

A Multiplex High-Throughput Targeted Proteomic Assay To Identify Induced Pluripotent Stem Cells

Anna Baud(1)§, Frank Wessely(2)§, Francesca Mazzacuva(1), James McCormick(1), Stephane Camuziaux(1), Wendy E. Heywood(1), Daniel Little(3), Jane Vowles(4,7), Marianne Tuefferd(5), Olukunbi Mosaku(3), Majlinda Lako(8), Lyle Armstrong(8), Caleb Webber(2), M. Zameel Cader(6), Pieter Peeters(5), Paul Gissen(3)¥, Sally A. Cowley(4,7)¥ and Kevin Mills(1)¥*.

(1) Centre for Translational Omics, UCL Great Ormond Street Institute of Child Health, London, WC1N 1EH, UK (2) Department of Physiology, Anatomy & Genetics, Oxford University, Oxford, OX1 3PT, UK (3) MRC Laboratory for Molecular Cell Biology, University College London, London, WC1E 6BT, UK (4) Oxford Parkinson's Disease Centre, University of Oxford, Oxford, OX1 3QX, UK (5) Janssen Research and Development, Beerse, 2340, Belgium (6) The Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK (7) James Martin Stem Cell Facility, Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK (8) Institute of Genetic Medicine, Newcastle University, Newcastle, NE1 3BZ, UK * e-mail: kevin.mills@ucl.ac.uk Tel +44(0)207 905 2873

ABSTRACT: Induced pluripotent stem cells have great potential as a human model system in regenerative medicine, disease modeling and drug screening. However, their use in medical research is hampered by laborious reprogramming procedures that yield low numbers of induced pluripotent stem cells. For further applications in research, only the best, competent clones should be used. The standard assays for pluripotency are based on genomic approaches, which take up to 1 week to perform and incur significant cost. Therefore, there is a need for a rapid and cost-effective assay able to distinguish between pluripotent and non-pluripotent cells. Here, we describe a novel multiplexed, high-throughput and sensitive peptide-based multiple reaction monitoring mass spectrometry assay, allowing for the identification and absolute quantitation of multiple core transcription factors and pluripotency markers. This assay provides simpler and high-throughput classification into either pluripotent or non-pluripotent cells in 7-minutes analysis while being more cost-effective than conventional genomic tests.

Induced pluripotent stem cells (iPSCs) are somatic cells, which are reprogrammed to the pluripotent state and acquire properties comparable to embryonic stem (ES) cells (1). Initially generated from human dermal fibroblasts by retroviral transfection of four transcription factors OCT4, SOX-2, c-MYC and KLF-4 (2), iPSCs can be now obtained by numerous different strategies, such as plasmids (3), non-integrating viruses (Sendai virus) (4, 5), drug-inducible systems (6), transposition (7), protein transduction (8) or mature microRNAs (9). Despite multiple reprogramming techniques available this process remains time-consuming and relatively inefficient. For further applications in research, only the best, competent clones need to be selected and used. The quality control assays vary from unreliable visualization of colony-morphology using light microscopy or extracellular marker staining, through more sophisticated embryoid body or teratoma formation. The latter *in vivo* test is not only time-consuming and expensive, but also lacks the standardization and raises ethical concerns (10). Therefore the current trend is to move towards assays based on gene expression, i.e. PluriTest (11) or ScoreCard (12). However these tests require isolation of genetic material prior to analysis, take up to 1 week to perform and incur significant cost.

In order to increase the simplicity and efficiency of iPSCs characterization, we have developed a multiplexed peptide based multiple reaction monitoring - mass spectrometry assay (MRM-LC-MS/MS), enabling quantitation of 15 pluripotency

markers in just 7 minutes. This test provides a simple and high-throughput measurement of pluripotency signature using as little as 200,000 cells. This work describes the development of a novel targeted proteomic assay and comparison of the mass spectrometry-based test with conventionally used tests based on gene expression profiles (ScoreCard and PluriTest).

EXPERIMENTAL PROCEDURES

Cell lines and culture. Cell lines used in this study are listed in Table S-1. iPSC lines (prefixed with SFC) were derived as part of the EU IMI-funded programme, StemBANCC, from donors who had given signed informed consent for derivation of hiPSC lines from skin biopsies. The SFC826 lines were derived from SF826 donor fibroblasts collected by the University of Lübeck. The study was approved by the ethics committee of the University of Lübeck, and all participants gave written informed consent prior to their participation in the study. SBAD iPSC lines were derived and characterised at the University of Newcastle from Lonza fibroblasts CC-2511, Lot 264781, Tissue Acquisition number 23447. All other StemBANCC iPSC lines were derived and characterised at the University of Oxford, James Martin Stem Cell Facility. The human ES cell line HUES-2 (passages 16–38) was obtained from the HUES Facility, Harvard University. Ethical approval for work on all hES cell lines was reviewed and approved by the UK Stem Cell Bank Steering Committee (Medical Research Council, London UK, 20.10.2005), and work using this

line was funded by the Oxford Martin School and the Wellcome Trust.

For derivation of iPSC lines, skin biopsies were cultured as previously described (13) to promote outgrowth of fibroblasts, in ADMEM (Invitrogen) supplemented with USDA-approved foetal bovine serum (10%; Sigma) and penicillin/streptomycin (1%) in a humidified incubator (37 °C, 5% CO₂). Fibroblasts were reprogrammed at passage 3-5, using Cytotune Sendai virus reprogramming kit (Invitrogen) according to the manufacturer's instructions (scaled down to 50,000 fibroblasts). Transduced fibroblasts were seeded onto CF1 outbred mouse embryonic feeder cells (MEF, Merck) on 0.1% gelatin coated plates (Sigma) on day 7, and cultured in KnockOut serum replacement medium (Knock-out DMEM (Invitrogen), KO-Serum Replacement (20% Invitrogen), Glutamax-I (2 mM Invitrogen), non-essential amino acids (1%, Invitrogen), penicillin (100 U/mL Invitrogen), streptomycin (100 µg/mL Invitrogen), 2-ME (55 µM Invitrogen) and bFGF (10 ng/mL R&D), substituting with MEF-conditioned medium from day 10 onwards. Colonies displaying iPSC morphology were picked and passaged on MEFs by manual dissection every 5-7 days. iPSC lines were adapted to feeder-free culture conditions in mTeSR™1 (StemCell Technologies), on Matrigel coated plates (BD Matrigel hESC-qualified Matrix), passaging as clumps using 0.5 mM EDTA in PBS (Beers et al., 2012). Cells were frozen in SNP-QCed batches at p15-25 and used for experiments within a minimal number of passages post-thaw to ensure consistency.

iPSC lines that have not been previously published are characterised in Figure S-1.

Initial assessment of pluripotent markers used antibodies to TRA-1-60 (B119983, IgM-488, Biolegend) and Nanog (2985S, IgG-647, Cell Signaling), with appropriate isotype control, at the same concentration, from the same supplier. Cells were fixed for 10 minutes in 2% paraformaldehyde in PBS (Alfa Aesar), permeabilised in 100% methanol at -20 °C for at least 30 min before staining. Fluorescence was measured using a FACS Calibur (Becton Dickinson), and data analysed with FlowJo software.

RT-PCR was used to assess clearance of Cytotune Sendai virus-delivered reprogramming genes. RNA was isolated using an All-Prep kit (Qiagen) and reverse transcribed using a RetroScript kit (Ambion), with 2 µg template RNA in a 20 µL reaction. 2 µl of 1:10 dilution of cDNA product was used in a 25 µl RT-PCR reaction, performed according to the manufacturer's instructions and run on a 1.5% agarose gel with Log2 ladder (NEB). Positive controls (fibroblasts infected 5 days previously) were always run in parallel. Primers were SeV F: GGATCACTAGGTGATATCGAGC, R: ACCAGACAAGAGTTTAAGAGATATGTATC 181bp; SOX2 F: ATGCACCGCTACGACGTGAGCGC, R: AATGTATCGAAGGTGCTCAA 451bp; KLF4 F: TTCCTGCATGCCAGAGGAGCC, R: AATGTATCGAAGGTGCTCAA 410bp; c-MYC F: TAACTGACTAGCAGGCTTGTCG, R: TCCACATACAGTCCTGGATGATGATG 532bp; OCT4 F: CCCGAAAGAGAAAGCGAACCAG, R: AATGTATCGAAGGTGCTCAA 483bp; β-Actin control Eurogentec 92bp.

Genomic DNA was made using an All-Prep kit (Qiagen). Genome integrity was assessed by Illumina Human CytoSNP-12v2.1 beadchip array (~300,000 markers) or OmniExpress24 array (700,000 markers) and analysed using GenomeStudio

software (Illumina). Ancestry plots (a tracking QC to confirm that the iPSC lines derive from the parental fibroblasts) and karyograms were generated using a customised SNP analysis pipeline in StemDB, which hosts StemBANCC datasets (<https://www.stemdb.org>).

iPSCs used for the blind test were grown in feeder free conditions in Essential 8 medium (ThermoFisher) on Matrigel coated plates (BD Matrigel hESC-qualified Matrix). Cells were harvested using 0.5 mM EDTA in PBS, centrifuged, and the resulting pellets were taken for analysis. Three different fibroblast lines were cultured in DMEM (Sigma) containing 10% fetal calf serum, 2 mM L-glutamine, and non-essential amino acids. Fibroblasts were harvested with 0.05% trypsin-EDTA, centrifuged, and the resulting pellets were taken for analysis. The collected cell pellets were labeled 1-7 so that the person performing the test was unaware of the cell type of each sample.

Differentiation to embryoid bodies. iPSC were cultured as above, and set up for both harvesting and differentiation. Harvesting was by washing with PBS, incubating for 5 mins with TrypLE express (Life Tech), then diluted 1/10 in PBS, centrifuged at 400g for 5 minutes, supernatant aspirated and the pellet frozen immediately at -80 °C. Cells were pelleted at ~2 million cells/pellet, 3 pellets per sample: 1 pellet for proteomics, 2 for RNA. For differentiation cultures, 10,000 PSC were seeded per Aggrewell 800 as per manufacturer's protocol (StemCell Technologies). Seeding was in mTeSR with 10 µM Rock inhibitor Y-27632 (AbCAM), then medium changed at 24h into ScoreCard differentiation medium as per ScoreCard manufacturer's protocol (Life Technologies), fed daily and harvested at day 4. For 14 day EBs, d4 Aggrewell EBs were plated to matrigel as per ScoreCard protocol, and fed every 2d with ScoreCard medium for 10d (undirected differentiation).

Targeted proteomics: MRM-based triple quadrupole mass spectrometry. All materials were of analytical and mass-spectrometry grade. DL-dithiothreitol (DTT), iodoacetamide, ASB-14, Tris base and urea were all purchased from Sigma-Aldrich. UPLC-MS grade acetonitrile (ACN), formic acid (FA) and water were obtained from Fluka, and sequencing-grade modified porcine trypsin from Promega. All buffers and solutions were prepared using ultra-pure 18 MΩ water (MilliQ) and UPLC solvents using UPLC-MS grade water.

Potential pluripotency biomarkers were determined from the literature and current pluripotency tests (staining markers and self-renewal markers included in ScoreCard assay). Representative quantotypic peptides for each protein were selected using the open source online global proteome machine MRM database at www.thegpm.org (14) and SRM atlas database at www.srmatlas.org (15). Custom synthesised standard peptides (Genscript, USA) were used to create the transitions list, optimize the peptide detection and determine the retention time. Details of confirmed marker peptides are given in Table S-2.

Harvested cell pellets (~2 million cells/pellet) were dissolved in lysis buffer containing 100mM Tris HCl, 6M Urea, 2M Thiourea, 2% ASB-14, and spiked with 20 pmol of stable isotope labelled (SIL) analog of standard peptides with trypsin tag (Thermo-Scientific) in order to monitor the efficiency of digestion. Subsequently cell lysates were digested using the sequencing-grade trypsin as described previously (16), and purified using C18 Isolute 96-well plate (Biotage) according to the manufacturer's instructions. Eluted peptides were freeze-dried and resuspended in 100 µL of 0.1% formic acid in water. 10 µL of sample was loaded onto a Waters CORTECS UPLC

C18 Column, 90 Å, 1.6 µm, 2.1 mm x 50 mm attached to a C18 VanGuard pre-column. UPLC-MS/MS analysis was performed on a Acquity UPLC system (Waters) coupled to a Xevo™ TQ-S triple quadrupole mass spectrometer (Waters) equipped with electrospray source operating in positive ionization mode as described previously (16). Analytical UPLC chromatography was performed using the flow rate 0.8 mL/min, and a 7-min linear gradient starting at 100% solvent A (0.1% formic acid in water), ramping to 40% solvent B (0.1% formic acid in acetonitrile) in 3.5 min, then to 99.9% Solvent B over 0.25 min (held for 1.5 min) and finally decreased to 100% solvent A in 0.25 min and reconditioned for 1.5 min. Dynamic Multiple Reaction Monitoring of a multiplex of 15 markers (at least two peptides per protein and two transitions per peptide were monitored- one for quantitation and one for confirmation) was performed over a 7 min gradient with a dwell time set up to automatic, and minimum 8 data points per peak. Only one NANOG peptide was monitored, due to the primary sequence of this protein, characterized by low number of tryptic sites and multiple phosphorylation sites, decreasing number of possible candidate peptides, which could be used in the assay. Additionally, only one transition of TRIM6 peptide VIPMTLRR was monitored due to the large number of non-specific transitions generated during method development. Chromatograms were analysed using Waters TargetLynx Software V.4.1. Standard peptides were dissolved in lysis buffer, spiked with SIL standard peptides, digested and purified using C18 Isolute 96-well plate (Biotage) in the same manner as cell pellets to encounter matrix effect on peptide retention time. Eluted standard peptides were freeze-dried and resuspended in 0.1% formic acid in water to concentrations used in standard curve.

Peptides were standardised by using a spiked SIL peptide and absolute levels were obtained from standard curves. A standard curve of 0-1 pmol/µL of each peptide was exported to Excel and GraphPad Prism for statistical analysis. The linearity of $R^2 > 0.90$ was achieved for all calibration curves, except the Podocalyxin, for which $R^2=0.87$ was achieved (Figure S-2). Limit of quantitation (LOQ) was determined for each compound using 5 repeat measurements for 5 consecutive days of standard peptides. LOQ defined as the compound concentration for which calculated coefficient of variation (CV) was below 20%, was found to be 50 fmol/µL for PODXL and GAPDH, 10 fmol/µL for LIN28 and 5 fmol/µL for SOX-2, OCT4 and CD44. The intra-batch variation was determined to be between 0.4 and 15%, and inter-batch variation being between 1.28 and 19.5% (n=5 for 5 consecutive days).

Transcriptomic profiling. RNA extraction was done using RNeasy mini kit (Qiagen). Out of the total RNA yield collected, 1µg was used for microarray profiling. Genomewide analysis of gene expression was done using Illumina's HumanHT-12-v4 expression BeadChip according to Illumina's protocol. Pluripotency assessment was performed using PluriTest algorithm (11) (pluritest.org). When available, the leftover RNA was profiled by multiplex qPCR profiling, using Taqman® hPSC ScoreCard™ panel (17) (ThermoFisher Scientific) for pluripotency and differentiation assessment according to the manufacturer's protocol.

Data analysis. All data analyses were carried out in R. Results presented in this study can be reproduced by the Rmarkdown script available from http://www.fgu.anat.ox.ac.uk/downloads/compbio_projects/CW025_WESSELY_MRM/

Samples used in the PluriTest and ScoreCard assay are listed in Table S-3. PluriTest classification results and scores were obtained from the analysis pipeline implemented as a web-service and available from <http://www.pluritest.org/> (accessed October 2015). Raw data from five Illumina HumanHT-12 V4.0 expression beadchips each containing 12 samples were individually uploaded in form of .idat files. The five corresponding results tables were downloaded and merged to obtain PluriTest results for 49 samples (remaining 11 samples not part of this study).

Normalised probe expression levels of selected genes of interest, which were used to compare measurements between the three assays, were obtained from probe expression profiles generated with GenomeStudio version 1.9.0. Data was pre-processed in the same way as the standardised PluriTest web-service by using the lumi R package (18, 19) with variance stabilising transformation (VST) followed by robust spline normalisation (RSN). Gene annotation of microarray probes was based on GenomeStudio mapping. If several probes mapped to the same gene, the probe with the highest Spearman's correlation coefficient with proteomic MRM assay levels was chosen.

ScoreCard qPCR assay results were received from Life-Tech/ThermoFisher for a total of 42 samples. Scores and classification of samples as positive, negative or borderline for self-renewal potential were extracted from the report. Fold changes of self-renewal genes also analysed by the MRM assay (OCT4/POU5F1 and SOX-2) were used as provided in the report. CD44, used as a negative control for mouse contamination in the qPCR assay, was expectedly not detected in any sample (CT values = 40).

RESULTS AND DISCUSSION

Reprogrammed cell lines. A total of 14 newly reprogrammed cell lines (3 clones each from 5 donors, except for one donor with 2 clones) and an established human embryonic stem cell line HUES-2 (20) at three stages of differentiation: iPSCs (undifferentiated), 4-day and 14-day embryoid body (EB) formation were included in the study. Additionally, the donor fibroblast cell lines were used as negative control (Table S-1 and Figure S-1). Cell lines included in the study carried different mutations related with Parkinson's disease: PTEN-induced putative kinase 1 (PINK1) mutation, A53T mutation in α -synuclein, triplication in the α -synuclein gene (SNCA), and the N370S mutation at the acid beta-glucosidase (GBA) gene, which is associated with Gaucher disease (Table S-1).

The differentiation potential of the reprogrammed cell lines was confirmed by the observed trends of increasing germ-layer scores when iPSCs were compared to 4-day and 14-day embryoid bodies, although not all samples at the embryoid body stage were classified as positive for specific germ layers (Figure S-3). Differentiation potential was found to be most pronounced for mesoderm layer and least pronounced for endoderm layer. The trends for differentiation potential observed here largely correspond to the pattern previously observed: rapid activation of ectoderm markers, gradual activation of mesoderm markers and delayed activation of endoderm markers (17).

Development of MRM-based pluripotency signature assay. We primarily included the core transcription factors (TFs) as candidate biomarkers: OCT4, SOX-2, KLF4 and c-MYC in the MRM-based assay, and subsequently added other factors qualified in PluriTest and ScoreCard assays as pluripotency-

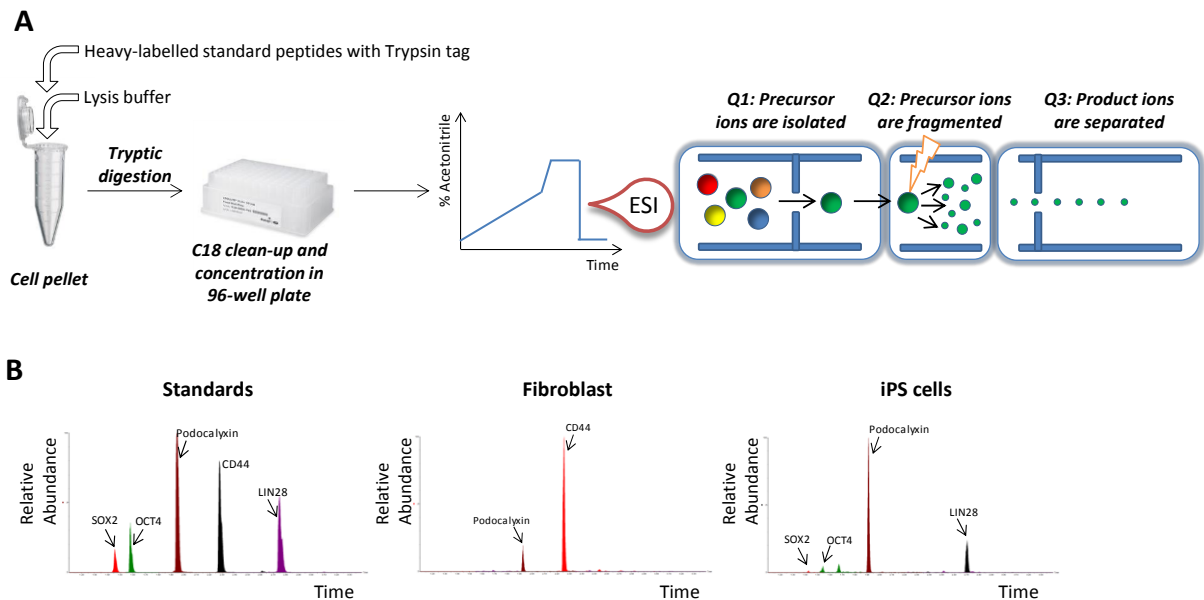


Figure 1. Schematic representation of a targeted proteomic mass spectrometry test for pluripotency. (A) The workflow of the assay: a cell pellet is spiked with SIL standard peptides, trypsin digested, desalted and concentrated. Resulting peptides are then separated by UPLC and analysed by a triple quadrupole mass spectrometer. (B) Typical and representative result of a sample analysis. The overlaid chromatogram of the standard peptides included in the multiplexed targeted proteomic assay and detected in cell lysates of iPSc and fibroblasts. Only 5 markers used for quantitation are displayed.

associated markers (12, 21). The final assay was developed for identifying and quantitating Myc proto-oncogene (c-MYC), Kruppel-like factor 4 (KLF4), Transcription Factor SOX-2, POU domain transcription factor 1 (OCT4, also known as POU5F1), Homeobox protein NANOG, Protein LIN28 homolog A, Zinc finger protein 42 (REX1), Tripartite motif-containing protein 6 (TRIM6), Steroid hormone receptor ERR2 (ESRRB) and Zinc finger protein GLIS1. Additionally, histochemical pluripotency markers: Podocalyxin (PODXL-transmembrane glycoprotein shown to have binding activities with TRA-1-60 and TRA-1-81 (22)), Alkaline Phosphatase, Neural cell adhesion molecule 1 (CD56), CD44 antigen and a housekeeping protein- glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were added to the multiplexed MRM panel. For each marker, we selected at least two proteotypic peptides, determined two optimum MRM transitions (pairs of parent and product ion) and instrument parameters (Table S-2). These 15 candidate pluripotency biomarkers were multiplexed into a 7-minute targeted peptide MRM-based assay (Figure 1).

Analysis of reprogrammed cell lines by MRM-based assay. The targeted proteomic analyses of reprogrammed cell lines, fibroblasts and human embryonic stem cell lines revealed that the four biomarkers OCT4, SOX-2, LIN28 and PODXL allowed for the measurement of a pluripotency signature. The mean concentration of these markers normalised to GAPDH, are significantly different between fibroblasts and iPSCs ($p < 0.0005$, Wilcoxon rank sum test) and allowed for clear classification of a given cell line into one of these groups (Table 1 and Figure 2C). The accuracy of using GAPDH for normalization was confirmed by parallel normalization with two other housekeeping proteins: DNA replication licensing factor MCM2 and DNA mismatch repair protein MSH2 (data not shown). No significant difference was observed and thus we retained the normalization to GAPDH, being the standard protein to normalise against in biochemistry. Another normalization method, such as the quantitative protein assay could not

be performed due to the high concentration of urea and thiourea in the lysis buffer, being incompatible with the assay.

The pluripotency signature was also confirmed by quantitating the negative marker, CD44, a surface glycoprotein, which functions as a receptor for hyaluronic acid and was reported previously as highly expressed in human fibroblasts and absent in pluripotent cells (23). The concentration of CD44 measured by proteomic assay in fibroblasts was almost 95-fold higher than in iPSCs. Augmenting this marker into our MRM assay allows for identification of non-fully reprogrammed clones, which still express fibroblast surface markers.

The principal components analysis based on OCT4, SOX-2, LIN28, PODXL and CD44 markers (Figure 2A) clearly shows the separation of fibroblasts from the rest of the samples and highlights the transition from undifferentiated cells to 14 day EBs. No significant difference in expression of c-MYC marker was found between fibroblasts and iPSCs lines. In line with our results, c-MYC was shown to be expressed strongly in pre-iPS cells (24). However, the remaining candidate biomarkers included in the assay (KLF4, NANOG, REX1, TRIM6, ESRRB, GLIS1, Alkaline Phosphatase and CD56) were largely not detected in the samples. The reason for the absence of these potential biomarkers is unclear but it is well documented that there is a poor relationship between mRNA and protein content of the cell (25). Another possible reason is that the abundance levels of those proteins cannot be detected by this method. However, the detected OCT4, SOX-2, LIN28, PODXL and CD44 demonstrate sufficient potential for use in clearly discriminating fibroblasts from iPSCs. Therefore the lack of detection of these other markers is not needed for the purpose of this assay.

Analysis of reprogrammed cell lines by ScoreCard and PluriTest. Technical replicates of reprogrammed cell lines submitted to the MRM pluripotency assay were also analysed using the commercially available TaqMan® hPSC ScoreCard™ Assay and the open-access bioinformatic assay PluriTest. Both tests provide a number of scores that are used to

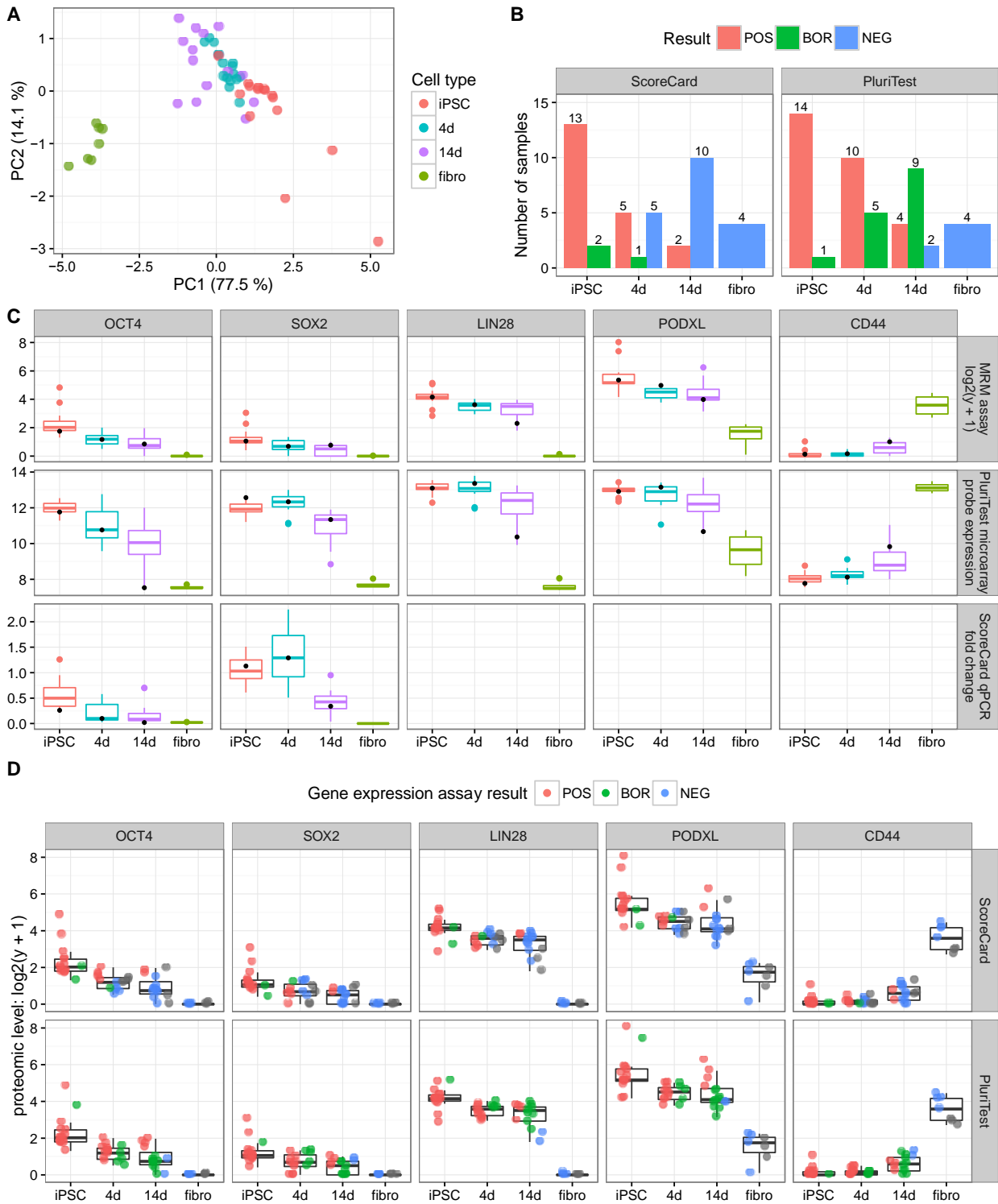


Figure 2. Results from proteomic and genomic pluripotency tests. (A) Principal components analysis (PCA) of proteomic MRM measurements of five biomarkers (OCT4, SOX-2, LIN28, PODXL and CD44) from 52 samples (see Table S-1 for the sample list). The percentage of variance explained by the first and second principal components (PC1 and PC2) is shown in brackets. (B) ScoreCard's self-renewal and PluriTest's classification of samples from different cell types into three categories: positive (POS), negative (NEG) or borderline (BOR). See Table S-3 for genomic test results. (C) Boxplots of assay measurements from individual pluripotency tests for five biomarkers. Boxes represent inter-quartile ranges (IQR) with a horizontal line for the median, lower and upper whiskers extend to the lowest and highest value within $1.5 \times \text{IQR}$ and measurements beyond the end of whiskers are plotted as points. The black points highlight the measurements of the embryonic stem cell line HUES2. Seven and four fibroblast samples were analysed by the proteomic and genomic assays, respectively. The number of samples of other cell types is 15 in all three assays except for ScoreCard embryoid bodies (4d: n=11, 14d: n=12). (D) MRM measurements are shown as boxplots equivalent to the top row in c and compared to the genomic test results using the same colour code as in b. Samples not analysed by ScoreCard or PluriTest are shown in grey. Abbreviations: 4d- 4-day embryoid body; 14d- 14-day embryoid body; y- fmol of biomarker/fmol of GAPDH.

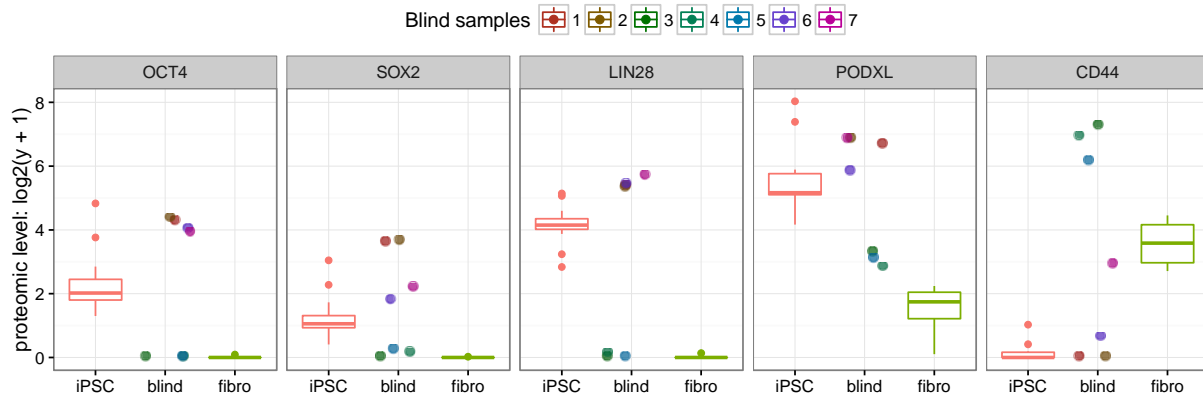


Figure 3. Results of the MRM blind test. Measurements of five biomarkers allowing for the classification into either iPSC or fibroblast range. Blind 1, 2 and 6 were correctly classified as iPSC, Blind 3, 4 and 5 as fibroblast. Blind 7, expressing both iPSC and fibroblast markers, corresponds to the iPSC cell line with spontaneous differentiation. As a comparison, boxplots are shown for MRM assay measurements of n=15 iPSC samples (including HUES2) and n=7 fibroblast samples. These boxplots are equivalent to the ones shown in Figure 2C. Blind samples are listed in Table S-4. Abbreviations: y, fmol of biomarker/fmol of GAPDH.

Table 1. Concentration ranges of biomarkers OCT4, SOX-2, PODXL, LIN28 and CD44 in Fibroblasts and iPSC cells detected by targeted proteomic assay.

Biomarker	Fibroblast n=7 (fmol of biomarker/ fmol of GAPDH)			iPSC n=15 (fmol of biomarker/ fmol of GAPDH)		
	Mean	Median	Min-Max	Mean	Median	Min-Max
OCT4	0.01	0.00	0-0.06	5.42	3.06	1.46-27.45
SOX-2	0.002	0.00	0-0.02	1.71	1.08	0.33-7.27
PODXL	2.17	2.35	0.07-3.73	60.89	34.97	16.98-260.87
LIN28	0.01	0.00	0-0.1	18.04	16.75	6.15-34.36
CD44	12.14	11.01	5.54-20.91	0.13	0.03	0-1.04

classify samples as being positive or negative for self-renewal or borderline (PluriTest labels the latter samples as 'further'). ScoreCard classified 86% of iPSCs lines as positive for self-renewal potential, whereas the remaining samples were classified as borderline. 45% of 4d EBs and 17% of 14d EBs also showed self-renewal potential (Figure 2B and Table S-3). PluriTest confirmed a pluripotency signature for 93% of iPSCs samples, including the two samples described as borderline by ScoreCard. The microarray-based test classified more EB samples positively than ScoreCard, including all the samples apart from one classified as pluripotent by ScoreCard (Figure 2B and Table S-3). However, it should be noted that PluriTest is usually not used for embryoid body samples. Both genomic tests show a high correlation between pluripotency scores (Spearman's correlation 0.8).

Comparison of MRM-based test results with ScoreCard and PluriTest. Direct comparison of the targeted proteomic test and gene-based pluripotency tests can be challenging since previous reports have shown that in general, protein abundances correlate with mRNA very poorly (26), which can be explained by translation and degradation of proteins occurring after transcription. Proteomic expression of the OCT4, SOX-2,

LIN28, PODXL and CD44 biomarkers measured by the MRM assay were compared to corresponding gene expression data measured by the two genomic tests (Figure 2C). The genome-wide microarray used by PluriTest includes all corresponding genes, whereas ScoreCard's qPCR assay includes only two of our selected set of biomarkers (OCT4 and SOX-2). The trends across the four different stages of cell differentiation observed by targeted proteomics were also found for gene expression measurements (Figure 2C). It is noteworthy that the embryonic stem cell samples (HUES2) used in our study, shows very similar results to the 14 iPSC cell lines. A high degree of correlation was found between MRM assay and both genomic assays, with Spearman's correlation coefficients ranging from 0.6 to 0.9 ($p < 0.00001$), with the highest correlation observed for OCT4. Proteomic levels were additionally compared to the classification results of both gene-expression based tests (Figure 2D). One of two iPSCs samples classified as borderline by ScoreCard shows consistently lower proteomic measurements outside the inter-quartile range for all four self-renewal markers. On the other hand, the single borderline iPSCs sample by PluriTest shows consistently higher proteomic levels outside the inter-quartile range, providing evidence for classifying this

sample as pluripotent by the MRM assay. Also, the 14d EBs classified as positive by ScoreCard (n=2) and PluriTest (n=4) show increased readouts of proteomic levels.

Note that two more biomarkers (NANOG and c-MYC) that were included in the initial MRM assay design were monitored by ScoreCard and PluriTest. Although NANOG was measured as differentially expressed between iPSCs and fibroblasts based on microarray and qPCR data, protein abundance was only detected in one iPSC sample. A potential reason for this could be due to the characteristic primary sequence of NANOG, containing tryptic cleavage sites clustering together in the center of the protein, therefore, decreasing number of potential peptides, which could be used in MRM assay. Additionally, NANOG has 11 phosphorylation sites (27) that could hinder the utility of modified peptides for MRM assay. Heavy phosphorylation results in the net negative charge of large regions of the protein, hampering the digestion efficiency of trypsin, which recognizes regions of net positive charge. Together this leads to tryptic fragments being too large and having poor propensity for ionization, thus hindering detection of these peptides. c-MYC, for which protein levels were detected in half of the iPSCs samples, showed no significant difference between iPSCs and fibroblasts, which was also observed for the gene expression data.

Additionally, the accuracy of the MRM test in assessing the pluripotency signature was confirmed by submission of 7 samples to a blind test (Table S-4). All iPSCs and fibroblast cell lines were properly classified into one of these groups (Figure 3). Interestingly, one sample expressed the OCT4, SOX-2, PODXL and LIN28 within the iPSCs range, but CD44 was found to be close to the lower limit of the fibroblast range. Metadata revealed that this iPSCs sample differentiated spontaneously during cell culture, leading to the mixed expression of biomarkers.

Logistic regression was used to obtain a classifier that can predict pluripotent versus non-pluripotent class assignment (data not shown). The MRM values of the five discriminative biomarkers for iPSC samples positively evaluated by ScoreCard (including HUES2 and excluding borderline samples; n=13) for self-renewal and the fibroblasts (negatively evaluated by ScoreCard; n=4) were used as training data. The seven blind samples were then used to evaluate the performance of the classifier. However, the training set represents a perfectly separable dataset, which is also true by using individual biomarkers as a single predictor variable, resulting in perfect classification accuracy and preventing the estimation of regression parameters.

CONCLUSION

Our targeted proteomic and mass-spectral based test demonstrates that the measurement of the expression of proteins OCT4, SOX-2, LIN28, PODXL and CD44 can be successfully applied to assess the self-renewal potential of reprogrammed cells. The pluripotency signature measured by the mass spectrometry was confirmed by the generation of three-germ layer embryonic bodies and the successful terminal differentiation of these clones into neuronal cells (13, Hanseler et al., Melguzo et al. in preparation).

Due to the novel nature of our MRM-based test, we could not exploit publicly available transcriptional profiles of iPSCs from different laboratories, as it was done for the evaluation of genomic-based tests (11). To overcome this problem, we reprogrammed 14 new cell lines and submitted them to the pro-

teomic and two genomic tests. The comparison of MRM-based results with TaqMan® hPSC ScoreCard™ Assay and PluriTest performed on the same cell lines has shown a high correlation between results based on transcriptomic and proteomic data, and allowed for the validation of our proteomic approach for measuring pluripotency signature. In all analyses, the targeted proteomic assay agreed with both genomic-based tests. The overall goal of our study was to generate a streamlined, economic and faster way to validate pluripotency of iPSCs, without compromising on the reliability of the assay. The main advantage of a targeted proteomic test is the low cost, high-throughput format and requirement of only 200,000 cells for analysis. The sample digestion and purification can be performed in 96-well plates thus significantly increasing the speed of sample preparation. The results can be obtained within only one day, including sample preparation, mass-spectrometry analysis and data interpretation. To compare, ScoreCard and PluriTest assays require about 1 week to be performed, including sample preparation. Moreover, we estimate the cost price of the targeted proteomic approach to be less than \$20 per sample (assuming access to a mass spectrometer), which is significantly lower compared to the transcriptomic approach, where the average price per sample is \$150 for PluriTest and \$175 for ScoreCard (17) (Figure S-4). The development of our new MS-based test is even more important in light of the announcement of Illumina to discontinue the HumanHT-12 v4 Expression BeadChip Kit by the end of 2016. This will hamper access to the PluriTest assay, which is built upon this type of gene expression array.

Our test can be implemented in any laboratory, which has access and expertise in using triple quadrupole mass spectrometers. Although targeted mass spectrometry approach has been already successfully applied in both biology and medicine (28–30), the proteomics of iPSCs is still an emerging field, with the main interest in deep profiling of cells (31–34). Here we report for the first time the application of targeted proteomics to measure the pluripotency of iPSCs, which can potentially be translated into a diagnostic test.

In the current design, this mass-spectral based test allows for assessment of self-renewal potential. However, it does not provide information about the differentiation efficiency into three germ layers as ScoreCard does, but this improvement is currently under development.

In summary, we have developed a high throughput, rapid and cost-effective assay capable of directly measuring core pluripotency biomarkers used to assess pluripotency potential. This assay is significantly simpler and faster than conventional genomic tests.

ASSOCIATED CONTENT

Supporting Information Available:

Fig. S-1: Characterization of previously unpublished cell lines. Fig. S-2: Representative standard curves. Fig. S-3: Overview of ScoreCard trends. Fig. S-4: Comparison of three pluripotency assays. Table S-1: List of samples included in the study. Table S-2: List of selected peptides and transitions of candidate pluripotency biomarkers. Table S-3: Overview of the ScoreCard and PluriTest results. Table S-4: List of samples included in the MRM-based blind test. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

* Dr Kevin Mills e-mail: kevin.mills@ucl.ac.uk Tel +44(0)207 905 2873

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

§ Authors contributed equally (joint first authors)

¥ Authors contributed equally (joint last authors)

ACKNOWLEDGMENT

The authors declare no competing financial interests.

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115439, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. This publication reflects only the author's views and neither the IMI JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein.

J.V. is supported by the Monument Trust Discovery Award from Parkinson's UK. The Wellcome Trust (WTISSF121302) and The Oxford Martin School (LC0910-004) provide financial support to the James Martin Stem Cell Facility, including work with HUES cell lines. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics, Oxford (Funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G0900747 91070) for the generation of Illumina genotyping and transcriptome data.

We thank Christine Klein and Philip Seibler, University of Lubec, for provision of the SF826 fibroblasts. We thank David Piper and Rex Lacambacal, Thermo Fisher Scientific, for performing the ScoreCard experiment.

REFERENCES

- (1) Takahashi, K.; Yamanaka, S. *Cell* 2006, 126 (4), 663–676.
- (2) Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K.; Yamanaka, S. *Cell* 2007, 131 (5), 861–872.
- (3) Yu, J.; Hu, K.; Smuga-otto, K.; Tian, S.; Stewart, R.; Igor, I.; Thomson, J. A. *Science* (80-.), 2009, 324 (5928), 797–801.
- (4) Fusaki, N.; Ban, H.; Nishiyama, A.; Saeki, K.; Hasegawa, M. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* 2009, 85 (8), 348–362.
- (5) Nishimura, K.; Sano, M.; Ohtaka, M.; Furuta, B.; Umemura, Y.; Nakajima, Y.; Ikehara, Y.; Kobayashi, T.; Segawa, H.; Takayasu, S.; Sato, H.; Motomura, K.; Uchida, E.; Kanayasu-Toyoda, T.; Asashima, M.; Nakauchi, H.; Yamaguchi, T.; Nakanishia, M. *J. Biol. Chem.* 2011, 286 (6), 4760–4771.
- (6) Hockemeyer, D.; Soldner, F.; Cook, E. G.; Gao, Q.; Mitalipova, M.; Jaenisch, R. 2008, 346–353.
- (7) Woltjen, K.; Michael, I. P.; Mohseni, P.; Desai, R.; Mileikovsky, M.; Hämmäläinen, R.; Cowling, R.; Wang, W.; Liu, P.; Gertsenstein, M.; Kaji, K.; Sung, H.-K.; Nagy, A. *Nature* 2009, 458 (April), 766–770.
- (8) Kim, D.; Kim, C.; Moon, J.; Chung, Y.; Chang, M.; Han, B.; Ko, S.; Yang, E.; Cha, K. Y.; Lanza, R.; Kim, K. *Stem Cell* 2009, 4 (6), 472–476.
- (9) Miyoshi, N.; Ishii, H.; Nagano, H.; Haraguchi, N.; Dewi, D. L.; Kano, Y.; Nishikawa, S.; Tanemura, M.; Mimori, K.; Tanaka, F.; Saito, T.; Nishimura, J.; Takemasa, I.; Mizushima, T.; Ikeda, M.; Yamamoto, H.; Sekimoto, M.; Doki, Y. *Stem Cell* 2011, 8 (6), 633–638.
- (10) Buta, C.; David, R.; Dressel, R.; Emgård, M.; Fuchs, C.; Gross, U.; Healy, L.; Hescheler, J.; Kolar, R.; Martin, U.; Mikkers, H.; Müller, F.; Schneider, R. K.; Seiler, A. E. M.; Spielmann, H.; Weitzer, G. *Stem Cell Res.* 2013, 11 (1), 552–562.

- (11) Müller, F.-J.; Schuldt, B. M.; Williams, R.; Mason, D.; Altun, G.; Papapetrou, E. P.; Danner, S.; Goldmann, J. E.; Herbst, A.; Schmidt, N. O.; Aldenhoff, J. B.; Laurent, L. C.; Loring, J. F. *Nat. Methods* 2011, 8 (4), 315–317.
- (12) Bock, C.; Kiskinis, E.; Verstappen, G.; Gu, H.; Boulting, G.; Smith, Z. D.; Ziller, M.; Croft, G. F.; Amoroso, M. W.; Oakley, D. H.; Gnirke, A.; Eggan, K.; Meissner, A. *Cell* 2011, 144 (3), 439–452.
- (13) Fernandes, H. J. R.; Hartfield, E. M.; Christian, H. C.; Emmanouilidou, E.; Zheng, Y.; Booth, H.; Bogetofte, H.; Lang, C.; Ryan, B. J.; Sardi, S. P.; Badger, J.; Vowles, J.; Evetts, S.; Tofaris, G. K.; Vekrellis, K.; Talbot, K.; Hu, M. T.; James, W.; Cowley, S. A.; Wade-Martins, R. *Stem Cell Reports* 2016, 6 (3), 342–356.
- (14) Craig, R.; Cortens, J. P.; Beavis, R. C. *J. Proteome Res.* 2004, 3 (6), 1234–1242.
- (15) Deutsch, E. W.; Lam, H.; Aebersold, R. *EMBO Rep.* 2008, 9 (5), 429–434.
- (16) Manwaring, V.; Heywood, W. E.; Clayton, R.; Lachmann, R. H.; Keutzer, J.; Hindmarsh, P.; Winchester, B.; Heales, S.; Mills, K. J. *Proteome Res.* 2013, 12 (5), 2013–2021.
- (17) Tsankov, A. M.; Akopian, V.; Pop, R.; Chetty, S.; Gifford, C. A.; Daheron, L.; Tsankova, N. M.; Meissner, A. *Nat. Biotechnol.* 2015, 33 (11), 1–15.
- (18) Lin, S. M.; Du, P.; Huber, W.; Kibbe, W. A. *Nucleic Acids Res.* 2008, 36 (2), 1–9.
- (19) Du, P.; Kibbe, W. A.; Lin, S. M. *Bioinformatics* 2008, 24 (13), 1547–1548.
- (20) Cowan, C. A.; Klimanskaya, I.; McMahon, J.; Atienza, J.; Witmyer, J.; Zucker, J. P.; Wang, S.; Morton, C. C.; McMahon, A. P.; Powers, D.; Melton, D. A. *N. Engl. J. Med.* 2004, 350 (13), 1353–1356.
- (21) Lenz, M.; Goetzke, R.; Schenk, A.; Schubert, C.; Veeck, J.; Hemeda, H.; Koschmieder, S.; Zenke, M.; Schuppert, A.; Wagner, W. *Sci. Rep.* 2015, 5, 8973.
- (22) Israel, B.; Medical, D. *Stem Cells* 2007;25: 723–730.
- (23) Quintanilla, R. H.; Asprer, J. S. T.; Vaz, C.; Tanavde, V.; Lakshminpathy, U. *PLoS One* 2014, 9 (1).
- (24) Sridharan, R.; Tchiew, R.; Mason, M. J.; Yachechko, R.; Kuoy, E. *Cell* 2009, 136 (2), 364–377.
- (25) Vogel, C.; Marcotte, E. M. *Nat. Rev. Genet.* 2012, 13 (4), 227–232.
- (26) Vogel, C.; Abreu, R. de S.; Ko, D.; Le, S.-Y. Y.; Shapiro, B. A.; Burns, S. C.; Sandhu, D.; Boutz, D. R.; Marcotte, E. M.; Penalva, L. O. *Mol. Syst. Biol.* 2010, 6 VN - re (400), 400.
- (27) Brumbaugh, J.; Russell, J. D.; Yu, P.; Westphall, M. S.; Coon, J. J.; Thomson, J. A. *Stem Cell Reports* 2014, 2 (1), 18–25.
- (28) Ebhardt, H. A.; Root, A.; Sander, C.; Aebersold, R. *Proteomics* 2015, 15 (18), 3193–3208.
- (29) Carr, S. a; Abbatiello, S. E.; Ackermann, B. L.; Borchers, C.; Domon, B.; Deutsch, E. W.; Grant, R. P.; Hoofnagle, A. N.; Hüttenhain, R.; Koomen, J. M.; Liebler, D. C.; Liu, T.; MacLean, B.; Mani, D. R.; Mansfield, E.; Neubert, H.; Paulovich, A. G.; Reiter, L.; Vitek, O.; Aebersold, R.; Anderson, L.; Bethem, R.; Blonder, J.; Boja, E.; Botelho, J.; Boyne, M.; Bradshaw, R. a; Burlingame, A. L.; Chan, D.; Keshishian, H.; Kuhn, E.; Kinsinger, C.; Lee, J. S. H.; Lee, S.-W.; Moritz, R.; Oses-Prieto, J.; Rifai, N.; Ritchie, J.; Rodriguez, H.; Srinivas, P. R.; Townsend, R. R.; Van Eyk, J.; Whiteley, G.; Wiita, A.; Weintraub, S. *Mol. Cell. Proteomics* 2014, 13, 907–917.
- (30) Heywood, W. E.; Galimberti, D.; Bliss, E.; Sirka, E.; Paterson, R. W.; Magdalinou, N. K.; Carecchio, M.; Reid, E.; Heslegrave, A.; Fenoglio, C.; Scarpini, E.; Schott, J. M.; Fox, N. C.; Hardy, J.; Bahtia, K.; Heales, S.; Sebire, N. J.; Zetterburg, H.; Mills, K. *Mol. Neurodegener.* 2015, 10 (1), 64.
- (31) Yamana, R.; Iwasaki, M.; Wakabayashi, M.; Nakagawa, M.; Yamanaka, S.; Ishihama, Y. *J. Proteome Res.* 2013, 12, 214–221.
- (32) Phanstiel, D. H.; Brumbaugh, J.; Wenger, C. D.; Tian, S.; Probasco, M. D.; Bailey, D. J.; Swaney, D. L.; Tervo, M. a; Bolin, J. M.; Ruotti, V.; Stewart, R.; Thomson, J. a; Coon, J. J. *Nat. Methods* 2011, 8 (10), 821–827.
- (33) Munoz, J.; Low, T. Y.; Kok, Y. J.; Chin, A.; Frese, C. K.; Ding, V.; Choo, A.; Heck, A. J. R. *Mol. Syst. Biol.* 2011, 7 (550), 1–13.

(34) Benevento, M.; Tonge, P. D.; Puri, M. C.; Hussein, S. M. I.; Cloonan, N.; Wood, D. L.; Grimmond, S. M.; Nagy, A.; Munoz, J.; Heck, A. J. R. Nat. Commun. 2014, 5, 1–11.

Insert Table of Contents artwork here

