Stem Cell Surface Marker Expression Defines Late Stages of Reprogramming to Pluripotency in Human Fibroblasts

JORDAN E. POMEROY,a SHELLEY R. HOUGH,b橼 KATHRYN C. DAVIDSON,c,f ALEX M. QUAAO,
JORDAN A. REEs,a MARTIN F. PERA,b,e,f

Key Words. Induced pluripotent stem cell • Cell surface marker • Reprogramming • CDH3 • Gene expression

ABSTRACT

Our current understanding of the induction of pluripotency by defined factors indicates that this process occurs in discrete stages characterized by specific alterations in the cellular transcriptome and epigenome. However, the final phase of the reprogramming process is incompletely understood. We sought to generate tools to characterize the transition to a fully reprogrammed state. We used combinations of stem cell surface markers to isolate colonies emerging after transfection of human fibroblasts with reprogramming factors and then analyzed their expression of genes associated with pluripotency and early germ lineage specification. We found that expression of a subset of these genes, including the cell-cell adhesion molecule CDH3, characterized a late stage in the reprogramming process. Combined live-cell staining with the antibody GCTM-2 and anti-CDH3 during reprogramming identified colonies of cells that showed gene expression patterns very similar to those of embryonic stem cell or established induced pluripotent stem cell lines, and gave rise to stable induced pluripotent stem cell lines at high frequency. Our findings will facilitate studies of the final stages of reprogramming of human cells to pluripotency and will provide a simple means for prospective identification of fully reprogrammed cells. *STEM CELLS TRANSLATIONAL MEDICINE* 2016;5:870–882

SIGNIFICANCE

Reprogramming of differentiated cells back to an embryonic pluripotent state has wide ranging applications in understanding and treating human disease. However, how cells traverse the barriers on the journey to pluripotency still is not fully understood. This report describes tools to study the late stages of cellular reprogramming. The findings enable a more precise approach to dissecting the final phases of conversion to pluripotency, a process that is particularly poorly defined. The results of this study also provide a simple new method for the selection of fully reprogrammed cells, which could enhance the efficiency of derivation of cell lines for research and therapy.

INTRODUCTION

The discovery of the induction of pluripotency by defined factors has raised new possibilities for the use of stem cells in disease modeling and regenerative medicine [1]. Since the initial description of induced pluripotent stem cells (iPSCs), advances in reprogramming technology have provided alternatives to retroviral integration and eliminated the need for permanent alteration to the cellular genome in the process [2]. However, concerns about the genetic integrity and differentiation capacity of iPSC remain [3, 4].

The reprogramming process has been extensively scrutinized at the molecular level, and many new insights into the epigenetic basis of induction of pluripotency have emerged [5]. Critical to this analysis has been the availability of molecular landmarks that reproducibly chart the progress of reprogramming within the cell. The availability of appropriate markers enhances efficient selection of bona fide reprogrammed cells or colonies and helps to facilitate mechanistic studies through identification of well-defined intermediate stages during the process. Although genetically engineered cell lines with reporters introduced into specific loci that undergo reactivation during reprogramming provide powerful tools for analysis [6], it is not practical to introduce such modifications into hundreds of fibroblast or blood cell isolates for large-scale derivation. Cell surface markers identified by monoclonal antibodies, therefore, have an important role in the identification and selection of reprogrammed
colonies and in monitoring progress of the reprogramming process itself [7].

Lowry et al. [8] were the first to demonstrate that live-cell staining with the antibody TRA-1-60 could be used to identify fully reprogrammed human iPSCs (hiPSCs). Chan et al. [9] confirmed and extended this observation; in this latter study, other markers of the pluripotent state, including the canonical stem cell surface marker SSEA-4, were not sufficient to distinguish fully reprogrammed cells from partially or incompletely reprogrammed cells. Only a combination of several markers plus assessment of intensity of staining with the DNA dye Hoechst 33258 could successfully identify reprogrammed cells. Another study described the use of a combination of SSEA-4, TRA-1-60, and CD30 to isolate fully reprogrammed cells from flow cytometry [10]. Tanabe et al. [11] showed that TRA-1-60-positive cells did not always progress to the fully reprogrammed state, and they identified maturation as a key roadblock in reprogramming. Recently Takahashi et al. [12] used TRA-1-60 staining to isolate cells at a late stage of reprogramming that expressed markers of the primitive streak, and they suggested these cells represented an intermediate in the reprogramming process, providing further evidence that cells expressing this marker have not completed the reprogramming process.

In this study, we use live-cell staining with several combinations of antibodies to isolate colonies that emerged following transfection of fibroblasts with a PiggyBac transposon vector (System Biosciences, Palo Alto, CA, https://www.systembio.com) containing the four reprogramming factors (POUSF1, SOX2, KLF4, CMYC) and examined gene expression in these colonies using a panel of genes including key pluripotency transcription factors, cell surface and secreted molecules involved in stem cell renewal, and lineage-specific transcription factors. We used previous analyses of gene expression in bulk populations and immunologically defined subsets of human pluripotent stem cells from our laboratory and others to identify a novel panel of markers that characterized different stages in a continuous hierarchy from self-renewing subpopulations to lineage primed cells [13–15]. The results showed that surface markers can be used to select colonies with specific patterns of gene expression at later stages of reprogramming, thereby defining a transition late in the reprogramming process and allowing facile selection of colonies of fully reprogrammed cells.

**MATERIALS AND METHODS**

Standard and xeno-free culture systems for human pluripotent stem cells, and for derivation and characterization of hiPSC including Fluidigm gene expression analysis (Fluidigm Corporation, South San Francisco, CA, https://www.fluidigm.com), are described in the supplemental online data.

**Marker Expression During Reprogramming**

Low passage human fetal dermal fibroblasts (hDFDs) were nucleofected with 4 μg of pPB.OSKM-puLtk transposon and 2 μg of pCyL43-PB transposase plasmids as described in the supplemental online data. After resuspension in fibroblast culture medium, nucleofected cells were plated in two-well chamber slides coated with 0.5 mg of vitronectin/fibronectin at a density of 1 × 10^4 cells per cm². Fibroblast medium was changed the day after transfection and every other day until day 5 when the medium was replaced with XF-KSR human embryonic stem cell (hESC) medium (ThermoFisher Scientific, Waltham, MA, http://thermofisher.com) with feeding continuing every other day. Nontransfected hDFDs were also plated at the same density as negative controls for antibody staining. Beginning on day 3 and ending on day 24, slides were fixed with 4% paraformaldehyde in phosphate-buffered saline at 3-day intervals. Fixed slides were stained with the following pluripotency marker combinations on days 3, 6, 9, 12, 15, 18, 21, and 24: TRA-1-60/GDF3, GCTM-2/EPACAM, POU5F1/DNMT3B, NANOG/CDH1, EPACAM/DNMT3B, and EPACAM/CDH1 (the sources of these antibodies are listed in the supplemental online data). Staining conditions were identical to those described previously.

Pluripotency marker expression was compared with the positive control hiPSC line J1. Three categories were utilized to score the staining: negative (matches negative control/background), positive (staining intensity above background with correct cellular localization), and intense (staining intensity resembles positive control staining). Quantification by integration of fluorescence above individual nuclei using ImageJ software (NIH, Bethesda, MD, https://imagej.nih.gov/ij) supported the three-category scoring for marker positivity, although assessment of marker staining was performed visually. All colonies present on the slides were scored according to the previously indicated criteria and were noted to be double-negative, single-positive, or double-positive for each marker pair. All intense staining colonies were also scored/included in the positive category.

**Selection of Colonies for Fluidigm BioMark HD Expression Profiling of Single Reprogramming Foci**

Low passage hDFDs were nucleofected with 4 μg of pPB.OSKM-puLtk transposon and 2 μg of pCyL43-PB transposase plasmids as described in Marker Expression During Reprogramming, and then they were plated at a density of 1 × 10^3 cells per cm² onto 35-mm cell culture dishes coated with 0.5 μg per cm² of vitronectin/fibronectin. Fibroblast medium was changed the day after transfection and every other day until day 5 when the medium was replaced with XF-KSR hESC medium with feeding continuing every other day. On days 10, 20, and 30, reprogramming dishes were stained by live-cell staining with the marker combinations GCTM-2/EPACAM and TRA-160/CDH1 and CDH3/GCTM-2 as follows. Primary and secondary antibodies were diluted into XF-KSR hESC medium and incubated consecutively for 20 minutes at 37°C with three sterile phosphate-buffered saline (PBS) washes following each incubation. One milliliter of sterile PBS was added to each 35-mm cell culture dish for imaging with a Zeiss Axio Vert.A1 (Zeiss, Jena, Germany, http://www.zeiss.com). Hoechst dye was added during the secondary antibody incubation for nuclear counterstaining. Primary and secondary antibody pairings were as follows: GCTM-2/Alexa Fluor 488 goat anti-mouse IgM (ThermoFisher Scientific); EPACAM/Alexa Fluor 568 goat anti-mouse IgG1 (ThermoFisher Scientific); TRA-160/Alexa Fluor 594 goat anti-mouse IgG1 (ThermoFisher Scientific); and CDH3/Alexa Fluor 488 goat anti-mouse IgM. In subsequent experiments, dishes containing reprogramming foci were stained at day 30 with a combination of anti-CDH3 (MAB 861; R&D Systems Inc., Minneapolis, MN, http://www.rndsystems.com) followed by Alexa Fluor 488 goat anti-mouse IgG1 and GCTM-2 followed by Alexa Fluor 594 goat anti-mouse IgM. Double-negative and double-positive colonies were imaged and isolated by manual dissection for expression analysis. Nontransfected hDFDs (negative control) and hiPSCs/hESCs (positive controls) were also live-cell stained and isolated.

Isolated colonies were immediately lysed in 250 μL of lysis buffer contained in the Qiagen RNeasy Micro Kit (Qiagen, Hilden, Germany).
Germany, https://www.qiagen.com) and RNA was extracted following the Qiagen protocol.

**Live-Cell Selection and Characterization of hiPSCs Using GCTM-2/EP CAM and TRA-1-60/CDH1**

Low-passage hDFs were nucleofected and plated as described in the previous section. After 30 days, reprogramming dishes were live-cell stained with the marker pairs GCTM-2/EP CAM and TRA-1-60/CDH1. Staining conditions were identical to the previous section except that Hoechst dye was not added. Double-positive and double-negative foci were imaged, isolated by manual dissection, and transferred to fresh XF culture dishes. hiPSCs exhibiting attachment and outgrowth were continually passaged as long as they were viable, and they were observed for morphological characteristics.

Reprogrammed colonies were plated in eight-well chamber slides for immunocytochemistry and stained for the following markers: TRA-1-60/GDF3, GCTM-2/EP CAM, POU5F1/DNMT3B, NANOG/CDH1, EPCAM/DNMT3B, EPCAM/CDH1, SSEA-4/SOX2, and alkaline phosphatase. Live-cell-selected hiPSCs were also prepared for spin-embryoid body (see supplemental online data). Differentiation assays.

**Selection of Reprogrammed Cells With CDH3 Live-Cell Staining**

Human W51 (ATCC, Manassas, VA, http://www.atcc.org) and BJ (Stemgent, Lexington, MA, https://www.stemgent.com) fibroblasts were cultured in Dulbecco’s modified Eagle’s medium (DMEM) with high glucose, 10% fetal bovine serum (FBS), 100 U/ml of penicillin, and 100 μg/ml of streptomycin. iPSCs were cultured on irradiated mouse embryonic fibroblasts (MEFs) (2.5 × 10^3 per cm^2) on gelatin-coated plates in knockout serum replacement (KSR) medium containing 20 ng/ml of FGF2 (EMD Millipore, Darmstadt, Germany, http://www.emdmillipore.com): DMEM/F-12 with GlutaMAX (ThermoFisher Scientific), 20% KSR, 0.1 mM of MEM-NEAA, 0.1 mM of β-mercaptoethanol, 50 U/ml of penicillin, and 50 μg/ml of streptomycin. iPSC colonies were subcultured weekly as small clusters using dispase (1.75 U/ml) and gentle dissociation with a Pasteur pipette. MEF carryover was minimized during passage by allowing iPSC clumps to settle under gravity in a conical tube, then aspirating the medium above the pellet before seeding. The optimal split ratio ranged from 1:5 to 1:20 and was determined empirically for each individual iPSC line. KSR media was replenished daily.

Fibroblasts were reprogrammed to iPSCs with nonintegrating episomal vectors containing six reprogramming factors (OCT4, SOX2, KLF4, L-MYC, LIN28, and shRNA for p53) according to Okita et al. [16]. Six hundred thousand fibroblasts were transfected with 1 μg each of pCXLE-hOCT4-shP53, pCXLE-hSK, and pCXLE-hUL plasmids (Addgene, Cambridge, MA, https://www.addgene.org) using nucleofection (Lonza Nucleofector; Lonza Group, Basel, Switzerland, http://www.lonza.com). Seven days postnucleofection, fibroblasts were trypsinized and 1 × 10^5 cells were replated onto 3- × 6-cm gelatin-coated plates containing irradiated MEFs. The following day, fibroblast media was changed to KSR medium with 20 ng/ml of FGF2; media was replenished every other day. Primary iPSC colonies emerged within approximately 20 days.

On day 26 postnucleofection, emerging colonies were coimmunolabeled with GCTM-2 (hybridoma supernatant, neat) and either mouse anti-human CDH3 (1:20) or mouse anti-human EPCAM (1:50) for 20 minutes at 37°C in a 5% CO₂ incubator (2 ml of antibody solution per 6-cm plate). Cells were rinsed twice with PBS then incubated for 20 minutes at 37°C with isotype-specific secondary antibodies: goat anti-mouse IgG1-Alexa Fluor 488 and goat anti-mouse IgM-Alexa Fluor 594 (both 1:500 in KSR medium). Cells were then rinsed twice in PBS, imaged on a Nikon Eclipse TE2000U inverted microscope with NIS Elements software (Nikon, Tokyo, Japan, http://www.nikon.com/), and double-positive colonies (GCTM-2+CDH3+ or GCTM-2+EPCAM+) were marked on each plate. These colonies were mechanically selected under a Leica M26 stereomicroscope in a biosafety hood using a 21-g needle and transferred to organ culture plates with MEFs (denoted P0 for passage 0). Clonal iPSC colonies were expanded for 9 days then subcultured by mechanical dissection each week until they exhibited stable, robust growth of morphologically undifferentiated colonies, after which point they were passaged with dispase as per conventional methods for propagating established iPSC lines (usually within 1 to 2 passages).

Characterization of selected iPSC lines, including GCTM-2/CD9 flow cytometry and multilineage differentiation in embryoid body assays, was carried out as described in the supplemental online data.

**RESULTS**

**Expression of Stem Cell Markers in Reprogramming Colonies Reaches Maximum at Day 21**

Many studies of reprogramming have analyzed the process by studying whole populations of cells or cellular subpopulations fractionated by flow cytometry using particular markers. In this work, we focused on analyzing growing colonies that emerged from a single clonal reprogramming event, for two reasons. First, the selection of human iPSCs is often carried out at the colony level, to maximize survival and recovery of reprogrammed cells. Second, in situ analysis of colonies can provide insight into how the progeny of single transfected cells evolve, a feature that is obscured by harvesting cells prior to analysis.

We monitored expression of cell surface glycoproteins (TRA-1-60, GCTM-2) cell adhesion molecules (CDH1, EP CAM); a growth factor (GDF3); and transcription factors (POU5F1, SOX2, NANOG, DNMT3B) in expanding foci of transfected cells (supplemental online Fig. 1) every 3 days (days 3, 6, 9, 12, 15, 18, 21, 24) following addition of reprogramming factors. The GCTM-2 antibody recognizes a cell surface glycoprotein that bears the TRA-1-60 and TRA-1-81 epitopes. The following marker combinations were used: TRA-1-60/GDF3 (Fig. 1A), EP CAM/GCTM-2 (Fig. 1B), DNMT3B/POU5F1 (supplemental online Fig. 2A), CDH1/NANOG (Fig. 1C), EP CAM/DNMT3B (supplemental online Fig. 2B), and EP CAM/CDH1 (Fig. 1D). Colonies of varying staining positivity and intensity for all markers were detected at each time point from day 9 onward, indicating nonuniform reprogramming. Most colonies showed heterogeneity of staining, with only subsets of cells within the colonies expressing the markers. As the reprogramming process progressed, the proportion of cells within a colony that stain positive for particular markers increased. Positive staining for POU5F1 (supplemental online Fig. 2A) in foci indicated that the majority of the expanding colonies had undergone transfection and were expressing the transgene. The selected colonies illustrated in these figures are representative of the strongest staining at each time point.

The time course of antibody staining, determined by counting the incidence and intensity of stained colonies, is depicted in
Figure 2. Staining was classified as positive (correct subcellular localization, above background) or intense (correct subcellular localization at a level similar to positive control hESC or iPSC lines). A colony was scored positive even if a small proportion of cells within it exhibited expression. As the reprogramming process progressed, the percentage of colonies staining positively for the markers increased. Intense POU5F1 staining was observed from day 3. GDF3- and NANOG-positive colonies appeared early (days 6–9, respectively). The cell surface markers including cell-cell adhesion molecules did not begin to appear until approximately day 12, as did DNMT3B. Cells that were double-positive for sets of markers began to appear on days 12–15, and intensely staining foci began to appear on days 15–18. Intense staining, including double-positive intense staining, was rarely seen in more than 10% of colonies and peaked by day 21. Each chamber slide well contained 100 to 200 reprogramming foci for analysis. With an initial seeding density of $5 \times 10^3$ cells/well, this represents a reprogramming efficiency of approximately 3%.

Figure 2. Staining was classified as positive (correct subcellular localization, above background) or intense (correct subcellular localization at a level similar to positive control hESC or iPSC lines). A colony was scored positive even if a small proportion of cells within it exhibited expression. As the reprogramming process progressed, the percentage of colonies staining positively for the markers increased. Intense POU5F1 staining was observed from day 3. GDF3- and NANOG-positive colonies appeared early (days 6–9, respectively). The cell surface markers including cell-cell adhesion molecules did not begin to appear until approximately day 12, as did DNMT3B. Cells that were double-positive for sets of markers began to appear on days 12–15, and intensely staining foci began to appear on days 15–18. Intense staining, including double-positive intense staining, was rarely seen in more than 10% of colonies and peaked by day 21. Each chamber slide well contained 100 to 200 reprogramming foci for analysis. With an initial seeding density of $5 \times 10^3$ cells/well, this represents a reprogramming efficiency of approximately 3%.

Coexpression of Cell Surface Markers Identifies Reprogramming Foci With Transcriptional Profiles Closer to Pluripotent Cells

To carry out transcriptional profiling of emerging reprogrammed colonies, we performed live-cell staining with the cell surface marker combinations GCTM-2/EPCAM and TRA-1-60/CDH1 (Fig. 3) at several time points to select double-positive emerging colonies of cells undergoing reprogramming (reprogramming foci) from fibroblasts. As a positive control, we used the embryonic stem cell line USC01 and hiPSC cell lines AA4 and AE1, which expressed markers of pluripotency and differentiated into derivatives of all three germ layers in embryoid body or teratoma assays [17]. These control cell lines exhibited strong positivity for all four markers (Figs. 3A, 3B, 3D, and 3E). GD30F (Fig. 3C) is an example of intense GCTM-2/EPCAM staining arising after 30 days of reprogramming and TD302B (Fig. 3F) is a reprogramming focus with intense CDH1/TRA-1-60 staining after 30 days.
Double-positive and double-negative colonies were harvested for RNA expression analysis at days 10, 20, and 30 following transfection. All colonies isolated at day 10 after live-cell staining were marker-negative, as expected from the reprogramming timeline study. By day 20, several colonies exhibited intense staining with increasing number and size of strongly staining colonies by day 30. We analyzed more than 150 foci along with positive and negative controls for expression of a panel of pluripotency and early lineage genes by quantitative RT-polymerase chain reaction using the Fluidigm system (Fig. 4). Hierarchical cluster analysis of the data provides two major clusters along the vertical axis. The top vertical cluster (pluripotent) groups the positive control hESC and hiPSC cell lines with a majority of the day 30 double-positive samples and a few day 20 double-positive samples. The bottom vertical cluster (fibroblastic) groups the negative control hFDF samples along with the majority of the day 20 double-negative, and all day 20/30 double-negative samples. A few major clusters also appear along the horizontal (gene) axis. Mouse *Pou5f1* is expressed strongly in some hiPSC and all reprogramming foci samples tested, indicating continued activity of the reprogramming transposon. Some genes were clearly upregulated in the pluripotency group but were also expressed in a subset of day 10 and day 20 negative colonies. These genes included *DPPA3*, *NANOG*, endogenous (human) *POU5F1*, *LEFTY1*, *NODAL*, *TDGF1*, *ERBB3*, and *DNMT3B* (gene clusters 1 and 2). This group of upregulated genes includes several
canonical pluripotency master regulators. Another cluster of genes exhibited strong expression in the positive controls but limited expression among some of the double-positive staining samples from days 20 and 30 in the pluripotent group: **FOXD3**, **CDH1**, **EPHA1**, **LCK**, **CDH3**, **SOX2**, **EDNRB**, and **HAS3** (gene cluster 3). Some of these genes are known pluripotency regulators (**FOXD3**, **SOX2**), although the functions of the remaining genes in the maintenance of the pluripotent state are largely unknown. Finally, some genes are activated in a subset of day 20 and day 30 positive foci and positive controls, including markers of mesendoderm such as **MIXL1**, **GOOSECOID**, **T**, and **CER1** (gene cluster 4).

A time course analysis of the proportion of foci expressing early upregulated genes (Fig. 5A), and the corresponding data for late upregulated genes (Fig. 5B), reflected the hierarchical clustering data. Although it is apparent that both classes of genes are expressed within an increasing proportion of foci over time, the early group demonstrates significant expression by day 10 with expression also noted in the day 20N and day 30N double-negative samples. The late group has very limited expression in double-negative foci (day 10N, day 20N, day 30N) and double-positive foci isolated at day 20. By contrast, approximately half of the double-positive foci isolated at day 30 show expression of these late genes. The expression differences between the fully reprogrammed hiPSC controls and day 30P double-negative foci are highly significant ($p < .01$) to very highly significant for the majority of late genes.

Examples of gene expression analysis for specific stained colonies are shown in supplemental online Figure 3.

While this work was in progress, Cacchiarelli et al. [18] carried out an extensive transcriptional and epigenomic analysis of the induction of pluripotency in human fibroblasts using an inducible transgene system. The time course of expression of our early and late marker gene panels in their study is shown in supplemental online Figure 4. The time course of their study is more compressed than ours, probably due to their use of an inducible transgene system, but the overall patterns are quite similar.

### Selection Using Late Markers of Reprogramming Increases hiPSC Derivation Efficiency

Colonies exhibiting dual positivity for either set of markers showed dramatically higher derivation efficiency compared with negative colonies at day 30 (supplemental online Table 1). None of the marker-negative colonies gave rise to fully reprogrammed hiPSCs (defined by continuous growth and positive marker expression). Of 53 marker-negative colonies, 30 attached and exhibited outgrowths resembling partially reprogrammed hiPSCs. Of the 73 marker-positive colonies subjected to live-cell selection, 19 attached and exhibited outgrowths. Fourteen of these lines could be serially propagated beyond five passages and exhibited a fully reprogrammed state on the basis of

---

**Figure 3.** Double-labeled indirect immunofluorescence micrographs showing live-cell staining for GCTM-2/EPCAM (A–C) and CDH1/TRA-1-60 (D–F). (A, D): Human embryonic stem cell line USC01. (B, E): Human induced pluripotent stem cell lines AA4 and AE1, respectively. (C, F): Day 30 colonies. Owing to the high cell density in reprogramming dishes in (C) and (F), the resolution of the live-cell staining is lower than in (A–D), but positive foci are easily detected above background. Scale bars = 20 μm.
Figure 4. Heat map and unsupervised hierarchical cluster analysis showing gene expression in growing colonies that were positive or negative for marker expression at 10, 20 and 30 days after gene transfection with reprogramming factors compared with parental FIBRO, hiPSC-P, hiPSC-F, or hESC. Color code indicates status of the colony. D20 and D30 positives showed dual staining for GCTM-2/EPCAM or TRA-1-60/CDH1; negative colonies lacked staining for either. All D10 colonies were negative for markers. The vertical axis shows clustering of colonies, and the horizontal axis shows clustering of genes. Colonies cluster into two main divisions, pluripotent and fibroblastic. Gene clusters 1 and 2 contain canonical... (Figure legend continues on next page.)
colony morphology and staining, with the other five outgrowths showing a partially reprogrammed identity. With 14/75 live-cell selected foci generating hiPSCs with morphology suggestive of a fully reprogrammed state, the selection-based derivation efficiency was 19.2%.

Several of these live-cell selected fully and partially reprogrammed colonies were propagated long-term and characterized by immunocytochemistry and embryoid body assays (supplemental online Figs. 5, 6). The live-cell selected, “fully” reprogrammed hiPSC line G4F demonstrates strong marker positivity for the pluripotency markers TRA-1-60, GCTM-2, GDF3, EPCAM, CDH1, SSEA-4, alkaline phosphatase, POU5F1, DNMT3B, NANOG, and SOX2 (supplemental online Fig. 5A–5H). The live-cell selected, “partially” reprogrammed hiPSC line G5D also demonstrates positivity for each of the markers in this study (i–p), but the surface markers show patchy staining across the colonies. The nuclear markers also show limited areas of positivity across each colony.

The pluripotency of the live-cell selected, fully reprogrammed hiPSC line G4F was confirmed by the ability to generate cells representative of the three primary germ layers in embryoid body assays (positive immunostaining for NESTIN, AFP, and SMA confirming the presence of ectoderm, endoderm, and mesoderm lineage cells, respectively) (supplemental online Fig. 6).

Live-Cell Selection With CDH3 (P-Cadherin) Identifies Colonies Expressing a Full Complement of Pluripotency Markers

CDH3 was among the genes expressed in common between the positive controls and late reprogrammed double-positive colonies. Expression of CDH3 was easily detected by immunocytochemical staining of fixed positive control cell lines (supplemental online Fig. 7A) or live-cell staining of reprogramming foci from day 20 onward (supplemental online Fig. 7B). We selected primary colonies using live-cell staining with an antibody against CDH3 combined with GCTM-2 at day 30 following transfection and analyzed gene expression in the positive versus negative colonies. The data were combined with the previous hierarchical cluster analysis from Figure 5 and are shown in Figure 6. Three clusters of colonies were identified: pluripotent, fibroblastic, and intermediate. It is clear that the GCTM-2/CDH3 double-positive colonies cluster with the positive controls. In particular, these double-positive colonies show high levels of expression of the late upregulated genes such as HAS3, END8B, SOX2, FOXD3, CDH1, EPHAl, LCK, and CDH3 (gene cluster 1). CDH3-positive colonies also express earlier pluripotency markers including NANOG, POU5F1, DNMT3B, LEFTY1, NODAL, EPCAM, and TGFβ1 (gene cluster 2) and show relatively consistent expression of markers associated with mesendoderm, including CER1, MIXL1, GSC, and T (gene cluster 3). By contrast, GCTM-2/EPCAM– or GCTM-2/CDH1-positive colonies at days 20 and 30 in the intermediate cluster showed less consistent expression of the genes in the late category.

Quantification of gene expression in selected colonies is shown in Figure 7. GCTM-2/CDH3 double-positive colonies live-cell selected at day 30 (GP D30) demonstrate expression of both early and late upregulated genes similar to that of the positive control hiPSC and hESC colonies. In contrast to GCTM-2/EPCAM and TRA-1-60/CDH1 colonies selected at day 30, GP D30 colonies had expression profiles more similar to the positive controls when comparing relative quantification values. The difference is most striking among the late upregulated gene set, which demonstrates a greater difference between GP D30 samples and those of all the remaining sample groups including the previous double-positive samples.

Live Colony Selection for CDH3 Staining Selects Fully Reprogrammed Cells

To determine whether gene expression in CDH3-positive colonies was predictive of their ability to yield permanent iPSC lines, and to determine whether CDH3 selection could be applied to other reprogramming methodologies, we compared selection using live-cell staining with CDH3 plus GCTM-2 to selection with EPCAM plus GCTM-2 following reprogramming with episomal vectors. Eight of eight CDH3/GCTM-2-positive colonies gave rise to permanent iPSC lines, whereas five of seven EPCAM/GCTM-2-positive colonies yielded permanent cell lines. All iPSC lines expressed pluripotent stem-cell surface markers and pluripotency associated genes and could differentiate into cells expressing markers of all three embryonic germ layers in embryoid body (supplemental online Fig. 8) and neurosphere assays, with the exception of one CDH3/GCTM-2-selected line. This line (CDH3-CLS) generated TUBB3+ neurons (ectoderm) in neurospheres but failed to form robust EBs.

DISCUSSION

The process of reprogramming to pluripotency by defined factors has been subject to detailed molecular analysis [19]. Many studies have performed sequential analyses of the transcriptome and epigenome in mouse fibroblast cells reprogrammed by inducible constructs that drive facile and tightly regulated expression of reprogramming factors. Using cell surface markers and reporter genes, earlier studies have identified early and late phases of the reprogramming process, marked by distinct patterns of epigenetic change. The early phase is associated with the mesenchymal to epithelial transition, followed by an intermediate stage of activation of a subset of pluripotency genes, then a late stabilization phase marked by the onset of SOX2 expression [7, 20]. In the mouse, downregulation of the fibroblast cell surface marker Thy-1, and upregulation of the pluripotent stem cell marker SSEA-1, identify early stages of reprogramming [21, 22]. Recent work on mouse cells shows that downregulation of CD44 and activation of ICAM1 mark a late stage in the reprogramming process [23]. In mouse or human studies, GDF3 and NANOG are activated relatively early, whereas the pluripotency factor SOX2 has been consistently identified as a factor whose expression is activated in the later stages of reprogramming.

In the human, most studies have focused on SSEA-4 and TRA-1-60 as a marker for the enrichment of reprogrammed cells [8, 9, 24]. The process of reprogramming to pluripotency by defined factors has been subject to detailed molecular analysis [19]. Many studies have performed sequential analyses of the transcriptome and epigenome in mouse fibroblast cells reprogrammed by inducible constructs that drive facile and tightly regulated expression of reprogramming factors. Using cell surface markers and reporter genes, earlier studies have identified early and late phases of the reprogramming process, marked by distinct patterns of epigenetic change. The early phase is associated with the mesenchymal to epithelial transition, followed by an intermediate stage of activation of a subset of pluripotency genes, then a late stabilization phase marked by the onset of SOX2 expression [7, 20]. In the mouse, downregulation of the fibroblast cell surface marker Thy-1, and upregulation of the pluripotent stem cell marker SSEA-1, identify early stages of reprogramming [21, 22]. Recent work on mouse cells shows that downregulation of CD44 and activation of ICAM1 mark a late stage in the reprogramming process [23]. In mouse or human studies, GDF3 and NANOG are activated relatively early, whereas the pluripotency factor SOX2 has been consistently identified as a factor whose expression is activated in the later stages of reprogramming.

In the human, most studies have focused on SSEA-4 and TRA-1-60 as a marker for the enrichment of reprogrammed cells [8, 9, 24].
24], more recently in conjunction with CD30 [10]. SSEA-4 is a cell surface glycolipid expressed on pluripotent stem cells and TRA-1-60 reacts with a glycolipid epitope present on large extracellular glycoprotein complex that contains the mucin-like protein podocalyxin [25]. CD30 is a member of the tumor necrosis factor receptor superfamily with very limited expression in adult tissues and whose expression in hiPSC is influenced by genetic and epigenetic alterations in the cells [26, 27]. These studies, like ours, consistently found that GDF3 and NANOG are activated early and that SOX2 expression occurs at a late stage of reprogramming.

We related expression of particular cell surface markers to activation of the pluripotency gene program during reprogramming of human fetal fibroblasts. We focused on the cell surface markers defined by monoclonal antibodies GCTM-2 and TRA-1-60, as well as

Figure 5. Summary of the percentage of colonies negative or positive for surface markers at 10, 20, and 30 days after reprogramming and expressing early upregulated genes (A) and late upregulated genes (B) compared with parental fibroblasts, or fully reprogrammed human induced pluripotent stem cells and human embryonic stem cells. Asterisks indicate statistically significant differences in the proportions of cells expressing the markers versus time. Numbers in parentheses refer to the number of biological replicates (colonies) analyzed in each group. Genes of interest were compared between groups by 2 × 2 contingency tables and Fisher’s exact test. Abbreviations: D, day; N, negative; P, positive.

©AlphaMed Press 2016
Figure 6. Heat map and unsupervised hierarchical cluster analysis showing gene expression in growing colonies that were positive for GCMT-2/CDH3 (P-CAD) at day 30 after gene transfection; positive for GCTM-2/EPCAM and TRA-1-60/CDH1 20 and 30 days after gene transfection (D30 POS or D20 POS); negative for GCMT-2/CDH3, GCTM-2/EPCAM, or TRA-1-60/CDH1 on day 30 after gene transfection (D30 NEG); negative for GCTM-2/EPCAM or TRA-1-60/CDH1 on day 10 (DAY 10 NEG) and day 20 (DAY 20 NEG) (data on colonies positive or negative for the latter two marker pairs are also shown in Figure 4): hESC; hiPSC-F; hiPSC-P; FIBROs. Data on plasmid used for reprogramming are also shown (PLASMID).

(Figure legend continues on next page.)
as EPCAM and CDH1. GCTM-2 and TRA-1-60 recognize related epitopes on a large cell surface glycoprotein complex of primate pluripotent stem cells [25]. EPCAM and CDH1 are well-characterized cell-cell adhesion molecules implicated in the maintenance of pluripotent stem cells in mice [28–30]. The genes we studied in our panel included known regulators of pluripotency as well as genes identified in our previous work as characteristic of a stem cell subpopulation with high self-renewal capacity, at the top of a hierarchy of cells showing a gradation in pluripotency gene expression [31, 32].

The reprogramming marker timeline experiment confirmed that intense staining for TRA-1-60 and EPCAM/CDH1 occurs only fairly late in the reprogramming process (TRA-1-60, human [9]; EPCAM and CDH1, mouse [28]). Likewise, the epitope defined by antibody GCTM-2 appears late during the reprogramming process.

The percentage of reprogramming foci exhibiting staining equivalent to positive controls reaches a maximum of approximately 10% for all surface markers by day 24. These data support the hypothesis that intense staining is representative of foci that are well on the pathway to reactivation of all the pluripotency networks necessary to reach a fully reprogrammed state. Expression of cell surface markers requires activation not only of the relevant transcriptional network but also appropriate posttranscriptional control and posttranslational modification.

Gene expression patterns can be related to cell surface marker staining within reprogramming foci. The early upregulated genes (POUSF1, NANOG, NODAL, CDH1, EPCAM, LEFT1, TDGF1, and DNMT3B) all have well known roles in the maintenance of pluripotency [29, 33–38]. These genes begin to be activated on approximately days 10–20 and are found in most foci by day 30, in contrast to the late upregulated group. All of the positive controls have high expression of these genes, but even at day 30, only some of the selected colonies exhibit expression of these genes. The activation of this group of genes, which includes FOXD3, CDH3, LCK, EPHA1, EDNRB, HAS3, and SOX2, marks a later stage in the reprogramming process and could thus be used to isolate colonies that have the highest probability of becoming fully reprogrammed hiPSCs.

Expression analysis of these genes indicates a highly statistically significant difference between day 30 double-positive foci (GCTM-2/EPCAM, TRA-1-60/CDH1) and the positive control fully reprogrammed hiPSCs and hESCs. More than half of the double-positive foci were negative for expression of late reprogramming genes, suggesting that they have not completed the reprogramming process at this point. It is possible that the foci lacking expression of these genes would later exhibit expression, but the absence of expression after 4–5 weeks of reprogramming indicates a prolonged process that could be abortive. The role of many of these genes in stem cell regulation is unknown, but FOXD3 has been reported to be important in the maintenance of pluripotency. SOX2 has a well-established role in pluripotency and also serves as a late marker of reprogramming in the mouse.

On the basis of results with GCTM-2, TRA-1-60, EPCAM, and CDH1, we identified CDH3 as a cell surface marker that could select cells that had fully activated the panel of genes in our study. Selection with CDH3 achieved this outcome, with almost all GCTM-2/CDH3-positive colonies at day 30 expressing levels of all pluripotency-related genes studied similar to positive controls. Consistent with this result, CDH3/GCTM-2 selection of colonies generated by transfection with episomal vectors encoding reprogramming factors gave rise to permanent iPSC lines with high frequency. Further comparative analysis with large-scale selection of reprogrammed foci will be required to determine whether selection with CDH3 provides a definitive advantage over other markers in the isolation of fully reprogrammed cells, but our data show clearly that CDH3-positive colonies have reached the end stage of the reprogramming process.

Recently Takahashi et al. [12] showed that during reprogramming of human fibroblasts, a subset of cells expressed markers of the primitive streak from approximately day 11 to day 28 after gene transfection. We too observed expression of MIXL1, T, CER1, and GSC in cell surface marker-positive colonies on days 20–30, although we saw a similar level of expression of these genes in hESC and fully reprogrammed iPSC in this study and in a single-cell analysis of hESC gene expression [13]. It is possible that expression of these mesendodermal genes in our system is indicative of a degree of spontaneous differentiation or lineage priming in fully reprogrammed pluripotent cells. Spontaneous differentiation or lineage priming may have been absent in the positive control iPSC of Takahashi et al. [12].

Cacchiarelli et al. [18] carried out extensive transcriptional and epigenomic analysis of the time course of reprogramming of human fibroblasts using an inducible transgene system. In their study, multidimensional scaling analysis distinguished SSEA-3-positive cells (appearing at day 10) from TRA-1-60-positive cells (appearing at day 20), and both of these categories from fully reprogrammed cells. Most of the reprogramming experiments in our study were carried out under xeno-free conditions using human fibroblast feeder cells. The studies using selection with CDH3 were carried out using either xeno-free conditions or using mouse embryo feeder cells and medium supplemented with conventional knockout serum replacer and FGF2. Cacchiarelli et al. [18] used mouse embryo feeder cells and medium supplemented with conventional knockout serum replacer and FGF2 but used an inducible system and immortalized fibroblasts. Because these various systems yielded similar time courses of induction of gene expression, it is likely that the classification of early and late genes here is intrinsic to the reprogramming process. The time course of expression of our late genes in the study of Cacchiarelli et al. [18] analysis suggests that despite the use of different modalities of reprogramming, our CDH3-positive population would likely equate to the end stage of their process.

(Figure legend continued from previous page.)

The vertical axis shows clustering of colonies, the horizontal axis, clustering of genes. Colonies fall into pluripotent, fibroblastic, or intermediate clusters. Gene cluster 1 is consistently expressed in CDH3/GCTM-2-positive colonies along with embryonic stem cells and fully reprogrammed iPSCs, but only in a subset of days 20 and GCTM-2/EPCAM- and TRA-1-60/CDH1-positive colonies, and is largely absent from fibroblasts or negative colonies. Gene cluster 2 contains canonical pluripotency genes that are expressed consistently in all colony types except fibroblasts or marker negative colonies. Gene cluster 3 contains mesendodermal genes and is expressed in a substantial fraction of CDH3/GCTM-2-positive colonies and days 20 and GCTM-2/EPCAM- and TRA-1-60/CDH1-positive colonies along with embryonic stem cells and fully reprogrammed induced pluripotent stem cells, but is absent from negative colonies and FiBROs. Cluster 1 genes, HAS3, ENDRB8, SOX2, FOXD3, CDH1, EPHA1, LCK, CHD3; cluster 2 genes, NANOG, POUSF1, LEFT1, NODAL, EPCAM, TDGF1; cluster 3 genes, CER1, GATA4, MIXL1, GSC, T. Color scale (top) shows JC1 values. Abbreviations: FiBRO, fibroblast; hESC, human embryonic stem cell; hiPSC-P, partially reprogrammed hiPSC; hiPSC-F, fully reprogrammed hiPSC; NEG, negative; POS, positive.
CONCLUSION

These results provide a means for discrimination and isolation of reprogrammed cell colonies at 20 to 30 days after transfection that may not have fully activated pluripotent stem cell gene expression (cells double-positive for GCTM-2/EPCAM or TRA-1-60/CDH1) from those that have (GCTM-2/CDH3-positive). Our results will facilitate a better understanding of the factors involved in the final stages of reprogramming of human cells to pluripotency, a critical rate-limiting phase of the process that is still poorly understood. Selection with CDH3 in combination with other markers might further enhance the yield of fully reprogrammed cells and could easily be incorporated into automated selection protocols for high throughput methodology.

ACKNOWLEDGMENTS

This work was supported by grants from the California Institute for Regenerative Medicine (New Cell Lines RL1-00667 and Basic Biology 1 RB1-01372), the Australian Research Council (Special Research Initiative in Stem Cell Sciences), and the National Stem Cell Foundation of Australia. K.C.D. is supported by a Gerard Crock Fellowship from the Eye Research Australia Foundation.

Figure 7. Summary of the relative quantification of gene expression in colonies that were positive (D20 and D30, day 20 and 30 GCTM-2/EPCAM- and TRA-1-60/CDH1-positive colonies; GP30, day 30 GCTM-2/CDH3-positive colonies); or negative (D20N, D30N day 20 and 30 GCTM-2/EPCAM- and TRA-1-60/CDH1-negative colonies; GP30N, day 30 GCTM-2/CDH3-negative colonies) for surface markers at 10, 20 and 30 days after reprogramming expressing early upregulated genes (A) and late upregulated genes (B) compared with parental fibroblasts, or fully reprogrammed hiPSCs and hESCs. Values are normalized to hESC. Abbreviations: D, day; hDFf, human dermal fetal fibroblasts; hESC, human embryonic stem cell; hiPSC, fully reprogrammed iPSC; N, negative; P, positive.
AUTHOR CONTRIBUTIONS

J.E.P.: conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript; J.C.D.: collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript; M.F.P.: conception and design, financial support, data analysis and interpretation, manuscript writing, final approval of manuscript.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors indicated no potential conflicts of interest.

REFERENCES

1. Okita K, Yamanaka S. Induction of pluripotency by defined factors. Exp Cell Res 2010;316: 2565–2570.
2. Malik N, Rao MS. A review of the methods for human iPSC derivation. Methods Mol Biol 2013;997:23–33.
3. Okano H, Nakamura M, Yoshida K et al. Steps toward safe cell therapy using induced pluripotent stem cells. Circ Res 2013;112:523–533.
4. Cao S, Loh K, Pei Y et al. Overcoming barriers to the clinical utilization of iPSCs: reprogramming efficiency, safety and quality. Protein Cell 2012;3:1334–1345.
5. Papp B, Plath K. Epigenetics of reprogramming to induced pluripotency. Cell 2013;152: 1324–1343.
6. Carey BW, Markoulaki S, Beard C et al. Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat Methods 2010; 7:56–59.
7. Patino JM, Andersen E, Walsh RM et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 2012;151:1617–1632.
8. Lowry WE, Richter L, Yachechko R et al. Generation of human induced pluripotent stem cells from dermal fibroblasts. Proc Natl Acad Sci USA 2008;105:2883–2888.
9. Chan EM, Ratnasiriwatwao S, Park IH et al. Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nat Biotechnol 2009;27:1033–1037.
10. Abujarour R, Valamehr B, Robinson M et al. Optimized surface markers for the prospective isolation of high-quality hiPSCs using flow cytometry selection. Sci Rep 2013;3:1179.
11. Tanabe K, Nakamura M, Narita M et al. Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc Natl Acad Sci USA 2013;110:12172–12179.
12. Takahashi K, Tanabe K, Ohnuki M et al. Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat Commun 2014;5:3678.
13. Hough SR, Thornton M, Mason E et al. Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human embryonic stem cells. Stem Cell Rep 2014;2:881–895.
14. Boheler KR, Bhattacharya S, Kropp EM et al. A human pluripotent stem cell surface N-glycoproteome resource reveals markers, extracellular epitopes, and drug targets. Stem Cell Rep 2014;3:185–203.
15. Koiie G, Ho M, Zhou Q et al. Identification of human embryonic stem cell surface markers by combined membrane-polyoxyme translation state array analysis and immunotranscriptional profiling. Stem Cells 2009;27:2446–2456.
16. Okita K, Matsumura Y, Sato Y et al. A more efficient method to generate integration-free human iPS cells. Nat Methods 2011;8:409–412.
17. Pomeroy J. Derivation and characterization of human embryonic stem (ES) cells and human induced pluripotent stem (iPS) cells in clinical grade conditions [Ph.D. thesis]. Los Angeles, CA: University of Southern California, 2012.
18. Cacchiarelli D, Trapnell C, Ziller MJ et al. Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell 2015;162:412–424.
19. Buganim Y, Faddah DA, Jaenisch R. Mechanisms and models of somatic cell reprogramming. Nat Rev Genet 2013;14:427–439.
20. Buganim Y, Faddah DA, Cheng AW et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 2012;150:1209–1222.
21. Bramblik T, Foreman R, Welstead GG et al. Single-cell expression profiling identifies key signature genes during mouse embryonic stem cell reprogramming. Stem Cell Rep 2014;2:230–240.
22. Stadtfeld M, Maherali N, Breault DT et al. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell 2008;2:230–240.
23. O’Malley J, Skyllaki S, Iwabuchi KA et al. High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature 2013;499:88–91.
24. Hockemeyer D, Soldner F, Cook EG et al. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell 2008;3:346–353.
25. Wright AJ, Andrews PW. Surface marker antigens in the characterization of human embryonic stem cells. Stem Cell Res (Amst) 2009;3:3–11.
26. Chung TL, Turner JP, Thaker NV et al. Ascorbate promotes epigenetic activation of CD30 in human embryonic stem cells. Stem Cells 2010;28:1782–1793.
27. Hersfeld D, Woltveten E, Langton-Bunker E et al. CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. Nat Biotechnol 2006;24:351–357.
28. Chen HF, Chuang CY, Lee WC et al. Surface marker epithelial cell adhesion molecule and E-cadherin facilitate the identification and selection of induced pluripotent stem cells. Stem Cell Rev 2011;7:722–735.
29. Lu TY, Lu RM, Liao MY et al. Epithelial cell adhesion molecule regulation is associated with the maintenance of the undifferentiated phenotype of human embryonic stem cells. J Biol Chem 2010;285:8719–8730.
30. Li J, Bennett SA, Wang L. Role of E-cadherin and other cell adhesion molecules in survival and differentiation of human pluripotent stem cells. Cell Adhes Migr 2012;6:59–70.
31. Hough SR, Laslett AL, Grimmond SB et al. A continuum of cell states spans pluripotency and lineage commitment in human embryonic stem cells. PLoS One 2009;4:e7708.
32. Laslett AL, Grimmond S, Gardiner B et al. Transcriptional analysis of early lineage commitment in human embryonic stem cells. BMC Dev Biol 2007;7:12.
33. James D, Levine AJ, Besser D et al. TGFBeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 2005;132:1273–1282.
34. Chambers I, Colby D, Robertson M et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 2003;113:643–655.
35. Redmer T, Diecke S, Grigoryan T et al. E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO Rep 2011;12:720–726.
36. Vallier L, Alexander M, Pedersen RA. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci 2005;118:4495–4509.
37. Li JY, Pu MT, Hirasawa R et al. Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Mol Cell Biol 2007;27:8748–8759.
38. Nichols J, Zevnik B, Anastassiadis K et al. Formation of pluripotent stem cells in the mammalian embryo depends on the Pou transcription factor Oct4. Cell 1998;95:379–391.

See www.StemCellsTM.com for supporting information available online.