Knowledge of the subcellular distribution of proteins is vital for understanding cellular mechanisms. Capturing the subcellular proteome in a single experiment has proven challenging, with studies focusing on specific compartments or assigning proteins to subcellular niches with low resolution and/or accuracy. Here we introduce hyperLOPIT, a method that couples extensive fractionation, quantitative high-resolution accurate mass spectrometry with multivariate data analysis. We apply hyperLOPIT to a pluripotent stem cell population whose subcellular proteome has not been extensively studied. We provide localization data on over 5,000 proteins with unprecedented spatial resolution to reveal the organization of organelles, sub-organelar compartments, protein complexes, functional networks and steady-state dynamics of proteins and unexpected subcellular locations. The method paves the way for characterizing the impact of post-transcriptional and post-translational modification on protein location and studies involving proteome-level locational changes on cellular perturbation. An interactive open-source resource is presented that enables exploration of these data.
luripotent mouse embryonic stem (ES) cells are self-renewing clonal populations derived from blastocysts, which can be differentiated into the ensemble of cell types of the organism\(^1\). Their study is central to developmental biology and the emerging field of regenerative medicine. Currently our understanding of the biology of ES cells is deeply rooted in our knowledge of their transcriptomes, epigenetics and underlying gene regulatory networks, which have created a foundation for understanding pluripotency and the transition to differentiation\(^2\,3\).

There is evidence that post-transcriptional events such as signalling, adhesion, protein turnover and post-translational modification make a significant contribution to the regulation of differentiation\(^4\,5\), yet their precise roles in this process, and how they interact with each other and with the transcriptional machinery, remain open questions. The transition from self-renewal to differentiation is also associated with major changes in cell morphology, and therefore some of the effects of these post-transcriptional processes must be associated with changes in intracellular organization. Understanding the subcellular distribution of proteins and other biomolecules, and how the distribution changes with cell state, is therefore of paramount importance for the delineation of post-transcriptional processes in ES cells.

Protein localization is typically determined by immunocytochemistry or by monitoring fluorescent fusion proteins by confocal microscopy. While these approaches are valuable and well-established, there are certain limitations to their applicability. Immunocytochemistry is dependent on the availability of high-specificity and high-sensitivity antibodies, while fluorescent fusion proteins are vulnerable to aberrant localization due to the effect of the fusion moiety on protein topology\(^6\,7\). These limitations can be overcome with complementary technologies such as protein mass spectrometry (MS), which offers the capability to assay thousands of proteins simultaneously and in their native state\(^8\).

Localization of organelle proteins by isotope tagging (LOPIT) is a quantitative proteomics method for the high throughput and simultaneous characterization of multiple subcellular compartments, without the requirement for total purification of compartments of interest\(^9\). LOPIT combines biochemical fractionation by density-gradient ultracentrifugation, sample multiplexing by in vitro covalent labelling, and liquid chromatography-mass spectrometry. In LOPIT, cells are first lysed under detergent-free conditions so that there is minimal disruption to organelle integrity. Membranes are then separated based on their characteristic buoyant densities by ultracentrifugation. Although organelles do not partition into discrete purified fractions, different organelles display distinct enrichment patterns. Fractions representing peak enrichment for organelles of interest are selected for proteolytic digestion. The resulting peptides are differentially labelled with amine-reactive tandem mass tag (TMT) reagents\(^10\), which allow peptides derived from each fraction to be distinguished by mass spectrometry. The relative abundance of a peptide can be determined by its TMT reporter ion profile, which recapitulates the distribution of the protein across the fractionation scheme. Proteins residing in the same subcellular niche would be expected to co-distribute, and therefore present similar TMT reporter ion profiles\(^11\). Classification algorithms are then used to assign proteins to subcellular compartments based on correlation with organelar marker proteins.

Here we significantly extend the LOPIT concept with novel approaches for sample preparation, mass spectrometry data acquisition and multivariate analysis. This new workflow, named hyperplexed LOPIT (hyperLOPIT), benefits from several recent technological advancements. First, the development of neutron-encoded isotopologue variants of TMT has increased the multiplexing capacity of isobaric tagging experiments to 10 samples\(^12\). These additional labels have enabled more subcellular fractions to be sampled, allowing for a more elaborate fractionation scheme that reaches sub-organellar levels of resolution. Second, quantitative accuracy of TMT-based applications is significantly improved by mass spectrometry data acquisition using synchronous precursor selection MS\(^3\) (SPS-MS\(^3\); see Supplementary Fig. 1 for a detailed overview of this method). Multivariate approaches such as hyperLOPIT represent particularly demanding applications of TMT quantification, as consistently high accuracy and precision are necessary for co-localized proteins to display correlated TMT distributions\(^13\). We have therefore incorporated SPS-MS\(^3\) acquisition on the Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) into the hyperLOPIT pipeline, and demonstrate that it greatly improves spatial resolution and the reliability with which protein localization may be determined. Finally, we have extended our data analysis platform to facilitate rapid interrogation of the data by providing an easy-to-use graphical user interface.

We apply the hyperLOPIT workflow to a population of self-renewing mouse ES cells. The result is comprehensive coverage of the subcellular proteome with unprecedented spatial resolution, enabling the mapping of components of organelles, transitory proteins, multi-protein complexes, signalling pathways and families of functionally related proteins. The presented data offer hitherto unknown subcellular detail about a population of self-renewing stem cells. We observe the dynamic nature of Golgi apparatus proteins, noting sets of Golgi marker proteins that are distributed amongst other subcellular structures, an observation we support by microscopy. We also demonstrate that additional insights into published interactomes and focused protein-localization studies can be attained by integrative analysis with LOPIT. HyperLOPIT offers a spatial scaffold onto which other high content proteomics data sets can be mapped yielding added value to complementary data. We give an example of this functionality by combining the data presented here with a recent data set of surface proteins captured using chemical tagging.

Results

Biochemical fractionation of mouse ES cells. An overview of the hyperLOPIT workflow is shown in Fig. 1. To create a method fit for the purpose of capturing cell-wide proteome localization data, we first improved the LOPIT workflow as follows: to increase subcellular resolution, extensive subcellular fractionation of pluripotent E14TG2a cells was performed. First, crude membranes were separated from the soluble fraction, enriched in cytosolic proteins. Crude membranes were then separated into organelle-enriched fractions by equilibrium density gradient centrifugation, within which organelles adopt specific distribution profiles consistent with their respective buoyant densities. A portion of the cell culture was also used to enrich chromatin-associated proteins with a parallel workflow based on detergent permeabilization\(^14\). Ten fractions were chosen to best represent peak organelle densities, labelled with TMT reagents and processed as described in the Methods section. We acquired three biological replicates of hyperLOPIT data, each with a slightly different selection of subcellular fractions for TMT 10-plex labelling. The first replicate placed greater emphasis on resolving low density-gradient fractions that are enriched in secretory pathway components, while the second and third replicates placed greater emphasis on separation of the denser organelles such as the mitochondrion and peroxisome.
SPS-MS$^3$ enhances quantitative performance. TMT quantification by conventional tandem mass spectrometry (MS$^2$) suffers from impaired quantitative accuracy and precision due to interference from contaminant peptides with similar chromatographic and mass-to-charge properties to the target peptide. Improved quantitative performance is achieved with an additional round of ion selection and fragmentation to purify the analyte from which TMT quantification is derived$^{17,18}$, but such MS$^2$-based methods result in substantially reduced sensitivity$^{19}$. SPS-MS$^3$ balances the quantitative gains of MS$^3$ quantification with the sensitivity required for proteome-wide analysis$^{20,21}$. Whereas in conventional MS$^2$ a single peptide fragment ion is selected for quantification, by using isolation waveforms with multiple frequency ‘notches’ to collect multiple peptide fragments, SPS-MS$^3$ improves ion statistics for quantification (Supplementary Fig. 1).

To assess the impact of SPS-MS$^3$ on quantitative performance, we evaluated the effect of the number of SPS-MS$^3$ ‘notches’ by comparing quantification derived from conventional MS$^2$, conventional MS$^3$ and SPS-MS$^3$ data acquisition. Adjusting the number of SPS-MS$^3$ notches altered the balance between quantitative performance and sensitivity. Increasing the number of notches augmented the TMT reporter ion signal intensity, and therefore the proportion of quantifiable spectra, but also reintroduces contaminant ions that distort quantification. Fewer notches result in lower reporter ion signal, with concomitant reduction in the number of quantifiable spectra. We found SPS-MS$^3$ with 10 precursors ‘notches’ to represent a suitable balance for global analysis, with 92.8% of the acquired SPS-MS$^3$ spectra yielding TMT reporter ion counts $>1 \times 10^5$—a comparable figure to that obtained with conventional MS$^2$ acquisition (Supplementary Table 1).

We then compared E14TG2a hyperLOPIT data sets acquired with conventional MS$^2$, and SPS-MS$^3$ with 10 notches. When comparing peptides derived from proteins with well curated localization, we observed SPS-MS$^3$ acquisition resulted in a significant improvement in quantitative accuracy over conventional MS$^2$ acquisition (Fig. 2 and Supplementary Figs 2 and 3). This gain in quantitative performance resulted in greater resolution of organelles from one another, and from the median of the total peptide population; tendency towards a unified distribution being characteristic of distorted TMT quantification (Supplementary Fig. 18)$^{16,22}$.
SPS-MS3 quantification. The median positions of other organelles are all significantly increased in SPS-MS3 (Wilcoxon rank sum test $P$ value $<2.2 \times 10^{-16}$), indicating greater organellar resolution. Experiment 3 was not included, as little additional resolution was obtained by further data fusion. Using the pRoloc data analysis pipeline\(^{25}\), an initial application of novelty detection\(^{26}\) was conducted to identify and confirm the presence of organelle clusters in an unbiased data-specific manner, followed by supervised classification using a support vector machine (SVM) for final protein-organelle assignment. Applying SVM scoring thresholds based on concordance with gene ontology annotation, the steady-state localization of 2,855 out of a total of 5,032 protein groups were unambiguously determined (Fig. 3 and Supplementary Figs 6 and 7). Sub-nuclear resolution was also obtained, with proteins localized to chromatin, non-chromatin nuclear and nuclear lamina displaying distinct quantitative distributions. Manual curation of the data set also revealed co-localization of proteins for several other subcellular niches, including coatamer and clathrin-coated vesicles, and cytoskeletal fragments (Fig. 3).

Given the breadth of proteome coverage and subcellular resolution, hyperLOPIT can be used to investigate the organization of cells at multiple levels of scope. Insights into cell behaviour may be drawn from the localization of individual proteins, protein families and even functional networks, or by evaluating the protein content of particular subcellular compartments. In the following sections, we discuss how some of the layers of data can be interpreted, and the insights into the organization of pluripotent ES cells that may be drawn.

**Organelle membership.** The hyperLOPIT workflow enabled unambiguous assignment of 2,855 proteins to 14 discrete organelar and sub-organelar compartments in a single experiment. This number amounted to over 50% of the proteins identified, with the remainder displaying intermediate distributions as described in the Proteins in transit section. The classified proteins represent both residents of an organelle, as well as transient traffickers or cargo proteins that are captured in this position under these experimental conditions and in this particular cell type. The catalogue of organelle members is therefore context specific and must be viewed as a subcellular snapshot of a dynamic system.

Approximately 83% of proteins classified by hyperLOPIT carry localization-specific gene ontology annotation, with only 39% of these proteins annotated based on direct assay evidence. This data set therefore provides new experimental
evidence for the localization of 1,775 murine proteins, ~350 of which currently lack organelle-specific localization information in the UniProt database (Supplementary Data 1). It should be noted that these values include proteins that are annotated as cytoplasmic. Since the term cytoplasm incorporates the cytosol, cytoskeleton, organelles and other cellular features other than the nucleus and cell surface, such annotation lacks the specificity to be suitably informative in this context. HyperLOPIT is therefore a useful tool for supplementing database annotation of protein localization with high throughput. Another key strength of the approach is that protein localization may be determined in particular cell type(s) and cell state(s) of interest—information that is typically lacking in gene ontology annotation or computational prediction of localization.

We compared the protein localizations determined by hyperLOPIT with another recently published plasma membrane proteome for the same cell line. Bausch-Fluck et al. identified plasma membrane proteins by cell surface biotinylation and affinity purification-mass spectrometry. In this study the authors categorized their cell capture data into three groups: High confidence, for proteins with UniProt keywords including ‘Cell junction’, ‘Cell membrane’ and ‘Secreted’; putative, for proteins with predicted transmembrane domains, but none of the above keywords assigned; and non-specific, for other identified proteins, which were assumed to be abundant contaminant proteins. Proteins labelled as high confidence in this study corroborate with their localization in the hyperLOPIT data, as almost all such proteins localize to the plasma membrane or endosomes—suggesting extensive recycling of some of these surface proteins. HyperLOPIT also confirms plasma membrane localization for many proteins labelled as putative in the Bausch-Fluck study (Supplementary Fig. 8). The third category, non-specific proteins, consists mostly of proteins found in non-surface localizations by hyperLOPIT. The strong correlation of results demonstrates that hyperLOPIT may also be used as an orthogonal validation method for targeted localization studies.

Proteins in transit. Not all proteins are expected to partition into discrete subcellular compartments. While over half of the quantified proteins were classifiable to a single and unambiguous location, many proteins were found to have distributions that did not closely correlate with those of the 14 classified subcellular compartments. These less discrete distribution patterns occur for several reasons.

First, proteins present in multiple compartments will adopt quantitative distributions that reflect their steady-state subcellular enrichment. β-catenin for example, has been robustly classified to localize to the adherens junctions, cytosol and nucleus. Classification to the plasma membrane reflects its relative enrichment at the adherens junction in ES cells cultured under self-renewing conditions, which has been previously demonstrated by immunofluorescence microscopy.

There are cases where the steady-state localization is not so skewed towards one of multiple compartments. For example, many components of the nuclear import and export machinery display distributions consistent with mixed localization between the cytosol and nucleus (Supplementary Fig. 9). These proteins were not classified as unambiguous residents of either the nucleus or cytosol; their distribution patterns accurately reflect their true subcellular localization. Other proteins displaying mixed localizations include signalling cascade effectors, such as Erk-2 and the adaptor protein Grb-2, components of the MAP kinase signalling pathway that distribute between the plasma membrane and cytosol, reflecting the dynamic transitions between the two locations. For some proteins with mixed localization the distribution patterns are quite complex. Mcl-1, a Bcl-2 family member protein, displays a three-way mixed localization between the endoplasmic reticulum, mitochondrion and nucleus, and therefore falls between the three organelle clusters in principal component analysis (PCA) space; this broad distribution is consistent with previous confocal microscopy analysis of its human orthologue.

The steady-state distribution of transitory proteins can provide information about the state of the cell population. For example,
Tfe3, a helix–loop–helix family transcription factor and modulator of the exit of ES cells from pluripotency, was observed with mixed localization between the cytoplasm and nucleus (Supplementary Fig. 9), consistent with immunocytochemistry data in ES cells. In the pluripotent state, Tfe3 localizes to both the nucleus and cytosol, and regulates the expression of key pluripotency factor Esrrb. In the early stages of differentiation, it is excluded from the nucleus.30

Second, intermediate subcellular distributions may represent proteins residing in organelles other than those used in the SVM classification. These compartments typically have their own physicochemical properties, and their resident proteins therefore co-localize, but are not sufficiently enriched for their distribution patterns to be fully resolved from those of other subcellular compartments. For example, known components of the clathrin-AP3 trafficking vesicles were found to co-localize away from other organelles and vesicles, but were not distinct from other proteins with mixed localization. Some proteins with correlated distributions to the clathrin-AP3 vesicles are plausible vesicular components, such as syntaxin-18, whereas others are transitory cytosolic and cytoskeletal proteins such as serine/threonine kinase Pak4.

Finally, the observed localization patterns may represent proteins that comprise or are tethered to the cytoskeleton. The fractionation pattern of the cytoskeleton is not easily predicted given its broad interconnectivity with other subcellular components. We observed groups of cytoskeletal proteins with a variety of distinct distributions, including actin, actin-capping modulators, microtubules, microtubule organizing centre components, dyneins, kinesins and myosins (Supplementary Fig. 10). While we did not perform classification on these smaller phenotypes, it is possible that uncharacterized proteins that co-localize with these niches are cytoskeletal components.

**Organelle structure.** The ability of hyperLOPIT to gain insight into very dynamic processes within the cell is exemplified by the Golgi apparatus. The position of well documented Golgi apparatus marker proteins reveals an unexpected observation about these cells. As can be seen from Fig. 4a, these markers fall into four categories; proteins that co-cluster with endoplasmic reticulum (ER) markers, proteins that lie between the ER/Golgi and endosomal compartments, proteins that cluster with plasma membrane markers, and proteins whose steady-state location does not correlate with any of the 14 classified compartments. The proteins that cluster along with ER markers are generally annotated as being cis-Golgi proteins, including alpha-mannosidase 2 (MA2A1)31 and members of the SNAP receptor complex

**Figure 4 | Steady-state location of Golgi apparatus marker proteins.** (a) PCA plot of data with Golgi marker proteins overlaid Golgi proteins (green) peripheral Golgi proteins (yellow), Golgins (pink), microtubular proteins (light blue), dyneins (purple), centrin 2 and associated proteins (red) centrin 3 and associated proteins (mid-blue), Golgi proteins with aberrant localization (orange). (b) Immunocytochemistry using anti-GM130 (red) on a self-renewing population of E14TG2a cultured in serum and LIF showing a punctate localization not consistent with the Golgi Apparatus. (c) Immunocytochemistry using anti-GM130 on a population of E14TG2a after 4 days in N2B27, a media promoting neural differentiation. (d) Immunocytochemistry on a self-renewing population of E14TG2a cultured in serum and LIF using anti-KDEL (green; ER lumenal retention signal) to verify typical endoplasmic reticular structures encapsulating the nucleus.
Proteins that cluster towards the endosomal markers are associated with the trans-Golgi, such as the copper-transporting ATPase (ATP7A) and vacuolar protein sorting-associated protein 45 (VSP45). Several proteins form part of the plasma membrane cluster including Ras-related protein Rab-6A (RAB6A) and Golgin 7 (GOGA7), which are thought to be involved in protein transport from the Golgi to the cell surface. In general, those proteins known to be peripherally associated with Golgi membranes lie in an intermediate position on the boundary of the ER/Golgi cluster, for example, the Golgi phosphoprotein GOLFP3 and oxytyrosine-binding protein 1.

It is curious to note that some Golgi marker proteins occupy a steady-state position far removed from the ER/Golgi or endosomal vesicles. GM130 (GolginA2) is a cis-Golgi matrix protein, and is thought to be involved in regulation of centrosomes during interphase. On the PCA plot (Fig. 4A), GM130 is intriguingly at a steady-state position far removed from the ER/Golgi or cell surface. In general, those proteins known to be peripherally associated with Golgi membranes lie in an intermediate position on the boundary of the ER/Golgi cluster, for example, the Golgi phosphoprotein GOLFP3 and oxytyrosine-binding protein 1.

Protein complexes. In addition to organelles, we observed that many protein complexes display highly correlated distribution patterns. Co-localization of highly abundant macromolecular protein complexes, such as ribosomal subunits and the mitochondrial ATP synthase complex by LOPIT has been previously demonstrated. The depth of proteome coverage in this study, coupled with the high subcellular resolution derived from precise quantification, enabled many more complexes to be detected. We selected 30 examples of highly curated protein complexes listed in KEGG and Reactome, and annotated their distributions in the hyperLOPIT data (Fig. 5). There are many other protein complexes represented in the data set that we have not curated, but which readers of this study can explore according to their particular interest using pRolocGUI.

Proteins with multiple functional roles within the cell may be involved in more than one protein complex. Neither the extent of this, nor the distribution of components between these complexes is typically captured in large-scale complex purification studies. Inspection of the steady-state position of complex components within hyperLOPIT data indicates that components have differing positions from the core complex members, and may even indicate those that function in a regulatory manner. For example, the steady-state location of the TFIID complex is nuclear in these data, however, Taf7 has a steady-state distribution that is distinct from the other complex components.
To demonstrate the utility of these data in evaluating other high-throughput data sets, we annotated the murine orthologues of proteins that were characterized in a census of human soluble protein complexes. By adding this subcellular context to the analysis, we see that well-established components of protein complexes tend to co-localize, whereas novel assignments and putative uncharacterized complexes display more varied localization (Supplementary Fig. 11). Novel assignments that co-localize with the 'core' complex might be assumed to be stable interactors, whereas those with different subcellular distributions may be transient interactors or false assignments. For example, two of eight novel components of the 39S mitochondrial ribosomal subunits assigned by Havugimana et al. were found to co-localize to the mitochondrion, whereas the remaining six were distributed in the nucleus, cytosol and secretory pathway (Supplementary Fig. 11C). The two co-localized proteins (Ict1 and Mrp63) both feature mitochondrial signal peptides and have molecular functions consistent with translation. The mouse and human variants of the six other putative interactors do not contain a distinct signal peptide, and are therefore probable false assignments in the census data set.

**Functional networks.** The high proteome coverage generated by hyperLOPIT enabled us to determine the localization of many proteins associated with pluripotency and differentiation, including components of the core transcriptional network of pluripotent cells such as Sox2, Oct4 and Nanog. Also, impressively, the subcellular distributions of the components of FGF/MAPK, canonical Wnt, Notch, BMP/SMAD, Nodal, Ras and Hippo signalling pathways are apparent, paving the way for using hyperLOPIT to determine modulation in the location of signalling proteins and their effectors on activation/deactivation of multiple signalling pathways during, for example, differentiation (Supplementary Fig. 12). During self-renewal, Wnt/β-catenin and FGF/MAPK signalling are maintained at a low level. Consistent with this, we observe that β-catenin is firmly located in the plasma membrane and that Sprouty proteins, negative regulators of FGF/ERK signalling, are also associated with the membrane, where they act to inhibit the early stages of FGF signalling. We also observe extranuclear localization of Smad2/5, consistent with the known low Nodal/Activin signalling. To our knowledge this is the first instance in which it is possible to have a snapshot of all signalling pathways in one cell.

To evaluate the spatial distribution of interaction partners for the core pluripotency transcription factor triad of Oct4, Sox2 and Nanog, we overlaid information from three high-throughput protein–protein interaction studies onto the E14TG2a hyperLOPIT data set. As might be expected, the seven interaction partners that were common to all three bait proteins displayed a canonical 'long' canonical isoform distribution, whereas the 'short' isoform was found with a steady-state distribution between the cytosol and plasma membrane (Supplementary Fig. 14A). Protein-localization algorithms based on primary sequence, such as WoLF PSORT (ref. 56), support the observation that the long isoform of Lap3 localizes to mitochondria, whereas the short isoform does not (although the specific localization of the short isoform within the cytoplasm is not predictable from sequence alone). This suggests that the two isoforms, while sharing a common catalytic activity, fulfill separate biological roles due to differential localization. The differential localization is achieved through alternative translation that incorporates or excludes an N-terminal mitochondrial target signal. Further credence is given to the dual localization determined by hyperLOPIT by the fact that predicted functional partners of Lap3 (Anpep, Cat, Ggt1, Gss, Hspd1 and Pycr), as reported by the STRING protein interaction database (v9.1 (ref. 57)), are found to localize to the mitochondrion, plasma membrane, and cytosol—the three subcellular compartments described by the steady-state localization of the two Lap3 isoforms.

The differential localization of some other isoforms is not as straightforward to interpret. Two isoforms of Dnmt1, a DNA (cytosine-5)-methyltransferase that modulates ES cell pluripotency by maintaining CpG methylation patterns were identified; a canonical 'long' isoform, and 'short' isoform with an 118 residue N-terminal truncation that removes the Dmap interacting domain. The canonical isoform distribution was consistent with localization to chromatin, and closely co-localizes with Dmap1. The short isoform also appears to localize to the nucleus, but with an atypical distribution profile that while most similar to chromatin, is distinct from the typical chromatin profile (Supplementary Fig. 14B). The significance of this differential nuclear distribution is not immediately clear.

While the number of isoforms pairs identified here is relatively modest, the depth of attainable proteome coverage will increase as the speed and sensitivity of high-resolution accurate mass

### Protein isoforms

HyperLOPIT can also provide information on the localization of protein isoforms. The impact of post-transcriptional modification on protein location has been previously documented; however observation of the differences in subcellular location of closely related isoforms has not been previously possible using LOPIT, as the accuracy of quantification was insufficient to reliably characterize the few peptides, or often single peptide, which distinguish protein isoforms. With the enhanced quantitative performance of TMT quantification by SPS-MS, measurements based on few peptides still yield reliable reporter ion distributions that allow us to characterize localization. We observed unique evidence for 25 and 26 pairs of protein isoforms (distinct proteins sharing the same gene name) in experiments 1 and 2, respectively (Supplementary Data 1). Some of the detected protein isoforms were found to co-localize, while others were found to be differentially localized. For example, two murine isoforms of Leucine aminopeptidase 3 (Lap3) have been reported to arise from alternative translational initiation codons in the same mRNA; a 'long' canonical isoform, and a 'short' isoform with a 31 residue N-terminal truncation. We observed the 'long' isoform of Lap3 with unambiguous mitochondrial localization, whereas the 'short' isoform was found with a steady-state distribution between the cytosol and plasma membrane (Supplementary Fig. 14A). Protein-localization algorithms based on primary sequence, such as WoLF PSORT (ref. 56), support the observation that the long isoform of Lap3 localizes to mitochondria, whereas the short isoform does not.
spectrometry continues to advance, and will allow more of these isoforms to be detected. Accurate quantification of proteins based on a single peptide measurement will also permit characterization of post-translationally modified peptides. The ability to functionally characterize post-translational variants of proteins is an exciting avenue of research, as such variants introduce an additional dimension of biochemical complexity that is not easily evaluated with high throughput.

Discussion

The hyperLOPIT technique is a systematic and accurate assay for characterizing the localization of thousands of proteins in a single experiment, and is applicable to many biological model systems. Application of this technology to a self-renewing population of ES cells has generated the most extensive analysis of protein localization in a stem cell line to date. The data contain information about organelle residency of proteins, sub-organelar structure, the impact of isoform status on location and dynamic localization of proteins. This creates a reference for mapping proteins relative to each other in pluripotency and differentiation.

HyperLOPIT data may also be used to provide a spatial context for pre-established protein complexes and functional networks. For example, our results provide molecular support for some pre-established protein complexes and functional networks.

Application of this technology to a self-renewing population of ES cells is applicable to many biological model systems. The hyperLOPIT technique is a systematic and accurate assay for characterizing the localization of thousands of proteins in a single experiment, and is applicable to many biological model systems.

Methods

Cell cultures: Murine pluripotent ES cells (cell line E14TG2a) were isolated from a 2-week-old C57BL/6J mouse E14.5 embryo. The E14TG2a data set we present acts both as a resource for interrogating the subcellular location of proteins of interest to researchers, and also acts as a scaffold onto which other high content data sets may be mapped to assist in their interpretation.

HyperLOPIT is a powerful tool for gaining insights into fundamental post-translational processes governing stem cell behaviour.

Immunocytochemistry: E14TG2a mES cells were plated and stained as described in ref. 42 using anti-GLI1 (AbCam—EP289Y) and KDEL (AbCam—ab50601), and imaged by confocal microscopy.

Cell lysis and subcellular fractionation: For density-gradient ultracentrifugation, cell pellets were resuspended in 15 ml lysis buffer (0.25 M sucrose, 10 mM HEPES pH 7.4, 2 mM EDTA, 2 mM magnesium acetate) containing protease inhibitors (Roche). Cell suspensions (10%) were reduced, alkylated and digested with trypsin. Briefly, each sample was made up to a concentration of 0.1% w/v and incubated on ice for 8 min to lyse the cells. Nuclei were pelleted by centrifugation at 1,300g, 5 min at 4 °C. The nuclear pellet was resuspended in chromatin buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM dithiothreitol) with protease inhibitors and incubated for 30 min on ice. Samples were then centrifuged at 1,700g, 5 min at 4 °C. The chromatin-enriched pellet was washed in chromatin buffer B, re-pelleted and stored at −20 °C.

Chromatin extraction and enrichment: Chromatin extracts were prepared as previously described61. Briefly, cells were resuspended in chromatin buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.34 M sucrose, 10% glycerol, 1 mM dithiothreitol) with protease inhibitors. Triton X-100 was added to a final concentration of 0.1% w/v and incubation on ice for 8 min to lyse the cells. Nuclei were pelleted by centrifugation at 1,300g, 5 min at 4 °C. The nuclear pellet was resuspended in chromatin buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM dithiothreitol) with protease inhibitors and incubated for 30 min on ice. Samples were then centrifuged at 1,700g, 5 min at 4 °C. The chromatin-enriched pellet was washed in chromatin buffer B, re-pelleted and stored at −20 °C.

Protein digestion and TMT 10-plex labelling: Protein concentrations were determined by BCA assay (Thermo Fisher Scientific) as per the manufacturer’s instructions. Optimal fractions were selected for TMT 10-plex labelling based on western blot verification of organellar marker proteins, protein concentration and refractive indices of membrane fractions. Protein (50 μg) from 10 differentially enriched subcellular fractions (8 membrane fractions from the density gradient, plus cytosol and chromatin-enriched fractions, Supplementary Table 3) was reduced, alkylated and digested with trypsin. Briefly, each sample was made up to a total volume of 50 μl with 25 mM TEAB and reduced and alkylated. Disulfide bonds were reduced with 5 μl of 200 mM tris(2-carboxyethyl)phosphine, 1 h at 37 °C, followed by alkylation of cysteine residues with 5 μl of 375 mM iodoacetamide, 30 min at room temperature. The samples were then diluted tenfold with 25 mM TEAB and digested with sequencing grade trypsin (Promega) for 1 h with a 1:40 enzyme:protein ratio. 37 °C. An additional aliquot of trypsin at 1:40 concentration was added and incubated overnight at 37 °C. Tryptic digest were centrifuged for 10 min at 13,000g to remove any insoluble matter, then reduced to dryness by vacuum centrifugation.

While the TMT tags were equilibrating to room temperature, peptide samples were resuspended in 30 μl 1 M TEAB and 70 μl isopropanol. The solubilized samples were transferred into the tag vials and placed on a shaker for 2 h at room temperature. The reaction was quenched by addition of 8% w/v hydroxyamine for 30 min. The labelled samples were then combined and reduced to dryness by vacuum centrifugation. C18 solid-phase extraction was performed using Sep-Pak cartridges (100 mg bed volume, Waters) and peptides were eluted in 70% acetonitrile + 0.05% acetic acid. The eluate was again reduced to dryness by vacuum centrifugation, and resuspended in 20 mM ammonium formate (pH 10.0), for high pH reversed-phase liquid chromatography.

Sample fractionation: Desalted peptides were resuspended in 0.1 ml 20 mM ammonium formate (pH 10.0) + 4% (v/v) acetonitrile. Peptides were loaded onto an Acquity bridged ethyl hybrid C18 UPLC column (Waters; 2.1 mm inner diameter × 150 mm, 1.7 μm particle size), and profiled with a linear gradient of 5–60% acetonitrile + 20 mM ammonium formate (pH 10.0) over 60 min, at a flow rate of 0.25 ml min−1. Chromatographic performance was monitored by sampling eluate with a diode array detector (Acquity UPLC, Waters) scanning between wavelengths of 200 and 400 nm. Fractions were collected at 1 min intervals. Twenty-four fractions representing peak peptide elution were selected for mass spectrometry analysis and resuspended in 0.05% trifluoroacetic acid.

Approximately 1 μg peptides were loaded per liquid chromatography–mass spectrometry run.

Mass spectrometry: All mass spectrometry experiments were performed on an Orbitrap Fusion coupled with a Proxeon EASY-nLC 1000 (Thermo Fisher Scientific). Peptides were separated on a Proxeon EASY-Spray column (Thermo
Scientific; 50 cm
ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9992
formic acid over 95 min at 300 nl min
collision energy, followed by analysis of fragment ions in the Orbitrap with
selected from full scans as described above. For each selected peptide ion, a
MS2 spectra were acquired between full scans was restricted to a duty cycle of 3 s.
To assess the effect of using different numbers of precursors for SPS, ions were
select from full scans as described above. For each selected peptide ion, a
sequence of six spectra was generated (conventional MS2, SPS with 15, 10, 5 and 2
precursors, and conventional MS3). The precursor ion for conventional MS2
was selected as the most intense ion within the 400-950 m/z range, excluding the
unfragmented peptide and TMTC ion series, and isolated with a 2 m/z selection
window. Conventional MS2 and SPS-MS2 were performed with the parameters
described previously. The duty cycle for the sequence of scans was fixed at 6 s,
with each sequence of six scans taking 1.5–2.5 s.

Data processing. Raw data files were processed using Proteome Discoverer
(v1.4, Thermo Fisher Scientific), interfaced with Mascot server (v2.3.02, Matrix
Science). Mascot searches were performed against SwissProt mouse database
(March 2013, 24,481 sequences), with carbamidomethylation of cysteine, and TMT
10-plex modification of lysine and peptide N termini set as modifications. For the SPS-MS3 method, quantification was performed at the MS3 level. Spectra with more than four missing reporter ion values were excluded from quantification, and remaining missing values were set to zero for downstream analysis. For protein-level reporting, protein grouping was enabled, and values were calculated from the median of all quantifiable PSMs for each group. TMT values were then reported as a ratio to the TMTC ion series 62. AGC targets and maximum accumulation times to 50 and 120 ms, for full and MS2 scans, respectively. Ions with +1 or +2 charge state were excluded from SPS-MS2 selection. Ions within a ±10 p.p.m. m/z window around selected ions for SPS-MS2 were excluded from further selection for fragmentation for 35 s.

For the SPS-MS2 method, the full scan parameters were identical to those for the MS3 method. The most intense ions above a threshold of 2 x 105 were selected for collision induced dissociation (CID)-MS2 fragmentation, with an AGC target and maximum accumulation time of 1 x 106 and 70 ms. Mass filtering was performed by the quadrupole with 1.5 m/z transmission window, followed by CID fragmentation in the linear ion trap with 35% normalized collision energy, SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–950 m/z range, and were set to preclude selection of the precursor ion and TMTC ion series 62. AGC targets and maximum accumulation times were set to 1 x 106 and 120 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 55% normalized collision energy, and were analysed in the Orbitrap with nominal resolution of 6 x 104. The number of HCD-MS3 events between full scans was determined on-the-fly so that the duty cycle was fixed at 3 s.

The automatic gain control (AGC) settings were 4 x 106 ions and 1 x 106 ions, and maximum ion accumulation times to 50 and 120 ms, for full and MS2 scans, respectively. Ions with +1 or +2 charge state were excluded from MS2 selection. Ions within a ±10 p.p.m. m/z window around selected ions for SPS-MS2 were excluded from further selection for fragmentation for 35 s.

Scientific; 50 cm
ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9992
formic acid over 95 min at 300 nl min
collision energy, followed by analysis of fragment ions in the Orbitrap with
selected from full scans as described above. For each selected peptide ion, a
MS2 spectra were acquired between full scans was restricted to a duty cycle of 3 s.
To assess the effect of using different numbers of precursors for SPS, ions were
select from full scans as described above. For each selected peptide ion, a
sequence of six spectra was generated (conventional MS2, SPS with 15, 10, 5 and 2
precursors, and conventional MS3). The precursor ion for conventional MS2
was selected as the most intense ion within the 400-950 m/z range, excluding the
unfragmented peptide and TMTC ion series, and isolated with a 2 m/z selection
window. Conventional MS2 and SPS-MS2 were performed with the parameters
described previously. The duty cycle for the sequence of scans was fixed at 6 s,
with each sequence of six scans taking 1.5–2.5 s.

Data processing. Raw data files were processed using Proteome Discoverer
(v1.4, Thermo Fisher Scientific), interfaced with Mascot server (v2.3.02, Matrix
Science). Mascot searches were performed against SwissProt mouse database
(March 2013, 24,481 sequences), with carbamidomethylation of cysteine, and TMT
10-plex modification of lysine and peptide N termini set as modifications. For the SPS-MS3 method, quantification was performed at the MS3 level. Spectra with more than four missing reporter ion values were excluded from quantification, and remaining missing values were set to zero for downstream analysis. For protein-level reporting, protein grouping was enabled, and values were calculated from the median of all quantifiable PSMs for each group. TMT values were then reported as a ratio to the TMTC ion series 62. AGC targets and maximum accumulation times to 50 and 120 ms, for full and MS2 scans, respectively. Ions with +1 or +2 charge state were excluded from SPS-MS2 selection. Ions within a ±10 p.p.m. m/z window around selected ions for SPS-MS2 were excluded from further selection for fragmentation for 35 s.

For the SPS-MS2 method, the full scan parameters were identical to those for the MS3 method. The most intense ions above a threshold of 2 x 105 were selected for collision induced dissociation (CID)-MS2 fragmentation, with an AGC target and maximum accumulation time of 1 x 106 and 70 ms. Mass filtering was performed by the quadrupole with 1.5 m/z transmission window, followed by CID fragmentation in the linear ion trap with 35% normalized collision energy, SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–950 m/z range, and were set to preclude selection of the precursor ion and TMTC ion series 62. AGC targets and maximum accumulation times were set to 1 x 106 and 120 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 55% normalized collision energy, and were analysed in the Orbitrap with nominal resolution of 6 x 104. The number of HCD-MS3 events between full scans was determined on-the-fly so that the duty cycle was fixed at 3 s.

The automatic gain control (AGC) settings were 4 x 106 ions and 1 x 106 ions, and maximum ion accumulation times to 50 and 120 ms, for full and MS2 scans, respectively. Ions with +1 or +2 charge state were excluded from MS2 selection. Ions within a ±10 p.p.m. m/z window around selected ions for SPS-MS2 were excluded from further selection for fragmentation for 35 s.

For the SPS-MS2 method, the full scan parameters were identical to those for the MS3 method. The most intense ions above a threshold of 2 x 105 were selected for collision induced dissociation (CID)-MS2 fragmentation, with an AGC target and maximum accumulation time of 1 x 106 and 70 ms. Mass filtering was performed by the quadrupole with 1.5 m/z transmission window, followed by CID fragmentation in the linear ion trap with 35% normalized collision energy, SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–950 m/z range, and were set to preclude selection of the precursor ion and TMTC ion series 62. AGC targets and maximum accumulation times were set to 1 x 106 and 120 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 55% normalized collision energy, and were analysed in the Orbitrap with nominal resolution of 6 x 104. The number of HCD-MS3 events between full scans was determined on-the-fly so that the duty cycle was fixed at 3 s.

The automatic gain control (AGC) settings were 4 x 106 ions and 1 x 106 ions, and maximum ion accumulation times to 50 and 120 ms, for full and MS2 scans, respectively. Ions with +1 or +2 charge state were excluded from MS2 selection. Ions within a ±10 p.p.m. m/z window around selected ions for SPS-MS2 were excluded from further selection for fragmentation for 35 s.

For the SPS-MS2 method, the full scan parameters were identical to those for the MS3 method. The most intense ions above a threshold of 2 x 105 were selected for collision induced dissociation (CID)-MS2 fragmentation, with an AGC target and maximum accumulation time of 1 x 106 and 70 ms. Mass filtering was performed by the quadrupole with 1.5 m/z transmission window, followed by CID fragmentation in the linear ion trap with 35% normalized collision energy, SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–950 m/z range, and were set to preclude selection of the precursor ion and TMTC ion series 62. AGC targets and maximum accumulation times were set to 1 x 106 and 120 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 55% normalized collision energy, and were analysed in the Orbitrap with nominal resolution of 6 x 104. The number of HCD-MS3 events between full scans was determined on-the-fly so that the duty cycle was fixed at 3 s.

The automatic gain control (AGC) settings were 4 x 106 ions and 1 x 106 ions, and maximum ion accumulation times to 50 and 120 ms, for full and MS2 scans, respectively. Ions with +1 or +2 charge state were excluded from MS2 selection. Ions within a ±10 p.p.m. m/z window around selected ions for SPS-MS2 were excluded from further selection for fragmentation for 35 s.

For the SPS-MS2 method, the full scan parameters were identical to those for the MS3 method. The most intense ions above a threshold of 2 x 105 were selected for collision induced dissociation (CID)-MS2 fragmentation, with an AGC target and maximum accumulation time of 1 x 106 and 70 ms. Mass filtering was performed by the quadrupole with 1.5 m/z transmission window, followed by CID fragmentation in the linear ion trap with 35% normalized collision energy, SPS was applied to co-select 10 fragment ions for HCD-MS3 analysis. SPS ions were all selected within the 400–950 m/z range, and were set to preclude selection of the precursor ion and TMTC ion series 62. AGC targets and maximum accumulation times were set to 1 x 106 and 120 ms. Co-selected precursors for SPS-MS3 underwent HCD fragmentation with 55% normalized collision energy, and were analysed in the Orbitrap with nominal resolution of 6 x 104. The number of HCD-MS3 events between full scans was determined on-the-fly so that the duty cycle was fixed at 3 s.
8. Shyh-Chang, N. & Daley, G. Q. Lin28: primal regulator of growth and metabolism in stem cells. Cell stem cell 12, 395–406 (2013).
9. Uhlen, M. et al. The UPR-UPer database: a knowledge-based human protein atlas. Nat. Biotechnol. 28, 1248–1250 (2010).
10. Stadler, C. et al. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nat Methods 10, 313–323 (2013).
11. Christoforou, A., Mulvey, C. M., Breckels, L. M., Gatto, L. & Lilley, K. S. in Quantitative Proteomics New Developments in Mass Spectrometry (eds Evers, C. & Gaskell, Simon) Ch. 9 185–210 (Royal Society of Chemistry, 2014).
12. Dunkley, T. P. J. et al. Mapping the arabidopsis organelle proteome. Proc. Natl Acad. Sci. USA 103, 6518–6523 (2006).
13. 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).
14. de Duve, C. Tissue fractionation. J. Cell Biol. 50, 20D–55D (1971).
15. McAlister, G. C. et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal. Chem. 84, 7469–7478 (2012).
16. Christoforou, A. & Lilley, K. Isobaric tagging approaches in quantitative proteomics: the ups and downs. Anal. Bioanal. Chem. 404, 1029–1037 (2012).
17. Ting, L., Rad, R., Gygi, S. G. & Haas, W. M53 eliminates ratio distortion in isobaric labeling multiplexed quantitative proteomics. Nat. Methods 9, 937–940 (2011).
18. Wenger, C. D. et al. Gas-phase purification enables accurate, large-scale, multiplexed proteome quantification. Nat. Methods 8, 933–935 (2011).
19. Dayan, I., Sonder, B. & Kussmann, M. Combination of gas-phase fractionation and MS3 acquisition modes for relative protein quantification with isobaric tagging. J. Proteome Res. 11, 5081–5089 (2012).
20. Vachet, R. W. & McElvany, S. W. Application of external customized waveforms to a commercial quadrupole ion trap. J. Am. Soc. Mass Spectrom. 10, 355–359 (1999).
21. 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 (2014).
22. Ow, S. Y. et al. iTRAQ underestimation in simple and complex mixtures: “the good, the bad and the ugly”. J. Proteome Res. 8, 5347–5355 (2009).
23. Trotter, M. W. B., Sadowski, P. G., Dunkley, T. P. J., Groen, A. J. & Lilley, K. S. Improved sub-cellular resolution via simultaneous analysis of organelle proteomes data across varied experimental conditions. Proteomics 10, 4213–4219 (2010).
24. Groen, A. J. et al. Identification of trans-Golgi network proteins in arabidopsis thaliana root tissue. J. Proteome Res. 13, 763–776 (2013).
25. Gatto, L., Breckels, L. M., Wieczorek, S., Burger, T. & Lilley, K. S. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. Bioinformatics 30, 1322–1324 (2014).
26. Breckels, L. M. et al. The effect of organelle discovery upon sub-cellular protein localisation. J. Proteomics 88, 129–140 (2013).
27. Bausch-Fluck, D. et al. A Mass spectrometric-derived cell surface protein atlas. PLoS One 10, e0121314 (2015).
28. Faustner, F. et al. Membrane-associated β-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. Development 140, 1117–1183 (2013).
29. Krajewski, S. et al. Immunohistochemical analysis of Mcl-1 and Bel-2 proteins in normal and neoplastic lymph nodes. Am. J. Pathol. 145, 515–525 (1994).
30. Betschinger, J. et al. Improved sub-cellular resolution via simultaneous analysis of organelle proteomes. J. Proteome Res. 35, W585–W587 (2007).
31. Franceschini, A. et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, D808–D815 (2013).
32. Moor, K. N., Ding, F. & Chaillet, J. R. Distinct roles of DMAP1 in mouse development. Mol. Cell. Biol. 31, 1861–1869 (2011).
33. Grohmann, M. et al. Restricted mobility of Dnr1/TOM40 in mouse embryonic stem cells. J. Cell Biol. 178, 668–699 (2008).
34. Ahmad, Y., Boisvert, F.-M., Lundberg, E., Uhlen, M. & Lamond, A. I. Systematic analysis of protein pools, isoforms, and modifications affecting turnover and subcellular localization. Mol. Cell. Proteomics 11, 1–11 doi:10.1074/mcp.M111.013680 (2012).
35. Horton, P. et al. WolFL PSORT: protein localization predictor. Nucleic Acids Res. 35, W585–W587 (2007).
36. Stehbens, S. J. Mapping organelle proteins and protein complexes in drosophila melanogaster. Nucleic Acids Res. 35, W585–W587 (2007).
37. Montano, A. CentrosomeDB: a human centrosomal proteins database. Nucleic Acids Res. 40, D175–D180 (2012).
38. Andersen, J. S. et al. Protemic characterization of the human centrosome by protein correlation profiling. Nature 426, 570–574 (2003).
39. Das, A., Diques-Cadenas, R., Abascal, F., Diez-Perez, J., Carazo, J. M. & Pascual-Montano, A. CentrosomeDB: a human centrosomal proteins database. Nucleic Acids Res. 37, D175–D180 (2009).
40. Jakobsen, L. et al. Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. EMBO J. 30, 1520–1535 (2011).
41. Gatto, L. & Lilley, K. S. The Golgi and the centrosome: building a functional partnership. J. Cell Biol. 188, 621–628 (2010).
42. Turner, D. A., Trot, J., Hayward, P., Ruc, P. & Martinez Arias, A. An interplay between extracellular signalling and the dynamics of the exit from pluripotency drives cell fate decisions in mouse ES cells. Biol. Open 3, 614–626 (2014).
43. Coronado, D. et al. A short G1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency. Nat. Cell Biol. 10, 731–739 (2008).
44. Tan, D. J. L. et al. Mapping organelle proteins and protein complexes in drosophila melanogaster. J. Proteome Res. 8, 2667–2678 (2009).
45. Kanelisa, M. et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199–D205 (2014).
46. Milacic, M. et al. Annotating cancer variants and anti-cancer therapeutics in React2, Cancers 4, 1180–1211 (2012).
47. Croft, D. et al. The reactome pathway knowledgebase. Nucleic Acids Res. 42, D472–D477 (2014).
48. Gogonea, A. et al. TFF4 component TAF7 functionally interacts with both TFF4 and T-Fe65. Proc. Natl Acad. Sci. USA 105, 5367–5372 (2008).
49. Das, A. et al. RAla promotes a direct exocytosis–Part interaction to regulate polarity in neuronal development. J. Cell Sci. 127, 686–699 (2014).
50. Havugimana, P. C. A census of human soluble protein complexes. Cell 150, 1088–1081 (2012).
51. van den Berg, D. L. C. et al. An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 6, 369–381 (2010).
52. Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric labeling multiplexed quantitative proteomics. J. Proteome Res. 10, 8435–8441 (2011).
53. Stehbens, S. J. The Golgi protein GM130 regulates centrosome morphology and function. Mol. Biol. Cell. 19, 745–753 (2008).
of the Orbitrap Fusion. We also thank Mike Deery for assistance with checking sample integrity on the mass spectrometers in the Cambridge Centre for Proteomics on equipment purchased via a Wellcome Trust grant (099135/Z/12/Z), and Brian Hendrich of the Wellcome Trust-MRC Stem Cell Institute in Cambridge and Sean Munro of the MRC Laboratory of Molecular Biology in Cambridge for insightful comments about the data. A.C. was supported by BBSRC grant (BB/D526088/1), C.M.M. and L.G. were supported by European Union 7th Framework Program (PRIMEXS project, grant agreement number 262067), L.M.B. was supported by a BBSRC Tools and Resources Development Fund (Award BB/K00137X/1), and P.C.H. was supported by an ERC Advanced Investigator grant to A.M.A. A.G. was funded through the Alexander S. Onassis Public Benefit Foundation, the Foundation for Education and European Culture (IPEP) and the Embiricos Trust Scholarship of Jesus College Cambridge. T.H. was supported by Commonwealth Split Site PhD Scholarship. T.N. was supported by an ERASMUS Placement scholarship.

Author contributions
A.C., C.M.M., A.M.A and K.S.L. designed and planned the experiments. A.C., C.M.M., A.G., T.H. and P.C.H. performed the cell culture and sample preparation. A.C., C.M.M. and R.V. performed the mass spectrometry. A.C., L.M.B. and L.G. performed the multivariate data analysis. L.M.B., T.N. and L.G. created tools and tutorials for data visualization. All authors contributed to the technical and biological interpretation of data, and to the writing of the manuscript.

Additional information
Accession codes: The mass spectrometry data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD001279.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Christoforou, A. et al. A draft map of the mouse pluripotent stem cell spatial proteome. Nat. Commun. 7:8992 doi: 10.1038/ncomms9992 (2016).