, corresponding secondary antibodies were incubated for an hour at room temperature. The nuclei

were stained by DAPI (Roche). Antibodies are listed in Additional file 2: Table S2.

#### Western blot analysis

Cells were lysed with RIPA lysis buffer supplemented with protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche). An equal amount of protein was separated on an 8% or 10% SDS-polyacrylamide gel and then was transferred to a 0.45- $\mu$ m pore PVDF membrane (Merck Millipore). After being blocked in 5% skim milk for an hour, the membrane bars were incubated in primary antibodies overnight at 4 °C. Then, the membrane bars were washed with 0.1% TBST for three times and incubated in secondary antibodies for an hour at room temperature. Antibodies are listed in Additional file 2: Table S2. After being washed in 0.1% TBST three times, the protein signals were detected with an ECL Western Blotting Substrate kit (Thermo Scientific) on Amersham Imager 600 (GE Healthcare Life Sciences) system.

#### G-banding karyotype analysis

Cells mostly grown under logarithmic phase were used for G-banding karyotype analysis. Confluent cells were treated with 0.2  $\mu$ g/ml colchicine (Sigma-Aldrich) for 2 h and then digested into single cells with Accutase. The single cells were suspended in 0.075 mol/l potassium chloride for 25 min at 37 °C and fixed for 30 min at room temperature under a solution consisting of methyl alcohol and glacial acetic acid at the volume ratio of 2:1. Then, the karyotypes were analyzed in Da An Health Testing Center (Shanghai, China).

#### Directed differentiation in vitro

Directed differentiation to neuronal cells was performed by PSC Neural Induction Medium (Life Technologies). Cells with 15–20% confluency were changed into the complete neural induction medium, including Neurobasal Medium and Neural Induction Supplement, for 7 days. On day 7, single cells of NSCs (passage 0) were harvested by utilizing Accutase and re-seeded on a Matrigel-coated plate under complete Neural Expansion Medium which included Neurobasal Medium, Advanced DMEM/F-12, and Neural Induction Supplement at the volume ratio of 49:49:2. Rock inhibitor Y27632 was added into the medium for 1 day at the final concentration of 5  $\mu$ M during the first 4 passages of NSCs. Endodermal differentiation of hPSCs was performed by culturing cells under the KnockOut DMEM (Gibco) medium supplemented with 20% fetal bovine serum (FBS, Hyclone), 1% MEM NEAA (Gibco), 1% L-glutamine (Gibco), 1% penicillin-streptomycin (Gibco), 20 ng/ml Activin A (PeproTech), and 50 ng/ml BMP4 (PeproTech) for 7 days. hPSCs were differentiated into mesodermal cells with STEMdiff™ Mesoderm Induction Medium (STEM-CELL). Cells were dissociated into single cells and seeded

onto a Matrigel-coated 24-well plate at a density of  $1 \times 10^5$  cells per well under TeSR-E8 medium supplemented with rock inhibitor Y27632 (10  $\mu$ M). From the next day, cells were changed into STEMdiff™ Mesoderm Induction Medium for 4 days and then harvested for subsequent analyses.

#### Teratoma formation

Human PSCs grown to 70–80% confluency were scraped off the plates, centrifuged, and re-suspended in 100  $\mu$ l TeSR-E8 medium per well of a 6-well plate. The 100- $\mu$ l cells were used to inject subcutaneously into a dorsal flank site of a non-obese diabetic severe combined immunodeficiency (NOD/SCID) mouse. About 7–11 weeks after injection, the tumors were dissected, fixed in 4% paraformaldehyde, sectioned, and then stained by hematoxylin and eosin.

#### RNA-seq library construction and sequencing

For RNA-seq, total RNA of samples was extracted by RNAiso Plus (Takara) following the product's descriptions. RNA-seq libraries were constructed by Novogene (Beijing, China) and sequenced on Hiseq-4000 system (Illumina).

#### Chromatin immunoprecipitation

Cells were incubated in 1% formaldehyde for 10 min at room temperature, and the cross-linking process was terminated by adding glycine at the final concentration of 125 mM. The cells were lysed and then digested with micrococcal nuclease (NEB) for 20 min at 37 °C. After the cells were sonicated, the extracts were immunoprecipitated by antibodies against H3K4me3 (ab8580, Abcam) or H3K27me3 (39155, ACTIVE MOTIF) overnight at 4 °C. The extracts were then incubated in magnetic beads for 2 h at 4 °C, and the beads were washed. Finally, the DNA products were eluted and purified. ChIP-seq libraries were constructed by Novogene (Beijing, China) and sequenced with the Novaseq-PE150 system (Illumina).

#### Public datasets

The datasets (GSE60945, E-MTAB-2031, E-MTAB-4461) used in this study, including three human reset PSC samples, three 3iL hESC samples, six conventional primed hPSC samples, and nine HNES samples, were downloaded from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) and ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress>).

#### Read mapping and transcripts assembly of RNA-seq data

Reads of RNA-seq were aligned to human reference genome (hg19) by the spliced read aligner Hisat2 (version 2.1.0) [47]. The gene abundances were estimated and normalized to fragments per kilobase of transcript per million fragments mapped (FPKM) using StringTie (version 2.2.1) [48] with default settings, and read counts of genes were summarized by HTSeq (version 0.7.2) [49].

#### Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) [50] in R package was used in this study to construct the weighted gene co-expression network of multiple samples. A major step in the module-centric calculation is to cluster genes of similar interconnection expression patterns into network modules. To construct a signed weighted co-expression network, we first calculate the adjacency matrix, which preserves the underlying continuous correlation instead of a dichotomizing one, of weighted Pearson correlation of all pair-wise genes with an index  $\beta = 20$ . Then, we use the topological overlap measure (TOM) as proximity, which is used as an input of hierarchical clustering, to combine the correlation between pair-wise genes with that of these two genes share with other “third party” genes, and this algorithm conforms to the actual situation of intergenic relationships in biological pathways.

#### Statistical analysis, gene function enrichment analysis, and gene set enrichment analysis

Differential expression analysis of genes in this study was identified by R Bioconductor package DESeq2 [51], and genes with adjusted  $p$  value  $< 0.05$  were considered as differentially expressed. ClusterProfiler [52] in R was used for GO and KEGG term gene function enrichment analysis and visualization. The significance test of prior defined functional gene set between genes expressed by different groups of samples was performed by gene set enrichment analysis (GSEA) [53, 54].

#### ChIP-seq data analysis

The ultrafast, memory-efficient read aligner bowtie2 (version 2.2.6) [55] was used to align the reads of ChIP-seq to the human reference genome (hg19). The output of bowtie2 was sorted and indexed by SAMtools [56]. Peak calling was performed by MACS2 (version 2.1.1.20160309). DeepTools [57] was used for the alignment of reads as input and normalized to reads per kilobase and million mapped reads (RPKM) and computing signals over a set of regions from 3 kb upstream of the TSS to 3 kb downstream of the TES of all genes and developmental genes. The signal visualization of interested genes was conducted by DeepTools and Integrative Genome Viewer (IGV) [58]. Read count frequency of regions which range from 3 kb upstream to 3 kb downstream of the TSS was performed by ChIPseeker [59].

## Results

#### Candidate lncRNAs tightly correlated with human pluripotency were screened through bioinformatic and experimental approaches

We initially attempted to characterize lncRNAs that support pluripotency maintenance of human PSCs. To screen lncRNAs that were specifically expressed in human PSCs, we collected a total of 55 RNA-seq data,

which contained 21 human iPSC (hiPSC) samples, 15 human ESC (hESC) samples, and 19 human somatic tissue samples, from GEO datasets (Fig. 1a, Additional file 3: Table S3). Unsupervised hierarchical clustering of the 55 RNA-seq data (FPKM  $\geq$  0.1) demonstrated that despite occasional intersections, all 21 hiPSC and 15 hESC samples clustered together and were clearly separated from the 19 human somatic tissues (Fig. 1b).

Next, all 55 RNA-seq data were subjected to weighted gene co-expression network analyses (WGCNA). We obtained multiple gene network modules representing properties of different human somatic tissues and identified top 3 notable modules (MEred, MElightcyan, and MEpurple) shared by all human PSC samples (Fig. 1c, d; Additional file 4: Figure S1a). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of these three modules revealed that their functions were primarily associated with RNA metabolism and processing, DNA replication and repair, chromatin organization and modification, oocyte meiosis and cell cycle, and protein catabolic processes (Additional file 4: Figure S1b-d).

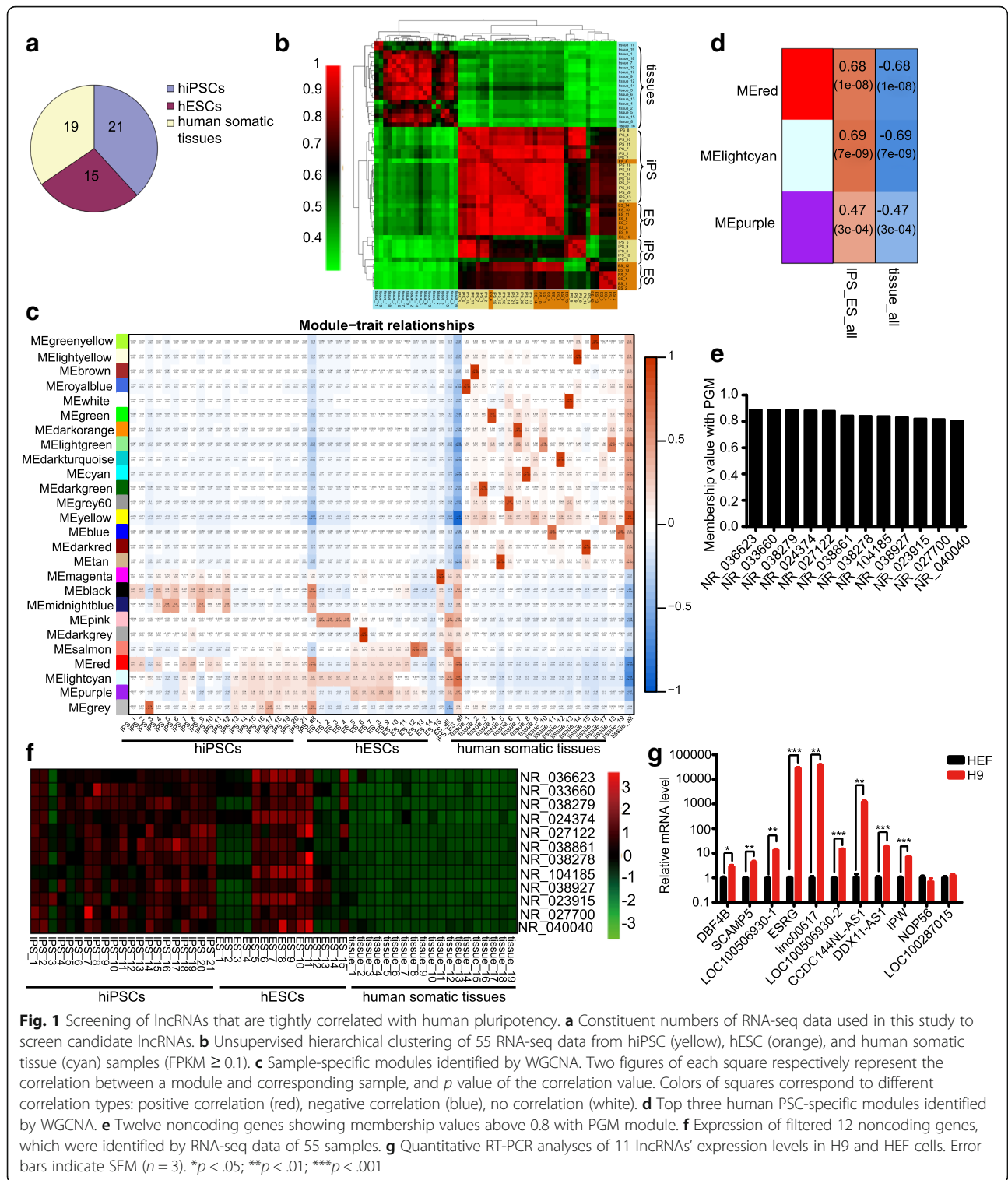
To obtain candidate lncRNAs tightly correlated with human pluripotency, we took 32 pluripotency-associated genes as a module (PGM) (Additional file 4: Figure S1e) and calculated the membership values for all noncoding genes from MEred, MElightcyan, and MEpurple modules with it. Then, 12 lncRNAs whose membership values were above 0.8 stood out (Fig. 1e). The transcriptional levels of these 12 lncRNAs in most human PSC samples were obviously higher than that in somatic tissues according to RNA-seq data (Fig. 1f). Furthermore, we performed qPCR experiments to examine their expression in one human ESC line (H9) and human embryonic fibroblast (HEF) cells. Leaving out a pseudogene NR\_024374, nine lncRNAs were validated to be significantly highly expressed in H9 cells relative to HEF cells (Fig. 1g). It should be noted that functions of five validated lncRNAs were already annotated, such as *DBF4B* and *LINC00617(TUNAR)*, the former played an important role in DNA replication and cell proliferation, and the latter was tested to be essential in maintaining pluripotency and the neural differentiation of embryonic stem cells [35, 60, 61]. The encompassment of known lncRNAs for pluripotency maintenance in our result suggested that our method was effective. Therefore, we focused on the remaining four uncharacterized lncRNAs in the following experiments, including *LOC100506930-1* (NR\_038278), *LOC100506930-2* (NR\_038279), *CCDC144NL-AS1* (NR\_104185), and *DDX11-AS1* (NR\_038927).

#### ***CCDC144NL-AS1* knockdown in human PSCs drives them to a naïve-like pluripotency state**

To functionally identify the potential roles of four lncRNAs in regulating human PSCs' pluripotency, we designed short hairpin RNAs (shRNAs) for them. At

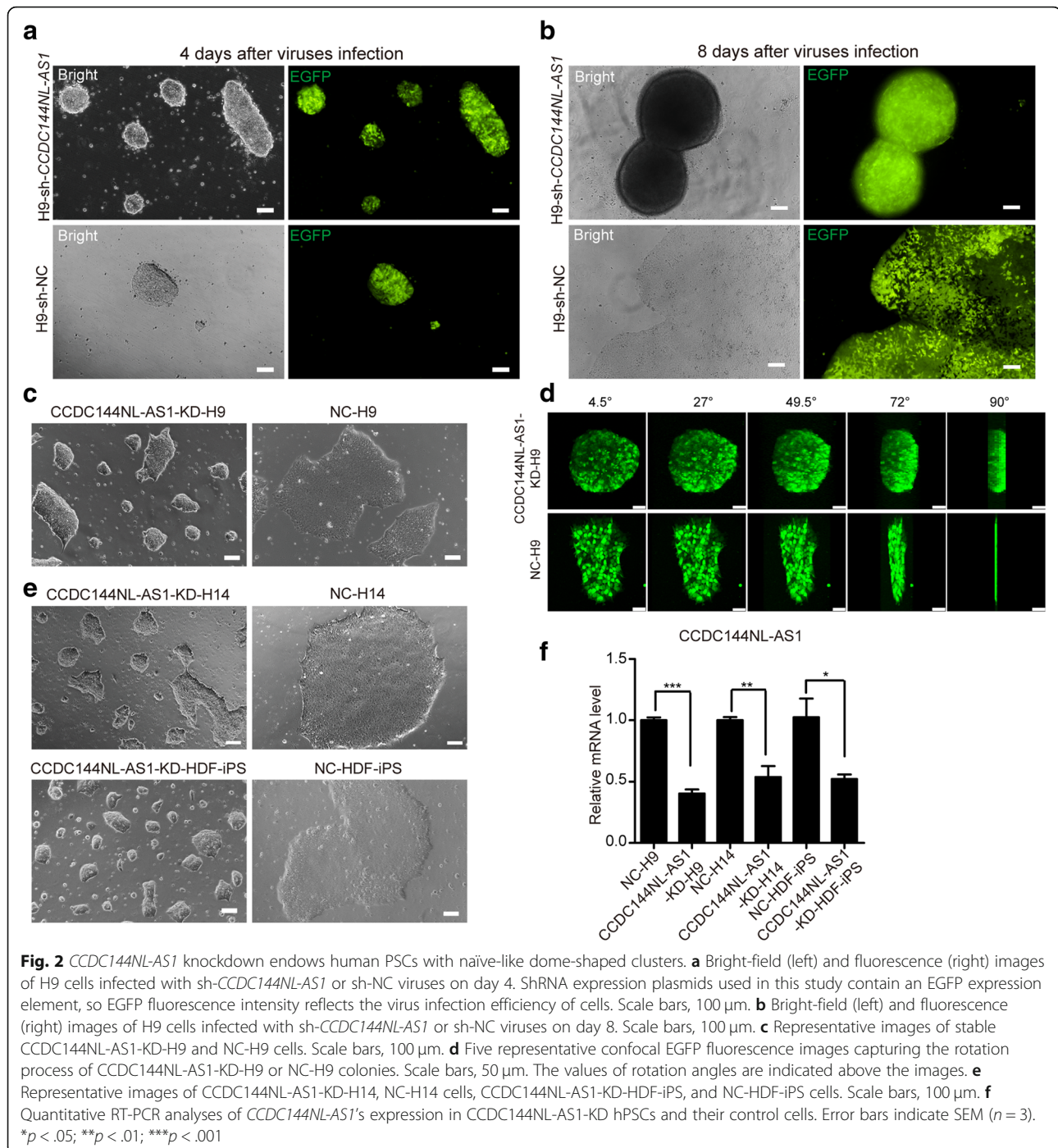
first, we focused on *CCDC144NL-AS1*, whose expression in H9 cells was hundreds-fold higher than that in HEF (Fig. 1g). When we infected human ES cell line H9 cells with shRNA-expressing viruses, we found that the expression of shRNA targeting *CCDC144NL-AS1* drove the edges of H9 flat clones to curl up on day 4 (Fig. 2a). And the clones were further rolled up, even completely became spheres on day 8 (Fig. 2b). It should be noted that the cells needed to be passaged between day 4 and day 8 to avoid excessive cell death. We passaged the cells and obtained stable *CCDC144NL-AS1-KD-H9* and *NC-H9* cells after several rounds of puromycin resistance screening (Fig. 2c). The effectively decreased RNA levels of *CCDC144NL-AS1* in *CCDC144NL-AS1-KD-H9* cells were confirmed by qPCR analyses (Fig. 2f). From the above morphology of ES cells, we can see that the downregulation of *CCDC144NL-AS1* drove H9 to naïve-like dome-shaped colonies from the primed flat disc ones. We also performed confocal layer scanning of these two different cluster types accompanied by 90° horizontal rotation (Additional file 5: Video S1 and Additional file 6: Video S2.). The videos and five representative images from each group clearly exhibited that *CCDC144NL-AS1-KD-H9* clusters were much more compact and thicker than single flat layers of *NC-H9* (Fig. 2d). For H9 is a female human ES cell line, we therefore examined whether *CCDC144NL-AS1* knockdown would engender similar effects in a male human ES cell line (H14) and a male human iPSC cell line (HDF-iPS) derived from human dermal fibroblasts. Likewise, we observed primed-to-naïve-like morphology changes in H14 and HDF-iPS cells after *CCDC144NL-AS1* knockdown (Fig. 2e, f).

Next, we performed a series of experiments to characterize the pluripotent state of *CCDC144NL-AS1-KD-H9* cells. Our qPCR analyses revealed that the expression of core pluripotency factors *OCT4*, *SOX2*, and *NANOG* changed differently in *CCDC144NL-AS1-KD-H9* cells. *SOX2* was significantly decreased, and *NANOG* was markedly elevated and no obvious changes in *OCT4* levels (Fig. 3a). Among them, *NANOG* was considered essential for the establishment of naïve pluripotency [62]. We also examined the expression levels of lineage commitment factors and found that both endodermal marker *ALB* and mesodermal marker *BRACHYURY* were evidently decreased in *CCDC144NL-AS1-KD-H9* cells relative to *NC-H9* cells (Fig. 3b). Furthermore, we observed obviously elevated expression of naïve pluripotency-associated transcription factors, such as *KLF2*, *KLF4*, *DPPA2*, *DPPA3*, and *TFCP2L1* (Fig. 3c). In addition, *CCDC144NL-AS1-KD-H9* cells stained positive for pluripotency markers *OCT4*, *SOX2*, *NANOG*, *SSEA3*, *SSEA4*, *TRA-1-60*, and naïve pluripotency-specific *TFCP2L1* (Fig. 3d, e). And western blot results further corroborated the expression changes of *SOX2*,



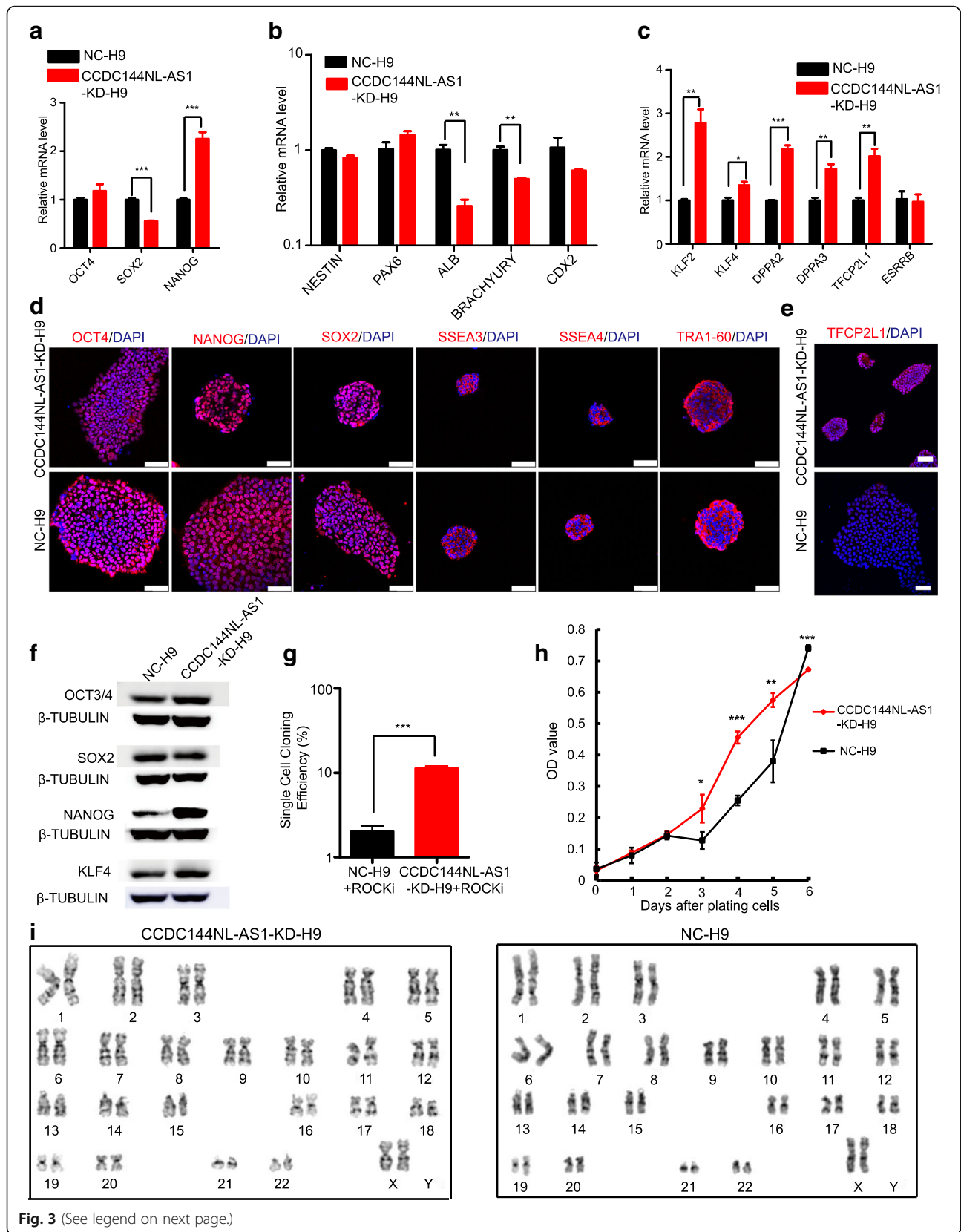
NANOG, and KLF4 at the protein level (Fig. 3f). In addition, we also examined the expression levels of these genes in CCDC144NL-AS1-KD-H14 and CCDC144NL-AS1-KD-HDF-iPS cells. In CCDC144NL-AS1-KD-H14 cells, we observed obviously increased expression of naive

pluripotency-associated genes *NANOG*, *KLF2*, *DPPA2*, *DPPA3*, and *TFCP2L1* and evidently decreased neuroectodermal marker *PAX6* and trophectoderm marker *CDX2* (Additional file 7: Figure S2a, b). In CCDC144NL-AS1-KD-HDF-iPS cells, we observed obviously increased *NANOG*,



*KLF2*, *KLF4*, *DPPA2*, *DPPA3*, and *TFCP2L1* and decreased neuroectodermal marker *PAX6* and endodermal marker *ALB* (Additional file 7: Figure S2c, d). Furthermore, both of these cells stained positive for pluripotency markers OCT4, SOX2, NANOG, SSEA3, SSEA4, TRA-1-60, and naïve pluripotency-specific *TFCP2L1* (Additional file 7: Figure S2e, f). And western blot results further validated the expression changes of NANOG and *KLF4* in these cells (Additional file 7: Figure S2g).

We then measured the single-cell cloning efficiency of these cells and found that *CCDC144NL-AS1*-KD-H9, *CCDC144NL-AS1*-KD-H14, and *CCDC144NL-AS1*-KD-HDF-iPS cells presented much higher single-cell cloning efficiency than control cells (Fig. 3g, Additional file 8: Figure S3 a). Although the absolute values are not identical to those in previous reported human naïve PSCs, possibly owing to the differences in cell culture and technical systems in the analyses, the upregulated trend





(See figure on previous page.)

**Fig. 3** Naïve-like pluripotency validation of *CCDC144NL-AS1*-downregulated human ESCs. Quantitative RT-PCR analyses of three core pluripotency genes (a), lineage commitment factor genes (b), and naïve pluripotency-associated genes (c) in CCDC144NL-AS1-KD-H9 and NC-H9 cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . Immunostaining images of pluripotency-associated markers OCT4, NANOG, SOX2, SSEA3/4, and TRA-1-60 (d) and naïve pluripotency-specific marker TFCP2L1 (e). Scale bars, 75  $\mu\text{m}$ . f Western blot detection of naïve pluripotency-related transcription factors, such as NANOG and KLF4, and shared pluripotency transcription factors, such as OCT4 and SOX2, in CCDC144NL-AS1-KD-H9 and NC-H9 cells.  $\beta$ -Tubulin was used as an endogenous control. g Single-cell cloning efficiencies of CCDC144NL-AS1-KD-H9 and NC-H9 cell lines. Error bars indicate SEM ( $n = 3$ ); \*\*\*,  $p < .001$ . h Growth curves of CCDC144NL-AS1-KD-H9 and NC-H9 cell lines. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . i Karyotype analysis of CCDC144NL-AS1-KD-H9 and NC-H9 cell lines

relative to the primed PSCs is similar. Naïve PSCs are characterized by high rates of proliferation, so we performed cell growth curve analysis. From our growth curve results, we can see that CCDC144NL-AS1-KD-H9 cells showed obviously increased OD values, which reflected elevated cell numbers, from day 3 to day 5 than NC-H9 cells, but they were surmounted by control cells on day 6 (Fig. 3h). Remarkably, CCDC144NL-AS1-KD-H9 cells entered the exponential phase of growth earlier than NC-H9 cells; however, we suppose, their gradually increasing layers of cells piled up within limited space led to lowered cell growth on day 6. Whereas, primed H9 cells maintained a single layer growth pattern from the beginning to end in this experiment. The growth curves indicate that *CCDC144NL-AS1* downregulation drives H9 cells to proliferate faster, and the tendency is concordant with the naïve pluripotency property. Additionally, we assessed the cell cycle on day 4 after plating the cells. CCDC144NL-AS1-KD-H9 cells revealed obviously increased S+G2/M phase cells relative to the control, indicating that they grow faster than primed cells (Additional file 8: Figure S3 b-c). Whereas, CCDC144NL-AS1-KD-H14 and CCDC144NL-AS1-KD-HDF-iPS did not reveal remarkably increased S+G2/M phase cells than control, suggesting that they did not grow much faster than control cells. Besides, both CCDC144NL-AS1-KD-H9 and NC-H9 cells maintained a normal karyotype (Fig. 3i).

It should be clear that the naïve degree of published human naïve ESCs generated from different conditions varies from each other. Obviously, CCDC144NL-AS1-KD-H9 cells own the naïve-like colonies, express higher naïve pluripotency-related genes, and possess naïve-like cell growth patterns, despite that the naïve pluripotency degree of them is somewhat restricted.

#### CCDC144NL-AS1-downregulated human PSCs exhibit elevated developmental potential in vitro and in vivo

Some human naïve induction medium endowed their naïve PSCs with more efficient differentiation abilities [18, 63]. To investigate the differentiation capacity of our CCDC144NL-AS1-KD-H9 cells in vitro, we directed them and control cells towards neuronal, endodermal, and mesodermal cells under lineage-specific differentiation medium

(Additional file 9: Figure S4a). Quantitative PCR analyses revealed that CCDC144NL-AS1-KD-H9-derived neural precursor cells (NPCs) expressed much higher levels of neuronal/neuroectodermal-specific marker genes *SOX1*, *PAX6*, and *SOX2* and much lower levels of endodermal-specific gene *GATA6* compared with control cells (Fig. 4a), illustrating that *CCDC144NL-AS1* downregulation endowed H9 cells with more homogeneity and higher efficiency towards neuronal differentiation. Western blot analyses further confirmed the elevated SOX1 and PAX6 protein levels in CCDC144NL-AS1-KD-H9-derived NPCs (Fig. 4b). Upon differentiation of endoderm, we also observed efficient differentiation potential of *CCDC144NL-AS1* knock-down H9 cells relative to their control counterparts, with RNA levels of endodermal-specific *GATA4* and *GATA6* significantly increased (Fig. 4c) and protein levels of endodermal-specific E-CADHERIN and *GATA6* remarkably elevated (Fig. 4d). On mesodermal differentiation, CCDC144NL-AS1-KD-H9-derived mesodermal cells expressed obviously higher levels of mesodermal-specific gene *BRACHYURY* and lower levels of endodermal-specific gene *GATA6* relative to control cells (Fig. 4e). It was hard to detect the expression of neuroectodermal-specific *SOX1* and *PAX6*, because both groups expressed very low levels of them. In addition, mesodermal-specific *BRACHYURY* was verified higher in protein levels in CCDC144NL-AS1-KD-H9-derived mesodermal cells compared with control cells (Fig. 4f). Both results demonstrated that *CCDC144NL-AS1*-downregulated H9 cells differentiated to mesodermal cells with more homogeneity and higher efficiency.

Furthermore, we also assessed the differentiation capacity of CCDC144NL-AS1-KD-H14 and CCDC144NL-AS1-KD-HDF-iPS cells in vitro. Upon neuroectoderm differentiation, CCDC144NL-AS1-KD-HDF-iPS showed much more efficient differentiation capacity with obvious higher neuronal/neuroectodermal-specific genes *SOX1*, *PAX6*, and *SOX2* expression relative to their control cells (Additional file 9: Figure S4 b,c), while the expression levels of *SOX1*, *PAX6*, and *SOX2* in CCDC144NL-AS1-KD-H14-derived NPCs were lower than those in their counterparts (Additional file 9: Figure S4 d,e). On endodermal differentiation, in accordance with CCDC144NL-AS1-KD-H9, both of these two cells revealed evidently increased endodermal-specific *GATA4*

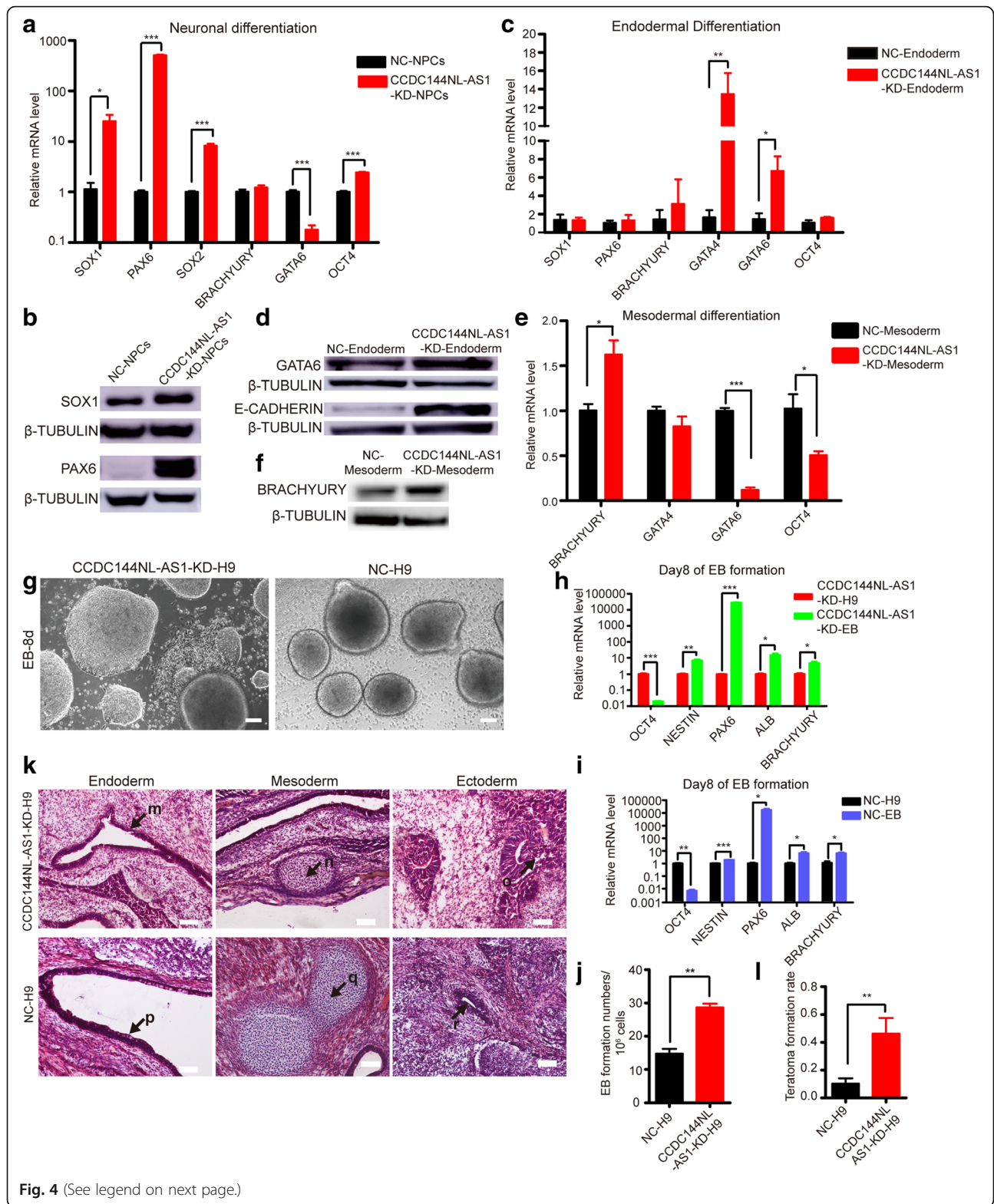


Fig. 4 (See legend on next page.)

(See figure on previous page.)

**Fig. 4** Developmental potential of CCDC144NL-AS1-downregulated human ESCs in vitro and in vivo. Quantitative RT-PCR analyses of expression levels of lineage-specific genes in CCDC144NL-AS1-KD-H9- and NC-H9-derived neuronal (**a**), endodermal (**c**), and mesodermal (**e**) cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . Western blot analyses of lineage-specific genes' expression in CCDC144NL-AS1-KD-H9- and NC-H9-derived neuronal (**b**), endodermal (**d**), and mesodermal (**f**) cells.  $\beta$ -Tubulin was used as an endogenous control. **g** Morphologies of embryoid bodies differentiated from CCDC144NL-AS1-KD-H9 and NC-H9 cells. Scale bars, 100  $\mu$ m. Quantitative RT-PCR analyses of lineage-related genes in embryoid bodies generated from CCDC144NL-AS1-KD-H9 (**h**) and NC-H9 (**i**) cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . **j** Embryoid body numbers generated from per  $10^6$  CCDC144NL-AS1-KD-H9 and NC-H9 cells. Error bars indicate SEM ( $n = 3$ ). \*\* $p < .01$ . **k** Hematoxylin and eosin staining of teratomas containing tissues of all three embryonic germ layers derived from CCDC144NL-AS1-KD-H9 and NC-H9 cells. (m, p) Gut-like epithelium (endoderm). (n, q) Cartilage (mesoderm). (o, r) Neural tissue (ectoderm). Scale bars, 500  $\mu$ m. **l** Teratoma formation rate of CCDC144NL-AS1-KD-H9 and NC-H9 cells. Error bars indicate SEM ( $n = 13$ ). \*\* $p < .01$

and *GATA6* gene expression compared with control cells (Additional file 9: Figure S4 f-i). When differentiated into mesodermal cells, CCDC144NL-AS1-KD-H14 showed significantly elevated expression of mesodermal-specific *BRACHYURY* relative to the control (Additional file 9: Figure S4 j-k), while CCDC144NL-AS1-KD-HDF-iPS cells did not reveal increased mesodermal differentiation compared with control cells (Additional file 9: Figure S4 l-m). The results above indicated that when directed differentiation in vitro, human PSCs may respond differently due to *CCDC144NL-AS1* downregulation.

Next, we performed in vitro spontaneous differentiation of CCDC144NL-AS1-KD-H9 and NC-H9 cells by generating embryoid bodies (EBs). After 8 days of suspending ES clones in TeSR-E6 differentiation medium, we obtained EBs that expressed significantly upregulated marker genes of three germ layers (Fig. 4g-i). It is noteworthy that there were always some EBs being attached to the plate bottom at day 8 in CCDC144NL-AS1-KD-H9 cells (Fig. 4g, left), whereas control cells were not (Fig. 4g, right). We speculated that the much more compact cell mass of CCDC144NL-AS1-KD-H9 endowed them with stronger propensities to form clones. In addition, we also examined the expression levels of some important cell adhesion molecules and found that CCDC144NL-AS1-KD-H9-derived EBs expressed evidently higher cell adhesion-associated genes *E-CAD*, *ZO-1*, *CTNNA1*, *MUC1*, and *DSG2* than NC-H9-derived EBs, in accordance with the phenomenon we observed (Additional file 10: Figure S5 a). Moreover, CCDC144NL-AS1-KD-H9 cells displayed markedly elevated levels in our assay of EB formation numbers per  $10^6$  cells in contrast with control cells (Fig. 4j). This result was concordant with those of high rates of proliferation and increased single-cell cloning efficiencies of CCDC144NL-AS1-KD-H9 cells (Fig. 3g, h), all of which reflected their higher cell viability. We also evaluated the EB formation ability of CCDC144NL-AS1-KD-H14, CCDC144NL-AS1-KD-HDF-iPS, and their counterparts. All of them could form EBs, and CCDC144NL-AS1-KD-H14 and CCDC144NL-AS1-KD-HDF-iPS showed significantly increased EB numbers per  $10^6$  cells relative to control cells (Additional file 10: Figure S5 b-g).

For differentiation in vivo, human CCDC144NL-AS1-KD-H9 and NC-H9 cells were injected subcutaneously into NOD/SCID mice for teratoma formation. Both groups obtained teratomas containing tissues from all three germ layers (Fig. 4k, Additional file 11: Figure S6 a, b), whereas H9 cells whose *CCDC144NL-AS1* expression was downregulated showed significantly elevated teratoma formation rate, suggesting that the downregulation of *CCDC144NL-AS1* improved in vivo developmental capacity of H9 cells (Fig. 4l). For CCDC144NL-AS1-KD-H14, CCDC144NL-AS1-KD-HDF-iPS, and their control cells, we also performed teratoma formation experiments. All of them could obtain teratomas containing tissues from all three germ layers (Additional file 11: Figure S6 c). In accord with the CCDC144NL-AS1-KD-H9 cells, CCDC144NL-AS1-KD-H14 and CCDC144NL-AS1-KD-HDF-iPS revealed obviously elevated teratoma formation rate (Additional file 11: Figure S6 d,e).

In summary, CCDC144NL-AS1-KD-H9 cells exhibited higher efficiency in directed differentiation to neuroectoderm, endoderm, and mesoderm relative to control cells. Although CCDC144NL-AS1-KD-H14 and CCDC144NL-AS1-KD-HDF-iPS cells behaved differently in some directed differentiation aspect, lowered expression of *CCDC144NL-AS1* rendered human pluripotent stem cells to differentiate more efficiently in spontaneous EB formation in vitro and teratoma formation in vivo.

#### The transcriptome of CCDC144NL-AS1-knockdown human PSCs distinguishes from that of conventional human PSCs and resembles those of human naïve PSCs

To gain more insights into the molecular features of *CCDC144NL-AS1*-downregulated human PSCs, we assessed the transcriptional profiles of CCDC144NL-AS1-KD-H9 cells, CCDC144NL-AS1-KD-H14 cells, CCDC144NL-AS1-KD-HDF-iPS cells, and their corresponding negative control cells using RNA-seq. Unsupervised hierarchical clustering of these samples revealed that they were clearly separated into two groups due to *CCDC144NL-AS1* knockdown despite of different cell lines (Fig. 5a).

We next conducted WGCNA to detect gene co-expression network modules of CCDC144NL-AS1-KD human PSCs. Obviously, two gene co-expression modules,

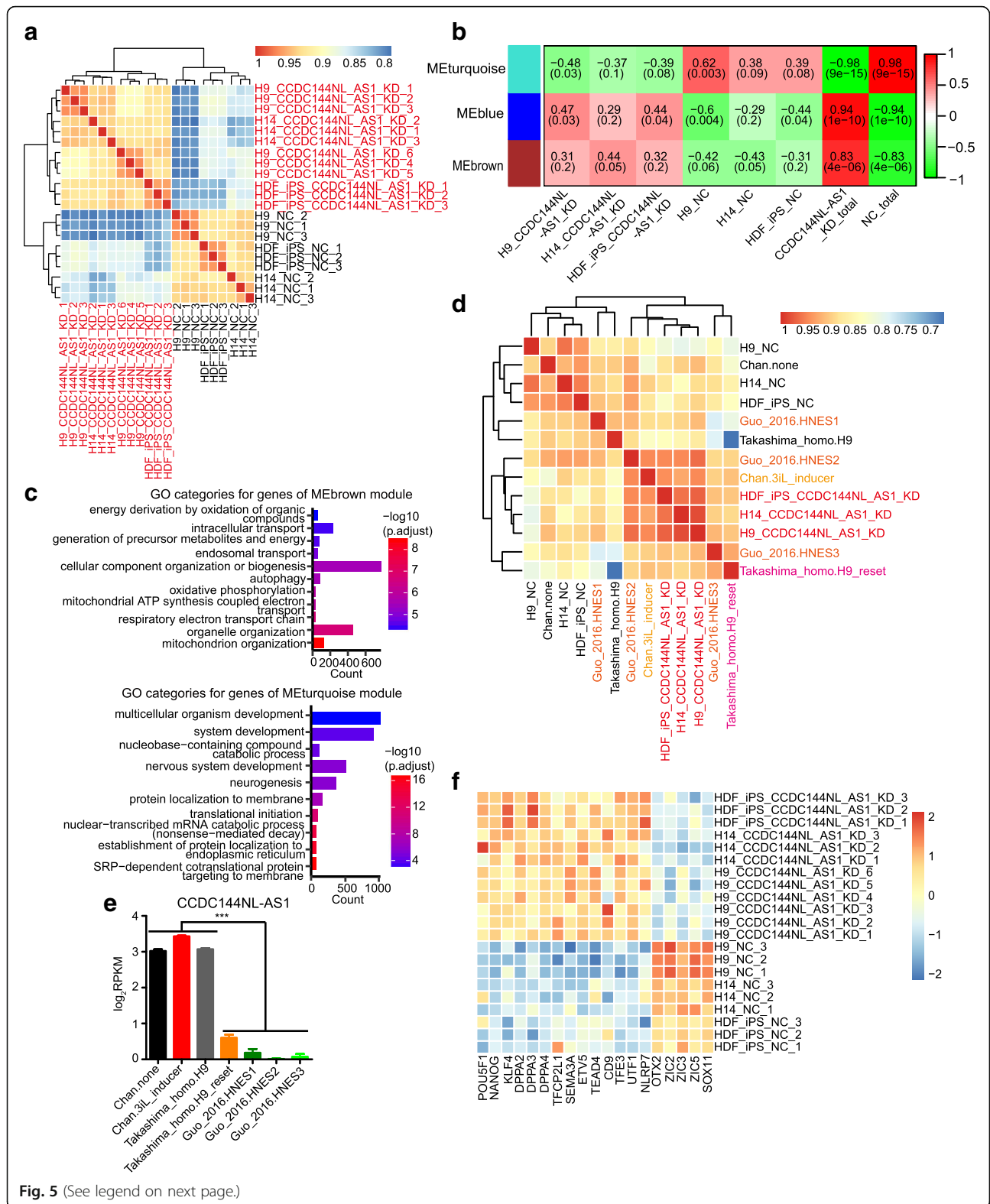


Fig. 5 (See legend on next page.)

(See figure on previous page.)

**Fig. 5** The transcriptome characteristics of CCDC144NL-AS1-knockdown hPSCs. **a** Unsupervised hierarchical clustering of RNA-seq data from CCDC144NL-AS1-KD-H9, CCDC144NL-AS1-KD-H14, CCDC144NL-AS1-KD-HDF-iPS, and their corresponding negative control cells. CCDC144NL-AS1-KD hPSC sample names are indicated in red, and all control sample names are indicated in black. **b** Three sample-specific modules identified by WGCNA. Two figures of each square respectively represent the correlation between a module and a corresponding sample, and  $p$  value of the correlation value. Colors of squares correspond to different correlation types: positive correlation (red), negative correlation (green), no correlation (white). **c** Gene Ontology (GO) analyses of MEbrown (upper panel) and METurquoise (lower panel) module genes. **d** Unsupervised hierarchical clustering of RNA-seq data from CCDC144NL-AS1-KD hPSC, negative control hPSC, and samples of three other studies Chan et al. [18], Takashima et al. [15], and Guo et al. [19]. Sample names from different experiments are indicated in different colors, and all control sample names are indicated in black. **e** The expression data of *CCDC144NL-AS1* in naïve hPSCs from Chan et al. [18], Takashima et al. [15], and Guo et al. [19] and in control hPSCs. Displayed are  $\log_2$  RPKM values. **f** Expression of genes identified by others as associating with the human naïve and primed states in CCDC144NL-AS1-KD hPSCs and control cells. The expression levels are determined by RNA-seq data.

MEbrown and MEblue, showed the highest correlation with total CCDC144NL-AS1-KD hPSC samples (Fig. 5b). On the other side, METurquoise presented high correlation with negative control hPSC samples and low correlation with CCDC144NL-AS1-KD human PSCs (Fig. 5b). GO classification and enrichment analyses exhibited that MEbrown module involved terms mainly about mitochondrion organization, respiratory electron transport chain, oxidative phosphorylation, autophagy, cellular component organization or biogenesis, and intracellular transport (Fig. 5c, upper panel). MEblue was associated with RNA processing, RNA splicing, and cellular macromolecule metabolic process (Additional file 12: Figure S7a). It is noteworthy that the terms, like “oxidative phosphorylation” and “respiratory electron transport chain,” enriched in CCDC144NL-AS1-KD cells were consistent with previous observations in mouse ESCs and human naïve PSCs, whereas mouse EpiSCs and human ESCs relied principally on aerobic glycolysis in energy generation [64–66]. Our gene set enrichment analysis (GSEA) further validated the upregulated enrichment of CCDC144NL-AS1-KD-H9 cells for respiratory chain (Additional file 12: Figure S7b). On the other hand, METurquoise module, which was negatively correlated with CCDC144NL-AS1-KD cells, unveiled GO categories related to SRP-dependent cotranslational protein targeting to the membrane, nuclear-transcribed mRNA catabolic process, translational initiation, neurogenesis, nervous system development, and system development (Fig. 5c, lower panel).

Afterwards, we compared the transcriptional profiles of our CCDC144NL-AS1-KD human PSCs with those of previously reported human naïve PSCs. We selected three groups of sequencing data coming from Chan et al. [18], Takashima et al. [15], and Guo et al. [19]. Unsupervised hierarchical clustering of these samples and ours revealed that CCDC144NL-AS1-KD hPSCs preferred to cluster with the naïve datasets of Chan et al. and Takashima et al. and HNES2 and HNES3 samples from Guo et al., while conventional PSCs showed high correlation with the primed parts (Fig. 5d). Naïve cells from Takashima et al. were demonstrated as one of the representative naïve cell types similar to human pre-implantation embryos [67].

Human naïve embryonic stem (HNES) cells from Guo et al. were generated directly from isolated cells of human inner cell mass and cultured in “t2iLGöY” medium. However, we can also see that HNES1 from Guo et al. was intermixed with conventional PSCs, and the other two HNES cell lines were closer to the naïve parts (Fig. 5d). This may have resulted from the internal individual properties of HNES1 samples. Additionally, we examined the expression pattern of *CCDC144NL-AS1* in these previously reported human naïve PSCs and found that *CCDC144NL-AS1* was obviously downregulated in naïve cells from Takashima et al. and Guo et al. but not in naïve cells from Chan et al. (Fig. 5e). This result suggested that the relatively lower expression level of *CCDC144NL-AS1* might be an important aspect for human naïve PSCs which were more similar to human pre-implantation embryos.

Next, in agreement with the expression patterns of genes associated with human naïve pluripotency, our RNA-seq data confirmed significantly elevated levels of naïve hallmarks in CCDC144NL-AS1-KD hPSCs relative to control cells, such as *NANOG*, *KLF4*, *DPPA2*, *DPPA3*, *DPPA4*, *TFCP2L1*, and *SEMA3A* (Fig. 5f). Meanwhile, the obviously reduced expression of primed pluripotency regulators such as Zinc finger of the cerebellum (*ZIC*) protein family transcription factors, *ZIC2*, *ZIC3*, *ZIC5*, and other lineage-priming factors, *OTX2* and *SOX11*, were also observed in CCDC144NL-AS1-KD hPSCs compared with control cells [68, 69] (Fig. 5f).

Generally speaking, *CCDC144NL-AS1* downregulation endowed human PSCs with distinct transcriptome characteristics from the conventional ones, and more importantly, CCDC144NL-AS1-KD human PSCs exhibited a lot of resemblances with human naïve PSCs.

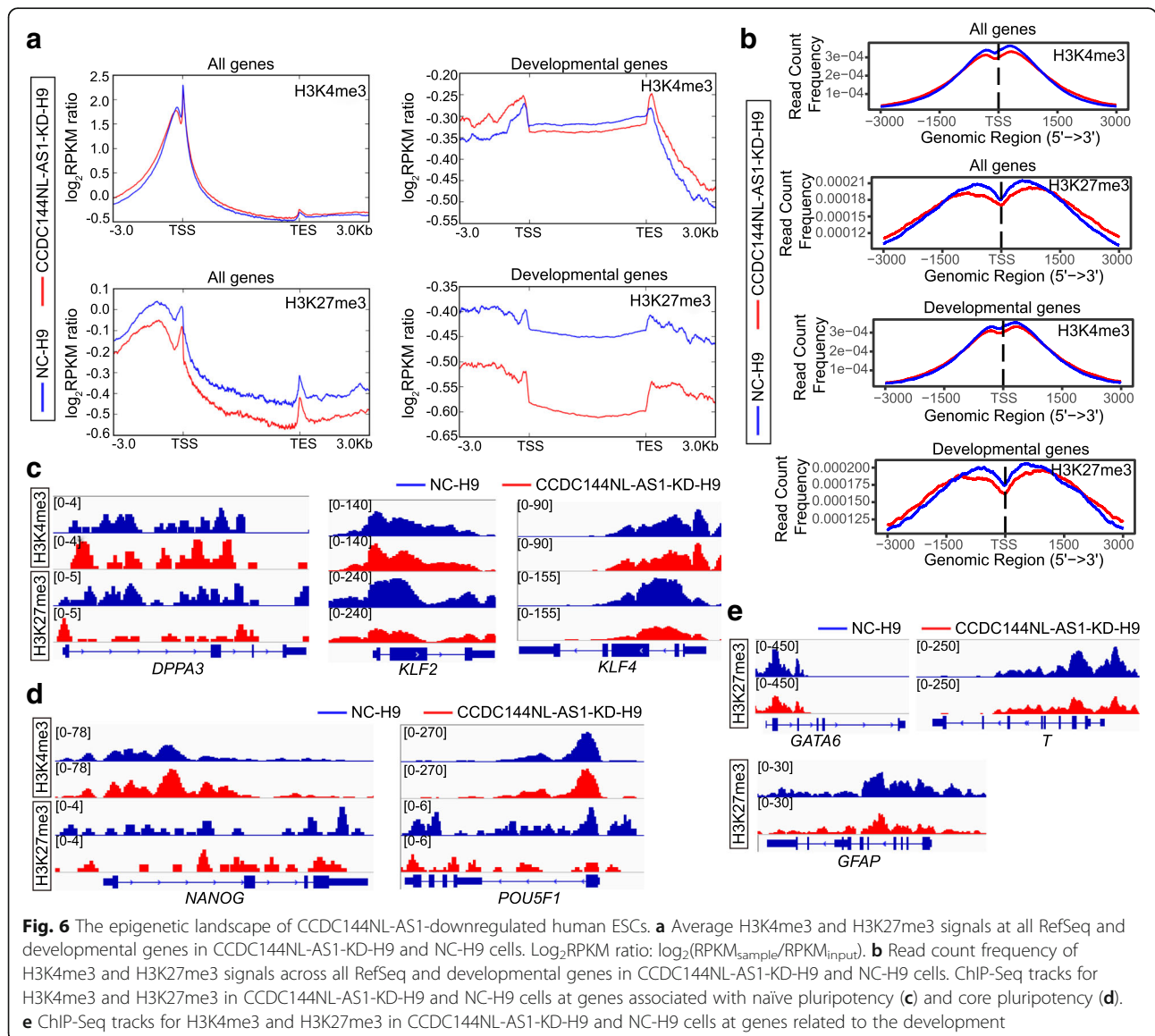
#### CCDC144NL-AS1-downregulated human ESCs possess some epigenetic characteristics similar to naïve pluripotency

To evaluate the epigenetic profiles of CCDC144NL-AS1-KD-H9 cells, we performed chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq)

analyses of trimethylation of histone 3 lysine 4 (H3K4me3) and trimethylation of histone3 lysine 27 (H3K27me3), which marked the transcriptional active and silencing of chromatin state, respectively. Compared to negative control H9 cells, the genome-wide signals of H3K4me3 showed no obvious change in CCDC144NL-AS1-KD-H9 cells, whereas a significant reduction of genome-wide H3K27me3 signals was observed in CCDC144NL-AS1-KD-H9 cells (Fig. 6a). In addition, the decrease of H3K27me3 signals was also found over developmental genes in CCDC144NL-AS1-KD-H9 cells as compared to that in control cells (Fig. 6a). This epigenetic pattern of these two chromatin marks in CCDC144NL-AS1-KD-H9 cells showed similarities to that in human naïve stem cells [14]. The read count frequency of H3K4me3 signals across the global and

developmental genes both revealed that there were slight rises over transcription start site (TSS) nearby the regions, but no change over the gene body or intergenic regions in CCDC144NL-AS1-KD-H9 cells compared with control cells (Fig. 6b). The reductions were more obvious for H3K27me3 read count frequency over TSS nearby regions of global and developmental genes in CCDC144NL-AS1-H9 cells relative to control cells (Fig. 6b).

We then examined the histone methylation profile at different gene loci. Naïve pluripotency-associated genes, such as *DPPA3*, *KLF2*, and *KLF4*, revealed apparently decreased distribution of H3K27me3 in CCDC144NL-AS1-KD-H9 cells as compared to that in control cells (Fig. 6c). The downregulation of H3K27me3 signals in CCDC144NL-AS1-KD-H9 cells was also observed at the



loci of core pluripotency factors *OCT4* (*POU5F1*) and *NANOG* (Fig. 6d). On the contrary, H3K4me3 exhibited no obvious change at most of these gene loci, except for enhanced signals at *DPPA3* and *NANOG* (Fig. 6c, d). Developmental genes in human primed PSCs were known as bivalent genes marked by the emergence of H3K4me3 and H3K27me3 simultaneously. In CCDC144NL-AS1-KD-H9 cells, genes involved in the development, such as *GATA6*, *T*, and *GFAP*, revealed evidently decreased H3K27me3 deposition relative to NC-H9 cells, just like they do in human naïve PSCs relative to the primed (Fig. 6e).

The above results illustrated that although the epigenetic profiles of CCDC144NL-AS1-KD-H9 cells are not the same as human naïve PSCs, such as not remarkably higher H3K4me3 signals at most naïve pluripotency-associated genes, downregulation of *CCDC144NL-AS1* indeed triggered a genome-wide decrease of suppressive chromatin modifications in H9 cells, which was similar to human naïve PSCs.

#### **Rescued expression of CCDC144NL-AS1 partially sets back the naïve-like state transition of human ESCs triggered by CCDC144NL-AS1 downregulation**

Next, we sought to investigate whether rescued expression of *CCDC144NL-AS1* in CCDC144NL-AS1-KD hESCs could set back the naïve-like conversion of the primed. When we rescued the expression of *CCDC144NL-AS1* in CCDC144NL-AS1-KD-H9 cells by infecting the cells with *CCDC144NL-AS1*-overexpression viruses, we observed the partial recovery of primed flat-disc colony morphology from naïve-specific dome-like clusters (Fig. 7a). CCDC144NL-AS1-KD-H9 cells infected with empty viruses were used as control cells. The rescued expression of *CCDC144NL-AS1* was confirmed by qPCR analysis (Fig. 7b). What is more, obviously reduced *NANOG* protein level was observed along with the primed pluripotency recovery (Fig. 7c, Additional file 13: Figure S8a). Our results indicated that the naïve-like conversion of primed hESCs indeed resulted from downregulation of *CCDC144NL-AS1*, and the conversion was partially reversible.

#### **CCDC144NL-AS1-downregulated human PSCs demonstrate signaling pathway dependency analogous to naïve pluripotency**

Previous studies reveal that naïve pluripotency state possesses specific signaling pathway patterns distinct from the primed state [70], so we investigated some naïve-specific signaling pathways in *CCDC144NL-AS1*-downregulated human PSCs. First of all, we investigated two pathways, including MAPK/ERK signaling pathway and canonical Wnt signaling pathway. These two pathways were involved in almost all previously published protocols to induce or maintain human naïve pluripotency by

utilizing two small molecular kinase inhibitors (2i) [13–15, 17, 18, 65]. One inhibitor is PD0325901 (PD), which inhibits the activity of mitogen-activated protein kinase kinase (MAPKK, MEK) and therefore gives rise to the suppression of MAPK/ERK signaling cascade; the other is CHIR99021 (CHIR), which targets glycogen synthase kinase-3 (GSK3) and results in augmented activity of Wnt signaling.

Although we observed obviously increased protein levels of total MAPK (ERK) in CCDC144NL-AS1-KD-H9 cells compared with NC-H9 cells, the active phospho-MAPK levels were greatly reduced (Fig. 7d), suggesting that the MAPK/ERK signaling was inhibited in CCDC144NL-AS1-KD-H9 cells. In line with this, the expression levels of MAPK/ERK signaling pathway-related genes fibroblast growth factor 2 (*FGF2*) and fibroblast growth factor receptors (*FGFR1*, *FGFR2*, and *FGFR3*) were much lower in CCDC144NL-AS1-KD hPSCs than those in control cells (Fig. 7e). As to the reasons why phosphorylation levels of MAPK were reduced despite the elevated total MAPK levels, we speculated that the phosphorylation process was inhibited through some particular way. In effect, we discovered the upregulation of a group of dual-specificity phosphatase (DUSP) family members from RNA-seq data (Fig. 7e). The protein levels of DUSP3, DUSP6, and DUSP7 were validated to be obviously higher in CCDC144NL-AS1-KD-H9 cells than those in control (Fig. 7f). DUSPs can dephosphorylate the threonine and the tyrosine residue sites of MAPKs and lead to their inactivation [71].

CCDC144NL-AS1-KD-H9 cells also showed increased protein levels of active unphosphorylated  $\beta$ -catenin relative to NC-H9 cells, indicating the higher activation degree of the Wnt signaling pathway (Fig. 7g). In consistent with the above, the expression of transcription factor 3 (*TCF3*) was reduced in CCDC144NL-AS1-KD-H9 and CCDC144NL-AS1-KD-H14 cells compared with control cells according to the RNA-seq data (Fig. 7h). *TCF3* is a transcriptional repressor of pluripotency factors and  $\beta$ -catenin can directly abrogate *TCF3* repression on the pluripotency network, so the inhibition of *TCF3* is a necessary downstream effect of Wnt signaling [72, 73]. In addition, the expression levels of Wnt target genes, such as *MYC*, *JUN*, and *FOSL1*, were elevated in CCDC144NL-AS1-KD hPSCs relative to the control, further validating the stimulated Wnt signaling pathway (Fig. 7h).

Our CCDC144NL-AS1-KD hPSCs were derived and maintained with TeSR-E8 medium. TeSR-E8 is supplemented with basic fibroblast growth factor (bFGF) and TGF- $\beta$ , these two components were confirmed to play indispensable roles in sustaining primed pluripotency in mouse EpiSCs and human ESCs, but were considered unnecessary for mouse ground state pluripotency maintenance [70]. Whereas in some human naïve PSCs

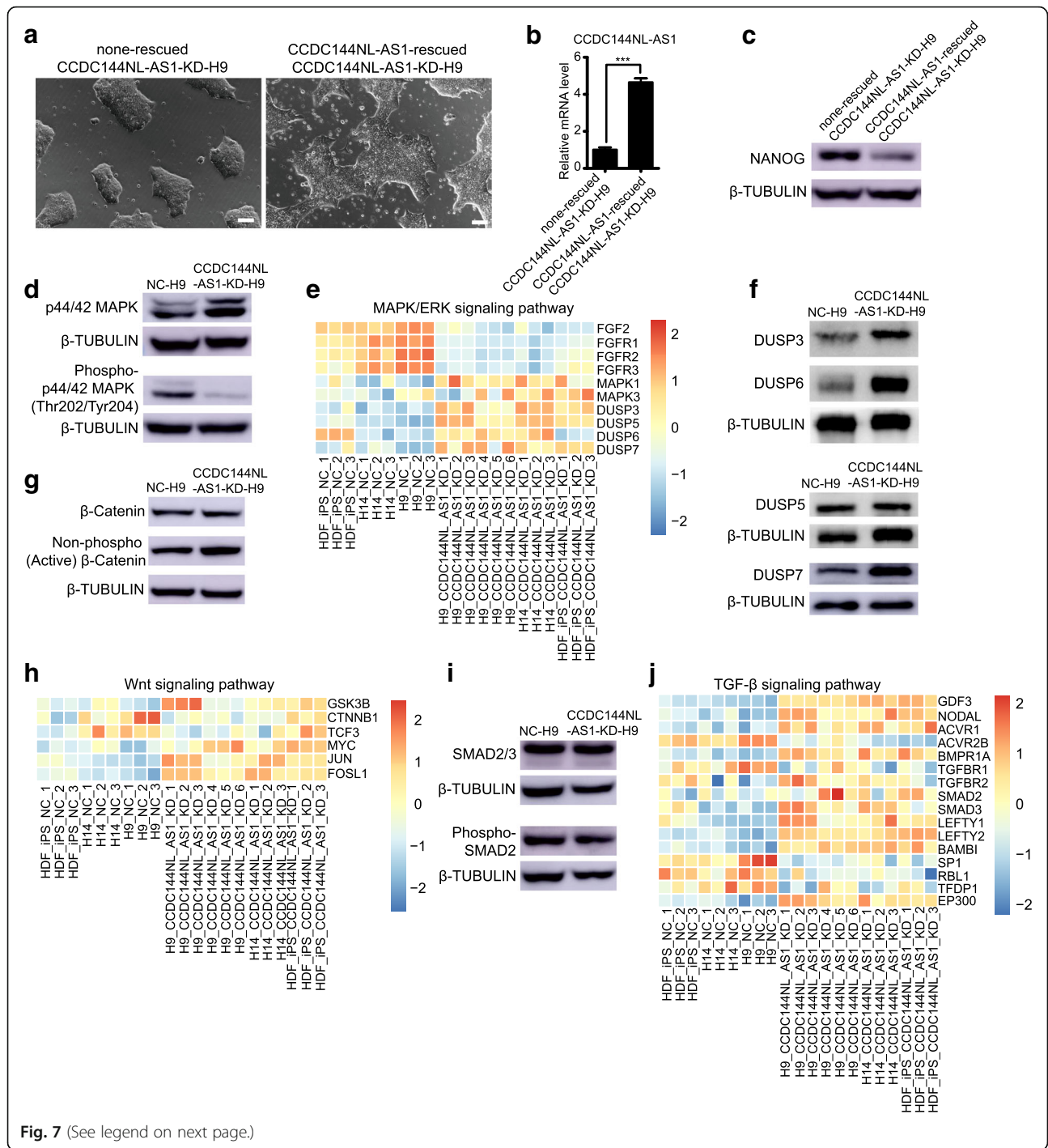


Fig. 7 (See legend on next page.)



(See figure on previous page.)

**Fig. 7** Rescued knockdown of *CCDC144NL-AS1* in CCDC144NL-AS1-KD hESCs and the signaling pathway state of CCDC144NL-AS1-KD hPSCs. **a** Morphologies of CCDC144NL-AS1-KD-H9 clusters before and after expression of *CCDC144NL-AS1* was rescued. Scale bars, 100  $\mu$ m. **b** qPCR validation of *CCDC144NL-AS1*'s rescued expression in CCDC144NL-AS1-KD-H9 cells. CCDC144NL-AS1-KD-H9 cells infected with empty overexpression viruses were used as control cells. Error bars indicate SEM ( $n = 3$ ). \*\*\* $p < .001$ . **c** Western blot analyses of naïve pluripotency-related NANOG protein levels in CCDC144NL-AS1-rescued CCDC144NL-AS1-KD-H9 cells and CCDC144NL-AS1-KD-H9 cells infected with empty overexpression viruses were used as control cells. **d** Western blot analyses of total and phosphorylated protein levels of MAPK in CCDC144NL-AS1-KD-H9 and NC-H9 cells. **e** Heatmap of genes involved in MAPK/ERK signaling pathway in CCDC144NL-AS1-downregulated hPSCs and control cells. **f** Western blot analyses of DUSP3, DUSP6, DUSP5, and DUSP7 protein levels in CCDC144NL-AS1-KD-H9 and NC-H9 cells. **g** Western blot analyses of total and unphosphorylated (active)  $\beta$ -catenin protein levels in CCDC144NL-AS1-KD-H9 and NC-H9 cells. **h** Heatmap of genes involved in the Wnt signaling pathway in CCDC144NL-AS1-downregulated hPSCs and control cells. **i** Western blot analyses of total SMAD2/3 and phosphorylated SMAD2 protein levels in CCDC144NL-AS1-KD-H9 and NC-H9 cells. **j** Heatmap of genes involved in the TGF- $\beta$  signaling pathway in CCDC144NL-AS1-downregulated hPSCs and control cells. The expression levels of genes in heatmaps are determined by RNA-seq data.  $\beta$ -Tubulin was used as an endogenous control in all western blot analyses

induction methods, bFGF and TGF- $\beta$  were added to their culture medium [14, 17, 65, 74], possibly implying the requirements for growth factors to stabilize human naïve PSCs. We then explored the TGF- $\beta$  signaling pathway state in our cells. Western blot analysis revealed that phospho-SMAD2 protein levels were slightly decreased in CCDC144NL-AS1-KD-H9 cells compared with NC-H9 cells (Fig. 7i). Nevertheless, when we added a selective activin receptor inhibitor SB 431542, which could effectively block TGF- $\beta$  signaling without toxicity [75], to the culture medium at the final concentration of 10  $\mu$ M to examine the TGF- $\beta$  signaling pathway dependency of our cells, we found both CCDC144NL-AS1-KD-H9 cells and NC-H9 cells underwent differentiation in 7 days (Additional file 13: Figure S8b). These results indicated that CCDC144NL-AS1-KD-H9 cells could not tolerate the TGF- $\beta$  signaling pathway blockage caused by SB 431542. RNA-seq data demonstrated that TGF- $\beta$  signaling pathway ligands *GDF3* and *NODAL*; receptors *ACVR1*, *BMPRIA*, and *TGFBR2*; and signal transducers *SMAD2* and *SMAD3* expressed higher levels in CCDC144NL-AS1-KD hPSCs relative to control cells (Fig. 7j). At the same time, there were two receptors *ACVR2B* and *TGFBRI* being downregulated and three negative regulators *LEFTY1*, *LEFTY2*, and *BAMBI* being elevated in these cells, reflecting the simultaneous existence of a stimulation state and a suppressive feedback loop of TGF- $\beta$  signaling pathway in CCDC144NL-AS1-KD hPSCs than control cells (Fig. 7j). Consistent with the intricate state, TGF- $\beta$  signaling target genes *SPI1*, *RBL1*, and *TFDPI* expressed lower levels and another *EP300* expressed higher levels in CCDC144NL-AS1-KD hPSCs compared with control cells (Fig. 7j).

Another nonnegligible signaling pathway implicated in the naïve pluripotency sustaining is the LIF/STAT3 signaling pathway. Reinforcement of STAT3 activity in combination with 2i/LIF can revert human PSCs from the primed pluripotency to a naïve-like pluripotency state [16]. We also analyzed the expression levels of STAT3 target genes and found that eight target genes (*ICAM1*, *IERS3*, *ZFP36L1*,

*SULF1*, *SPRY2*, *SMAD7*, *PIM3*, *DACT1*) were upregulated in CCDC144NL-AS1-KD hESCs compared to control cells, implying the relatively high activity of STAT3 signaling (Additional file 13: Figure S8c), whereas a few STAT3 target genes were not significantly reduced in CCDC144NL-AS1-KD-HDF-iPSCs, and this might be caused by individual properties of this iPSC cell line (Additional file 13: Figure S8c).

It was reported that naïve human ESCs induced by small chemical molecules relied on mTORC2 subunit to maintain pluripotency [76]. So, we also analyzed the mTORC2 dependency of CCDC144NL-AS1-KD-H9 cells. We added rapamycin (mTORC1 inhibitor) and PP242 (mTORC1/2 inhibitor) to the culture medium of the cells separately. Both of CCDC144NL-AS1-KD-H9 and NC-H9 cells exhibited apoptosis in rapamycin and PP242 condition (Additional file 13: Figure S8d). After 48 h treatment, the cells were fixed and immunofluorescence stained with OCT4 and NANOG. The results showed that the expression of NANOG in CCDC144NL-AS1-KD-H9 was significantly reduced after rapamycin treatment, while it was not in NC-H9 cells (Additional file 13: Figure S8e, upper panel). On the other hand, PP242 condition made the CCDC144NL-AS1-KD-H9 cells lose OCT4 and NANOG expression simultaneously, while NC-H9 revealed obviously reduced expression in NANOG (Additional file 13: Figure S8e, lower panel). The above results suggested that, compared to NC-H9 cells, CCDC144NL-AS1-KD-H9 cells relied more on mTORC1/2, especially mTORC2, to maintain their pluripotency.

Collectively, our results demonstrated that downregulation of *CCDC144NL-AS1* endowed human PSCs with an analogous signaling pathway dependence pattern to the naïve pluripotent stem cells, suggestive of a significant role of *CCDC144NL-AS1* in human alternative pluripotency states conversion.

## Discussion

In this study, we identify a lncRNA, *CCDC144NL-AS1*, whose downregulation drives human PSCs to a pluripotency

state with characteristics resembling human naïve PSCs. *CCDC144NL-AS1*-downregulated hPSCs exhibit naïve-like dome-shaped clones, enhanced proliferation and single-cell cloning efficiency, increased developmental capacity, transcriptional profiles analogous to human naïve PSCs, genome-wide reduction of suppressive chromatin modification signals, and similar signaling pathway dependency patterns with human naïve pluripotency. Previous studies of naïve pluripotency derivation generally centered around transcription factors, small molecular inhibitors, and growth factors [14, 15, 17, 18, 65]. Only recently, two lncRNAs were reported to take part in naïve pluripotency preservation of mouse or human ESCs [77, 78]. *CCDC144NL-AS1* is the first identified lncRNA being involved in naïve-like pluripotency conversion from human primed PSCs without supplying additional transcription factors or small chemical molecules except for TeSR-E8 medium. It engenders our pondering over lncRNA roles in human naïve pluripotency regulation or early process of human embryonic development and provides new insights into the molecular switch between human primed and naïve pluripotency.

Over the last decade, lncRNAs have been functionally characterized in widespread biological processes implicating cell cycle regulation, immune response, pluripotency maintenance, lineage development, and multiple kinds of diseases, such as cancers and neurological and psychiatric disorders [29, 38, 79, 80]. Notably, thousands of known or novel lncRNAs were identified from RNA-seq data of mouse cleavage-stage embryos and human preimplantation embryos, and a large part of them exhibited developmental stage-specific expression patterns, unveiling their probable essential roles in mammalian early embryogenesis [81, 82]. From this point of view, it is not difficult to comprehend that only the modulation of a lncRNA's expression, for instance, *CCDC144NL-AS1* knocking down here, in primed culture condition can initiate the process of human primed-to-naïve pluripotency conversion. In addition, we think the important step used to narrow down candidate lncRNAs in our screening strategy is the calculation of the membership values between pluripotency-associated genes module (PGM) and human PSCs' enriched noncoding genes from WGCNA (Fig. 1c–d, Additional file 4: Figure S1e). This screening method enabled us to obtain the top 12 lncRNAs highly correlated with human pluripotency to conduct subsequent experiments (Fig. 1e). Hence, our study provides an efficient screening model to extract potentially useful genes from the vast RNA-seq data.

*CCDC144NL-AS1*-KD hPSCs display distinct transcriptional and epigenetic characteristics from the primed cells, but the exact underlying mechanisms are not completely understood. It is known that the usage of small molecular kinase inhibitors and growth factors in the generation of naïve pluripotency can lead to naïve-

specific transcriptome and epigenome [7, 14, 17, 18]. Thus, we investigated the signaling pathways in our cells. Our data demonstrate that in *CCDC144NL-AS1*-KD-H9 cells, the activity of MAPK/ERK signaling pathway is inhibited, active  $\beta$ -catenin is accumulated, and some LIF-STAT3 target genes are upregulated, all of which are concordant with that in human naïve PSCs (Fig. 7d, e, g, h; Additional file 13: Figure S8c). Because we did not add any small molecular inhibitors to the culture conditions except for two growth factors (bFGF and TGF $\beta$ ) included in TeSR E8 medium, the exact mechanisms underlying the naïve-like pluripotency conversion of human PSCs caused by *CCDC144NL-AS1* knockdown are probably different from the usual induction methods of human naïve PSCs, such as “3iL” or “5iL/A” protocols [14, 18]. In effect, a group of protein phosphatases is activated in *CCDC144NL-AS1*-KD hPSCs, such as some DUSP family members (Fig. 7e, f). We speculate that these protein phosphatases are stimulated in some way and play a crucial role in the suppression of protein kinases involved in signaling pathways. LncRNA *lincU* was reported to take part in naïve pluripotency preservation of mouse and human ESCs by binding and stabilizing DUSP9 protein and then constitutively inhibits the ERK1/2 signaling pathway [78]. Thus, *CCDC144NL-AS1* may promote the opposite processes, such as proteasome pathway or other protein degradation pathways; its knockdown then leads to the accumulation of DUSP protein and results in changes in signaling pathway activity state. On the other hand, lncRNA *lincenc1* was reported to maintain naïve states of mouse ESCs by promoting the glycolysis pathway [77]. GO classification and enrichment analyses also revealed that our *CCDC144NL-AS1*-KD-hPSCs were highly correlated with energy generation terms (Fig. 5c, upper panel). So, *CCDC144NL-AS1* may also play roles in this process. Additionally, we found totally four lncRNAs in the first lncRNA screening part. Our examination results of the expression changes of the other three lncRNAs after *CCDC144NL-AS1* knockdown in H9 showed that *LOC100506930-1* (NR\_038278) and *LOC100506930-2* (NR\_038279) were significantly increased (Additional file 13: Figure S8f). Hence, *LOC100506930-1* (NR\_038278) and *LOC100506930-2* (NR\_038279) may be involved in the pluripotency state conversion of hPSCs triggered by *CCDC144NL-AS1* downregulation. Generally speaking, our *CCDC144NL-AS1*-KD hPSCs provide a platform to further investigate what other kinds of intermediate regulators are involved in human early pluripotency regulation.

Of note, some expression of *CCDC144NL-AS1* is required in naïve-like type hPSCs. The expression of *CCDC144NL-AS1* in H9 cells was hundreds-fold higher than that in HEF (Fig. 1g) and was about 2.5-fold higher

than that in CCDC144NL-AS1-KD-H9 (Fig. 2f). The *CCDC144NL-AS1* mRNA levels in CCDC144NL-AS1-KD-hESCs was obviously higher than that in HEF. Actually, the expression data of *CCDC144NL-AS1* in naïve H9 and its control H9 cells from Takashima et al. [15], and in 55 samples we used in the lncRNA screening process which contained 21 hiPSC samples, 15 hESC samples, and 19 human somatic tissue samples, revealed that *CCDC144NL-AS1* was expressed significantly lower in naïve H9 than that in human PSCs, but obviously higher than that in human somatic tissues (Additional file 13: Figure S8 g).

Previous studies and our research all suggest that the acquisition of naïve pluripotency involves a quite intricate regulatory network, but a substantial part of it is still undiscovered. Our study unveils a lncRNA's unexpected role in the complex network of naïve-like pluripotency transition from human PSCs, supplies a new perspective to further understand the lncRNA's effects in the early process of human embryonic development, and provides a platform to decipher unknown aspects of human naïve pluripotency. We think that *CCDC144NL-AS1* can be potentially valuable for future research on deriving higher quality naïve state hPSCs and promoting their therapeutic applications.

## Conclusions

*CCDC144NL-AS1* knockdown induces naïve-like state conversion of hPSCs without supplying transcription factors or small molecular inhibitors. CCDC144NL-AS1-KD hPSCs exhibit dome-shaped clones, increased developmental capacity, transcriptional profiles analogous to human naïve PSCs, genome-wide reduction of suppressive chromatin modification signals, and similar signaling pathway dependency patterns with human naïve pluripotency. Our study supplies a new perspective to further understand the lncRNA's effects in the early process of human embryonic development and provides a platform to decipher unknown aspects of human naïve pluripotency.

## Additional files

**Additional file 1: Table S1.** Summary of primer sequences for quantitative real-time PCR in this study. (XLSX 12 kb)

**Additional file 2: Table S2.** Antibodies used in immunofluorescence staining and western blot analysis. (XLSX 11 kb)

**Additional file 3: Table S3.** RNA-seq datasets utilized in this paper for screening lncRNAs that were specifically highly expressed in human PSCs. (XLSX 10 kb)

**Additional file 4: Figure S1.** Functional categories of genes from three human PSC-specific modules enriched by WGCNA. a WGCNA dendrogram indicating expression of different gene modules in all 55 samples. b-d Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)

analyses of genes in MEred (b), MElightcyan (c), and MEpurple (d) modules. e List of pluripotency-associated genes in PGM module. (PDF 3714 kb)

**Additional file 5: Video S1.** Confocal laser scanning of CCDC144NL-AS1-KD-H9 cluster accompanied by 90° horizontal rotation. (MP4 1823 kb)

**Additional file 6: Video S2.** Confocal laser scanning of NC-H9 cluster accompanied by 90° horizontal rotation. (MP4 1745 kb)

**Additional file 7: Figure S2.** Pluripotency validation of *CCDC144NL-AS1* downregulated H14 and HDF-iPS cells. a-d Quantitative RT-PCR analyses of three core pluripotency and naïve pluripotency genes (a, c) and lineage commitment factor genes (b, d) in CCDC144NL-AS1-KD-H14, NC-H14, CCDC144NL-AS1-KD-HDF-iPS, and NC-HDF-iPS cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . e-f Immunostaining images of pluripotency-associated markers OCT4, NANOG, SOX2, SSEA3/4, and TRA-1-60 (e) and naïve pluripotency-specific marker TFCP2L1 (f) for CCDC144NL-AS1-KD-H14, NC-H14, CCDC144NL-AS1-KD-HDF-iPS, and NC-HDF-iPS cells. Scale bars, 100  $\mu$ m. g Western blot detection of naïve pluripotency-related transcription factors, such as NANOG and KLF4, and shared pluripotency transcription factors, such as OCT4 and SOX2, in CCDC144NL-AS1-KD-H14, NC-H14, CCDC144NL-AS1-KD-HDF-iPS, and NC-HDF-iPS cells.  $\beta$ -Tubulin was used as an endogenous control. (PNG 860 kb)

**Additional file 8: Figure S3.** Cell growth pattern analysis of *CCDC144NL-AS1*-downregulated H9, H14, and HDF-iPS cells. a Single-cell cloning efficiencies of CCDC144NL-AS1-KD-H14, NC-H14, CCDC144NL-AS1-KD-HDF-iPS, and NC-HDF-iPS cells. Error bars indicate SEM ( $n = 3$ ); \*\* $p < .01$ . b-c After plating cells for 4 days, the cell cycle of CCDC144NL-AS1-KD-H9 and NC-H9 was analyzed by flow cytometry. The representative flow cytometry analysis results (b) and the percentages of cells in G0/G1 and S+G2/M phases (c) are shown. d-g After plating cells for 4 days, the cell cycle of CCDC144NL-AS1-KD-H14, NC-H14, CCDC144NL-AS1-KD-HDF-iPS, and NC-HDF-iPS cells was analyzed by flow cytometry. The representative flow cytometry analysis results (d, f) and the percentages of cells in G0/G1 and S+G2/M phases (e, g) are shown. (PDF 660 kb)

**Additional file 9: Figure S4.** Developmental potential of *CCDC144NL-AS1*-downregulated H14 and HDF-iPS cells in vitro. a Morphologies of CCDC144NL-AS1-KD-H9- and NC-H9-derived neuronal, endodermal, and mesodermal cells. b, f, l Quantitative RT-PCR analyses of expression levels of lineage-specific genes in CCDC144NL-AS1-KD-HDF-iPS and NC-HDF-iPS-derived neuronal (b), endodermal (f), and mesodermal (l) cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . c, g, m Western blot analyses of lineage-specific genes' expression in CCDC144NL-AS1-KD-HDF-iPS- and NC-HDF-iPS-derived neuronal (c), endodermal (g), and mesodermal (m) cells.  $\beta$ -Tubulin was used as an endogenous control. d, h, j Quantitative RT-PCR analyses of expression levels of lineage-specific genes in CCDC144NL-AS1-KD-H14- and NC-H14-derived neuronal (d), endodermal (h), and mesodermal (j) cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . e, i, k Western blot analyses of lineage-specific genes' expression in CCDC144NL-AS1-KD-H14- and NC-H14-derived neuronal (e), endodermal (i), and mesodermal (k) cells.  $\beta$ -Tubulin was used as an endogenous control. (PNG 1548 kb)

**Additional file 10: Figure S5.** Embryoid body formation analysis of *CCDC144NL-AS1*-downregulated H14 and HDF-iPS. a Quantitative RT-PCR analyses of cell adhesion-associated genes in embryoid bodies generated from CCDC144NL-AS1-KD-H9 and NC-H9 cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . b-e Quantitative RT-PCR analyses of lineage-related genes in embryoid bodies generated from CCDC144NL-AS1-KD-H14 (b), NC-H14 (c), CCDC144NL-AS1-KD-HDF-iPS (d), and NC-HDF-iPS (e) cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . f-g Embryoid body numbers generated from per  $10^6$  CCDC144NL-AS1-KD-H14 and NC-H14 (f) and CCDC144NL-AS1-KD-HDF-iPS and NC-HDF-iPS (g) cells. Error bars indicate SEM ( $n = 3$ ). \* $p < .05$ ; \*\* $p < .01$ . (PDF 448 kb)

**Additional file 11: Figure S6.** a-b Representative immunostaining images of three embryonic germ layer markers for teratomas derived from CCDC144NL-AS1-KD-H9 (a) and NC-H9 (b) cells.  $\beta$  III TUBULIN stands for ectoderm, E-CADHERIN stands for endoderm, and VIMENTIN represents mesoderm. Scale bars, 100  $\mu$ m. c Hematoxylin and eosin staining of teratomas containing tissues of all three embryonic germ

layers derived from CCDC144NL-AS1-KD-H14, NC-H14, CCDC144NL-AS1-KD-HDF-iPS, and NC-HDF-iPS cells. f, l, o Gut-like epithelium (endoderm). g, j, m, p Cartilage (mesoderm). h, k, n, q Neural tissue (ectoderm). Scale bars, 50  $\mu$ m. d-e Teratoma formation rate of CCDC144NL-AS1-KD-H14 and NC-H14 cells (d) and CCDC144NL-AS1-KD-HDF-iPS and NC-HDF-iPS cells (e). Error bars indicate SEM ( $n = 6$ ). \* $p < .05$ . (PNG 3193 kb)

**Additional file 12: Figure S7.** Functional categories of genes from CCDC144NL-AS1-KD human PSC-specific MEblue module and upregulated enrichment of respiratory chain in CCDC144NL-AS1-KD-H9 by GSEA. a GO analyses of MEblue module genes. b The enriched upregulated respiratory chain by gene set enrichment analysis (GSEA) in CCDC144NL-AS1-KD-H9 cells relative to NC-H9 cells. Vertical black bars represent the position of genes involved in the respiratory chain. (PDF 439 kb)

**Additional file 13: Figure S8.** Self-renewal of CCDC144NL-AS1-KD hESCs is TGF- $\beta$  and mTORC1/2 dependent and exhibit relative activated LIF-STAT3 signaling pathway state. a The relative NANOG protein intensity of western blot analyses in CCDC144NL-AS1-rescued CCDC144NL-AS1-KD-H9 and their control cells.  $\beta$ -Tubulin was used as an endogenous control. Error bars indicate SEM ( $n = 3$ ). \*\* $p < .01$ . b Images of differentiated cell morphologies of CCDC144NL-AS1-KD-H9 and NC-H9 after adding SB 431542 into the culture medium for 7 days. Scale bars, 100  $\mu$ m. c Heatmap of genes implicated in LIF-STAT3 signaling pathway in CCDC144NL-AS1-KD hPSCs and control cells. The expression levels of genes are determined by RNA-seq data. d Images of NC-H9 and CCDC144NL-AS1-KD-H9 cell morphologies after 48 h treatment of rapamycin (6  $\mu$ M) and PP242 (2.5  $\mu$ M). Scale bars, 100  $\mu$ m. e Immunostaining images of pluripotency-associated markers OCT4 and NANOG for NC-H9 and CCDC144NL-AS1-KD-H9 cells after 48 h treatment of rapamycin (6  $\mu$ M) and PP242 (2.5  $\mu$ M). Scale bars, 100  $\mu$ m. f Quantitative RT-PCR analyses of *LOC100506930-1* (NR\_038278), *LOC100506930-2* (NR\_038279), and *DDX11-AS1* in CCDC144NL-AS1-KD-H9 and NC-H9 cells. \*\* $p < .01$ . g The expression data of *CCDC144NL-AS1* in naïve H9 and its control H9 cells from Takashima et al. [15], and in 55 samples, we used in the lncRNA screening process which contained 21 hiPSC samples, 15 hESC samples, and 19 human somatic tissue samples. Displayed are FPKM values. (PNG 1613 kb)

## Abbreviations

bFGF: Basic fibroblast growth factor; ChIP-Seq: Chromatin immunoprecipitation followed by deep sequencing; DUSP: Dual-specificity phosphatase; EBs: Embryoid bodies; EpiSCs: Post-implantation epiblast-derived stem cells; FGF2: Fibroblast growth factor 2; FGFR: Fibroblast growth factor receptor; FPKM: Fragments per kilobase of transcript per million fragments mapped; GO: Gene Ontology; GSEA: Gene set enrichment analysis; GSK3: Glycogen synthase kinase-3; HEF: Human embryonic fibroblast; hESCs: Human embryonic stem cells; hiPSCs: Human induced pluripotent stem cells; HNES: Human naïve embryonic stem; hPSCs: Human pluripotent stem cells; KD: Knockdown; KEGG: Kyoto Encyclopedia of Genes and Genomes; KLF4: Kruppel-like factor 4; LIF: Leukemia inhibitory factor; lncRNAs: Long noncoding RNAs; MAPK or ERK: Mitogen-activated protein kinase; MAPKK or MEK: Mitogen-activated protein kinase kinase; NOD/SCID: Non-obese diabetic severe combined immunodeficiency; NPCs: Neural precursor cells; OCT4: Octamer-binding protein 4; PGM: Pluripotency-associated gene module; PSCs: Pluripotent stem cells; qPCR: Quantitative real-time PCR; ROCK: Rho-associated protein kinase; RPKM: Reads per kilobase per million mapped reads; shRNAs: Short hairpin RNAs; TCF3: Transcription factor 3; TGF- $\beta$ : Transforming growth factor beta; TOM: Topological overlap measure; TSS: Transcription start site; WGCNA: Weighted gene co-expression network analysis; ZIC: Zinc finger of the cerebellum

## Acknowledgements

Not applicable.

## Authors' contributions

YL and SL were responsible for the conception and design and financial support. YW was responsible for the collection and assembly of the data, data analysis and interpretation, and manuscript writing. BG, HL, and YS were

responsible for the data analysis and interpretation. XZ, ZX, YL, XG, JY, ZS, and XL were responsible for the collection and assembly of the data. All authors read and approved the final manuscript.

## Funding

This project was supported by grants from the National Natural Science Foundation of China (31830111, 81771333, 31571405) and National Key R&D Program of China (2017YFE9126600).

## Availability of data and materials

All data generated or analyzed during this study are included in this published article. Data of RNA-seq and ChIP-seq in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE111929.

## Ethics approval and consent to participate

All animal experiments were performed in accordance with institutional and national guidelines and the Guide for the Care and Use of Laboratory Animals promulgated by NIH and were approved by the animal experimentation ethics committee of Tongji University.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Stem Cell Translational Research Center, Tongji Hospital, Tongji University School of Medicine, Shanghai 200065, China. <sup>2</sup>College of Life Sciences, Nanchang University, Nanchang 330031, China. <sup>3</sup>The Children's Hospital, School of Medicine, Zhejiang University, Hangzhou 310052, China. <sup>4</sup>Human Aging Research Institute and School of Life Science, Nanchang University, Nanchang 330031, China. <sup>5</sup>Collaborative Innovation Center for Brain Science, Tongji University, Shanghai 200092, China.

Received: 17 February 2019 Revised: 18 June 2019

Accepted: 2 July 2019 Published online: 29 July 2019

## References

- Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. *Science* (New York, NY). 1998;282:1145–7.
- 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;131:861–72.
- Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* (New York, NY). 2007;318:1917–20.
- Pera MF, Trounson AO. Human embryonic stem cells: prospects for development. *Development*. 2004;131:5515–25.
- Kimbrel EA, Lanza R. Pluripotent stem cells: the last 10 years. *Regen Med*. 2016;11:831–47.
- Brook FA, Gardner RL. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A*. 1997;94:5709–12.
- Ying QL, Wray J, Nichols J, Battle-Morera L, Doble B, Woodgett J, et al. The ground state of embryonic stem cell self-renewal. *Nature*. 2008;453:519–23.
- Nichols J, Smith A. Naïve and primed pluripotent states. *Cell Stem Cell*. 2009;4:487–92.
- Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. 2007;448:191–5.
- Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 2007;448:196–9.
- Davidson KC, Mason EA, Pera MF. The pluripotent state in mouse and human. *Development*. 2015;142:3090–9.
- Rossant J. Mouse and human blastocyst-derived stem cells: vive les differences. *Development*. 2015;142:9–12.

13. Hanna J, Cheng AW, Saha K, Kim J, Lengner CJ, Soldner F, et al. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A*. 2010;107:9222–7.
14. Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell*. 2014;15:471–87.
15. Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*. 2014;158:1254–69.
16. Chen H, Aksoy I, Gonnot F, Osteil P, Aubry M, Hamela C, et al. Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nat Commun*. 2015;6:7095.
17. Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature*. 2013;504:282–6.
18. Chan YS, Goke J, Ng JH, Lu X, Gonzales KA, Tan CP, et al. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell*. 2013;13:663–75.
19. Guo G, von Meyenn F, Santos F, Chen Y, Reik W, Bertone P, et al. Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Reports*. 2016;6:437–46.
20. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*. 2005;122:947–56.
21. Jaenisch R, Young R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*. 2008;132:567–82.
22. Chia NY, Chan YS, Feng B, Lu X, Orlov YL, Moreau D, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*. 2010;468:316–20.
23. Zhang X, Yalcin S, Lee DF, Yeh TY, Lee SM, Su J, et al. FOXO1 is an essential regulator of pluripotency in human embryonic stem cells. *Nat Cell Biol*. 2011;13:1092–9.
24. Wang Z, Oron E, Nelson B, Razis S, Ivanova N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell*. 2012;10:440–54.
25. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006;125:315–26.
26. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*. 2006;441:349–53.
27. Spivakov M, Fisher AG. Epigenetic signatures of stem-cell identity. *Nat Rev Genet*. 2007;8:263–71.
28. Adamo A, Sese B, Boue S, Castano J, Paramonov I, Barrero MJ, et al. LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat Cell Biol*. 2011;13:652–9.
29. Ghosal S, Das S, Chakrabarti J. Long noncoding RNAs: new players in the molecular mechanism for maintenance and differentiation of pluripotent stem cells. *Stem Cells Dev*. 2013;22:2240–53.
30. Rosa A, Ballarino M. Long noncoding RNA regulation of pluripotency. *Stem Cells Int*. 2016;2016:1797692.
31. Melton C, Judson RL, Blleloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*. 2010;463:621–6.
32. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. 2008;18:1433–45.
33. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011;477:295–300.
34. Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J*. 2012;31:522–33.
35. Lin N, Chang KY, Li Z, Gates K, Rana ZA, Dang J, et al. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell*. 2014;53:1005–19.
36. Carrinci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science (New York, NY)*. 2005;309:1559–63.
37. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447:799–816.
38. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–66.
39. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22:1775–89.
40. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43:621–9.
41. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142:409–19.
42. Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol*. 2013;14:1190–8.
43. Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G. Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron*. 2015;88:861–77.
44. Chakraborty D, Paszkowski-Rogacz M, Berger N, Ding L, Mircetic J, Fu J, et al. lncRNA Panct1 maintains mouse embryonic stem cell identity by regulating TOBF1 recruitment to Oct-sox sequences in early G1. *Cell Rep*. 2017;21:3012–21.
45. Khalil AM, Guttman M, Huarte M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009;106:11667–72.
46. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet*. 2010;42:1113–7.
47. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12:357–60.
48. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–5.
49. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
50. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
51. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
52. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
53. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–50.
54. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34:267–73.
55. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
56. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
57. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42:W187–91.
58. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178–92.
59. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382–3.
60. Yoshizawa-Sugata N, Ishii A, Taniyama C, Matsui E, Arai K, Masai H. A second human Dbf4/ASK-related protein, Drf1/ASKL1, is required for efficient progression of S and M phases. *J Biol Chem*. 2005;280:13062–70.
61. Tsuji T, Lau E, Chiang GG, Jiang W. The role of Dbf4/Drf1-dependent kinase Cdc7 in DNA-damage checkpoint control. *Mol Cell*. 2008;32:862–9.
62. Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, et al. Nanog is the gateway to the pluripotent ground state. *Cell*. 2009;138:722–37.

63. Duggal G, Warriar S, Ghimire S, Broekaert D, Van der Jeught M, Lierman S, et al. Alternative routes to induce naive pluripotency in human embryonic stem cells. *Stem cells*. 2015;33:2686–98.
64. Zhou W, Choi M, Margineantu D, Margaretha L, Hesson J, Cavanaugh C, et al. HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J*. 2012;31:2103–16.
65. Ware CB, Nelson AM, Mecham B, Hesson J, Zhou W, Jonlin EC, et al. Derivation of naive human embryonic stem cells. *Proc Natl Acad Sci U S A*. 2014;111:4484–9.
66. Yang Y, Zhang X, Yi L, Hou Z, Chen J, Kou X, et al. Naive induced pluripotent stem cells generated from beta-thalassemia fibroblasts allow efficient gene correction with CRISPR/Cas9. *Stem Cells Transl Med*. 2016;5:8–19.
67. Huang K, Maruyama T, Fan G. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell*. 2014;15:410–5.
68. Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*. 2014;14:838–53.
69. Yang SH, Kalkan T, Morissroe C, Marks H, Stunnenberg H, Smith A, et al. Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. *Cell Rep*. 2014;7:1968–81.
70. Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol*. 2016;17:155–69.
71. Kidger AM, Keyse SM. The regulation of oncogenic Ras/ERK signalling by dual-specificity mitogen activated protein kinase phosphatases (MKPs). *Semin Cell Dev Biol*. 2016;50:125–32.
72. Wray J, Kalkan T, Gomez-Lopez S, Eckardt D, Cook A, Kemler R, et al. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol*. 2011;13:838–45.
73. Yi F, Pereira L, Hoffman JA, Shy BR, Yuen CM, Liu DR, et al. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nat Cell Biol*. 2011;13:762–70.
74. Qin H, Hejna M, Liu Y, Percharde M, Wossidlo M, Blouin L, et al. YAP induces human naive pluripotency. *Cell Rep*. 2016;14:2301–12.
75. Granier C, Gurchenkov V, Perea-Gomez A, Camus A, Ott S, Papanayotou C, et al. Nodal cis-regulatory elements reveal epiblast and primitive endoderm heterogeneity in the peri-implantation mouse embryo. *Dev Biol*. 2011;349:350–62.
76. Warriar S, Van der Jeught M, Duggal G, Tilleman L, Sutherland E, Taelman J, et al. Direct comparison of distinct naive pluripotent states in human embryonic stem cells. *Nat Commun*. 2017;8:15055.
77. Sun Z, Zhu M, Lv P, Cheng L, Wang Q, Tian P, et al. The long noncoding RNA Lncenc1 maintains naive states of mouse ESCs by promoting the glycolysis pathway. *Stem Cell Reports*. 2018;11:741–55.
78. Jiapaer Z, Li G, Ye D, Bai M, Li J, Guo X, et al. LincU preserves naive pluripotency by restricting ERK activity in embryonic stem cells. *Stem Cell Reports*. 2018;11:395–409.
79. Lin C, Yang L. Long noncoding RNA in cancer: wiring signaling circuitry. *Trends Cell Biol*. 2018;28:287–301.
80. Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci*. 2012;13:528–41.
81. Zhang K, Huang K, Luo Y, Li S. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC Genomics*. 2014;15:845.
82. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol*. 2013;20:1131–9.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

