A quantitative comparison of human embryonic and induced stem cell proteomes

Alejandro J. Brenes[1,2], Eva J. Griesser[1,3], Linda V. Sinclair[2], Hao Jiang[1], Harunori Yoshikawa[1,4], Lindsay Davidson[5], Melpomeni Platani[1], Jason Swedlow[1], Doreen A. Cantrell[2] & Angus I. Lamond[1*]*

[1] Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dow St, Dundee, DD1 5EH, United Kingdom

[2] Cell Signalling & Immunology, School of Life Sciences, University of Dundee, Dow St, Dundee, DD1 5EH, United Kingdom

[3] Present address: Drug Discovery Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

[4] Present address: Division of Cell Signalling, Fujii Memorial Institute of Medical Sciences, Institute of Advanced Medical Sciences, Tokushima University, 3-18-15 Kuramoto, Tokushima, 770-8503, Japan

[5] Human Pluripotent Stem Cell Facility, School of Life Sciences, University of Dundee, Dow St, Dundee DD1 5EH, United Kingdom

[*]correspondence to: a.i.lamond@dundee.ac.uk, ajbrenesmurillo@dundee.ac.uk

KEYWORDS: proteomics, iPSC, mass spectrometry, hESC, TMT, stem cells

Abstract

Human induced pluripotent stem cells (hiPSCs) have great potential to be used as alternatives to embryonic stem cells (hESCs) in regenerative medicine and disease modelling. However, a clear overview of their differences at the protein level is still incomplete. In this study we characterise the proteomes of hiPSC and hESC lines, where we find that they express a similar set of proteins but show consistent quantitative differences that can be masked by the normalisation methods. hiPSCs have a higher protein content, with over 1,500 proteins showing over two-fold increased expression. They also display proteomic differences in their mitochondria, with increased expression of mitochondrial transporters and metabolic proteins as well as mitochondrial translation machinery. The hiPSCs also show higher expression of important amino acid transporters, secreted proteins, and growth factors with potential to affect neighbouring cells, coupled with a systematic reduction in the expression levels of H1 histone variants. We conclude that despite hiPSCs and hESCs being highly similar cell types, they show important differences in protein expression that may be relevant for their use in clinical research.

Introduction

Human embryonic stem cells (hESC) are derived from the inner cell mass of a pre-implantation embryo[1]. They show prolonged undifferentiated potential, as well as the ability to differentiate into the three main embryonic germ layers[2], making them excellent models for studying disease mechanisms, development and differentiation. However, their use remains restricted by regulations, based in part upon ethical considerations[3].

Over a decade ago, methods allowing the induction of pluripotent stem cells from fibroblast cultures, in both human and mice, were developed[4,5]. These reports showed that by exogenously expressing a small set of key transcription factors (Oct4, Sox2, c-Myc and Klf4), a somatic cell could be reprogrammed back into a pluripotent state, characterised by their capacity for self-renewal and ability to differentiate into the three main germ layers. These human induced pluripotent stem cells (hiPSCs) show many key features of their

physiological embryonic stem cell (hESC) counterparts, while avoiding many of the ethical issues regarding the use of stem cells derived from embryos.

Since the discovery of reprogramming methods, hiPSC lines have attracted great interest, particularly for their potential use as alternatives to hESCs in regenerative medicine[6] and disease modelling, including studies on monogenic disorders[7,8], and some late onset diseases[9]. However, to understand the value of using hiPSCs in clinical research, drug development and/or studies of disease mechanisms, it is important to establish how similar hiPSCs are to hESCs at the molecular and functional levels. To address this, multiple studies have compared hiPSCs and hESCs, using a variety of assays, including methylation analysis[10], transcriptomics[11,12] and even quantitative proteomics[13]. It should be noted, however, that these earlier studies were mostly performed at a time when reprogramming protocols were less robust[14] and when the depth of proteome coverage and quantitative information that could be obtained was lower than today.

In this study, we have addressed the similarity of hiPSCs to hESCs by performing a detailed proteomic analysis, comparing a set of hiPSC lines derived from human primary skin fibroblasts[15] of independent, healthy donors, with several independent hESC lines. The data highlight that while both types of stem cell lines have similar global protein abundance profiles, they also show some specific and significant quantitative differences in protein expression. In particular, the reprogramed iPSC lines consistently display higher total protein levels, predominantly reflecting higher expression of a subset of protein types, including ribosomal, mitochondrial, secreted proteins and amino acid transporters.

Results

Quantitative analysis

A mass spectrometry-based strategy, involving MS3-based, synchronous precursor selection[16] (SPS) tandem mass tagging (TMT)[17], was used to characterise the proteomes of independent sets of hESC and hiPSC lines derived from different donors within a single 10-plex. To optimise quantification accuracy, each sample was allocated to a specific isobaric tag in a way that minimised cross-population reporter ion interference (Fig.1a), as previously described[18]. In total 8,491 protein groups (henceforth referred to as ‘proteins’), were detected at 1% FDR, with >99% overlap between the proteins detected from both the hESC and hiPSC lines (Fig. 1b).

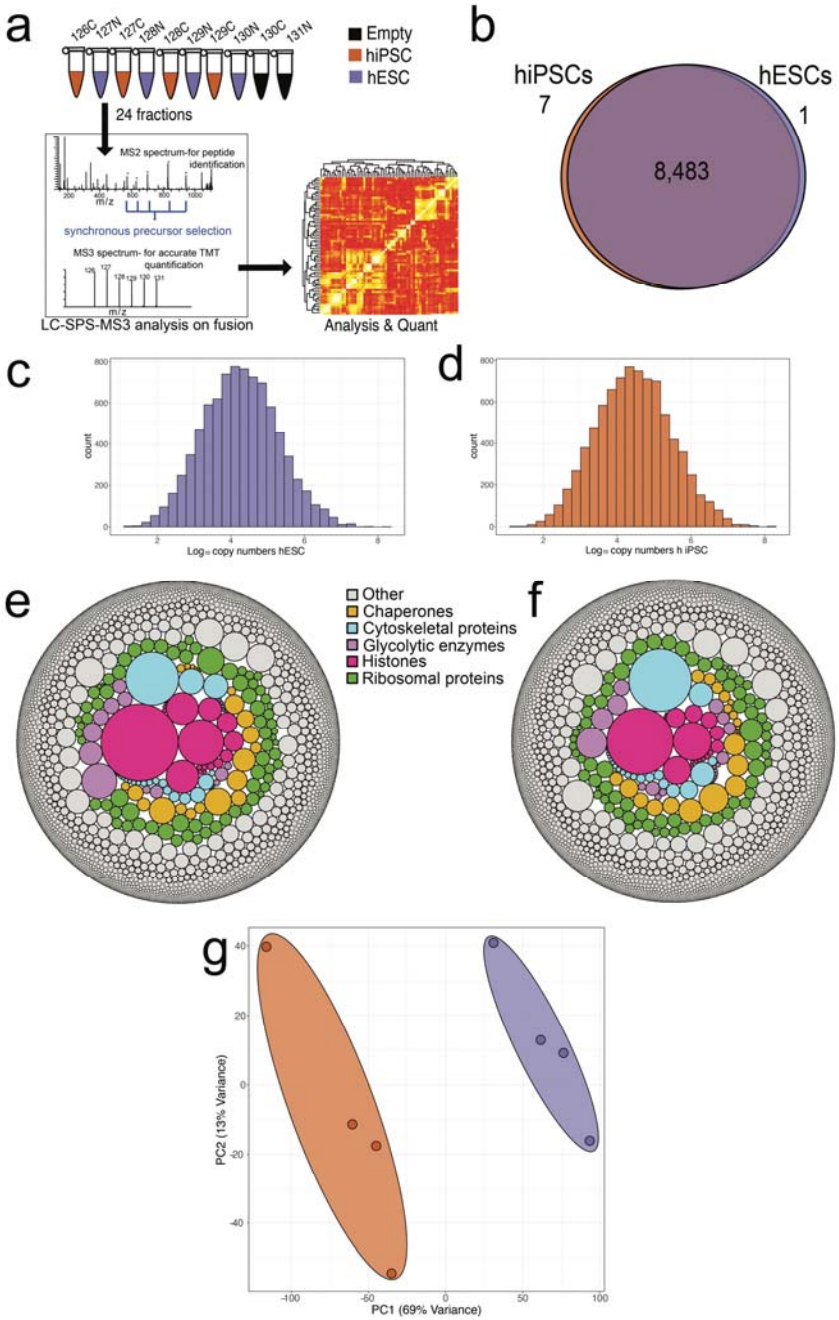

Figure 1- Proteomic overview: (a) Diagram showing the SPS-MS3 TMT proteomic workflow used for the experiment. **(b)** Venn diagram showing the overlap of proteins identified within the hiPSC and hESC populations. **(c)** Average copy number histogram for the hESCs. **(d)** Average copy number histogram for the hiPSCs. **(e)** Bubble plot showing proteins coloured by specific categories where the size is represented by the average hESC estimated protein copy numbers. **(f)** Bubble plot showing proteins coloured by specific categories where the size is represented by the average hiPSC estimated protein copy numbers. **(g)** PCA plot based on the $\log_{10}$ copy numbers for all 8 replicates. hESCs are shown in purple and hiPSCs in orange.

To provide a quantitative comparison of the respective hESC and hiPSC proteomes, we focussed on analysing the 7,878 proteins that were detected with at least 2

unique and razor peptides. Protein copy numbers were estimated via the “proteomic ruler”[19], which revealed that both the hESC (Fig. 1c) and hiPSC (Fig. 1d) proteomes display a similar wide dynamic range, with estimated protein copy numbers extending from a median of less than 100 copies, to over 100 million copies per cell. Furthermore, the composition of the respective hiPSC and hESC proteomes also appear very similar. Both populations display high expression levels of ribosomal proteins, protein chaperones and glycolytic enzymes (Fig. 1e&f), consistent with both being primed pluripotent stem cells, which are heavily dependent on glycolysis for energy generation[20].

It is only when the quantitative data are examined in more detail that differences between the cell types become apparent (Fig. 1g). A principal component analysis (PCA), based on the protein copy numbers, revealed a clear separation between the two stem cell populations within the main component of variation, which accounted for 69% of variance. The PCA clearly showed that the independent iPSC lines were more similar to each other than to any of the hESC lines, and vice versa.

Protein content differences masked by data normalisation

To assess potential population-scale effects, we next compared the hESC and hiPSC proteomes using two different normalisation approaches, along with differential expression analysis. First, a concentration-based approach was used, as typically applied to proteomics datasets (see methods). It should be noted that, unlike protein copy number estimates, this normalisation strategy does not account for potential changes in either cell size, or total protein content, between the

populations being compared. Using this concentration based methodology (Fig. 2a), no major differences in protein expression were detected between the hESC and hiPSC lines, i.e. with no significant changes seen for ~95% of all proteins (see methods), consistent with previous reports[13].

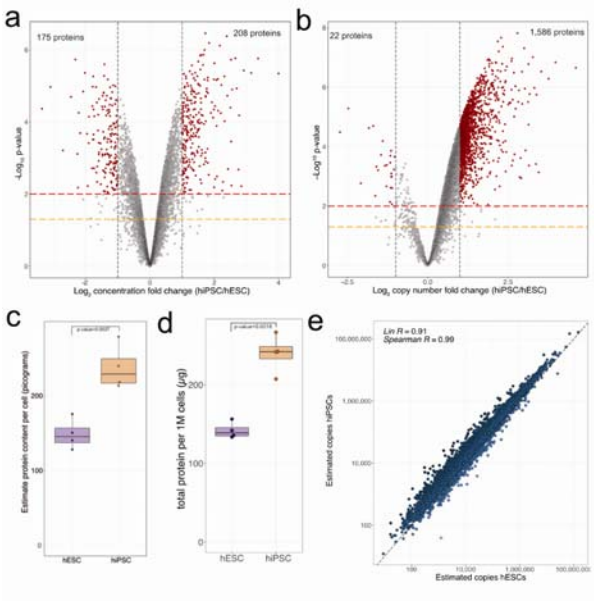

Figure 2 Protein content and translation related: (a) Concentration based volcano plot showing the -$\log_{10}$ p-value and the $\log_2$ fold change comparing hiPSCs to hESCs. Elements shaded in red are considered significantly changed. All dots above the orange line have a p-value lower than 0.05 and all dots above the red line have a p-value lower than 0.01 **(b)** Copy number-based volcano plot showing the -$\log_{10}$ p-value and the $\log_2$ fold change comparing hiPSCs to hESCs. Elements shaded in red are considered significantly changed. All dots above the orange line have a p-value lower than 0.05 and all dots above the red line have a p-value lower than 0.01 **(c)** Box plot showing the MS based estimated protein content for all hESCs and hiPSCs. **(d)** Box plot showing the protein amount per million cells derived from the EZQ Protein Quantification Kit for all hESCs and hiPSCs **(e)** Scatter plot showing the median estimated copy numbers for both hiPSCs and hESCs along with the corresponding Lin and Spearman correlation coefficients

However, when a differential expression analysis is based on protein copy numbers instead (Fig. 2b), systematic differences between the two populations were distinguished. Thus, 20% (1,587/7,878) of all proteins detected showed over two-fold higher expression in hiPSCs than in hESCs (Fig. 2e, p-value < 0.01). In contrast, only 22 proteins (0.3%) showed significantly lower expression levels in hiPSCs. Estimations of the total protein content for both populations, based upon the MS data, indicated that hiPSCs have a median increase of 57% in total protein content,

compared to hESCs (Fig. 2c). To validate this observation, an independent assay (EZQ^{TM} assay; see methods), was used to measure the total protein yield from similar numbers of freshly grown hiPSC and hESC cells. From these experiments, the calculated protein amount per million cells was 74% higher (Fig. 2d; p-value=0.0018) in hiPSC cells, relative to hESCs.

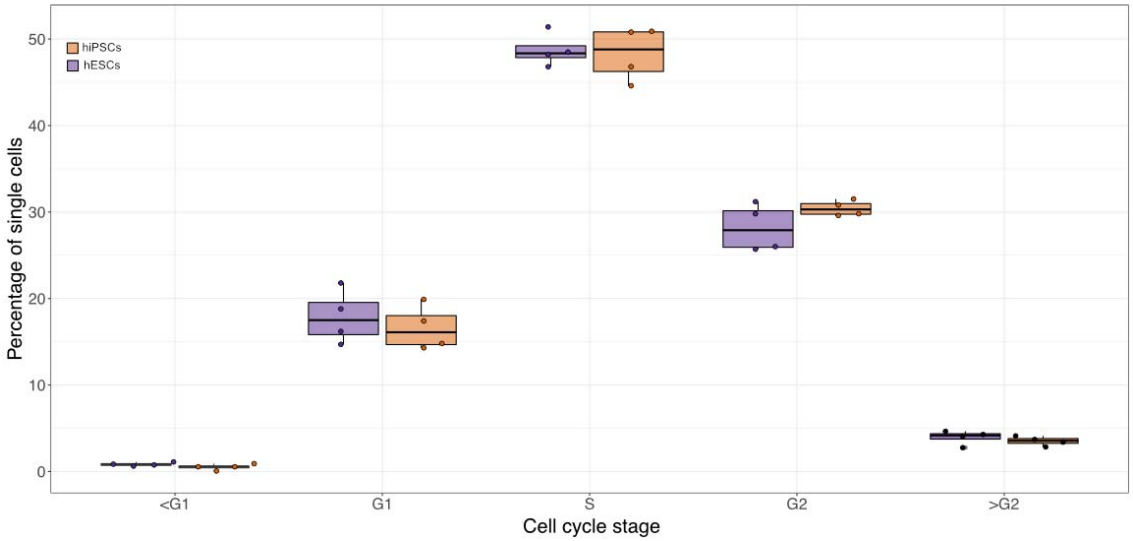

Figure EV1 – Cell cycle overview: (a) Box plot showing the percentage of single cells (FACS derived) across the different cell cycle stages for hESCs and hiPSCs.

Next, to check if these differences in total protein content reflected differences in cell cycle distributions between the two populations, FACS analyses were performed. This showed no significant differences in the percentage of cells at each cell cycle stage between the hiPSC and hESC lines (Fig. EV1). We conclude that there is a consistent difference in total protein expression between hiPSCs and hESCs, independent of cell cycle effects.

To explore further the similarities and differences between the respective hiPSC and hESC proteomes, we next compared Spearman and Lin's concordance correlation coefficients. The Spearman rank correlation was used to compare protein ranking

based on the estimated copy numbers in both populations (Fig. 2e). This showed a correlation coefficient of 0.99, demonstrating that the protein expression profiles are nearly identical between the two populations. However, Lin's correlation coefficient, which measures the degree of agreement between two populations, essentially evaluating if the results are identical[21], was notably lower, at 0.91 (Fig. 2e).

These data indicate that hiPSCs and hESCs have very similar rank profiles, i.e., the most abundant proteins are essentially the same in both populations. However, there are nonetheless quantitative differences between the hiPSC and hESC proteomes.

Subcellular Proteome scaling

Having detected many proteins that were significantly increased in expression in hiPSCs compared with hESCs, we next checked specifically whether this included key primed pluripotency markers (Fig. 3a). The data showed no significant differences (p-value <0.01) in expression of SOX2, NANOG (detected as NANOGP8), and OCT4 (detected as POU5F1), across the independent hiPSC and hESC lines (Fig. 3a).

To test whether the protein abundance difference was related to phenotypic variations between hiPSCs and hESCs in one or more specific subcellular compartments, we used an overrepresentation analysis (ORA), of the cellular compartments using WebGestalt[22]. The analysis focussed on all the proteins showing significantly increased expression in hiPSCs, compared to hESCs (see methods) and it showed the highest enrichment for organelle and plasma membrane related localisations, proteins localised to the Golgi apparatus and proteins in pre-ribosomes (Fig. 3b).

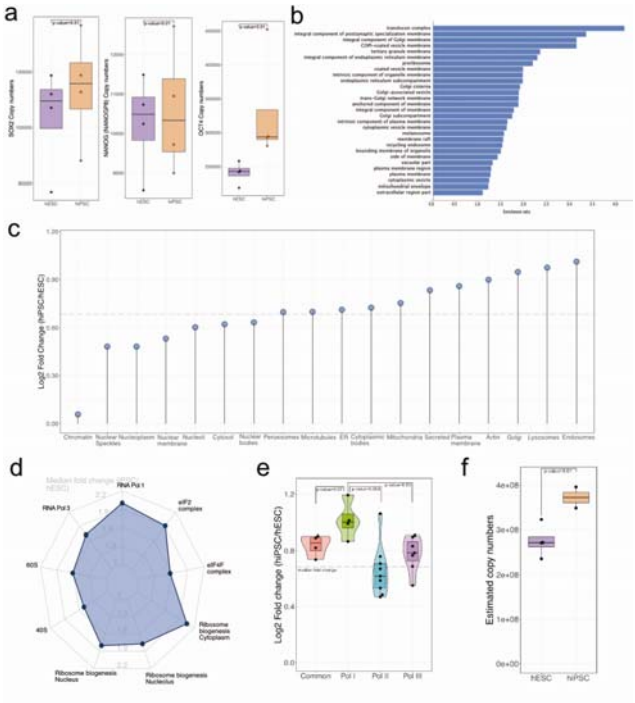

Figure 3- Subcellular enrichment: (a) Boxplots showing the protein copy numbers for NANOG, SOX2 and OCT4 within iPSCs and ESCs. **(b)** Bar chart showing the results of the cellular compartment enrichment analysis, all terms have an FDR < 0.05 **(c)** Lollipop plot showing the median fold change for different subcellular compartments, the gray dotted line represents the median fold change of all proteins **(d)** Radar plot showing the median fold change (iPSC/ESC) for protein categories which are related to the pre-ribosomes **(e)** Boxplot showing the log2 fold change (iPSC/ESC) for all the subunits of the different RNA polymerases classified by the complex they belong to **(f)** Boxplot showing the copy numbers for the cytoplasmic ribosomes in both iPSCs and ESCs.

To characterise the subcellular compartment protein scaling, total copy numbers for all proteins in each compartment were compared between the hiPSC and hESC lines. Interestingly, the data showed that nuclear proteins had the highest similarities between the cell types, with chromatin-associated proteins being virtually unchanged between hiPSCs and hESCs (Fig. 3c). However, the data showed higher than median fold changes for mitochondrial, secreted, plasma membrane, Golgi, lysosome and endosome associated proteins.

We next focussed on the only non-membrane related ‘compartment’ that was enriched in the ORA, the pre-ribosome. Hence, we next looked at proteins directly related to ribosome subunit biogenesis and associated processes, such as pre-rRNA

processing (Fig. 3d). Proteins linked with ribosome subunit biogenesis, as defined by Kegg[23], showed considerable increases in expression in hiPSC, as compared with hESC lines.

For example, RNA Polymerase I subunits (responsible for transcription of rRNA genes), showed the highest increase in hiPSCs. In contrast, subunits unique to RNA Polymerase II, which is responsible for transcription of mRNA, snRNA and microRNA genes, showed a more modest increase (Fig. 3e). And finally, the end point of ribosome biogenesis, the ribosomal proteins themselves, also had significantly higher expression in hiPSCs compared to hESCs (Fig. 3f). These specific observations can account, at least in part, for the higher protein content detected in hiPSCs.

Upscaling of mitochondrial translation and ribosomes

Next, another ORA was performed, this time focused on biological processes, rather than cellular compartments. This revealed a clear enrichment in specific terms relating predominantly to mitochondrial translation, transmembrane transport, extracellular structure and rRNA metabolic process (Fig. 4a). The highest enrichment score was seen for proteins involved in mitochondrial translation, with nearly all proteins in the pathway showing significantly increased expression (p-value <0.01) in iPSCs. Specifically, all proteins involved in mitochondrial pre-rRNA processing, translation initiation, translation elongation, together with 66% of the proteins involved in translation termination, all had higher copy numbers per cell in hiPSCs (Fig. 4b). Mitochondrial ribosome proteins were also increased in expression by 74% (Fig. 4c), with a significantly altered ratio of small to large subunit proteins (Fig. 4d).

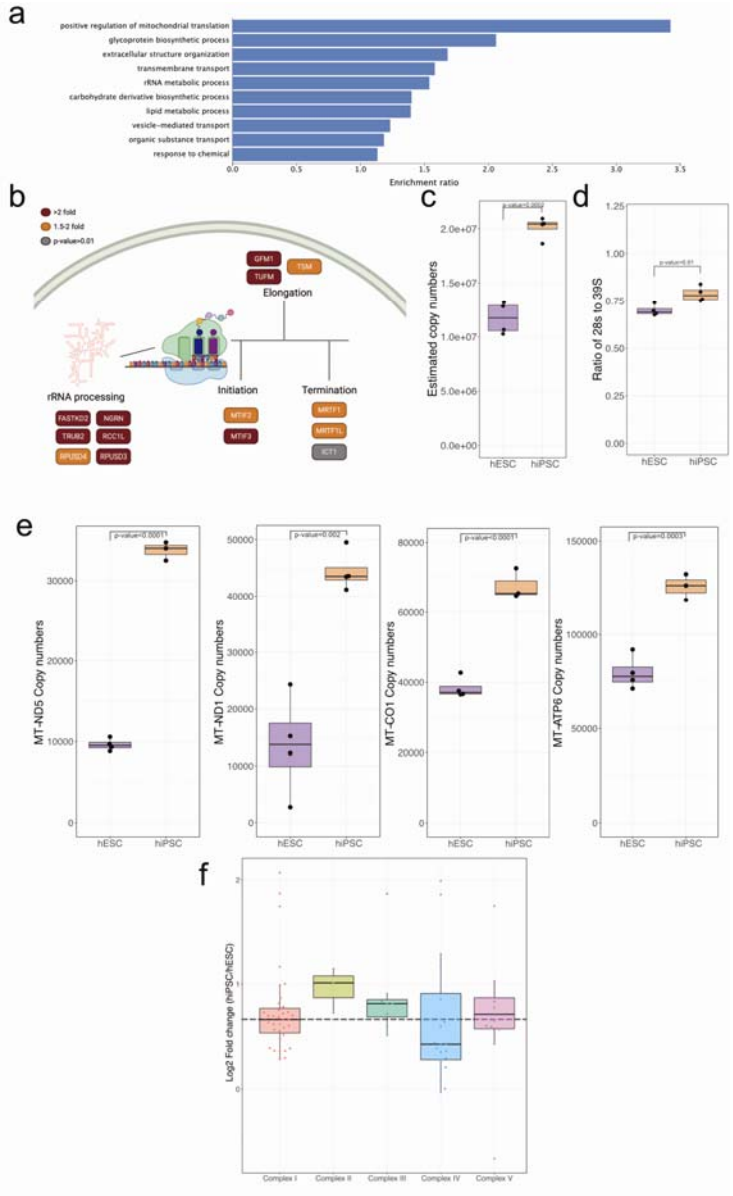

Figure 4- Mitochondrial translation: (a) Barplot showing the results of the biological process enrichment analysis **(b)** Schematic showing the log2 fold change (iPSC/ESC) for all proteins related to mitochondrial translation **(c)** Boxplot showing the estimated copy numbers for the mitochondrial ribosomes for both iPSCs and ESCs. **(d)** Boxplot showing the ratio of small subunit (28S) to large subunit (39S) proteins for both iPSCs and ESCs **(e)** Boxplot showing the log2 fold change (iPSC/ESC) for all the subunits of the different RNA polymerases classified by the complex they belong to **(f)** Boxplot showing the copy numbers for the cytoplasmic ribosomes in both iPSCs and ESCs.

Mitochondrial ribosomes translate proteins that are encoded on the mitochondrial genome[24]. Of these mitochondrial encoded proteins, 4 out of 5 that were quantified showed significantly increased expression in hiPSCs and with very high fold changes (Fig. 4e). These proteins are important subunits of the electron transport chain

(ETC). Further analysis of ETC complexes showed that all 5 complexes were also significantly increased in hiPSCs (Fig. 4f), with the highest median fold change seen in complex II, succinate dehydrogenase, which also functions as part of the tricarboxylic acid cycle (TCA).

Increased expression of mitochondrial and amino acid transporters

Another biological process that was highly enriched in the ORA analysis was 'transmembrane transport'. To study its component members in more detail, we focussed on the solute carrier (SLC) family of transporters, highlighting the 15 SLC family members showing the highest fold change (fold change >3 & with a p-value <0.01). These proteins clustered in two main categories, i.e. plasma membrane amino acid transporters and mitochondrial transporters (Fig. 5a).

Amongst the mitochondrial family of SLC transporters, while they vary considerably in total abundance (Fig. 5b), all 26 showed significantly increased expression in hiPSCs. This included SLC25A6, which is the most abundant member of the family and represents the main ATP/ADP transporter for both hiPSCs and hESCs. Notably, SLC25A6 was present in almost 11 million copies per cell in iPSCs, representing an increase in expression of ~83% over hESCs.

Furthermore, 11/12 of the main cellular amino acid transporters located in the plasma membrane also showed significantly increased expression in hiPSCs. This included the most abundant transporter SLC3A2 (Fig. 5c), which is a subunit of several heterodimeric amino acid transporter complexes, whose substrates vary

according to the specific subunits within the complex. We note that the 3 main glutamine transporters[25,26], i.e., SLC38A1, SLC38A2 and SLC1A5, were all increased by >2-fold in hiPSCs (Fig. 5d). This suggested that hiPSCs have higher potential capacity for glutamine transport, compared to hESCs.

To test this hypothesis experimentally, the uptake of radio-labelled glutamine was measured for both hiPSCs and hESCs (see methods). These data showed that hiPSCs had a median 93% higher uptake of glutamine, compared to hESCs (Fig. 5e). This experiment provides independent confirmation of the functional significance of the quantitative protein expression data determined by MS analysis and supports one of the key conclusions concerning phenotypic and functional differences between hiPSC and hESC lines.

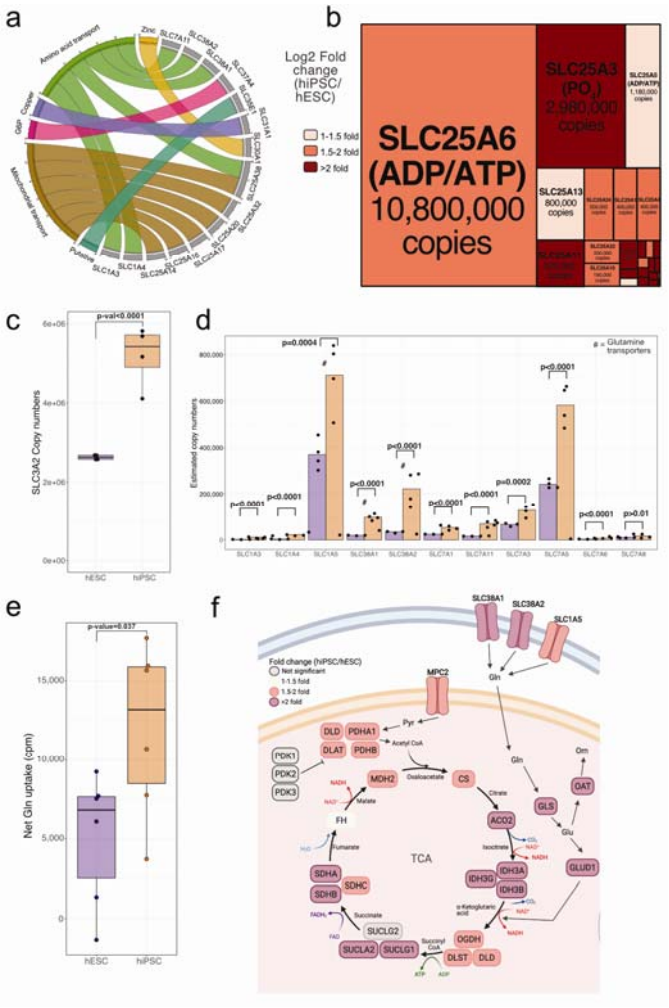

Figure 5- Mitochondrial and amino acid transporters: (a) Chord diagram showing the 15 transporters with highest fold change along with their protein category **(b)** Treemap showing all mitochondrial transporters, the size is representative of the copy numbers in iPSCs and the colour represent the log2 fold change (iPSC/ESC) **(c)** Boxplot showing the estimated copy numbers for SLC3A2 in both iPSCs and ESCs **(d)** Barplot showing the copy numbers of the main amino acid transporters **(e)** Boxplot showing the net glutamine uptake in both iPSCs and ESCs. **(f)** Schematic showing the log 2 fold change (iPSC/ESC) for all proteins involved in the glutaminolysis and TCA pathways.

To explore the potential consequences of the increased glutamine uptake in hiPSCs, we focused on the glutamine catabolism pathway. Both the GLS and GLUD1 proteins show over two-fold higher expression in hiPSCs, compared to hESCs, with a p-value <0.01. Glutaminolysis has been shown to be vital for human PSCs as it can provide ATP via the TCA[27], as well as the aforementioned biosynthetic

precursors required to sustain growth and proliferation. Our data show most TCA-related enzymes are increased in expression >1.5-fold in hiPSCs, compared with hESCs, with the biggest change seen for the isocitrate dehydrogenase 3 and succinate dehydrogenase complexes (Fig. 5f).

Secreted proteins and extracellular matrix

Amongst secreted proteins and proteins related to extracellular matrix organisation, multiple growth factors of relevance to primed pluripotent stem cells showed significantly increased expression in hiPSCs, including FGF2, FGF1, TGFB1 and NODAL. FGF and Activin/Nodal, which have important roles in differentiating cells, are also vital components of signalling pathways that maintain pluripotency within human primed stem cells[28-30]. Both show >3-fold higher expression in hiPSCs, compared to hESCs (Fig. 6a).

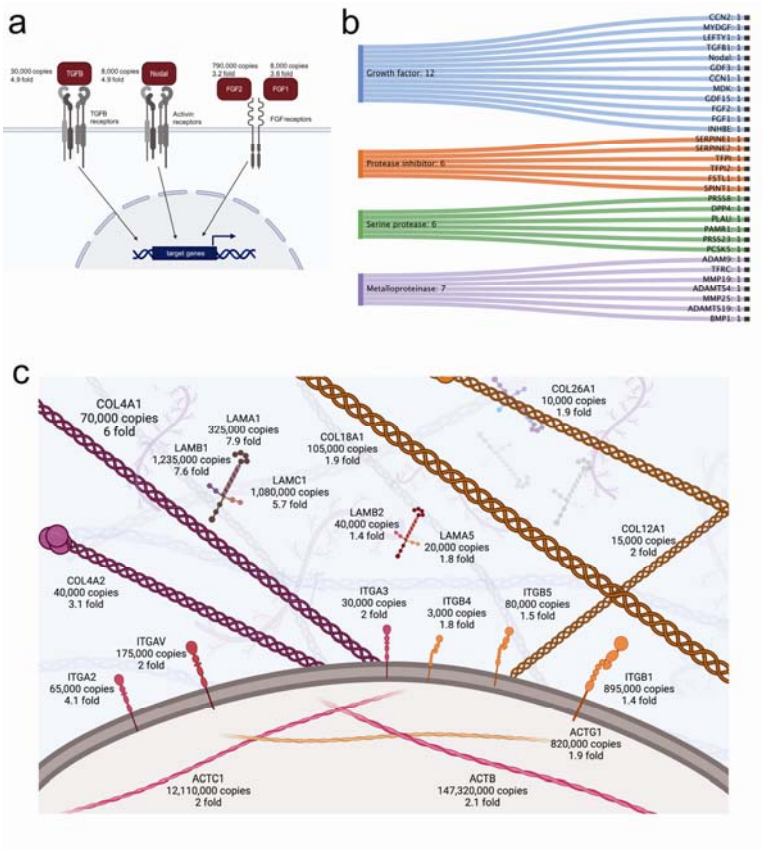

Figure 6 – Extracellular matrix related proteins: (a) Diagram showing the copy numbers and fold change (hiPSC/hESC) of vital growth factors for human primed pluripotent stem cells. **(b)** Sankey diagram showing the main class of proteins represented within the secreted proteins which were significantly increased in expression **(c)** Schematic showing ECM associated proteins along with their copy numbers in iPSCs as well as their fold change (iPSC/ESC).

A further 8 growth factors were identified as having significantly increased expression in iPSCs, including MYDGF, which is reported to promote cardiac protection[31] and MDK, which has been reported to promote inflammation by recruiting macrophages and neutrophils[32]. The data also highlighted 6 protease inhibitors, some linked to thrombosis, like TFPI[33], along with 6 serine proteases and 7 metalloproteinases (Fig. 6b).

Focussing then on the extracellular matrix (ECM) proteins and those known to interact with its components, hiPSCs showed the highest increase for type IV collagen (COL4A1 and COL4A2), along with alpha integrins (ITGA2 and ITGAV) (Fig 6c). Increased expression was also seen for the most abundant laminins (LAMA1,

LAMB1 and LAMC1), which are all part of the Laminin 111 complex[34]. This complex has been reported to interact with Actin, of which all 3 main proteins (ACTB, ACTC1, ACTG1) are significantly increased in expression, to mediate growth and quiescence[35].

Histone variants

Very few proteins (<1%; 52/7,878) showed decreased expression (p-value<0.01) within hiPSCs, in comparison with hESCs. An ORA showed that the proteins decreased in abundance were enriched for GO terms related to DNA recombination, nucleosome positioning and chromatin silencing (Fig. 7a). Notably, this included four Histone H1 proteins. Histone H1 proteins are linker histones, which do not form part of the core histone octamer, but instead sit on top of the nucleosome and bind DNA entry and exit sites (Fig. 7b). They have been linked with influencing nucleosomal repeat length[36] and stabilising chromatin structures[37].

Our data show that the most abundant variant in hESCs, HIST1H1E, which is present at almost 25 million copies per cell, is significantly decreased in expression in hiPSCs to just over 7.5 million copies per cell. A similar case was also seen for HIST1H1C, HIST1H1D and H1FX, which were all decreased in expression within hiPSCs compared to hESCs (Fig. 7c).

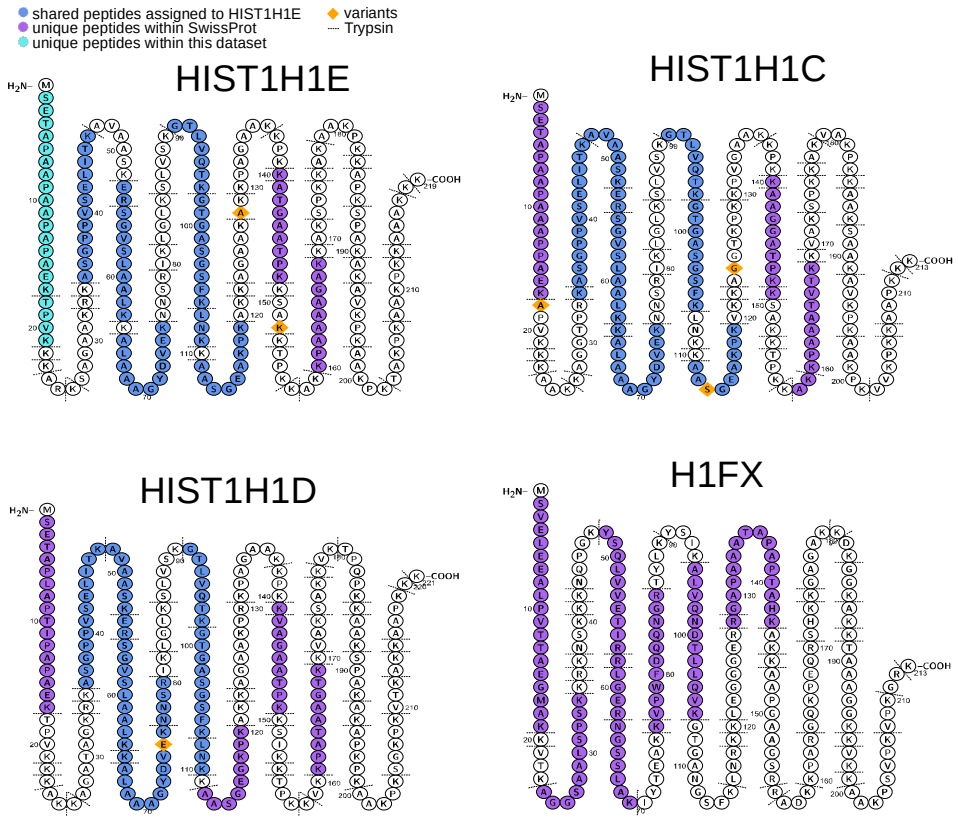

Figure EV2 – Unique and shared peptides for the H1 histones: Schematic showing the unique and shared peptides detected for HISTH1E, HIST1H1C, HIST1H1D and H1FX

These specific histone variants have high sequence similarity; hence some peptide sequences will be unique to each protein, but some will be shared by many. HIST1H1E, HIST1H1D and HIST1H1C have 75% sequence identity. Thus, while unique peptides were identified for all 3 proteins, a pool of shared peptides were also identified that could belong to all 3 of them (Fig. EV2). Due to limitations in how the Andromeda algorithm[38] assigns peptides to a protein, all of the shared peptides were assigned to HIST1H1E, potentially distorting the abundance of the 3 histones by overestimating HIST1H1E and underestimating HIST1H1D and HIST1H1C. Despite the potential issues in reliably estimating total abundance, the data for both the

unique and shared peptides is consistent and supports that all three of these H1 variant proteins were reduced in expression in iPSCs.

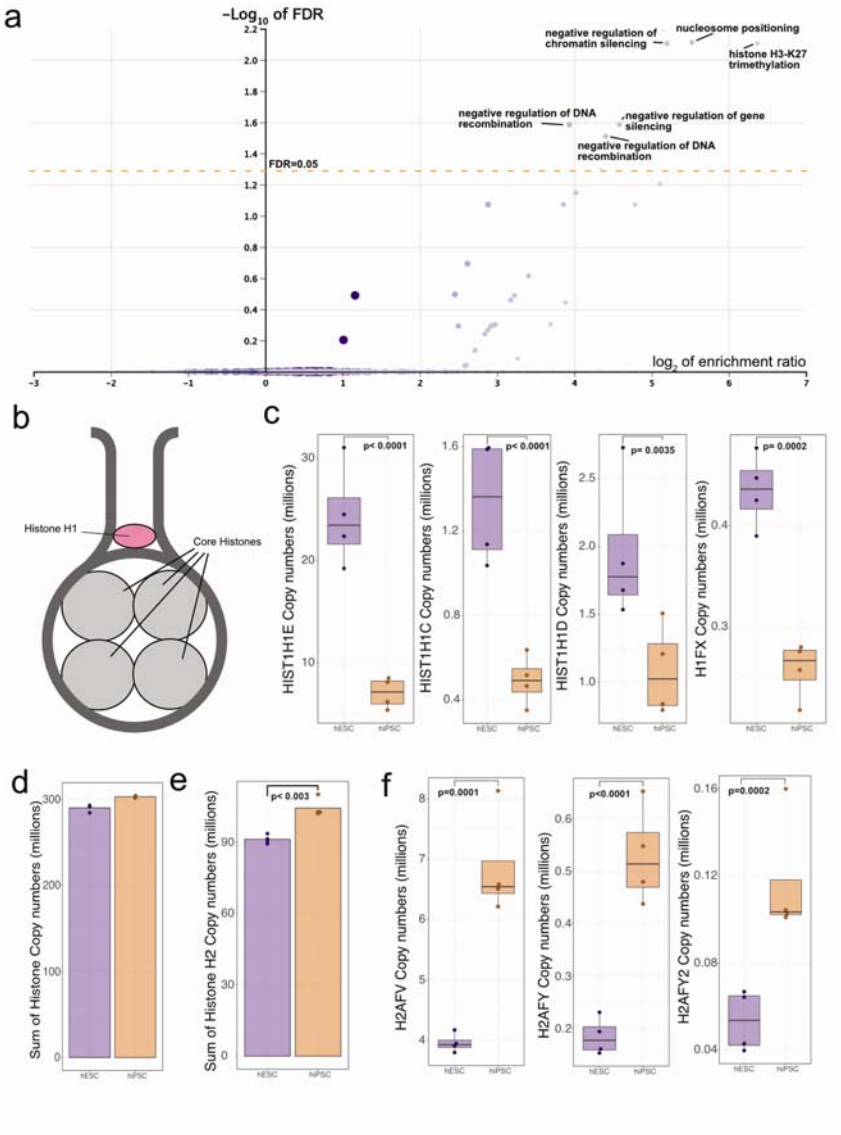

Figure 7 – Changes within histones: (a) Volcano plot showing the -log10 FDR and log2 enrichment for all biological process GO terms relating to proteins that were significantly decreased in expression (see methods) in hiPSCs. **(b)** Schematic showing the positioning of H1 and core histones within the nucleosome. **(c)** Boxplots showing estimated copy numbers for all histone H1 variants for both hESCs and hiPSCs. **(d)** Bar plot showing the sum of all histone copy numbers for both hESCs and hiPSCs. (e) Bar plot showing the sum of all of H2 histone copy numbers for both hESCs and hiPSCs **(f)** Boxplots showing estimated copy numbers for histones H2AFV, H2AFY and H2AFY2 variants for both hESCs and hiPSCs.

This reduced expression seen in H1 histones was not seen for protein members of the other histone families. Evaluating the estimated protein copy numbers across all histones, for example, showed no significant differences between hiPSCs and hESCs (Fig. 7d). However, there was a difference in expression for H2 histones, with hiPSCs showing significantly (p-value=0.003) higher expression compared to hESCs (Fig. 7e). Moreover, it is particularly interesting that the core H2 histones, HIST1H2BK, HIST1H2AC and HIST1H2BJ, were unchanged in expression between the hiPSC and hESC populations. Rather, altered expression was seen for the H2 variants, H2AFV, H2AFY, H2AFY2, which were all significantly (p-value<0.01) increased in expression within hiPSCs compared to hESCs (Fig. 7f). As with H1 histones, these H2 variants share similar sequences. For example, H2AFY and H2AFY2 have 68% sequence identity. Nonetheless, both the shared and unique peptides display congruent behaviour, with all showing increased expression in hiPSCs (Fig. EV3).

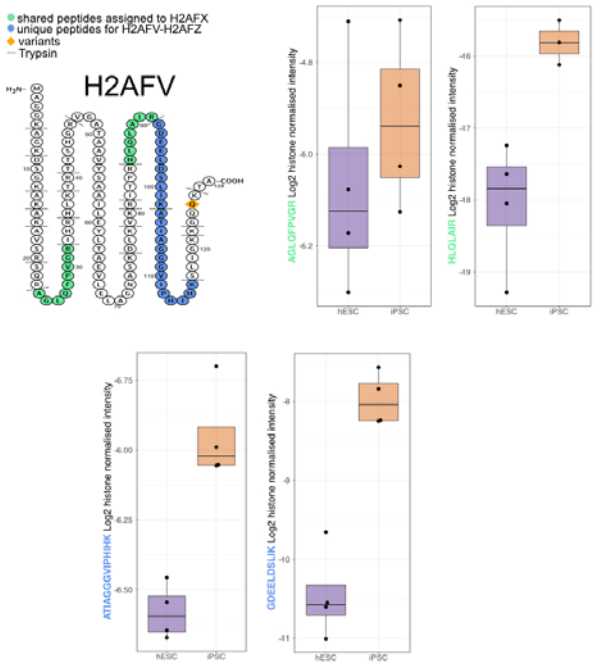

Figure EV3 – H2AFV peptides: Schematic showing the unique and shared peptides detected for H2AFV as well as box plots showing the Log2 histone normalised intensity for both hESC and hiPSCs.

Discussion

Induced pluripotent stem cells can provide vital models for clinical research and future therapies, which makes understanding their similarities and any specific differences with embryo-derived human stem cells all the more important. Our data have highlighted that while multiple, independent hESC and hiPSC lines express a near identical set of proteins, with similar abundance ranks, they also display important quantitative differences in the copy numbers with over 20% of all proteins quantified were significantly increased (fold change>2 and p-value <0.01) in expression within hiPSCs, compared to hESCs. Consequently, estimation of the total protein content per cell, as calculated from the MS analysis, showed that hiPSCs had a median increase of ~60% when compared to the hESCs. The conclusion that hiPSCs have a higher protein expression level than hESCs was subsequently confirmed using an orthogonal EZQ assay, independent of the MS data, which indicated ~75% higher total protein levels in iPSCs.

An important technical point that emerges from this study is that the normalisation approach used to analyse the data has to be carefully considered. Thus, by using a standard median normalisation (concentration-based approach), instead of the proteomic ruler[19], the difference in total protein content between the cell types, involving the increased expression of thousands of proteins, is not apparent. This results in an erroneous conclusion that there is little to no change in protein expression between hiPSCs and hESCs, while the MS data analysed for protein copy number and validated via independent, non-MS methods, shows that protein levels are significantly higher across all of the independent hiPSC lines.

The copy number data showed that the increase in total protein levels in iPSCs resulted specifically from enhanced expression of a subset of protein families. One of the most prominent protein families showing increased expression in hiPSCs were the amino acid transporters, which are known to play important roles to fuel cell growth and protein synthesis[39]. The largest increase in transporter expression was detected in the 3 main glutamine transporters, i.e., SLC1A5, SLC38A1 and SLC38A2. The functional significance of this increased expression was shown by performing a radio labelled glutamine uptake assay, which revealed that hiPSCs had ~94% higher glutamine uptake. Furthermore, there was also increased expression of proteins involved in the downstream glutaminolysis pathway, including GLS and GLUD1. These two proteins are involved in the conversion of glutamine to glutamate and subsequently to alpha-ketoglutarate, an important intermediate for the TCA pathway, which in turn can provide additional energy required to fuel high protein synthesis rates in hiPSCs. It has been reported that when cells preferentially use the glycolytic pathway, as seen in both primed pluripotent stem cells and many transformed tumour cells, there is increased demand for biosynthetic precursors and NADPH, which can be supplied by glutaminolysis[40-42]. The higher uptake of glutamine is potentially fuelling the increased protein mass seen in hiPSCs.

The data also showed that specific subsets of mitochondrial proteins were increased in hiPSCs. Virtually all proteins involved in mitochondrial rRNA processing, translation initiation, elongation and termination of the mitochondrial genome-encoded proteins, were significantly increased in expression. We also detected increased expression of proteins encoded in the mitochondrial genome, which are all hydrophobic membrane proteins that are components of the electron transport chain (ETC)[43]. Similarly, all ETC complexes were significantly increased in expression in

hiPSCs with the highest fold change seen on complex II, which is also a part of the TCA. The mitochondrial protein differences extended also to mitochondrial transporter proteins, all of which were significantly increased in expression, including SLC25A6, the main ATP/ADP transporter within both cell types.

Differences between the mitochondria in hiPSCs and hESCs have been previously reported. For example, it has been shown that hESCs have globular mitochondria with few cristae, while hiPSCs show a mixture of the phenotypes with some globular and some elongated mitochondria, similar to somatic cells[44]. Furthermore, it has been shown that hiPSCs have higher oxygen consumption rate and reserve capacity[44], which is congruent with our proteomic data, as we also see increased expression of the OXPHOS related machinery and transporters.

The other major class of proteins showing increased expression between hiPSCs and hESCs were secreted and ECM-related proteins. These proteins can exert a wide range of effects beyond the cell that produced them. For example, MDK has been reported to promote inflammation by recruiting macrophages and neutrophils[32], while increased expression of SERPINE1 (PAI-1) is linked to higher risk of deep vein thrombosis[45] and a strong risk factor for stroke in the elderly[46]. Furthermore, FGF2 overexpression has been linked with breast cancer[47], gastric cancer[48] and gliomas[49].

FGF2 is of relevance to pluripotent cells as it has been shown that sustained increased exposure to FGF2 better promotes ERK activation in human primed pluripotent cells[50]. ERK2, which also is significantly increased in expression in hiPSCs, has been shown to promote protein synthesis via multiple mechanisms, including through mTORC1 activation[51], eIF4E phosphorylation[52] and PDCD4 inhibition via RSK1/2[53]. The increased expression of FGF2 could be a feedforward loop driving/sustaining growth in hiPSCs.

The only family of proteins which were expressed at significantly lower levels in iPSCs, as compared with hESCs, were H1 histone variants. H1 histones are often referred to as 'linker histones'. They do not form nucleosomes directly, but bind to nucleosomes and are reported to compact chromatin[54]. It has been reported that upon differentiation of hESCs, the expression levels of the H1 histone variants are increased[55]. In other cell types it has also been shown that changes in H1 histone variants are linked with modified differentiation potential[56]. It therefore would be of interest in future to study if the variations in expression of these histones affect the differentiation potential of hiPSCs into different lineages, compared to hESCs.

In summary, our data show that hiPSCs and hESCs, despite their clear similarities, are not identical. These data help define the specific differences between these cells at the protein level and will assist researchers in developing strategies to mitigate for these differences as hiPSCs continue to be used in clinical applications and as disease models.

Acknowledgements

This work was funded by the Wellcome Trust/MRC (098503/E/12/Z) and Wellcome Trust grants (073980/Z/03/Z and 105024/Z/14/Z) and supported by a UK Research Partnership Infrastructure Fund award to the Centre for Translational and Interdisciplinary Research.

Author contributions

A.J.B conceived the study, analysed and interpreted the data. E.G executed all the proteomic sample preparation, and the mass spectrometry experiments. L.V.S performed the glutamine uptake assay and assisted with data interpretation. H.J. performed the EZQ assay and assisted with data interpretation. L.D cultured the hESC and hiPSCs and performed the FACS analysis. H.Y., M.P and J.S helped to

interpret the data. A.I.L and D.A.C supervised the project and helped to interpret the data. The paper written be A.J.B and A.I.L and edited by all authors.

Declaration of interests

E.G now works for Boehringer Ingelheim.

Methods

hiPSC and hESC Cell Culture

Human iPS cells (bubh_3, kucg_2, oaqd_3 and wibj_2) and human hESCs (SA121 and SA181, H1, H9) were both grown in identical conditions, maintained in TESR medium[57] supplemented with FGF2 (Peprotech, 30 ng/ml) and noggin (Peprotech, 10 ng/ml) on growth factor reduced geltrex basement membrane extract (Life Technologies, 10 µg/cm2) coated dishes at 37°C in a humidified atmosphere of 5% CO_2 in air.

Cells were routinely passaged twice a week as single cells using TrypLE select (Life Technologies) and replated in TESR medium that was further supplemented with the Rho kinase inhibitor Y27632 (Tocris, 10 µM) to enhance single cell survival. Twenty-four hours after replating Y27632 was removed from the culture medium. For proteomic analyses cells were plated in 100 mm geltrex coated dishes at a density of $5x10^4$ cells cm^{-2} and allowed to grow to for 3 days until confluent with daily medium changes.

Protein extraction

Cell pellets were resuspended in 300 µL extraction buffer (4% SDS in 100 mM triethylammonium bicarbonate (TEAB), phosphatase inhibitors (PhosSTOP™, Roche)). Samples were boiled (15 min, 95 °C, 350 rpm) and sonicated for 30 cycles

in a bath sonicator (Bioruptor® Pico bath sonicator, Diagenode, Belgium; 30 s on, 30 s off) followed by probe sonication for 50 s (20 s on, 5 s off). 2 µL Benzonase® nuclease HC (250 U/µL, Merck Millipore) was added and incubated for 30 min (37 °C, 750 rpm). Reversibly oxidized cysteines were reduced with 10 mM TCEP (45 min, 22 °C, 1,000 rpm) followed by alkylation of free thiols with 20 mM iodoacetamide (45 min, 22 °C, 1,000 rpm, in the dark). Proteins were quantified using the fluorometric EZQ™ assay (Thermo Fisher Scientific).

Protein digestion using the SP3 method

Protein extracts were cleaned and digested with the SP3 method as described previously with modifications [58,59]. Briefly, 50 µL of a 20 µg/µL SP3 bead stock (Sera-Mag SpeedBead carboxylate-modified magnetic particles; GE Healthcare Life Sciences) and 500 µL acetonitrile (ACN; final concentration of 70%) were added to 150 µL of protein extract and incubated for 10 min (1000 rpm). Tubes were mounted on a magnetic rack, supernatants were removed and beads were washed twice with 70% ethanol and once with ACN (1 mL each). Beads were resuspended in 80 µL 100 mM TEAB and digested for 4 h with LysC followed by tryptic digestion overnight (1:50 protease:protein ratio, 37 °C, 1,000 rpm). Peptides were cleaned by addition of 3.5 µL formic acid (final concentration of 4%) and 1700 µL ACN (final concentration of 95%) followed by incubation for 10 min. After spinning down (1,000 g) tubes were mounted on a magnetic rack and beads were washed once with 1.5 mL ACN. Peptides were eluted from the beads with 100 µL 2% DMSO and acidified with 5.2 µL 20 % formic acid (final concentration of 1%) followed by centrifugation (15,000 g). Peptide amounts were quantified using the fluorometric CBQCA assay (Thermo Scientific).

TMT labelling

The required volume for 15 µg peptides per sample were dried *in vacuo* in a Concentrator plus (Eppendorf) and resuspended in 50 µL 200 mM EPPS pH 8.5. TMT10plex tags (Thermo Scientific) were dissolved in anhydrous ACN and added to the peptide sample in a 1:10 peptide:TMT ratio. Additional anhydrous ACN was added to a final volume of 22 µL. Samples were incubated for 2 h (22 °C, 750 rpm). Unreacted TMT was quenched by incubation with 5 µL 5% hydroxylamine for 30 min. Samples were combined, dried *in vacuo* and resuspended in 1% TFA followed by clean-up with solid-phase extraction using Waters Sep-Pak tC18 50 mg. Samples were loaded, washed five times with 1 mL 0.1% TFA in water and peptides were eluted with 70% ACN/0.1% TFA (1 mL) and dried *in vacuo* in a Concentrator plus (Eppendorf).

High pH reversed phase fractionation

TMT labelled samples were fractionated using off-line high pH reversed phase chromatography. Dried samples were resuspended in 5% formic acid and loaded onto a 4.6 x 250 mm XBridge BEH130 C18 column (3.5 µm, 130 Å; Waters). Samples were separated on a Dionex Ultimate 3000 HPLC system with a flow rate of 1 mL/min. Solvents used were water (A), ACN (B) and 100 mM ammonium formate pH 9 (C). While solvent C was kept constant at 10%, solvent B started at 5% for 3 min, increased to 21.5% in 2 min, 48.8% in 11 min and 90% in 1 min, was kept at 90% for further 5 min followed by returning to starting conditions and re-equilibration for 8 min. Peptides were separated into 48 fractions, which were concatenated into

24 fractions and subsequently dried *in vacuo*. Peptides were redissolved in 5% formic acid and analysed by LC-MS.

LC-MS analysis

TMT labelled samples were analysed on an Orbitrap Fusion Tribrid mass spectrometer coupled to a Dionex RSLCnano HPLC (Thermo Scientific). Samples were loaded onto a 100 µm × 2 cm Acclaim PepMap-C18 trap column (5 µm, 100 Å) with 0.1% trifluoroacetic acid for 7 min and a constant flow of 4 µL/min. Peptides were separated on a 75 µm × 50 cm EASY-Spray C18 column (2 µm, 100 Å; Thermo Scientific) at 50 ºC over a linear gradient from 10% to 40% B in 153 min with a flow rate of 200 nL/min. Solvents used were 0.1% formic acid (A) and 80% ACN/0.1% formic acid (B). The spray was initiated by applying 2.5 kV to the EASY-Spray emitter. The ion transfer capillary temperature was set to 275 °C and the radio frequency of the S-lens to 50%. Data were acquired under the control of Xcalibur software in a data-dependent mode. The number of dependent scans was 12. The full scan was acquired in the orbitrap covering the mass range of *m/z* 350 to 1,400 with a mass resolution of 120,000, an AGC target of $4x10^5$ ions and a maximum injection time of 50 ms. Precursor ions with charges between 2 and 7 and a minimum intensity of $5x10^3$ were selected with an isolation window of *m/z* 1.2 for fragmentation using collision-induced dissociation in the ion trap with 35% collision energy. The ion trap scan rate was set to “rapid”. The AGC target was set to $1x10^4$ ions with a maximum injection time of 50 ms and a dynamic exclusion of 60 s. During the MS3 analysis, for more accurate TMT quantification, 5 fragment ions were co-isolated using synchronous precursor selection in a window of *m/z* 2 and further fragmented with a HCD collision energy of 65%. The fragments were then analysed

in the orbitrap with a resolution of 50,000. The AGC target was set to $5x10^4$ ions and the maximum injection time was 105 ms.

Radiolabelled glutamine uptake (protocol was adapted from[60])

Two hiPSC lines (wibj_2 and oaqd_3) with 3 technical replicates each were compared to two hESC lines (SA121 and SA181) with 3 technical replicates of each. Both hiPSCs and hESCs were plated in 6-well plates 2 days before the transport assay (5e4 cells/cm2 – this gives 1e6 cells/well on "uptake day"). The cell growth media was carefully aspirated so as not to disturb the adherent monolayer of cells. They were washed gently by pipetting on 5 mls preheated (37°C) uptake solution (HBSS (pH 7.4), GIBCO) and aspirating off, this was repeated 3 times. They were then incubated with 0.5 ml of uptake solution containing [^{3}H]glutamine (5 µCi/ml ; perkin elmer, NET 55100) in the presence or absence of L-glutamine (5 mM; sigma) for 2 min.

The Uptake was stopped by removing the uptake solution and washing cells with 2 ml of ice-cold stop solution (HBSS with 10 mM nonradioactive L-glutamine) three times. After the third wash, the cells were lysed in 200 µl of 0.1% SDS and 100 mM NaOH, and 100 µl was used to measure the radioactivity associated with the cells. Finally 100 µl sample was added to scint vials containing 3 mls scintillant (*OptiPhase HiSafe* 3, Perkin Elmer). β-radioactivity was measured with Tri-Carb 4910TR liquid scintillation counter.

Net glutamine CPM

The net glutamine CPM values where calculated by subtracting the Quench CPM values from the Glutamine CPM values.

hiPSC vs hECs proteomics pipeline

The data were searched and quantified with MaxQuant[61] (version 1.6.3.3) against the human SwissProt database from UniProt[62] (June 2019). The data was searched with the following parameters: carbamidomethylation of cysteine, as well as TMT modification on peptide amino termini and lysine side chains, were fixed modifications; methionine oxidation and acetylation of amino termini of proteins were variable modifications. The false discovery rate was set to 1% for positive identification at the protein and peptide spectrum match (PSM) level.

Data filtering

All protein groups identified with less than 2 razor or unique peptides or labelled as ‘Contaminant’, ‘Reverse’ or ‘Only identified by site’ were removed from the analysis.

Copy number calculations

Protein copy numbers were estimated following the “proteomic ruler” method[19] but adapted to work with TMT MS3 data. The summed MS1 intensities were allocated to the different experimental conditions according to their fractional MS3 reporter intensities.

Protein content estimations

The protein content was estimated using the following formula: CN × MW and then converting the data from Daltons to picograms, where CN is the protein copy number and MW is the protein molecular weight (in Da).

28S to 39S ratios

For each hiPSC and hESC line the sum of the estimated copy numbers for all subunits of the 28S complex were divided by the sum of estimated copy numbers of all 39S subunits.

Differential expression analysis

Fold changes and P-values were calculated in R, for individual proteins the p-values were calculated with the bioconductor package LIMMA[63] version 3.7. The Q-values provided were generated in R using the "qvalue" package version 2.10.0. P-values for protein families and protein complexes were calculated using Welch's T-test.

Subcellular localisation data

The subcellular analysis was performed using the database obtained from the Human Protein Atlas[64] in their subcellular location dataset version 20.1

hiPSC vs hESC overrepresentation analysis

All overrepresentation analysis were done on WebGestalt. The first analysis selected proteins where the $\log_2$ fold change was greater than 1 and a p-value lower than 0.01. The second analysis selected proteins whose fold change was lower than median minus one standard deviation (0.195) and a p-value lower than 0.01. Both analyses used all identified proteins with 2 or more razor and unique peptides as a background and required an FDR lower than 0.05.

References

1 Smith, A. G. Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* **17**, 435-462, doi:10.1146/annurev.cellbio.17.1.435 (2001).

2 Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147, doi:10.1126/science.282.5391.1145 (1998).

3 Volarevic, V. *et al.* Ethical and Safety Issues of Stem Cell-Based Therapy. *Int J Med Sci* **15**, 36-45, doi:10.7150/ijms.21666 (2018).
4 Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872, doi:10.1016/j.cell.2007.11.019 (2007).
5 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).
6 Kimbrel, E. A. & Lanza, R. Current status of pluripotent stem cells: moving the first therapies to the clinic. *Nat Rev Drug Discov* **14**, 681-692, doi:10.1038/nrd4738 (2015).
7 Ebert, A. D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277-280, doi:10.1038/nature07677 (2009).
8 Lee, G. *et al.* Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461**, 402-406, doi:10.1038/nature08320 (2009).
9 Liu, G. H. *et al.* Progressive degeneration of human neural stem cells caused by pathogenic LRRK2. *Nature* **491**, 603-607, doi:10.1038/nature11557 (2012).
10 Mallon, B. S. *et al.* Comparison of the molecular profiles of human embryonic and induced pluripotent stem cells of isogenic origin. *Stem Cell Res* **12**, 376-386, doi:10.1016/j.scr.2013.11.010 (2014).
11 Mallon, B. S. *et al.* StemCellDB: the human pluripotent stem cell database at the National Institutes of Health. *Stem Cell Res* **10**, 57-66, doi:10.1016/j.scr.2012.09.002 (2013).
12 Guenther, M. G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7**, 249-257, doi:10.1016/j.stem.2010.06.015 (2010).
13 Munoz, J. *et al.* The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol Syst Biol* **7**, 550, doi:10.1038/msb.2011.84 (2011).
14 Vitale, A. M. *et al.* Variability in the generation of induced pluripotent stem cells: importance for disease modeling. *Stem Cells Transl Med* **1**, 641-650, doi:10.5966/sctm.2012-0043 (2012).
15 Kilpinen, H. *et al.* Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370-375, doi:10.1038/nature22403 (2017).
16 McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem* **86**, 7150-7158, doi:10.1021/ac502040v (2014).
17 Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
18 Brenes, A., Hukelmann, J., Bensaddek, D. & Lamond, A. I. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Mol Cell Proteomics* **18**, 1967-1980, doi:10.1074/mcp.RA119.001472 (2019).
19 Wisniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics* **13**, 3497-3506, doi:10.1074/mcp.M113.037309 (2014).
20 Folmes, C. D. *et al.* Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab* **14**, 264-271, doi:10.1016/j.cmet.2011.06.011 (2011).
21 Watson, P. F. & Petrie, A. Method agreement analysis: a review of correct methodology. *Theriogenology* **73**, 1167-1179, doi:10.1016/j.theriogenology.2010.01.003 (2010).

22 Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199-W205, doi:10.1093/nar/gkz401 (2019).
23 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).
24 Taanman, J. W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* **1410**, 103-123, doi:10.1016/s0005-2728(98)00161-3 (1999).
25 Bhutia, Y. D. & Ganapathy, V. Glutamine transporters in mammalian cells and their functions in physiology and cancer. *Biochim Biophys Acta* **1863**, 2531-2539, doi:10.1016/j.bbamcr.2015.12.017 (2016).
26 Broer, A., Rahimi, F. & Broer, S. Deletion of Amino Acid Transporter ASCT2 (SLC1A5) Reveals an Essential Role for Transporters SNAT1 (SLC38A1) and SNAT2 (SLC38A2) to Sustain Glutaminolysis in Cancer Cells. *J Biol Chem* **291**, 13194-13205, doi:10.1074/jbc.M115.700534 (2016).
27 Tohyama, S. *et al.* Glutamine Oxidation Is Indispensable for Survival of Human Pluripotent Stem Cells. *Cell Metab* **23**, 663-674, doi:10.1016/j.cmet.2016.03.001 (2016).
28 Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* **137**, 3351-3360, doi:10.1242/dev.050146 (2010).
29 Xu, R. H. *et al.* NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell* **3**, 196-206, doi:10.1016/j.stem.2008.07.001 (2008).
30 Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* **17**, 155-169, doi:10.1038/nrm.2015.28 (2016).
31 Korf-Klingebiel, M. *et al.* Myeloid-derived growth factor (C19orf10) mediates cardiac repair following myocardial infarction. *Nat Med* **21**, 140-149, doi:10.1038/nm.3778 (2015).
32 Weckbach, L. T. *et al.* The cytokine midkine supports neutrophil trafficking during acute inflammation by promoting adhesion via beta2 integrins (CD11/CD18). *Blood* **123**, 1887-1896, doi:10.1182/blood-2013-06-510875 (2014).
33 Winckers, K., ten Cate, H. & Hackeng, T. M. The role of tissue factor pathway inhibitor in atherosclerosis and arterial thrombosis. *Blood Rev* **27**, 119-132, doi:10.1016/j.blre.2013.03.001 (2013).
34 Aumailley, M. *et al.* A simplified laminin nomenclature. *Matrix Biol* **24**, 326-332, doi:10.1016/j.matbio.2005.05.006 (2005).
35 Fiore, A. *et al.* Laminin-111 and the Level of Nuclear Actin Regulate Epithelial Quiescence via Exportin-6. *Cell Rep* **19**, 2102-2115, doi:10.1016/j.celrep.2017.05.050 (2017).
36 Woodcock, C. L., Skoultchi, A. I. & Fan, Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res* **14**, 17-25, doi:10.1007/s10577-005-1024-3 (2006).
37 Robinson, P. J. & Rhodes, D. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* **16**, 336-343, doi:10.1016/j.sbi.2006.05.007 (2006).
38 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805, doi:10.1021/pr101065j (2011).
39 Broer, S. Amino Acid Transporters as Targets for Cancer Therapy: Why, Where, When, and How. *Int J Mol Sci* **21**, doi:10.3390/ijms21176156 (2020).

40 Jin, L., Alesi, G. N. & Kang, S. Glutaminolysis as a target for cancer therapy. *Oncogene* **35**, 3619-3625, doi:10.1038/onc.2015.447 (2016).
41 Wise, D. R. & Thompson, C. B. Glutamine addiction: a new therapeutic target in cancer. *Trends Biochem Sci* **35**, 427-433, doi:10.1016/j.tibs.2010.05.003 (2010).
42 Reitzer, L. J., Wice, B. M. & Kennell, D. Evidence that glutamine, not sugar, is the major energy source for cultured HeLa cells. *J Biol Chem* **254**, 2669-2676 (1979).
43 Reinecke, F., Smeitink, J. A. & van der Westhuizen, F. H. OXPHOS gene expression and control in mitochondrial disorders. *Biochim Biophys Acta* **1792**, 1113-1121, doi:10.1016/j.bbadis.2009.04.003 (2009).
44 Varum, S. *et al.* Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One* **6**, e20914, doi:10.1371/journal.pone.0020914 (2011).
45 Tang, J., Zhu, W., Mei, X. & Zhang, Z. Plasminogen activator inhibitor-1: a risk factor for deep vein thrombosis after total hip arthroplasty. *J Orthop Surg Res* **13**, 8, doi:10.1186/s13018-018-0716-2 (2018).
46 Hoekstra, T. *et al.* 4G/4G genotype of PAI-1 gene is associated with reduced risk of stroke in elderly. *Stroke* **34**, 2822-2828, doi:10.1161/01.STR.0000098004.26252.EB (2003).
47 Giulianelli, S. *et al.* FGF2 induces breast cancer growth through ligand-independent activation and recruitment of ERalpha and PRBDelta4 isoform to MYC regulatory sequences. *Int J Cancer* **145**, 1874-1888, doi:10.1002/ijc.32252 (2019).
48 Li, Y., Guo, X. B., Wang, J. S., Wang, H. C. & Li, L. P. Function of fibroblast growth factor 2 in gastric cancer occurrence and prognosis. *Mol Med Rep* **21**, 575-582, doi:10.3892/mmr.2019.10850 (2020).
49 Sooman, L. *et al.* FGF2 as a potential prognostic biomarker for proneural glioma patients. *Acta Oncol* **54**, 385-394, doi:10.3109/0284186X.2014.951492 (2015).
50 Lotz, S. *et al.* Sustained levels of FGF2 maintain undifferentiated stem cell cultures with biweekly feeding. *PLoS One* **8**, e56289, doi:10.1371/journal.pone.0056289 (2013).
51 Ma, L., Chen, Z., Erdjument-Bromage, H., Tempst, P. & Pandolfi, P. P. Phosphorylation and functional inactivation of TSC2 by Erk implications for tuberous sclerosis and cancer pathogenesis. *Cell* **121**, 179-193, doi:10.1016/j.cell.2005.02.031 (2005).
52 Pelletier, J., Graff, J., Ruggero, D. & Sonenberg, N. Targeting the eIF4F translation initiation complex: a critical nexus for cancer development. *Cancer Res* **75**, 250-263, doi:10.1158/0008-5472.CAN-14-2789 (2015).
53 Galan, J. A. *et al.* Phosphoproteomic analysis identifies the tumor suppressor PDCD4 as a RSK substrate negatively regulated by 14-3-3. *Proc Natl Acad Sci U S A* **111**, E2918-2927, doi:10.1073/pnas.1405601111 (2014).
54 Bednar, J. *et al.* Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell* **66**, 384-397 e388, doi:10.1016/j.molcel.2017.04.012 (2017).
55 Terme, J. M. *et al.* Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *J Biol Chem* **286**, 35347-35357, doi:10.1074/jbc.M111.281923 (2011).
56 Sollberger, G. *et al.* Linker histone H1.2 and H1.4 affect the neutrophil lineage determination. *Elife* **9**, doi:10.7554/eLife.52563 (2020).
57 Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nat Methods* **3**, 637-646, doi:10.1038/nmeth902 (2006).

58 Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **10**, 757, doi:10.15252/msb.20145625 (2014).
59 Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **14**, 68-85, doi:10.1038/s41596-018-0082-x (2019).
60 Yeramian, A. *et al.* Arginine transport via cationic amino acid transporter 2 plays a critical regulatory role in classical or alternative activation of macrophages. *J Immunol* **176**, 5918-5924, doi:10.4049/jimmunol.176.10.5918 (2006).
61 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).
62 The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).
63 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
64 Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, doi:10.1126/science.aal3321 (2017).