A Cell Surfaceome Map for Immunophenotyping and Sorting Pluripotent Stem Cells

Rebekah L. Gundry†§¶, Daniel R. Riordon¶, Yelena Tarasova¶, Sandra Chuppa‡§, Subarna Bhattacharya‡§, Ondrej Juhasz¶, Olena Wiedemeier‡§, Samuel Milanovich**¶, Fallon K. Noto‡‡, Irina Tchernyshyov¶, Kimberly Raginski¶, Damaris Bausch-Fluck§§, Hyun-Jin Tae¶, Shannon Marshall¶, Stephen A. Duncan‡‡, Bernd Wollscheid§§, Robert P. Wersto¶, Sridhar Rao**‡‡, Jennifer E. Van Eyk¶¶, and Kenneth R. Boheler¶¶¶

Induction of a pluripotent state in somatic cells through nuclear reprogramming has ushered in a new era of regenerative medicine. Heterogeneity and varied differentiation potentials among induced pluripotent stem cell (iPSC) lines are, however, complicating factors that limit their usefulness for disease modeling, drug discovery, and patient therapies. Thus, there is an urgent need to develop nonmutagenic rapid throughput methods capable of distinguishing among putative iPSC lines of variable quality. To address this issue, we have applied a highly specific chemoproteomic targeting strategy for de novo discovery of cell surface N-glycoproteins to increase the knowledge-base of surface exposed proteins and accessible epitopes of pluripotent stem cells. We report the identification of 500 cell surface proteins on four embryonic stem cell and iPSCs lines and demonstrate the biological significance of this resource on mouse fibroblasts containing an oct4-GFP expression cassette that is active in reprogrammed cells. These results together with immunophenotyping, cell sorting, and functional analyses demonstrate that these newly identified surface marker panels are useful for isolating iPSCs from heterogeneous reprogrammed cultures and for isolating functionally distinct stem cell subpopulations. Molecular & Cellular Proteomics 11: 10.1074/mcp.M112.018135, 303–316, 2012.

Stem cells retain the abilities of self-renewing division and differentiation into one or many specialized cell types. Pluripotent stem cells (PSCs)1, in particular, differentiate to all cell types of an embryo proper and may serve as an inexhaustible source of cell progeny useful for regenerative medicine. The best characterized and accepted standard for PSCs are embryonic stem cells (ESCs) derived from the inner cell mass of the mammalian blastocyst (reviewed in (1)). Alternatively, experimentally derived PSCs, known as induced PSCs (iPSCs), are generated in vitro from somatic cells through forced expression of pluripotency-promoting transcription factors such as OCT4, SOX2, KLF4, and c-MYC (2–4). Although ESC lines demonstrate differences among clones and in germline transmission, iPSC lines show much greater interline variation, ranging from variable quantities of essential transcription factors, residual epigenetic memory, sporadic point mutations, differential DNA methylation patterns, variable degrees of tumorogenicity, and in mouse, differential chimerism and low germ-line transmission (5–13). Importantly, manual selection of putative iPSC colonies is routinely based on morphology followed by indirect methodologies analyzing intracellular transcription factors, selected signaling molecules, and epigenetic states to establish stem cell identity and function. The final determination of the pluripotent phenotype ultimately relies on chimera formation with germline transmission (mouse) and teratoma formation (mouse and human). Although guidelines have been proposed for the derivation and characterization of PSCs (14, 15), no system is available to characterize PSCs analogous to hematopoietic stem cell (HSC) immunophenotyping where cell surface proteins or epitopes serve as surrogate markers of a cell’s phenotype to define potency (CD34/
CD133 or c-KIT (CD117), function (ALDH enzyme activity), or drug efflux (SP cell analysis).

Although molecular approaches employing expressed fluorescent or tagged proteins are experimentally valuable for analyzing PSC populations, immunophenotyping is vector-independent, nonmutagenic, and can be applied broadly in both clinical and experimental settings. This approach relies principally on antibodies against cluster-of-differentiation (CD) molecules, and it is routinely employed in clinical hematology to isolate subsets of bone marrow-derived HSCs, and myeloid and lymphoid progeny for therapeutic interventions and quantitative assessments (16, 17). Although markers like stage specific embryonic antigen-1 (SSEA-1) for mouse (18) and SSEA-3 and SSEA-4 in human aid in the identification of PSCs, very few known surface markers and corresponding application-specific antibodies are specific for the pluripotent state. Sorted SSEA-1 mouse ESC (mESC) populations are at best heterogeneous (19, 20), and sorted Thy1−SSEA-1+ cells only partially enrich for mouse fibroblasts poised to become iPSCs (21). The Tra-1−81 surface marker also allows for the identification of human iPSC colonies (22), but like SSEA-3, −4, and Tra-1−80, it is not specific to the undifferentiated state (23, 24). The fundamental lack of cell surface markers for isolating homogeneous populations of PSCs analogous to that described for HSCs significantly restricts the clinical implementation of iPSCs for regenerative medicine.

Several experimental approaches are available to identify cell surface proteins (selected reviews (25–27)), but most are either constrained by the limited availability of antibodies or are inefficient for unambiguous identification of cell surface proteins. Chemical tagging and/or plasma membrane (PM) enrichment based strategies have partially evaluated the cell surface proteome of mouse and human PSCs (28–36); however, these studies did not confirm the utility of these identified surface proteins to functionally define the pluripotent phenotype. Except for one publication (29), these reports relied principally on published data, publicly available database annotations, or immunological-based methods to predict or indicate the subcellular localization of putative surface proteins. Consequently, targeted analytical approaches that experimentally verify extracellular domains in an antibody-independent manner will be advantageous for more rapidly defining the PSC surface landscape and accelerating the development of new and informative stem cell surface markers.

Here we have employed discovery-driven Cell Surface Capturing (CSC Technology) and mass spectrometry to selectively label, capture, and identify extracellular-exposed cell surface N-glycoproteins from PSCs. The CSC Technology data were integrated with cell imaging, flow cytometry, and protein and RNA quantitation to generate a cell surface proteome (surfaceome) resource for defining the state of pluripotency. The value of this new resource is exemplified in several applications where surface protein markers proved useful for identifying novel ESC subpopulations and for isolating newly generated iPSCs from heterogeneous cultures of reprogrammed cells.

**EXPERIMENTAL PROCEDURES**

**Cell Culture and iPSC Generation—**Mouse pluripotent stem cell lines (R1, D3, 2D4, TTF1 lines) were maintained on mitotically inactivated mouse embryonic fibroblasts (MEFs) prior to cultivation under feeder free conditions as described (37). After passaging onto gelatin, mESC colony morphology was compact and characterized by clusters of small cells with a high nuclear to cytoplasmic ratio as expected (1). All cell lines contained high levels of OCT4 and SOX2, were positive for alkaline phosphatase activity (not shown), and expressed the cell surface antigen SSEA-1. For differentiation, mESCs were cultured without LIF for 7 days. MEF reprogramming experiments were performed similar to (2) and early passage (3, 4) MEFs were obtained from commercially available transgenic mice expressing EGFP under the control of the Pous5f1 (Oct4) promoter and distal enhancer (Jackson Labs, West Grove, PA; Stock # 004654). 8 × 10^5 MEFs were plated onto a 10-cm plate with mitotically inactivated MEFs, and infected twice over 2 days with ecotropic retroviruses (pMX-Oct4, pMX-Klf4, pMX-Sox2, pMX T58A, available from ADDGENE) in the presence of 4 μg/ml polybrene. Twenty-four hours after the second round of infection, media was changed to standard ESC media with two exceptions: KO serum (Invitrogen, Carlsbad, CA) in the place of FCS, and 2 × LIF (2 × 10^3 U/ml) were used. Ecotropic retroviruses were generated by transiently transfecting HEK 293T cells with the viral vectors and a packaging plasmid (EcoPack), with virus collection 48 and 72 h after the initial transfection. Viruses were collected, filtered to remove any contaminating cells with a 0.45 μm filter, and mixed in a 1:1:1:1 ratio prior to infection.

**Cell Surface Capturing Technology (CSC Technology)—**Approximately 1E8 cells per biological replicate (n = 3) of each established cell line (R1, D3, 2D4, TTF1) were taken through the CSC Technology workflow as reported previously (38, 39) with slight modifications. Undifferentiated ESCs were allowed to detach for 30 min at 4 °C in enzyme-free cell dissociation solution (Millipore, Billerica, MA). To ensure that proteins observed on the cell surface after using the enzyme-free cell dissociation solution were not a result of exposing the cells to this solution, immunoblotting was used to validate that samples obtained by scraping versus the enzyme-free cell dissociation solution resulted in similar levels of both pluripotency markers as well as cell surface proteins (data not shown). Cells were washed with phosphate-buffered saline (PBS) pH 7.4 followed by treatment for 15 min in 1 mM sodium metaperiodate (Pierce, Rockford, IL) in phosphate-buffered saline (PBS) pH 7.4 at 4 °C followed by 2.5 mg/ml biocytin hydrazide (Biotium, Hayward, CA) in PBS pH 6.5 for 1 h at 4 °C. Cells were homogenized in 0.5% Triton X-100 in 100 mM NH4HCO3,5% streptavidin PLUS (Pierce, Rockford, IL) for 1 h at 25 °C. Beads were sequenced for 300 μl 100 mM NH4HCO3, 5 μM Tris (2-carboxyethyl) phosphine (Sigma, St. Louis, MO), and 0.1% (v/v) Ripigest (Waters, Milford, MA) with continuous vortexing and proteins were allowed to reduce for 10 min at 25 °C followed by alklylation with 10 mM iodoacetamide for 30 min. The sample was incubated with 1 μg glycerol-free endoproteinase Lys-C (Calbiochem, San Diego, CA) at 37 °C for 6 h followed by 20 μg proteomics grade trypsin (Promega, Madison, WI) at 37 °C for 16 h at 4 °C to collect the membranes. The supernatant was removed and the membrane protein pellet was resuspended in 300 μl 100 mM NH4HCO3, 5 mM Tris (2-carboxyethyl) phosphine (Sigma, St. Louis, MO), and 0.1% (v/v) Rapigest (Waters, Milford, MA) with continuous vortexing and proteins were allowed to reduce for 10 min at 25 °C followed by alklylation with 10 mM iodoacetamide for 30 min. The sample was incubated with 500 μl bead slurry of UltraLink Immobilized Streptavidin PLUS (Pierce, Rockford, IL) for 1 h at 25 °C. Beads were sequentially washed with 0.5% Triton X-100 in 100 mM NH4HCO3, 5
100% homology) to only the protein reported. Membrane topology peptide sequence for any “single peptide identification” was assays (Figs. 3, 4). To ensure specificity for the proteins reported, the EFNA2, each of which provided positive data in antibody-based data for these identifications. Examples of proteins identified by a false discovery rate (FDR)

Denmark) and filtered to contain only proteins corresponding to a files were input into ProteinCenter (Proxeon Bioinformatics, Odense, each biological replicate. The ProteinProphet interact-prot.xml result
tions. Peptide mass tolerance was set to 50 ppm, fragment mass
tation (Cys), and deamidation (Asn) allowed as differential modifica-
trypsin (after Lys or Arg) with up to two missed cleavages; monoiso-
was originally attached to a biotin-labeled oligosaccharide structure,
glycosylation site is within a tryptic peptide of suitable

protein "contaminants " derived from intracellular membranes (i.e.

glycosylation consensus amino acid sequence motif (Nxs/T) (Fig. 1C). These criteria allowed for the efficient filtering of protein “contaminants ” derived from intracellular membranes (i.e. did not contain “tag”) and determination of the orientation of transmembrane proteins within the membrane. In this type of sample processing, the number of peptides identified for each protein is dependent upon the number of occupied glycosylation sites and whether the glycosylation site is within a tryptic peptide of suitable m/z for MS analysis. For this reason, proteins identified by a single peptide (either by a single or multiple spectra) were not automatically excluded; rather, care was taken to appropriately and manually evaluate the data for these identifications. Examples of proteins identified by a single peptide sequence include PRSS50 (Tsp50), CDH11, and EFNA2, each of which provided positive data in antibody-based assays (Figs. 3, 4). To ensure specificity for the proteins reported, the peptide sequence for any “single peptide identification” was searched (via BLAST) against NCBI to ensure that it mapped (with 100% homology) to only the protein reported. Membrane topology predictions were obtained from TMAP (42), which is integrated into ProteinCenter. GO term annotations are also integrated into ProteinCenter. MS data are publicly available in PRIDE (http://www.ebi.ac.uk/pride/), accession numbers 16446–47 and all protein and peptide information is provided in the supplemental Table S1.

Immuno-based Assays and qRT-PCR—Antibodies and experimental conditions used for immunoblotting and flow cytometry are provided in the supplemental Table S4. qRT-PCR experiments were performed as previously described (43).

Flow Cytometry—For flow cytometry, cell dissociation solution was used for cell detachment as described above. 1 × 10⁶ cells were washed twice with wash buffer (cold 1.0% fetal bovine serum in PBS) and were blocked with 15 μl 10% goat serum (Invitrogen) for 15 min at 4 °C. Primary antibody was allowed to bind for 45 min at 4 °C and subsequently cells were washed twice with wash buffer. Isotype controls were used at the same concentration as the primary antibody. Where necessary, secondary antibody was allowed to bind for 30 min at 4 °C and cells were washed twice with wash buffer. For co-staining experiments, live cells were stained with antibodies against cell surface proteins followed by fixation for 15 min at 4 °C (BD Cytofix, BD Biosciences, San Jose, CA), permeabilization (BD Phosflow Perm Buffer III, BD Biosciences) for 30 min at 4 °C and staining with antibodies against intracellular proteins. For analysis (without sorting), cells were resuspended in 1.0% fetal bovine serum in hank’s buffered saline solution (HBSS), or 50% fetal bovine serum in HBSS (sorting experiments), and filtered prior to analysis on a BD FACSCanto II flow cytometer (BD Biosciences) and 50,000 events were acquired.

For cell sorting of ESCs, live cells were immunolabeled with either anti-SSEA-1 or anti-Cd130, and following a sterile cell sort on a Beckman Coulter MoFlo High Speed Sorter, 800,000 cells/plate were plated onto feeder layers. Colony counts of plated cells represent an average of 10–25 unique locations imaged across the plate observed. Where larger cell numbers were required for subsequent analytical analyses (immunoblotting or flow cytometry), cells were passaged two to five times at mitotically inactivated MEFs after sorting. For co-staining, live cells were incubated with anti-Cd130 and then fixed (BD Cytofix fixation buffer, 10 min, 4 °C), permeabilized (BD Phosflow Perm Buffer, 30 min, 4 °C), and incubated with an anti-OCT4 antibody. For cell sorting of reprogrammed MEFs, cells were labeled with antibodies to cell surface proteins as above, and sorted on a BD FACSAria Ilu. Cells (~30,000) were collected directly for qRT-PCR or were plated (~50,000) onto MEFs for colony formation assays. For all flow cytometry experiments, data were analyzed using FCExpress V3 (DeNovo Software, Los Angeles, CA). The percent positive cells reported are based on gated cells with background contribution of 2% and an average of at least three biological replicates.

Teratoma Assay—Cd130+ and Cd130− expressing cells were sorted (n = 2 biological replicates) by flow cytometry to high purity and 2 × 10⁶ cells injected into the left flanks of allogeneic 129sv mice (n = 2 mice per population). Mice were sacrificed after 6 weeks and tumors examined by histological methods. All animal work was conducted according to relevant national and international guidelines.

RESULTS

Pluripotent Cell Surface N-glycoproteome—We employed the CSC Technology (Fig. 1A) to define the average N-glycoprotein surfaceome among four genetically distinct pluripotent cell lines from mouse (mESC: D3, R1; and mPSC: 2D4, TTF1). As shown in Fig. 1B, biotin labeling of surface glycoproteins on PSCs was specific and all reported proteins detected by this technology were identified by peptides containing a chemical deamidation (result of PNGaseF) at the
consensus amino acid sequence motif (NxS/T) for N-glycosyla-
tion. A representative spectrum illustrating the data “tag,”
which provides experimental evidence that the protein origi-
nated from the cell surface, is shown in Fig. 1C. Using CSC
Technology, 500 nonredundant N-glycoproteins were identi-
fied within the cell surface location (supplemental Table S1a,
Fig. 2) including predicted transmembrane (89%), GPI-linked
(2%), and extracellular matrix proteins (3%). Compared with
published proteomic studies of PSCs (i.e. 34 studies of mouse
and human PSC lines; herein referred to as the “PSC pro-
teome” (Gundry et al., 2011), 187 proteins (37% of the current
data) have not been reported among the 7487 proteins pre-
viously described in mouse PSCs (Fig. 1D), and 128 proteins
(25% of the current data) were identified in only a single other
study. For 175 proteins, UniProt annotations for experimental
data do not exist at the protein level in mouse (i.e. either
predicted, inferred by homology, or evidence at transcript
level; supplemental Table S1). A complete catalog of extra-
cellular glycopeptides and glycosites is listed in supplemental
Table S1b-e. This list contains more than 900 sites of
N-glycosylation, including many never previously identified exper-
imentally. These data illustrate the ability of the chemopro-
teomic CSC Technology to efficiently access a subproteome
that is under-sampled by other proteomic methods and to
confirm the extracellular domains of transmembrane proteins
not yet crystallized.

Of the 443 surfaceome proteins identified that are predicted
transmembrane proteins based on the TMAP algorithm (42),
gene ontology (GO) annotations reveal that 386 (77%) proteins are currently classified as “cell surface,” “plasma membrane,” or “extracellular” proteins, and 63 are annotated with the generic “membrane” term or are not annotated for localization (supplemental Table S1a). Only 166 proteins are annotated in UniProt with topology predictions supported by prior experimental evidence (supplemental Table S1a). Thus, for the remaining 277 predicted transmembrane proteins, this resource provides new evidence regarding transmembrane orientation and cell surface accessibility, and, for the 63 “generically” annotated proteins, the data should result in their reclassification to “plasma membrane” or “cell surface.” Additionally, 10% of the proteins identified here have no GO term associated with the PM. Some of these may represent intracellular “contaminants,” but others correspond to proteins with incomplete GO annotations. For example, testes-specific protease 50 (PRSS50), which is annotated as localized to the endoplasmic reticulum, was found to be present on the cell surface in mouse PSCs by the CSC-Technology. Our resource thus expands the known information about this biologically important but technically challenging subproteome and allows for identification of surface proteins and their accessible epitopes that can be targeted for development of new reagents for immunophenotyping live cells.

Surfaceome Comparisons Reveal Putative PSC-Restricted Markers—To identify putative “PSC-restricted” cell surface proteins, ProteinCenter software was used to compare the current surfaceome with data from the Cell Surface Protein Atlas (D. Bausch-Fluck, manuscript in preparation), consisting of CSC Technology data from over 80 human and mouse primary cells and cell lines (details in supplemental Table S2). The comparison of our data to 14 non-PSC mouse cell types within the Atlas yielded 127 proteins which were absent in non-PSC mouse cells (Fig. 2, supplemental Table S1). Of these 127 “PSC-restricted” proteins, 89 were either absent or observed in only a single other proteomic study within the previous PSC proteome (supplemental Table S2). Of the 53 “PSC-restricted” proteins not identified by other proteomic studies, 22 were statistically expressed below the detection limit in a microarray analysis of R1 ESCs (supplemental Table S3, Zhan et al., in preparation). Further analysis of publicly available gene expression profile data, (http://www.ebi.ac.uk/gxa/experiment/E-AFMX-4) revealed that 33 of 127 “PSC-restricted” proteins showed high levels of transcripts in fetal/embryonic tissues or were quantitatively more abundant in developmentally related tissues like uterus, testes, umbilical cord and placenta (supplemental Table S2). Altogether, the PSC-restricted proteins newly identified in this study represent a novel set of putatively informative proteins for PSCs.

To verify the utility of our surfaceome resource and the relevance of these PSC-restricted markers for characterizing PSCs, proteins not observed in non-PSC cell types within the Atlas and for which antibodies were available were selected for immunocytochemistry. Standard intracellular markers of pluripotency (OCT4 and NANOG) were used for comparison.
Fig. 3 shows the informative data obtained for Cadherin-11 (CDH11), Ephrin-A2 (EFNA2), Glypican-3 (GPC3), and Testis-specific protein 50 (PRSS50). With the exception of PRSS50, public transcriptome data show each are quantitatively higher in tissues associated with development, though neither Cdh11 nor Efn2a transcripts were present in our R1 ESCs at levels above background. By immunostaining, CDH11 and GPC3 were weakly to strongly detectable on the surface of cells throughout the colony. EFNA2 was more prevalent on the edges of the colonies; however, among cells found in the interior of the colony, EFNA2 was only observed intracellularly as focal points of positive staining. PRSS50, a protein previously annotated as localized to the endoplasmic reticulum but shown by the CSC-Technology to be localized to the cell surface, was also prominent on the surface of cells located at the edge of PSC colonies. However, cells in the interior of the colony either poorly expressed or did not express this surface protein. Consistently, by flow cytometry, PRSS50 was detected at similar levels in ~25% of ESCs and iPSCs, but it was not detectable in MEFs (Fig. 3B). These data suggested that PRSS50 could be either a potentially useful marker for PSCs or a marker for a distinct sub-population of PSCs; however, in flow experiments, the polyclonal antibody against PRSS50 under all conditions tested had significant overlap with the isotype control (Fig. 3B). Sorting experiments were therefore not pursued further. Similarly, the polyclonal antibodies to CDH11, EFNA2, and GPC3 when compared with isotype controls proved unsuitable for flow cytometry. Despite these polyclonal antibody-dependent limitations, the immunostaining data demonstrate that some surface proteins are informative of distinct cell populations within PSC colonies, and that some of the proteins may be dynamically cycled to the surface depending on its location within PSC colonies. Moreover, the development of monoclonal antibodies may allow future testing of these markers for PSC specificity in the future. Examples of CD molecules identified by the CSC Technology in mouse PSCs that were also detected on human H9 ESCs by flow cytometry are shown in Fig. 4C.

The Pluripotent CD Molecule “Barcode” Reveals Distinctions Among Cell Types Useful For Sorting New iPSC Cells—Although the future impact of this resource is expected to derive from the many newly identified proteins (i.e. uncharacterized, novel markers), monoclonal antibodies suitable for live cell sorting were mostly available to CD antigens. Thus, proof-of-concept studies to show that proteins identified via the CSC-Technology can serve as surrogate markers for phe-
notypically distinct PSC populations focused on CD molecules. In this study, 90 CD annotated glycoproteins were identified via the CSC Technology, including 27 not previously reported in the PSC proteome. When combined with published proteomic data, our CSC Technology data revealed a qualitative “working” CD molecule profile for mouse PSCs (Fig. 4A). Included in this list are stem cell-associated molecules, like CD31 (PECAM-1), CD118 (LIFR), CD90 (THY1), CD117 (C-KIT), and CD202b (TIE2); and CD molecules not typically associated with undifferentiated mESCs, including CD13 (aminopeptidase N), CD56 (NCAM), CD140a (PDGF receptor α), CD140b (PDGF receptor β), CD282 (TLR2), CD339 (Jagged-1), and CD320. To determine any quantitative differences between ESCs and iPSCs, as well as distinctions between iPSCs and their somatic precursors (MEFs), flow cytometry was used in conjunction with monoclonal antibodies to 11 surface proteins, including three not previously identified in the published PSC proteome (CD19, CD123, CD172a; Fig. 4B). Flow cytometry analysis revealed that CD71, CD90, CD98, CD130, and CD172a were positive on all three cell types, and, excluding CD130, were not examined further. Most interestingly, CD19, CD31, CD49f, CD123, and CD326 were present in both PSC populations and absent in MEFs and were thus considered potentially useful for distinguishing reprogrammed cells from their somatic precursors.

Based on these data that showed quantitative differences between MEFs and PSCs, antibody pairs were chosen for cell sorting of iPSCs from heterogeneous cultures of reprogrammed fibroblasts and included antibodies that showed the greatest quantitative difference among iPSCs and MEFs (Fig. 4B). In a pilot experiment, antibodies to CD326, CD112, and CD31 were used to sort live cells using an established iPSC line (referred to as iPS-13; (44)) to determine whether any of these markers could sort iPSCs at a higher efficiency than the current gold-standard surface marker SSEA-1. Sorted populations of CD326pos/CD31pos and CD112hi/CD31pos demonstrated empirically higher plating efficiencies than SSEA-1pos and excellent colony morphology (data not shown). Having established the possibility of sorting viable iPSCs using surface markers other than SSEA-1, we therefore tested whether surface protein markers could successfully sort newly generated iPSCs from a heterogeneous culture of reprogrammed cells without manual preselection of colonies. To do this, iPSCs were generated from oct4-EGFP MEFs (45, 46). 16 to 18 days after the start of reprogramming, EGFP+ colonies were observed and cells were sorted using antibodies against four sets of surface markers: SSEA-1, CD49f/CD31, CD112/
CD31, and CD326/CD31. EGFP expression was monitored but not used for sorting and cells sorted based on antibody selection were collected directly for qRT-PCR analyses or plated onto MEFs for colony formation assays. All data were compared with those sorted by the current gold standard surface marker for mouse, SSEA-1. In agreement with the earlier experiments, the results revealed that live EGFP+ cells could be isolated by FACS from the heterogeneous mixture of reprogrammed cells using antibodies against only these surface proteins (Fig. 5A). Each of the three selected surface marker pairs provided a higher selectivity for EGFP+ cells than SSEA-1 alone (Fig. 5A) and all successfully gave rise to EGFP+ colonies with high quality colony morphology (supplemental Fig. S1). qRT-PCR analysis of the sorted populations revealed quantitative differences in pluripotency markers Nanog, Oct4, Sall4, and Rex1 (Fig. 5B) among the populations. CD326/CD31 demonstrated the highest levels of each pluripotency marker among all the populations examined, but CD49fpos/CD31pos and CD112pos/CD31pos were similarly higher than SSEA1pos for all markers, with the exception of Nanog in the CD112pos/CD31pos. Based on these data, both CD326/CD31 and CD49f/CD31 sorted cells yielded iPSC colonies at an improved efficiency to those isolated with SSEA-1 alone. The antibody combinations tested here are not intended to be an exhaustive examination of all possible marker panels, but rather to provide proof-of-concept data that it will be possible to develop surface marker panels for sorting more homogeneous populations of newly reprogrammed iPSCs based on surfaceome maps from proteomic studies.

**CD130 Subpopulations**—During the flow cytometry analyses of mouse PSCs, an unexpected bimodal distribution was observed using a CD130 antibody (Figs. 6A, 4B). This distribution (designated CD130hi, CD130lo) suggested the presence of two phenotypically distinct populations in undifferentiated cells. This phenotypic heterogeneity was further supported by the observed changes in CD130 profiles in mESCs following LIF withdrawal (Fig. 6B). Undifferentiated mESCs had a high percentage of CD130lo cells (60%, n = 4), which shifted to a CD130intermediate and/or to a CD130hi population with differentiation by LIF withdrawal (Fig. 6B). The CD130 bi-modal distribution described here was dynamic in vitro. Re-plating of either CD130lo or CD130hi sorted cells led to the re-establishment of a bimodal profile within one passage (Fig. 6H). There was, however, a prevalence of CD130lo cells expanded from CD130hi sorted cells, and conversely a higher percentage of CD130hi cells expanded from the CD130lo sorted population. Following sorting to isolate specific cell populations identifiable by SSEA-1 or CD130, the plating efficiency and survival of SSEA1pos and CD130hi cells was significantly higher than either the SSEA1neg or CD130lo cells. Altogether ~12-fold more ESC colonies were visible on the plates containing SSEA1pos and ~sixfold more from the CD130lo cells compared with SSEA1neg and CD130hi cells, respectively (Fig. 6C). Although high quality colony morphology was observed among the surviving colonies of CD130lo cells (Fig. 6D), some colonies had poorly defined edges and exhibited evidence of differentiation, similar to that observed for sorted SSEA1pos cells (19). Empirically, the incidence of poor colony morphology was observed at a higher frequency for sorted SSEA1pos cells than the CD130lo cells.

These data indicate that CD130, akin to SSEA-1, can be employed to characterize populations of mESCs, but subsequent experiments demonstrated that CD130lo and CD130hi are unique subpopulations. In co-staining experiments with an intracellular marker of pluripotency (OCT4), CD130lo cells contained a higher intensity signal for OCT4 and a higher percentage of OCT4pos cells than the CD130hi population (Fig. 6E: 93% ± 0.3% of CD130lo compared with 53% ± 1.0% of CD130hi are OCT4pos). CD130lo sorted cells were strongly positive for surface pluripotency markers SSEA-1 and PECAM-1 (Fig. 6F). Quantitative immunoblotting of cells sorted by CD130lo or SSEA-1pos revealed that the CD130lo population contained significantly higher total protein levels of pluripotency markers NANOG, SOX2, and REX1 (47) than the SSEA1pos population (Fig. 6G). Finally, teratoma assays revealed that sorted CD130lo, but not sorted CD130hi, led to robust formation of tumors containing multiple cell types from all three germ layers (Fig. 6I). Thus, CD130 is an accessible cell surface protein that can be employed as a marker to distinguish between two novel subpopulations of mESCs and provides another example of the utility of combining surfaceome discovery studies with flow cytometry for more in-depth characterization of the heterogeneity within PSC cultures.

**DISCUSSION**

The present study begins to address the fundamental lack of cell surface markers for isolating more homogeneous populations of PSCs and their derivatives. Using the CSC-Technology, we have focused on the pluripotent state and identified 500 experimentally validated cell surface glycoproteins. Through bioinformatic comparisons and use of standard immunophenotyping techniques, we have been able to identify surface proteins that are informative of PSC subpopulations and antibody panels that are useful for the isolation of iPSCs from mixed populations of reprogrammed oct4-EGFP MEFs. In these sorting experiments, EGFP was not employed for sorting, but instead was used to monitor the efficacy of the antibodies used to isolate reprogrammed MEFs. Because the sorted and plated cells had colony characteristics of PSCs and expressed markers of pluripotency at levels equal to or above those isolated with the mouse “gold-standard” SSEA-1, we conclude that the cells isolated by flow cytometry and independent of cell morphology, are in fact iPSCs. We cannot yet fully comment on the genetic or epigenetic state of these isolated cells, but these proof-of-principal studies serve as the basis for future studies designed to identify surface proteins as surrogate markers of pluripotency or differentiated phenotypes.
**Fig. 5. Sorting reprogrammed OCT4-EGFP fibroblasts using combinations of CD molecules.** A. Schematic workflow showing sorting of reprogrammed MEFs. Representative images showing culture heterogeneity after MEF reprogramming and prior to FACS (top). Size marker = 500 μm. Cells were sorted by four marker sets: SSEA-1^pos^, CD49f^pos^/CD31^pos^, CD112^hi^/CD31^pos^, and CD326^pos^/CD31^pos^. On each dot blot, EGFP^+^ (green) and EGFP^+^ cells (red) are indicated and the average %EGFP^+^ and standard error of the mean within each gated population is shown (middle). The thresholds for the sorting gates were established by first analyzing isotype controls for the surface markers (i.e. not based on EGFP). B. qRT-PCR analysis of pluripotency markers in each population of cells sorted by the surface markers.
The surfaceome resource described here expands the known cell surface accessible proteome of mouse PSCs and establishes the experimental framework for developing accessible surface molecule barcodes for isolating more homogeneous, better defined, and therapeutically relevant populations of human PSCs for regenerative medicine. Although

![Fig. 6. CD130 subpopulations within mESCs. A, Histogram and contour plot illustrating two subpopulations of CD130 positive cells in undifferentiated R1 mESCs. B, Histograms of CD130 populations in D3 mESCs with differentiation via LIF withdrawal. C, Distribution of ratios of the number of colonies observed after plating sorted CD130 and SSEA-1 populations. Chart shows ratios of the number of colonies observed for 3 and 2 replicate comparisons, respectively, of CD130<sup>lo</sup>:CD130<sup>hi</sup> (red circle) and SSEA-1<sup>pos</sup>:SSEA-1<sup>neg</sup> (green triangle). Standard deviation (error bars) and average (black line) are indicated. D, Phase contrast images of colonies resulting from plating sorted CD130<sup>lo</sup> populations. E, Dot blot showing co-staining of CD130 and OCT4 by flow cytometry. Percent double positive cells is average of three biological replicates. F, Histograms showing analysis of PECAM-1 (top/green) and SSEA-1 (bottom/blue) in the sorted/plated CD130<sup>lo</sup> population. G, Bar graph illustrating total protein quantification of pluripotency and differentiation markers in SSEA-1<sup>pos</sup> and CD130<sup>lo</sup> populations via immunoblotting. Data are an average of three biological replicates. Significant differences (p < 0.05) are indicated by asterisk (*). H, Histograms showing analysis of CD130 (red) before FACS (left panel) and after (right two panels) live cell sorting of R1 mESCs. Each cell subpopulation (CD130<sup>lo</sup> and CD130<sup>hi</sup>) was plated, passaged once and then re-analyzed by flow cytometry. The results demonstrated that the subpopulations dynamically transition between these two states. The percent of CD130<sup>lo</sup> and CD130<sup>hi</sup> for each sorted subpopulation following plating is given in the inset (n = 3). I, Histological analysis of teratomas formed from the CD130<sup>lo</sup> population. i) Overview and magnified tissue sections (H&E stain) consistent with endoderm; ii, iii) intestine, iv) trachea; mesoderm: v) muscle, vi) bone, chondrocyte; ectoderm: vii) skin and hair follicle. Scale bar sizes are indicated.]

The surfaceome resource described here expands the known cell surface accessible proteome of mouse PSCs and establishes the experimental framework for developing accessible surface molecule barcodes for isolating more homogeneous, better defined, and therapeutically relevant populations of human PSCs for regenerative medicine. Although
previous studies with antibody arrays have revealed some of the proteins identified here (48), by starting with an antibody-independent discovery approach, we were able to simultaneously uncover CD and non-CD molecules, independent of antibody availability, quality, or specificity, and report surface accessibility. Consequently, this approach is invaluable for the identification and preselection of molecules of interest prior to obtaining antibodies for subsequent analysis of PSCs or PSC-derived progeny.

Data comparisons revealed four critical findings associated with this study. First, of the 500 surfaceome N-glycoproteins described, 187 have not been reported in the published PSC proteome, highlighting the ability of this strategy to access an underrepresented subproteome for expanding the available targets for immunophenotyping PSCs. Second, the unambiguous identification of N-glycosylation sites represents a resource of experimentally annotated extracellular domains useful for topology determinations and affinity reagent development, which is especially important in instances where predicted and publicly available information is limited or contradictory with regards to surface localization and transmembrane orientation. Thus, by providing experimental evidence regarding the surface accessibility of a protein domain, this resource is expected to reveal previously overlooked or novel candidates (e.g. PRSS50) and to facilitate the development of new reagents for the many proteins identified for which there are no available or flow compatible antibodies. Third, a subset of the proteins identified here are informative of the pluripotent state, whereas others, when experimentally verified, are informative of colony and cell line heterogeneity, potential stages of early differentiation, and cell type differences. Quantitative differences between mESCs and miPSCs shown by flow cytometry for CD19 and CD123 require further investigation, especially as quantitative differences in CD34 have proven critical for discrimination among HSC subpopulations. Fourth, several putative “PSC-restricted” proteins identified here are known to play important roles in pre-implantation embryo development (e.g., LPAR2 (49), PDGFRα (50), and PTGS2/COX2 (51)), whereas others not classified as “PSC-restricted” are present in pre-implantation blastocysts, including FGFR1, IGF1R, INSR, LIFR, and NOTCH1. Thus, it is expected that further interrogations of proteins identified in this resource will provide insights relevant to early embryonic developmental processes. Collectively, the compilation and comparison of our original data described here with that of more than 40 other datasets including PSCs and non-PSCs, provides an unprecedented insight into this subproteome and its usefulness for defining the pluripotent or differentiated cell state of stem cells.

Although these data are, to our knowledge, the most comprehensive study of the PSC surfaceome available, the PSC barcode is a work in progress. The CSC Technology employed in the present study focused only on N-linked glycoproteins. Although a majority of surface proteins are predicted glycoproteins, nonglycosylated and exclusively O-linked proteins were not targeted for identification. Separately, failure to identify a protein or qualitative differences among cell lines may result from the MS-level analysis (e.g. failure to select a peptide for fragmentation, failure to generate sufficient fragmentation for identification) or inherent properties of the protein (e.g. differences in glycosylation which affect our ability to capture the peptides). These are exacerbated in the case of low abundance proteins; consequently, some proteins reported in a single PSC line may be present in multiple lines. For example, two developmentally associated proteins (ZP3 and PRSS50) were only detected in mESCs via the CSC Technology but, by immunological-based methods, both were present in miPSCs (ZP3 data not shown). Finally, we expect that protein quantity and protein modifications, such as glycosylation, are likely to become critical components of future cell type-specific surface marker “barcodes,” analogous to that recently shown for CD133 in cancer (52, 53) and other examples from pluripotent populations (54–58).

The biological significance of this PSC surface proteome was established in two relevant applications by focusing on experimentally verified CD glycoproteins and selected markers preferentially expressed on PSCs. First, we unequivocally demonstrated the utility of surface marker pairs identified from our surfaceome to isolate iPSCs from heterogeneous mixtures using mouse MEFs containing an oct4-EGFP expression construct. In this study, CD112, CD326, or CD49f with CD31 all proved highly effective at isolating EGFP+ iPSCs, independently of EGFP or manual selection of colonies. Although the finding that CD326 is useful for isolating PSCs is consistent with recent reports (59–61), and CD49f has been reported in human repopulating spermatogonial stem cells, highly myogenic cell populations in muscle tissue, and in HSCs with long-term multipotent progenitor capacity (62–64), our use of marker pairs is likely to be more informative than strategies targeting a single protein. Moreover, to our knowledge, CD112 has not been previously described as an informative marker for PSCs. Combinations of CD49f, CD112, CD326, and CD31 with negative selection markers may ultimately prove even more effective in the isolation of highly purified iPSCs. Perhaps most importantly, these proof-of-principle findings demonstrate that surface accessible proteins can be used to isolate iPSCs independently of expressed and tagged intracellular proteins, thus overcoming current restrictions to the unbiased isolation of iPSCs for human therapeutic trials and regenerative medicine. Of note, the proteins that proved most useful for sorting iPSCs within the context of reprogramming were not PSC-restricted when compared with the Cell Surface Protein Atlas (Fig. 2). These data emphasize the importance of considering the biological context when developing useful marker panels.

Separately, multiple surface markers (EFNA2, GPC3, CD130) revealed a high degree of heterogeneity not previ-
Cell Surface Markers for Pluripotent Stem Cells

ously observed in PSCs, including one (CD130) that could be used in sorting experiments to show functional differences. Specifically, two unique populations of CD130 mESCs were identified that dynamically transitioned back and forth between the lo and hi CD130 states. While both populations were positive for pluripotency markers, CD130lo cells had a higher plating efficiency and was more strongly OCT4, SOX2, NANOG, and REX1 positive than CD130hi cells. Most importantly, only the CD130lo cell population robustly formed teratomas expressing multiple cell types from all three germ layers. Because CD130 forms a heterodimer with CD118 and is critical for signal transduction events in mouse PSCs, the differences in phenotype observed here may reflect altered CD118/CD130 signaling. CD130hi cells could also be in a unique substate awaiting a signal to differentiate, consistent with previous reports of the nonstatic nature of mESCs in culture (65, 66). Alternatively, the CD130 bimodal distribution could be related to epitope availability (not quantity), which can be affected by monomeric/heterodimeric structures, internalization, protein modifications, or binding of LIF (1). Although further functional studies are required to assess these populations in vitro and to determine their relevance during embryonic development, the ability of the CD130lo population to plate more efficiently, survive and form high grade teratomas is indicative of the enhanced pluripotent status of this ESC subpopulation. CD130 is not however useful for the isolation of iPSCs from reprogrammed MEFs, as it is also expressed on mouse fibroblasts.

In summary, this study provides a reference surfaceome “barcode” that will pave the way for a more rapid selection and analysis of candidate markers for defining the state of pluripotency, evaluating heterogeneity, isolating pluripotent cells from mixed cell populations, and for identifying accessible epitopes useful for generating antibodies against surface proteins of interest from mouse and human. This surfaceome is not only a valuable resource to investigators interested in iPSC generation and improved efficiency for isolating reprogrammed cells, these data also likely could prove valuable to studies of drug targets, mammalian development, and in the case of directed reprogramming strategies, this resource could assist in developing marker panels useful for determining whether rare pluripotent cells transiently exist in vivo. Altogether, the workflow and results described here, especially when applied to human PSCs, are expected to have a substantial impact upon future stem cell studies by increasing the efficiency of iPSC generation as well as providing new reagents required to overcome fundamental limitations to the clinical applications of stem cells.

Acknowledgments—We thank Cuong Nguyen (NIA), Tonya Wolf (NIA), Jade Scheers (NIA) and Hope Campbell (Blood Research Institute of WI) for assistance with flow cytometry and Attila Csordas and David Ovelliero at PRIDE for assistance with data conversion and upload. Special thanks to Dr. Paul Burridge (Stanford) for expert advice and insightful discussions and Dr. Konrad Hochedlinger (Harvard) for kindly providing 2D4 and TTF-1 cells. We report no conflict of interest.

This work was supported by National Institutes of Health (NIH) grant K99-L094708-01/4R00L094708-03, BD Biosciences Research Grant Award, the Innovation Center at the Medical College of Wisconsin (RLG); the Intramural Research Program of the NIH/NIA, NIH Induced Pluripotent Stem Cell Center/Center for Regenerative Medicine Research Study Award and Research Grants Council of Hong Kong (Theme-based Research Scheme T13-706) (KRB); NIH 5R01HL085434-02, AHA grant in aid (JVE); NIH 5K08HL87951, Blood Center Research Foundation (SR); NIH 1F31AA019874-01 (FN); NIH HL089471, NIH DK55743, NIH HL107434 (SAD); and National Center of Competence in Research Neural Plasticity and Repair (BW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

This article contains supplemental Table S1.

To whom correspondence should be addressed: Medical College of Wisconsin, Department of Biochemistry, Basic Science Building, 373, 8701 Watertown Plank Road, Milwaukee, WI 53226. Tel.: 414-955-2825; Fax: 414-955-6568; E-mail: rgundry@mcw.edu. Molecular Cardiology and Stem Cell Unit, NIH, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, MD 21224; Tel: 410-558-8055; Fax: 410-558-8150; E-mail: bohelerkgc.nia.nih.gov.

REFERENCES

1. Wobus, A. M., and Boheler, K. R. (2005) Embryonic stem cells: prospects for developmental biology and cell therapy. Physiol. Rev. 85, 635–678
2. Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676
3. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920
4. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Takahashi, C., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872
5. Banito, A., Rashid, S. T., Acosta, J. C., Li, S., Pereira, C. F., Geti, I., Pinho, S., Silva, J. C., Azuara, V., Walsh, M., Vallier, L., and Gil, J. (2009) Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev. 23, 2134–2139
6. Daley, G. Q., Lensch, M. W., Jaenisch, R., Meissner, A., Plath, K., and Yamanaka, S. (2009) Broaden implications of defining standards for the pluripotency of iPSCs. Cell Stem Cell 4, 200–201, author reply 202
7. Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rhodain, L., Loewer, S., Miller, J., Schlaeger, T., Daley, G. Q., and Feinberg, A. P. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. 41, 1350–1353
8. Feng, Q., Lu, S. J., Klimanskaya, I., Gomes, I., Kim, D., Chung, Y., Honig, G. R., Kim, K. S., and Lanza, R. (2010) Hemangioblastetic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. Stem Cells 28, 704–712
9. Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B. S., Li, P., Lim, S. L., Cao, S., Tay, J., Orlov, Y. L., Lufkin, T., Ng, H. H., Tam, W. L., and Lim, B. (2010) Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature 463, 1096–1100
10. Kim, P. G., and Daley, G. Q. (2009) Application of induced pluripotent stem cells to hematologic disease. Cytotherapy 11, 980–989
11. Osafune, K., Caron, L., Borowiak, M., Martinez, R. J., Fitz-Gerald, C. S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J. A., Evans, R. M., and Ecker, J. R. (2011) Hotspots of aberrant epigenomic reprogramming during human in vitro differentiation. Nat. Biotechnol. 29, 313–315
12. Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., O’Malley, R., Castanon, R., Kugman, S., Dornings, M., Yu, R., Stewart, R., Ren, B., Thomson, J. A., Evans, R. M., and Ecker, J. R. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471, 68–73
13. Gore, A., Li, Z., Fung, H. L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E., Lee, J. H., Loh, Y. H., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872

Molecular & Cellular Proteomics 11.8
32. Intoh, A., Kurisaki, A., Yamanaka, Y., Hirano, H., Fukuda, H., Sugino, H., and Asahima, M. (2009) Proteomic analysis of membrane proteins expressed specifically in pluripotent murine embryonic stem cells. Proteomics 9, 126–137

33. Dormeyer, W., van Hoof, D., Braam, S. R., Heck, A. J., Mummery, C. L., and Krijgsfeld, J. (2008) Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells. J. Proteome Res. 7, 2936–2951

34. Van Hoof, D., Braam, S. R., Dormeyer, W., Van-oostwaard, D., Heck, A. J., Krijgsfeld, J., and Mummery, C. L. (2008) Feeder-free monolayer cultures of human embryonic stem cells express an epithelial plasma membrane protein profile. Stem Cells 26, 2777–2781

35. Dormeyer, W., van Hoof, D., Mummery, C. L., Krijgsfeld, J., and Heck, A. J. (2008) A practical guide for the identification of membrane and plasma membrane proteins in human embryonic stem cells and human embryonal carcinoma cells. Proteomics 8, 4036–4053

36. Gerwe, B. A., Angel, P. M., West, F. D., Hasneen, K., Young, A., Orlando, R., and Stice, S. L. (2011) Membrane proteome signatures of karyotypically normal and abnormal human embryonic stem cell lines and derivatives. Proteomics 11, 2515–2527

37. Tarasova, Y., Riordor, D., Tarasov, K., and Boehler, K. (2006) Differentiation of mouse ES cells into muscle lineages. In: Evans, E. N. a. M., ed. Derivation and Manipulation of Embryonic Stem Cells: a Practical Approach, pp. 130–188, IRL Press, Oxford

38. Gundry, R. L., Runache, N., Brehm, C. R., and Boehler, K. R. (2011) Pluripotent stem cell heterogeneity and the evolving role of proteomic technologies in identifying biologically relevant differences in cell behavior. Proteomics Clin. Appl. 5, 3947–3961

39. Chan, E. M., Ratanasirintrawoot, S., Park, I. H., Manos, P. D., Loh, Y. H., Tchieu, J., Sridharan, J., Ward-van Oostwaard, D., Heck, A. J., Krijgsveld, J., and Mummery, C. L. (2008) Feeder-free monolayer cell culture N-linked glycoprotein: identification, glycosite occupancy, and membrane orientation. Mol. Cell. Proteomics 7, 2555–2569

40. Hofmann, A., Gerrits, B., Schmidt, A., Bock, T., Bausch-Fluck, D., and Boehler, K. R. (2010) Proteomic cell surface phenotyping of differentiating acute myeloid leukemia cells. Blood 116, 26–34

41. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383–5392

42. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646–4658

43. Persson, B., and Argos, P. (1997) Prediction of membrane protein topology utilizing multiple sequence alignments. J. Protein Chem. 16, 453–457

44. Rao, S., Zhen, S., Roumiantev, S., McDonald, L. T., Yuan, G. C., and Orkin, S. H. (2010) Differential roles of Sal1 isoforms in embryonic stem cell pluripotency. Mol. Cell. Biol. 30, 5364–5380

45. Si-Tayeb, K., Noto, F. K., Nagaoka, M., Li, J., Battle, M. A., Doris, C., North, P. E., Dalton, S., and Duncan, S. A. (2010) Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. Hepatology 51, 297–305

46. Szabo, P. E., Hubner, K., Scholer, H. R., and Mann, J. R. (2002) Allele-specific labeling and mass spectrometry of mouse C2C12 myoblast cell surface N-linked glycoproteins: identification, glycosite occupancy, and membrane orientation. Mol. Cell. Proteomics 1, 130–168, IRL Press, Oxford
Kemper, K., Sprick, M. R., de Bree, M., Scopelliti, A., Vermeulen, L., Hoek, M., Zelista, J., Pals, S. T., Mehmet, H., Stassi, G., and Medema, J. P. (2010) The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. *Cancer Res.* 70, 719–729

Alvarez-Manilla, G., Warren, N. L., Atwood, J., 3rd, Orlando, R., Dalton, S., and Pierce, M. (2010) Glycoproteomic analysis of embryonic stem cells: identification of potential glycobiomarkers using lectin affinity chromatography of glycopeptides. *J. Proteome Res.* 9, 2062–2075

An, H. J., Gip, P., Kim, J., Wu, S., Park, K. W., McVaugh, C. T., Schaffer, D. V., Bertozzi, C. R., and Lebrilla, C. B. (2012) Extensive determination of glycans in enriched plasma membranes of human embryonic stem cells. *Mol. Cell. Proteomics* 11, 10.1074/mcp.M111.010660

Chan, C. K., Rolle, M. W., Potter-Perigo, S., Braun, K. R., Van Biber, B. P., Laflamme, M. A., Murry, C. E., and Wight, T. N. (2010) Differentiation of cardiomyocytes from human embryonic stem cells is accompanied by changes in the extracellular matrix production of versican and hyaluronan. *J. Cell. Biochem.* 111, 585–596

Hennen, E., Czopka, T., and Faissner, A. (2011) Structurally distinct LewisX glycans distinguish subpopulations of neural stem/progenitor cells. *J. Biol. Chem.* 286, 16321–16331

Natunen, S., Satomaa, T., Pitkänen, V., Salo, H., Mikkola, M., Natunen, J., Otonkoski, T., and Valmu, L. (2011) The binding specificity of the marker antibodies Tra-1–60 and Tra-1–81 reveals a novel pluripotency-associated type 1 lactosamine epitope. *Glycobiology* 21, 1125–1130

Chen, H. F., Chuang, C. Y., Lee, W. C., Huang, H. P., Wu, H. C., Ho, H. N., Chen, Y. J., and Kuo, H. C. (2011) Surface marker epithelial cell adhesion molecule and E-cadherin facilitate the identification and selection of induced pluripotent stem cells. *Stem Cell Rev.* 7, 722–735

Ng, V. Y., Ang, S. N., Chan, J. X., and Choo, A. B. (2010) Characterization of epithelial cell adhesion molecule as a surface marker on undifferentiated human embryonic stem cells. *Stem Cells* 28, 29–35

Lu, T. Y., Lu, R. M., Liao, M. Y., Yu, J., Chung, C. H., Kao, C. F., and Wu, H. C. (2010) Epithelial cell adhesion molecule regulation is associated with the maintenance of the undifferentiated phenotype of human embryonic stem cells. *J. Biol. Chem.* 285, 8719–8732

Izadyar, F., Wong, J., Maki, C., Pacchiarotti, J., Ramos, T., Howerton, K., Yuen, C., Greilach, S., Zhao, H. H., Chow, M., Chow, Y. C., Rao, J., Barritt, J., Bar-Chama, N., and Copperman, A. (2011) Identification and characterization of repopulating spermatogonial stem cells from the adult human testis. *Hum. Reprod.* 26, 1296–1306

Notta, F., Doulatov, S., Laurenti, E., Poepl, A., Jurisica, I., and Dick, J. E. (2011) Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* 333, 218–221

Wilschut, K. J., van Tol, H. T., Arkestijn, G. J., Haagsman, H. P., and Roelen, B. A. (2011) Alpha 6 integrin is important for myogenic stem cell differentiation. *Stem Cell Res.* 7, 112–123

Canham, M. A., Sharov, A. A., Ko, M. S., and Brickman, J. M. (2010) Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* 8, e1000379

Payer, B., Chua de Sousa Lopes, S. M., Barton, S. C., Lee, C., Saitou, M., and Surani, M. A. (2006) Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genesis* 44, 75–83