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 Guo[2], 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] *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] *Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou 510005, China*

[7] *Institute for 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*

[#] Co-first authors

[*] Corresponding authors

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

Running title:

Yuan L et al / Lysine Crotonylome in Pluripotent Cells

ORCID

0000-0003-1260-5800 (Lv Y)

0000-0002-1377-6858 (Bu C)

0000-0003-1857-2843 (Meng J)

0000-0003-0889-9025 (Ward C)

0000-0001-5000-6951 (Volpe G)

0000-0001-5698-270X (Guo L)

0000-0001-6905-3694 (Hu J)

0000-0001-7484-6891 (Jiang M)

0000-0001-5168-7074 (Chen J)

0000-0002-1426-6809 (Esteban MA)

0000-0003-0389-4233 (Bao X)

0000-0002-8764-2019 (Cheng Z)

Total counts of **words**: 9331

Total counts of **figures**: 4

Total counts of **tables**: 1

Total counts of **supplementary figures**: 6

Total counts of **supplementary tables**: 4

Total counts of **references**: 56

Total counts of **references since 2014**: 35

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 state, as well as metastable PSCs undergoing early pluripotency exit. We successfully identified 3628 high-confidence sites in 1426 proteins. These crotonylated proteins are enriched for factors involved in functions related to pluripotency such as RNA biogenesis, central carbon metabolism and proteasome function. Moreover, we found that increasing the cellular levels of crotonyl-CoA through crotonic acid treatment promotes proteasome activity in metastable PSCs and delays their differentiation, consistent with previous observations showing that enhanced proteasomal 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.

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

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 ICM cells in specific conditions, which then produces PSCs. Accordingly, PSCs are self-renewing, can be maintained indefinitely *in vitro* and differentiated on demand upon exposure to specific signalling cues. Studying the regulation of pluripotency maintenance and exit in PSCs is important not only to understand embryonic development but also for cell fate transitions in other contexts (*e.g.*, somatic cell reprogramming and transdifferentiation, cancer, or aging). Mouse PSCs are a well-studied pluripotency model and can be cultured in different conditions: ground state (in serum-free medium with MEK and GSK3 inhibitors, or 2i embryonic stem cells [ESCs]), metastable state (with serum and leukemia inhibitory factor [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 downwards progression from the early preimplantation blastocyst to the post-implantation embryo. PSCs in 2i more closely resemble the pre-implantation ICM and produce chimeras upon blastocyst complementation more efficiently than metastable PSCs. On the contrary, EpiSCs resemble the post-implantation ICM 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 function is controlled by different signalling pathways [5, 6]. Regarding metabolism, naive PSCs rely on both glycolysis and oxidative phosphorylation (through the tricarboxylic, TCA, cycle) for producing energy, whereas EpiSCs mostly rely on glycolysis [6, 7].

Besides providing energy, cell metabolism regulates myriad cellular functions comprehensively including modifications in 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, among others) have been shown to be substrate for lysine acylation of not only histones but also non-histone proteins involved in many cellular processes [10, 11]. These post-translational modifications 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 3 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 be crotonylated too [14, 15]. For example, crotonylation of RPA1 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 report, we have constructed an atlas of Kcr, with focus on non-histone proteins, in different mouse pluripotent states (2i ESCs, S/L ESCs, and EpiSCs) and also S/L ESCs allowed to exit pluripotency by removing LIF. We have identified 3628 high-confident Kcr sites on 1426 proteins among these four cell states. These sites reveal links between crotonylation and cell functions relevant for pluripotency maintenance, including RNA biogenesis, carbon metabolism, and the proteasome. Consistently, we observed that increased protein crotonylation caused by crotonic acid administration up-regulates pluripotency gene expression, delays differentiation, and enhances proteasome activity, a function relevant for 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 differentiated cells.

Results and discussion

Quantitative lysine crotonylome analysis in different pluripotent states and differentiated cells

We aimed to investigate protein crotonylation in mouse PSCs in different states: 2i ESCs, S/L ESCs, EpiSCs, and also S/L ESCs triggered to exit pluripotency by LIF withdrawal for 4 days (hereafter referred to as differentiating cells) (**Figure 1**A). 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 showed bright GFP fluorescence as well as the characteristic domed colony shape, whereas EpiSCs and differentiated cells had lost the GFP signal and displayed flat and flat/irregular colony shape, respectively (Figure S1A). To further verify the different cell identities, we performed reverse transcription-quantitative PCR (RT-qPCR) for general pluripotency markers (*Nanog* and *Oct4/Pou5f1*), pluripotent state-specific regulators (*Lin28a* and *Myc*) [21,

22] and EpiSC/differentiation markers (*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 above-mentioned cell states.

Next, we quantified the lysine crotonylome of cells in these four culture conditions. To this end, we labelled the digested peptides from different conditions with tandem mass tags (TMT). We combined and fractionated the labelled peptides with high-performance liquid chromatography, and then enriched them with anti-Kcr antibodies [14] before conducting mass spectrometry analysis. Altogether, we identified 8,102 different sites in 2578 proteins among the four cell types, which were divided into high- (3628 sites in 1426 proteins) or low-confidence (4474 sites in 1152 proteins) candidates (Table 1 and S1A). High-confidence proteins were defined as quantified in at least two experiments out of three, 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 S1C and 1C).

We then integrated our total proteome with the publicly available ‘core ESC-like gene module’ and the ‘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 the lowest in differentiating cells, whereas the ‘adult tissue stem module’ was the lowest in ground state PSCs but the highest in the 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 and S1G). The

largest Euclidean distances in the crotonylome PCA were between 2i ESCs and the differentiating cells, consistent with the functional divergence between them. These observations 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 known crotonylation writers and erasers (also crotonylation readers) do not change noticeably between the four cell states (Figure S1E). Likewise, global crotonylation was at an equivalent level too (Figure S1F).

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

Characteristics of the crotonylated proteins in the four cell states

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 crotonylated sites, whereas 30 showed more than 12 modification sites (**Figure 2**A and 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, cytoplasm, plasma membrane, cytoskeleton, and mitochondria (Figure 2B and Table S1C). We also compared our complete Kcr dataset with published crotonylomes in different human cells including H1299 [15] , HeLa [23], A549 [27], and HCT116 cells [28], and also peripheral blood from patients with kidney failure [29], all of which described identified crotonylated proteins without selection of high-confidence ones. Our dataset was the largest, with 886 out of the total 2578 crotonylated proteins (high- and low-confidence candidates) being unique to it (Figure 2C and Table S1D). To validate

the identified candidates, we selected and overexpressed 15 high-confidence proteins and one low-confidence protein with a FLAG tag in HEK293T cells, followed by immunoprecipitation and western blotting with Kcr antibodies. The substrate for crotonyl-CoA generation, crotonic acid, was added to enhance the detection. Importantly, 11 out of 16 candidates showed increased basal Kcr signal, crotonic acid treatment boosted the basal signal in most of the selected candidates (15 out of 16) (Figure 2D and S2). Three out of the 15 validated proteins, including the pluripotency regulator LIN28A and the RNA m^6A reader YTHDF2 [30], were exclusive to our own dataset. These results support the reliability of our proteomic experiments and analysis.

RNA binding proteins are highly represented in the four cell states crotonylome

We then performed gene ontology (GO) and KEGG pathway analysis of the 1426 high-confidence crotonylated proteins. Interestingly, GO analysis showed enrichment in RNA metabolism-related and ribosome-related terms (Figure 2E and 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 and 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 m^6A modification on RNA [36], was identified as a high-confidence Kcr site. This may affect its binding affinity to m^6A modified RNA and in turn regulate cell fate, as this RNA modification is

243 necessary for pluripotency exit [37]. As for the ribosome, it is known that
244 post-translational modifications of ribosome subunits influence protein translation
245 [38], also suggesting a potential link of this function with cell metabolism. Notably,
246 translational control has been implicated in pluripotency maintenance and
247 differentiation [39, 40]. KEGG analysis showed enrichment in terms related to carbon
248 metabolism and the proteasome (Figure 2F and Table S2A).

249 Because lysine residues are subjected to other posttranslational modifications in
250 addition to crotonylation, we also compared our Kcr sites with reported acetylation,
251 malonylation, succinylation, and ubiquitination mouse datasets from the Protein
252 Lysine Modifications Database (PLMD) [41]. Our high-confidence Kcr sites
253 significantly overlapped with other modifications, but a large number of Kcr sites
254 were unique (Figure S3C and Table S2C). Moreover, we observed that the average
255 amino acid distribution around crotonylated lysines is enriched in glutamic acid (E),
256 in agreement with a previous crotonylome study [15] (Figure S3D). The negative
257 charge of glutamic acid residues might synergize with the crotonylation-mediated
258 suppression of lysine's positive charge, changing the affinity of RBPs for substrates.
259 Interestingly, it was reported that the glutamic acid-lysine (EK) rich region of the
260 splicing factor SREK1/SRrp86 can function as an inhibitor of splicing [42]. Two
261 lysine residues in this same domain of SREK1/SRrp86 were identified as
262 low-confidence candidate sites in our dataset (Figure S3E and S3F).

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

266

267 **Dynamics of protein crotonylation in the four cell states**

268 We next compared the individual crotonylomes of the three pluripotent states and
269 differentiating cells to gain insight into the potential functional consequences of
270 differential protein crotonylation levels. For this purpose, we applied variance

analysis (ANOVA with a cut-off of $q < 0.01$) 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 3**A and B; Table S3A).

Cluster I corresponded to proteins highly crotonylated in differentiating cells compared to the three pluripotent states. This category included proteins related to protein translation including RPL7, LIN28A, and PDCD11. Cluster II included proteins displaying the lowest crotonylation level in EpiSCs, and was particularly enriched in proteins related to RNA localization/transport (*e.g.*, NPM1), protein folding (*e.g.*, CCT8), and telomere maintenance (*e.g.*, XRCC5). Regarding the latter, crotonylation has been shown to facilitate telomere maintenance in chemically induced reprogramming [43], suggesting a potential cause-effect link. Cluster III corresponded to proteins that displayed the highest crotonylation level in 2i ESCs compared to the other three cell states, and included proteins related to mRNA processing (*e.g.*, U2AF2) or DNA repair (*e.g.*, KDM1A). Cluster IV consisted of proteins more highly crotonylated in S/L ESCs and differentiating cells compared to 2i ESCs and EpiSCs, and included regulators of protein folding (*e.g.*, HSPA8) or tRNA aminoacylation (*e.g.*, SARS). The last category, cluster V, were proteins more highly crotonylated in EpiSCs, and contained 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 (TCA cycle, pentose phosphate pathway [PPP], and glycolysis) in the four cell stats. The crotonylation levels of these enzymes were highly dynamic among the different cell types (Figure 3C). Notably, we noticed enhanced crotonylation of glycolytic enzymes in EpiSCs compared to naïve ESCs, whereas TCA cycle and PPP enzymes were more

enriched in 2i ESCs compared to EpiSCs. These data are consistent with increased glycolytic activity being a major route for energy production in EpiSCs and with the enhanced mitochondrial activity in naive 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 function as an adaptive mechanism to environmental changes [44, 45].

We also observed that many crotonylated peptides were 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 tendency between 2i ESCs and S/L ESCs, with K23 more crotonylated in 2i ESCs and K188 less (Figure S4A and S4B). This suggests: a) that crotonylation of different lysine residues of the same protein could have diverse functional consequences depending on the cell state, and b) 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 that experience changes in their crotonylation status might be more critically involved in determining shifts in cell identity between the four cell states. After normalization, we identified 485 differential sites in 330 proteins between the four cell states (Figure S4C and S4D; Table S3C). These sites and proteins distributed in five clusters. Cluster I included proteins highly crotonylated in differentiating cells and was 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 consisted of proteins highly crotonylated in 2i PSCs and was enriched in GO terms related to mRNA processing, ribonucleoprotein complex assembly, and chromatin remodelling.

The former GO terms included the classical RBP HNRNPU [47] and the deacetylase/decrotonylase HDAC1[48], which has been reported to regulate pluripotency. Cluster III represented proteins highly crotonylated in EpiSCs and contained proteins involved into RNA splicing and transport (*e.g.*, DHX9). Cluster IV represented proteins displaying lower crotonylation in EpiSCs and it was only enriched for proteins related to protein folding (*e.g.*, HSPA9). Cluster V contained proteins more crotonylated in EpiSCs compared to the other cell states. This cluster had higher crotonylation in EpiSCs than Cluster III; however, there were no enriched GO terms from these 20 proteins.

Histone crotonylation is highly dynamic during ESCs differentiation and is required for self-renewal [49]. Besides the analysis of non-histone proteins, we identified 25 crotonylation sites in histone proteins (Figure S5 and Table S3D). While extracting the histone for mass spectrometry might increase the number of identified histones crotonylation 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 extend beyond the better studied histone targets, likely play relevant roles in controlling the transition between different pluripotent states and pluripotency exit.

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 genes *Fgf5* and *Pax6* compared to the control (**Figure 4**A). S/L ESCs treated with crotonic acid that were allowed to gradually differentiate upon LIF withdrawal for four days also exhibited higher

pluripotency gene and lower differentiation gene levels than non-treated ESCs (Figure 4B), suggesting that enhanced protein crotonylation maintains pluripotency and slows pluripotency exit. It has been reported that crotonic acid treatment in mouse ESCs downregulates a group of pluripotency genes while upregulating two-cell (2C)-like state genes [43]. This 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 PSC cultures 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 (~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]. This 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 and Table S4A). Indeed, most of the proteasome complex subunits were crotonylated (29/34 total and 22/34 high-confidence) in our dataset (Figure 4C and Table S4B). It has been shown that proteasome activity is positively associated with pluripotency maintenance [17, 18], and our crotonylated 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 or untreated with crotonic acid. Adding crotonic acid significantly enhanced proteasome activity compared to the control in crude lysate (Figure 4D) and also in the native gel assay (Figure 4E and

4F). 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 crotonylation signals caused by adding crotonic acid associate 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.

Taken together, these experiments suggested 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].

Conclusions

Our global atlas of protein crotonylation in four different pluripotent cell states 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 impact of crotonylation on the function of these proteins, and also how these effects crosstalk with the consequences on gene expression of histone crotonylation [49]. Hence, our 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 previous study [52]. For 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, United Kingdom] 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.

ESC and HEK293T 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, 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, Nuaillé, 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 four days. HEK293T cells were cultured in DMEM-high glucose supplemented with 10% fetal bovine serum.

Protein extraction and trypsin digestion for LC-MS/MS

Cells were lysed on ice in a buffer containing 8 M urea, 10 mM DTT, 50 mM nicotinamide, 3 μM TSA, and 1× protease inhibitor cocktail (Catalog No. 11697498001, Roche, Basel, Switzerland). Lysates were centrifuged (20,000 g) at 4

°C for 10 minutes. Supernatants were precipitated with cold 15% trichloroacetic acid at -20 °C for 2 hours, following 20,000 g centrifugation at 4 °C for 10 minutes. The precipitated proteins were dissolved in a pH 8.0 buffer containing 8 M urea and 100 mM triethylammonium bicarbonate buffer. The protein solution was reduced with 10 mM DTT at 37 °C for 1 hour followed by alkylation with 20 mM iodoacetamide at room temperature for 45 minutes 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 hours.

TMT labelling

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 labelled 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 hours at room temperature and then the samples were desalted.

Peptide fractionation

TMT-labelled 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, labelled peptides were separated with a gradient of 2-60% acetonitrile in 10 mM ammonium bicarbonate into 80 fractions for 80 minutes. 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 antiKcr 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 minutes, 20% to 35% in 8 minutes and climbing to 80% in 5 minutes then holding at 80% for the last 3 minutes, all at a constant flow rate of 300 nl/minute 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-terminal of the protein were specified as variable modifications. FDR threshold 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 were used default setting.

Statistical analysis

All statistical analyses were performed using R (3.5.1). For protein homology analysis, protein sequences in identified human crotonylated proteins 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 < 1*e*-10 were considered as overlapping proteins in the human crotonylated proteins 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).

522

523 **Cellular compartment analysis**

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

529

530 **GO and KEGG annotation**

531 GO and KEGG annotations were performed using the R package "clusterProfiler"
532 (3.10.1) [54] with the 'enrichGO' and 'enrichKEGG' function, respectively.

533

534 **Co-modification analysis**

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

539

540 **Protein domain analysis**

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

546

547 **Crotonylation motif analysis**

548 The frequency of amino acids surrounding crotonylated lysine residues was generated
549 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, 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 (Catalog No. B01020001 Diagenode, Ougrée, Belgium) sonicator (low power, 30 second ON/OFF, 5 cycles), followed by 10,000 g centrifugation at 4 °C for 10 minutes to remove the undissolved particles. Then, they were incubated with pre-washed anti-FLAG M2 magnetic beads (Catalog No. M8823, Sigma, Darmstadt, Germany) at 4 °C overnight with rotation, and eluted by heating at 95 °C for 5 minutes 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 minutes to ensure inactivation of deubquitinase enzymes. Denatured lysates were sonicated with a Bioruptor sonicator (low power, 30 second ON/OFF, 5 cycles), followed by 10,000 g centrifugation at 4 °C for 10 minutes 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, United Kingdom) 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 minutes, the supernatants were incubated with proteasome substrate (Suc-LLVY-AMC) at 37 °C for 20 minutes. The released free AMC (7-amino-3-methylcoumarin) fluorescence was detected on a microplate fluorometer (350 nm excitation, 440 nm emission) after 30 minutes and then 60 minutes at 37 °C. Background signal was corrected by subtracting the 30-minute reading from the 60-minute reading and data were then normalized against the control sample.

Native gel assay for proteasome activity

Native gel assay for proteasome activity was performed according to the 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 $MgCl_2$, 5 mM ATP, 1 mM DTT, and 10% glycerol) and lysate by passing 10 times through the syringe with 27-G needled. After 12,000 g centrifugation at 4 °C for 10 minutes, the supernatants were loaded on 3–12% NativePAGE Bis-Tris gel (Catalog No. BN1003BOX, Invitrogen, Carlsbad, CA) in NativePAGE running buffer (Catalog No. BN2001, Invitrogen) containing 5 mM $MgCl_2$ and 1 mM ATP, and run at 4 °C for 150 min at 150 V. The native gel was then incubated with 300 μM Suc-LLVY-AMC (Catalog No. HY-P1002, MedChemExpress, Monmouth Junction, NJ) 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 minutes followed by incubation with transfer buffer for 10 minutes. 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 under licence number 2007007.

Data availability

Mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [56] partner repository with the dataset identifier PXD017121. MS/MS spectrum data of all identified crotonylated peptides have been deposited to the ProteinProspector database. They can be accessed in http://msviewer.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer by the following search keys: Replicate 1: zxlfv5pcbb; Replicate 2: g7bxkypz5c; Replicate 3: rdqvecojfj.

Authors' contributions

ZC and XB conceived the original idea and YL contributed to it. YL performed most of the experiments, CB performed the LC-MS/MS experiments, JM, JH, and MJ contributed to the experiments, YL performed the bioinformatic analyses with help from CW. MAE contributed to all data analysis. LG isolated the EpiSC cell line under the supervision of JC. XB, ZC, and MAE provided financial support and supervised the project. MAE, XB, and YL wrote the manuscript. CB, CW, and GV helped to

revise the manuscript. All authors read and approved the final version of the 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.

Acknowledgements

We thank all members of the Laboratory of Integrative Biology for their support. We also thank Xiaofei Zhang, from the Guangzhou Institutes of Biomedicine and Health, for constructive advice on proteomic data analyses. This work was supported by the National Key Research and Development Program of China (2016YFA0100102, 2016YFA0100701, and 2018YFA0106903), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502), the Natural Science Foundation of Guangdong Province (2018B030306042), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2015294), and an Innovation Team Project grant from the Guangzhou Regenerative Medicine and Health Guangdong Laboratory (2018GZR110103001). Carl Ward is supported by a Zhujiang Talent-Overseas Postdoctoral Funding Grant and a President's International Fellowship Initiative grant from the Chinese Academy of Sciences.

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: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol 2016;17:155–69.

[3] Nichols J, Smith A. Naive 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, Menafra R, Denissov 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 naive-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. HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. EMBO J 2012;31:2103–16.

[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, Ratanasirintrawoot 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 D, 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 D, Chung D, Tang Z, Huang H, Dai L, et al. Metabolic Regulation of Gene Expression by Histone Lysine beta-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 JS, 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 SM, 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, Tsafou 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, Fohr 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 DC, Wilusz JE, Garcia BA, 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 CK, Horos R, Alleaume AM, Foehr S, et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol Cell 2016;63:696–710.
[35] Ye J, Blelloch 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 L, 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, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, et al. Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science 2015;347:1002–6.
[38] Simsek D, Tiu GC, Flynn RA, Byeon GW, Leppek 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. Nat1 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 Reports 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

779 critical for stem cell pluripotency via promoting mitochondrial fission. EMBO J
780 2017;36:1330–47.
781 [47] Vizlin-Hodzic D, Johansson H, Ryme J, Simonsson T, Simonsson S. SAF-A has a
782 role in transcriptional regulation of Oct4 in ES cells through promoter binding. Cell
783 Reprogram 2011;13:13–27.
784 [48] Kidder BL, Palmer S. HDAC1 regulates pluripotency and lineage specific
785 transcriptional networks in embryonic and trophoblast stem cells. Nucleic Acids Res
786 2012;40:2925–39.
787 [49] Wei W, Liu X, Chen J, Gao S, Lu L, Zhang H, et al. Class I histone deacetylases
788 are major histone decrotonylases: evidence for critical and broad function of histone
789 crotonylation in transcription. Cell Res 2017;27:898–915.
790 [50] Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al.
791 Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature
792 2012;487:57–63.
793 [51] Myeku N, Clelland CL, Emrani S, Kukushkin NV, Yu WH, Goldberg AL, et al.
794 Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented
795 early in disease by activating cAMP-PKA signaling. Nat Med 2016;22:46–53.
796 [52] Liu J, Han Q, Peng T, Peng M, Wei B, Li D, et al. The oncogene c-Jun impedes
797 somatic cell reprogramming. Nat Cell Biol 2015;17:856–67.
798 [53] Cox J, Mann M. MaxQuant enables high peptide identification rates,
799 individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.
800 Nat Biotechnol 2008;26:1367–72.
801 [54] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing
802 biological themes among gene clusters. OMICS 2012;16:284–7.
803 [55] Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved
804 visualization of protein consensus sequences by iceLogo. Nat Methods 2009;6:786–7.
805 [56] Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu
806 DJ, et al. The PRIDE database and related tools and resources in 2019: improving
807 support for quantification data. Nucleic Acids Res 2019;47:D442–D50.

808

809

Figure legends

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 crotonylome profiling. **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 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, colours in the heatmap indicate the pairwise Pearson correlation between the different samples (n = 3628). Epi: EpiSCs; Diff: Differentiating cells. **D**. Left: Principal component analysis 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.

Figure 2 Characteristics of the crotonylated proteins identified in three pluripotent states and differentiating cells.

A. Bar plot indicating the number of crotonylated sites in each high-confidence protein in our crotonylome datasets. **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 administration of 10 mM crotonic acid for 24 hours. Cell lysates were immunoprecipitated with anti-FLAG magnetic beads and analyzed by immunoblotting with anti-PAN Kcr antibodies. **E**. GO analysis of the high-confidence crotonylated proteins, the top 6 terms with smallest adjusted *P* value are shown

838 (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$). **F**. KEGG analysis of
839 the high-confidence crotonylated proteins, the top 11 terms with smallest adjusted *P*
840 value are shown (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$).
841
842 **Figure 3 Functional clustering of the crotonylated sites in three pluripotent**
843 **states and differentiating cells.**
844 **A**. Heatmap of crotonylated sites in high-confidence proteins shows distinct patterns
845 in the four cell states (n = 2306, ANOVA test, FDR < 0.01). Normalized $\log_2$ intensity
846 were grouped by hierarchical clustering into five clusters. **B**. GO analysis for all five
847 clusters (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$, scaled without
848 centering). **C**. Heatmap of differentially crotonylated enzymes involved in central
849 carbohydrate metabolism (glycolysis, TCA cycle, and PPP pathway) (Fisher's exact
850 test, Benjamini-Hochberg corrected $P < 0.01$). **D**. Box plot indicating the number of
851 crotonylated proteins overlapping between two clusters. Arrows indicate the direction
852 of data processing. **E**. Crotonylation levels of specific lysine sites in the overlapped
853 proteins between different clusters.
854
855 **Figure 4 Crotonic acid promotes pluripotency and enhances proteasome**
856 **activity.**
857 **A**. Relative expression of pluripotency and differentiation genes measured by
858 RT-qPCR in ESCs cultured in S/L with or without 10 mM crotonic acid. Data are
859 presented as the mean ± S.E.M. (n = 3 biological replicates with three technical
860 replicates each, two-tailed unpaired Student's *t*-test). $*P < 0.05$, n.s.: not
861 significant. **B**. Relative expression of representative pluripotency and differentiation
862 genes measured by RT-qPCR in ESCs cultured in S/L and subjected to differentiation
863 by withdrawing LIF over a time-course of 4 days. Data are presented as the mean ±
864 S.E.M. (n = 3 biological replicates with three technical replicates each, two-tailed
865 unpaired Student's *t*-test). $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$, n.s.: not

significant. **C**, Schematic representation of the proteasome complex and the components identified in our crotonylome. Gray (dashed) ellipse: not detected; blue ellipse: low-confidence; red ellipse: high-confidence. **D**. Chymotrypsin-like proteasome activity measurement in ESCs cultured in S/L with or without 10 mM crotonic acid treatment for 48 hours (n = 4 biological replicates, two-tailed unpaired Student's *t*-test). *$P < 0.05$. **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 hours. Samples were resolved by denatured SDS-PAGE and western blotting for analysis the 20S proteasome level and loading control. **F**. Quantified densitometry of the result shown in **E**, data are presented as the mean ± S.E.M. (n = 4 biological replicates, two-tailed unpaired Student's *t*-test). *$P<0.05$. **G**. Representative western blotting of lysine crotonylation, poly-ubiquitination, and K-48 ubiquitination in ESCs cultured in S/L and untreated or treated with different concentrations of crotonic acid for 48 hours. ACTIN was the loading control.

Table 1 Summary of proteome and crotonylome profiling results.

Supplementary material

Supplementary Figure S1 Characterization of different cell states. *Related to Figure 1*.

A. Representative images showing PSCs cultured in different conditions, bright-field (phase) and green fluorescence for the same fields are shown. Scale bar = 100 μm. **B**. Relative expression of selected pluripotency and differentiation genes measured by RT-qPCR in the four different cell states. Data are presented as mean ± S.E.M. (n = 3 biological replicates with 3 technical replicates each, two-tailed unpaired Student's *t*-test). *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$, n.s.: not significant. *Gapdh* was used as the housekeeping control gene. **C**. Hierarchical clustering of the normalized $\log_2$ intensity for the total proteome of three different LC-MS/MS experiments (biological replicates) in four different cell states, colours in the heatmap indicate pairwise Pearson correlation between the different datasets (n = 4728). **D**. Normalized $\log_2$ intensity of 'Core ESC-like gene module' and 'Adult tissue stem module' gene sets from MSigDB in the four different cell states (n = 3 biological replicates, two-tailed unpaired Student's *t*-test). **$P < 0.01$, and ***$P < 0.001$, n.s.: not significant. **E**. Heatmap showing the protein expression levels of crotonylation regulators in our total proteome analysis. **F**. Western blotting showing the global protein crotonylation levels in the four different cell conditions. **G**. Left: PCA of the total proteome in the four different cell states is shown; the variance was calculated using normalized $\log_2$ intensity. Right: Euclidean distances of proteome profiles between the indicated cell conditions.

Supplementary Figure S2 MS/MS spectrum of identified Kcr peptides of the 15 western blotting-validated proteins in Figure 2D.

Supplementary Figure S3 Functional characterization of the crotonylome in different cell states. *Related to Figure 2*.

A. Protein domain enrichment analysis of the high-confidence crotonylated proteins, the top 10 terms with the smallest adjusted *P* value are shown (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$). **B**. Venn diagrams comparing our crotonylated datasets with published mouse ESCs RBP datasets. **C**. Comparison of the numbers of the overlapped lysine sites between our high-confidence crotonylome and other lysine modification datasets in the PLMD database. Ac: acetylation; ubi: ubiquitination; suc: succinylation; mal: malonylation; glu: glutarylation. **D**. Consensus sequence motif extracted from all high-confidence Kcr sites in our datasets. **E**. Schematic showing lysine crotonylation sites in the EK rich region of SREK1/SRrp86. **F**. MS/MS spectrum of K321cr (left) and K334cr (right) for SREK1/SRrp86.

Supplementary Figure S4 Crotonylation of SET protein in different cell states. ***Related to Figure 3.***

A. Crotonylation levels of two lysine sites of SET protein in the four different cell states. Data are presented as the mean ± S.E.M. (n = 3 biological replicates, two-tailed unpaired Student's *t*-test). $*P < 0.05$, and $***P < 0.001$. **B**. Schematic showing the high-confidence Kcr sites of SET. Green circle: higher crotonylation level in S/L ESCs than in 2i ESCs; red circle: higher crotonylation level in EpiSCs than in 2i ESCs. **C**. Heatmap of differential crotonylated sites in the four different cell states after being normalized by protein abundance (n = 485, ANOVA test, FDR < 0.01). **D**. GO biological process analysis for the different clusters of proteins in **C** (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$).

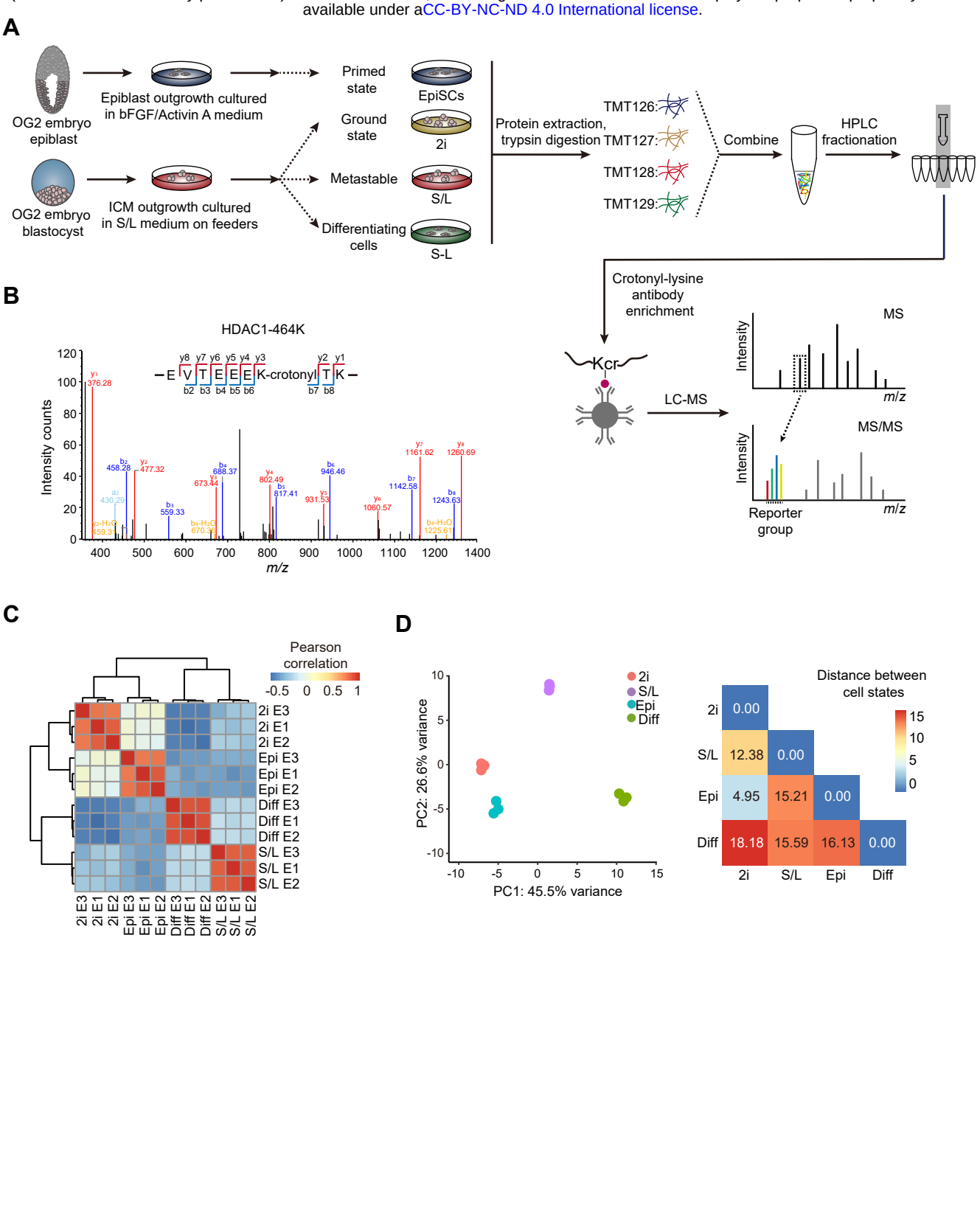
Supplementary Figure S5 Histone crotonylation analysis in different cell states. ***Related to Figure 4.***

Diagrams showing the quantitative information of histone Kcr sites in the four cell states of our study. For the isoform-specific histone sites that occurred at the same position, the mean normalized intensity has been used.

Supplementary Figure S6 Enrichment for ESC-specific proteins in our crotonylome. ***Related to Figure 4*****.**

A. Venn diagram comparing our crotonylome datasets with the 'ESC-specific gene sets'. **B**. GO terms for the 466 overlapping proteins from (**A**) (Fisher's exact test, Benjamini-Hochberg corrected $P < 0.01$).

Figure 1



Supplementary Figure 1

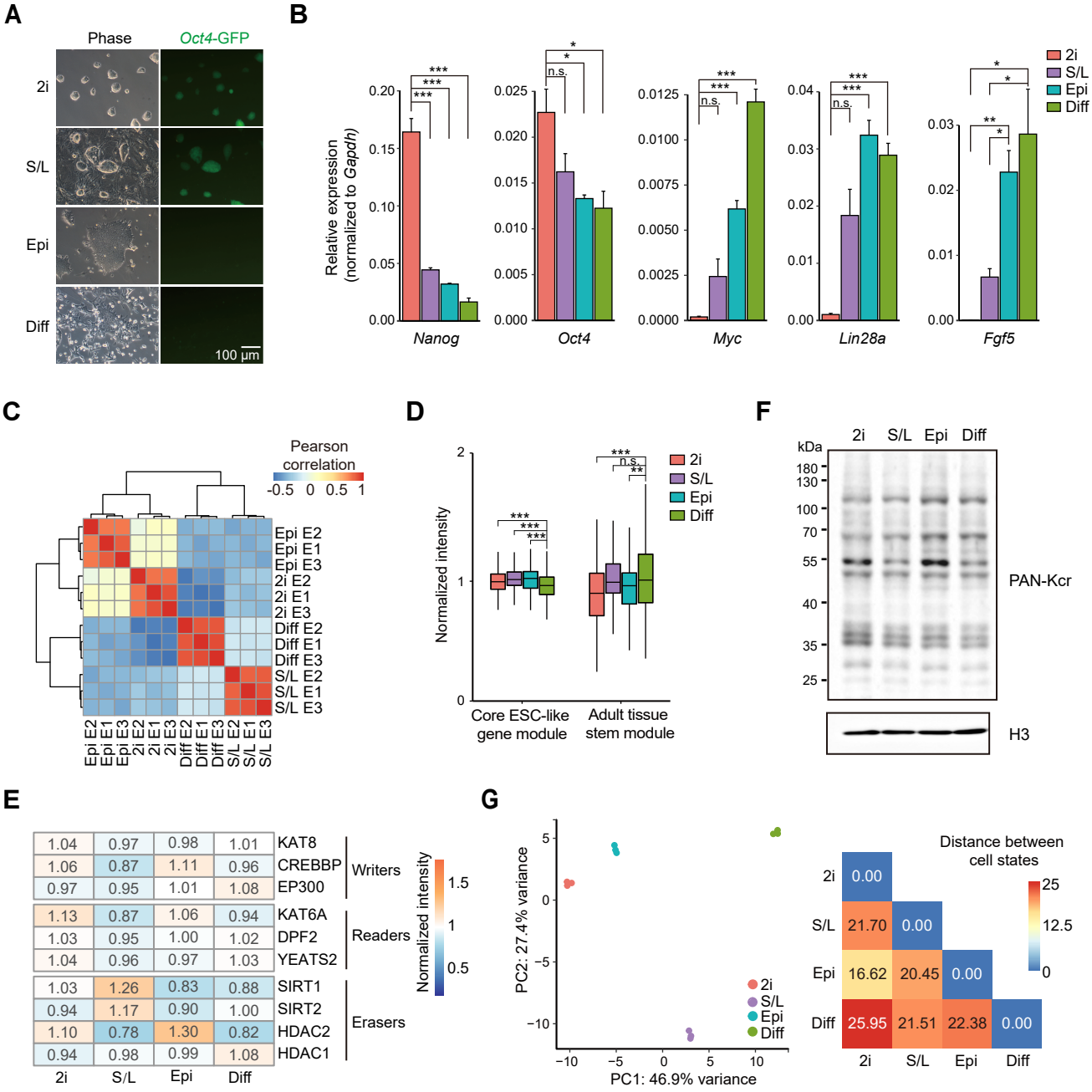
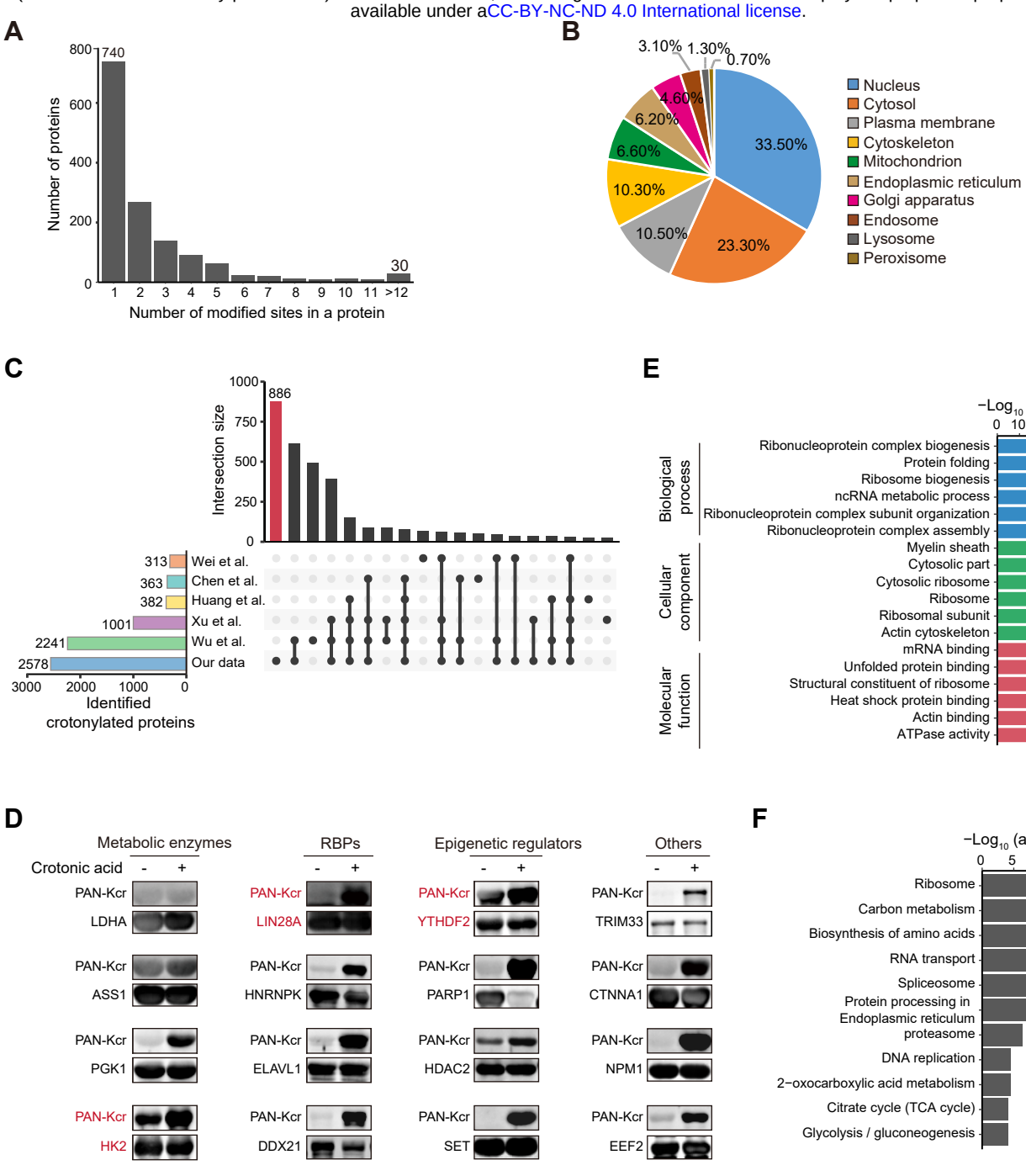
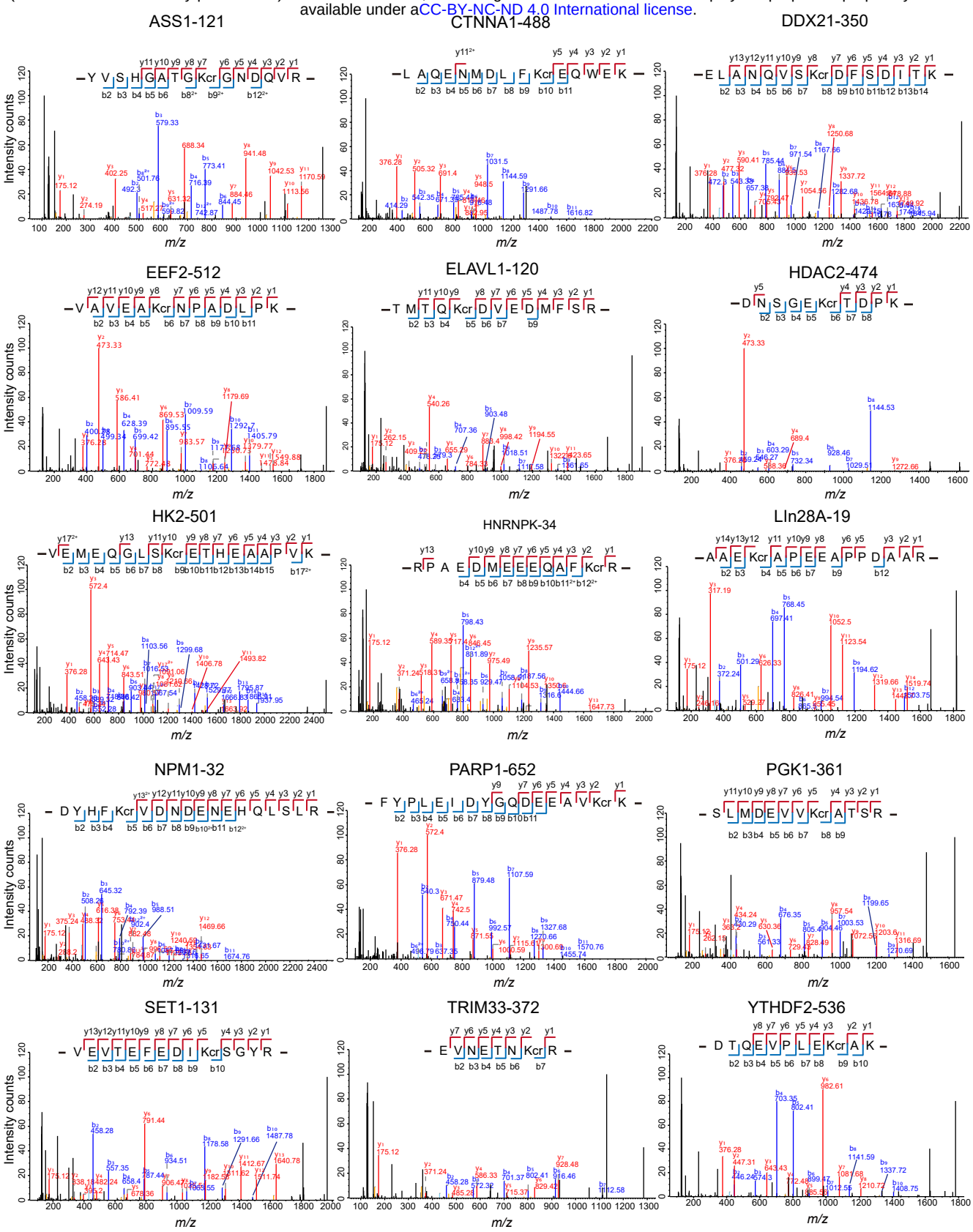


Figure 2



Supplementary Figure 2



Supplementary Figure 3

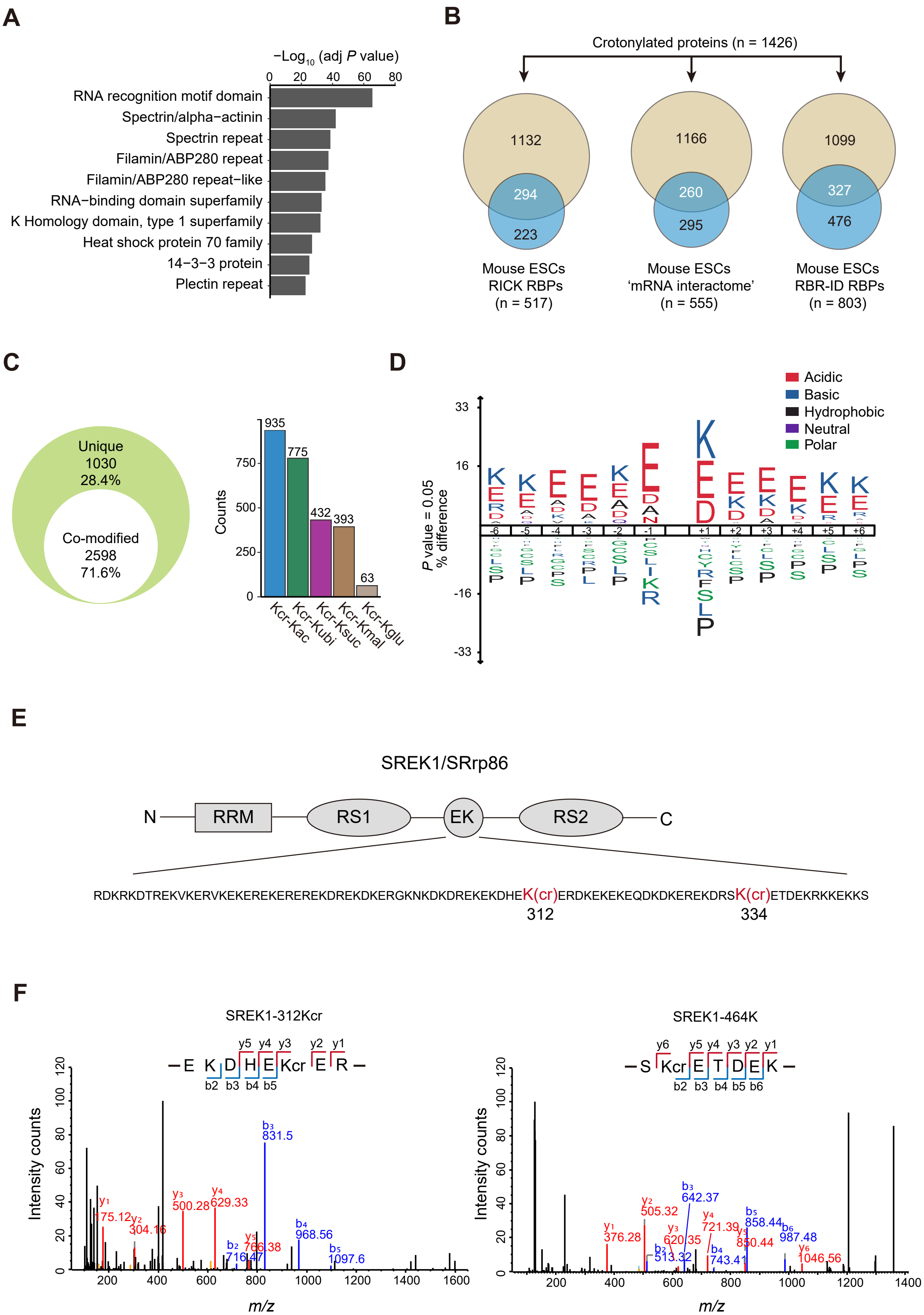
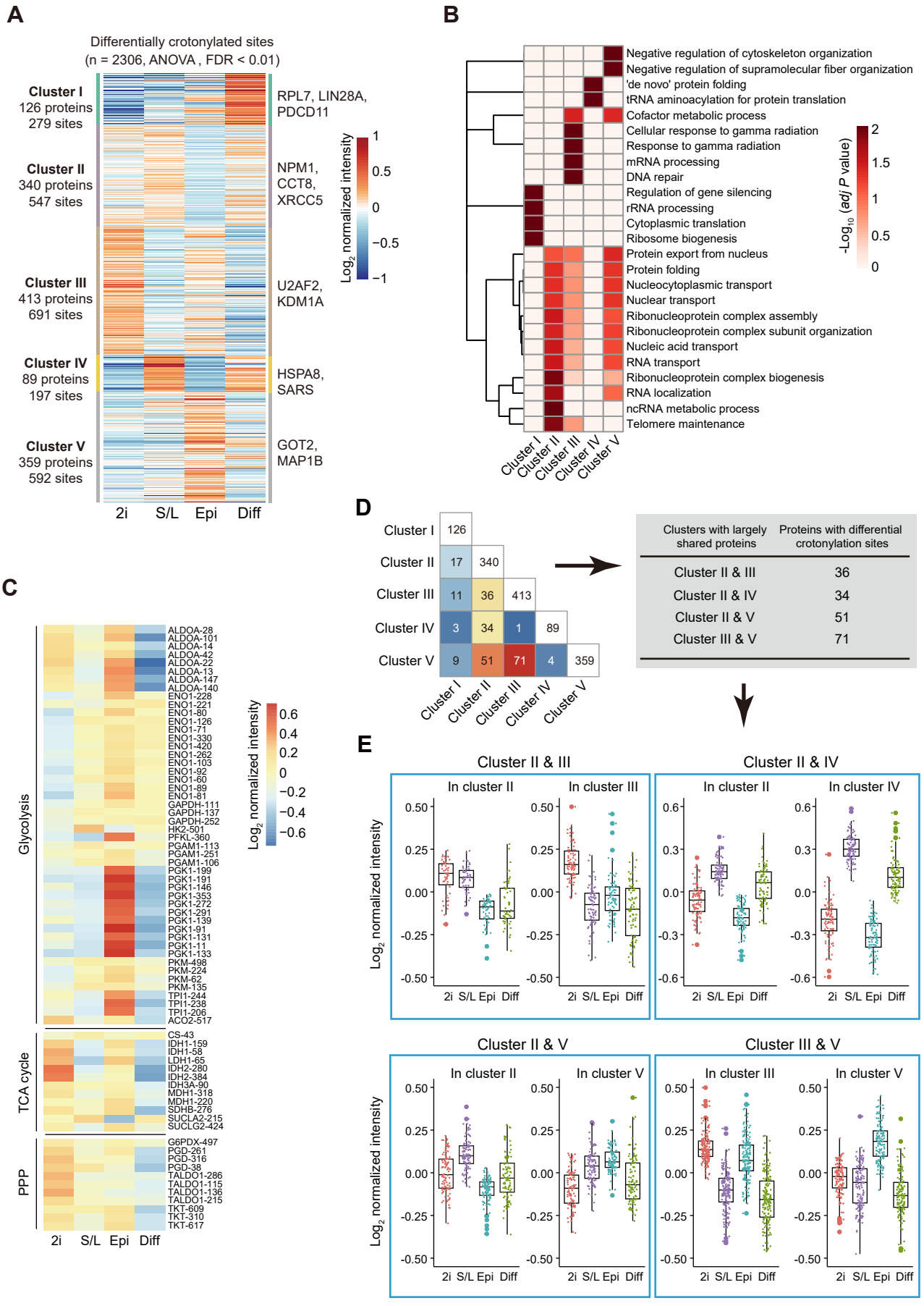
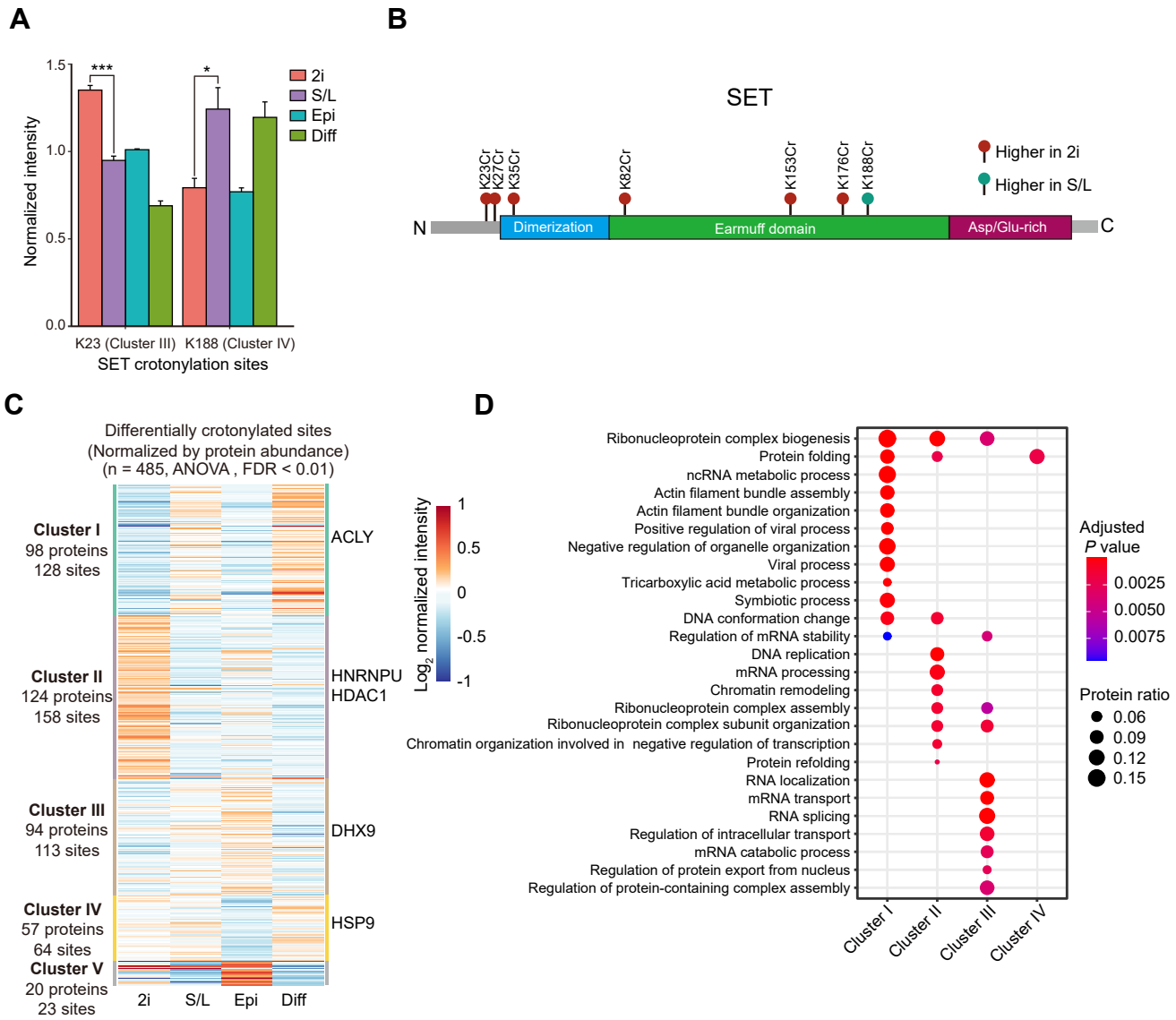


Figure 3



Supplementary Figure 4



Supplementary Figure 5

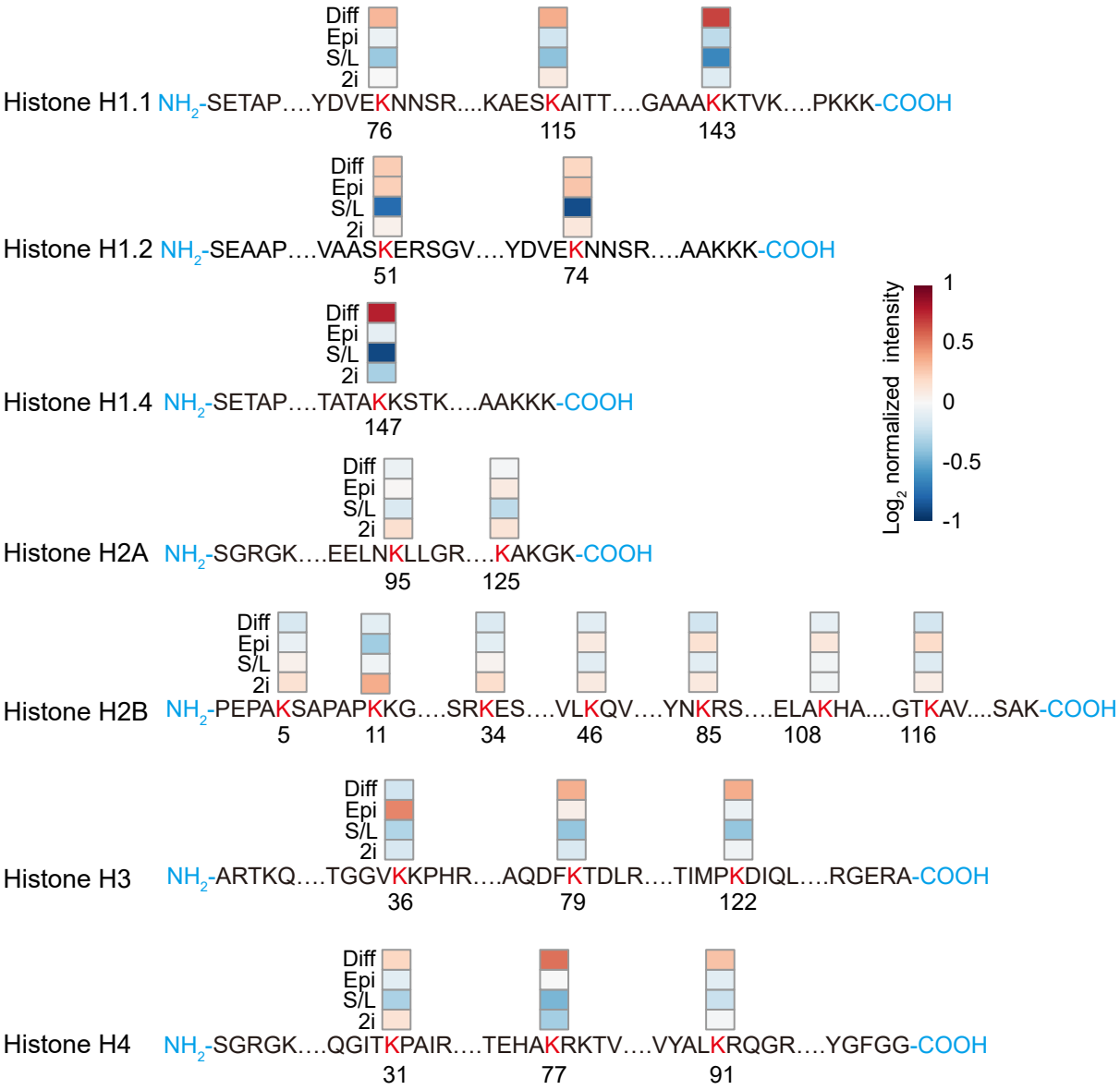
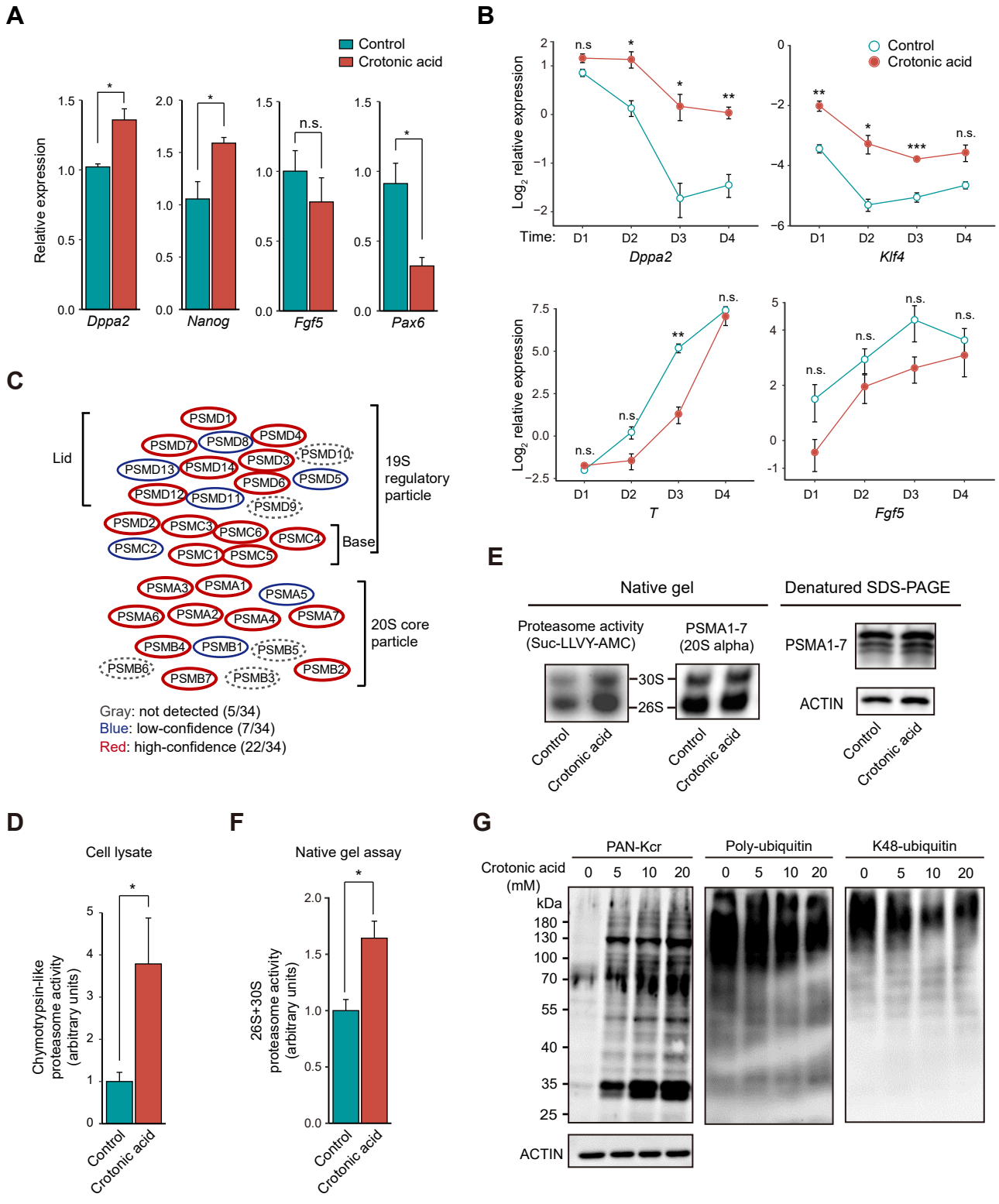


Figure 4



Supplementary Figure 6