



OPEN

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

SUBJECT AREAS:
INDUCED PLURIPOTENT
STEM CELLS
IMAGE PROCESSING
NUCLEAR ORGANIZATION

Kazuaki Tokunaga^{1,2}, Noriko Saitoh^{1,2}, Ilya G. Goldberg³, Chiyomi Sakamoto¹, Yoko Yasuda¹, Yoshinori Yoshida⁴, Shinya Yamanaka^{4,5} & Mitsuyoshi Nakao^{1,2}

Received
19 June 2014

Accepted
22 October 2014

Published
11 November 2014

Correspondence and requests for materials should be addressed to N.S. (norikos@kumamoto-u.ac.jp) or M.N. (mnakao@gpo.kumamoto-u.ac.jp)

¹Department of Medical Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, 2-2-1 Honjo, Chuo-ku, Kumamoto 860-0811, Japan, ²Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Tokyo, Japan, ³Image Informatics and Computational Biology Unit, Laboratory of Genetics, National Institute on Aging, National Institutes of Health, 251 Bayview Boulevard, Suite 100, Baltimore, MD 21224, USA, ⁴Center for iPSC Cell Research and Application, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto, 606-8507, Japan, ⁵Gladstone Institute of Cardiovascular Disease, 1650 Owens Street, San Francisco, CA 94158, USA.

Non-invasive evaluation of cell reprogramming by advanced image analysis is required to maintain the quality of cells intended for regenerative medicine. Here, we constructed living and unlabelled colony image libraries of various human induced pluripotent stem cell (iPSC) lines for supervised machine learning pattern recognition to accurately distinguish bona fide iPSCs from improperly reprogrammed cells. Furthermore, we found that image features for efficient discrimination reside in cellular components. In fact, extensive analysis of nuclear morphologies revealed dynamic and characteristic signatures, including the linear form of the promyelocytic leukaemia (PML)-defined structure in iPSCs, which was reversed to a regular sphere upon differentiation. Our data revealed that iPSCs have a markedly different overall nuclear architecture that may contribute to highly accurate discrimination based on the cell reprogramming status.

The generation of human induced pluripotent stem cells (iPSCs) is simple and highly reproducible¹. However, only a small proportion of cells become pluripotent after introduction of the reprogramming factors, possibly resulting in a mixture of bona fide iPSCs and partially reprogrammed cells^{2,3}. It is essential to develop reliable methods to select completely reprogrammed iPSCs by eliminating the contamination of non-iPSCs⁴. Previous studies have shown changes in gene expression, DNA methylation, and histone modifications during iPSC reprogramming^{1,5}. Furthermore, reporter genes have been integrated into the genomic loci of pluripotency genes to visualize bona fide iPSCs⁴. However, there are no non-invasive methods that reliably identify live human iPSCs in large and heterogeneous populations of reprogramming cells.

Recent advances in automated biological image analyses enable objective measurements of cellular morphologies⁶. A supervised machine learning algorithm, *wndchrm* (weighted neighbour distances using a compound hierarchy of algorithms representing morphology), has been developed for automated image classification and mining of image similarities or differences⁷. It is a flexible, multi-purpose image classifier that can be applied to a wide range of bio-image problems. Unlike conventional image analysis, where users are required to specify target morphologies, choose specific algorithms, and try different parameters depending on the imaging problem, *wndchrm* users define classes by providing example images for each class; completely reprogrammed cells or partially reprogrammed cells, for example. Once classes are defined, classifications and similarity measurements are performed automatically. As the first step of the classification, *wndchrm* computes a large set of image features for each image in the defined classes and then selects image features that are informative for discrimination of the groups and constructs a classifier in an automated fashion^{6,7}. Next, the dataset is tested by multiple rounds of cross validation to measure classification accuracy (CA) as well as class similarity, which can be visualized with phylogenetic tree. The *wndchrm* algorithm has been successfully used for early detection of osteoarthritis⁸, measurement of muscle decline with aging, sarcopenia⁹, classification of malignant lymphoma¹⁰, and many other applications¹⁰.

Nuclear structure and function are closely linked to cellular reprogramming and epigenomic regulation⁵. During cell differentiation, nuclear structures are reconfigured dynamically. Previous studies have identified numerous distinct nuclear bodies^{11–13}. For example, promyelocytic leukaemia (PML) nuclear bodies typically exist as small spheres of 0.3–1.0 μm in diameter, and are implicated in various cellular pathways including



chromatin organisation, viral response, DNA replication, repair, and transcriptional regulation^{11,13}. Cajal Bodies are prominent in highly metabolically active cells such as neurons and cancers, and are implicated in the assembly or modification of transcriptional and splicing machinery¹⁴. The perinucleolar compartment (PNC) accumulates polypyrimidine tract binding protein¹⁵ and several polymerase III RNAs, which appears in virtually all types of solid tumours¹⁶. These bodies have been studied intensively in somatic cells^{11–13}, but much less is known about them in human iPSCs¹⁷.

Here, we established an accurate classification method to identify iPSCs using images of unlabelled live iPSC colonies. A combination of *wndchrn* and specific morphology quantification suggested that signals contributing to morphological discrepancies reside in nuclear sub-domains.

Results

Colony morphologies reflect proper reprogramming, which can be measured by pattern recognition. To build image classifiers to differentiate variously reprogrammed human cells, we first collected phase contrast images of live colonies formed by standard iPSC lines (201B7 and 253G1)^{2,3}, newly generated iPSC lines (1H–4H), non-iPSC lines (15B2 and 2B7), and somatic cells (human mammary epithelial cells, HMECs) (Fig. 1a). 253G1 and 201B7 cells were the initially established iPSC lines that were generated from human fibroblasts by introduction of four factors (Oct3/4, Sox2, Klf4, and c-Myc) and three factors (Oct3/4, Sox2, and Klf4), respectively^{2,3}. New iPSCs and non-iPSCs were derived from HMECs and human fibroblasts by a Sendai virus (SeV) carrying the four factors¹⁸. We confirmed that these iPSCs maintained pluripotency and could differentiate into three lineages *in vitro* (Supplementary Figs. S1 and S2)^{2,3}. In contrast, 15B2 and 2B7 cells lacked pluripotency, probably because of failure to silence the transgenes and activate endogenous stemness genes¹. The resultant image libraries included 60 colony images (1024 × 767 pixels) for each of the nine cell lines (Supplementary Fig. S3). In *wndchrn*, pattern recognition is based on ability to distinguish different classes, not pre-defined objects. Therefore, other than manually centering colonies, we used the entire colony image with no prior segmentation, as input.

The image classifier must be trained with a sufficient number of images. To optimize the classification capacities, we measured CA using different number of training data set of iPSC (1H) and non-iPSC (15B2), and found that the accuracy reached a plateau with more than 40 images (Supplementary Fig. S4a).

In addition, dividing a large image into multiple equal-sized tiles can sometimes provide better classification, particularly when numerous cells are distributed throughout an image. Treating the tiled images independently is expected to improve classification ability as the size of dataset increases^{19,20}. Therefore, we measured CA with and without tiling and found that the accuracy was improved by breaking an image into more than 16 images (Supplementary Fig. S4b).

Under these optimized classification condition, we classified several cell lines against iPSCs (1H cells) and compared CAs (Fig. 1b). In this binary classification, CA reflects the degree of morphological dissimilarities⁷. If the morphologies of two cell types are very distinct, the classifier is expected to show higher rate of accurate cross validations, at the maximum CA of 1.0. On the other hand, the CA value of random classification between two cell types with no feature differences is expected to be 0.5. The results showed that the CA against iPSCs (1H cells) was 0.66 for 2H cells, 0.68 for 201B7 cells, and 0.63 for 253G1 cells, but it was significantly high for non-iPSCs, 15B2 cells (0.87), and HMECs (0.96) (Fig. 1b). Therefore, *wndchrn* analysis was effective for discrimination of iPSC and non-iPSC colonies.

A set of informative image features extracted from each *wndchrn* test is summarized in Figure 1c and Supplementary Table S1. The Fisher discriminant values were clearly small between iPSC lines (1H

vs. 2H, and 1H vs. 201B7), while they were remarkably large for non-iPSCs (15B2 cells) and HMECs that exhibited a common feature pattern (Fig. 1c). In addition, most of the image features that contributed to the accurate classifications were based on transformed images (Fig. 1c, black bars)^{6,7}. Thus, *wndchrn* analyses are effective and objective for discrimination of iPSC and non-iPSC colonies.

We further examined the morphological similarities among the cell lines. The phylogeny in Fig. 1d was generated based on the pairwise class similarity (Supplementary Table S2), and showed that various iPSC lines, particularly those reprogrammed with the four factors, were closely clustered, whereas non-iPSCs (15B2) and HMECs were distantly positioned from them (Fig. 1d, and Supplementary Figs. S4c and S4d). A set of most informative image features extracted for this *wndchrn* test is listed in Supplementary Table S4.

Consistently, classifications between any combination of iPSC lines (1H–4H) resulted in low CA, which suggests that their morphologies are similar to each other (Supplementary Fig. S4e). Furthermore, binary classifications using another iPSC line (4H) as a reference (Supplementary Fig. S4f) resulted in a similar pattern to Fig. 1b.

Besides the above-mentioned studies on the cell lines, *wndchrn* analysis was effective to classify partially reprogrammed and fully reprogrammed mouse cells grown in the same dish as a mixed population (Supplementary Fig. S5).

Reprogrammed cells grow as large colonies. We investigated the nature of image features that discriminate iPSCs and non-iPSCs. As mentioned above, we classified iPSCs (1H) and non-iPSCs (15B2) with and without tiling the colony images (Supplementary Fig. S4b). By doing so, an image is broken into equally sized rectangles that are treated independently for successive training and test. An important feature as a single entity is lost, while the one distributed throughout the image is maintained, and size of the data set increases. We found that the CA was improved by tiling (Supplementary Fig. S4b), suggesting that the signals to identify iPSCs (1H) and non-iPSCs (15B2) were scattered in the image, and the colony morphology *per se* is not critical.

We further localised image features that discriminate iPSCs (1H) and non-iPSC (15B2) by tiling the colony images into 64 tiles and measuring the CA in each of them (Fig. 1e). As expected, a large part of the predictive signal came from areas containing iPSCs. Interestingly, the higher signals (CA ≥ 0.75) were positioned inside of the colony region (Inside), rather than the periphery with the local edge shape (Periphery) or outside of the colony (MEF), suggesting that unique features of the internal structure of the colony contribute most to the distinction.

Because nuclear morphology changes during differentiation status^{21–23}, we searched nuclear sub-structures that are different in iPSCs and non-iPSCs. Among the nuclear structures tested in this study, lamin A/C, the major component of the nuclear lamina²⁴, was expressed in the peripheral cells of iPSC colonies, while it was detected in most of the cells in non-iPSC colonies (Fig. 1f). In addition, transcription factor Sp1 (specificity protein 1)²⁵ was highly expressed inside of iPSC colonies (Supplementary Fig. S6). Such a metastatic state of iPSCs in the colony may be recognized by *wndchrn*.

The linear form of the PML-defined structure is characteristic of appropriately reprogrammed iPSCs. Extensive immunofluorescence analyses of ~20 distinct nuclear structures revealed that the PML body^{26,27}, Cajal body²⁷, and PNC²⁷ were characteristic of bona fide iPSCs, non-iPSCs, and cancerous HeLa cells, respectively (Figs. 2a–b, and Supplementary Fig. S7). The frequency of PML body formation in iPSC lines (1H, 201B7, and 253G1) was less than that in non-iPSC lines (15B2 and 2B7), and somatic cell lines (HMECs and IMR90 fibroblasts). Cajal body formation in non-iPSCs (0 ± 0.2 per

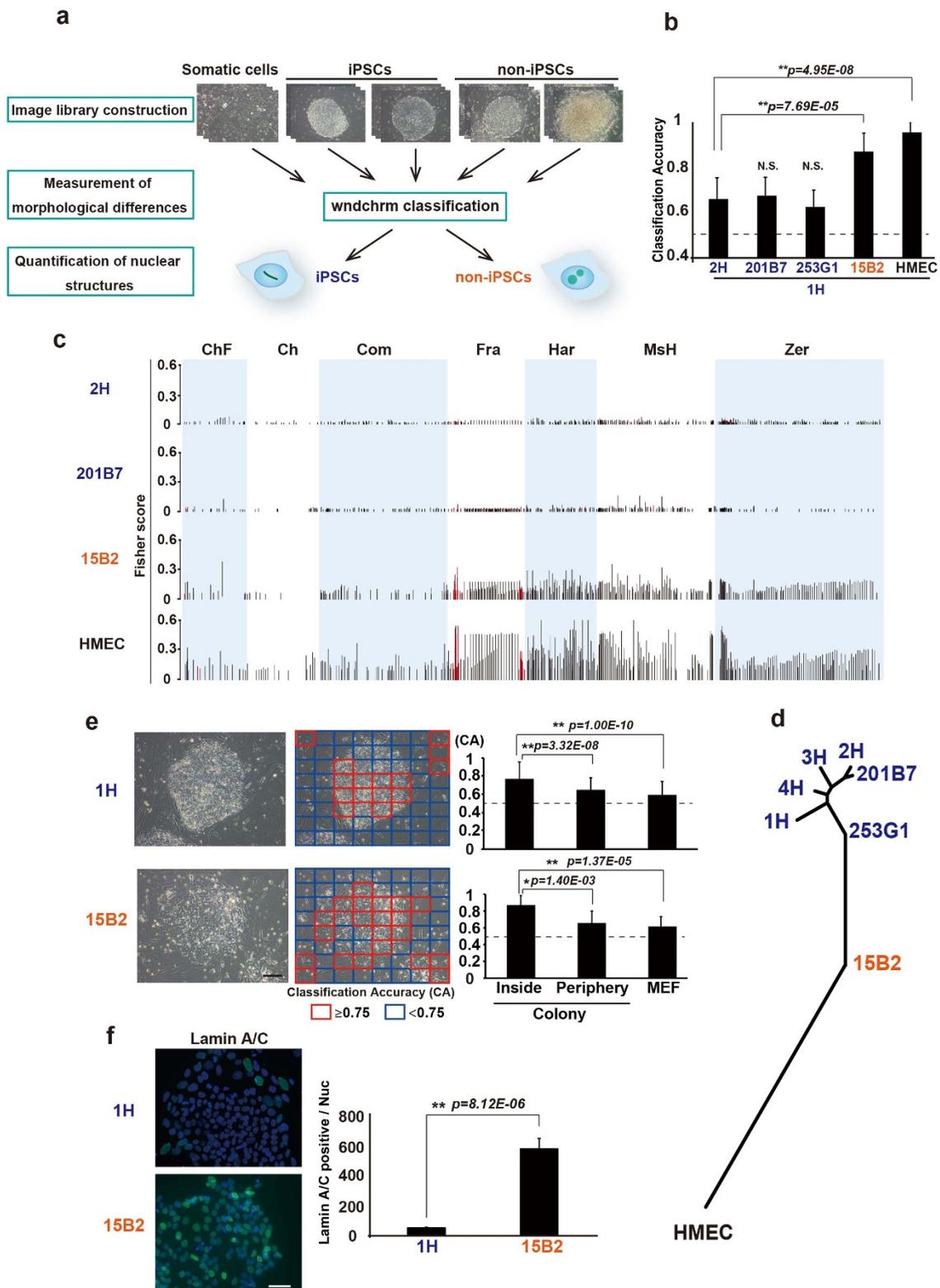


Figure 1 | Quantitative classification of completely and incompletely reprogrammed human iPSC colonies. (a) Experimental Overview. iPSCs and non-iPSCs are indicated as blue and red, respectively. (b) Binary classification of colony images against iPSCs (1H). Classification accuracy (CA) indicates the level of morphological differences between two cell types. CA value of two cell types with no feature differences is expected to be 0.5 (dotted line). The values are the means and standard deviation (s.d.) from 10 independent tests. N.S., not significant. (c) Fisher discriminant scores assigned to the 2773 features for each test in Fig. 1b. The values were calculated from raw (red bars) and transformed images (black bars). The name of each feature group is described in Supplementary Table S1. (d) Phylogeny based on morphological similarities. (e) Specification of the areas that distinguish iPSC (1H) and non-iPSC (15B2) colonies. CA values for each sub-image are shown as high (red) and low (blue). Average CA values inside, at the periphery, and outside of the colony (MEF, mouse embryonic fibroblast) are shown on the graph. (f) Selective expression of lamin A/C in the periphery of the iPSC colony. Immunofluorescence images of lamin A/C (green) and DAPI (blue), and quantified intensities are shown at the right ($n > 600$). Values are the means and s.d. *, $p < 0.05$; **, $p < 0.01$. Scale bars, 200 μm .

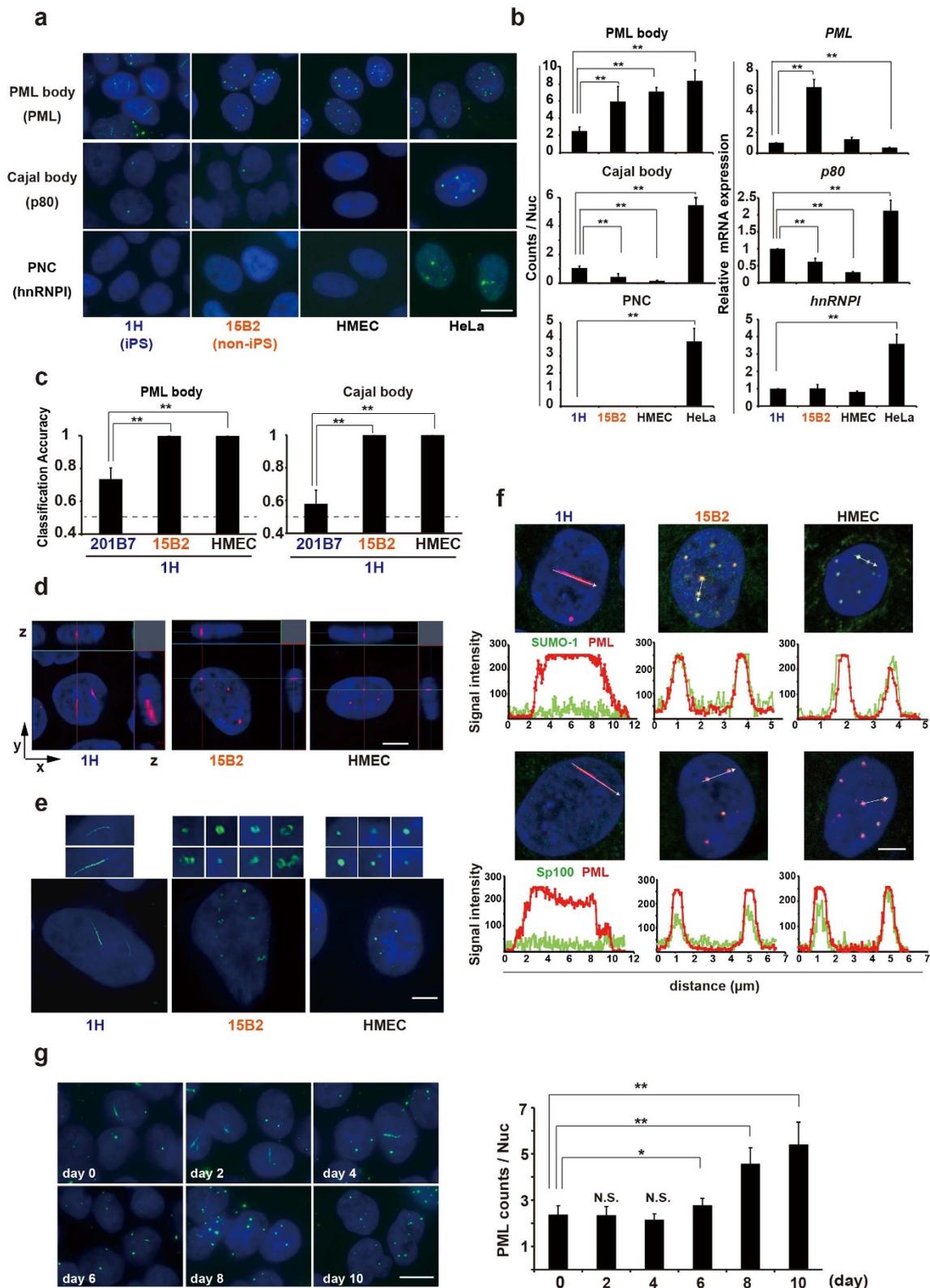


Figure 2 | Quantitative assessment of nuclear structures in completely and incompletely reprogrammed human iPSCs. (a) Identification of nuclear structure characteristics of iPSCs. Immunostaining was performed to identify the PML body (PML), Cajal body (p80 coilin), and perinucleolar compartment (PNC) (hnRNPI) (green). Nuclei were stained with DAPI (blue). (b) Quantification of nuclear structure formation ($n > 200$, left) and the mRNA levels of the corresponding components in the structures ($n = 3$, right). (c) *wndchr*m classifications against iPSCs (1H) using immunofluorescence images of PML and Cajal bodies ($n = 10$). (d) Detection of linear PML structures by three-dimensional confocal microscopy. (e) Detection of PML structural variation by structured illumination microscopy (100 nm resolution). Enlarged images of PML structures are shown in the upper boxes. (f) Lack of SUMO-1 and Sp100 in the linear PML structures of bona fide iPSCs. The signal intensity along the arrow is shown below. PML, red; SUMO-1 and Sp100, green. (g) Transition of PML structures from linear to round during differentiation. The number of PML structures is shown at the right ($n > 300$). Values are the means and s.d. *, $p < 0.05$; **, $p < 0.01$. Scale bars, 5 μm .



nucleus in 15B2 and 2B7 cells) and somatic cells (HMECs and IMR90 fibroblasts) was less frequent than that in iPSCs (1 ± 0.15 per nucleus in 1H, 201B7, and 253G1 cells), implying that hypoplasia of Cajal bodies is a feature of non-iPSCs. The PNC was only observed in HeLa cells (mean number = 4 ± 0.77), which is consistent with its specific appearance in cancerous cells¹⁶ and high expression of one of its components, hnRNPI, in HeLa cells (Fig. 2a–b).

Using the wndchrn image libraries constructed from immunofluorescence of the PML and Cajal bodies (Supplementary Fig. S8), bona fide iPSCs (1H cells) were discriminated from non-iPSCs (15B2 cells) by extremely high CA values (~ 1.0) (Fig. 2c).

PML-defined structures in iPSCs were especially striking. Linear PML structures were found uniquely in bona fide iPSCs (Fig. 2a and Supplementary Fig. S7a). Three-dimensional imaging by confocal microscopy revealed approximately straight, rod-like PML structures traversing within the nuclei of iPSCs (Fig. 2d). In addition, more detailed structures visualized by structured illumination microscopy showed at least three classes of PML structures: linear and connected bead-like in iPSCs (1H), irregular ring-like in non-iPSCs (15B2), and normal spheres in HMECs (Fig. 2e). In terms of protein composition, the linear PML structure in bona fide iPSCs was distinct from that in somatic cells (Fig. 2f). In somatic cells such as HMECs, the PML protein and its SUMO modification are required for PML body formation and colocalisation with other components such as Sp100²⁶. In contrast, the linear PML structure in iPSCs evidently lacked enrichment of SUMO-1 and Sp100. Finally, we found that the PML-defined structure in iPSCs transited to a somatic sphere PML body under differentiation conditions (day 6–10) in parallel with an increased number of the bodies (Fig. 2g). The resulting PML bodies coexisted with SUMO-1 and Sp100 on day 10 (Supplementary Fig. S9). Thus, the PML structure is dynamically regulated in iPSCs during their differentiation, indicating that the linear form of the PML body is one of the hallmarks of fully reprogrammed iPSCs.

In summary, we report the morphometric characteristics of human iPSCs by quantitative assessment of colony and nuclear structures.

Discussion

In the present study, we developed a new non-invasive method to distinguish nascent reprogrammed iPSC and non-iPSC colonies based on their morphologies. Previously, mouse reprogramming studies have used reporters integrated into the genomic loci of pluripotency genes *Fbx15*, *Oct4*, or *Nanog*^{28–30} to identify reprogrammed cells. However, there have been no methods to reliably identify human iPSCs in a population of fibroblasts and imperfectly reprogrammed cells without cell labelling⁴. Our analysis using a collection of cell lines including standard iPSC lines (201B7 and 253G1)^{2,3}, newly generated iPSC lines (1H–4H), non-iPSC lines (15B2 and 2B7), and somatic cells (human mammary epithelial cells, HMECs) demonstrated that wndchrn analysis is effective and objective for discrimination of iPSC and non-iPSC colonies.

Quantitative measurement of morphological differences can be very complex, and it is sometimes difficult to analyse only pre-defined features^{6,7,9}. Therefore, we used a supervised machine learning system, wndchrn, which has been developed to automatically mine for morphological similarities in a wide variety of objects^{7,9}. As the first step, wndchrn computes a large number of features of both raw images and ones processed for transformations including Fourier, Wavelet and Chebyshev. Total calculated feature numbers are 2873 at maximum, which are expected to cover general image features. They include polynomial decompositions, high contrast features, pixel statistics and textures, and so on^{7,19}. Next, wndchrn selects an informative feature by rejecting noisy feature by using Fisher Linear Discriminant algorithm. The Fischer scores are also used as feature weights, so that less discriminating features have a reduced effect on the classifier. The classifier used is WND5 which is

related to the k-nearest neighbors type, except that it uses a negative exponential in a weighted feature space rather than a simple linear distance equation. While the classification system is composed of several linear elements, it is not justified to characterize this system as somehow limited in the complexity of the feature relationships it can exploit. In fact, wndchrn outperforms over a dozen state of the art classifiers which are constructed as either general, problem specific, linear or non-linear programming^{7,31}. Because the method used here is versatile and not limited to any particular type of cell or image, it would be applicable to classify cells in various states, providing that a sufficient numbers of images and their retrospective metadata is available.

In this study, we found image features that discriminate between iPSCs and non-iPSC (Fig. 1e) within colony components. It is interesting that the value of morphological similarity does not depend on the original cell types or reprogramming procedures. For example, 201B7, 253G1, and 15B2 cells were all derived from human fibroblasts, However, 201B7 cells were closely clustered with 1H–4H cells that were derived from HMECs, suggesting that the cells have lost their lineage identities with morphological features during reprogramming. It is also interesting that the morphologies of the two non-iPSC lines (15B2 and 2B7) were not only distinct from iPSCs, but also from each other, implying that they have failed reprogramming at different stages or for different reasons (Supplementary Fig. S4d).

Prior to the colony classifications, we analyzed the cell lines for their undifferentiated states and differentiation abilities to find that they formed relatively homogeneous colonies (Supplementary Figs S1 and S2). However, reprogrammed cells generally could form a colony that is composed of different types of cells. A question of how the classification technology works on the heterogeneous colonies remains to be investigated. It is of interest that tiling an image was effective to determine the source of the classification signal in cultured colonies (Fig. 1e) and cells¹⁰. It was also successfully applied to determine the location of the predictive osteoarthritis signal in X-Rays³².

We found that nuclear structures changed during reprogramming dynamically and specifically. Using immunofluorescence images of the PML and Cajal bodies, bona fide iPSCs (1H cells) were discriminated from non-iPSCs (15B2 cells) by extremely high accuracy, almost 100% accuracy, with wndchrn (Fig. 2c). This result suggests that one can use the information on the different nuclear bodies to discriminate between cell types effectively.

Among the nuclear structures, the linear PML-defined structure is an attractive indicator that represents the reprogramming state of iPSCs. The linear form of PML bodies has been observed in human embryonic stem (ES) cells³³, but not in mouse ES cells, possibly because human ES cells correspond to mouse-derived epiblast stem cells^{34,35}. The ring-shaped PML body found in human non-iPSCs may represent a transition state from somatic cells to iPSCs. Collectively, the linear PML structure is likely to be characteristic of the pluripotent state of human cells.

In conclusion, the present study indicates that the reprogramming states of live human iPSCs can be evaluated precisely by machine learning technologies for image analyses. This versatile method is applicable to other cell types, and may be valuable for quality control of cells intended for regenerative medicine, as well as for basic research. Our findings will also significantly advance our knowledge of the nuclear landscape of iPSCs.

Methods

Generation and maintenance of iPSC lines. To generate human iPSC lines, four reprogramming factors (*Oct3/4*, *Sox2*, *Klf4*, and *c-Myc*) were introduced into HMECs and primary human fibroblasts using SeV vectors according to the manufacturer's protocol (Cyto tune-iPS DNAVEC). Briefly, 2×10^5 cells were plated and infected with SeV vectors, and then cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum. After ES cell-like



colonies appeared, the medium was changed to primate ES cell medium supplemented with 5 ng/ml basic fibroblast growth factor (bFGF) (ReproCELL). Growing colonies were picked up mechanically, expanded, and maintained on mouse embryonic fibroblasts (MEFs) to avoid spontaneous differentiation. Standard human iPSC lines, 201B7 and 253G1, were provided by Kyoto University and RIKEN BioResource Center, Japan, respectively. These cell lines were cultured on MEFs in Repro Stem medium supplemented with 5 ng/ml bFGF and penicillin/streptomycin.

In vitro differentiation of human iPSCs. For embryoid body (EB) formation, iPSCs were treated with a dissociation solution (CTK, ReproCELL), and clumps of cells were cultured in DMEM/F12 containing 20% knockout serum replacement (Invitrogen) supplemented with 9.2 mM L-glutamine, 1×10^{-4} M non-essential amino acids, 1×10^{-4} M 2-mercaptoethanol, and penicillin/streptomycin. The medium was changed every other day. After 8 days of floating culture, EBs were transferred to a gelatin-coated plate and cultured in the same medium for another 8 days.

Immunofluorescence analysis. Cells were fixed with 4% paraformaldehyde in PBS for 10 min at room temperature, washed, and permeabilized with PBS containing 0.5% Triton X-100 for 5 min on ice. The cells were then incubated with primary antibodies for 1 h, followed by secondary antibodies for 1 h. Images were obtained under a microscope (IX-71; Olympus) equipped with a $60 \times$ NA1.0 Plan Apo objective lens and a cooled charged-coupled device camera (Hamamatsu). Alternatively, images were captured with a confocal laser-scanning microscope (LSM 710, Carl Zeiss) with a $63 \times/1.4$ Plan-Apochromat objective lens and a cooled charged-coupled device camera (Carl Zeiss). For immunofluorescences of the nuclear structures (PML body, Cajal body and PNC), image stacks containing three-dimensional datasets were collected at 1.0 μ m intervals through the z axis, and projected onto two dimensions using imaging software (Lumina Vision; Mitani Corp). For structured illumination microscopic analyses, we used a microscope (Ti-E; Nikon) equipped with a $100 \times$ NA1.49 CFI Apo TIRF objective lens, an electron multiplying charged-coupled device camera (iXon Em-CCD, Andor), and image acquisition software (Nikon).

Image quantification. For image classifications, we used wndchrm ver. 1.31⁶⁷. Images used were; phase contrast of colonies (1024×767 pixels) (Supplementary Fig. S3), and immunofluorescences of nuclear structures (1280×1024 pixels) (Supplementary Fig. S8). No further segmentation was done prior to wndchrm analysis. The numbers of training/test images were 60/6 (Fig. 1), 40/8 (PML body in Fig. 2), 35/7 (Cajal body in Fig. 2), and 26/7 (Supplementary Fig. S5). The options used were a larger feature set of 2873 (-1), tiling of an image into 4 (-t4) for Fig. 1b–1d, and (-1, -t8) for Fig. 1e. Fisher scores were automatically computed for each feature in the following groups: ChF, Chebyshev-Fourier Statistics; Ch, Chebyshev Statistics; Com, Combined First Fourier Moments; Fra, Edge and Fractal Statistics; Har, Haralick Texture; MsH, Multiscale Histogram; Zer, Zernike Moments (Fig. 1c and Supplementary Fig. S4c). Pairwise class similarity values in Supplementary Table S2 were computed from the average of the marginal probabilities of all of the test images in each class. The per-class marginal probabilities were used as coordinates in a marginal probability space, where pairwise inter-class distances were computed using the Euclidean distance formula. Morphological similarity is the inverse of morphological distance between classes. Phylogenies were computed using the Fitch-Margoliash method implemented in the PHYLIP package, which is based on pairwise class similarity values reported by wndchrm ver 1.3³⁶. For imaging cytometry analyses, CELAVIEW RS100 (Olympus) was used to capture images, and quantitative analyses was done using CELAVIEW analysis software as following. For automated quantification of PML body, Cajal body and PNC, each fluorescent image was segmented in CELAVIEW (Olympus) using the DAPI channel to define the nuclei. The nuclear bodies were automatically segmented based on their intensity and area using CELAVIEW as previously reported³⁷.

RNA isolation and quantitative PCR analysis. Total RNA was isolated using TRIzol (Invitrogen). For cDNA synthesis, 1 μ g of total RNA was reverse transcribed with a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Quantitative PCR of target cDNAs was performed using Power SYBR Green PCR Master Mix (Applied Biosystems). Each experiment was performed at least three times. Relative fold changes were quantified by normalisation to β -actin expression. Primer sequences are listed in Supplementary Table S3.

Antibodies. The primary antibodies used were: mouse PML (1 : 500, SC-966, Santa Cruz), p80 coilin (1 : 300, #612074, BD Biosciences), hnRNPI (1 : 300, sc-16547, Santa Cruz), lamin A/C (1 : 300, sc-7292, Santa Cruz), Sp1 (1 : 300, sc-16547, Santa Cruz), and lamin B (1 : 300, sc-6216, Santa Cruz). The following secondary antibodies were used: Alexa 488-conjugated donkey anti-mouse IgG (1 : 250), Alexa 488-conjugated donkey anti-rabbit IgG (1 : 250) (Molecular Probes), Cy3-conjugated donkey anti-mouse IgG (1 : 1000), Cy3-conjugated donkey anti-rat IgG (1 : 1000), and Cy3-conjugated donkey anti-rabbit IgG (1 : 1000) (Jackson ImmunoResearch).

1. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).

2. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
3. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101–106 (2008).
4. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPSC cells from partially reprogrammed cells. *Nat Biotechnol* **27**, 1033–1037 (2009).
5. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* **28**, 1079–1088 (2010).
6. Eliceiri, K. W. *et al.* Biological imaging software tools. *Nat Methods* **9**, 697–710 (2012).
7. Orlov, N. *et al.* WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recog Lett* **29**, 1684–1693 (2008).
8. Grisendi, S. *et al.* Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* **437**, 147–153 (2005).
9. Johnston, J., Iser, W. B., Chow, D. K., Goldberg, I. G. & Wolkow, C. A. Quantitative image analysis reveals distinct structural transitions during aging in *Caenorhabditis elegans* tissues. *PLoS One* **3**, e2821 (2008).
10. Shamir, L., Orlov, N., Mark Eckley, D., Macura, T. J. & Goldberg, I. G. IICBU 2008: a proposed benchmark suite for biological image analysis. *Med Biol Eng Comput* **46**, 943–947 (2008).
11. Scaglioni, P. P. *et al.* Translation-dependent mechanisms lead to PML upregulation and mediate oncogenic K-RAS-induced cellular senescence. *EMBO Mol Med* **4**, 594–602 (2012).
12. Zhao, R., Bodnar, M. S. & Spector, D. L. Nuclear neighborhoods and gene expression. *Curr Opin Genet Dev* **19**, 172–179 (2009).
13. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol* **8**, 1006–1016 (2007).
14. Morris, G. E. The Cajal body. *Biochim Biophys Acta* **1783**, 2108–2115 (2008).
15. Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C. & Dreyfuss, G. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res* **20**, 3671–3678 (1992).
16. Norton, J. T. & Huang, S. The perinuclear compartment: RNA metabolism and cancer. *Cancer Treat Res* **158**, 139–152 (2013).
17. Fussner, E. *et al.* Constitutive heterochromatin reorganization during somatic cell reprogramming. *The EMBO J* **30**, 1778–1789 (2011).
18. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 348–362 (2009).
19. Shamir, L. *et al.* Wndchrm - an open source utility for biological image analysis. *Source Code Biol Medicine* **3**, 13 (2008).
20. Shamir, L., Delaney, J. D., Orlov, N., Eckley, D. M. & Goldberg, I. G. Pattern recognition software and techniques for biological image analysis. *PLoS Comp Biol* **6**, e1000974 (2010).
21. Gasser, S. M. Visualizing chromatin dynamics in interphase nuclei. *Science* **296**, 1412–1416 (2002).
22. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews. Molec Cell Biol* **7**, 540–546 (2006).
23. Aoto, T., Saitoh, N., Ichimura, T., Niwa, H. & Nakao, M. Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. *Dev Biol* **298**, 354–367 (2006).
24. Constantinescu, D., Gray, H. L., Sannak, P. J., Schatten, G. P. & Csoka, A. B. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells* **24**, 177–185 (2006).
25. Safe, S. & Abdelrahim, M. Sp transcription factor family and its role in cancer. *Eur J Cancer* **41**, 2438–2448 (2005).
26. Zhong, S., Salomoni, P. & Pandolfi, P. P. The transcriptional role of PML and the nuclear body. *Nat Cell Biol* **2**, E85–90 (2000).
27. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet* **27**, 295–306 (2011).
28. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
29. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
30. Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
31. Orlov, N. V., Eckley, D. M., Shamir, L. & Goldberg, I. G. Improving class separability using extended pixel planes: a comparative study. *Machine Vis Applic* **23**, 1047–1058 (2012).
32. Shamir, L. *et al.* Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarth Cartilage/OARS, Osteoarth Res Society* **17**, 1307–1312 (2009).
33. Butler, J. T., Hall, L. L., Smith, K. P. & Lawrence, J. B. Changing nuclear landscape and unique PML structures during early epigenetic transitions of human embryonic stem cells. *J Cell Biochem* **107**, 609–621 (2009).
34. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
35. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
36. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).



37. Saitoh, N. *et al.* The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Mol Biol Cell* **23**, 1115–1128 (2012).

Acknowledgments

We thank Dr. Tamiyo Kobayashi (Olympus, Japan) for technical assistance. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Japan Science and Technology Agency (CREST) (M.N. and N.S.).

Author contributions

K.T., N.S. and M.N. designed and performed the experiments, and prepared the manuscript, together with I.G.G. for the wndchrn analyses, Y.Y. and S.Y. for the images of iPSC colonies, and C.S. and Y.Y. for technical assistance.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: S.Y. is a member without salary of the scientific advisory board of iPS Academia Japan. Y.Y. is a founder of iPS Portal.

How to cite this article: Tokunaga, K. *et al.* Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. *Sci. Rep.* **4**, 6996; DOI:10.1038/srep06996 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>