**Supplementary information**

# Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells

Kazuaki Tokunaga[1,2], Noriko Saitoh[1,2]*, Ilya G. Goldberg[3], Chiyomi Sakamoto[1], Yoko Yasuda[1], Yoshinori Yoshida[4], Shinya Yamanaka[4], and Mitsuyoshi Nakao[1,2]*

[1]Department of Medical Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, 2-2-1 Honjo, Chuo-ku, Kumamoto 860-0811, Japan,
[2]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Tokyo, Japan,
[3]Image Informatics and Computational Biology Unit, Laboratory of Genetics, National Institute on Aging, National Institutes of Health, 333 Cassell Drive, Suite 4000, Baltimore MD 21224, USA,
[4]Center for iPS Cell Research and Application, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan.

*Address correspondence to:

Mitsuyoshi Nakao M.D., Ph.D. (mnakao@gpo.kumamoto-u.ac.jp)
Noriko Saitoh Ph.D. (norikos@kumamoto-u.ac.jp)
Department of Medical Cell Biology,
Institute of Molecular Embryology and Genetics, Kumamoto University
2-2-1 Honjo, Chuo-ku, Kumamoto 860-0811, Japan.
Tel: +81-96-373-6800; Fax: +81-96-373-6804

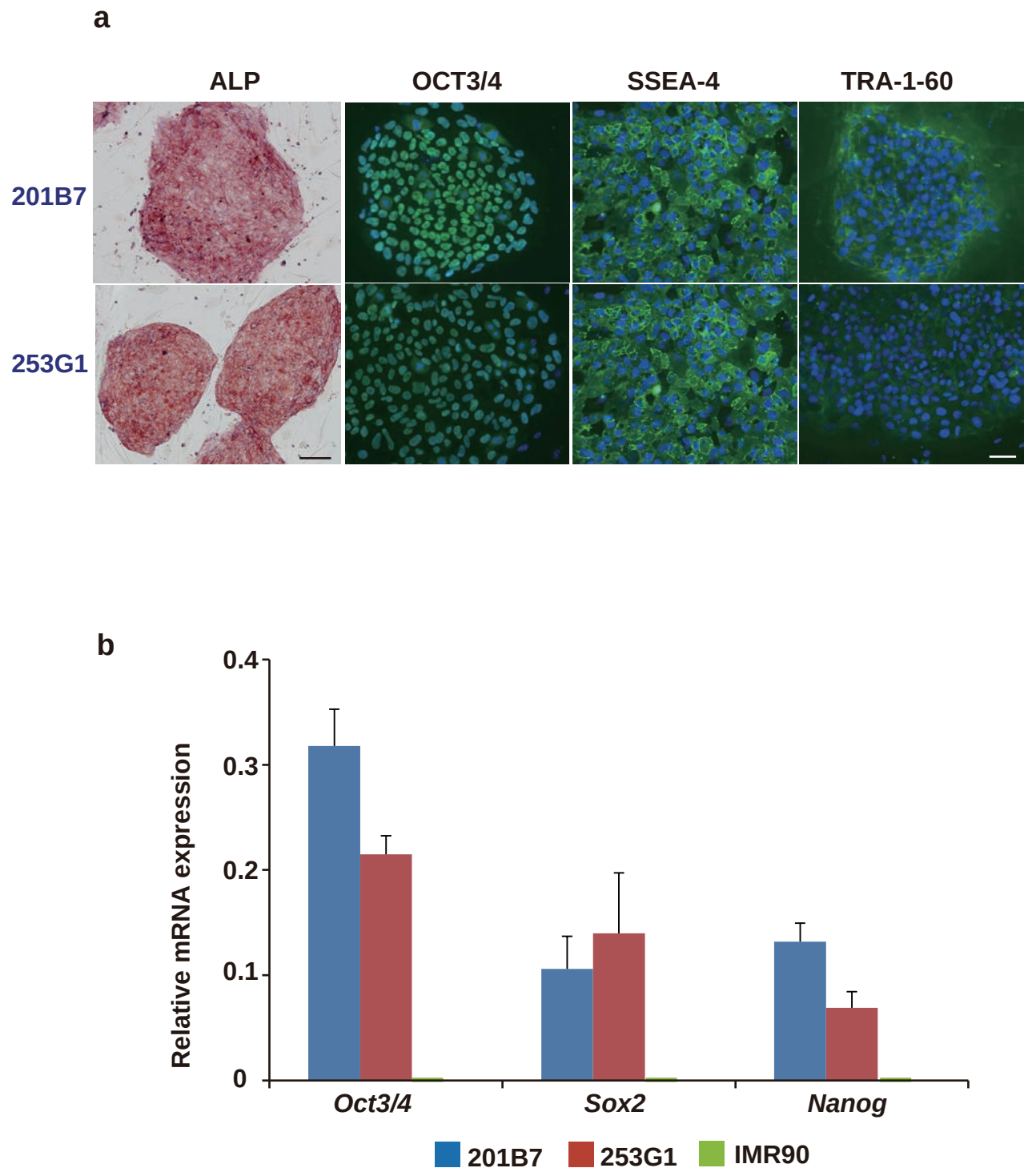


**Figure S1. Maintenance of human iPSC lines in an undifferentiated state**

**(a)** Immunostaining analyses of alkaline phosphatase (ALP, red), Oct3/4, SSEA-4, and TRA-1-60 (green) in established human iPSC lines (201B7 and 253G1). Nuclei were stained with DAPI (blue). Scale bars, 100 μm (left) and 20 μm (right). **(b)** Quantitative RT-PCR analyses of stemness genes. Expression levels were normalised to that of *β-actin*. Values are the means and s.d. of three independent experiments.

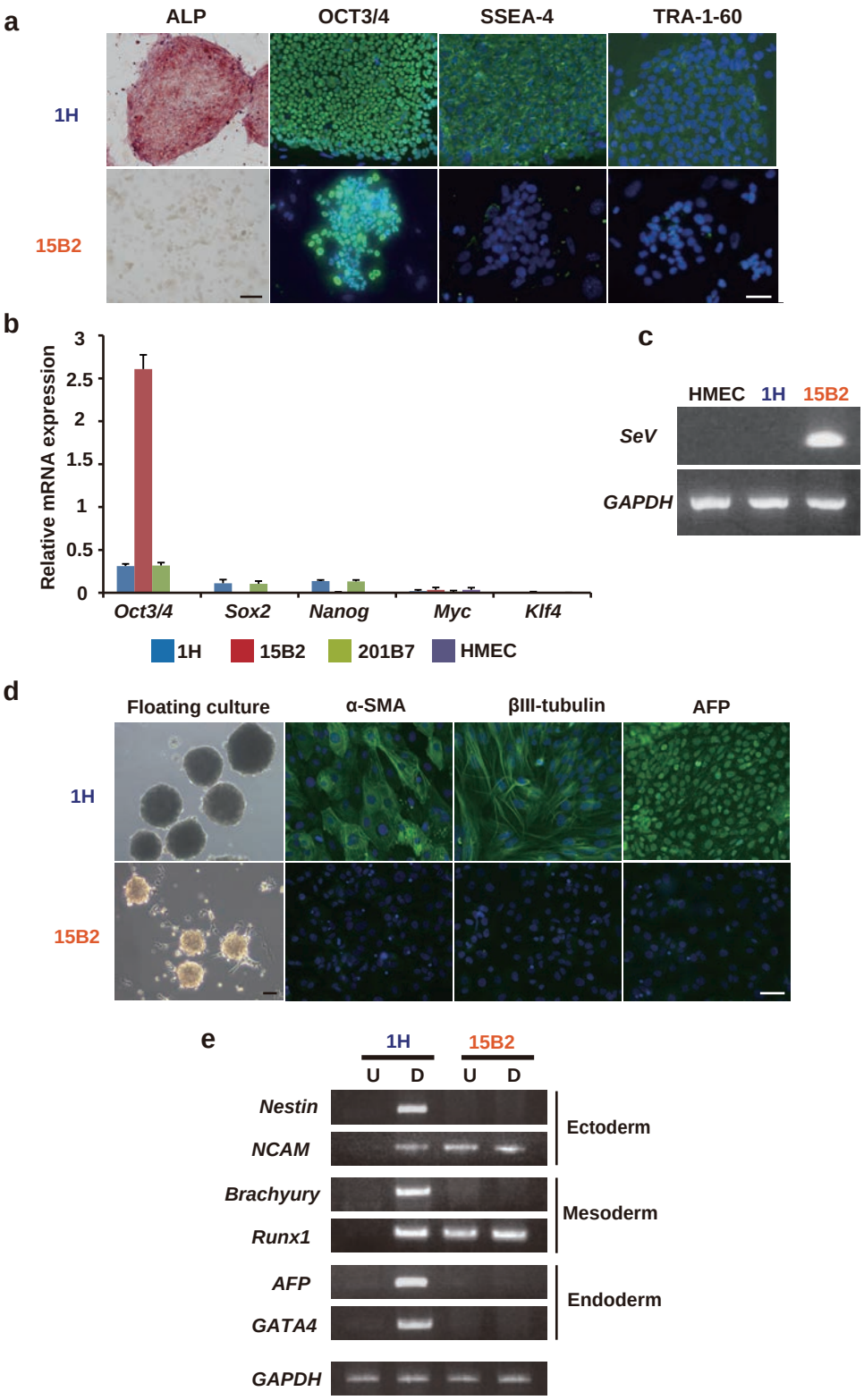


**Figure S2. Generation of human iPSCs and non-iPSCs**

**(a)** Generation of reprogrammed cells from HMECs or fibroblasts by Sendai viruses (SeV) expressing four factors. These cells were immunostained as described in Supplementary Fig. S1. In contrast to 1H cells, 15B2 cells (non-iPSCs) failed to express ALP, SSEA-4, and TRA-1-60. Scale bars, 100 μm (ALP) and 20 μm (right). **(b)** Quantitative RT-PCR analyses of reprogramming factor genes and Nanog. Expression levels were normalised to that of *β-actin*. Values are the means and s.d. of three independent experiments. **(c)** RT-PCR analysis of viral transcripts derived from SeV. 15B2 cells failed to evict the SeV during reprogramming. **(d)** Characterisation of embryoid body formation and differentiation markers in reprogrammed cells in vitro. Cells were immunostained for mesoderm smooth muscle actin (α-SMA), ectoderm β-III-tubulin, and endoderm α-fetoprotein (AFP). Scale bars, 50 μm. (e) RT-PCR analyses of differentiation markers for the three germ layers.

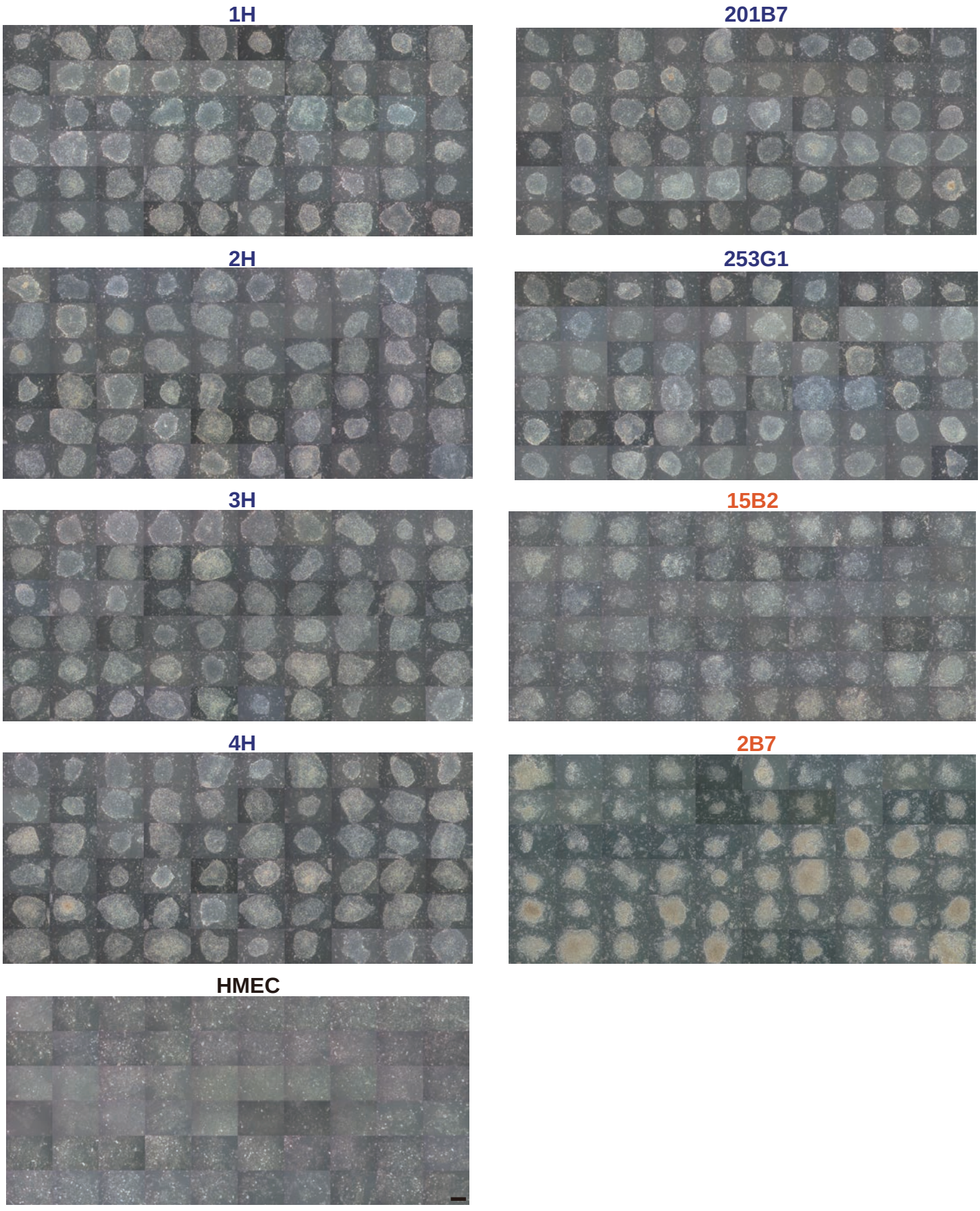


**Figure S3. Gallery of colony images used for the wndchrm analysis**

201B7 and 253G1 cells, which are standard bona fide iPSCs, were previously established from human fibroblasts with four factors (Oct3/4, Sox2, Klf4, and c-Myc) and three factors excluding c-Myc, respectively. 1H, 2H, 3H, and 4H cells are iPSCs reprogrammed from human mammary epithelial cells (HMECs) using the four factors. 15B2 and 2B7 cells are non-iPSCs reprogrammed from human fibroblasts using the four factors. Scale bar, 500 μm.

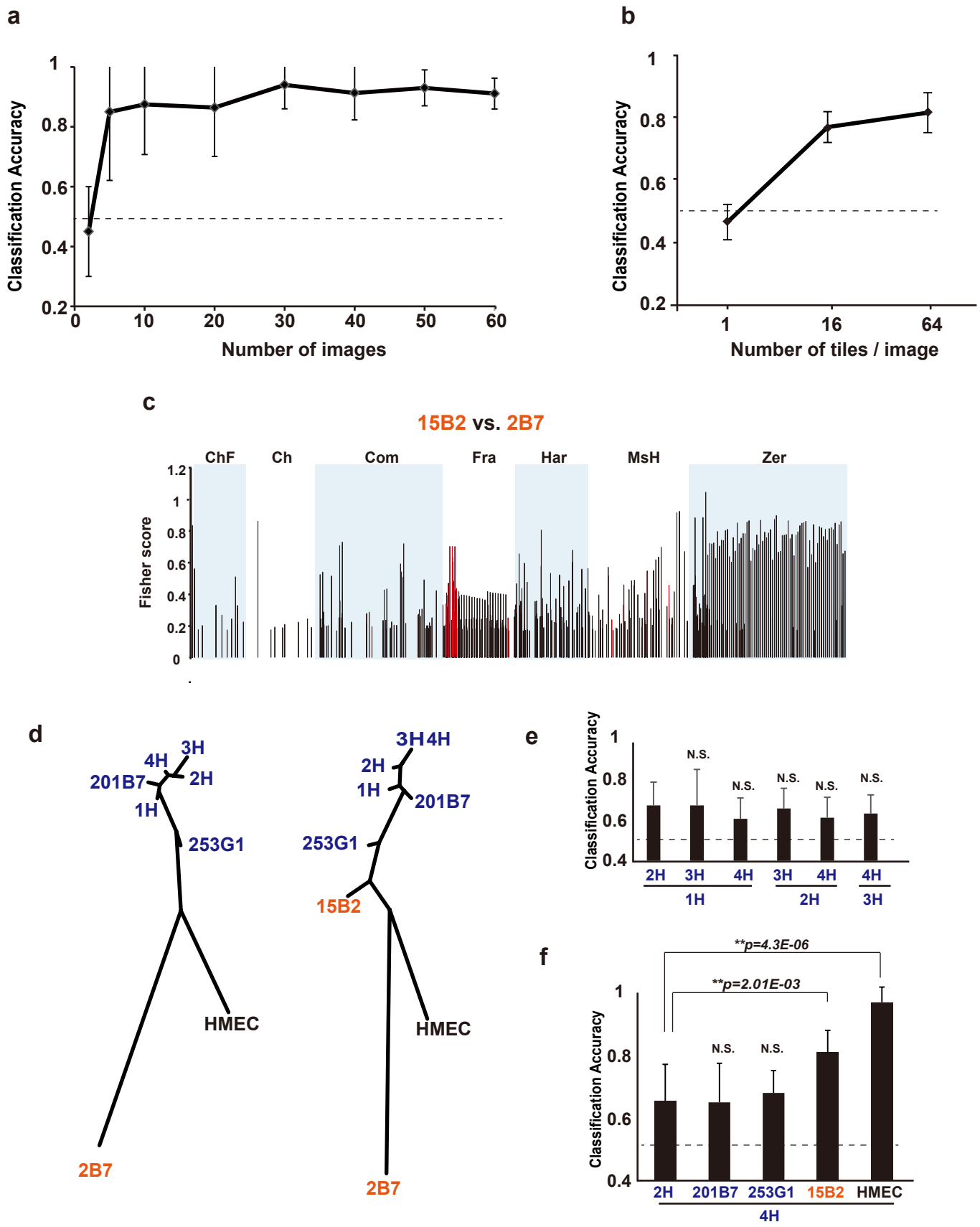


**Figure S4. Classification of bona fide iPSCs and non-iPSCs by the wndchrm program**

**(a** and **b)** Assessment of the optimum numbers of training images and tilings for classification of iPSC (1H cell) and non-iPSC (15B2 cell). Classification accuracy reached a plateau using more than 40 images (a) and 16 tilings (b). Values are the means and s.d. from 10 independent wndchrm tests. **(c)** Fisher discriminant scores assigned for 4008 image features in the classification of non-iPSCs (15B2 and 2B7 cells). See other comparisons (shown in Fig. 1c). **(d)** Phylogeny of colony morphologies. Bona fide iPSCs (1H–4H, 201B7, and 253G1 cells) are clustered, while non-iPSCs (2B7 and 15B2) are distant from the iPSC cluster. **(e** and **f)** Binary classifications of colony images among iPSCs (1H-4H) **(e)**, and against 4H **(f)**. Classification accuracy indicates the level of morphological differences between two cell types. The values are the means and standard deviation (s.d.) from 10 independent tests. N.S., not significant.

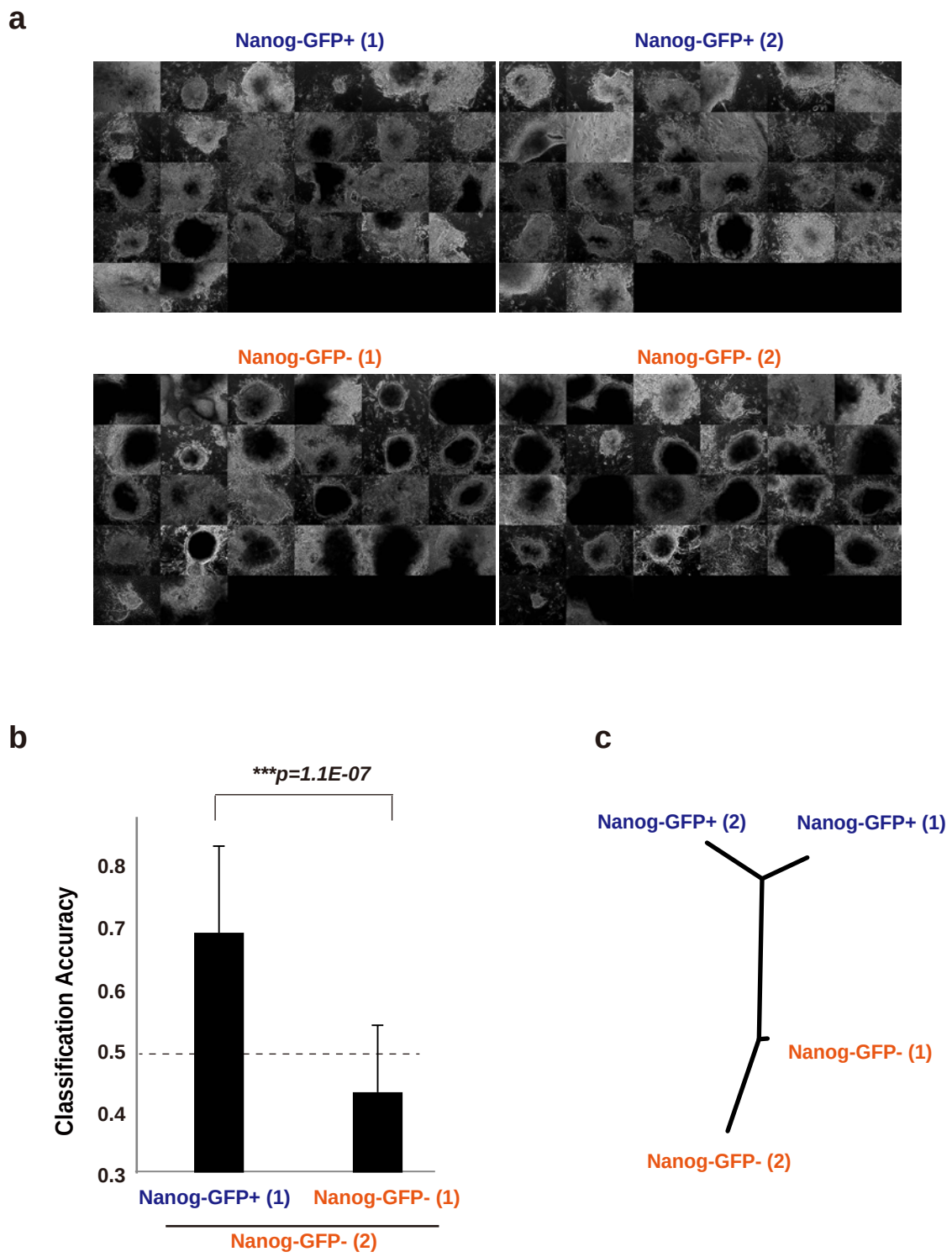


**Figure S5. Wndchrm classifications of heterogeneous populations of mouse iPSCs**

**(a)** Colony images of reprogrammed mouse embryonic fibroblasts (MEFs) carrying the Nanog-GFP-Ires-Puro cassette[23, 24]. Four factors (Oct3/4, Sox2, Klf4, and c-Myc) were introduced with retroviral vectors. In this system, fully reprogrammed cells give rise to green fluorescent protein (GFP)-positive (Nanog-GFP+). Phase contrast images of generated colonies corresponding to Nanog-GFP+ (iPSCs) or Nanog-GFP- (non-iPSCs) were collected and each was divided into two sub-groups. Overlapped colonies were avoided as much as possible. **(b)** Binary wndchrm classification showing a significant morphological dissimilarity between non-iPSCs and iPSCs (left column). **(c)** Phylogeny of mouse colony morphologies.

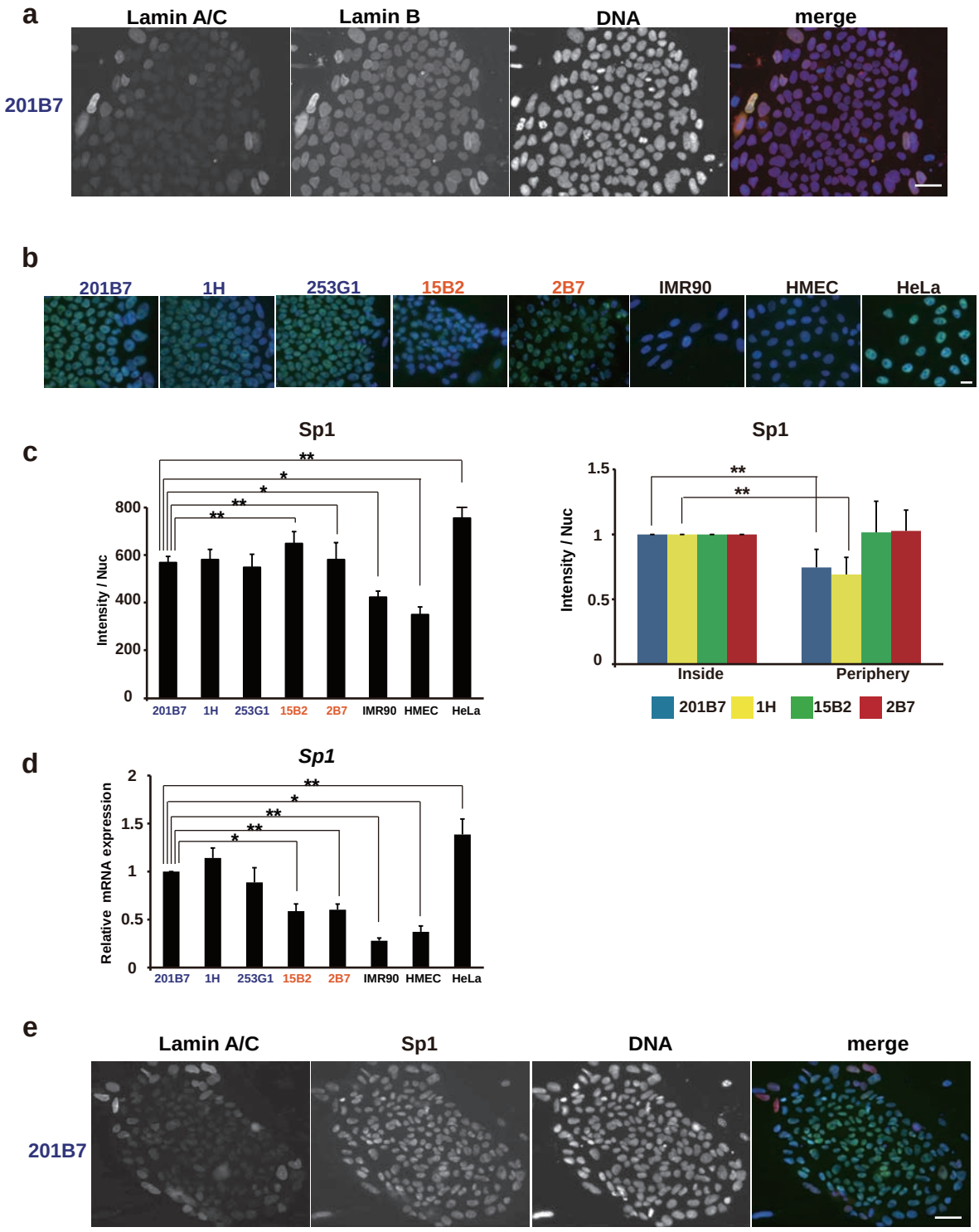


**Figure S6. Unique expression pattern of lamin A/C and Sp1 in a bona-fide iPSC colony**

**(a)** Immunofluorescence analyses of lamin A/C and lamin B in an iPSC colony. Lamin A/C (green) was expressed only in the peripheral cells of the colony, while lamin B (red) was evenly detected in the whole colony. Nuclei were stained with DAPI (blue). **(b)** Immunofluorescent analyses of Sp1 in various cell types. **(c)** Quantification of Sp1 signals in Supplementary Fig. S6b. Sp1 staining was intense in non-iPSCs (15B2 and 2B7 cells) and HeLa cells (left). In bona fide iPSCs (1H and 201B7 cells), Sp1 signals were higher inside of the colony (right). Values are the means and s.d. (n>200). *, p<0.05; **, p<0.01. **(d)** Quantitative RT-PCR analyses of *Sp1*. **(e)** Dual immunostaining of lamin A/C and Sp1 in iPSCs. Lamin A/C (red) and Sp1 (green) were strongly stained at the periphery and inside of the colony, respectively. Scale bars, 20 μm (a), 5 μm (b), and 20 μm (e).

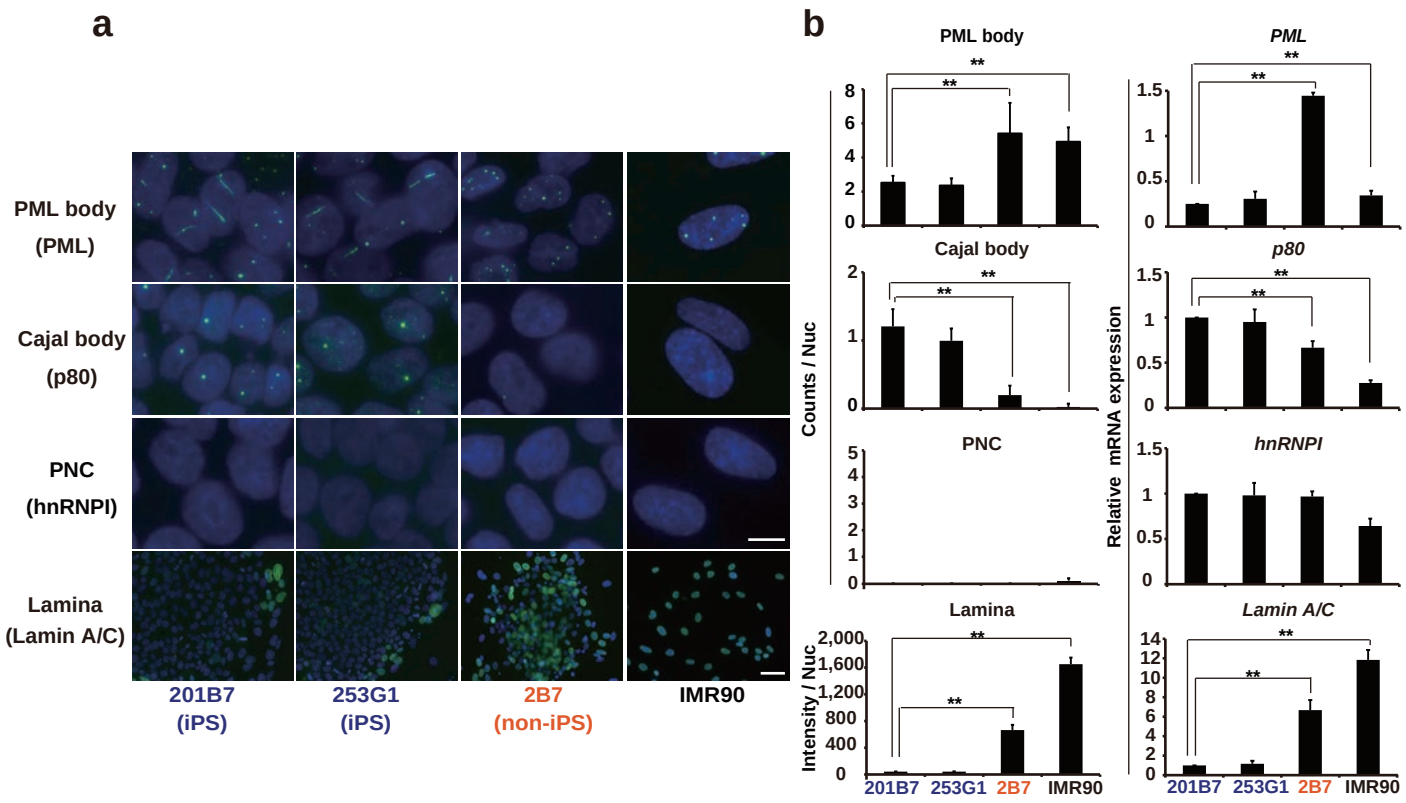


**Figure S7. Quantitative analyses of nuclear structures of iPSCs and non-iPSCs**

**(a)** Immunostaining of nuclear structures with antibodies against each marker protein in parentheses (green). Nuclei were stained with DAPI (blue). Scale bars, 5 μm (upper) and 20 μm (lower). **(b)** Quantification of nuclear structures in Supplementary Fig. S6a (n>300, left) and quantitative RT-PCR analyses of the corresponding nuclear markers (n=3, right). Values are the means and s.d. **, p<0.01.

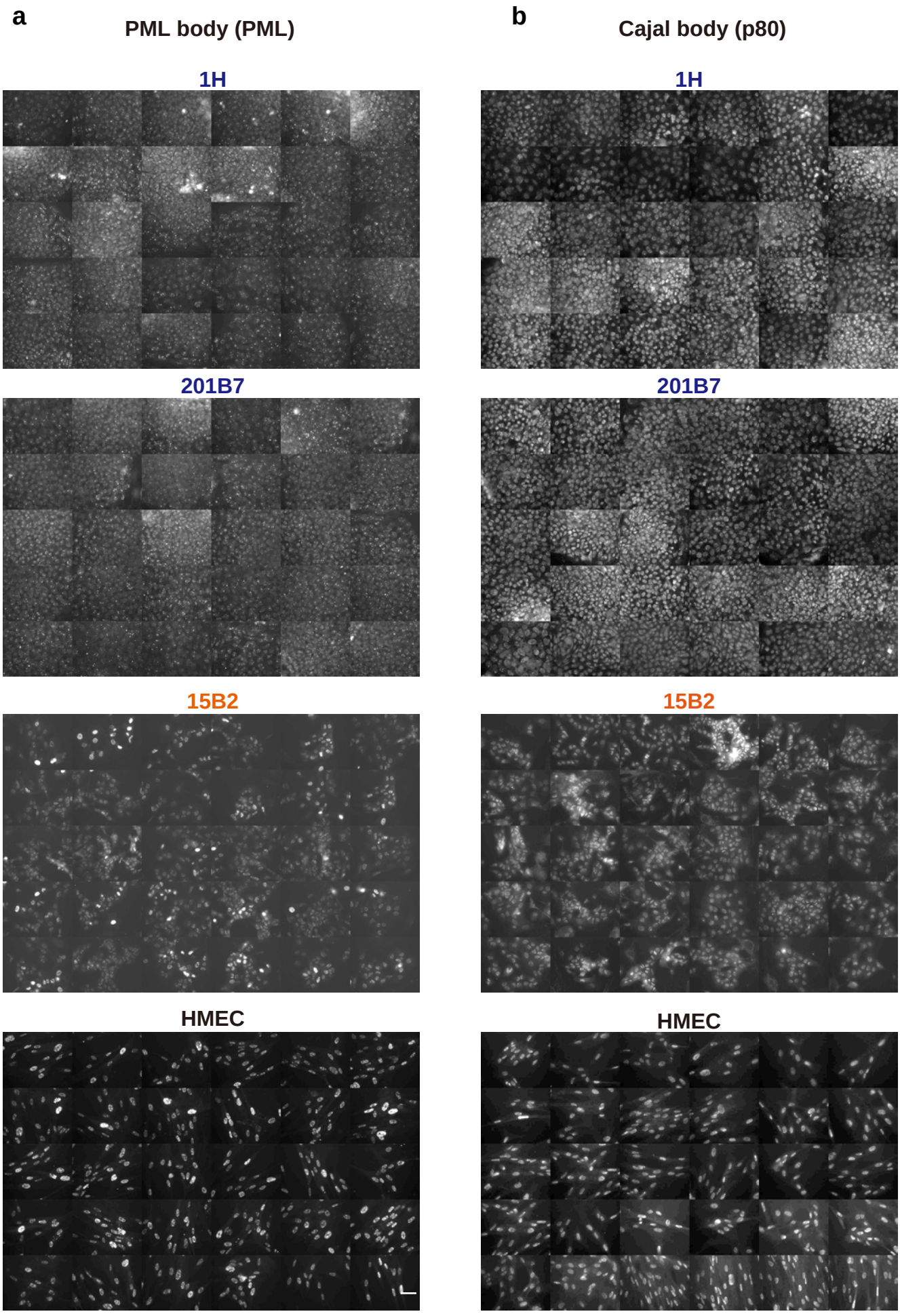


**Figure S8. Gallery of PML and Cajal body images used for the wndchrm classification in Fig. 2c**

Reprogrammed iPSCs (1H and 201B7 cells) and non-iPSCs (15B2 and HMEC cells) were immunostained to visualize PML (a) and Cajal (b) bodies. Scale bar, 20 μm.

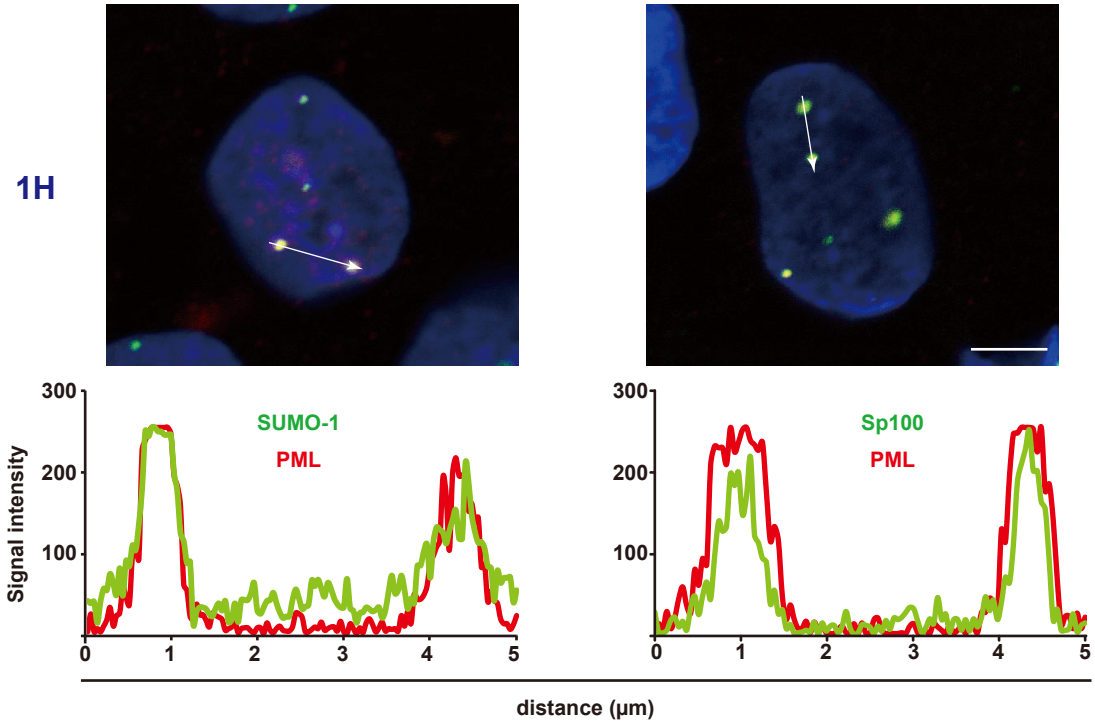


**Figure S9. Localisation of SUMO-1 and Sp100 to PML bodies in differentiated cells derived from iPSCs**

Immunofluorescence analyses were performed using iPSCs (1H cells) after differentiation (day 10). Line-scan histograms are shown below. Scale bar, 5 µm.