



ORIGINAL RESEARCH

Global Profiling of the Lysine Crotonylome in Different Pluripotent States



Yuan Lv^{1,2,3,#}, Chen Bu^{4,#}, Jin Meng^{1,2,5}, Carl Ward^{1,2}, Giacomo Volpe^{1,2},
 Jieyi Hu^{1,2,3}, Mengling Jiang^{1,2}, Lin Guo², Jiekai Chen^{2,3,5,6},
 Miguel A. Esteban^{1,2,5,6,7,*}, Xichen Bao^{2,6,8,*}, Zhongyi Cheng^{4,*}

¹ Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

² CAS Key Laboratory of Regenerative Biology and Guangdong Provincial Key Laboratory of Stem Cells and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Jingjie PTM BioLab (Hangzhou) Co. Ltd, Hangzhou 310018, China

⁵ Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health and Guangzhou Medical University, Guangzhou 511436, China

⁶ Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510005, China

⁷ Institute of Stem Cells and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

⁸ Laboratory of RNA Molecular Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

Received 16 December 2019; revised 21 December 2020; accepted 27 January 2021

Available online 19 March 2021

Handled by Minjia Tan

KEYWORDS

Metabolism;
 Crotonylation;
 Pluripotency;
 RNA-binding proteins;
 Proteasome

Abstract Pluripotent stem cells (PSCs) can be expanded *in vitro* in different culture conditions, resulting in a spectrum of cell states with distinct properties. Understanding how PSCs transition from one state to another, ultimately leading to lineage-specific differentiation, is important for developmental biology and regenerative medicine. Although there is significant information regarding gene expression changes controlling these transitions, less is known about post-translational modifications of proteins. Protein **crotonylation** is a newly discovered post-translational modification where lysine residues are modified with a crotonyl group. Here, we employed affinity purification of crotonylated peptides and liquid chromatography–tandem mass spectrometry (LC–MS/MS) to systematically profile protein crotonylation in mouse PSCs in different states including ground, metastable, and primed states, as well as metastable PSCs undergoing early **pluripotency**

* Corresponding authors.

E-mail: miguel@gibh.ac.cn (Esteban MA), bao_xichen@gibh.ac.cn (Bao X), zhongyi_cheng@ptm-biolab.com (Cheng Z).

Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2021.01.004>

1672-0229 © 2021 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

exit. We successfully identified 3628 high-confidence crotonylated sites in 1426 proteins. These crotonylated proteins are enriched for factors involved in functions/processes related to pluripotency such as RNA biogenesis, central carbon **metabolism**, and **proteasome** function. Moreover, we found that increasing the cellular levels of crotonyl-coenzyme A (crotonyl-CoA) through crotonic acid treatment promotes proteasome activity in metastable PSCs and delays their differentiation, consistent with previous observations showing that enhanced proteasome activity helps to sustain pluripotency. Our atlas of protein crotonylation will be valuable for further studies of pluripotency regulation and may also provide insights into the role of metabolism in other cell fate transitions.

Introduction

Pluripotency is a transient state of the developing embryo. Through this property, cells in the inner cell mass (ICM) of the blastocyst have the capacity to differentiate into all tissues that compose the body [1]. *In vitro*, pluripotency can be maintained by culturing early embryonic cells in specific conditions, which then produces pluripotent stem cells (PSCs). Accordingly, PSCs are self-renewing, and can be maintained indefinitely *in vitro* and differentiated on demand upon exposure to specific signaling cues. Studying the regulation of pluripotency maintenance and exit in PSCs is important for understanding not only embryonic development but also cell fate transitions in other contexts (e.g., somatic cell reprogramming and transdifferentiation, cancer, and aging). Mouse PSCs are a well-studied pluripotency model and can be cultured in different conditions: ground state [in serum-free medium with two inhibitors (MEK and GSK3) and leukemia inhibitory factor (LIF), or 2i embryonic stem cells (ESCs)], metastable state (in a classical condition containing serum and LIF, or S/L ESCs), and primed state [in serum-free medium with bFGF and Activin A, or epiblast stem cells (EpiSCs)] [2,3]. Each of these pluripotent states represents a downward progression from the early pre-implantation blastocyst to the post-implantation embryo. PSCs in ground state more closely resemble the pre-implantation embryo and produce chimeras upon blastocyst complementation more efficiently than metastable PSCs. On the contrary, EpiSCs resemble the post-implantation epiblast and have negligible capacity to contribute to chimeras. The term “metastable” refers to the tendency of S/L ESCs to spontaneously oscillate between more naïve (ICM-like) conditions and the primed state [4]. Notably, cells in these three pluripotent states display differences in their epigenome, transcriptome, and metabolome, as their functions are controlled by different signaling pathways [5,6]. Regarding metabolism, naïve PSCs rely on both glycolysis and oxidative phosphorylation [through the tricarboxylic acid (TCA) cycle] for producing energy, whereas EpiSCs mostly rely on glycolysis [6,7].

Besides providing energy, cell metabolism regulates myriad cellular functions comprehensively including the modifications of DNA/histones, the transcriptome, and the proteome. Because metabolic features are variable in different cell types, changing metabolism can influence cell fate transitions. For instance, a rapid decrease of glycolysis during mouse ESC differentiation leads to reduced abundance of acetyl-coenzyme A (acetyl-CoA) and a consequent decline in histone lysine acetylation (an epigenetic mark associated with gene activation) at pluripotency loci [8]. Likewise, down regulation of SAM (S-adenosyl-methionine) caused by threonine depletion results in decreased trimethylation of histone H3 lysine 4, which slows

PSC proliferation and facilitates differentiation [9]. Recently, a group of novel metabolites derived from short-chain fatty acids (e.g., propionyl-CoA, crotonyl-CoA, butyryl-CoA, and myristoyl-CoA) have been shown to be substrates for lysine acylation of not only histones but also non-histone proteins involved in many cellular processes [10,11]. Post-translational modifications based on lysine acylation are functionally relevant and distinct from protein lysine acetylation mediated by acetyl-CoA, further strengthening the link between metabolism and cellular functions beyond energy production. For example, during starvation histone H3 lysine 9 β -hydroxybutyrylation activates responsive genes in the mouse liver to induce adaption [12]. Similarly, lysine myristoylation of germline proteins can modulate the MPK-1/MAPK pathway to impact sex determination and reproductive development in *Caenorhabditis elegans* [13]. Moreover, lysine crotonylation (Kcr) in histones activates gene expression through yet unclear mechanisms, and a large repertoire of non-histone proteins can also be crotonylated [14,15]. For example, crotonylation of RPA1 (replication protein A1) promotes homologous recombination-mediated DNA repair by enhancing its interaction with single-strand DNA [16]. Yet, comprehensive analysis of these novel post-translational modifications, including crotonylation, in different cell types and during cell fate transitions is largely lacking.

In this study, we have constructed an atlas of Kcr, with focus on non-histone proteins, in mouse PSCs in different states (2i ESCs, S/L ESCs, and EpiSCs) and also in S/L ESCs triggered to exit pluripotency by removing LIF. We have identified 3628 high-confidence Kcr sites on 1426 proteins among these four cell states. These sites reveal links between crotonylation and cell functions relevant to pluripotency maintenance, including RNA biogenesis, carbon metabolism, and proteasome. Consistently, we observed that increased protein crotonylation caused by crotonic acid treatment up-regulates pluripotency gene expression, delays differentiation, and enhances proteasome activity, a function relevant to pluripotency maintenance [17,18]. Our findings provide a useful resource for understanding how metabolism regulates cell identity through crotonyl-CoA production, in particular the transition between different pluripotent states and differentiating cells.

Results and discussion

Derivation of PSCs in different pluripotent states and differentiating cells for quantitative lysine crotonylome analysis

We aimed to investigate protein crotonylation in mouse PSCs in different states: 2i ESCs, S/L ESCs, and EpiSCs, and also in

S/L ESCs triggered to exit pluripotency by LIF withdrawal for 4 days (hereafter referred to as differentiating cells) (**Figure 1A**). To monitor the different PSC states, we used ESCs and EpiSCs derived from crossed offspring of 129 female mice and OG2

transgenic male mice. These cells contain multiple copies of an *Oct4* distal enhancer-driven GFP reporter activated only in naïve pluripotency conditions [19,20]. ESCs in 2i and S/L and S/L showed bright GFP fluorescence as well as the characteristic

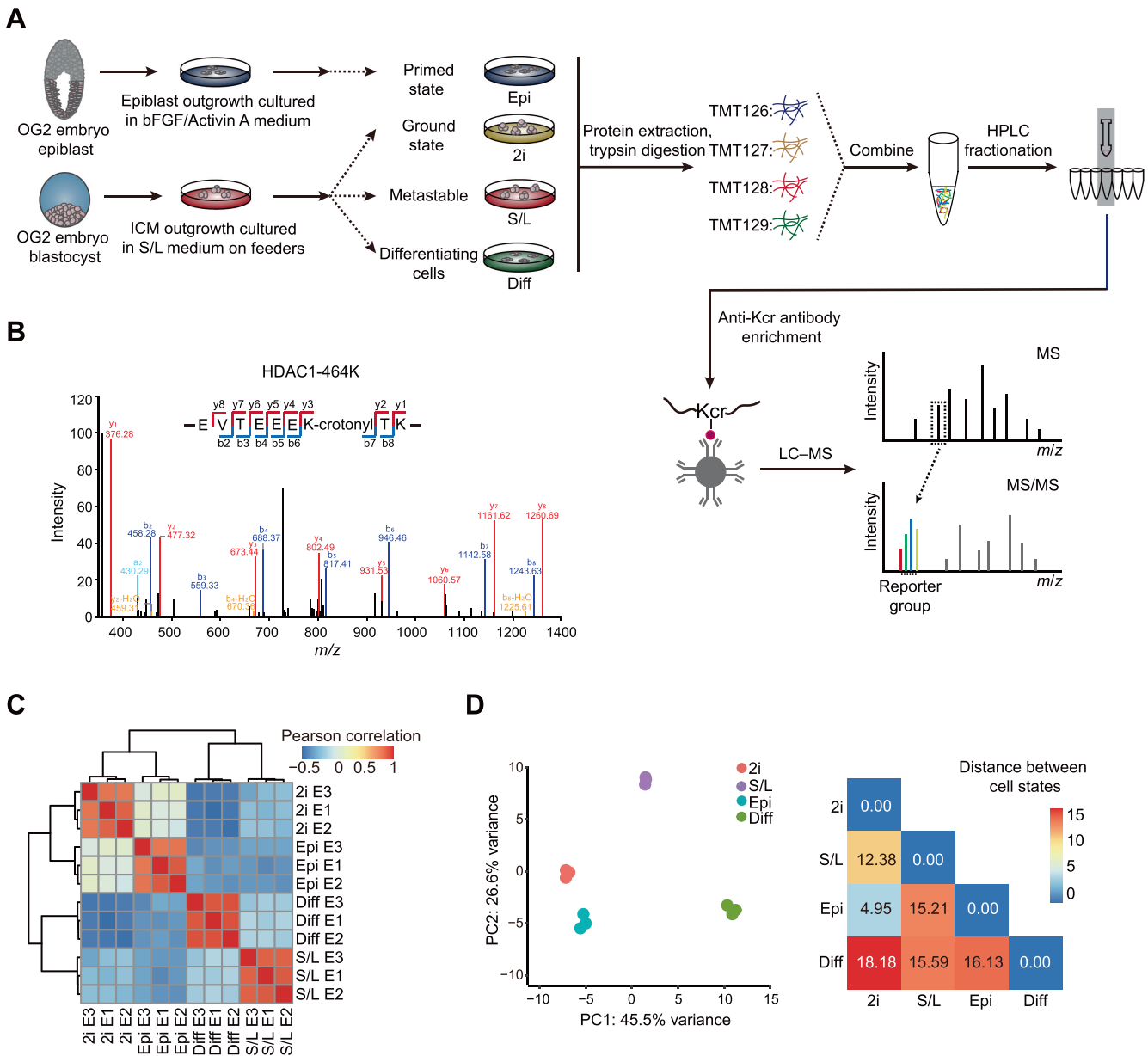


Figure 1 Generation of a high-confidence crotonylome map in three pluripotent states and differentiating cells

A. Schematic view showing the culture of cells in four different cell states and the workflow for the crotonylation profiling. 2i represents ground state mouse ESCs cultured in serum-free medium with two inhibitors (MEK and GSK3) and LIF; S/L represents metastable state mouse ESCs cultured in classical conditions containing serum and LIF; Epi represents primed state mouse EpiSCs cultured in serum-free medium with cytokines bFGF and Activin A; Diff represents differentiating cells by LIF withdrawal for 4 days in S/L ESCs. **B.** Identification of HDAC1 crotonylation in LC-MS/MS. The MS/MS spectrum at m/z 1174.62 Da matches with the peptide “EVTEEEKTK” being crotonylated at K7 (K7-crotonyl, 68.023 Da). **C.** Hierarchical clustering of the normalized \log_2 intensity for crotonylated peptides of three LC-MS/MS experiments (biological replicates) in the four different cell states. Colors in the heatmap indicate the pairwise Pearson correlation between the different samples ($n = 3628$). **D.** Left: PCA of the crotonylome in the four different cell states. The variance was calculated using normalized \log_2 intensity. Right: Euclidean distances of crotonylation profiles between the indicated cell states. ESC, embryonic stem cell; LIF, leukemia inhibitory factor; ICM, inner cell mass; TMT, tandem mass tag; HPLC, high-performance liquid chromatography; Kcr, lysine crotonylation; LC-MS/MS, liquid chromatography–tandem mass spectrometry; PCA, principal component analysis.

domed colony shape, whereas EpiSCs and differentiating cells had lost the GFP signal and displayed flat and flat/irregular colony shapes, respectively (Figure S1A). To further verify the different cell identities, we performed reverse transcription-quantitative PCR (RT-qPCR) for two general pluripotency markers (*Nanog* and *Oct4/Pou5f1*), two pluripotent state-specific regulators (*Lin28a* and *Myc*) [21,22], and a EpiSC/differentiation marker (*Fgf5*). EpiSCs and differentiating cells showed higher expression of *Fgf5* compared to naïve ESCs (2i or S/L conditions), and 2i ESCs showed the lowest expression of *Myc*, *Lin28a*, and *Fgf5*, but the highest expression of *Nanog* and, to a lesser extent, *Oct4* (Figure S1B). These results demonstrate that our culture conditions truly represent the aforementioned cell states.

Next, we quantified the lysine crotonylome of cells in these four culture conditions. To this end, we labeled the digested peptides from different conditions with tandem mass tags (TMTs). We combined and fractionated the labeled peptides with high-performance liquid chromatography (HPLC), and then enriched them with anti-Kcr antibodies [14] before conducting liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis. Altogether, we identified 8102 different Kcr sites in 2578 proteins among the four cell types, which were divided into high-confidence (3628 sites in 1426 proteins) and low-confidence (4474 sites in 1152 proteins) candidates, respectively (Table 1, Table S1A). High-confidence crotonylated proteins were defined as quantified in at least two out of three experiments, and low-confidence as quantified only in one experiment. For example, we detected the chromatin regulator HDAC1 (Histone deacetylase 1), a known target of crotonylation [23], among the high-confidence crotonylated proteins (Figure 1B). In addition, we analyzed the total proteome (Table 1). Both the total proteome and the lysine crotonylome showed high reproducibility in three biological replicates (Figure 1C, Figure S1C).

We then integrated our total proteome with the publicly available ‘core ESC-like gene module’ and ‘adult tissue stem module’, which represent pluripotency regulators and differentiation regulators enriched in mouse ESCs and adult stem cells, respectively [24]. As anticipated, the ‘core ESC-like gene module’ scored lowest in differentiating cells, whereas the ‘adult tissue stem module’ scored lowest in 2i ESCs but highest in differentiating cells (Figure S1D). Importantly, principal component analysis (PCA) of our crotonylome and proteome datasets could also distinguish cells in the four cell states (Figure 1D, Figure S1G). The largest Euclidean distances in the crotonylome and proteome PCA were between the 2i ESCs and differentiating cells (Figure 1D, Figure S1G), consistent with the functional divergence between them. These observa-

tions support the notion that protein crotonylation may be functionally important for maintaining cell identity in these different culture conditions. In this regard, protein crotonylation is controlled by the balance between writers and erasers [25], whose changes in expression could explain putative differences in crotonylation between the four cell states. However, analysis of the total proteome showed that the protein expression levels of known crotonylation writers and erasers, as well as readers, do not change noticeably between the four cell states (Figure S1E). Likewise, global protein crotonylation was at an equivalent level (Figure S1F).

In summary, we have successfully profiled global protein crotonylation in different pluripotent states including 2i ESCs, S/L ESCs, and EpiSCs, and also in early differentiating cells.

Crotonylation is a widely prevalent modification in PSCs in different pluripotent states and differentiating cells

We next studied the features of the 1426 high-confidence crotonylated proteins identified among the four cell states. Of note, 686 proteins showed at least two Kcr sites, whereas 30 showed more than 12 Kcr sites (Figure 2A; Table S1B). Analysis of the subcellular localization using the COMPARTMENTS database [26] showed that these crotonylated proteins exert different functions in multiple compartments including nucleus, cytosol, plasma membrane, cytoskeleton, and mitochondria (Figure 2B; Table S1C). We also compared our complete Kcr dataset with published crotonylomes from different human cell lines including H1299 [15], HeLa [23], A549 [27], and HCT116 cells [28], as well as peripheral blood from patients with kidney failure [29], all of which described identified crotonylated proteins without selection of high-confidence ones. Our dataset is the largest, with 886 out of the total 2578 crotonylated proteins (high- and low-confidence candidates) being unique (Figure 2C; Table S1D). To validate the identified candidates, we selected and overexpressed 15 high-confidence proteins and 1 low-confidence protein with a FLAG tag in HEK293T cells, followed by immunoprecipitation and Western blotting with anti-PAN Kcr antibodies. The substrate for crotonyl-CoA generation, crotonic acid, was added to enhance the detection. Importantly, 11 out of 16 candidates showed basal Kcr signals and crotonic acid treatment boosted the basal signals in most of the selected candidates (15 out of 16) (Figure 2D, Figure S2). In total 3 out of the 15 validated proteins, including the pluripotency regulator LIN28A and the RNA m⁶A reader YTHDF2 [30], were exclusive to our dataset. These results

Table 1 Summary of proteome and crotonylome profiling results

Type	No. of crotonylated proteins	No. of crotonylated sites	Proteome
High-confidence	1426	3628	4728
Low-confidence	1152	4474	1158
Total identified	2578	8102	5886

Note: High-confidence, quantified in at least two out of three experiments; Low-confidence, quantified only in one experiment.

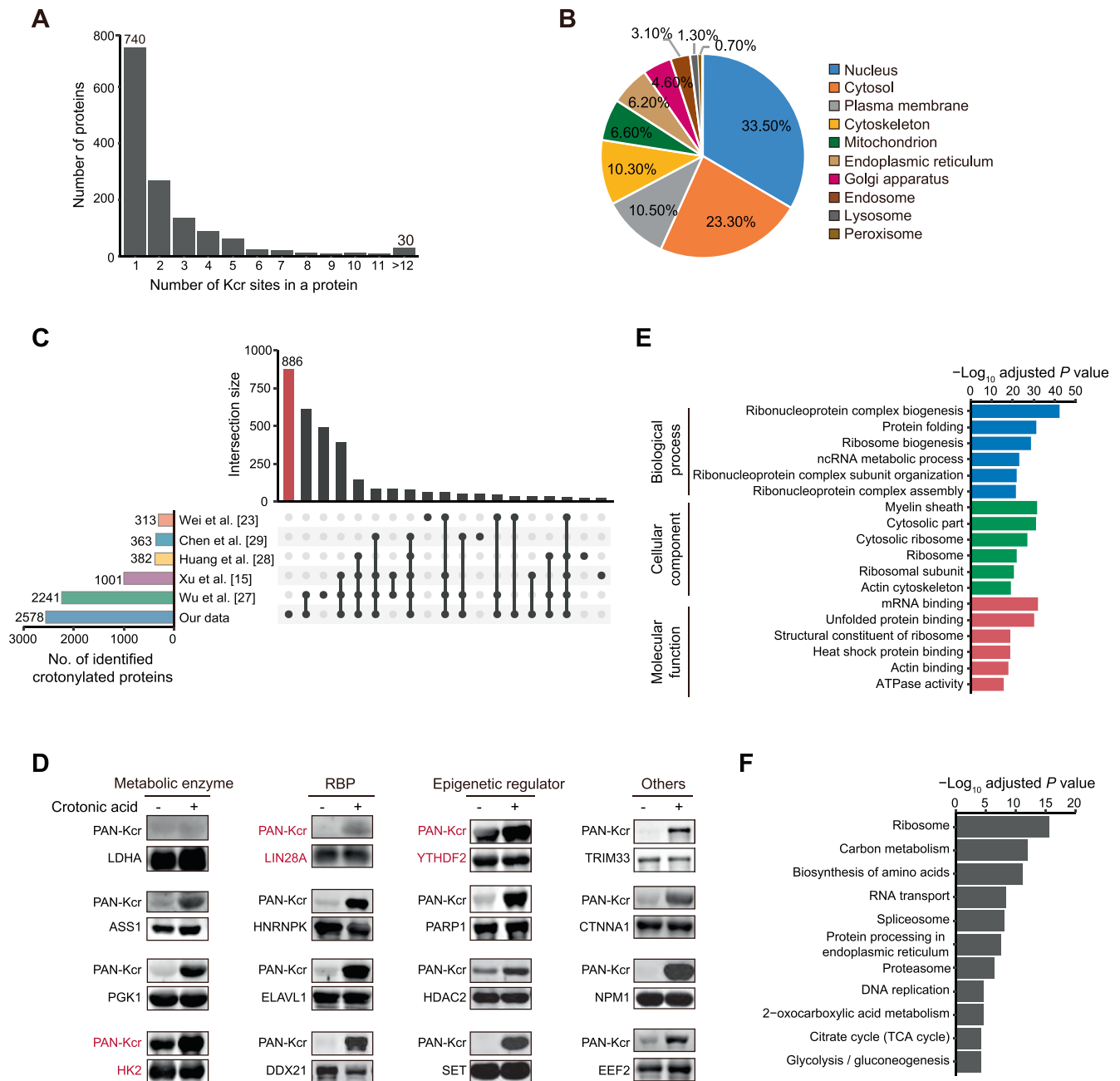


Figure 2 Crotonylation affects multiple proteins in different compartments and with different functions in three pluripotent states and differentiating cells

A. Bar plot indicating the number of Kcr sites in each high-confidence protein in our crotonylome dataset. **B.** Pie chart indicating the subcellular localization of the high-confidence crotonylated proteins. The localization was annotated using the COMPARTMENTS database in ‘knowledge’ channel. **C.** Comparison of our identified crotonylome data with crotonylome datasets from other published studies using human cells. **D.** Validation of protein crotonylation using ectopically expressed FLAG-tagged proteins in HEK293T cells with or without treatment of 10 mM crotonic acid for 24 h. Cell lysates were immunoprecipitated with anti-FLAG magnetic beads and analyzed by immunoblotting with anti-PAN Kcr antibodies. Three crotonylated proteins exclusive to our dataset are shown in red. **E.** GO analysis of the high-confidence crotonylated proteins. The top 6 terms with smallest adjusted P values are shown (Fisher’s exact test, Benjamini-Hochberg corrected $P < 0.01$). **F.** KEGG analysis of the high-confidence crotonylated proteins. The top 11 terms with smallest adjusted P values are shown (Fisher’s exact test, Benjamini-Hochberg corrected $P < 0.01$). GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; RBP, RNA-binding protein.

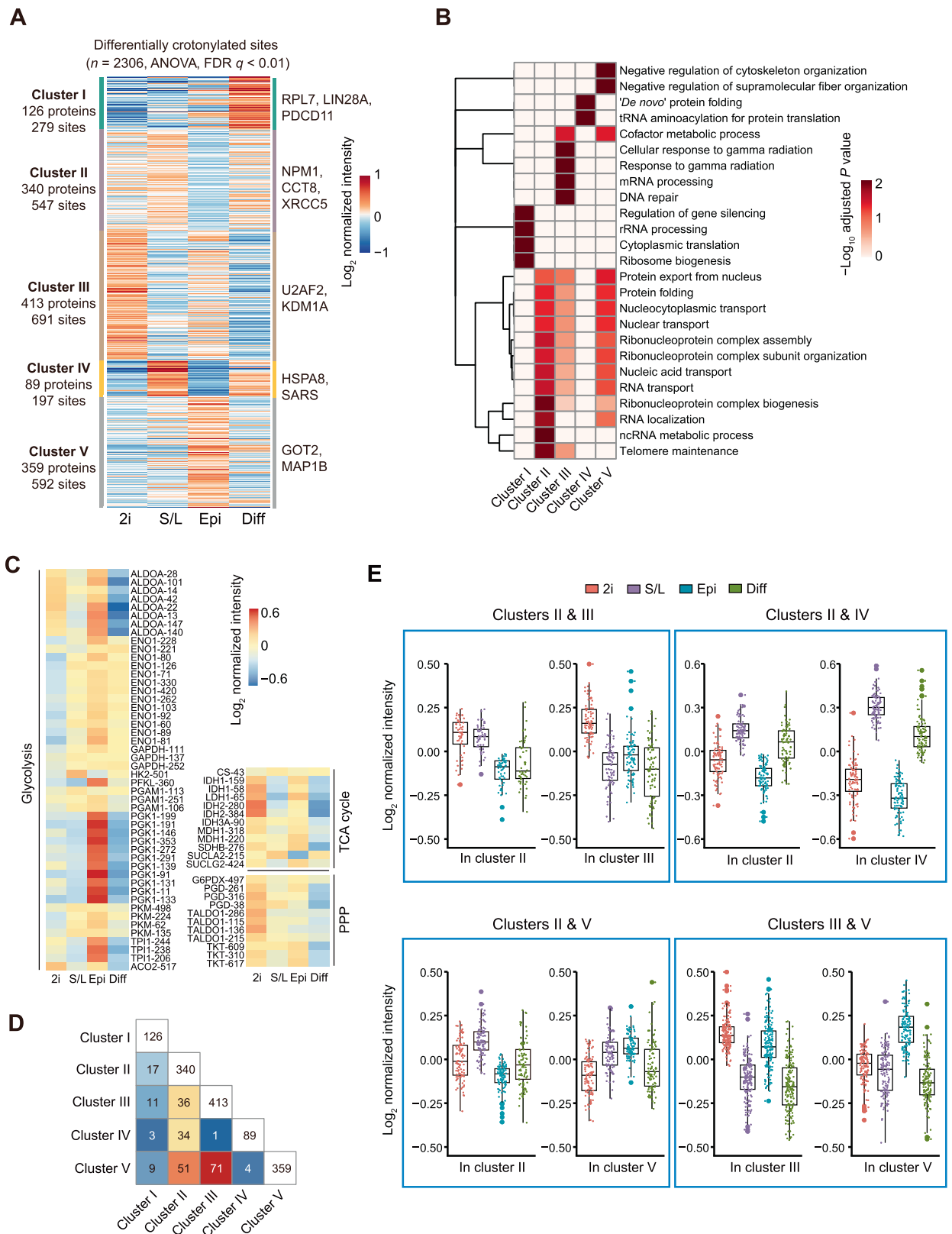


Figure 3 Characterization of crotonylation dynamics and their functional prediction in three pluripotent states and differentiating cells

A. Heatmap of crotonylated sites in high-confidence non-histone proteins shows distinct patterns in the four cell states ($n = 2306$, ANOVA test, FDR $q < 0.01$). Crotonylated proteins and sites were grouped into five clusters based on \log_2 normalized intensity values using hierarchical clustering. **B.** GO biological process analysis for all five clusters (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$, scaled without centering). **C.** Heatmap of differentially crotonylated enzymes involved in central carbon metabolism (including glycolysis, TCA cycle, and PPP). **D.** Number of crotonylated proteins overlapping between two clusters. **E.** Crotonylation levels of specific lysine sites in the overlapping proteins between different clusters. ANOVA, analysis of variance; FDR, false discovery rate; TCA cycle, tricarboxylic acid cycle; PPP, pentose phosphate pathway.

support the reliability of our proteomic experiments and analyses.

RNA-binding proteins are highly represented in the crotonylome of PSCs in different pluripotent states and differentiating cells

We then performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of the 1426 high-confidence crotonylated proteins. Interestingly, GO analysis showed enrichment in RNA metabolism-related and ribosome-related terms (Figure 2E; Table S2A). Consistent with the former functional enrichment analysis, classical RNA-binding domains (Figure S3A) and experimentally identified RNA-binding proteins (RBPs) from other studies [31–33] (Figure S3B; Table S2B) were enriched in our high-confidence crotonylome dataset. This is perhaps not surprising because lysine is one of the most commonly enriched amino acids in the RNA-binding domains of RBPs [34]. Thus, RBP crotonylation may be a widespread mechanism for regulating RNA–protein interactions by neutralizing the positive charge on lysine residues, in which case cell metabolism would appear as a major regulator of these interactions. This is particularly important for PSC function because many RBPs are essential regulators of pluripotency maintenance and exit [35]. For instance, K536 in the YTH domain of YTHDF2, which selectively recognizes the m⁶A modification on RNA [36], was identified as a high-confidence Kcr site. The crotonylation of this site may affect its binding affinity to m⁶A-modified RNA and in turn regulate cell fate, as this RNA modification is necessary for pluripotency exit [37]. As for the ribosome, it is known that post-translational modifications of ribosome subunits influence protein translation [38], also suggesting a potential link of this function with cell metabolism. Notably, translational control has been implicated in pluripotency maintenance and differentiation [39,40]. KEGG pathway analysis showed enrichment in terms related to carbon metabolism and proteasome (Figure 2F; Table S2A).

Because lysine residues are subjected to other post-translational modifications in addition to crotonylation, we also compared our crotonylome dataset with reported acetylation, malonylation, succinylation, and ubiquitination mouse datasets from the Protein Lysine Modifications Database (PLMD) [41]. Our high-confidence Kcr sites significantly overlapped with other modifications, but a large number of Kcr sites were unique (Figure S3C; Table S2C). Moreover, we observed that the average amino acid distribution around crotonylated lysine is enriched in glutamic acid (E), in agreement with a previous crotonylome study [15] (Figure S3D). The negative charge of glutamic acid residues might synergize with the crotonylation-mediated suppression of lysine's positive charge, changing the affinity of RBPs for substrates. Interestingly, it was reported that the glutamic acid-lysine (EK) rich region of the splicing factor SREK1/SRrp86 can function as an inhibitor of splicing [42]. Two lysine residues in this same domain of SREK1/SRrp86 were identified as low-confidence Kcr sites in our dataset (Figure S3E and F).

Overall, these results demonstrate that protein crotonylation is a widespread lysine modification in mouse PSCs and early differentiating cells, affecting proteins with relevant functions in pluripotency maintenance and differentiation.

Dynamics of protein crotonylation in PSCs in different pluripotent states and differentiating cells

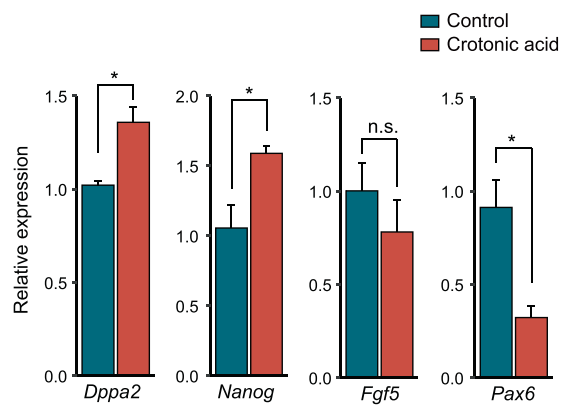
We next compared the individual crotonylomes of PSCs in three pluripotent states and differentiating cells to gain insight into the potential functional consequences of differential protein crotonylation levels. For this purpose, we applied variance analysis and hierarchical clustering to the respective high-confidence crotonylated peptide datasets. We initially focused on non-histone proteins, which led to the identification of five categories (clusters I to V) among the four cell states (Figure 3A and B; Table S3A).

Cluster I corresponds to proteins highly crotonylated in differentiating cells compared to the three pluripotent states. This category includes proteins related to protein translation such as RPL7, LIN28A, and PDCD11. Cluster II includes proteins displaying the lowest crotonylation level in EpiSCs and is particularly enriched in proteins related to RNA localization/transport (*e.g.*, NPM1), protein folding (*e.g.*, CCT8), and telomere maintenance (*e.g.*, XRCC5). Interestingly, crotonylation has been shown to facilitate telomere maintenance in chemically induced reprogramming [43], suggesting a potential cause-effect link. Cluster III corresponds to proteins that display the highest crotonylation level in 2i ESCs compared to the other three cell states, and includes proteins related to mRNA processing (*e.g.*, U2AF2) or DNA repair (*e.g.*, KDM1A). Cluster IV consists of proteins more highly crotonylated in S/L ESCs and differentiating cells compared to 2i ESCs and EpiSCs, and includes regulators of protein folding (*e.g.*, HSPA8) or tRNA aminoacylation (*e.g.*, SARS). The last category, cluster V, includes proteins more highly crotonylated in EpiSCs, and contains cofactors of metabolic processes (*e.g.*, GOT2) and cytoskeletal regulators (*e.g.*, MAP1B). We also observed a significant enrichment of ribonucleoprotein complex-related terms in clusters III and IV, albeit to a lesser extent than in cluster II.

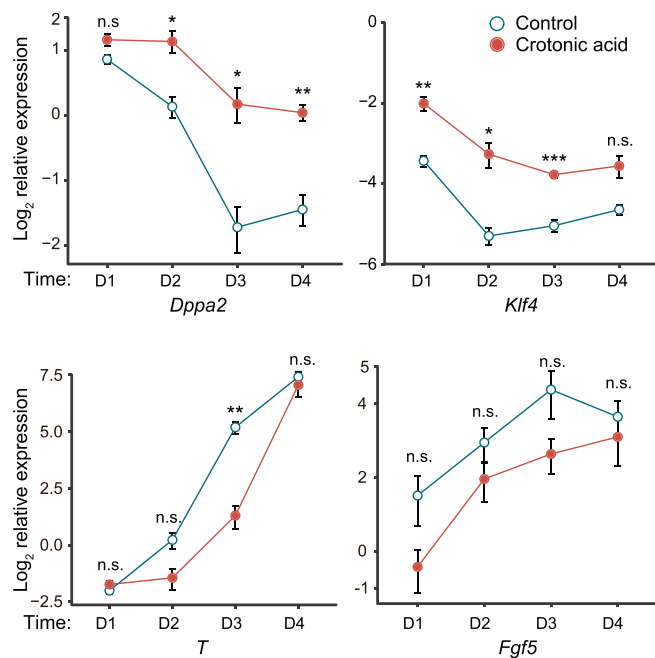
We looked in more detail into the enrichment of crotonylated peptides belonging to proteins involved in central carbon metabolism-related KEGG terms [including TCA cycle, pentose phosphate pathway (PPP), and glycolysis] in the four cell states. The crotonylation levels of these enzymes were highly dynamic among the four different cell states (Figure 3C). Notably, we noticed enhanced crotonylation of glycolytic enzymes in EpiSCs compared to naïve ESCs, whereas TCA cycle and PPP-related enzymes were more enriched in 2i ESCs compared to EpiSCs. These data are consistent with the increased glycolytic activity being a major route for energy production in EpiSCs and with the enhanced mitochondrial activity in naïve ESCs compared to EpiSCs [6,7]. This suggests that differential protein crotonylation in these cell states could actually be a contributing cause rather than a consequence of the observed metabolic differences. In this regard, it has been reported that acetylation of metabolic enzymes changes their functions as an adaptive mechanism to environmental changes [44,45].

We also observed that many crotonylated peptides are shared between clusters albeit with different intensity in each respective cluster (Figure 3D and E; Table S3B). For instance, the high-confidence K23 and K188 residues of the apoptosis regulator protein SET showed an opposed crotonylation ten-

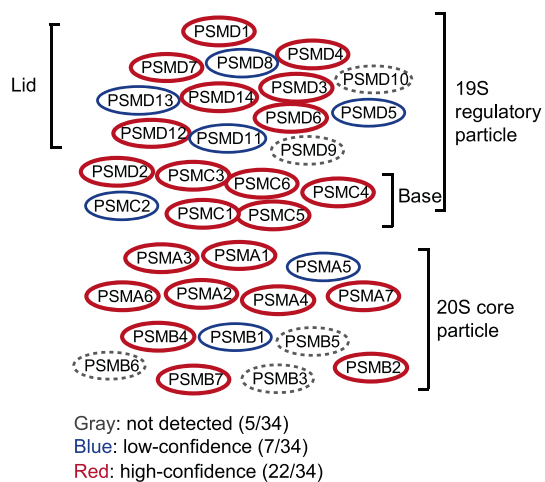
A



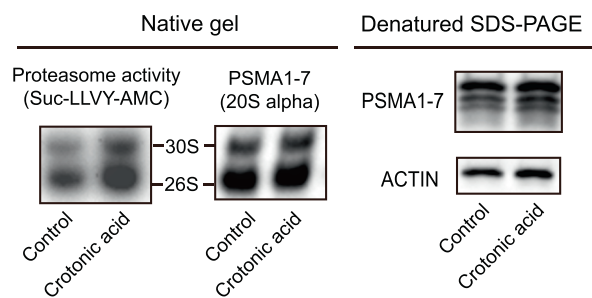
B



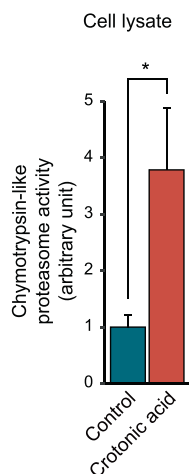
C



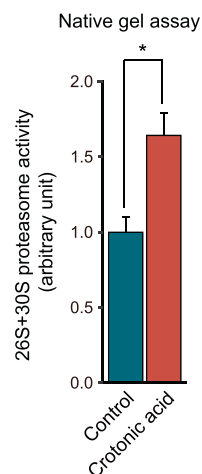
E



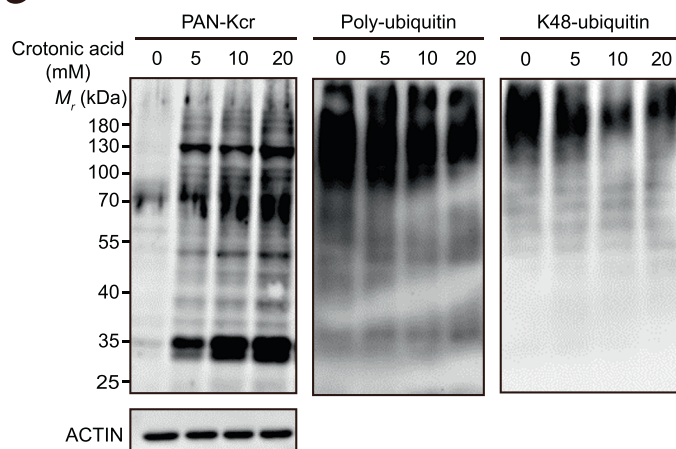
D



F



G



dependency between 2i ESCs and S/L ESCs, with K23 more crotonylated and K188 less in 2i ESCs (Figure S4A and B). This suggests 1) that crotonylation of different lysine residues of the same protein could have diverse functional consequences depending on the cell state, and 2) that differential crotonylation of the same lysine in different cell states may be involved in causing distinctive functional features.

Because protein levels vary among the four cell states, we also performed the differential crotonylation analysis after calibrating the crotonylation level by protein abundance. Proteins with stable expression but experiencing changes in their crotonylation statuses might be more critically involved in the determination of cell identity shifts between the four cell states. After normalization, we identified 485 differential sites in 330 proteins between the four cell states (Figure S4C and D; Table S3C). These sites and proteins distributed in five clusters. Cluster I includes proteins highly crotonylated in differentiating cells and is enriched in proteins related to the TCA cycle. Among these, we noticed for example ACLY, which participates in fatty acid synthesis to regulate pluripotency [46]. Cluster II consists of proteins highly crotonylated in 2i ESCs and is enriched in GO terms related to mRNA processing, ribonucleoprotein complex assembly, and chromatin remodelling. The mRNA processing term includes the classical RBP HNRNPU [47] and the deacetylase/decrotonylase HDAC1 [48], which have been reported to regulate pluripotency. Cluster III represents proteins highly crotonylated in EpiSCs and contains proteins involved in RNA splicing and transport (e.g., DHX9). Cluster IV represents proteins displaying lower crotonylation in EpiSCs and it is only enriched for proteins related to protein folding (e.g., HSPA9). Cluster V contains proteins more crotonylated in EpiSCs compared to the other cell states. This cluster has higher crotonylation in EpiSCs than Cluster III; however, there were no enriched GO terms from these 20 proteins.

Histone crotonylation is highly dynamic during ESC differentiation and is required for self-renewal [49]. Besides the analysis of non-histone proteins, we identified 21 Kcr sites in histone proteins (Figure S5; Table S3D). While extracting the histone for mass spectrometry might increase the number of identified histone Kcr sites, our dataset is nevertheless valuable for studying the role of histone crotonylation in pluripotency and differentiation.

In summary, our data support that dynamic changes in protein crotonylation, which extends beyond the better studied histone targets, likely play relevant roles in controlling the transition between different pluripotent states and pluripotency exit. Though our proteome data are of good reproducibility, bias may happen during sample preparation. In the future, it will be important to validate these changes using specific antibodies for each protein target and also different crotonylated residues that are not available at the moment. In this regard, our dataset constitutes a valuable resource for further studies.

Crotonic acid enhances pluripotency gene expression associated with increased proteasome activity

Given the dynamic changes in protein crotonylation among the four cell states, we next asked whether modulating protein crotonylation would affect PSC pluripotency or differentiation. To test this, we treated S/L ESCs with crotonic acid. We observed significant upregulation of the pluripotency genes *Nanog* and *Dppa2*, and a concomitant downregulation of the differentiation gene *Pax6* as well as a tendency of lower expression of the differentiation gene *Fgf5*, compared to the control (Figure 4A). S/L ESCs treated with crotonic acid that were allowed to gradually differentiate upon LIF withdrawal for 4 days also exhibited higher expression of pluripotency genes (*Dppa2* and *Klf4*) and a tendency of lower expression of differentiation genes (*T* and *Fgf5*) than non-treated S/L ESCs (Figure 4B), suggesting that enhanced protein crotonylation helps maintain pluripotency and slow pluripotency exit. It has been reported that crotonic acid treatment in mouse ESCs downregulates specific pluripotency genes (*Oct4*, *Sox2*, and *Nanog*) while upregulating two-cell (2C)-like state genes [43]. The 2C-like state refers to cells closer to early post-fertilization time points, where embryonic cells have totipotent characteristics (can produce trophectoderm and extraembryonic mesoderm rather than only the cell types coming from the three embryonic germ layers). 2C-like cells are known to express lower levels of pluripotency factors than S/L or 2i ESCs from which they spontaneously originate [50]. The discrepancy with our data could be due to the use of different cell lines (N33 versus OG2 ESCs) or unnoticed variations in the culture conditions. Considering the very low number



Figure 4 Crotonic acid promotes pluripotency and enhances proteasome activity

A. Relative expression of pluripotency and differentiation genes measured by RT-qPCR in ESCs cultured in S/L with or without 10 mM crotonic acid. Data are presented as mean \pm SEM ($n = 3$ biological replicates with 3 technical replicates each). **B.** Relative expression of representative pluripotency and differentiation genes measured by RT-qPCR in ESCs cultured in S/L with or without 10 mM crotonic acid and subjected to differentiation by withdrawing LIF over a time-course of 4 days. Data are presented as mean \pm SEM ($n = 3$ biological replicates with 3 technical replicates each). **C.** Schematic representation of the proteasome complex and the components identified in our crotonylome. **D.** Chymotrypsin-like proteasome activity measurement in ESCs cultured in S/L with or without 10 mM crotonic acid treatment for 48 h ($n = 4$ biological replicates). **E.** Chymotrypsin-like proteasome activity measurement in pre-purified proteasome by native gel electrophoresis. ESCs were cultured in S/L with or without 10 mM crotonic acid treatment for 48 h. Samples were resolved by denatured SDS-PAGE and Western blotting for analyzing the 20S proteasome level. ACTIN was used as a loading control. **F.** Quantified densitometry of the result shown in (E). Data are presented as mean \pm SEM ($n = 4$ biological replicates). **G.** Representative Western blotting of lysine crotonylation, poly-ubiquitination, and K48-ubiquitination in ESCs cultured in S/L and untreated or treated with different concentrations of crotonic acid for 48 h. ACTIN was the loading control. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n.s., not significant (two-tailed unpaired Student's *t*-test).

(~ 0.5%) of 2C-like cells in PSC cultures, single-cell gene expression studies would be necessary to fully clarify this. Regardless, the trend of the changes in both studies is in the same direction (increased potency).

Then, we sought to understand how crotonylation affects pluripotency. To this end, we first compared the crotonylated proteins in the four states with a reported ESC-specific gene set (genes highly expressed in mouse/human ESCs) [33]. The result showed that 466 ESC-specific gene-encoded proteins are crotonylated (Figure S6A). GO cellular component analysis of these proteins showed enrichment in terms related to 'proteasome complex' as well as 'peptidase complex', among others (Figure S6B; Table S4A). Indeed, most of the proteasome complex subunits were crotonylated (29/34 total and 22/34 high-confidence) in our dataset (Figure 4C; Table S4B). It has been shown that proteasome activity is positively associated with pluripotency maintenance [17,18] and our crotonylome dataset included the two proteasomal subunits, PSMD11 and PSMD14, reported as pluripotency regulators. Thus, we tested chymotrypsin-like proteasome activity in S/L ESCs treated with or without crotonic acid. Adding crotonic acid significantly enhanced proteasome activity compared to the control in the crude lysate (Figure 4D) and also in the native gel assay (Figure 4E and F). The latter experiment can separate active proteasome complexes from other biomacromolecules based on the molecular weight [51]. Moreover, Western blotting showed that the increase of protein Kcr signals caused by adding crotonic acid was associated with a decrease of poly-ubiquitin and K48-ubiquitin signals in cell lysates (Figure 4G). This further points to an increase in proteasome activity after crotonic acid treatment. To study whether the effect on ESCs is specific, we added crotonic acid to two different types of somatic (and hence differentiated) cells, HEK293T and NIH3T3 cells. There was no significant increase in proteasome activity measured using crude lysates (Figure S6C). This may be explained by PSC-specific expression of some proteasome subunits targeted by crotonylation (*e.g.*, PSMD11 and PSMD14) [17,18].

Taken together, these experiments suggest a link between crotonylation of proteasome subunits in PSCs and pluripotency maintenance. Additional work will be necessary to clarify whether this link is causal and how it relates to the crotonylation of other proteins involved in pluripotency control including histones [14].

Conclusion

Our global atlas of protein crotonylation in three pluripotent states and differentiating cells has identified targets essential for pluripotency regulation. Among these, it is worth noting factors involved in RNA biogenesis, protein translation, metabolic factors, and proteasome subunits. Further work will be important to validate the specific impacts of crotonylation on the functions of these proteins, and also how these effects crosstalk with the consequences on gene expression of histone crotonylation [49]. Hence, this dataset will be an important resource for future studies on pluripotency and early pluripotency exit. Our results may also be helpful to understand how changes in metabolism influence cell function in other contexts through variation in the levels of crotonyl-CoA.

Materials and methods

Isolation of ESCs and EpiSCs

OG2 ESCs were isolated in a previous study [52]. For the derivation of OG2 EpiSCs, the late epiblast layer was dissected intact from a pre-gastrulation stage mouse blastocyst (E5.75), maintained in FA medium [DMEM/F12 (Catalog No. SH30023.01B, Hyclone, Buckinghamshire, UK) and Neurobasal medium (Catalog No. 21103049, Gibco, Carlsbad, CA) mixed 1:1, supplemented with 1× non-essential amino acids (Catalog No. 11140050, Gibco), 1× GlutaMAX (Catalog No. 35050061, Gibco), 1 mM sodium pyruvate (Catalog No. 11360070, Gibco), 50 U/ml penicillin/streptomycin (Catalog No. SV30010, Hyclone), 1× N-2 (Catalog No. 17502048, Gibco), 1× B-27 (Catalog No. 17504044, Gibco), 12.5 ng/ml bFGF (Catalog No. 233-FB-050, R & D, Minneapolis, MN), and 20 ng/ml Activin A (Catalog No. 338-AC-050, R & D)] on Matrigel (Catalog No. 354248, Corning, Corning, NY)-coated plates and passaged 4–8 days after isolation.

Cell culture

OG2 ESCs were cultured in the presence of 1× non-essential amino acids, 1× GlutaMAX, 1 mM sodium pyruvate, 50 U/ml penicillin/streptomycin, and 1000 U/ml LIF (Catalog No. ESG1107, Millipore, Darmstadt, Germany), in either DMEM-high glucose medium (Catalog No. SH30022.01, Hyclone) supplemented with 15% fetal bovine serum (Catalog No. S1580, Biowest, Nuaille, France) on feeder (mitomycin-treated mouse embryonic fibroblasts)-coated plates (for the S/L condition) or DMEM/F12 and Neurobasal medium mixed 1:1, supplemented with 1× N-2, 1× B-27, 1 μM PD0325901, and 3 μM CHIR99021 on 0.1% gelatin (Catalog No. ES-006-B, Millipore)-coated plates (for the 2i condition). For differentiating cells, ESCs were cultured in the same medium as S/L ESCs but without LIF and on 0.1% gelatin-coated plates for 4 days. HEK293T and NIH3T3 cells were cultured in DMEM-high glucose supplemented with 10% fetal bovine serum (Catalog No. SFBE, NATOCOR, Villa Carlos Paz, Argentina).

Protein extraction and trypsin digestion for LC-MS/MS

Cells were lysed on ice in a buffer containing 8 M urea, 10 mM DTT (Catalog No. 10197777001, Sigma, Darmstadt, Germany), 50 mM nicotinamide, 3 μM Trichostatin A (Catalog No. T1952, Sigma), and 1× protease inhibitor cocktail (Catalog No. 11697498001, Roche, Basel, Switzerland). Lysates were centrifuged (20,000 *g*) at 4 °C for 10 min. Supernatants were precipitated with cold 15% trichloroacetic acid at –20 °C for 2 h, following 20,000 *g* centrifugation at 4 °C for 10 min. The precipitated proteins were dissolved in a pH 8.0 buffer containing 8 M urea and 100 mM triethylammonium bicarbonate. The protein solution was reduced with 10 mM DTT at 37 °C for 1 h followed by alkylation with 20 mM iodoacetamide at room temperature for 45 min protected from light. The alkylated protein samples were diluted by adding 100 mM triethylammonium bicarbonate buffer. Trypsin (Catalog No. V5111, Promega, Madison, WI) was added at 1:50

(w/w) trypsin to protein ratio overnight, and then at 1:100 ratio for 4 h for digestion.

TMT labeling

The digested proteins were desalted by running on a Strata X C18 SPE column (Catalog No. 8B-S100-AAK, Phenomenex, Torrance, CA) and then vacuum-dried. Desalted peptides were labeled with TMTsixplex™ Isobaric Label Reagent Set (Catalog No. 90061, ThermoFisher Scientific, Waltham, MA) following the manufacturer's protocol. Briefly, TMT was added at 1:1 (U/mg) TMT reagent to protein ratio for 2 h at room temperature and then the samples were desalted.

Peptide fractionation

TMT-labeled peptides were fractionated by high pH reverse-phase HPLC using an ZORBAX Extend-C18 column (Catalog No. 770450-902, Santa Clara, CA) (5 μm particles, 4.6 mm inner diameter, 250 mm length). Briefly, labeled peptides were separated with a gradient of 2%–60% acetonitrile in 10 mM ammonium bicarbonate into 80 fractions for 80 min. Fractionated peptides were combined into 18 fractions for total proteome analysis or 8 fractions for crotonylome analysis.

Kcr enrichment

Kcr-containing peptides were dissolved in pH 8.0 NETN buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl, and 0.5% NP-40). Dissolved peptides were incubated with anti-Kcr antibody-coated agarose beads (Catalog No. PTM-503, PTM Biolabs, Hangzhou, China) overnight at 4 °C. Beads were then washed with NETN buffer, and bound peptides were eluted with 0.1% trifluoroacetic acid. The resulting peptides were desalted with C18 ZipTips (Catalog No. ZTC18S008, Millipore) before LC–MS/MS analysis.

LC–MS/MS analysis

Enriched peptides were dissolved in 0.1% formic acid and loaded onto a home-made reverse-phase analytical column (15 cm length, 75 μm inner diameter). Peptide samples were eluted with a gradient elution system: an increasing gradient of solvent A (0.1% formic acid in 98% acetonitrile) from 7% to 20% over 24 min, 20% to 35% in 8 min, climbing to 80% in 5 min, and then holding at 80% for the last 3 min. And all at a constant flow rate of 300 nl/min on an EASY-nLC 1000 UPLC system (Catalog No. LC120, ThermoFisher Scientific). The eluted peptide samples were analyzed by a Q Exactive Plus™ hybrid quadrupole-Orbitrap™ mass spectrometer (Catalog No. IQLAAEGAAPFALGMAZR, ThermoFisher Scientific). The electrospray voltage was set to 2.0 kV. The *m/z* scan range was fixed from 350 to 1800 for full scan, and intact peptides were detected at a resolution of 70,000. Peptides were then selected for MS/MS using NCE setting as 30 and the fragments were detected at a resolution of 17,500. Data-dependent acquisition was used for MS data collection.

LC–MS/MS data analysis

The resulting TMT data were processed using MaxQuant (v.1.4.1.2) with integrated Andromeda search engine [53]. Tandem mass spectra data were searched against non-redundant mouse protein amino acid sequence from Uniprot databases (<https://www.uniprot.org/>) concatenated with reverse decoy database. Trypsin/P was selected as cleavage enzyme permitting up to two missing cleavages per peptide for total proteome analysis or four missing cleavages per peptide for crotonylome analysis. Mass error was set to 10 ppm for precursor ions and 0.02 Da for fragmented ions. Carbamidomethylation on cysteine was specified as fixed modification; oxidation on methionine, crotonylation on lysine, and acetylation on the N-terminus of the protein were specified as variable modifications. False discovery rate (FDR) thresholds for proteins, peptides, and modification sites were set at less than 1%. Minimum peptide length was set at 7. For peptide quantification, TMT6plex was selected. All other parameters in MaxQuant used the default setting.

Statistical analysis

All statistical analyses were performed using R (3.5.1). For protein homology analysis, protein sequences in identified human crotonylated protein datasets [15,23,27–29] were inputted into 'blastp' function in blast+ (2.8.1) with the mouse crotonylated proteins we discovered. Resulting homologous proteins with E value < 1E–10 were considered as overlapping proteins in the human crotonylated protein datasets. The 'upset' function in R package "UpSetR" (1.4.0) was used to visualize the number of identified crotonylated proteins and intersections of this study with the various human crotonylated protein datasets. Pairwise experimental Pearson correlation matrix was generated using 'cor' function in R with the option "use = all.obs". Correlation matrix was clustered using hierarchical clustering by Euclidean distance and complete linkage method. Cell state variation was defined by performing PCA with 'prcomp' function in R. For analysis of the differentially modified sites, we used ANOVA test to generate *P* values for each high-confidence peptide. The generated *P* values were corrected by multiple hypotheses with FDR (Benjamini-Hochberg). Pearson correlation results and expression data were visualized using the 'pheatmap' function within the R package "pheatmap" (1.0.12).

Cellular compartment analysis

Subcellular localization of proteins was analyzed using the COMPARTMENTS database [26] in 'knowledge' channel. Only the major compartments in eukaryotic cells (nucleus, cytosol, plasma membrane, cytoskeleton, mitochondrion, endoplasmic reticulum, Golgi apparatus, endosome, lysosome, and peroxisome), with possibility score > 1, were selected.

GO and KEGG annotations

GO and KEGG annotations were performed using the R package "clusterProfiler" (3.10.1) [54] with the 'enrichGO' and 'enrichKEGG' function, respectively.

Co-modification analysis

For analysis of crosstalk between Kcr sites and other lysine modifications, we used the PLMD database [41], which lists 20 types of lysine modifications in multiple species. Co-modified sites were generated by merging our dataset with mouse PLMD dataset.

Protein domain analysis

Protein domains of crotonylated proteins were annotated using InterProScan (<http://www.ebi.ac.uk/interpro/>) based on protein sequence alignment method with default settings. Fisher's exact test and standard FDR control method were used to test the enrichment of the annotated domains from crotonylated proteins against proteome-wide domains.

Crotonylation motif analysis

The frequency of amino acids surrounding crotonylated lysine residues was generated using iceLogo [55]. Briefly, high-confidence Kcr-central peptide sequences (6 amino acids upstream and downstream of the modified site) were used as input against the precompiled *Mus musculus* protein sequences from Swiss-Prot, and the percent difference was set as scoring system with a significance cut-off of $P = 0.05$.

Validation of crotonylated proteins by co-immunoprecipitation

HEK293T cells were transfected with plasmids producing FLAG-tagged proteins. After transfection, cells were lysed with TNE lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% NP-40, 1 mM EDTA, 10 mM sodium butyrate, and 1× protease inhibitor cocktail). Lysates were sonicated with a Bioruptor sonicator (Catalog No. B01020001 Diagenode, Ougrée, Belgium) with low power (30 s ON/OFF, 5 cycles), followed by 10,000 g centrifugation at 4 °C for 10 min to remove the undissolved particles. Then, they were incubated with pre-washed anti-FLAG M2 magnetic beads (Catalog No. M8823, Sigma) at 4 °C overnight with rotation, and eluted by heating at 95 °C for 5 min with SDS-PAGE loading buffer. The following primary antibodies were used for immunoblotting: anti-FLAG (Catalog No. F7425, Sigma) and anti-Kcr (Catalog No. PTM-501, PTM Biolabs).

Western blotting

Western blotting was performed using standard procedure after lysing cells with a buffer containing 10 mM Tris-HCl pH 7.4, 10 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS, 10 mM N-ethylmaleimide, and 1× protease inhibitor cocktail. Lysates were boiled for 10 min to ensure inactivation of deubiquitinase enzymes. Denatured lysates were sonicated with a Bioruptor sonicator (low power, 30 s ON/OFF, 5 cycles), followed by 10,000 g centrifugation at 4 °C for 10 min to remove undissolved particles. Lysates were then subjected to immunoblotting with the following primary antibodies: anti-polyubiquitin (Catalog No. 14220, Cayman, Ann Arbor, MI), anti-K48-linkage specific polyubiquitin (Catalog No. 8081S, CST, Danvers, MA), anti-NANOG (Catalog

No. A300-397A, Bethyl, Montgomery, MA), and anti- β -ACTIN (Catalog No. A2228, Sigma).

Proteasome activity assay

Chymotrypsin-like proteasome activity was measured using a Proteasome Activity Assay Kit (Catalog No. ab107921, Abcam, Cambridge, UK) following the manufacturer's protocol. In brief, cells with or without crotonic acid treatment were collected in 0.5% NP-40 and homogenized by pipetting up and down a few times. After 10,000 g centrifugation at 4 °C for 10 min, the supernatants were incubated with the proteasome substrate Suc-LLVY-AMC (Catalog No. HY-P1002, MedChemExpress, Monmouth Junction, NJ) at 37 °C for 20 min. The released free 7-amino-3-methylcoumarin (AMC) fluorescence was detected on a microplate fluorometer (350 nm excitation, 440 nm emission) after 30 min and then 60 min at 37 °C. Background signal was corrected by subtracting the 30-min reading from the 60-min reading, and data were normalized against the control sample.

Native gel assay for proteasome activity

Native gel assay for proteasome activity was performed according to a previous publication [17]. Briefly, the cells with or without crotonic acid treatment were collected in proteasome activity assay buffer (50 mM Tris-HCl pH 7.5, 5 mM MgCl₂, 5 mM ATP, 1 mM DTT, and 10% glycerol) and lysed by passing 10 times through the syringe with 27-G needle. After 12,000 g centrifugation at 4 °C for 10 min, the supernatants were loaded on 4%–12% NativePAGE Bis-Tris gel (Catalog No. BN1003BOX, Invitrogen, Carlsbad, CA) in NativePAGE running buffer (Catalog No. BN2001, Invitrogen) containing 5 mM MgCl₂ and 1 mM ATP, and run at 4 °C for 150 min at 150 V. The native gel was then incubated with 300 μ M Suc-LLVY-AMC diluted in NativePAGE running buffer. Chymotrypsin-like proteasome activity was detected on a FluorChem E system (Catalog No. 92-14860-00, ProteinSimple, San Jose, CA) in UV channel with 460 nm filter. The gel was then incubated with transfer buffer with 1% SDS for 10 min followed by incubation with transfer buffer for 10 min. The denatured gel was then transferred to PVDF membrane and incubated with anti-Proteasome 20S alpha antibody (Catalog No. ab22674, Abcam).

Ethical statement

Animal experiments in this study were compliant with all relevant ethical regulations for animal research and conducted under the approval of the Animal Care and Use Committee of the Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, under licence number 2007007.

Data availability

Mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (ProteomeXchange: PXD017121) [56], which are publicly accessible at <http://proteomecentral.proteomexchange.org>. MS/MS spectrum data of all identified crotonylated peptides

have been deposited to the ProteinProspector database, which can be accessed in <http://msviewer.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer> by the following search keys: zxlfv5pcbb (Replicate 1), g7bxkypz5c (Replicate 2), and rdqve-cjff (Replicate 3).

CRedit author statement

Yuan Lv: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization. **Chen Bu:** Methodology, Resources. **Jin Meng:** Investigation, Validation. **Carl Ward:** Methodology. **Giacomo Volpe:** . **Jieyi Hu:** Investigation. **Mengling Jiang:** Investigation. **Lin Guo:** Resources. **Jiekai Chen:** Resources. **Miguel A. Esteban:** Conceptualization, Resources, Writing - original draft, Supervision, Project administration. **Xichen Bao:** Conceptualization, Writing - original draft, Supervision, Project administration. **Zhongyi Cheng:** Conceptualization, Resources, Supervision, Project administration. All authors read and approved the final manuscript.

Competing interests

ZC is co-founder and chief executive officer of PTM Bio Inc., and CB is an employee. Other authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant Nos. 2018YFA0106903, 2016YFA0100102, and 2016YFA010070), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA16030502), the Natural Science Foundation of Guangdong Province, China (Grant No. 2018B030306042), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (Grant No. 2015294), the Innovation Team Project grant from the Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), China (Grant No. 2018GZR110103001), and the Science and Technology Planning Project of Guangdong Province, China (Grant No. 2020B1212060052). Carl Ward was supported by a Zhujiang Talent-Overseas Postdoctoral Funding Grant, China and a President's International Fellowship Initiative grant from the Chinese Academy of Sciences. We thank all members of the Esteban laboratory for their support. We also thank Xiaofei Zhang (Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences) for constructive advice on proteomic data analysis.

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2021.01.004>.

ORCID

0000-0003-1260-5800 (Yuan Lv)
0000-0002-1377-6858 (Chen Bu)
0000-0003-1857-2843 (Jin Meng)

0000-0003-0889-9025 (Carl Ward)
0000-0001-5000-6951 (Giacomo Volpe)
0000-0001-6905-3694 (Jieyi Hu)
0000-0001-7484-6891 (Mengling Jiang)
0000-0001-5698-270X (Lin Guo)
0000-0001-5168-7074 (Jiekai Chen)
0000-0002-1426-6809 (Miguel A. Esteban)
0000-0003-0389-4233 (Xichen Bao)
0000-0002-8764-2019 (Zhongyi Cheng)

References

- [1] De Los Angeles A, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, et al. Hallmarks of pluripotency. *Nature* 2015;525:469–78.
- [2] Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: naïve to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* 2016;17:155–69.
- [3] Nichols J, Smith A. Naïve and primed pluripotent states. *Cell Stem Cell* 2009;4:487–92.
- [4] Hanna J, Markoulaki S, Mitalipova M, Cheng AW, Cassady JP, Staerk J, et al. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* 2009;4:513–24.
- [5] Marks H, Kalkan T, Menafrá R, Denisov S, Jones K, Hofemeister H, et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 2012;149:590–604.
- [6] Sperber H, Mathieu J, Wang Y, Ferreccio A, Hesson J, Xu Z, et al. The metabolome regulates the epigenetic landscape during naïve-to-primed human embryonic stem cell transition. *Nat Cell Biol* 2015;17:1523–35.
- [7] Zhou W, Choi M, Margineantu D, Margaretha L, Hesson J, Cavanaugh C, et al. HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J* 2012;31:2103–6.
- [8] Moussaieff A, Rouleau M, Kitsberg D, Cohen M, Levy G, Barasch D, et al. Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab* 2015;21:392–402.
- [9] Shyh-Chang N, Locasale JW, Lyssiotis CA, Zheng Y, Teo RY, Ratanasirintrao S, et al. Influence of threonine metabolism on S-adenosylmethionine and histone methylation. *Science* 2013;339:222–6.
- [10] Choudhary C, Weinert BT, Nishida Y, Verdin E, Mann M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* 2014;15:536–50.
- [11] Sabari BR, Zhang Di, Allis CD, Zhao Y. Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol* 2017;18:90–101.
- [12] Xie Z, Zhang Di, Chung D, Tang Z, Huang He, Dai L, et al. Metabolic regulation of gene expression by histone lysine β -hydroxybutyrylation. *Mol Cell* 2016;62:194–206.
- [13] Tang H, Han M. Fatty acids regulate germline sex determination through ACS-4-dependent myristoylation. *Cell* 2017;169:457–69. e13.
- [14] Tan M, Luo H, Lee S, Jin F, Yang J, Montellier E, et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 2011;146:1016–28.
- [15] Xu W, Wan J, Zhan J, Li X, He H, Shi Z, et al. Global profiling of crotonylation on non-histone proteins. *Cell Res* 2017;27:946–9.
- [16] Yu H, Bu C, Liu Y, Gong T, Liu X, Liu S, et al. Global crotonylome reveals CDYL-regulated RPA1 crotonylation in homologous recombination-mediated DNA repair. *Sci Adv* 2020;6:eaay4697.
- [17] Vilchez D, Boyer L, Morantte I, Lutz M, Merkwirth C, Joyce D, et al. Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. *Nature* 2012;489:304–8.

- [18] Buckley S, Aranda-Orgilles B, Strikoudis A, Apostolou E, Loizou E, Moran-Crusio K, et al. Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell* 2012;11:783–98.
- [19] Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, et al. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* 2010;6:71–9.
- [20] Zhuang Q, Li W, Benda C, Huang Z, Ahmed T, Liu P, et al. NCoR/SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. *Nat Cell Biol* 2018;20:400–12.
- [21] Tsanov KM, Pearson DS, Wu Z, Han A, Triboulet R, Seligson MT, et al. LIN28 phosphorylation by MAPK/ERK couples signalling to the post-transcriptional control of pluripotency. *Nat Cell Biol* 2017;19:60–7.
- [22] Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* 2010;143:313–24.
- [23] Wei W, Mao A, Tang B, Zeng Q, Gao S, Liu X, et al. Large-scale identification of protein crotonylation reveals its role in multiple cellular functions. *J Proteome Res* 2017;16:1743–52.
- [24] Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* 2008;2:333–44.
- [25] Wan J, Liu H, Chu J, Zhang H. Functions and mechanisms of lysine crotonylation. *J Cell Mol Med* 2019;23:7163–9.
- [26] Binder JX, Pletscher-Frankild S, Tsaou K, Stolte C, O'Donoghue SI, Schneider R, et al. Compartments: unification and visualization of protein subcellular localization evidence. *Database* 2014;2014:bau012.
- [27] Wu Q, Li W, Wang C, Fan P, Cao L, Wu Z, et al. Ultradeep lysine crotonylome reveals the crotonylation enhancement on both histones and nonhistone proteins by SAHA treatment. *J Proteome Res* 2017;16:3664–71.
- [28] Huang H, Wang DL, Zhao Y. Quantitative crotonylome analysis expands the roles of p300 in the regulation of lysine crotonylation pathway. *Proteomics* 2018;18:e1700230.
- [29] Chen W, Tang D, Xu Y, Zou Y, Sui W, Dai Y, et al. Comprehensive analysis of lysine crotonylation in proteome of maintenance hemodialysis patients. *Medicine* 2018;97:e12035.
- [30] Du H, Zhao Y, He J, Zhang Y, Xi H, Liu M, et al. YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat Commun* 2016;7:12626.
- [31] Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, et al. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* 2013;20:1122–30.
- [32] He C, Sidoli S, Warneford-Thomson R, Tatomer D, Wilusz J, Garcia B, et al. High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol Cell* 2016;64:416–30.
- [33] Bao X, Guo X, Yin M, Tariq M, Lai Y, Kanwal S, et al. Capturing the interactome of newly transcribed RNA. *Nat Methods* 2018;15:213–20.
- [34] Castello A, Fischer B, Frese C, Horos R, Alleaume A-M, Foehr S, et al. Comprehensive identification of RNA-binding domains in human cells. *Mol Cell* 2016;63:696–710.
- [35] Ye J, Billeloch R. Regulation of pluripotency by RNA binding proteins. *Cell Stem Cell* 2014;15:271–80.
- [36] Zhu T, Roundtree IA, Wang P, Wang X, Wang Li, Sun C, et al. Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res* 2014;24:1493–6.
- [37] Geula S, Moshitch-Moshkovitz S, Domimissini D, Mansour AA, Kol N, Salmon-Divon M, et al. m⁶A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* 2015;347:1002–6.
- [38] Simsek D, Tiu GC, Flynn RA, Byeon GW, Leppeck K, Xu AF, et al. The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. *Cell* 2017;169:1051–65.e18.
- [39] Bulut-Karslioglu A, Biechele S, Jin H, Macrae TA, Hejna M, Gertsenstein M, et al. Inhibition of mTOR induces a paused pluripotent state. *Nature* 2016;540:119–23.
- [40] Sugiyama H, Takahashi K, Yamamoto T, Iwasaki M, Narita M, Nakamura M, et al. NatI promotes translation of specific proteins that induce differentiation of mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 2017;114:340–5.
- [41] Xu H, Zhou J, Lin S, Deng W, Zhang Y, Xue Y. PLMD: an updated data resource of protein lysine modifications. *J Genet Genomics* 2017;44:243–50.
- [42] Li J, Barnard DC, Patton JG. A unique glutamic acid-lysine (EK) domain acts as a splicing inhibitor. *J Biol Chem* 2002;277:39485–92.
- [43] Fu H, Tian CL, Ye X, Sheng X, Wang H, Liu Y, et al. Dynamics of telomere rejuvenation during chemical induction to pluripotent stem cells. *Stem Cell Rep* 2018;11:70–87.
- [44] Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 2009;325:834–40.
- [45] Wang Q, Zhang Y, Yang C, Xiong H, Lin Y, Yao J, et al. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* 2010;327:1004–7.
- [46] Wang L, Zhang T, Wang L, Cai Y, Zhong X, He X, et al. Fatty acid synthesis is critical for stem cell pluripotency via promoting mitochondrial fission. *EMBO J* 2017;36:1330–47.
- [47] Vizlin-Hodzic D, Johansson H, Ryme J, Simonsson T, Simonsson S. SAF-A has a role in transcriptional regulation of *Oct4* in ES cells through promoter binding. *Cell Reprogram* 2011;13:13–27.
- [48] Kidder BL, Palmer S. HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. *Nucleic Acids Res* 2012;40:2925–39.
- [49] Wei W, Liu X, Chen J, Gao S, Lu L, Zhang H, et al. Class I histone deacetylases are major histone deacetylases: evidence for critical and broad function of histone crotonylation in transcription. *Cell Res* 2017;27:898–915.
- [50] Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 2012;487:57–63.
- [51] Myeku N, Clelland CL, Emrani S, Kukushkin NV, Yu WH, Goldberg AL, et al. Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented early in disease by activating cAMP-PKA signaling. *Nat Med* 2016;22:46–53.
- [52] Liu J, Han Q, Peng T, Peng M, Wei B, Li D, et al. The oncogene *c-Jun* impedes somatic cell reprogramming. *Nat Cell Biol* 2015;17:856–67.
- [53] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008;26:1367–72.
- [54] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- [55] Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 2009;6:786–7.
- [56] Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 2019;47:D442–50.