UCLA
UCLA Electronic Theses and Dissertations

Title

Investigating the Mechanism of Somatic cell Reprogramming and Developing Methodologies in Bottom-up Proteomics

Permalink

https://escholarship.org/uc/item/90p2s1sf

Author

Deng, Weixian

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Investigating the Mechanism of Somatic cell Reprogramming and Developing

Methodologies in Bottom-up Proteomics

A dissertation submitted in partial satisfaction of the

requirements for a degree Doctor of Philosophy

in Molecular Biology

by

Weixian Deng

2022

ABSTRACT OF THE DISSERTATION

Investigating the Mechanism of Somatic Cell Reprogramming and Developing

Methodologies in Bottom-up Proteomics

by

Weixian Deng

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2022

Professor Kathrin Plath, Co-Chair

Professor James A. Wohlschlegel, Co-Chair

Ectopically expressing the transcription factors (TFs) Oct4, Sox2, Klf4, and c-Myc (OSKM) leads to the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs). iPSC reprogramming takes several weeks and yields pluripotent cells only at low frequencies indicating that the reprogramming factors need to overcome barriers established in somatic cells to preserve cell identity. Yet, the mechanisms driving the successful decommissioning of the starting somatic program and the activation of the target pluripotency program are currently unclear. A recent study from the Plath lab has begun to determine how OSKM induce the remodeling of enhancers and induce the

transition from somatic to pluripotency enhancers during the reprogramming of mouse embryonic fibroblasts (MEFs) to iPSCs. The key finding was that the collaborative binding of OSK is essential for the step-wise selection and activation of pluripotency enhancers (PEs) throughout reprogramming. Consequently, the target sites of OSKM gradually change during reprogramming to mediate the step-wise induction of pluripotency enhancers. Intriguingly, the reprogramming factors also act on MEF enhancers (MEs). Based on ChIP-seq results, it was suggested that OSK redirect somatic (endogenously expressed) TFs away from MEs to new sites opened by the reprogramming factors. Concomitantly, the active enhancer mark H3K27ac is decreasing at MEs, suggesting that the movement of somatic TFs is critical for the inactivation of MEs. The key questions in the field now are to understand how OSKM, at a mechanistic level, induce (i) the redistribution of somatic TFs and (ii) the decommissioning of MEs in the early stage of reprogramming.

In my graduate work, I am addressing these questions by taking advantage of co-mentorship in the Plath and Wohlschlegel labs. In Chapter 2, I hypothesize that both protein-protein interactions (PPIs) of the reprogramming factors with somatic TFs and the newly opened sites containing somatic TFs' binding motifs are critical for the redistribution of somatic TFs binding and the following somatic enhancer decommissioning. Consequently, I am combining functional experiments with mass spectrometry (MS) methodologies to i) define co-binding between OSK and somatic TFs, ii) to distinguish mechanisms of somatic TFs redistribution through direct PPIs, cooperative binding, and open-sites free binding and iii) identify the mechanism of how active histone mark is removed from MEs. However, since TFs are often of low abundance in cells, MS

approaches with a large dynamic range are required for the identification and quantification. Therefore, another aspect of my graduate work is to develop and optimize cutting-edge bottom-up proteomics methodologies for the assessment of low abundance proteins.

In Chapter 3, to achieve the better identification and quantification of lowly abundant proteins in complex protein mixtures, I developed a bead-based off-line peptide fractionation method termed: CMMB (Carboxylate Modified Magnetic Bead) -based isopropanol gradient peptide fractionation or 'CIF'. CIF provides an effective but low material loss alternative to other fractionation methods.

In Chapter 4, by combining optimized proteomics and cell biology approaches, we uncovered an understudied mechanism of nuclear proteome regulation: activity-dependent proteasome-mediated degradation. We found that the tumor suppressor protein PDCD4 undergoes rapid stimulus-induced degradation in the nucleus of neurons. We demonstrate that degradation of PDCD4 is required for normal activity-dependent transcription and that PDCD4 target genes include those encoding proteins critical for synapse formation, remodeling, and transmission.

In Chapter 5, for improving the quantification of proteins of interest, targeted proteomics assays are often used to pursue more accurate quantitation and better sensitivity. The recently launched High-field asymmetric waveform ion mobility spectrometry (FAIMS) device enables the possibility of improving conventional targeted proteomics assay data quality, while such improvement relies heavily on tuning the parameters of the FAIMS settings, in my thesis work, I investigated the molecular determinants underlying peptide separation by FAIMS and demonstrate that the machine learning model can be used to

predict optimized FAIMS settings for peptides which significantly improves targeted proteomics workflows.

The dissertation of Weixian Deng is approved.

Siavash K. Kurdistani

Jeffrey Aaron Long

April Dawn Pyle

Joseph Ambrose Loo

James A. Wohlschlegel, Committee Co-Chair

Kathrin Plath, Committee Co-Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENT

It has been an amazing journey doing Ph.D. in such two amazing labs, and it's my privilege to thank all the people accompanying me throughout the training both inside and outside of the lab.

Time flies, it has almost been 12 years since the first time I got to learn the term molecular biology. I'd like to thank Drs. Zhenming Yang, Mingdi Bian, Jie Gao, Shengzhong Su and Chentao Lin for bringing me into thinking about life from molecular scale. A special shout out to Dr. Zecheng Zuo for giving me opportunity to explore various different technologies and especially for opening the door of learning mass spectrometry to study protein. Without you, I would not have the chance to be trained at UCLA.

I'm always grateful to have two such great mentors, Dr. Kathrin Plath and Dr. James Wohlschlegel. They have been super supportive to my training, projects, learning skills and developing my career. Their distinct training styles have made a great chemistry so that I learned to be careful on making conclusions, to view biological experiments more quantitatively, to have the courage to try different ideas and to take risk of making mistakes. These are invaluable assets, and they will be beneficial for my entire life. I will always remember their guidance. Beyond my supervisors, I'm extremely grateful to my thesis committee members too, Dr. Siavash Kurdistani, Dr. Jeff Long, Dr. April Pyle and Dr. Joseph Loo, I always appreciate them for not only their scientific inputs in my projects, but also career support as well, most importantly encouragement when I was feeling unconfident.

I would like to thank fantastic lab members of both labs. I'm thankful to Dr. William Barshop for being a model of how an excellent graduate student should be like and

VITA

Education

M.Sc.in Plant Molecular Biology and Biochemistry	2013-2016

Jilin University- Changchun, China

B.Sc. in Agriculture, Plant Biology	2009-2013

Jilin University- Changchun, China

Publications and Manuscripts

1. **CMMB-based isopropanol gradient peptide fractionation (CIF) enables rapid and robust off-line peptide mixture fractionation in bottom-up proteomics.**
   <u>**Weixian Deng**</u>**,** Jihui Sha, Kathrin Plath, James A. Wohlschlegel**. (2021)** *Molecular Cellular Proteomics*

2. **The transcription factor code in iPSC reprogramming**.
   <u>**Weixian Deng**</u>***,** Elsie C Jacobsen*, Amanda J Collier*, Kathrin Plath. **(2021)** *Curr. Opin. Genet. Dev.*

3. **Neuronal activity regulates the nuclear proteome to promote activity-dependent transcription.**
   Wendy A. Herbst, <u>**Weixian Deng**</u>, James A. Wohlschlegel, Jennifer M Achiro, Kelsey C. Martin**. (2021)** *J Cell Biol*.

4. **Molecular basis for blue light-dependent phosphorylation of Arabidopsis cryptochrome 2.**
   Qing Liu*, Qin Wang*, <u>**Weixian Deng**</u>***,** Xu Wang, Dawei Cai, Mingxin Piao, Yaxing Li, William Barshop, Xiaolan Yu, Bin Liu, Yoshito Oka, James Wohlschlegel, Zecheng Zuo, Chentao Lin**. (2017)** *Nature Communications*

5. **Using HEK293T Expression System to Study Photoactive Plant Cryptochromes.**
   Liang Yang*, Xu Wang*, <u>**Weixian Deng**</u>, Weiliang Mo, Jie Gao, Qing Liu, Zhang C, Wang Q,Lin C and Zuo Z. **(2016)** *Frontiers in Plant Science*

6. **Trp triad-dependent rapid photoreduction is not required for the function of Arabidopsis CRY1.**
   Jie Gao*, Xu Wang*, Meng Zhang*, Mingdi Bian*, <u>**Weixian Deng**</u>, Zecheng Zuo, Zhenming Yang, Dongping Zhong and Chentao Lin. **(2015)** *Proc Natl Acad Sci U S A*

Manuscripts:

1. **High-field asymmetric waveform ion mobility spectrometry interface enhances parallel reaction monitoring on an Orbitrap mass spectrometer.**
   <u>**Weixian Deng**</u>***,** Jihui Sha*, Yasaman Jami-Alahmadi, Fanglei Xue, Kathrin Plath, James Wohlschlegel. **(2022)** *Analytical Chemistry*, *in review.*

2. **RAP-MS 2.0: A new method in RNA proteomics.**
   Drew Honson*, <u>**Weixian Deng**</u>*****, Mario Blanco, Ali Palla, James Wohlschlegel, Mitchell Guttman, Dev Majumdar. *in preparation*.

3. **The mechanism and functions of the blue light-induced phosphorylation of Arabidopsis CRY1.**
   Lin Gao*, Qing Liu*, Nannan Zeng, <u>**Weixian Deng**</u>, Yaxing Li, Yu Zhang, Dong Wang, Siyuan Liu, Qin Wang. **(2022)** *Journal of Integrative Plant Biology*, *in review*.

4. **Combined physical and in silico immunopeptidomic profiling of the cancer antigen prostatic acid phosphatase reveals peptide targets enabling cognate TCR isolation**.
   Zhiyuan Mao, Pavlo A. Nesterenko, Jami McLaughlin, <u>**Weixian Deng**</u>, Giselle Burton Sojo, Donghui Cheng, Miyako Noguchi, William Chour, Diana C. DeLucia, Kathryn A. Finton, Yu Qin, Matthew B. Obusan, Wendy Tran, Liang Wang, Nathanael J. Bangayan, Lisa Ta, Chia-Chun Chen, Christopher S, Seet, Gay M. Crooks, John W. Phillips, James R. Heath, Roland K. Strong, John K. Lee, James A. Wohlschlegel, Owen N. Witte. **(2022)** *Proc Natl Acad Sci U S A*, In review.

5. **Enhancing PD-L1 Degradation by ITCH during MAPK Inhibitor Therapy Suppresses Acquired Resistance.**
   Zhentao Yang, Yan Wang, Sixue Liu, <u>**Weixian Deng**</u>, Shirley H. Lomeli, Gatien Moriceau, James Wohlschlegel, Marco Piva, Roger S. Lo. **(2022)** *Cancer Discovery*, *In review*.

* Co-first author

Conference and Workshop Presentations

Weixian Deng, Advanced quantitative proteomics: DIA and PTMs. (Expected in Feb. 2022). Instructor of **Skyline online workshop,** session 4.

Weixian Deng, Advanced quantitative proteomics: DIA and PTMs. (Apr. 2021). Instructor of **Skyline online workshop**, session 4.

Weixian Deng, Jihui Sha, Fanglei Xue, Yasaman Jami-Alahmadi, Kathrin Plath, James Wohlschlegel**.** (November 2021) FAIMS pro and machine learning prediction of CV (compensation voltage) improve PRM performance on both Orbitrap and IonTrap mass analyzers. Poster presented at **ASMS**, Philadelphia, PA.

Weixian Deng, Jihui Sha, Kathrin Plath, James A. Wohlschlegel. (July 2020) CMMB-based isopropanol gradient peptide fractionation (CIF) enables rapid and robust off-line peptide mixture fractionation in bottom-up proteomics. Poster presented at **ASMS**, Online.

Weixian Deng, William D. Barshop, Constatinos Chronis, Kathrin Plath, James Wohlschlegel. (June 2019). Characterizing essential reprogramming factors' interactomes during cellular reprogramming towards pluripotency through multiple optimized proteomics approaches. Poster presented at **ASMS**, Atlanta, GA.

Weixian Deng, Constatinos Chronis, James Wohlschlegel, Kathrin Plath. (October 2018). The Study of Transcription Factor Movement and Enhancer Reorganization across Reprogramming Stages. Poster presented at **Wellcome Trust Genome Campus**, Cambridge, UK.

Fellowships and Awards

Audree V. Fowler Fellowship in Protein Science	2021-2022
UCLA	
Whitcome Predoctoral Fellowship	2021-2022
Molecular Biology Interdepartmental graduate program, UCLA	
Whitcome Predoctoral Fellowship	2020-2021
Molecular Biology Interdepartmental graduate program, UCLA	
Biological Chemistry Trainee Innovation Award	2020
Department of Biological Chemistry, UCLA	

Chapter 1

Introduction: Transcription factor code during iPSC reprogramming

Transcription factor code during iPSC reprogramming

Weixian Deng[1,2,*], Elsie Jacobson[1,*], Amanda J. Collier[1,*] and Kathrin Plath[1,2,#]

1 Department of Biological Chemistry, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095, USA

2 Molecular Biology Interdepartmental Graduate Program, University of California Los Angeles, Los Angeles, CA 90095, USA

3 Molecular Biology Institute, Jonsson Comprehensive Cancer Center, Brain Research Institute, Graduate Program in the Biosciences, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA 90095, USA

* co-first authors

Abstract

Transcription factor (TF)-induced reprogramming of somatic cells across lineages and to induced pluripotent stem cells (iPSCs) has revealed a remarkable plasticity of differentiated cells and presents great opportunities for generating clinically-relevant cell types for disease modeling and regenerative medicine. The understanding of iPSC reprogramming provides insights into the mechanisms that safeguard somatic cell identity,

drive epigenetic reprogramming, and underlie cell fate specification in vivo. The combinatorial action of TFs has emerged as the key mechanism for the direct and indirect effects of reprogramming factors at enhancers. The interplay of TFs in iPSC reprogramming also yields trophectoderm- and extraembryonic endoderm-like cell populations, uncovering an intriguing plasticity of cell states and opening new avenues for exploring cell fate decisions during early embryogenesis.

Introduction: overview of iPSC reprogramming

TFs are master regulators of development that determine gene expression programs and understanding how they define gene expression programs is one of the central goals of developmental biology [1,2]. In 2006, Shinya Yamanaka's laboratory stunned the developmental biology community by showing that ectopic expression of the TFs OCT4, SOX2, KLF4, and cMYC (OSKM) could reactivate the pluripotency gene network in terminally differentiated cells and establish iPSCs that carry the same features as embryonic stem cells [3-5]. Since then, iPSC reprogramming has offered a unique experimental system to explore the basic principles by which TFs drive cell fate specification.

Landmark studies to uncover the features of the transition of somatic cells into iPSCs have integrated genomics approaches such as RNA-sequencing, mapping of chromatin accessibility, chromatin marks and TF binding, the isolation of reprogramming intermediates, and applied single cell transcriptomics [6-19]. These studies revealed that fibroblasts gradually progress through a continuum of states toward a mesenchymal-to-

epithelial transition state from which a small proportion of cells continues to successfully reprogram to iPSCs. Many cells stall along this path or diverge from it to alternative cell states. Early in reprogramming, these alternative trajectories produce cells with a strong stromal identity characterized by increased expression of extracellular matrix genes, and later in the process trophectoderm-, extraembryonic endoderm-, and neural-like cells can arise in parallel to iPSCs [10,11,19,20] (Figure 1A). These findings indicate that cell fate specification is highly plastic during OSKM-induced somatic cell reprogramming, and that one reprogramming factor cocktail can result in numerous distinct gene expression programs. Along the iPSC path, cells lose the somatic gene expression program and activate the pluripotency expression program, which culminates in the hierarchical activation of pluripotency-related TFs (Figure 1B). These changes are accompanied by transient expression of genes from unrelated lineages [5,17,19] (Figure 1B). This pattern applies to iPSC reprogramming regardless of starting cell type and species [5,21]. In this review, we discuss the emerging general principles that allow the reprogramming factors to disassemble diverse somatic cell states and to activate the pluripotency program as well as alternative cell fates.

OSK-mediated rewiring of the enhancer landscape

Enhancers play a central role in cell type-specific gene expression as binding platforms that integrate the function of multiple TFs [2,22]. Therefore, understanding how the reprogramming factors act on enhancers is essential for deriving the logic of their action. The genome-wide reorganization of enhancer usage during reprogramming is predominantly driven by OSK, without cMYC, which predominantly acts at promoters

[18,23,24]. OSK action at enhancers leads to the inactivation of somatic enhancers, temporary activation of transient enhancers, and activation of pluripotency enhancers [6,7,16] (Figure 2A). During the earliest step of reprogramming, OSK predominantly bind somatic and transient enhancers, and engage only a small fraction of pluripotency enhancers [6] (Figure 2A). Over time, somatic and transient enhancers become silenced and no longer bound by OSK, concomitant with OSK binding to additional pluripotency enhancers [6,7,16] (Figure 2A). Thus, OSK bind to somatic, transient, and pluripotency enhancers, and produce different outcomes at these elements. As we discuss in detail below, OSK open chromatin by direct DNA binding and close chromatin active in somatic cells through indirect mechanisms, and both processes are linked through interactions with a small set of somatic TFs.

Somatic enhancer inactivation

How OSK repress somatic enhancers is not as well understood as pluripotency enhancer activation, yet extensive genomics approaches combined with loss- and gain-of-function experiments are beginning to shed light on this process [6-8,10,16,25]. A critical observation is that very early in reprogramming, somatic enhancers are perturbed genome-wide and display decreased levels of active enhancer marks (p300, H3K27ac), decreased chromatin accessibility, and decreased binding of somatic TFs [6-8]. These initial changes at somatic enhancers arise without a dramatic change in somatic TF expression levels and occur across most enhancers regardless of whether they are bound by OSK or not [6] (Figure 2A). Thus, global destabilization of somatic enhancers is not predominantly driven through their direct interaction with OSK.

Chronis et al found that the rapid loss of somatic TFs from somatic enhancers is accompanied by their redistribution to transient and early-engaged pluripotency enhancers [6] (Figures 2A/B). These new sites are bound by OSK and carry canonical motifs for OSK and somatic TFs [6-8,10]. The most parsimonious model explaining somatic enhancer inactivation therefore is that OSK redirect somatic TF binding by recruiting them to their target sites in newly opening enhancers and simultaneously removing them from somatic enhancers (Figures 2A/B), leading to widespread somatic gene silencing. Supporting the idea that the loss of somatic TFs is causal for somatic enhancer inactivation, the overexpression of somatic TFs that redistribute early in reprogramming (AP-1, CEBP, ETS, TEAD, RUNX family TFs, see below) blocks reprogramming whereas depletion enhances reprogramming [6,7,19,25]. A similar mechanism was later uncovered in T-cell development to explain a genome-wide gene expression switch of TFs guided by the master TF PU.1 [26], suggesting that the TF redistribution mechanism is broadly employed to induce cell fate transitions. Yet, many features of the redistribution process remain to be clarified. For instance, it is unknown whether redistribution of somatic TFs requires protein-protein interactions between somatic TFs and reprogramming factors, is critical for the opening of pluripotency enhancers together with OSK, or results from passive exploration of sites newly opened by OSK (Figure 2C). The subset of somatic enhancers bound by OSK may readily loose OSK during reprogramming because the reprogramming factors lack strong DNA binding motifs in those enhancers and may solely bind via interactions with somatic TFs or co-

factors [6-8,10] (Figure 2A), enabling their disengagement upon loss of the somatic factors.

OSKM can induce reprogramming across different somatic cell types and species [5,10]. The identity of somatic TF co-binding with OSK at transient and pluripotency enhancers is beginning to shed light on how OSK can universally disassemble different somatic networks. Although cells express hundreds of TFs, a relatively small set of somatic TF motifs are associated with sites closing and opening early in reprogramming, including AP-1, ETS, RUNX, TEAD, CEBP TF motifs, regardless of species, starting cell type, and reprogramming method [6-8,10] (Figure 2A). Intriguingly, these TFs are expressed in many cell types. For example, AP-1 family TFs are ubiquitous transcriptional effectors with a broad role in differentiation and proliferation [27]. They are critical for enhancer selection in many cell types and collaborate with cell-type specific TFs to activate respective cell-type-specific enhancers [21,28-31]. We hypothesize that broadly expressed TF that are critical for enhancer selection in vastly diverse cell types are exploited by OSK to shut off the starting cell program. Since somatic enhancer activation requires the collaborative action of AP-1 as well as somatic cell-specific TFs [1,21,26,28], OSK-mediated redeployment of a broadly acting somatic TF such as AP-1 may induce the loss of additional, somatic cell-specific TFs that depend on AP-1 co-occupancy (Figure 2B). Taken together, OSK may be highly effective reprogramming TFs because they can redistribute broadly acting somatic TFs. Although somatic TF redistribution appears to be the predominant mode of somatic enhancer destabilization, additional mechanisms are at play (Box1).

Pluripotency enhancer opening

In contrast to somatic enhancers, OSK sites in pluripotency enhancers are strongly enriched for their cognate DNA motifs (Figure 2A), indicating that sequence-specific binding is critical for their selection. Some pluripotency enhancers become bound by the reprogramming factors early during reprogramming while others become bound only later (Figure 2A). Differences in chromatin accessibility are not responsible for this difference since both early and late-occupied pluripotency enhancers are embedded in nucleosomes in starting cells [6-8,10]. The recognition of nucleosomal binding sites and eviction of histone octamers are therefore required to establish the nucleosome-free region at pluripotency enhancers that is permissive for extensive TF binding and nucleation of transcriptional machinery observed in the pluripotent end state. Accessing pluripotency enhancers in closed chromatin appears to be a critical barrier, since most TFs cannot bind nucleosomal DNA [22]. Consistent with nucleosomes representing an obstacle, suppression of the histone chaperone CAF-1 enhances reprogramming by reducing the density of nucleosomes chromatin [32].

A small number of TFs, called pioneer factors, are able to access DNA motifs wrapped in nucleosomes and to induce histone octamer displacement and to expose binding sites for additional TFs [33]. Early reports indicated that OSK are pioneer factors that can bind to nucleosomal DNA both in vivo and in vitro [24,34]. However, collaborative binding is required despite the pioneer factor activity of each reprogramming factor. Specifically, it was observed that early-engaged pluripotency enhancers are typically co-

bound by OSK, and O, S, or K cannot access these sites when expressed alone in fibroblasts [6], indicating that OSK can compete with nucleosomes only when acting together. A distinguishing feature of early- and late-engaged pluripotency enhancers is that early sites are co-bound by O, S, and K, whereas late sites tend to be bound by only O and S, without K, which correlates with motif presence (Figure 2A). Thus, early in reprogramming, O and S are not sufficient to compete with nucleosomes at late pluripotency enhancers, implying that additional TFs are required for binding site selection. One such stage-specific TF is Esrrb, which only becomes expressed late in reprogramming and co-binds late pluripotency enhancers with O and S [6,8,35] (Figure 2A). Similarly, since the opening of pluripotency (and transient) enhancers by OSK in early reprogramming coincides with somatic TF recruitment, somatic TFs may be required for the selection of these enhancers (Figure 2). During developmental cell fate decisions, selection of new enhancer elements also requires combinatorial TF action [1,21,26,28,36], confirming iPSC reprogramming as a useful model for understanding the general logic of TF-guided cell fate decisions.

TFs can collaboratively bind nucleosomal DNA in multiple ways [37]. For instance, cooperativity can arise from protein-protein interactions between them, which can be enhanced by close spacing of binding sites or allosteric interactions on DNA. Consistent with this mechanism, OCT4 and SOX2 can dimerize on DNA, and this protein-protein interaction is required for reprogramming; furthermore, an Oct-Sox composite motif with juxtaposed binding sites is highly enriched in pluripotency enhancers [6-8,10,38,39] (Figure 2A). In an alternative mode, several TFs can compete with the histone octamer

without the need of direct protein interactions, when their motifs are contained within the DNA sequence that is covered by the nucleosome [40].

Recent studies uncovered a range of binding modes for O and S at target sites within nucleosome-covered DNA, increasing the complexity of how these TFs open chromatin. Imaging studies with O and S engaging with nucleosomal DNA showed that binding of one factor often precedes the other [41,42]. The order is debated, and the presence of one TF can have synergistic and antagonistic effects on the binding of the other, depending on motif arrangement and position along the nucleosome [41,42]. Exciting structural studies revealed that O and S induce local DNA distortions and the detachment of DNA from the histone octamer to increase the accessibility to DNA [43,44]. Interestingly, O harbors two DNA binding domains, POU-S and POU-HD, of which the POU-S domain is sufficient to engage nucleosomal targets together with S [44]. Upon displacement of the histone octamer, it is thought that the POU-HD domain can engage the other half of the Oct4 motif [44]. Partial motifs recognized by the POU-S domain of OCT4 are enriched within the sequences curated for reprogramming factor binding sites that maintain nucleosomes in reprogramming cells [34], confirming this mode of action.

The maintenance of the open chromatin state by TFs is surprising given that TF occupancy at a binding site is intermittent with TFs cycling constantly on and off. Therefore, once the histone octamer is evicted, re-formation of nucleosomes may be inherently slow [36,41] or require other, active mechanisms, such as the action of ATP-dependent chromatin remodelling complexes. Indeed, reprogramming requires the OSK-

10

mediated recruitment of the BAF chromatin remodeling complex [6,45]. BAF is critical for maintaining a nucleosome-free enhancer site, reinforces the binding of OSK and promotes the removal of flanking nucleosomes to enable the binding of additional TFs nearby [46,47]. Intriguingly, O, S, and K can also bind methylated DNA and induce demethylation through passive mechanisms or the recruitment of Tet enzymes [48-50], highlighting those diverse mechanisms are exploited by the reprogramming factors to open closed chromatin.

Alternative fates in reprogramming cultures

The emergence of various cell fates is an intriguing feature of iPSC reprogramming (Figure 1). For example, endodermal genes, including those encoding the TFs Gata4 or Gata6, become upregulated in OSKM-induced mouse fibroblast reprogramming cultures, and extraembryonic endoderm stem cells (iXENs) can be obtained when supporting culture medium is added [17,20]. Depletion of these endodermal TFs reduced the number of iXEN colonies while increasing iPSC colony number, indicating that iXEN formation occurs in parallel and competes with the iPSC reprogramming branch [20] (Figure 3A). It is likely that endodermal TFs collaborate with the reprogramming factors to alter enhancer site selection and modulate the reprogramming outcome. In other studies, a small population of cells exhibiting a trophectoderm gene expression signature was detected during iPSC reprogramming of human fibroblasts [10,51]. Again, with a timely switch to appropriate culture conditions, this population can give rise to stable induced trophoblast stem cell (iTSC)-like cell lines [10,51]. However, continued reprogramming in fibroblast or iPSC medium extinguished the TSC-like identity [10,51]. Together, these results

demonstrate that the signaling cues provided by the culture medium, and their downstream TFs, ultimately permit the stabilization and propagation of alternative cell fates such as iTSCs and iXENs.

Although the rules underlying this cellular plasticity of reprogramming cells are still unknown, the levels and stoichiometry of the reprogramming factors and other TFs appear critical, which is consistent with the observation that SOX2 levels define the developmental potential of early embryonic cells [14,52]. Reprogramming experiments with a non-OSKM TF cocktail also support this idea. The TFs GATA3, EOMES, TFAP2C, cMYC, and ESRRB can reprogram three stable stem cell types from mouse fibroblasts: iPSCs, iTSCs, and iXENs [53]. The balance of these TFs is the predominant factor determining the cellular outcome, with high levels of Eomes inducing iTSC identity, and Esrrb favoring iPSC and iXEN reprogramming (Figure 3B).

Conclusions

iPSC reprogramming is a rich model for understanding the TF code underlying cell fate changes in general. The future development and application of single cell multi-omics technologies combined with new lineage recording methods will provide many opportunities to address open questions. It remains to be shown how OSK interact with somatic TFs to induce their redistribution away from somatic enhancers to pluripotency enhancers; what the role of transient enhancers and the transient gene expression program is; how transient enhancers, bound directly by both somatic TFs and OSK, are silenced to give rise to iPSCs; whether somatic TFs are critical for pluripotency enhancer

selection or, alternatively, may interfere with the full transcriptional activation of these sequences [31]. Similarly, whether comparable mechanisms for the decommissioning of the starting cell program also apply to direct reprogramming processes from one somatic cell into another, is an interesting question for the future. Finally, the derivation of human iPSCs, iXENs and iTSCs from one reprogramming culture has paved the way for the development of new cell models for the study of human embryogenesis [54], highlighting that insights gained from reprogramming studies will also be relevant for our understanding of early embryonic development.

BOX1. Additional mechanisms of somatic program inactivation

In addition to somatic TF redeployment, various other mechanisms are involved in controlling the activity of somatic enhancers. For instance, the recruitment of the histone deacetylase HDAC1 occurs specifically at OSK-bound somatic enhancers, which might shift the balance to co-factors towards repression [6]. Additionally, the co-repressor complex Sin3A is upregulated during reprogramming, required for iPSC induction, predominantly binds to promoters and contributes to the repression of critical somatic TFs [8]. In opposition, other mechanisms contribute to the maintenance of somatic enhancers and are barriers of reprogramming. A large number of somatic TFs, including AP-1, RUNX and CEBP, and the reprogramming factors, are modified by SUMO (small ubiquitin-like modifier). Since many TFs also contain SUMO-interacting domains, protein interaction networks are formed that stabilizes TF binding at somatic enhancers and secures somatic cell identity [55]. Accordingly, SUMO perturbation dramatically increases reprogramming efficiency [56,57]. Although the precise mechanism of how SUMO depletion enhances

reprogramming is still unknown, its depletion may promote the redistribution of somatic TFs. Somatic enhancers also require the continuous recruitment of chromatin modifiers to stay active. One example is that the inhibition of MLL1, a histone H3K4 methyltransferase, results in efficient reprogramming via loss of the active histone H3K4me1 enhancer mark [58]. The loss of active chromatin modifiers and chromatin remodelers due to somatic TF redistribution may further destabilize somatic enhancers [28,46]. Regardless, these findings overall show that the inactivation of the somatic program is critical for iPSC induction and have highlighted mechanisms that maintain somatic cell identity.

Figure Legends

Figure 1: Cell state transitions and global gene expression changes in iPSC reprogramming

A) Roadmap of iPSC induction from somatic cells. Upon expression of OSKM, a diminishing pool of cells transitions through sequential stages towards the iPSC state. In addition to cells stalling along the productive reprogramming path, the formation of alternative cell states explains the low efficiency of iPSC generation. The proportion of cell states at each stage is strongly influenced by the experimental reprogramming system and culture medium [10-12,15].

B) Key gene expression dynamics during OSKM-induced reprogramming. Regardless of the starting somatic cell type, three broad gene expression changes occur on the productive path to iPSCs: somatic program silencing, transient program expression, and

pluripotency program activation. Pluripotency program activation occurs gradually with the upregulation of cell cycle, biosynthesis, chromatin remodeling genes, and culminates in the activation of endogenous pluripotency-related TFs. Somatic gene repression and pluripotency gene activation, previously thought to be separated temporally, can overlap in individual cells [12]. It is still largely unclear how the expression of the transient program relates to the silencing of the somatic program and the activation of the pluripotency program.

Figure 2: Enhancer reorganization during reprogramming is linked to distinct TF binding and motif patterns

A) Key enhancer and associated TF binding changes during reprogramming. Very early in reprogramming, OSK bind a fraction of somatic enhancers as well as transient enhancers and a subset of pluripotency enhancers. At transient and early-engaged pluripotency enhancers, OSK co-bind with somatic TFs. Over time, early engaged pluripotency enhancers gain the binding of additional TFs throughout reprogramming (such as NANOG), which replaces the binding of somatic TFs. The majority of pluripotency enhancers is engaged later in the process (late-engaged) by O and S (without K) and requires additional TFs (for instance ESRRB) that are activated during the reprogramming process. In starting fibroblasts, both early and late-engaged pluripotency enhancers lack hypersensitivity (based on ATAC-seq) and reprogramming factor binding coincides with substantial nucleosome removal. Based on the presence and absence of DNA sequence motifs (as shown on the right), it is thought that OSK

engage transient and pluripotency enhancers through direct DNA binding and interact with somatic enhancers largely indirectly.

B) Somatic TF redistribution model. Early in reprogramming, OSK recruit broadly expressed somatic TFs such as AP-1, CEBP, TEAD (orange symbols) to new sites in transient and pluripotency enhancers, depleting them from somatic enhancers. Since somatic cell-specific TF occupancy at somatic enhancers depends on the presence of broadly expressed somatic TFs, the binding of somatic cell-specific TFs (yellow symbols) is also decreased in this process. The redistribution of broad somatic TFs and the loss of somatic cell-specific TFs lead to the destabilization of fibroblast enhancers and the repression of the somatic gene program. In this model, OSK inactivate somatic enhancers indirectly, without the need for direct binding to somatic enhancers.

C) Putative roles for somatic TF binding at early-engaged pluripotency enhancers. From left to right: broadly expressed somatic TFs may collaborate with OSK to remove nucleosomes if their binding sites are within one nucleosome length, and therefore be required for enhancer opening early in reprogramming; somatic TFs passively bind to DNA in regions opened by OSK; and somatic TFs indirectly bind through protein-protein interaction with OSK or co-factors. In the latter two cases, somatic TFs may not have a specific function or, alternatively, may block the activation of these enhancers.

Figure 3: Strategies for producing iPSCs, iTSCs and iXENs

A) The existence of cells expressing endodermal TFs such as GATA4 and GATA6 in OSKM-induced iPSC reprogramming cultures can be exploited to, in addition to iPSCs, derive iXENs by exposing the reprogramming culture to a culture medium that supports

iXENs [20]. Gata6 expression is required for iXEN formation. Similarly, iTSCs and iPSCs can be derived from human OSKM reprogramming cultures [10] (not shown).

B) iTSCs, iXENs and iPSCs can also be derived from a reprogramming culture upon expression of an alternative TF cocktail (GETMS), when appropriate media are supplied [53]. Whether XEN-like cells arise in parallel to or on the path to iPSCs remains unclear. The balance of EOMES and ESRRB influences which cell states are formed during reprogramming. High EOMES levels favor iTSC induction, whilst high ESRRB favors iXEN and iPSC induction.

Declaration of Interest

The authors declare no conflict of interest.

Acknowledgements

References

1. Heinz S, Romanoski CE, Benner C, Glass CK: The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol 2015, 16:144-154.

2. Spitz F, Furlong EE: Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 2012, 13:613-626.

3. Takahashi K, Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006, 126:663-676.

4. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R, et al.: Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 2007, 1:55-70.

5. Nefzger CM, Rossello FJ, Chen J, Liu X, Knaupp AS, Firas J, Paynter JM, Pflueger J, Buckberry S, Lim SM, et al.: Cell Type of Origin Dictates the Route to Pluripotency. Cell Rep 2017, 21:2649-2660.

6. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K: Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell 2017, 168:442-459 e420.

7. Knaupp AS, Buckberry S, Pflueger J, Lim SM, Ford E, Larcombe MR, Rossello FJ, de Mendoza A, Alaei S, Firas J, et al.: Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. Cell Stem Cell 2017, 21:834-845 e836.

8. Li D, Liu J, Yang X, Zhou C, Guo J, Wu C, Qin Y, Guo L, He J, Yu S, et al.: Chromatin Accessibility Dynamics during iPSC Reprogramming. Cell Stem Cell 2017, 21:819-833 e816.

9. Schwarz BA, Cetinbas M, Clement K, Walsh RM, Cheloufi S, Gu H, Langkabel J, Kamiya A, Schorle H, Meissner A, et al.: Prospective Isolation of Poised iPSC

Intermediates Reveals Principles of Cellular Reprogramming. Cell Stem Cell 2018, 23:289-305 e285.

10. Liu X, Ouyang JF, Rossello FJ, Tan JP, Davidson KC, Valdes DS, Schroder J, Sun YBY, Chen J, Knaupp AS, et al.: Reprogramming roadmap reveals route to human induced trophoblast stem cells. Nature 2020, 586:101-107.

* In this study, the authors define the molecular path of iPSC induction from human fibroblasts and demonstrate that alternative cell states can be exploited to derive stable iTSCs in addition to naïve and primed iPSCs, by exposing the reprogramming culture to medium that supports a given cell type.

11. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Gould J, Liu S, Lin S, Berube P, et al.: Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. Cell 2019, 176:928-943 e922.

12. Tran KA, Pietrzak SJ, Zaidan NZ, Siahpirani AF, McCalla SG, Zhou AS, Iyer G, Roy S, Sridharan R: Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency. Cell Rep 2019, 27:1726-1741 e1725.

13. Guo L, Lin L, Wang X, Gao M, Cao S, Mai Y, Wu F, Kuang J, Liu H, Yang J, et al.: Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell RNA-Seq. Mol Cell 2019, 73:815-829 e817.

14. Francesconi M, Di Stefano B, Berenguer C, de Andres-Aguayo L, Plana-Carmona M, Mendez-Lago M, Guillaumet-Adkins A, Rodriguez-Esteban G, Gut M, Gut IG, et al.: Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming. Elife 2019, 8.

15. Zunder ER, Lujan E, Goltsev Y, Wernig M, Nolan GP: A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell 2015, 16:323-337.

16. Chen J, Chen X, Li M, Liu X, Gao Y, Kou X, Zhao Y, Zheng W, Zhang X, Huo Y, et al.: Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. Cell Rep 2016, 14:1540-1554.

17. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A: Dissecting direct reprogramming through integrative genomic analysis. Nature 2008, 454:49-55.

18. Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K: Role of the murine reprogramming factors in the induction of pluripotency. Cell 2009, 136:364-377.

19. Xing QR, El Farran CA, Gautam P, Chuah YS, Warrier T, Toh CXD, Kang NY, Sugii S, Chang YT, Xu J, et al.: Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. Sci Adv 2020, 6.

20. Parenti A, Halbisen MA, Wang K, Latham K, Ralston A: OSKM Induce Extraembryonic Endoderm Stem Cells in Parallel to Induced Pluripotent Stem Cells. Stem Cell Reports 2016, 6:447-455.

21. Heinz S, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK: Effect of natural genetic variation on enhancer selection and function. Nature 2013, 503:487-492.

22. Long HK, Prescott SL, Wysocka J: Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 2016, 167:1170-1187.

23. Zviran A, Mor N, Rais Y, Gingold H, Peles S, Chomsky E, Viukov S, Buenrostro JD, Scognamiglio R, Weinberger L, et al.: Deterministic Somatic Cell Reprogramming Involves Continuous Transcriptional Changes Governed by Myc and Epigenetic-Driven Modules. Cell Stem Cell 2019, 24:328-341 e329.

24. Soufi A, Donahue G, Zaret KS: Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 2012, 151:994-1004.

25. Liu J, Han Q, Peng T, Peng M, Wei B, Li D, Wang X, Yu S, Yang J, Cao S, et al.: The oncogene c-Jun impedes somatic cell reprogramming. Nat Cell Biol 2015, 17:856-867.

26. Hosokawa H, Ungerback J, Wang X, Matsumoto M, Nakayama KI, Cohen SM, Tanaka T, Rothenberg EV: Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. Immunity 2018, 48:1119-1134 e1117.

27. Eferl R, Wagner EF: AP-1: a double-edged sword in tumorigenesis. Nat Rev Cancer 2003, 3:859-868.

28. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, Roberts CWM, Greenberg ME: AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol Cell 2017, 68:1067-1082 e1012.

29. Hogan NT, Whalen MB, Stolze LK, Hadeli NK, Lam MT, Springstead JR, Glass CK, Romanoski CE: Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. Elife 2017, 6.

30. Li QV, Dixon G, Verma N, Rosen BP, Gordillo M, Luo R, Xu C, Wang Q, Soh CL, Yang D, et al.: Genome-scale screens identify JNK-JUN signaling as a barrier for pluripotency exit and endoderm differentiation. Nat Genet 2019, 51:999-1010.

31. Markov GJ, Mai T, Nair S, Shcherbina A, Wang YX, Burns DM, Kundaje A, Blau HM: AP-1 is a temporally regulated dual gatekeeper of reprogramming to pluripotency. Proc Natl Acad Sci U S A 2021, 118.

**   Exploring the heterokaryon system to define reprogramming mechanisms, the study shows the requirement of AP-1 for the control of somatic enhancers, and reveals that AP-1 also functions as repressor at the distal enhancer Oct4 in a phosphorylation-dependent manner.

32. Cheloufi S, Elling U, Hopfgartner B, Jung YL, Murn J, Ninova M, Hubmann M, Badeaux AI, Euong Ang C, Tenen D, et al.: The histone chaperone CAF-1 safeguards somatic cell identity. Nature 2015, 528:218-224.

33. Zaret KS, Carroll JS: Pioneer transcription factors: establishing competence for gene expression. Genes Dev 2011, 25:2227-2241.

34. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS: Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 2015, 161:555-568.

35. Adachi K, Kopp W, Wu G, Heising S, Greber B, Stehling M, Arauzo-Bravo MJ, Boerno ST, Timmermann B, Vingron M, et al.: Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. Cell Stem Cell 2018, 23:900-904.

36. Sonmezer C, Kleinendorst R, Imanci D, Barzaghi G, Villacorta L, Schubeler D, Benes V, Molina N, Krebs AR: Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. Mol Cell 2021, 81:255-267 e256.

37. Morgunova E, Taipale J: Structural perspective of cooperative transcription factor binding. Curr Opin Struct Biol 2017, 47:1-8.

38. Malik V, Glaser LV, Zimmer D, Velychko S, Weng M, Holzner M, Arend M, Chen Y, Srivastava Y, Veerapandian V, et al.: Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. Nat Commun 2019, 10:3477.

39. Tapia N, MacCarthy C, Esch D, Gabriele Marthaler A, Tiemann U, Arauzo-Bravo MJ, Jauch R, Cojocaru V, Scholer HR: Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. Sci Rep 2015, 5:13533.

40. Miller JA, Widom J: Collaborative competition mechanism for gene activation in vivo. Mol Cell Biol 2003, 23:1623-1632.

41. Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, et al.: Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell 2014, 156:1274-1285.

42. Li S, Zheng EB, Zhao L, Liu S: Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. Cell Rep 2019, 28:2689-2703 e2684.

43. Dodonova SO, Zhu F, Dienemann C, Taipale J, Cramer P: Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature 2020, 580:669-672.

44. Michael AK, Grand RS, Isbel L, Cavadini S, Kozicka Z, Kempf G, Bunker RD, Schenk AD, Graff-Meyer A, Pathare GR, et al.: Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science 2020, 368:1460-1465.

* In this study, the authors use a genomics readout to map all OS binding combinations on nucleosomal DNA in combination with stuctural studies to show how OS access sites in closed chromatin.

45. Chen K, Long Q, Xing G, Wang T, Wu Y, Li L, Qi J, Zhou Y, Ma B, Scholer HR, et al.: Heterochromatin loosening by the Oct4 linker region facilitates Klf4 binding and iPSC reprogramming. EMBO J 2020, 39:e99165.

46. Iurlaro M, Stadler MB, Masoni F, Jagani Z, Galli GG, Schubeler D: Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat Genet 2021, 53:279-287.

* Studying OCT4 binding, the authors demonstrate that chromatin remodelling complexes are continously needed to preserve and restore chromatin accessibility and enable rapid TF cycling.

47. King HW, Klose RJ: The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife 2017, 6.

48. Vanzan L, Soldati H, Ythier V, Anand S, Braun SMG, Francis N, Murr R: High throughput screening identifies SOX2 as a super pioneer factor that inhibits DNA methylation maintenance at its binding sites. Nat Commun 2021, 12:3337.

49. Sardina JL, Collombet S, Tian TV, Gomez A, Di Stefano B, Berenguer C, Brumbaugh J, Stadhouders R, Segura-Morales C, Gut M, et al.: Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell 2018, 23:727-741 e729.

50. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, et al.: Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 2017, 356.

51. Castel G, Meistermann D, Bretin B, Firmin J, Blin J, Loubersac S, Bruneau A, Chevolleau S, Kilens S, Chariau C, et al.: Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells. Cell Rep 2020, 33:108419.

52. Tremble KC, Stirparo GG, Bates LE, Maskalenka K, Stuart HT, Jones K, Andersson-Rolf A, Radzisheuskaya A, Koo BK, Bertone P, et al.: Sox2 modulation increases naive pluripotency plasticity. iScience 2021, 24:102153.

53. Benchetrit H, Jaber M, Zayat V, Sebban S, Pushett A, Makedonski K, Zakheim Z, Radwan A, Maoz N, Lasry R, et al.: Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts. Cell Stem Cell 2019, 24:983-994 e987.

*   This study shows that a non-OSKM reprogramming factor cocktail can induce iPSCs, iTSCs and iXENs from mouse fibroblasts and that the levels of the TFs EOMES and ESRRB determines which cell state is formed.

54. Liu X, Tan JP, Schroder J, Aberkane A, Ouyang JF, Mohenska M, Lim SM, Sun YBY, Chen J, Sun G, et al.: Modelling human blastocysts by reprogramming fibroblasts into iBlastoids. Nature 2021, 591:627-632.

**   This study exploits iPSCs and iTSCs derived from human fibroblasts upon OSKM expression to form blastoids that resemble the human blastocyst and will enable to study of human embryogenesis at scale.

55. Theurillat I, Hendriks IA, Cossec JC, Andrieux A, Nielsen ML, Dejean A: Extensive SUMO Modification of Repressive Chromatin Factors Distinguishes Pluripotent from Somatic Cells. Cell Rep 2020, 32:108146.

56. Cossec JC, Theurillat I, Chica C, Bua Aguin S, Gaume X, Andrieux A, Iturbide A, Jouvion G, Li H, Bossis G, et al.: SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. Cell Stem Cell 2018, 23:742-757 e748.

57. Borkent M, Bennett BD, Lackford B, Bar-Nur O, Brumbaugh J, Wang L, Du Y, Fargo DC, Apostolou E, Cheloufi S, et al.: A Serial shRNA Screen for Roadblocks to Reprogramming Identifies the Protein Modifier SUMO2. Stem Cell Reports 2016, 6:704-716.

58. Zhang H, Gayen S, Xiong J, Zhou B, Shanmugam AK, Sun Y, Karatas H, Liu L, Rao RC, Wang S, et al.: MLL1 Inhibition Reprograms Epiblast Stem Cells to Naive Pluripotency. Cell Stem Cell 2016, 18:481-494.

Figure 1

A

Starting cell (fibroblast)

productive path

iPSC

O
K
S
M

alternative fates

stromal

trophectoderm -like

endoderm -like

neural

B

somatic gene expression

transient gene expression

pluripotency- related gene expression

relative expression level

fibroblast → productive reprogramming path → iPSC

28

Figure 2

A

fibroblasts | early intermediate | iPSCs

enriched motifs
somatic TFs · OSK · pluripotency TFs

somatic enhancers

AP1 Runx Tead Cebp NF1 Ets Hox

weak Oct4 Sox2 Klf4 motifs at OSK bound sites

transient enhancers

AP1 Cebp Tead

Oct4 Sox2 Klf4 Oct-Sox

pluripotency enhancers

early-engaged

Tead AP1 Cebp

Oct4 Sox2 Klf4 Oct-Sox

late-engaged

Oct4 Sox2 Oct-Sox

Esrrb, Tcfcp2l1

● somatic TFs   ● pluripotency TFs

B

somatic enhancers   transient and early-engaged pluripotency enhancers

fibroblasts

early stage

● broadly expressed somatic TFs (AP-1, CEBP etc)

● cell-type specific somatic TFs

destabilization   opening

C

competitve nucleosome removal by OSK and somatic TFs | secondary binding of somatic TFs to chromatin opened by OSK | indirect binding of somatic TFs | early-engaged pluripotency (early reprogramming stage)

● broadly expressed somatic TFs (AP-1, CEBP etc)

Figure 3

A

O K S M → fibroblast

no GATA6 → iPSCs

↑ GATA6 → iXENs

cell culture medium for cell fate stabilization and expansion

B

G E T M S → fibroblast

G GATA3
E EOMES
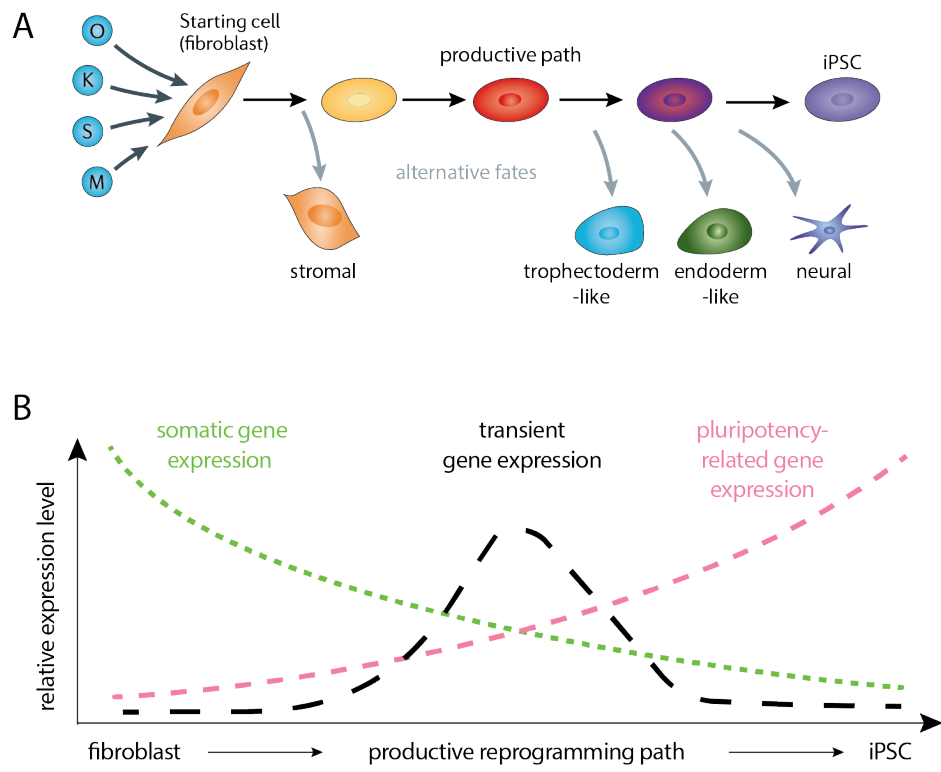T TFAP2C
M cMYC
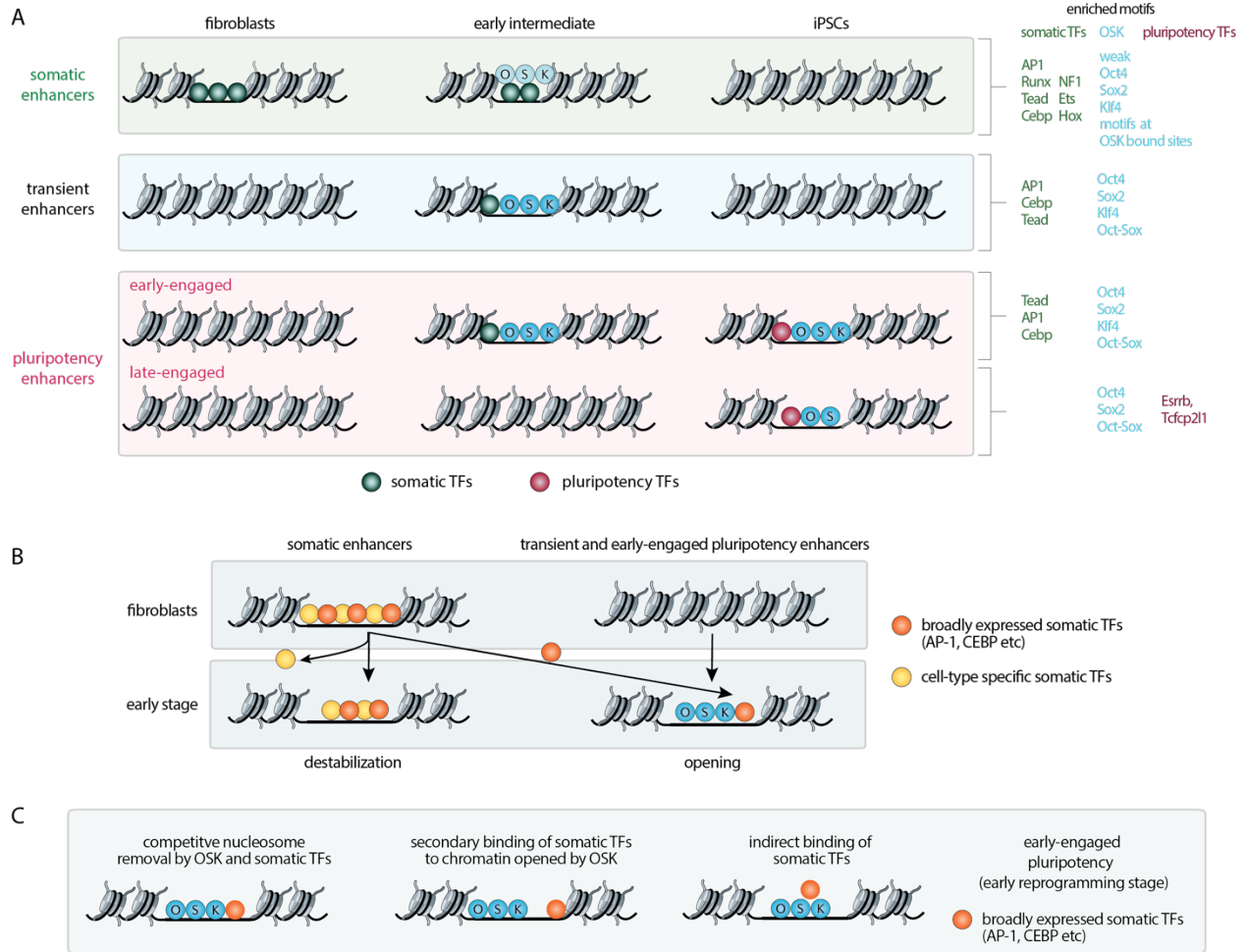S ESRRB

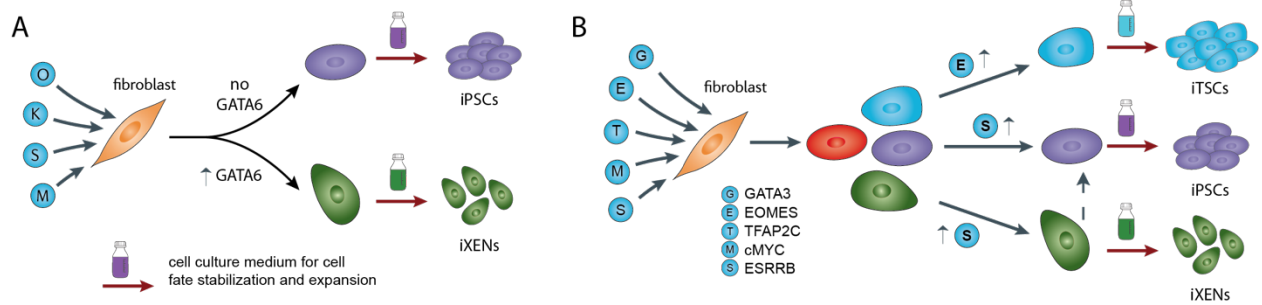E ↑ → iTSCs

S ↑ → iPSCs

S ↑ → iXENs

29

Chapter 2

Essential Reprogramming Factors Redistribute Somatic Transcription Factors in

Early Reprogramming through Multiple Molecular Mechanisms

Essential Reprogramming Factors Redistribute Somatic Transcription Factors in Early Reprogramming through Multiple Molecular Mechanisms

Weixian Deng*, Elsie Jacobson*, Jihui Sha, Constatinos Chronis, Amanda Collier, James Wohlschlegel[#], Kathrin Plath[#]

*: contribute equally to the work

#: co-corresponding author

Abstract

Ectopically expressed Oct4, Sox2, Klf4 and c-Myc (OSKM), also known as Yamanaka factors, can reprogram terminally differentiated cells to pluripotency. Among them, Oct4, Sox2 and Klf4 are indispensable and mainly bind to distal regulatory regions. The cooperative binding of them leads to opening of somatically closed chromatin sites in early stage of reprogramming, and concomitantly, some somatic TFs are observed to switch their binding sites from mouse embryonic fibroblast (MEF) specific enhancers or MEs, to OSK bound, newly opened enhancer regions in the early stage of reprogramming. In this work, we confirmed the co-binding of somatic TFs with O, S or K through several orthogonal approaches, identified somatic TFs co-bind with OSK through both direct protein-protein interaction (PPI) and other indirect manners. And lastly, we found the removal of active enhancer histone mark on MEs is not led by the binding of Oct4.

Introduction

Uncovering the mechanisms underlying cell fate changes is critical for understanding how cell identities are established and maintained and how they are destabilized or altered during disease processes. I am studying these mechanisms in the context of the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs), which changed

31

the way of how human diseases can be studied in the laboratory and is paving the way to personalized cell replacement therapies. The ectopic expression of the transcription factors (TFs) Oct4, Sox2, Klf4, and c-Myc (OSKM) in somatic cells leads to the induction of iPSCs[1,2]. iPSC reprogramming takes several weeks and occurs at low frequencies, indicating that the reprogramming factors need to overcome barriers established in somatic cells to preserve their identity. Thus, a mechanistic understanding of how OSKM induce the transcriptional changes that lead to pluripotency will reveal strategies to overcome these barriers and yield fundamental insights into basic principles by which cell identity can be manipulated. The Plath lab[3] previously found that the collaborative binding of OSK is essential for the selection and activation of pluripotency enhancers (PEs). Intriguingly, despite being transcriptional activators, OSK also have been suggested to mediate the silencing of the enhancers that are active in the starting somatic cell type, such as fibroblasts. The silencing of somatic enhancers correlates with the redistribution of somatic TFs to sites that are newly opened and bound by OSK early in reprogramming. Thus, OSK may achieve the decommissioning of somatic enhancers by redistributing somatic regulators, which represents a new paradigm of cell fate control. Yet, it remained unknown if OSK and somatic TFs indeed co-occupy these new sites. To address this question, I optimized and applied the ChIP-SICAP (chromatin immunoprecipitation combined with selective isolation of chromatin-associated proteins) method. I found TFs of the AP-1 family on the same chromatin fragment as O, S, and K, including the family members Jun, JunB and JunD, as well as other somatic TFs, such as Cebpa/b, and Runx2. Moreover, intriguingly, these TFs are expressed in most somatic/differentiated cell types. Since it is thought that all somatic cells can be

reprogrammed by OSK, the biochemical interaction of OSK with these TFs may underlie the ability to be universal reprogramming factors across species and cell types. The key question of this work is to understand how OSKM mechanistically induce this somatic TF redistribution. We hypothesized that protein-protein interactions (PPIs) of OSKM with somatic TFs are critical for this process. To test this idea, we leveraged dual-genetic background MEF cell line to test OSK binding dependency to somatic TFs *in vivo* to shorten the target list and followed by PAQMAN and *in vitro* pull-down assay to further unveil the interaction types between OSK and somatic TFs in the presence and absence of DNA/chromatin.

Lastly, to assay if MEs are decommissioned by somatic TF redistribution or proactive recruitment of Hdac1 by OSK, we dissected the Oct4's function by testing whether Oct4 nucleosome binding mutant can remove H3K27ac from MEs through binding to MEs without opening new sites.

Results

ChIP-SICAP confirms the co-binding of somatic TFs and O/S/K

From previous study[3], Chronis et. al. has demonstrated that the overexpression of OSKM in the early stage of reprogramming causes somatic TFs redistributing from their original binding sites mainly on MEs to OSK occupied newly opened transient enhancers (TEs) and pluripotency enhancers (PEs), and OSK's binding sites are not restricted to newly opened sites but also found on MEs despite the low density of their binding motifs. These observations were obtained through chromatin immune-precipitation-Sequencing (ChIP-seq) techniques. Yet, the how OSK make somatic TFs' redistribution is not unveiled.

Due to the low motif density of some somatic TFs including AP-1, CEBP, TEAD and RUNX family TFs on newly gained sites at 48hrs post-induction of OSKM expression and low motif density of OSK on MEs, we hypothesized that protein-protein interaction between somatic TFs and OSK may contribute to both somatic TFs' binding to newly gained sites and OSK's binding to MEs. To test this hypothesis, we first must confirm whether OSK and somatic TFs co-colocalized on the chromatin at the protein level. To confirm their co-existence on chromatin, we employed a proteomics approach named chromatin immnunoprecipitation- selective isolation of chromatin associated proteins (ChIP-SICAP)[4], through which we are able to find co-binding somatic TFs to OSK respectively.

As shown in Figure 1A, we isolated MEF cells from tetO-OSKM stemcca mouse embryos and added doxycycline to induce the expression of OSKM for 48hrs, then crosslinked and harvested cells for ChIP-SICAP. Since streptavidin heavily contaminates bottom-up proteomics samples due to its excessive amount of signal masking peptide-of-interest's signal in the mass spectrometer (MS), we optimized the published workflow by introducing nuclease elution (Figure 1B) to maximize the detection to TFs which are of low abundance in nature.

Next, we performed ChIP-SICAP using O S and K antibodies respectively, and comparing to IgG negative control, we identified 1288, 808 and 1326 proteins co-bind with O S and K on the chromatin respectively. In accordance with Chronis et.al. ChIP-seq results, O S and K showed highly similar co-binding proteome (Figure 1C). Among all the OSK co-binding proteins, there are 130 TFs, 74 of them showed decrease to slight increase RNA expression (ES:MEFs<=1.5) during the reprogramming process from MEFs to iPSCs, and

59 of them showed only decrease trend (ES:MEFs<=1). In the 59 somatic TFs, we found

7 TFs belong to basic leucine zipper factors including Jun JunB JunD, Fosl2, Cebpβ,

Cebpδ, 16 TFs from homeo domain factors including Hox-A13, 6 from C2H2 zinc finger

factors including Klf4 and Mecom, and many others e.g. RUNX1 and 2 from Runt domain

factors (Figure 1E). And according to STRING database[5], many of these identified

proteins also have been reported having PPIs to each other (Figure 1F). Other than TFs,

we have also identified chromatin remodeler complexes e.g. NurD, SWI/SNF, LSD1 and

PRC2 and so on. In this section, we confirmed that O S and K co-bind to somatic TFs as

well as chromatin remodelers on the chromatin suggesting there could be both direct and

indirect PPIs between O S and K with somatic TFs.

OSK have somatic TF motif binding dependency in early stage of reprogramming

Moving forward, to test if O S and K bind to chromatin through somatic TFs *in vivo*, we

decided to establish a dual-genetic background tetO-OSK MEFs system. The dual-

genetic background system has previously reported to be used for characterizing

cooperative binding of TFs[6–8]. Having two sets of chromosomes from different genetic

background introduces many natural single nucleotide polymorphism events (SNPs),

these SNPs disrupt or weaken TF binding motifs *in vivo*, therefore, we can leverage these

natural mutations to assess the effects of the loss of some TFs' binding.

We crossed wild-type (WT) PWK/PhJ mouse with homozygotic tetO-OSK in C57BL/6J

background, therefore all the F1 offspring have tetO-OSK and each one of the two genetic

backgrounds (Figure 2A). We took the isolated MEFs from F1 embryos and performed

CUT&RUN and CUT&Tag to measure the binding of O S and K and characterized

whether O, S or K lose their binding due to the disruption of any somatic TF's motif.

Additionally, we performed ATAC-seq and H3K27ac CUT&Tag assayed the chromatin status pre and post OSK induction. Moreover, since O S and K are all identified as pioneer factors[9,10], if O S and K's binding upon a chromatin locus do not depend on any somatic TF, the chromatin accessibility should presumably be maintained by O S and K.

First, we assessed motif scores of HOMER identified motifs in both alleles under all the O S and K binding peaks as long as at least a peak is identified from one of the two alleles. We found weakened and disrupted motifs tend to have a SNP at +1 position to the motif center suggesting that position is essential to maintain motif's identity (Figure 2B).

Next, we identified all the binding peaks of O S and K on both alleles and found Oct4 has 22%, Sox2 has 24%, Klf4 has 23.5% peaks skewed to one of the two alleles with majority of peaks staying bi-allelic binding (Figure 2C).

Furthermore, we examined what disrupted or weakened motifs are enriched in the skewed binding peaks to the other allele and calculated how much more binding signals are observed in the corresponding unchanged motif (Figure 2D). In the Oct4 binding assay dataset, we identified AP-1, Klf4, CEBP and Sox2 motifs' disruption are enriched in the skewed peaks to the other allele in different chromatin regulatory regions and the binding signal fold change are all above 2 folds. Similarly, in Sox2, we identified Sox2 its own motif as well as AP-1 and CEBP, suggesting that Sox2's binding relies on not only its own motif but AP-1 and CEBP. Lastly, Klf4 showed binding dependency to its own motif along with AP-1, TEAD and Oct4. And the signal change magnitudes are all more than 2 folds. These results suggesting, O S and K have binding dependency to motifs of each other and also motifs of CEBP, AP-1 and TEAD *in vivo*.

Oct4 shapes the binding chromatin status in the early stage of reprogramming

Next, we assessed the chromatin accessibility and active histone mark H3K27ac skewness correlation with Oct4 binding skewness (Figure 3A). Since Oct4 is not expressed in 0hr, the number of shared peaks between ATAC/H3K27ac at 0hr and Oct4 at 48hr are much less than ATAC/H3K27ac at 48hr, more importantly, the correlation of shared peaks between ATAC/H3K27ac at 48hrs and Oct4 at 48hrs is much than them in MEFs, this is suggesting that Oct4 has strong impact on its binding sites chromatin status.

Oct4 binds to AP-1 motif without changing AP-1 proteins binding affinity to AP-1 motif

To assess if Oct4 can bind to somatic TF, specifically, the most influencing somatic TF family in reprogramming [11,12], AP-1 motif without its own motif in proximity, and in the meantime to measure if Oct4 influences AP-1 proteins binding to their own motif. We deployed Protein-nucleic acid affinity quantification by MAss spectrometry (PAQMAN)[13] to measure nuclear proteins' binding affinity to DNA sequence of interest regardless of chromatin status.

We immobilized biotinylated double stranded DNA probes containing AP-1 motif or negative control on streptavidin beads, and isolated nuclear extract from MEFs or stemcca MEFs +dox for 48hrs, then performed PAQMAN assay. As shown in Figure 3B, five of AP-1 family proteins are identified from both MEFs and stemcca MEFs 48hr post induction nuclear extract forming sigmoid curves as a function of increasing concentrations of the AP-1 motif probe (Figure 3B). Interestingly, even with this pure AP-1 motif DNA sequence in the absence of Oct4 motif, Oct4 showed sigmoid binding curve to it but not the negative control DNA (Figure 3C). Then we checked AP-1 proteins' binding curves to the AP-1 motif in the presence and absence of OSKM and found the

Kd$_{apparent}$ stay unchanged. Taken together, in this semi- in vitro system, Oct4 binds to the artificial AP-1 motif containing DNA sequence without negatively affecting AP-1 proteins binding affinity to the same probe, it is suggesting Oct4 binds to AP-1 motif through binding to AP-1 proteins.

In the future, we will assess more complicated situations by using genomic sequences extracted from sequencing data. It will include Oct4 Cebpb co-bound peak sequences consisting following scenarios: i) peaks only contain Oct4 motif; ii) peaks contain both Oct4 and CEBP motifs, we will test if Cebp family proteins can bind to these genomic sequences in the presence and absence of OSKM and whether the presence of OSKM binding can influence CEBP proteins' binding affinity to their own motif.

Oct4 has direct interaction with somatic TFs

After assessing OSK's PPIs with somatic TFs in various DNA co-existing systems, we want to confirm if there's DNA-free, direct PPIs between OSK and somatic TFs. Therefore, we used recombinant proteins of OSK, c-Jun, Fosl1, Tead1, Tead3, Cebpa, Cebpb purified from non-mammalian systems to perform *in vitro* pull-down assay. From current result, Oct4 as bait protein, as reported by others, has direct PPIs with Sox2 and Klf4, and interestingly, we found it also has direct PPIs with c-Jun, Cebp and b, but not Tead3 and Fosl1 (Figure 4A). We concluded that Oct4 has physical PPIs with somatic TFs c-Jun Cebpa/b but not with any Tead proteins. In accordance with this result, our Tead CUT&RUN only showed motif dependency to its own motif but not any of O S and K (Figure 4B).

We will further assess S and K's interaction with somatic TFs in the future with *in vitro* pull-down assay.

Oct4's binding cannot remove H3K27ac from MEs

Previously, Chronis et.al[3] has shown that Klf4 and Oct4-bound sites on MEs showed significantly higher Hdac1 binding signal when they are individually expressed in MEFs. We hypothesized that the recruitment of Hdac1 to MEs by Oct4 and Klf4 is the mechanism of ME decommissioning, the alternative hypothesis is that Oct4 and Klf4 open new sites and recruit somatic TFs away from MEs and leads to the loss of H3K27ac. To test this hypothesis, we started with Oct4. Recently, a Oct4 mutant Oct4^{del79}, was identified to retain DNA binding but lose nucleosome binding ability by deleting five amino acids in the linker region between the two Oct4 DNA binding domains[14]. We decided to leverage its incapability of opening new sites to assay if the overexpression of this Oct4^{del79} can also lead to loss of H3K27ac.

We first test if regular anti-Oct4 recognizing WT Oct4 could still detect Oct4^{del79} , we retrovirally infected MEFs with WT Oct4 or Oct4^{del79} and immunofluorescence staining showed that Oct4^{del79} is equally detectable by using this antibody (Figure 4C). Next, we performed ATAC-seq, H3K27ac and Oct4 CUT&RUN in WT or mutant expressed MEFs. As shown in Figure 4D, Oct4^{del79} failed to bind, open and establish enhancer activity on some sites as expected.

Finally, we did meta-analysis for H3K27ac signal on MEs with the expression of WT and mutant form of Oct4 (Figure 4E), we found with the expression of Oct4^{del79} , the H3K27ac signal on MEs remains at the same level as it is in MEFs which is higher than the one with WT form Oct4 expressed MEFs.

Taken together, we concluded that Oct4's binding doesn't remove H3K27ac from MEs, but it is also suggesting the depletion of the active histone marks are likely caused by Oct4's ability to open new sites.

Discussion

In this study, we first established O S and K's co-binding proteome network through optimized ChIP-SICAP techniques in the early reprogramming stage MEFs. And through this approach, we confirmed that O S and K co-exist with many somatic TFs including bZIP TFs e.g. AP-1, CEBP, Homeo domain TFs e.g. HOX-A13 etc. A lot of these somatic TFs also interact with each other. However, these are not yet clear which one's co-bind with OSK through direct or indirect PPIs. By leveraging dual-genetic background system, we established OSK binding dependency upon somatic TFs also including AP-1, CEBP and TEAD and therefore shortlisted the protein candidates which may have PPIs with OSK *in vivo*. Furthermore, we performed PAQMAN assay for AP-1 motifs in the presence or absence of OSKM and confirmed that Oct4 binds to AP-1 motifs without the co-existence of its own motif and doesn't affect AP-1 proteins' binding affinity to the DNA. This is suggesting Oct4 binds to AP-1 motif containing DNA *in vitro* likely through an indirect manner. Next, we performed *in vitro* pull-down assay and demonstrated that Oct4 has direct PPIs with c-Jun, Cebpa and b but not Tead proteins and Fosl1, and it showed that Oct4 can physically interact with some somatic TFs in the absence of DNA. Lastly, we tested that Oct4's binding is not enough to remove H3K27ac and suggested that the opening of new sites and recruitment of somatic TFs to newly opened sites by pioneer factors like Oct4 could be the main mechanism of somatic enhancer decommissioning.

With current data, we haven't been able to demonstrate that if OSK's binding has synergistic effect to somatic TFs to their motifs, and we haven't shown if the redistribution of somatic TFs through PPIs affects the progress of reprogramming. In the future work, we will perform PAQMAN assays with genomic sequences as probes to find whether OSK assists somatic TFs binding to newly opened sites when OSK's motifs are sitting adjacently. And we will also screen for Oct4 PPI dysfunction mutant to evaluate if it has any influence on reprogramming progress rate or efficiency.

Methods

Cell culture

The following cell line was used for the ChIP-SICAP of the reprogramming process at discrete stages: primary MEFs harboring a heterozygous R26-M2rtTA allele and a single dox-inducible polycistronic cassette coding for OSKM in the Col1A locus (tetO-OSKM)[15], derived from day 13.5 embryos of timed mouse pregnancies.

Immunofluorescence staining

Coverslips were rinsed twice with 1x PBS, and the cells were fixed in 2% PFA (Electron Microscopy Sciences) in 1x PBS for 10 min. After fixation, cells were permeabilized with 0.5% Triton X-100 (Acros) in 1x PBS for 10-15 min, before transferring to blocking buffer solution (1x PBST containing 2% bovine serum albumin (BSA) and 0.5% fish skin gelatin) for 1 hour. Cells were then incubated with primary antibody for 45 min, washed 3 times with 1x PBST, followed by incubating with secondary antibody for 30 min and washing 3 times with 1x PBST. Both antibody incubation steps were performed in a dark humidified chamber. Finally, cells were post-fixed with 4% PFA in 1x PBS for 10 min, stained with 1:10000 dilution of DAPI (0.5mg/ml) and mounted with Vectashield (Vector Labs) on glass

slides, before sealing with Biotium Covergrip coverslip sealant (ThermoFisher). All procedures were performed at RT, and used reagents were disposed according to standard laboratory safety practices.

CUT&Tag

CUT&Tag experiments were performed as described[16]. Briefly, for each sample, 100K cells were harvested, washed, and immobilized on concanavalin-A bead. Primary antibodies were diluted in antibody buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine, 0.05% Digitonin, 2mM EDTA and 0.1% BSA) at 1:100 ratio, and then incubated with cells for 2 hrs. Next, secondary antibodies were incubated with cells in Dig-wash buffer (20mM HEPES, 150mM NaCl, 0.5mM Spermidine, 0.05% Digitonin) at 1:100 ratio for 1 hr and washed by Dig-wash buffer. pA-Tn5 was then diluted with Dig-300 wash buffer (20mM HEPES, 300mM NaCl, 0.5mM Spermidine, 0.01% Digitonin) at 1:250 ratio and incubated with cells for 1 hr. After washing cells by Dig300 wash buffer, Tagmentation reaction was activated by adding Tagmentation buffer (20mM HEPES, 300mM NaCl, 0.5mM Spermidine, 0.01% Digitonin, 10mM MgCl2) to cells and incubated at 37C for 1 hr. Tagmentation reaction was terminated by adding EDTA to 16.7mM. Lastly, genomic DNA was extracted and proceeded to library amplification.

Figure legend

Figure 1: (A) Scheme of treatment for cells used in ChIP-SICAP experiment; **(B)**Scheme of optimized ChIP-SICAP workflow; **(C)** Proteins identified significantly enriched compared to IgG control in O/S/K ChIP-SICAP experiment; **(D)** Circle from outside to inside: total TF identified, TF with expression level (ES:MEF<1.5), TF with expression

level (ES:MEF<1); **(E)** Top 3 of most identified TF sub-families and representative motifs; **(F)** Known PPIs between identified TFs (ES:MEF<1.5) per SRING db.

Figure 2: (A) Scheme of dual-genetic background experiment design; **(B)** SNP density distribution on disrupted, stable or weakened motif sequences; **(C)** Peaks identified on different alleles from O/S/K CUT&RUN, CUT&Tag combined data; **(D)** Enriched motif in the peaks of disruption of the motif on one allele and gain binding on the other allele.

Figure 3: (A) PWK allele frequency in peaks shared between 0hr or 48hr post induction of OSK expression ATAC/H3K27ac and 48hr Oct4 CUT&RUN data; **(B)** AP-1 family proteins fraction bound change as a function of probe oligo concentration using nuclear extract from MEFs; **(C)** Oct4 fraction bound change as a function of probe oligo concentration using nuclear extract from 48hr post induction of OSKM MEFs; **(D)** Jun fraction bound change as a function of probe oligo concentration with nuclear extract pre and post induction of OSKM expression MEFs; **(E)** Cebpa and Cebpb binding sites in MEFs and 48hr post induction of OSKM expression, percentage represents proportion of sites also occupied by O/S/K; **(F)** Oct4 and Cebpb co-bound peaks' peak summit distance distribution.

Figure 4: (A) Oct4 bait *in vitro* pull-down assay blots; **(B)** Enriched motif in the peaks of disruption of the motif on one allele and gain binding on the other allele; **(C)** α-Oct4 immunofluorescence staining in WT or del79 mutant forms of Oct4 retro-virus infected MEF cells, 48hr post infection; **(D)** Genome browser view of WT Oct4 and Oct4[del79] ATAC-seq, Oct4 and H3K27ac CUT&RUN at Pou5f1 locus; **(E)** H3K27ac signal profiles on MEs in MEFs, WT Oct4 and Oct4[del79] expressed MEFs; **(F)** Schematic model of somatic TF redistribution and ME decommissioning mechanism.

Reference

1. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors**. *Cell* 2006, **126**:663–676.

2. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors**. *Cell* 2007, **131**:861–872.

3. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K: **Cooperative Binding of Transcription Factors Orchestrates Reprogramming**. *Cell* 2017, **168**:442-459 e20.

4. Rafiee MR, Girardot C, Sigismondo G, Krijgsveld J: **Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins**. *Mol Cell* 2016, **64**:624–635.

5. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al.: **The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.** *Nucleic Acids Res* 2021, **49**:D605–D612.

6. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, Roberts CWM, Greenberg ME: **AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection**. *Mol Cell* 2017, **68**:1067-1082.e12.

7. Link VM, Duttke SH, Chun HB, Holtman IR, Westin E, Hoeksema MA, Abe Y, Skola D, Romanoski CE, Tao J, et al.: **Analysis of Genetically Diverse Macrophages Reveals Local and Domain-wide Mechanisms that Control Transcription Factor Binding and Function**. *Cell* 2018, **173**:1796-1809.e17.

8. Link VM, Romanoski CE, Metzler D, Glass CK: **MMARGE: Motif mutation analysis for regulatory genomic elements**. *Nucleic Acids Res* 2018, **46**:7006–7021.

9. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS: **Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming**. *Cell* 2015, **161**:555–568.

10. Zaret KS: **Pioneering the chromatin landscape**. *Nat Genet* 2018, **50**:167–169.

11. Eferl R, Wagner EF: **AP-1: A double-edged sword in tumorigenesis**. *Nat Rev Cancer* 2003, **3**:859–868.

12. Liu J, Han Q, Peng T, Peng M, Wei B, Li D, Wang X, Yu S, Yang J, Cao S, et al.: **The oncogene c-Jun impedes somatic cell reprogramming**. *Nat Cell Biol* 2015, **17**:856–867.

13. Makowski MM, Gräwe C, Foster BM, Nguyen NV, Bartke T, Vermeulen M: **Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry**. *Nat Commun* 2018, **9**:1653.

14. Roberts GA, Ozkan B, Gachulincová I, O'Dwyer MR, Hall-Ponsele E, Saxena M, Robinson PJ, Soufi A: **Dissecting OCT4 defines the role of nucleosome binding in pluripotency**. *Nat Cell Biol* 2021, **23**:834–845.

15. Ho R, Papp B, Hoffman JA, Merrill BJ, Plath K: **Stage-Specific Regulation of Reprogramming to Induced Pluripotent Stem Cells by Wnt Signaling and T Cell Factor Proteins**. *Cell Reports* 2013, **3**:2113–2126.

16. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S: **Efficient low-cost chromatin profiling with CUT&Tag**. *Nat Protoc* 2020, **15**:3264–3283.
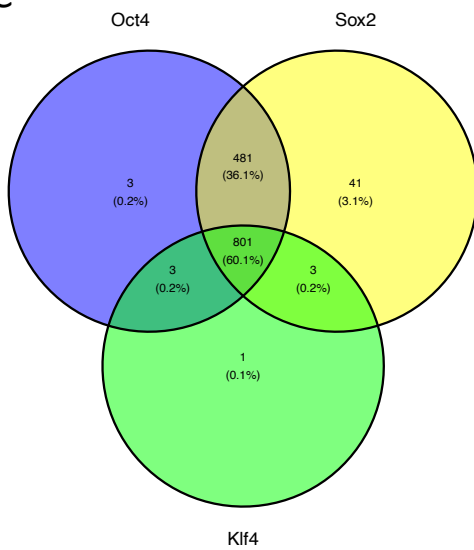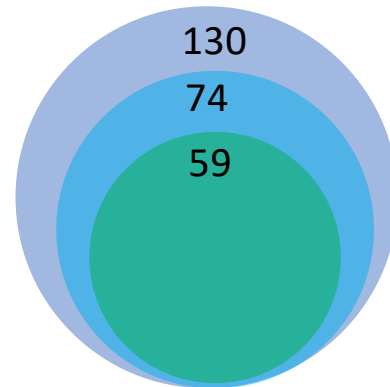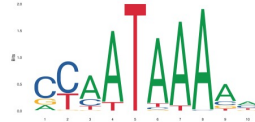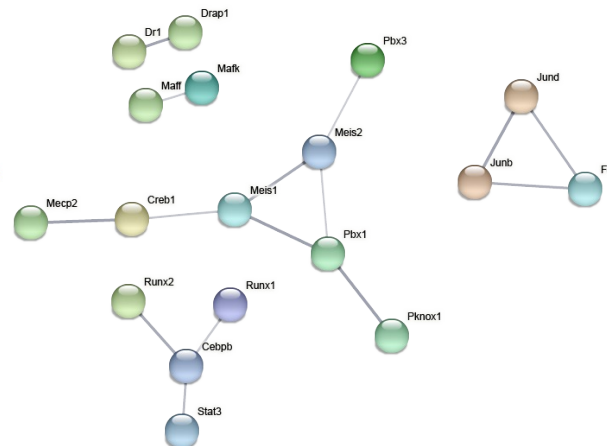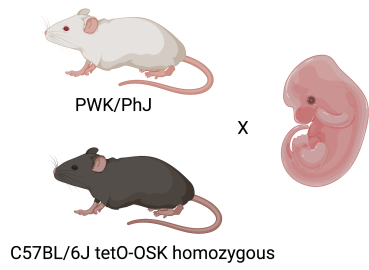
1A
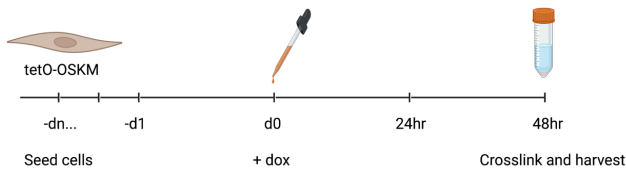
1B
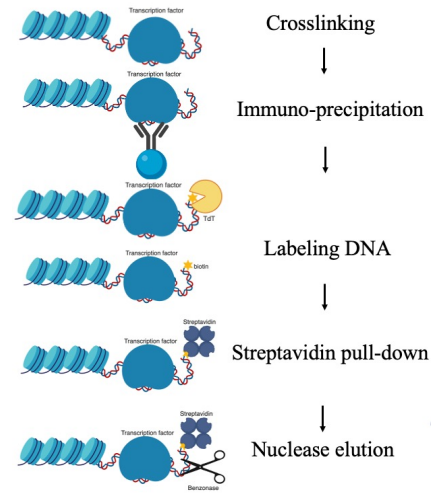
1C

1D

1E
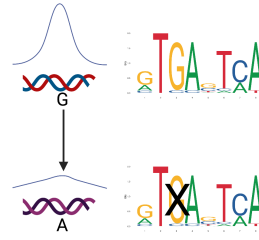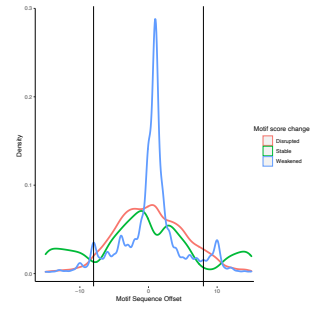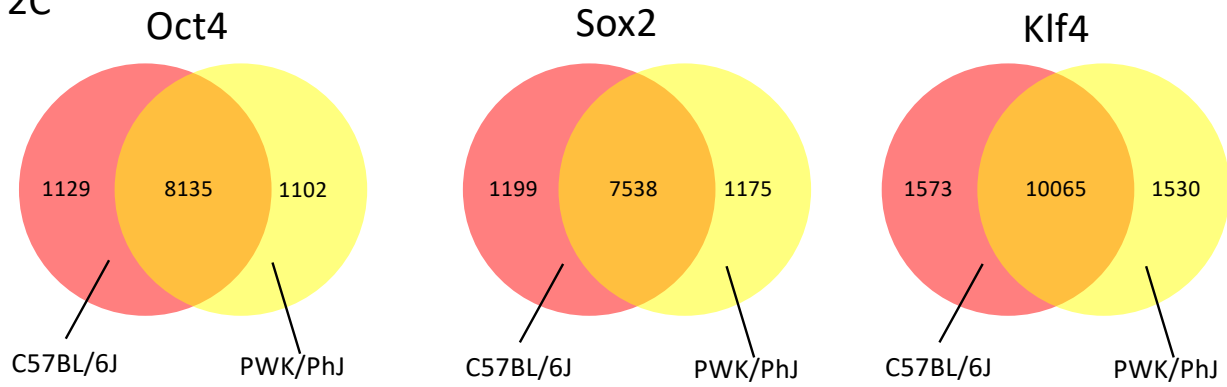7 TFs bZIP

16 TFs Home domain

6 TFs C2H2 zinc finger

1F

2A

PWK/PhJ

X

C57BL/6J tetO-OSK homozygous

OSK CUT&RUN
ATAC-seq
H3K27ac CUT&Tag

2B

2C

Oct4

1129 | 8135 | 1102

C57BL/6J   PWK/PhJ

Sox2

1199 | 7538 | 1175

C57BL/6J   PWK/PhJ

Klf4

1573 | 10065 | 1530

C57BL/6J   PWK/PhJ

2D

Oct4

MEF promoter

AP-1 — 1.66

Klf4 — 1.34

CEBP — 1.53

Skewed to the other allele with Motif disrupted

Sox2

MEF promoter

Sox2 — 1.86

CEBP — 2.21

AP-1 — 1.54

Skewed to the other allele with Motif disrupted

Klf4

MEF promoter

Klf4 — 1.34

AP-1 — 1.36

TEAD — 1.15

Oct4 — 1.51

Skewed to the other allele with Motif disrupted

Trensient Enhancers

Klf4 — 1.53

Sox2 — 1.14

Skewed to the other allele with Motif disrupted

3A

3B

AP-1 family members

3C

Oct4

3D

Jun

3E

Cebpa

MEFs 48h

9109 6621 10502

42% 87%

Cebpb

MEFs 48h

21482 16831 5333

20% 83%

3F

4A

Oct4   −   +

Sox2

Klf4

c-Jun

Cebpα

Cebpβ

4B

MEF Enhancers

TEAD   ACATTCCTcG   -   1.38

Log2FC

1.38

Skewed to the other allele with Motif disrupted

4C

Merge   α-Oct4   DAPI

WT

Del79

4D

MEF
- ATAC
- C&R
- H3K27ac

Oct4 del79
- ATAC
- C&R
- H3K27ac

Oct4 WT
- ATAC
- C&R
- H3K27ac

Pou5f1   Pou5f1

4E

H3K27ac signal MEF Enhancer profile

ME state5

del79_K27ac
Oct4_K27ac
MEF_K27ac

ME state6

−2.0   center   2.0Kb   −2.0   center   2.0Kb

4F

- Oct4
- Sox2
- Klf4
- Primary
- Secondary

Somatic Sites

Ac   TFsomatic   Ac   TFsomatic

Pluripotency Sites

TFsomatic   Ac   Ac   TFsomatic

Chapter 3

Carboxylate-Modified Magnetic Bead (CMMB)-Based Isopropanol Gradient Peptide Fractionation (CIF) Enables Rapid and Robust Off-Line Peptide Mixture Fractionation in Bottom-Up Proteomics
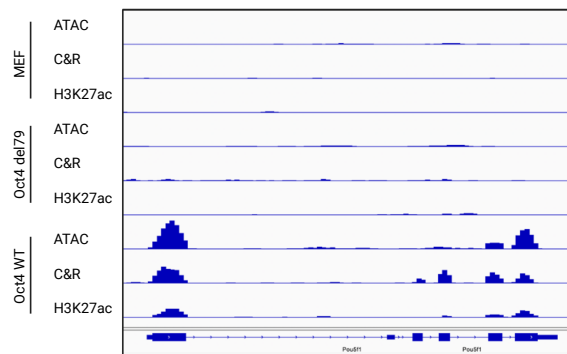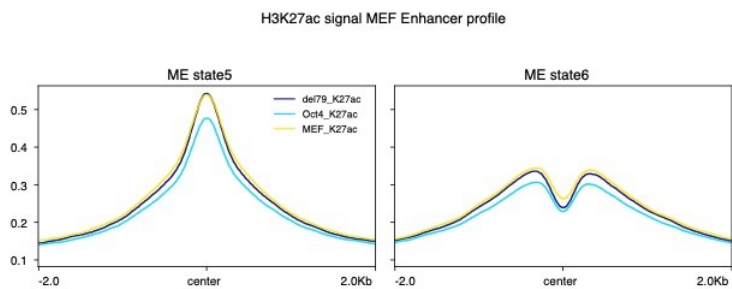
Carboxylate-Modified Magnetic Bead (CMMB)-Based Isopropanol Gradient Peptide Fractionation (CIF) Enables Rapid and Robust Off-Line Peptide Mixture Fractionation in Bottom-Up Proteomics

Authors

Weixian Deng, Jihui Sha, Kathrin Plath, and James A. Wohlschlegel

Correspondence

jwohl@ucla.edu

In Brief

We propose a novel method for fractionating tryptic peptide mixtures on carboxylate-coated magnetic beads. It is an extension of the previously reported SP3 (single-pot solid phase–enhanced sample preparation) protein and peptide cleanup method and provides an effective but complementary approach to other commonly used fractionation methods including strong cation exchange (SCX) and reversed phase (RP)-based chromatography.

Graphical Abstract
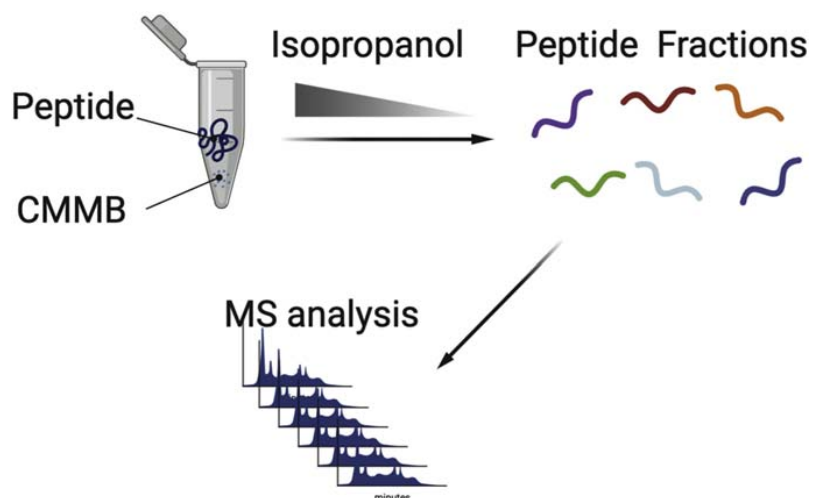
Highlights

- CIF provides a new on-bead offline peptide fractionation method based on HILIC.

- CIF offers a complementary dimension to RP for peptide fractionation.

- CIF can be seamlessly integrated with SP3 protein sample desalting.

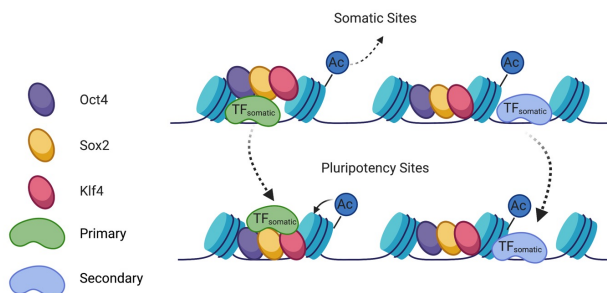- A machine learning model is able to predict CIF peptide fractionation patterns based on its peptide sequence.

Carboxylate-Modified Magnetic Bead (CMMB)-Based Isopropanol Gradient Peptide Fractionation (CIF) Enables Rapid and Robust Off-Line Peptide Mixture Fractionation in Bottom-Up Proteomics

Weixian Deng[1,2], Jihui Sha[1], Kathrin Plath[1], and James A. Wohlschlegel[1,*]

Deep proteome coverage in bottom-up proteomics requires peptide-level fractionation to simplify the complex peptide mixture before analysis by tandem mass spectrometry. By decreasing the number of coeluting precursor peptide ions, fractionation effectively reduces the complexity of the sample leading to higher sample coverage and reduced bias toward high-abundance precursors that are preferentially identified in data-dependent acquisition strategies. To achieve this goal, we report a bead-based off-line peptide fractionation method termed CIF or carboxylate-modified magnetic bead–based isopropanol gradient peptide fractionation. CIF is an extension of the SP3 (single-pot solid phase–enhanced sample preparation) strategy and provides an effective but complementary approach to other commonly used fractionation methods including strong cation exchange and reversed phase–based chromatography. We demonstrate that CIF is an effective offline separation strategy capable of increasing the depth of peptide analyte coverage both when used alone or as a second dimension of peptide fractionation in conjunction with high pH reversed phase. These features make it ideally suited for a wide range of proteomic applications including the affinity purification of low-abundance bait proteins.

Shotgun mass spectrometry has been the major strategy for bottom-up proteomics for decades (1). This technique involves analyzing a population of proteolytically digested peptides that are eluted from the reversed-phase (RP) separation into the mass spectrometer and then selecting the most intense ones for fragmentation to generate sequence information. Owing to limitations in scan speed, however, mass spectrometers are unable to fragment and scan all of the precursors eluting at a given time into the mass spectrometer, resulting in undersampling of low-abundance peptides. This

problem becomes more severe as the peptide mixture becomes more complex and ultimately results in reduced proteomic depth for the analysis of many complex biological samples. In addition, the large number of coeluting peptide precursors also leads to ion suppression, which further limits the ability to identify and quantify low-abundance precursors (2). These issues are typically addressed, at least in part, by using chromatographic methods to reduce the complexity of the mixture to enable the mass spectrometer to isolate and fragment the majority of peptides coeluting at any given retention time. Strong cation exchange (SCX) and high pH RP chromatography have emerged as the two most common strategies for reducing peptide complexity offline before analysis by LC-MS/MS (2, 3). In addition, owing to lack of accessibility to HPLC equipment, many laboratories use spin-column, stage-tip, or solid-phase extraction cartridges filled with matrix material (C18 for RP, benzenesulfonic acid bonded sorbent for SCX) instead of an HPLC, in essence, sacrificing some fractionation efficiency for speed and ease of use.

Although SCX fractionation methods have strong orthogonality to low pH RP chromatography (4), the use of salt in the mobile phase for elution requires an extra desalting procedure to make it compatible with LC-MS. Moreover, SCX chromatography often suffers from inefficient peptide recovery because of secondary interactions with the SCX sorbents that reduce the recovery of hydrophobic peptides (5). Compared with SCX, high pH RP fractionation requires no additional cleanup steps for fractionated products that limits sample loss. Like SCX, however, high pH RP suffers from incomplete peptide recovery with this material loss becoming more evident in samples with low amounts of peptide (6). Considering both offline high-pH RP and online low-pH RP utilize a similar peptide binding matrix and

buffers (besides pH), the orthogonality of fractions for low-pH RP is not ideal (7).

To circumvent the disadvantages of SCX and RP chromatography, a rapid, robust fractionation method that is compatible with low sample amounts and orthogonal to online low-pH RP chromatography is needed. In a previous study, it was shown that proteins and tryptic peptides can be immobilized on the hydrophilic surface of carboxylate-modified magnetic beads (CMMBs and also widely known as SP3) in an unbiased manner using a high concentration of the organic solvent (8–10). This method is derived from a mechanism similar to hydrophilic interaction liquid chromatography (11) or electrostatic repulsion hydrophilic interaction chromatography (12). It features high material recovery and high binding capacity and can be easily integrated into a variety of proteomics applications. Using the CMMB/SP3 technology, peptides are eluted from the beads when the acetonitrile (ACN) concentration is decreased below 90% with no detectable retention. However, owing to the narrow ACN concentration window over which peptide elution occurred, only limited success was reported when peptide fractionation was attempted using CMMB/SP3 (8). In this study, we describe a novel CMMB-based isopropanol gradient peptide fractionation method that we termed CIF that allows the elution of peptides into fractions using a step-wise isopropanol gradient. This strategy not only leverages the high binding capacity and low material loss advantages of CMMB but also achieves effective offline peptide-level fractionation, thus facilitating deeper proteomic coverage and improved analysis of high dynamic range samples.

EXPERIMENTAL PROCEDURES

Cell Culture and Tryptic Peptide Preparation

HEK293 cells were cultured in high glucose and glycine DMEM containing 10% FBS and 1% penicillin-streptomycin and then trypsinization for harvesting. Incubating cells in the lysis buffer (8 M urea, 0.1 M Tris HCl, pH 8.0) at 4 °C for 30 min followed by centrifugation to clarify the sample. Two milligrams of protein were reduced and alkylated by sequentially incubating with 5-m ris(2-carboxyethyl) phosphine and 10-mM iodoacetamide or 30 min at room temperature (RT) in the dark. The protein sample was then diluted fourfold with 0.1 M Tris HCl, pH 8.0, to reduce the final urea concentration to 2 M before incubating overnight with at 37 °C with trypsin protease at ratio of 1:100. SP3 was reported not suitable for cleanup of high quantity of proteins (high microgram or milligram quantities) (9). So, peptide digests were desalted using Pierce C18 tips (100-μl bed volume, cat. 87784), dried, and then reconstituted in water.

Peptide Recovery Assay

For each elution concentration tested, 1.7 μg of peptides were reconstituted in 10 μl of water, mixed with 5 μl of CMMB (GE Healthcare: 65152105050250, GE Healthcare: 45152105050250, mixed at 1-to-1 ratio) followed by 300 μl of ACN, which raises the final ACN concentration to 95% and allows peptide binding to CMMB. Peptides were eluted from CMMB by incubation with 30 μl of the elution buffer containing varying amounts of isopropanol (95%, 90%, 85%, 80% 75%, 70%, 0% of isopropanol) in a thermomixer for 15 min. Peptide

concentrations were determined for each sample by measuring their absorbance at 205 nm on a NanoDrop 3000 spectrophotometer.

Optimization of Bead Amount for CIF

To optimize the ratio of beads required to fractionate 20 μg of peptides using CIF, we tested the ability of 1 μl, 5 μl, 20 μl, and 50 μl of 50 μg/μl CMMB to fractionate 20 μg of peptides. Twenty microliter-beads (~1000 μg) showed the greatest number of identified peptides (supplemental Fig. 1, *A* and *B*), whereas 5 μl beads (~250 μg) showed only marginally fewer identified peptides. We conclude that CIF is effective over a broad range of peptide-to-bead mass ratios ranging from 1:12.5 to 1:50.

CIF

For each experiment, a defined amount of digested peptide (7 μg for the pH comparisons, 100 μg for comparisons between fractionation methods, and 20 μg for modeling the CIF elution concentration experiment) was bound to CMMB by incubating the peptide/CMMB mixture in 95% ACN for 10 min at RT. Peptides were eluted by sequentially incubating the CMMB with 30 μl of each elution buffer (90%, 85%, 80% 75%, 70%, 0% of isopropanol) and pipetting up and down 15 times (pipetting up and down can be replaced by shaking on a shaker at 1200 rpm for 10 min). Each elution step was repeated once to ensure no residual peptides were carried over into the next elution. Eluted peptides were dried by vacuum centrifugation and reconstituted in 5% formic acid for LC-MS/MS analysis (Fig. 1A). To test different pH conditions, 50-mM triethylamine bicarbonate was added to the isopropanol elution to create high-pH conditions, whereas 5% formic acid was added to the isopropanol elution to generate low-pH conditions.

High-pH RP Fractionation

High-pH RP fractionation was performed according to the manufacturer's instructions (Pierce High-pH Reversed-Phase Peptide Fractionation Kit, catalog number: 84868). Essentially, 100-μg peptides were bound to the resin in the spin column and then eluted by stepwise incubations with 300 μl of increased ACN concentrations. Fractions were then dried by vacuum centrifugation and reconstituted in 5% formic acid for mass spectrometry analysis.

RP-CIF 2D Fractionation

For eight-fraction RP-CIF 2D fractionation, 100 μg of the peptide mixture was bound to the resin in the spin column and then eluted stepwise with 300 μl of increasing ACN concentrations. Subsequently, eluate fractions 1 and 5, 2 and 6, 3 and 7, and 4 and 8 were combined pairwise into four total fractions, dried by vacuum centrifugation, and reconstituted in 30 μl of water. Each combined RP fraction was bound to 10-μl CMMB and eluted in two steps (80% isopropanol and water) as described in the CIF section. For 16-fraction RP-CIF 2D fractionation, conditions were identical except RP fractions were not combined before applying CIF.

APEX2-Based Proximity Labeling

The APEX2-Oct4 fusion protein coding sequence was cloned into the pMX retro-viral packaging vector and then transiently transfected into HEK293 cells with Lipofectamine 3000. Five hundred micro molar biotin-phenol was added to the media 18 h after transfection and incubated at 37 °C for 30 min. Then peroxidase reaction was activated by adding H_2O_2 to 1 mM and incubating at RT for 1 min. The reaction was quenched by washing cells three times with quencher containing PBS (10-mM sodium azide, 5-mM Trolox, 10-mM sodium ascorbate). Cells were harvested by trypsinization and then flash-frozen in liquid nitrogen.

Streptavidin Pull-Down

Cells are lysed in RIPA buffer (50-mM Tris HCl, pH 7.5, 150-mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100) supplemented with protease inhibitor cocktail (Roche) and Benzonase (1 µl of 250 U/µl) and incubated at 37 °C for 20 min. Lysates were clarified by centrifugation, quantitated using the Pierce 660-nm protein assay, and 1 mg of protein was incubated with 300 µl of high-capacity streptavidin (SA) beads (Thermo Fisher) for each sample at RT for 1 h. SA beads were then washed three times with RIPA buffer, once with 1 M KCl, once with 2 M urea in 25-mM Tris HCl, pH 8.0, and three more times with RIPA buffer. Bound proteins were then reduced, alkylated, and digested on beads with Lys-C and trypsin. The supernatant from the on-bead digestion was then transferred to another tube, bound to SP3/CMMB beads by the addition of ACN to a concentration of 95%, and either eluted from CMMB in water or fractionated using the CIF protocol into three fractions (85%, 75%, and 0% isopropanol elution steps). Although it is difficult to measure the protein abundance on beads in this type of analysis, we estimate that there is less than 1 µg of protein (not including SA) based on other experiments using comparable purification strategies.

LC-MS Data Acquisition

A 75-µm × 25-cm homemade C18 column was connected to a nano-flow Dionex Ultimate 3000 UHPLC system. The 70-min gradient of increasing ACN was delivered at a 200 nl/min flow rate as follows: 1% ACN phase from minutes 0 to 6, 6 to 25% ACN from minutes 6 to 55, 25 to 32% ACN from minutes 55 to 63.5, 32 to 80% ACN from minutes 63.5 to 67, and then 1% ACN from minutes 68 to 70. An Orbitrap Fusion Lumos Tri-brid mass spectrometer was used for data acquisition. Full MS scans were acquired at 120 K resolution with the automatic gain control target set to 2e5 and a maximum injection time set to 100 ms. MS/MS scans were collected at 15K resolution after isolating precursors with an isolation window of 1.6 m/z and HCD-based fragmentation using 35% collision energy. For data-dependent acquisition, a 3-s cycle time was used to acquire MS/MS spectra corresponding to peptide targets from the preceding full MS scan. Dynamic exclusion was set to 25 s.

Database Search

MS/MS database searching was performed using MaxQuant (1.6.17.0) against the human reference proteome from EMBL (UP000005640_9606 HUMAN *Homo sapiens*, 20600 entries, released in 2020_04). The search included carbamidomethylation on cysteine as a fixed modification and methionine oxidation and N-terminal acetylation as variable modifications. The digestion mode was set to trypsin and allowed a maximum of two missed cleavages. The precursor mass tolerances were to 20 and 4.5 ppm for the first and second searches, respectively, whereas a 20-ppm mass tolerance was used for fragment ions. Data sets were filtered at 1% false discovery rate at both the peptide spectral match (PSM) and protein level. Peptide quantitation was performed using MaxQuant's LFQ mode.

Modeling the CIF Elution Profile

We used peptide and elution information from the CIF data set to identify the physicochemical properties of peptides that determine their elution using CMMB. The R package, Peptides (v2.4.1), was used to calculate the aliphatic index (13), peptide charge at given pH, peptide isoelectric point (14), instability index (15), and hydrophobicity (16) as well as 18 parameters related to amino acid composition including the number and mole percentage of nine classes of amino acids. These variables were normalized using the Max-Min normalization method to ensure all the variables values were within (0,1). The model was generated using all three replicates of the CIF data that

were combined by assigning the isopropanol elution concentration in which the peptide displayed maximum LFQ intensity across all replicates and concentrations. Peptides eluted in 0% isopropanol were removed because it was not possible to determine a narrow range of isopropanol over which those peptides were eluted. After combining and preprocessing the data, we obtained a matrix with 39,225 peptide sequences or rows, 23 columns of variables, and an observed elution isopropanol concentration. We then separated the data set into two parts by randomly assigning two-thirds of the rows to the training set and one-third to the test set. Using the training set, we trained a Lasso regression with tenfold cross-validation using the R package glmnet (v3.0–2) (17). We selected the model where the lambda value provided the most regularized model such that error was within one standard error of the minimum and then removed features that contributed minimally to the model (Table 1). An R script that generates peptide properties and predicts the isopropanol elution concentration can be found here: https://github.com/weixiandeng/CIFpredictor.

Calculation of the Distribution Index

The distribution index is calculated by the following equation for which we partition the 70-min gradient into 70 equal bins and denote the number of PSM counts in each bin as P_j.

$$Distribution\ index = \frac{\sum_{i=1}^{70} P_i}{\sum_{j=1}^{70} P_j} \times 100\%$$

P_j is PSM count with in each retention time bin,

$$P_i = \begin{cases} P_i = P_j, P_i \leq \dfrac{\sum_{j=1}^{70} P_j}{70} \\ P_i = \dfrac{\sum_{j=1}^{70} P_j}{70}, P_i > \dfrac{\sum_{j=1}^{70} P_j}{70} \end{cases}$$

Experimental Design and Statistical Rationale

Peptide recovery experiments were performed in three technical replicates. CIF experiments exploring different pH conditions (Fig. 1, *C*

TABLE 1
Peptide fractionation pattern multivariable linear model

Variables		Coefficients
(Intercept)		76.62
Charged No.		−21.05
Charge		11.87
Hydrophobicity		8.30
Polar No.		−6.82
Acidic mole %		−5.34
a-Index		5.27
Tiny mole %		−3.83
Nonpolar mole %		0.85
Instability index		−0.19
RMSE		3.82
R^2		0.70

Physical chemistry peptide properties (gray) and peptide composition properties (green shaded) contribute to the model.
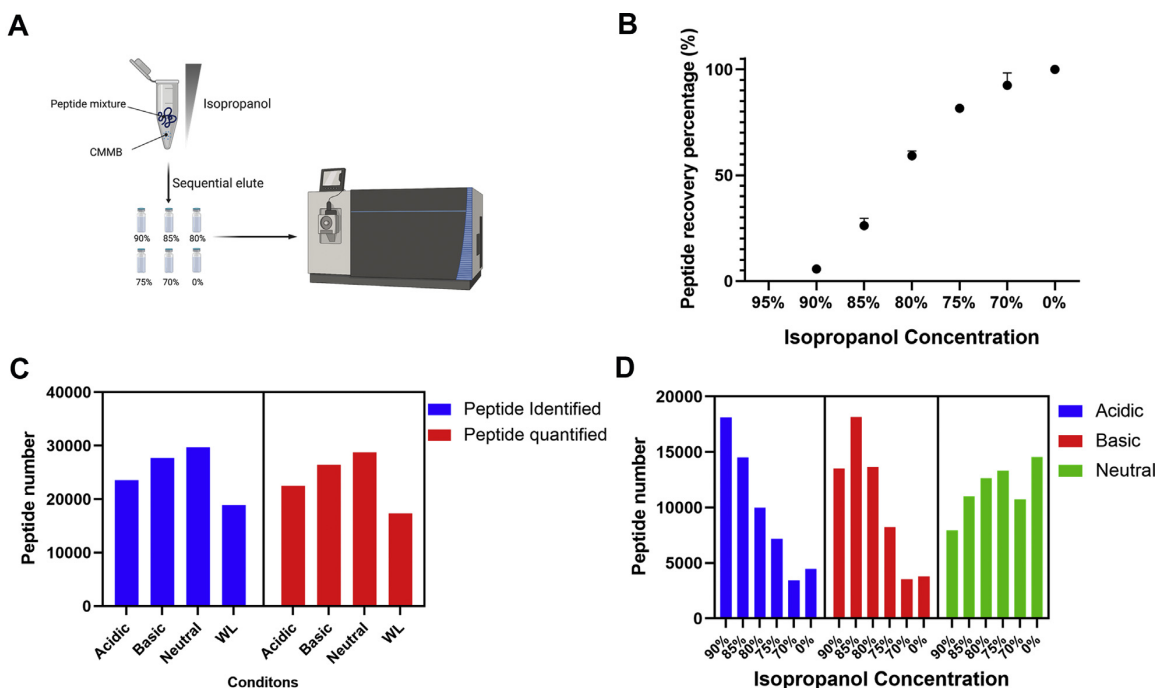
FIG. 1. **Effective stepwise elution of peptides from CMMB using isopropanol at neutral pH**. *A*, the schematic of CIF workflow. *B*, fraction of peptides eluted from CMMB at different isopropanol concentrations, 95% (mean: 0, SD: 0.39), 90% (mean: 5.74, SD: 0.74), 85% (mean: 26.11, SD: 3.35), 80% (mean: 59.13, SD: 2.32), 75% (mean: 81.61, SD: 1.05), 70% (mean: 92.32, SD: 5.83), 0% (mean: 99.88, SD: 0.64), (n = 3). *C*, the number of unique peptides identified or the number of peptides quantified after LC-MS/MS analysis of peptides fractionated by CIF in acidic (isopropanol solution with 5% formic acid), neutral (isopropanol solution with water) or basic (isopropanol solution with 50-mM triethylamine bicarbonate) pH conditions. WL is the unfractionated control. (n = 1). *D*, the number of unique peptides identified by LC-MS/MS analysis in each isopropanol fraction after fractionation by CIF under acidic, basic, and neutral pH conditions (n = 1). CIF, carboxylate-modified magnetic bead–based isopropanol gradient peptide fractionation; CMMB, carboxylate-modified magnetic beads; WL, whole lysate.

and *D*) were single replicate-relative comparisons that were validated in subsequent experiments. All comparisons between CIF, RP, and RP-CIF-2D fractionation experiments (Figs. 2 and 3) were performed using three technical replicates from the same HEK293 digested lysate. Statistical comparison of peptides identified in Figure 2*B* were conducted using an unpaired Student's *t* test. APEX2-based proximity labeling experiment was performed using two technical replicates. Nontransfected but otherwise identically treated cells served as the negative control for the analysis. MSStats (3.10) was used to analyze the MaxQuant LFQ data in the APEX2-Oct4 proximity labeling experiment to statistically assess protein enrichment. Equalized medians were used for normalization and the Tukey median polish method was used for protein summarization. *p*-Values for *t* tests were corrected for multiple hypothesis testing using the Benjamini-Hochberg adjustment. For proteins absent in either condition, the fold change was imputed based on its abundance in the detected condition.

RESULTS

Isopropanol-Based Fractionation of Peptides on CMMB

According to Hughes *et al*. (8), the majority of bound peptides are coeluted from CMMB/SP3 when the ACN concentration drops below 87%, providing only a narrow useful range

for stepwise elution. To identify other solvents capable of eluting peptides in a graded manner from CMMB, we tested methanol, ethanol, and isopropanol as candidates to replace ACN. For each solvent, we bound tryptic peptide digests from HEK293 whole-cell lysates to CMMB in the presence of 95% ACN (by volume). We then eluted the peptides from the CMMB by incubating with stepwise decreases in the solvent concentration (95%, 90%, 85%, 80%, 75%, 70%, and 0%) (Fig. 1*A*). Peptide recovery for each solvent at each elution condition was measured by peptide absorbance at 205 nm. Methanol, ethanol, and ACN showed similar elution profiles in which nearly all peptides were eluted across a narrow range of solvent (data not shown). In contrast, peptides were strongly bound to the beads in 95% ACN and eluted gradually from the beads by decreasing isopropanol concentrations (Fig. 1*B*). We also examined the ratio of beads to digested protein required for optimal separation using CIF and determined a bead-to-protein ratio of 1:12.5 to 1:50 resulted in optimal fractionation and peptide identification (supplemental Fig. 1, *A* and *B*). These characteristics suggested that isopropanol could be useful as a solvent for offline peptide fractionation on CMMB
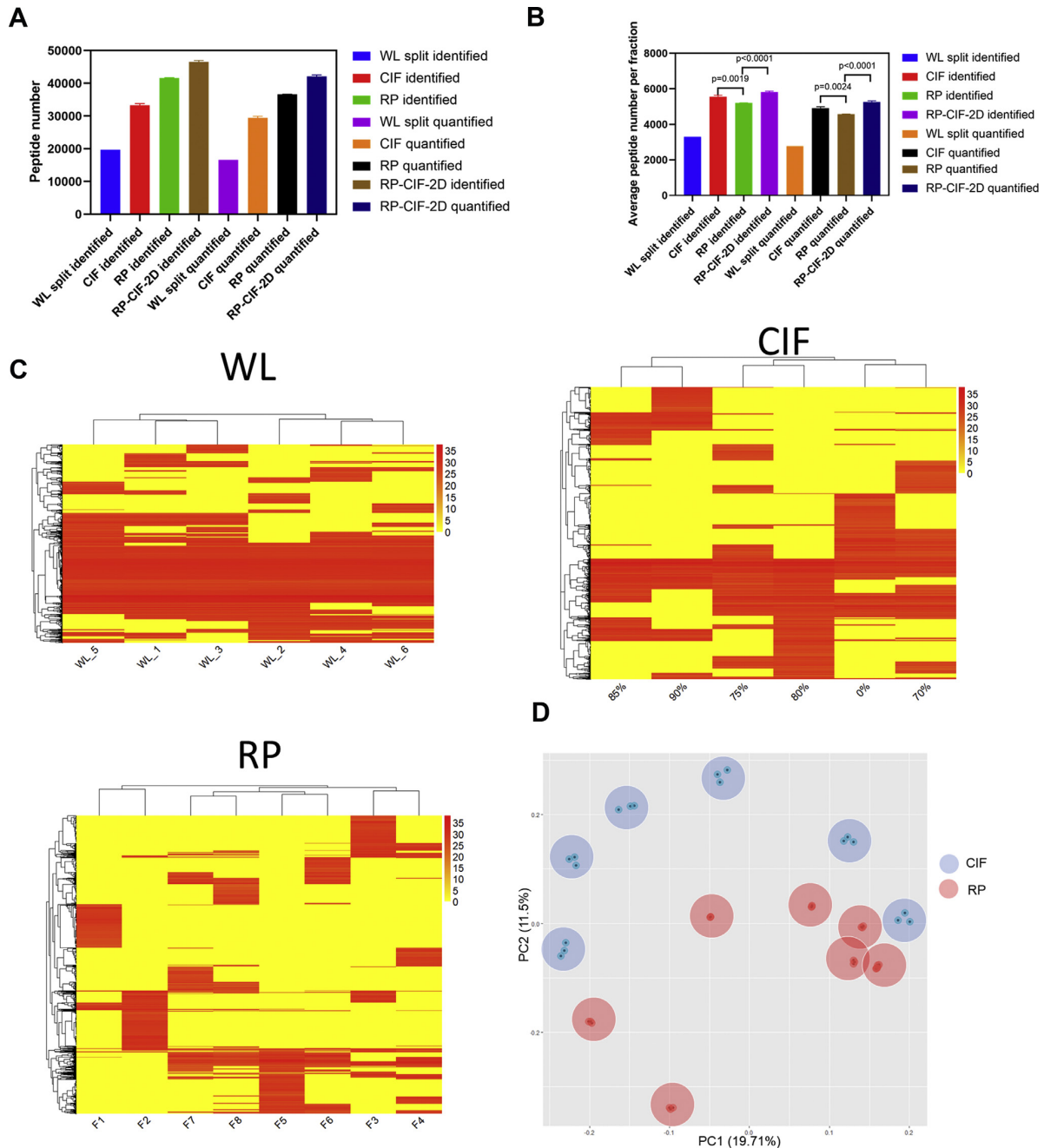
FIG. 2. CIF effectively fractionates complex peptide mixtures leading to improved depth of proteome coverage. *A*, the number of unique peptides identified or the number of peptides quantified by LC-MS/MS analysis of 100 μg of digested peptides either evenly split into six fractions (WL, no replicate) or fractionated using CIF RP or eight-fraction RP-CIF-2D (n = 3). *B*, identical to (*A*), except the number of identified peptides, the number of quantified peptides is normalized by the number of fractions analyzed. *C*, the heatmaps of quantified peptides in different fractions from WL (replicates of unfractionated sample), CIF (six fractions), and RP (eight fractions) samples. Each heatmap is derived from one representative sample. *D*, principal component analysis (PCA) of peptide intensities from each fraction for both CIF and RP experiments including replicates (n = 3). Each *dot* represents one replicate LC-MS/MS while each shaded *circle* represents all of the replicates corresponding to a specific fraction. CIFs are shaded in *red*, whereas RP fractions are shaded in *blue*. CIF, carboxylate-modified magnetic bead–based isopropanol gradient peptide fractionation; RP, reversed phase; WL, whole lysate.

before LC-MS analysis and that this fractionation worked over a wide bead:protein ratio.

We also examined peptide elution from CMMB under different pH environments because previous work on hydrophilic interaction liquid chromatography (18) found that pH can affect the binding affinity of peptides to a noncharged matrix. We bound HEK293-derived tryptic peptide digests to CMMB and then eluted peptides stepwise into six fractions using decreasing concentrations of isopropanol (90%, 85%, 80%, 75%, 70%, 0%) across a range of pH levels. Fractionated samples were subjected to LC-MS analysis and compared with unfractionated peptide digests as a control. In Figure 1C, we compared the number of unique peptides identified and quantified peptides across different pH conditions relative to a single LC-MS/MS run of the unfractionated control (whole lysate). Both identified and quantified peptides are higher in fractionated samples irrespective of pH. For pH comparisons, neutral pH fractionation generated more peptide identifications relative to the acidic and basic fractions (Fig. 1C). In addition, neutral pH showed the most even distribution of peptides across the isopropanol steps, whereas both acidic and basic fractionations resulted in the majority of the peptides being eluted in the early elution steps (Fig. 1D). These data suggest that the even distribution of peptides across isopropanol fractions at neutral pH relative to low/high pH explains its improved peptide identification rates. Overall, these findings indicate that neutral pH elution using isopropanol effectively fractionates CMMB/SP3 bound peptides across a broad concentration range and led us to explore its utility in proteomic applications.

Peptide Fractionation by CIF Increases Proteome Coverage

One of the main benefits of peptide fractionation is improved analyte coverage. Figure 1 showed that CIF produced a higher number of identified peptides and quantifiable peptides relative to an unfractionated sample. However, this increase could have been attributed primarily to the length of the analysis because data were collected for six 70-min gradients for the fractionation experiment (six fractions × 70 min per fraction) compared with only a single 70-min analysis of the unfractionated sample. To distinguish whether the increase in analyte coverage resulted from longer data acquisition or reduced complexity due to the fractionation itself, we compared peptide identifications between a sample in which HEK293-derived tryptic peptides were partitioned into six fractions using CIF and analyzed by LC-MS/MS (six fractions × 70 min of LC-MS/MS per fraction) and compared with a sample in which the tryptic peptides were evenly divided into six unfractionated parts that were analyzed by LC-MS/MS (six unfractionated samples × 70 min of LC-MS/MS per sample). For the CIF method, we identified on average 33,311 unique peptides per sample and could quantify 29,452 of them (Fig. 2A). For the unfractionated

samples, we identified 19,723 unique peptides of which 16,637 were quantifiable (Fig. 2A). These data demonstrate CIF collectively produces 69% more identified peptides and 77% more quantifiable peptides from the same amount of input material and analysis time compared with unfractionated samples. For analyses of individual fractions, we identified and quantified on average 3287 and 2773 peptides, respectively, from an unfractionated sample compared with 5815 and 4909 peptides on average, respectively, from a CIF (Fig. 2B). We conclude that CIF is an effective method for peptide fractionation and increases peptide coverage by reducing sample complexity.

We next compared the effectiveness of CIF with high pH RP fractionation, a widely used offline peptide fractionation method. Using the Pierce High-pH Reversed-Phase Fractionation Kit, we fractionated our digested peptide sample into eight fractions following the manufacturer's recommendations. LC-MS/MS analysis of these eight fractions resulted in the identification of 41,638 unique peptides and 36,583 quantifiable peptides compared with 33,311 and 29,452 peptides for the six-fraction CIF analysis (Fig. 2A). Although high-pH RP outperforms CIF under these conditions (25% and 24% increase in unique peptides identified and quantified, respectively), these increases likely result from the 33% longer analysis time (eight fractions *versus* six fractions). Consistent with this idea, if we normalize the number of identified and quantified peptides per 70-min gradient, then CIF modestly outperforms high-pH RP (5815 *versus* 5204 identified peptides and 4909 peptides *versus* 4573 quantifiable peptides per 70-min gradient). We conclude that CIF and high-pH RP have comparable efficiency in peptide-level offline fractionation experiments.

Next, we compared the peptide elution pattern across the different fractions for unfractionated (whole lysate) *versus* CIF *versus* high-pH RP. As expected, we find that a large number of abundant peptides are reproducibly identified in replicate analyses of unfractionated samples (Fig. 2C). However, CIF and high-pH RP fractionation both produced peptides that are predominantly enriched in only one fraction. Notably, this trend is weaker in CIF than in RP with a significant number of peptides also being enriched in two to three fractions, suggesting that the RP separation has better resolution than CIF. Nonetheless, both CIF and RP are effectively fractionating peptides with discrete populations of peptides being eluted at different concentrations of organic solvent. Importantly, both are effective at simplifying complex peptide mixtures to increase the depth of peptide coverage, making them well-suited for standard proteomic workflows.

To assess fraction-to-fraction reproducibility of CIF relative to high-pH RP, we performed the principal component analysis of the two methods. Consistent with the high degree of reproducibility in peptide identifications between replicates seen in Figure 2, A and B, the principal component analysis plot shown in Figure 2D demonstrates that replicate fractions

generated using the same concentration of organic solvent clustered well in both methods. We conclude that both methods are highly reproducible. In summary, we demonstrate that the CIF method is able to reproducibly increase peptide coverage when conducting proteomic analyses on complex peptide mixtures and is an effective alternative to high-pH RP yielding comparable results.

Peptides from CIFs Are Evenly Distributed Across RP Gradients

Distributing peptides across the entire chromatographic gradient is essential for performing MS/MS on as many different precursors as possible and thus maximize the efficiency of data acquisition. To evaluate how peptides from CIFs were distributed across the subsequent online RP gradient, we plotted the number of PSMs across the gradient for each fraction analyzed using 1-min bins. Figure 3*A* displays the PSM distribution across the LC gradient from each of the fractions CIF (left) and RP (right) fractionation experiment. As is shown in Figure 3*A* (left panel), all six fractions have almost identical shape with PSMs evenly distributed on the retention time space. In comparison, the fractions from an offline high-pH RP separation Figure 3*A* (right panel) have a more biased peptide elution pattern relative to CIF. Moreover, the peptide identifications are biased such that the earlier elution steps (F1 and F2) show more identifications in the early portion of the LC RP gradient while late fractions (F7 and F8) identify more peptides near the end of the gradient (Fig. 3*B*). These data demonstrate that CIF displays better orthogonality than high-pH RP to low-pH RP leading to a more even distribution of peptides across each gradient.

To analyze these chromatographic distributions in a more quantitative manner, we defined the distribution index as a metric for assessing the distribution of PSMs across the chromatographic gradient. Ideally, if PSMs are distributed evenly across the gradient, the PSM distribution would be flat and could be represented as a rectangle in which every bin has the same height (Fig. 3*C*, orange area). In practice, the number of identified PSMs can vary significantly across the chromatographic separation (Fig. 3*C*, blue area) and we can calculate a distribution index (see Experimental Procedures) that measures the extent to which the actual PSM distribution falls into the ideal distribution (Fig. 3*C*, overlapped area).

In Figure 3*D*, we plot the distribution index for each fraction from three replicate CIF or RP-based separations. This plot supports what we observed visually in Figure 3*A*. All six CIFs have a similar distribution index, whereas the distribution index for RP fractions was more variable—consistent with the idea that there is not perfect orthogonality between high-pH and low-pH RP chromatography. Figure 3*D* also shows a higher distribution index for CIFs relative to RP fractions, highlighting the even distribution of peptides across the RP gradient. Together, these data indicate that PSMs in CIFs are more evenly distributed on the ACN gradient axis of LC than

RP fractions, which may promote better LC-MS acquisition efficiency.

CIF Serves as the Second Dimension to RP to Further Increase Sequence Coverage in the Same Analysis Time

Based on the orthogonality between RP and CIF, we posited CIF could be used in a multidimensional fractionation strategy together with RP to further improve proteome coverage. To examine this possibility, we took HEK293 digested lysates fractionated using Pierce high-pH RP spin columns, combined them from eight into four fractions as described above, and then used CIF to fractionate each into two additional fractions using 80% isopropanol and water (eight fractions total). These fractions were analyzed by LC-MS/MS, and the peptide identifications were compared with peptide identifications obtained from eight high-pH RP fractions from the high-pH spin columns before they were combined. As demonstrated in Figure 2*A*, the RP-CIF-2D fractionation method identified and quantified on average 46,519 and 42,092 unique peptides, respectively, corresponding to 11.7% more identified peptides and 15% more quantified peptides compared with RP alone using the same amount of analysis time (eight × 70 min). We also analyzed a similar multidimensional fractionation experiment in which all high-pH fractions were subsequently separated using CIF into two additional fractions generating 16 fractions total. As shown in supplemental Figure 2*A*, the 16-fraction RP-CIF-2D fractionation identified and quantified 46.4% and 54% more peptides, respectively, than RP alone, further highlighting the improved depth of this approach.

CIF Improves Coverage Depth in AP-MS Workflows

Having demonstrated that CIF is a reliable method for peptide-level fractionation in standard proteomics workflows, we next explored applications that might specifically benefit from it. We and others have reported low sample loss as a major strength of CMMB peptide cleanup, leading us to examine whether CIF was suitable for fractionating low-abundance affinity-purified samples to increase depth and data quality. Here, we test CIF's utility in this regard using an APEX proximity labeling experiment. APEX is an engineered ascorbate peroxidase that can label nearby proteins with biotin in the presence of peroxide (19). Leveraging its proximity labeling capacity, researchers fuse it to their protein of interest or to specific localization signal for a cellular compartment to capture protein interactomes or compartment-specific proteomes, respectively. Although the high binding affinity between biotin and SA enables robust capture of labeled proteins and stringent washing conditions during the purification, the SA–biotin interaction creates technical challenges due to (1) the presence of endogenously biotinylated proteins that are captured by the immobilized SA and (2) the high affinity of the interaction often requires denaturation or tryptic digestion to efficiently elute captured proteins, which introduces highly
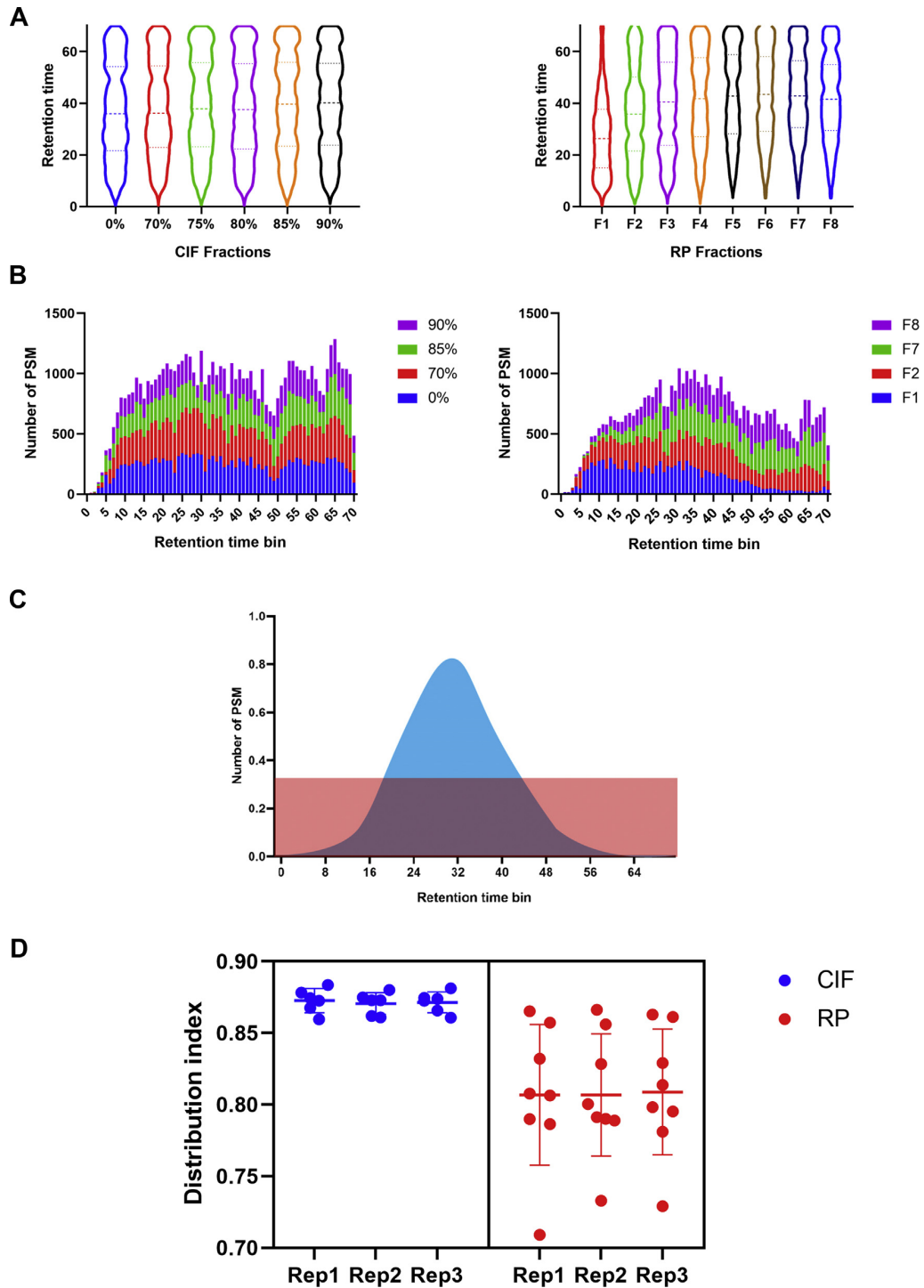
FIG. 3. **Peptides from CIFs are more evenly distributed than peptides from basic RP fractions across the LC-MS/MS online RP gradient**. *A*, the violin plot of PSM number distribution at corresponding retention time in each fraction of CIF (*left*) or RP (*right*) fractionation method. *B*, the histogram overlay of identified PSMs at the corresponding retention time divided into 1-min bins for the first two (*blue* and *red*) and last two (*green* and *purple*) fractions of the CIF-based (*left*) or RP-based (*right*) fractionation. These data are derived from one representative
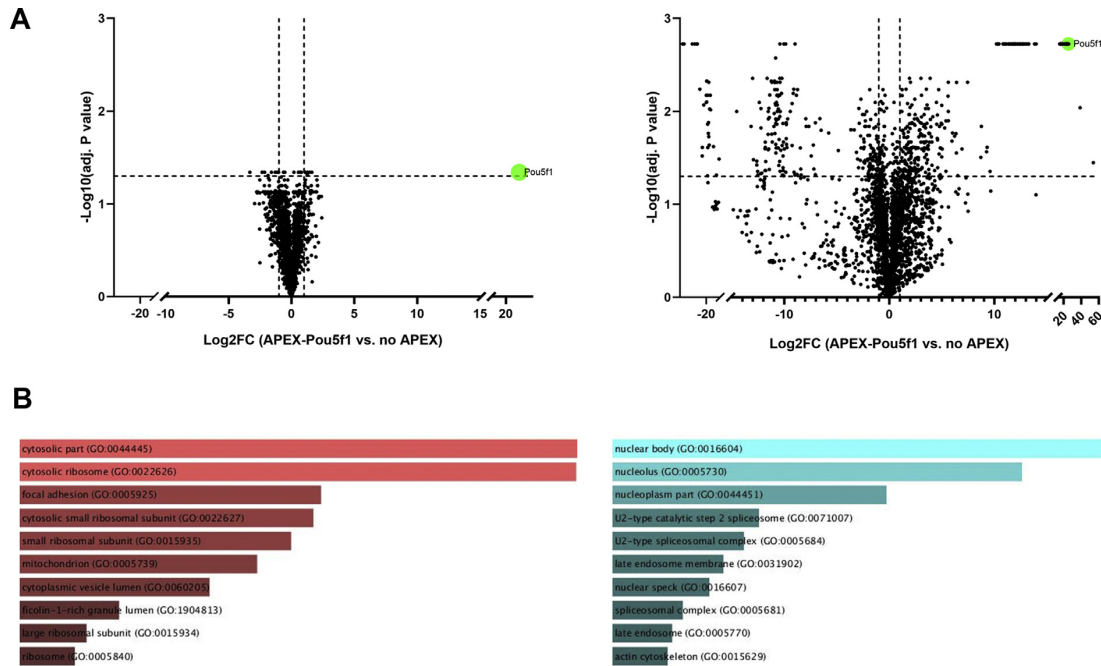
Fig. 4. **CIF improves proteome coverage of AP-MS samples.** *A*, the volcano plot of proteins enriched in APEX2-Oct4 (Pouf51) proximity labeling experiments. APEX labeling was performed in HEK293 cells expressing APEX2-Oct4 followed by streptavidin purification and LC-MS/MS analysis. Samples were analyzed without prefractionation (*left*) or with CIF (*right*). *B*, Gene ontology analysis (cellular component) of proteins identified as negatively enriched (*left*) or positively enriched (*right*) in APEX-Oct4 experiments CIF (*right*) samples. The samples in (*A*) and (*B*) are from two biological replicates. AP, affinity purification; CIF, carboxylate-modified magnetic bead–based isopropanol gradient peptide fractionation.

abundant SA into the sample for LC-MS analysis. Both endogenously biotinylated proteins and SA can significantly suppress signal from the less-abundant but physiologically relevant proteins in the sample making efficient fractionation a potential solution for improving the effectiveness of these approaches. However, owing to the low yield that is typical of affinity purification (AP) experiments, commonly used fractionation methods are not generally applied to AP-MS samples. To examine the utility of CIF for APEX-based AP experiments, we examined the proximal interaction of Oct4, one of the Yamanaka transcription factors involved in somatic cell reprogramming (20). Oct4 is a transcription factor that localizes strictly in the nucleus. We expressed APEX2-Oct4 in HEK293 cells and performed proximity labeling while using the parental HEK293 cells as negative control. After labeling, we performed SA AP, on-bead tryptic digestion and then bound the digested peptides to CMMB where peptides were sequentially eluted using three different isopropanol concentrations (85%, 75%, and 0%). In parallel, we conducted a control experiment using

CMMB to desalt the samples but without fractionation. Both samples were analyzed by LC-MS. Figure 4*A* shows that only six proteins were enriched in APEX2-Oct4 samples over the no APEX2 control (nontransfected HEK293 cells) (adjusted *p*-value ≤ 0.05, Log2 fold change ≥ 1) for the nonfractionated sample. Of these six proteins, one of them was Oct4 itself and four are localized in the nucleus, making them putative Oct4-proximal proteins. In the fractionated samples, 446 proteins were enriched in APEX2-Oct4 over the no APEX2 control that included Oct4 and 263 nuclear proteins (Fig. 4*B*). These results demonstrate that CIF is effective at fractionating low peptide amounts leading to major increases in sensitivity in proximity labeling experiments.

CIF Peptide Elution Properties Can Be Predicted by a Multi-Variable Linear Model

Figure 2, *C* and *D* show that CIF displays a distinct peptide fractionation pattern from RP. We next explored the

replicate out of three total replicates per condition. *C*, the schematic representing the distribution index where *red* indicated the idealized distribution if peptides elute evenly across the entire gradient, *blue* indicating a typical peptide elution distribution, and the area of the overlap representing the distribution index. *D*, the distribution index of replicate fractions from the CIF (*blue*) and RP (*red*) separations (n = 3). CIF, carboxylate-modified magnetic bead–based isopropanol gradient peptide fractionation; RP, reversed phase.
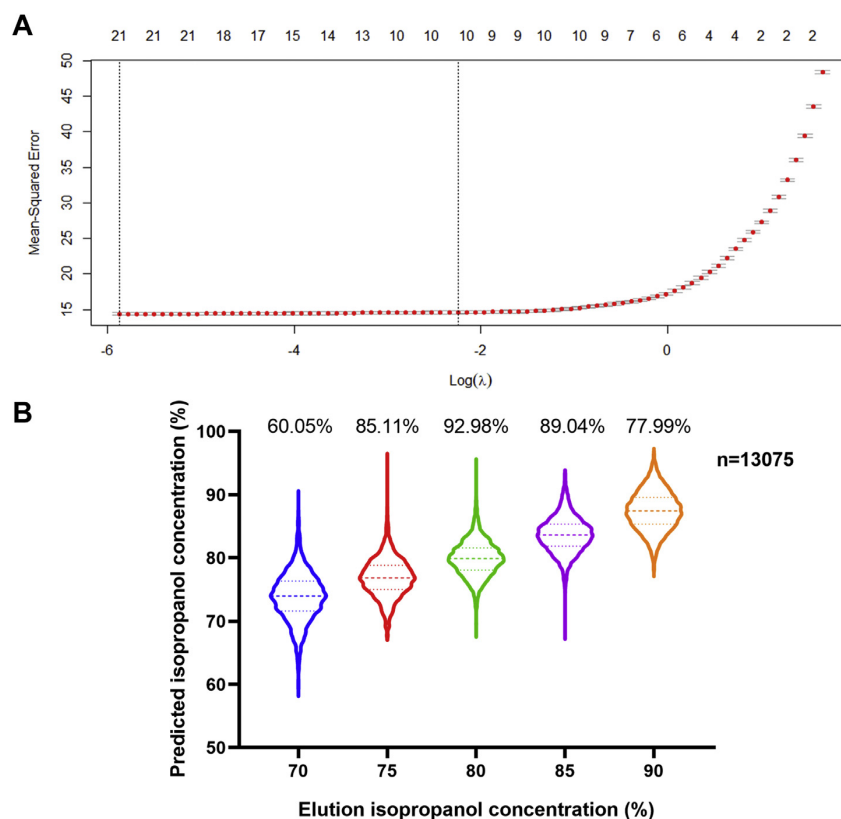
FIG. 5. **Peptide elution by isopropanol from CMMB is determined primarily by nine physicochemical properties of the peptides.** *A*, the Lasso regression model mean-squared error changes with change of the number of features (*top axis*) and regularization parameter (λ) (*bottom axis*). *B*, the violin plot of the predicted isopropanol concentration necessary to elute peptides from the test set compared with their experimentally determined elution profile from CIF. Test set data points (n = 13,075), the percentage of accurately predicted peptides in each experimentally measured fraction (on the *top* of each column). CIF, carboxylate-modified magnetic bead–based isopropanol gradient peptide fractionation; CMMB, carboxylate-modified magnetic beads.

determinants of this peptide elution pattern and assessed whether we could predict the peptide elution profile based on peptide sequence alone. To address this question, we attempted to model the elution profile based on peptide sequence. We used the R package Peptides to calculate five physical chemistry properties including aliphatic index, peptide charge at given pH, peptide isoelectric point, instability index, and hydrophobicity as well as 18 amino acid composition parameters that include the number and mole percentage of nine classes of amino acids for a training set of peptides with known elution properties for CIF. We then built a model to describe the relationship between the isopropanol concentration at which a peptide eluted from CMMB and 23 total peptide property variables to assess their relative contribution to the elution profiles (Fig. 5*A*). The model generated has nine variables that contribute significantly to the elution profile (Table 1). The R square for the training set equals 0.70 and the RMSE when cross-validated using a test set of peptides is 3.82, suggesting that the model describes the elution pattern very well. In Figure 5*B*, we plot the predicted isopropanol elution

concentration for each peptide in each fraction and considered the prediction correct if the predicted value fell within ± 5% of the observed isopropanol elution concentration. The final prediction accuracies for all fractions ranged from 60.05% to 92.98%, with an average of 81.03%.

Taken together, we built a multivariable linear model that predicts the concentration of isopropanol needed to elute a given peptide based on its sequence and identified the major physiochemical properties that determine this binding. The model indicates that the number of charged amino acids in the peptide sequence, the charge state of the peptide under neutral conditions, and the peptide's hydrophobicity are the key drivers determining which peptides are eluted at what isopropanol concentration.

DISCUSSION

In this study, we describe a CMMB-based peptide fractionation method that offers several features that provide significant utility in proteomics applications. First, desalting

and fractionation are performed in the same tube that minimizes sample loss and facilitates potential automation. Second, the binding capacity of CMMB is high, enabling small beads volumes and hence small elution volumes that also enhances sensitivity and limits losses. Third, CIF is orthogonal to SCX and high-pH RP allowing it to be easily integrated into multidimensional chromatographic schemes. Although we believe that high-pH RP chromatography is likely to remain the superior method for applications for requiring deep proteome coverage, the advantages of CIF outlined here support a strong complementary role for CIF in other common workflows that require high sample throughput or low amounts of starting material (*e.g.*, APs).

One major advantage to CIF is its orthogonality to the acidic online RP separations that are standard in the majority of LC-MS/MS workflows. Considering how evenly peptides are distributed across the LC gradient determines how efficiently data-dependent MS/MS acquisition occurs, the orthogonality of offline separations becomes a determining factor for the effectiveness of the analysis. Based on data in Figure 3, *A* and *C*, we demonstrate that CIF displays excellent orthogonality to RP chromatography in LC-MS/MS applications and is likely the reason for improved peptide identification and quantitation in CIFs. In addition, we take advantage of this orthogonality by demonstrating that CIF and high-pH RP can be used in 2D fractionation experiments to sequentially fractionate peptides offline before LC-MS/MS analysis to further increase peptide coverage.

Another major advantage of CIF is its scalability in terms of peptide input. Standard spin column-based high-pH RP kits typically used for offline fractionation separate input peptides ranging in amount from 10 to 100 µg. However, in the two applications we reported here, CIF is compatible with the fractionation of low-input affinity-purified samples. Specifically, for low-input samples that are particularly sensitive to material loss during processing and which limits fractionation options, we demonstrate CIF retains the ability to efficiency fractionate samples and can improve data quality at those peptide concentration regimes. Based on our experience, fractionation of affinity-purified samples at the level of either cell compartment or peptide significantly improves acquisition of reproducible and biologically meaningful data (data not shown here).

Finally, we built a linear model that predicts the elution properties of a peptide based on its sequence. This model not only sheds light on the mechanism of underlying CMMB peptide-protein binding but also provides a tool for enriching peptides with particular properties. Because the current model assigns very high weights to the number of charged amino acid residuals and peptides that are charged under pH 7, we speculate that CIF might have utility for fractionation of phosphopeptides.

DATA AVAILABILITY

The raw proteomics data are deposited in the MassIVE data repository (https://massive.ucsd.edu) under the identifier MSV000085458. All peptide identification and quantitation data are included in the supporting information.

Author contributions—W. D. and J. A. W. designed the study and analyzed the data; W. D. and J. S. conducted the experiments; W. D., K. P., and J. A. W. wrote the article.

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: AP, affinity purification; CIF, CMMB-based isopropanol gradient peptide fractionation; CMMB, carboxylate-modified magnetic bead; PSM, peptide spectral match; RP, reverse phase; SA, streptavidin; SCX, strong cation exchange.

REFERENCES

1. McDonald, W. H., and Yates, J. R., 3rd. (2002) Shotgun proteomics and biomarker discovery. *Dis. Markers* **18**, 99–105
2. Yang, F., Shen, Y., and Camp, D. G. (2012) 2nd & Smith, R. D. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev. Proteomics* **9**, 129–134
3. Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–5690
4. Gilar, M., Olivova, P., Daly, A. E., and Gebler, J. C. (2005) Orthogonality of separation in two-dimensional liquid chromatography. *Anal. Chem.* **77**, 6426–6434
5. Alpert, A. J., and Andrews, P. C. (1988) Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J. Chromatogr.* **443**, 85–96
6. Bobaly, B., Mikola, V., Sipko, E., Marta, Z., and Fekete, J. (2015) Recovery of proteins affected by mobile phase trifluoroacetic acid concentration in reversed-phase chromatography. *J. Chromatogr. Sci.* **53**, 1078–1083
7. Dwivedi, R. C., Spicer, V., Harder, M., Antonovici, M., Ens, W., Standing, K. G., Wilkins, J. A., and Krokhin, O. V. (2008) Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal. Chem.* **80**, 7036–7042
8. Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., and Krijgsveld, J. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757
9. Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., and Krijgsveld, J. (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85
10. Moggridge, S., Sorensen, P. H., Morin, G. B., and Hughes, C. S. (2018) Extending the compatibility of the SP3 paramagnetic bead processing approach for proteomics. *J. Proteome Res.* **17**, 1730–1740

11. Hemstrom, P., and Irgum, K. (2006) Hydrophilic interaction chromatography. *J. Sep. Sci.* **29**, 1784–1821
12. Alpert, A. J. (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **80**, 62–76
13. Ikai, A. (1980) Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895–1898
14. Sillero, A., and Maldonado, A. (2006) Isoelectric point determination of proteins and other macromolecules: Oscillating method. *Comput. Biol. Med.* **36**, 157–166
15. Guruprasad, K., Reddy, B. V., and Pandit, M. W. (1990) Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng.* **4**, 155–161
16. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132
17. Engebretsen, S., and Bohlin, J. (2019) Statistical predictions with glmnet. *Clin. Epigenetics* **11**, 123
18. Alpert, A. J. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr.* **499**, 177–196
19. Hung, V., Udeshi, N. D., Lam, S. S., Loh, K. H., Cox, K. J., Pedram, K., Carr, S. A., and Ting, A. Y. (2016) Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **11**, 456–475
20. Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676

Supplemental Figure 1

1A
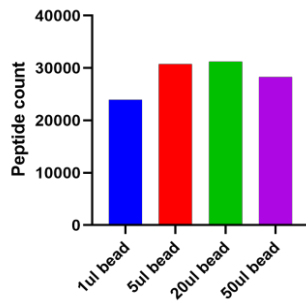
1B
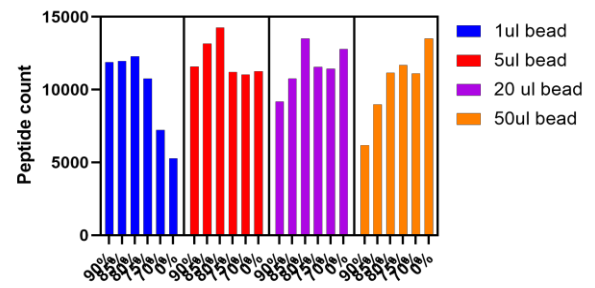
Supplemental Figure 1: (A) Total number of identified peptides in 20 µg CIF fractionated samples using 1 µl, 5 µl, 20 µl and 50 µl CMMB (n=1). **(B)** Number of peptides identified by each fraction of each bead amount CIF experiment.

Supplemental Figure 2
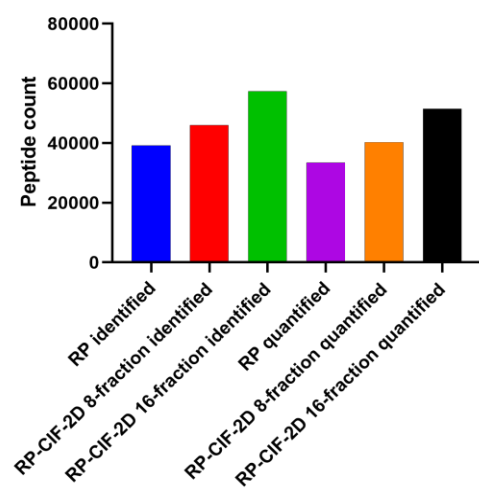
Supplemental Figure 2: Total number of identified peptides fractionated by RP, 8-fraction RP-CIF-2D, and 16-fraction RP-CIF-2D experiments (n=1).

Chapter 4

Neuronal activity regulates the nuclear proteome to promote activity-dependent transcription

ARTICLE

Neuronal activity regulates the nuclear proteome to promote activity-dependent transcription

Wendy A. Herbst[1,2], Weixian Deng[2], James A. Wohlschlegel[2], Jennifer M. Achiro[2], and Kelsey C. Martin[2]

The formation and plasticity of neuronal circuits relies on dynamic activity-dependent gene expression. Although recent work has revealed the identity of important transcriptional regulators and of genes that are transcribed and translated in response to activity, relatively little is known about the cell biological mechanisms by which activity alters the nuclear proteome of neurons to link neuronal stimulation to transcription. Using nucleus-specific proteomic mapping in silenced and stimulated neurons, we uncovered an understudied mechanism of nuclear proteome regulation: activity-dependent proteasome-mediated degradation. We found that the tumor suppressor protein PDCD4 undergoes rapid stimulus-induced degradation in the nucleus of neurons. We demonstrate that degradation of PDCD4 is required for normal activity-dependent transcription and that PDCD4 target genes include those encoding proteins critical for synapse formation, remodeling, and transmission. Our findings highlight the importance of the nuclear proteasome in regulating the activity-dependent nuclear proteome and point to a specific role for PDCD4 as a regulator of activity-dependent transcription in neurons.

Introduction

Stimulus-induced gene expression allows neurons to adapt their structure and function in response to a dynamically changing external environment (Yap and Greenberg, 2018; Holt et al., 2019; Gallegos et al., 2018). Activity-dependent transcription is critical to neural circuit function, from synapse formation during development (West and Greenberg, 2011; Flavell et al., 2006; Wayman et al., 2006; Polleux et al., 2007; Lin et al., 2008) to synaptic plasticity in the mature brain (Bloodgood et al., 2013; Chen et al., 2017; Ramanan et al., 2005; Tyssowski et al., 2018; Yap and Greenberg, 2018). Neuronal activity can regulate gene expression in the nucleus through chromatin modification and transcriptional regulation and in the cytoplasm through regulation of RNA localization, stability, and translation (Martin and Ephrussi, 2009). The extreme morphological polarity and compartmentalization of neurons pose a cell biological challenge to the activity-dependent regulation of gene expression: how are stimuli received at one site coupled to changes in gene expression within specific subcellular compartments? To produce activity-dependent changes in transcription, signals must be relayed from the site where the signal is received, at the synapse, to the nucleus. To better understand how neuronal activity is coupled with changes in transcription, we developed an assay to systematically identify activity-dependent changes in the nuclear proteome of neurons and thereby elucidate novel mechanisms by which neuronal activity alters the concentration of

specific proteins in the nucleus. To our knowledge, this study provides the first activity-dependent characterization of the nuclear proteome of neurons.

Neuronal activity can change the concentration of nuclear proteins via a variety of mechanisms, from nucleocytoplasmic shuttling of signaling proteins to synthesis and degradation of nuclear proteins (Ch'ng et al., 2012; Dieterich et al., 2008; Ma et al., 2014; Lin et al., 2008; Bayraktar et al., 2020; Upadhya et al., 2004). Although the activity-dependent transcriptome and translatome of neurons have been extensively characterized (Brigidi et al., 2019; Chen et al., 2017; Dörrbaum et al., 2020; Fernandez-Albert et al., 2019; Hrvatin et al., 2018; Lacar et al., 2016; Tyssowski et al., 2018), less work has been focused on the upstream changes that occur in the nucleus, specifically identifying the population of proteins that undergo activity-dependent changes in nuclear abundance due to regulated transport or stability, which can in turn influence transcription. In this study, we used nucleus-specific proteomic mapping to identify preexisting proteins that undergo activity-dependent changes in concentration in the nucleus. By characterizing stimulus-induced changes in the nuclear proteome, these data provide insight into how activity-dependent transcription is regulated.

Through our screen of nuclear proteins with activity-dependent changes in abundance, we discovered that programmed cell

death 4 (PDCD4) undergoes a significant reduction in nuclear concentration following neuronal stimulation. PDCD4 has been studied primarily in the context of cancer, where it has been found to function as a tumor suppressor and translational inhibitor in the cytoplasm (Matsuhashi et al., 2019; Wang and Yang, 2018; Yang et al., 2003). These studies have revealed that the abundance of PDCD4 protein is regulated at multiple levels, including via translation (Asangani et al., 2008; Frankel et al., 2008; Ning et al., 2014), proteasome-mediated degradation (Dorrello et al., 2006), and nucleocytoplasmic trafficking (Böhm et al., 2003), with decreases in PDCD4 correlating with invasion, proliferation, and metastasis of many types of cancer (Allgayer, 2010; Chen et al., 2003; Wang and Yang, 2018; Wei et al., 2012).

Despite being expressed at significant levels in the brain, especially in the hippocampus and cortex (Lein et al., 2007; Li et al., 2021), few studies have addressed the role of PDCD4 in the nervous system. PDCD4 expression in neurons is altered by injury and stress (Jiang et al., 2017; Narasimhan et al., 2013; Li et al., 2021), and recent work has shown that, as in cancer cells, PDCD4 may act as a translational repressor in neurons (Li et al., 2021; Di Paolo et al., 2020). However, the impact of neuronal activity on PDCD4 concentration and the function of nuclear, as opposed to cytoplasmic, PDCD4 remains unknown. Here, we show that neuronal activity stimulates PKC- and proteasome-mediated degradation of PDCD4 in the nucleus of neurons and that PDCD4 degradation is necessary for activity-dependent transcription.

In this study, we describe an assay that represents the first, to our knowledge, to identify activity-dependent changes in the nuclear proteome of neurons, and it does so in a manner that is independent of translation. Our results not only elucidate a novel mechanism by which activity can regulate the nuclear proteome but also support a role for the tumor suppressor protein PDCD4 during activity-dependent transcription in neurons.

Results

Identification of the nuclear proteome from silenced and stimulated neurons using APEX2 proximity biotinylation and mass spectrometry (MS)

To identify proteins that undergo activity-dependent changes in nuclear localization or abundance, we analyzed the nuclear proteomes of silenced and stimulated cultured rat forebrain neurons. In developing this assay, we used cAMP-response element binding protein–regulated transcriptional coactivator 1 (CRTC1) as a positive control because we have previously shown that glutamatergic activity drives the synapse-to-nucleus import of CRTC1 and that neuronal silencing decreases CRTC1 nuclear abundance (Ch'ng et al., 2012, 2015). We initially used nuclear fractionation to capture the nuclear proteins but discovered that CRTC1 leaked out of the nucleus during the assay. This suggested that nuclear fractionation was not a suitable method and led us to instead use ascorbate peroxidase 2 (APEX2) proximity biotinylation (Hung et al., 2016), an in situ proximity ligation assay, to identify activity-dependent changes in the nuclear proteome. To specifically label the nuclear proteome, we fused APEX2 to two SV40 NLSs (Kalderon et al., 1984; Fig. 1 A). APEX2 proximity ligation was advantageous for these experiments for the following reasons: (1) APEX2 biotinylated proteins can be captured directly by streptavidin pulldown, avoiding the need for subcellular fractionation; (2) biotinylation occurs rapidly (1-min labeling period); and (3) APEX2 can be expressed in a specific cell type of interest. We designed a neuron-specific nuclear-localized APEX2 construct (Fig. 1 A) and transduced cultured rat forebrain neurons with adeno-associated virus (AAV) expressing APEX2-NLS. Immunofluorescence of transduced neurons revealed that APEX2-NLS was expressed specifically in the nucleus (Fig. 1 B). We optimized the MOI of AAV to achieve high transduction efficiency without overexpression of the construct (important because higher doses of AAV led to APEX2-NLS expression in the cytoplasm). When all three components of the labeling reaction were supplied (APEX2-NLS, biotin-phenol, and H_2O_2), proteins were biotinylated specifically in neuronal nuclei (Fig. 1 B). No labeling was detected in the absence of APEX2-NLS, biotin-phenol, or H_2O_2.

To identify proteins that undergo activity-dependent changes in nuclear abundance, we silenced neurons for 1 h with the voltage-gated sodium channel antagonist tetrodotoxin (TTX) or stimulated neurons for 1 h with bicuculline (Bic), which inhibits $GABA_A$ receptors and drives glutamatergic transmission. We also inhibited protein synthesis using cycloheximide (CHX) in these experiments because many of the genes that are rapidly transcribed and translated in response to activity encode nuclear proteins (Yap and Greenberg, 2018; Heinz and Bloodgood, 2020; Alberini, 2009). We were concerned that the translation of activity-dependent genes would overshadow, and thereby hinder the detection of, the upstream changes that occur in the nuclear proteome resulting from alterations in nuclear protein localization or stability. We confirmed that the CHX treatment used throughout this study impairs protein translation, as detected by the strong reduction of new protein synthesis using L-azidohomoalanine (AHA) labeling (Fig. S1, A and B) as well as the complete inhibition of activity-dependent FOS induction (Fig. S1, C and D).

After the neurons were silenced or stimulated for 1 h and the 1-min labeling reaction was performed (Fig. 1 C), protein lysates were collected for analysis by Western blotting and MS. In neurons expressing APEX2-NLS, many proteins at a variety of molecular weights were biotinylated in both TTX and Bic conditions, whereas very few proteins were biotinylated in the no-APEX control, as detected by Western blot analysis (Fig. 1 D). The bands detected in the no-APEX control were at molecular weights of known endogenously biotinylated proteins (Hung et al., 2016). For all experiments, we also confirmed by immunocytochemistry (ICC) that the APEX2-NLS construct was localized to the nucleus during both experimental conditions (Fig. S1, E and F). For MS, biotinylated proteins were captured using streptavidin pulldown, and the nuclear proteomes were characterized using the tandem mass tag–MS/MS/MS–synchronous precursor selection (TMT-MS3-SPS) acquisition method (Ting et al., 2011) through liquid chromatography–MS. We detected 4,407 proteins, and, of those, 2,860 proteins were significantly enriched above the no-APEX negative control with log_2 fold
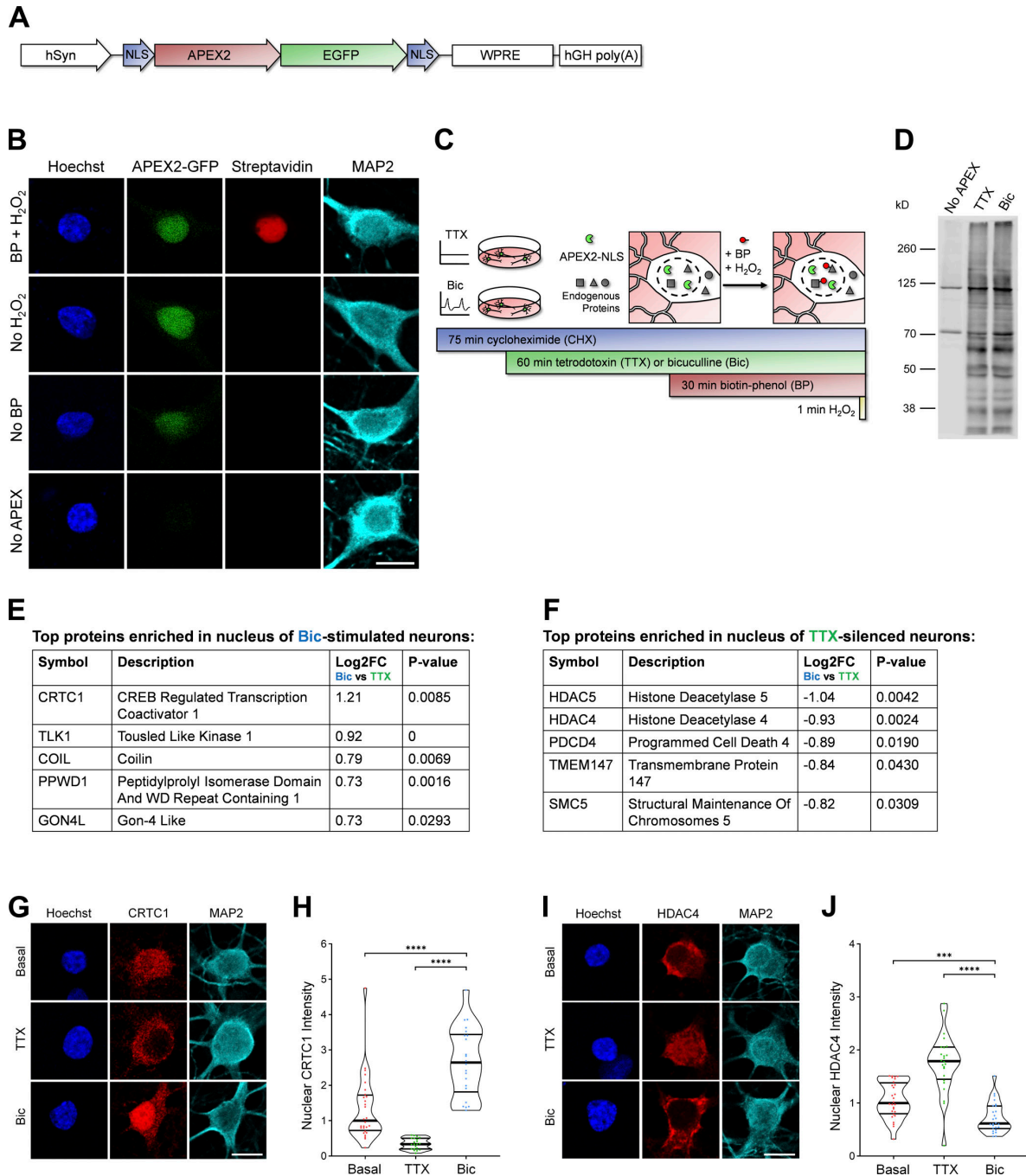
Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology   2 of 19
https://doi.org/10.1083/jcb.202103087

68

Figure 1. **Identification of the nuclear proteomes from silenced and stimulated neurons using APEX2 proximity biotinylation. (A)** Design of APEX2-NLS construct. hSyn, human synapsin promoter; NLS, SV40 NLS; WPRE, woodchuck hepatitis virus posttranscriptional regulatory element; hGH poly(A), human growth hormone polyadenylation sequence. **(B)** ICC of cultured rat forebrain neurons after APEX2 proximity biotinylation labeling. Nuclear proteins were biotinylated (streptavidin; red) by the combined presence of APEX2-NLS (GFP; green), biotin-phenol (BP), and H_2O_2. Scale bar, 10 µm. **(C)** Workflow for labeling nuclear proteins from silenced and stimulated neurons. APEX2 labeling diagram based on Hung et al. (2016). **(D)** Western blot of cultured neuron protein lysates from no-APEX, APEX+TTX, or APEX+Bic conditions, stained with streptavidin. **(E)** The top five proteins enriched in the nucleus of Bic-stimulated neurons, as ranked by Bic versus TTX log₂ FC. **(F)** The top five proteins enriched in the nucleus of TTX-silenced neurons, as ranked by Bic versus TTX log₂ FC. **(G)** CRTC1 ICC of basal, TTX-silenced, and Bic-stimulated neurons. Scale bar, 10 µm. **(H)** Violin plots of normalized nuclear CRTC1 ICC intensity. Basal, $n = 28$;

Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology 3 of 19
https://doi.org/10.1083/jcb.202103087

69

change (FC) >3 and adjusted P value <0.05 (Table S1). To assess nuclear enrichment of the samples, we performed gene ontology (GO) cellular component analysis on the list of 2,860 proteins that were enriched above no-APEX (which could be matched to 2,615 unique gene identifiers). Approximately 80% of the top 100 most abundant proteins detected in the study contained the GO term "nucleus," and >60% of the entire list of 2,615 proteins contained the GO term nucleus (Fig. S1 G). We performed GO cellular component enrichment analysis and found that most of the top enriched terms were components of the nucleus (Fig. S1 H and Table S2).

When comparing the Bic and TTX conditions, 23 proteins were differentially expressed in the nucleus with $\log_2$ FC >0.5 or $\log_2$ FC less than −0.5 and P value <0.05 (Table S1). The highest-ranked protein by $\log_2$ FC enriched in the Bic versus TTX nuclear proteome was the synapse-to-nucleus signaling protein, CRTC1 (Ch'ng et al., 2015; Nonaka et al., 2014; Ch'ng et al., 2012; Sekeres et al., 2012; Fig. 1 E). The highest-ranked proteins enriched in the TTX versus Bic nuclear proteome were histone deacetylase 4 (HDAC4) and HDAC5, both of which have been reported to undergo nuclear export following neuronal stimulation (Chawla et al., 2003; Schlumm et al., 2013; Fig. 1 F). Using ICC, we confirmed that CRTC1 increased in the nucleus and HDAC4 decreased in the nucleus following Bic stimulation (Fig. 1, G–J). These findings demonstrate that the nucleus-specific APEX2 proximity ligation assay successfully captured activity-dependent changes in the nuclear proteome of neurons.

Neuronal stimulation decreases PDCD4 protein concentration in the nucleus and cytoplasm of neurons
Among the proteins that underwent activity-dependent changes in nuclear concentration, the PDCD4 protein was significantly enriched in the TTX-treated nuclear proteome compared with the Bic-treated nuclear proteome. To validate this finding, we characterized the expression of PDCD4 protein in cultured neurons using ICC, and found that PDCD4 was present in both the nucleus and cytoplasm of neurons (Fig. 2 A and Fig. S2 A). Bic stimulation significantly decreased PDCD4 protein expression in the nucleus by ~50% (Fig. 2, A and B), with a smaller decrease of ~20% in the cytoplasm (Fig. 2 C). The decrease of PDCD4 occurred within 15 min of Bic stimulation, and PDCD4 protein levels continued to decrease further with longer incubations of Bic up to 3 h (Fig. 2 D). After washout of a 1-h Bic stimulation, PDCD4 protein expression gradually returned toward baseline levels, although PDCD4 protein concentration was still below baseline 24 h after washout (Fig. 2 E).

In complementary experiments, we transduced neurons with AAV expressing C-terminal HA-tagged PDCD4 (Fig. S2 B) and characterized PDCD4-HA expression by Western blot analysis (Fig. 2 F) and ICC (Fig. S2, C and D). By Western blot analysis, we found that total PDCD4-HA protein levels decreased by ~50%

following Bic stimulation (Fig. 2 F). By ICC, both nuclear and cytoplasmic PDCD4-HA decreased by ~40% (Fig. S2, C and D). To further validate the Bic-induced decrease in PDCD4, we also created an N-terminal V5-tagged PDCD4 plasmid (Fig. S2 B) and transfected the construct in neurons. Consistent with the results from endogenous PDCD4 and transduced C-terminally HA-tagged PDCD4, both nuclear and cytoplasmic V5-PDCD4 decreased by ~40% with Bic stimulation (Fig. S2, E and F). These results show that exogenously expressed PDCD4 concentration in neurons is regulated by activity and that the decrease in PDCD4 represents a decrease in the protein rather than cleavage into multiple fragments.

We also found that depolarization of neurons with 40 mM KCl for 5 min significantly decreased PDCD4 protein concentration by ~40% in the nucleus (Fig. 2 G) and ~30% in the cytoplasm (Fig. 2 H), as detected immediately after the 5-min treatment. Altogether, these results indicate that increases in glutamatergic transmission and depolarization lead to a rapid and long-lasting reduction in PDCD4 abundance in the nucleus and, to a lesser extent, the cytoplasm of neurons.

PDCD4 undergoes proteasome-mediated degradation following neuronal stimulation
In nonneuronal cells, PDCD4 has been reported to undergo miRNA-mediated translational repression (Asangani et al., 2008; Frankel et al., 2008; Ning et al., 2014), stimulus-induced nuclear export (Böhm et al., 2003), and stimulus-induced proteasome-mediated degradation (Dorrello et al., 2006). We sought to investigate the mechanisms of the neuronal activity-dependent decrease of PDCD4 (Fig. 3 and Fig. S3). Because the assay we used to detect activity-dependent changes in the nuclear proteome was conducted in the presence of the protein synthesis inhibitor CHX, we considered it unlikely that the Bic-induced decrease in PDCD4 was due to miRNA-mediated translational repression. To confirm this, we conducted PDCD4 ICC of TTX-silenced and Bic-stimulated neurons in the presence or absence of CHX and found that CHX did not block the Bic-induced decrease in PDCD4 (Fig. S3, A and B), thereby ruling out a role for activity-dependent miRNA-mediated translational regulation.

To investigate the mechanism underlying the decrease in nuclear PDCD4, we tested if the decrease in nuclear abundance was due to activity-dependent increases in nuclear export. Nuclear export of PDCD4 is mediated by the nuclear export protein, CRM1, and sensitive to the nuclear export inhibitor, leptomycin B (LMB; Böhm et al., 2003). Long incubation with LMB successfully caused nuclear accumulation of PDCD4 in unstimulated neurons (Fig. S3 C); however, LMB was unable to prevent the activity-dependent decrease of nuclear PDCD4 following stimulation (Fig. 3 A and Fig. S3 D). This result demonstrates that regulated nuclear export is not required for the activity-dependent decrease of PDCD4.
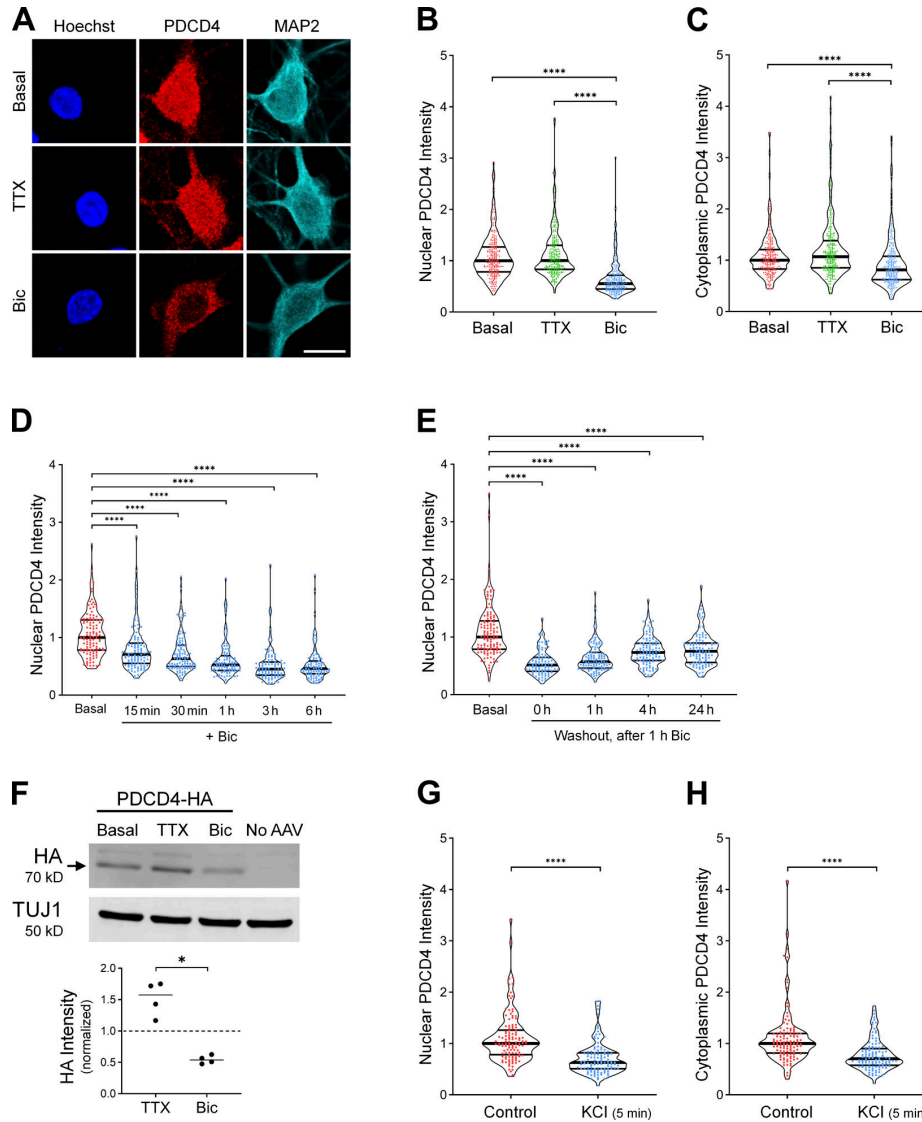
Figure 2. **Neuronal stimulation decreases PDCD4 protein concentration in the nucleus and cytoplasm of neurons. (A)** PDCD4 ICC of basal, TTX-silenced, and Bic-stimulated neurons. Scale bar, 10 µm. **(B)** Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, n = 226; TTX n = 227; and Bic n = 218 cells from 6 sets of cultures. Basal median = 1.00; TTX median = 1.00; Bic median = 0.55. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001. **(C)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in B. Basal median = 1.00; TTX median = 1.07; Bic median = 0.81. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001. **(D)** Violin plots of normalized nuclear PDCD4 ICC intensity after varying durations of Bic stimulation. Basal, n = 115; Bic 15 min, n = 111; Bic 30 min, n = 102; Bic 1 h, n = 97; Bic 3 h, n = 92; and Bic 6 h, n = 91 cells from three sets of cultures. Basal median = 1.00; Bic 15 min median = 0.71; Bic 30 min median = 0.63; Bic 1 h median = 0.52; Bic 3 h median = 0.45; Bic 6 h median = 0.46. Basal versus Bic 15 min, P < 0.0001; basal versus Bic 30 min, P < 0.0001; basal versus Bic 1 h, P < 0.0001; basal versus Bic 3 h, P < 0.0001; basal versus Bic 6 h, P < 0.0001. **(E)** Violin plots of normalized nuclear PDCD4 ICC intensity at various time points after removal of a 1-h Bic stimulation. Basal, n = 124; washout 0 h, n = 94; washout 1 h, n = 111; washout 4 h, n = 104; and washout 24 h, n = 99 cells from 3 sets of cultures. Basal median = 1.00; washout 0 h median = 0.51; washout 1 h median = 0.57; washout 4 h median = 0.73; washout 24 h median = 0.75. Basal versus 0 h, P < 0.0001; basal versus 1 h, P < 0.0001; basal versus 4 h, P < 0.0001; basal versus 24 h, P < 0.0001. **(F)** Top: Western blot of protein lysates from basal, TTX-silenced, and Bic-stimulated neurons transduced with PDCD4-HA AAV. The PDCD4-HA band is indicated with an arrow. We observed a faint nonspecific band above the HA band. Bottom: Quantification of Western blot from four sets of cultures. HA intensity was normalized to TUJ1 intensity. Within each experiment, all samples were normalized to the basal sample. TTX/basal median = 1.57; Bic/basal median = 0.54. TTX versus Bic, P = 0.0286. **(G)** Violin plots of normalized nuclear PDCD4 ICC intensity. Control n = 118 and KCl n = 110 cells from three sets of cultures. Control median = 1.00; KCl median = 0.63. Control versus KCl, P < 0.0001. **(H)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in G. Control median = 1.00; KCl median = 0.70. Control versus KCl, P < 0.0001. *, P < 0.05; ****, P < 0.0001; Mann-Whitney U test with Bonferroni correction.
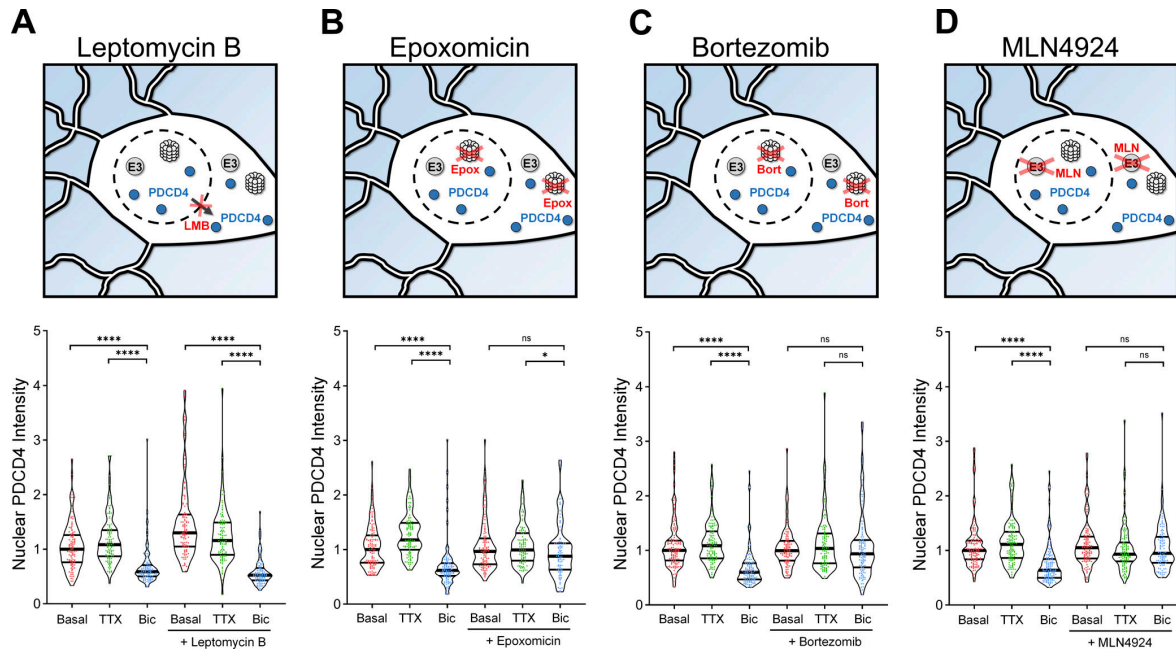
71

Figure 3. **PDCD4 undergoes proteasome-mediated degradation, not nuclear export, following neuronal stimulation. (A)** Top: Schematic of CRM1-mediated nuclear export inhibitor LMB experiments. Bottom: Violin plots of normalized PDCD4 ICC intensity. Basal, n = 109; TTX, n = 109; Bic, n = 109; LMB-basal, n = 106; LMB-TTX, n = 100; and LMB-Bic, n = 86 cells from 3 sets of cultures. Basal median = 1.00; TTX median = 1.08; Bic median = 0.59; LMB-basal median = 1.30; LMB-TTX median = 1.16; LMB-Bic median = 0.52. Basal versus Bic, $P < 0.0001$; TTX versus Bic, $P < 0.0001$; LMB-basal versus LMB-Bic, $P < 0.0001$; LMB-TTX versus LMB-Bic, $P < 0.0001$. **(B)** Top: Schematic of proteasome inhibitor Epox experiments. Bottom: Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, n = 113; TTX, n = 106; Bic, n = 103; Epox-basal, n = 107; Epox-TTX, n = 94; and Epox-Bic, n = 86 cells from 3 sets of cultures. Basal median = 1.00; TTX median = 1.18; Bic median = 0.62; Epox-basal median = 0.97; Epox-TTX median = 0.99; Epox-Bic median = 0.88. Basal versus Bic, $P < 0.0001$; TTX versus Bic, $P < 0.0001$; Epox-basal versus Epox-Bic, $P = 0.1662$; Epox-TTX versus Epox-Bic, $P = 0.0174$. **(C)** Top: Schematic of proteasome inhibitor Bort experiments. Bottom: Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, n = 131; TTX, n = 111; Bic, n = 120; Bort-basal, n = 100; Bort-TTX, n = 98; and Bort-Bic, n = 116 cells from three sets of cultures. Basal median = 1.00; TTX median = 1.09; Bic median = 0.60; Bort-basal median = 1.00; Bort-TTX median = 1.04; Bort-Bic median = 0.94. Basal versus Bic, $P < 0.0001$; TTX versus Bic, $P < 0.0001$; Bort-basal versus Bort-Bic, $P = 0.6224$; Bort-TTX versus Bort-Bic, $P = 0.2648$. **(D)** Top: Schematic of MLN experiments. MLN inhibits neddylation, preventing the activation of Cullin-RING E3 ubiquitin ligases. Bottom: Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, n = 130; TTX, n = 120; Bic, n = 120; MLN-basal, n = 115; MLN-TTX, n = 108; and MLN-Bic, n = 97 cells from three sets of cultures. Basal median = 1.00; TTX median = 1.11; Bic median = 0.64; MLN-basal median = 1.05; MLN-TTX median = 0.93; MLN-Bic median = 0.95. Basal versus Bic, $P < 0.0001$; TTX versus Bic, $P < 0.0001$; MLN-basal versus MLN-Bic, $P = 0.329$; MLN-TTX versus MLN-Bic, $P = 1$. *, $P < 0.05$; ****, $P < 0.0001$; Mann-Whitney U test with Bonferroni correction.

We next hypothesized that regulated ubiquitin proteasome-mediated degradation may explain the Bic-induced decrease of PDCD4. To test this idea, we incubated TTX-silenced and Bic-stimulated neurons with the proteasome inhibitors epoxomicin (Epox) or bortezomib (Bort). The proteasome inhibitors impaired Bic-induced decreases of PDCD4 in both the nucleus and cytoplasm of neurons (Fig. 3, B and C; and Fig. S3, E and F), indicating that neuronal activity decreases PDCD4 concentrations via proteasome-mediated degradation. The finding that the nuclear export inhibitor LMB did not block the Bic-induced decrease of PDCD4 in the nucleus (Fig. 3 A) indicates that activity regulates proteasome-mediated degradation of nuclear PDCD4 directly in the nucleus, rather than by nuclear export of PDCD4 followed by degradation in the cytoplasm.

The E3 ubiquitin ligases β-transducin repeat-containing protein 1/2 (βTRCP1/2) have been shown to be required for the proteasome-mediated degradation of PDCD4 in the T98G glioblastoma cell line (Dorrello et al., 2006). βTRCP1/2 belong to

the family of Cullin-RING E3 ubiquitin ligases, which require neddylation in order to be activated (Merlet et al., 2009). To determine if this family of ligases is involved in the activity-dependent decrease of PDCD4 in neurons, we used the neddylation inhibitor MLN4924 (MLN) and found that it blocked the Bic-induced decrease of PDCD4 in the nucleus (Fig. 3 D) and cytoplasm (Fig. S3 G) of neurons. This result further supports the finding that PDCD4 undergoes proteasome-mediated degradation following Bic stimulation, likely through ubiquitination by βTRCP1/2.

PDCD4 S71A mutation and PKC inhibition prevent the activity-dependent decrease of PDCD4

To understand how neuronal activity at synapses leads to proteasome-mediated degradation of PDCD4 in the nucleus, we focused on PDCD4's βTRCP-binding motif. Previous work in T98G glioblastoma cells has shown that phosphorylation of Ser67, Ser71, and/or Ser76 in the βTRCP-binding region of

PDCD4 is necessary for βTRCP binding and subsequent proteasome-mediated degradation of PDCD4 (Dorrello et al., 2006). To test if a mutation in PDCD4 at one of these sites would prevent the activity-dependent decrease of PDCD4 in neurons, we expressed WT and mutant (S71A) PDCD4-HA in cultured neurons using AAV (Fig. 4, A and B). Similar to endogenous PDCD4, we found that WT PDCD4-HA decreased following Bic stimulation, with an ~30% decrease in nuclear HA intensity and an ~15% decrease in cytoplasmic intensity, as detected by ICC (Fig. 4 C and Fig. S3 H). In contrast, the PDCD4-HA S71A mutant did not undergo an activity-dependent decrease in either the nucleus or cytoplasm, but showed a slight activity-dependent increase in nuclear intensity (Fig. 4 C). In complementary experiments, we found that Bic stimulation resulted in an ~40% decrease of WT PDCD4-HA signal by Western blot analysis, whereas PDCD4-HA S71A did not decrease after stimulation (Fig. 4 D). These results suggest that phosphorylation of S71 in PDCD4's βTRCP-binding motif is necessary for its activity-dependent degradation in neurons.

The βTRCP-binding region in PDCD4 contains canonical phosphorylation consensus sites for the kinases S6K1 and PKC (Fig. 4 B); Ser71 is a consensus site for PKC, and Ser67 and Ser76 are consensus sites for S6K1 (Dorrello et al., 2006; Matsuhashi et al., 2014, 2019; Nakashima et al., 2010). To test which kinase phosphorylates PDCD4 with neuronal activity, we incubated cultures in the S6K inhibitor LY2584702 or the PKC inhibitor Ro-31-8425. We found that the S6K inhibitor LY2584702 did not prevent the Bic-induced decrease of PDCD4 (Fig. 4 E and Fig. S3 I), despite its ability to potently inhibit the phosphorylation of a known S6K target, ribosomal protein S6 (Fig. S3 J). In contrast, the pan-PKC inhibitor Ro-31-8425 completely prevented the activity-dependent decrease (and slightly increased nuclear intensity) of PDCD4 (Fig. 4 F and Fig. S3 K). These results suggest that in response to glutamatergic activity, PKC phosphorylates PDCD4 at Ser71, which enables the ubiquitin ligase βTRCP1/2 to bind and promote the proteasome-mediated degradation of PDCD4.

Stimulus-induced degradation of PDCD4 regulates the expression of neuronal activity-dependent genes

The finding that synaptic activity dynamically regulates PDCD4 protein concentration in the nucleus in response to synaptic activity points to a nuclear function for PDCD4 in coupling neuronal activity with changes in transcription. To investigate a role for PDCD4 in the regulation of activity-dependent transcription, we performed RNA sequencing (RNA-seq) of forebrain cultures transduced with either WT PDCD4 or degradation-resistant PDCD4 (S71A) following neuronal silencing with TTX or stimulation with Bic for 1 h (Fig. 5 and Table S3). Given that PDCD4 has a well-known role in regulating translation, we sought to distinguish between direct PDCD4-mediated transcriptional changes in the nucleus and the changes in expression that are downstream of PDCD4-mediated translational changes in the cytoplasm by performing the experiments in either the presence or absence of CHX. We detected robust Bic-induced increases in normalized read counts of transcripts for canonical immediate early genes such as Npas4, Rgs2, and Egr4 in

all biological replicates (WT Bic versus TTX: Npas4 $\log_2$ FC = 6.5; Rgs2 $\log_2$ FC = 2.4; Egr4 $\log_2$ FC = 3.1; Fig. 5 A and Table S3). We identified 912 activity-dependent genes, defined as genes with significant differential expression between Bic and TTX for PDCD4 WT samples (459 up-regulated, 453 down-regulated; adjusted P value <0.05 for WT no CHX; Fig. 5, B–D). Clustering of activity-dependent genes by FC across sample types revealed that most activity-dependent genes showed similar FCs between PDCD4 WT and PDCD4 S71A samples (Fig. 5 B), especially for genes with relatively high activity-dependent FCs (Fig. 5 C). However, Fig. 5 B also shows that PDCD4 S71A altered activity-dependent changes in gene expression for a subset of genes. Specifically, we found that PDCD4 S71A led to a decrease in activity-induced differential expression for a substantial proportion of genes: 43% of activity-dependent up-regulated genes (198 genes) and 57% of activity-dependent down-regulated genes (260 genes) were not significantly up-regulated or down-regulated, respectively, in the PDCD4 S71A samples (Fig. 5 D). These results suggest that regulated degradation of PDCD4 is important for the expression of activity-dependent genes in neurons.

This inhibition of activity-dependent gene expression could be due to both a potential role for PDCD4 in transcriptional processes and secondary effects from PDCD4's regulation of translation of specific transcripts (Matsuhashi et al., 2019; Wang and Yang, 2018). To isolate effects at the transcription level, we focused on CHX-insensitive activity-dependent genes, that is, genes that showed activity-dependent differential expression in both the presence and absence of CHX (in WT, 459 genes after excluding 3 genes that showed differential expression in different directions with or without CHX; Fig. 6). We ranked CHX-insensitive genes by their change in activity-dependent FC between PDCD4 WT and PDCD4 S71A samples and identified 91 putative PDCD4 target genes that showed large differences in activity-dependent gene expression between WT and degradation-resistant PDCD4 samples (see Materials and methods; Fig. 6 A). We validated the effect of PDCD4 on activity-dependent gene expression with quantitative RT-PCR (RT-qPCR) for two of the genes with the largest change between PDCD4 WT and PDCD4 S71A (Scd1 and Thrsp; Fig. S4 A). To confirm a role for PDCD4 in regulating gene expression, we tested the effect of PDCD4 knockdown and overexpression on the transcript abundance of these putative PDCD4 target genes. We found that PDCD4 siRNA knockdown of endogenous PDCD4 significantly up-regulated the expression of Scd1 and Thrsp (Fig. S4 B), demonstrating that endogenous PDCD4 represses the expression of these target genes. Overexpression of PDCD4 using PDCD4 AAV had a smaller effect on the target genes, significantly down-regulating Thrsp but only slightly and nonsignificantly down-regulating Scd1 (Fig. S4 C). These findings support a role for PDCD4 in gene expression regulation and demonstrate that endogenous levels of PDCD4 are sufficient to repress the expression of target genes.

We performed motif analysis of promoter sequences of the putative PDCD4 target genes and found motifs similar to those in promoters of other CHX-insensitive activity-dependent genes (e.g., AP-1/TRE, ATF/CRE, Sp1/Klf motifs; Fig. S5), suggesting that there was not a specific transcription factor motif associated
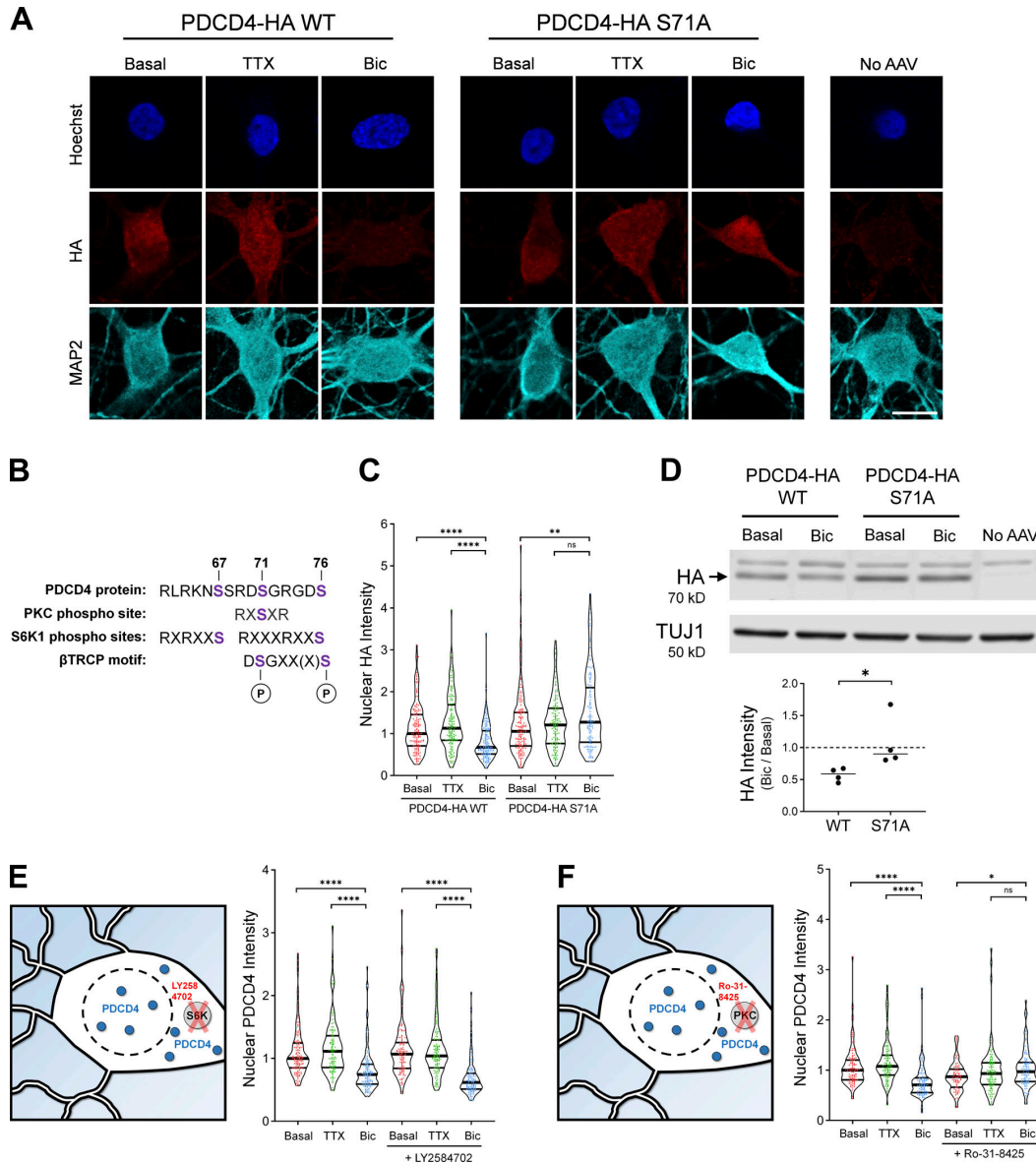
Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology   7 of 19
https://doi.org/10.1083/jcb.202103087

73

Figure 4. **PDCD4 S71A mutation and PKC inhibition prevent the activity-dependent decrease of PDCD4. (A)** HA ICC of basal, TTX-silenced, and Bic-stimulated neurons transduced with PDCD4-HA WT, PDCD4-HA S71A AAV, or no AAV (negative control). Scale bar, 10 μm. **(B)** PDCD4 protein sequence (aa 62–76). PKC and S6K1 phosphorylation sites are indicated in purple. Adapted from Matsuhashi et al. (2019). **(C)** Violin plots of normalized nuclear HA ICC intensity. WT-basal, n = 144; WT-TTX, n = 136; WT-Bic, n = 147; S71A-basal, n = 158; S71A-TTX, n = 140; and S71A-Bic, n = 122 cells from four sets of cultures. WT-basal median = 1.00; WT-TTX median = 1.13; WT-Bic median = 0.67; S71A-basal median = 1.05; S71A-TTX median = 1.21; S71A-Bic median = 1.27. WT-basal versus WT-Bic, P < 0.0001; WT-TTX versus WT-Bic, P < 0.0001; S71A-basal versus S71A-Bic, P = 0.0034; S71A-TTX versus S71A-Bic, P = 0.14. **(D)** Top: Western blot of protein lysates from basal and Bic-stimulated neurons transduced with PDCD4-HA WT or PDCD4-HA S71A. The PDCD4-HA band is indicated with an arrow. We observed a nonspecific band above the HA band. Bottom: Quantification of Western blot from four sets of cultures. HA intensity was normalized to TUJ1 intensity. Within each experiment, each Bic sample was normalized to its respective basal sample. WT Bic/Basal median = 0.59; S71A Bic/basal median = 0.90. WT versus S71A, P = 0.0286. **(E)** Left: Schematic of S6K inhibitor Ly2584702 (LY) experiments. Right: Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, n = 138; TTX, n = 104; Bic, n = 122, LY-basal, n = 112; LY-TTX, n = 112; and LY-Bic, n = 107 cells from 3 sets of cultures. Basal median = 1.00; TTX median = 1.11; Bic median = 0.74; LY-basal median = 1.07; LY-TTX median = 1.04; LY-Bic median = 0.62. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001; LY-basal versus LY-Bic, P < 0.0001; LY-TTX versus LY-Bic, P < 0.0001. **(F)** Left: Schematic of PKC inhibitor Ro-31-8425 (Ro) experiments. Right: Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, n = 101; TTX, n = 101; Bic, n = 96; Ro-basal, n = 81; Ro-TTX, n = 95; and Ro-Bic, n = 87 cells from three sets of cultures. Basal median = 1.00; TTX median = 1.08; Bic median = 0.71; Ro-basal median = 0.87; Ro-TTX median = 0.93; Ro-Bic median = 0.97. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001; Ro-basal versus Ro-Bic, P = 0.0114; Ro-TTX versus Ro-Bic, P = 0.3892. *, P < 0.05; **, P < 0.01; and ****, P < 0.0001; Mann-Whitney U test with Bonferroni correction.

74

A

B

Activity-dependent genes

C

Activity-dependent genes

D

Up with activity in **WT** (459 genes)

Down with activity in **WT** (453 genes)

- ■ up with activity in **S71A**
- ■ no change with activity in **S71A**
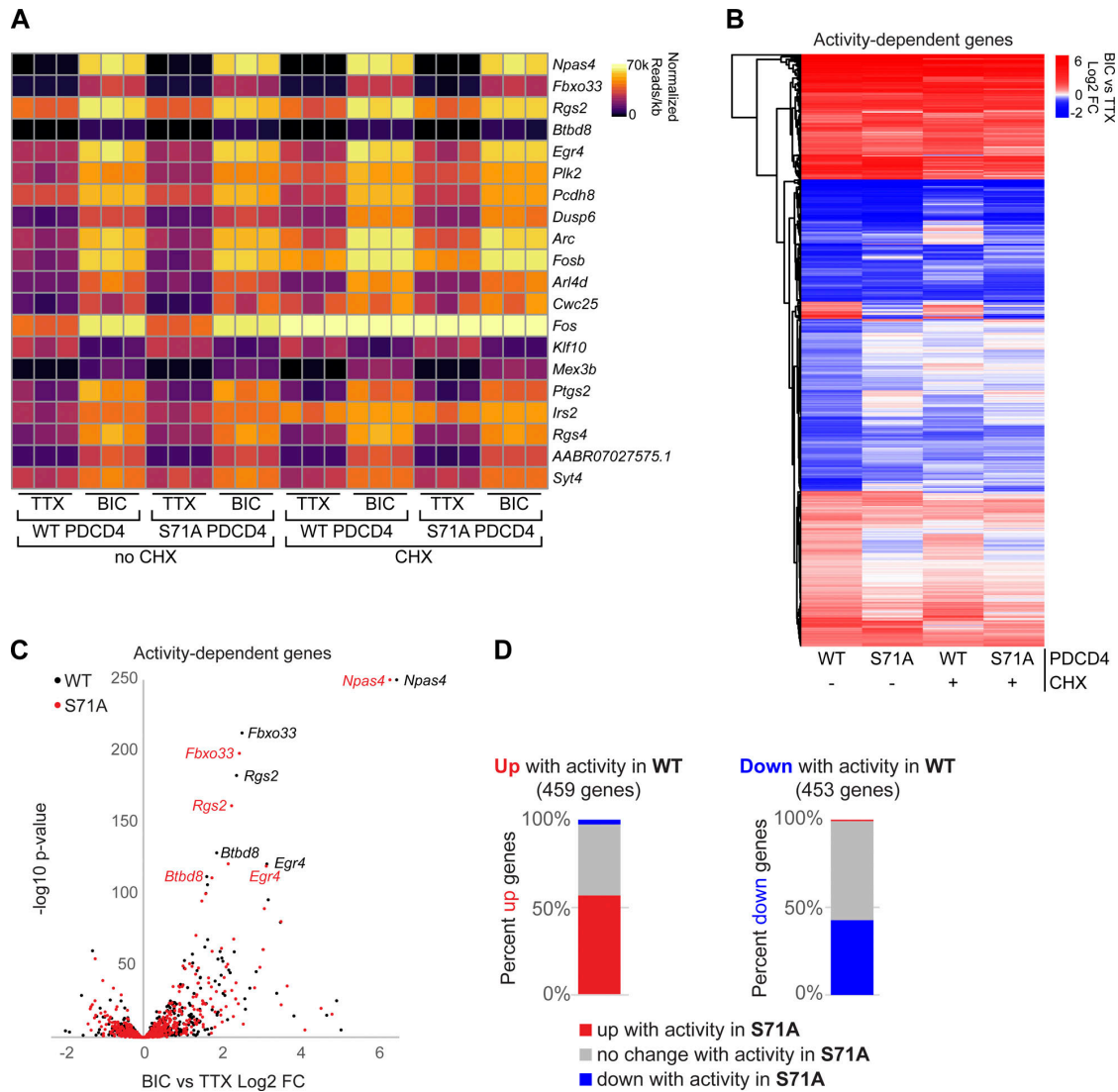- ■ down with activity in **S71A**

Figure 5. **Stimulus-induced degradation of PDCD4 regulates the expression of neuronal activity–dependent genes. (A)** For each biological replicate, normalized read counts (from DESeq2) divided by transcript length are shown for the top 20 activity-dependent genes, ranked by adjusted P value for PDCD4 WT no-CHX samples. Each row represents a gene, and each column represents a biological replicate. The color of each box indicates transcript abundance (note: color is not scaled linearly in order to display full range of read counts; see Table S3 for full dataset). **(B)** Stimulation-induced $\log_2$ FC for all 912 activity-dependent genes, clustered by FC across sample type. Each row represents a gene, and each column represents a sample type. The color legend represents Bic versus TTX $\log_2$ FC, with red representing up-regulation, blue representing down-regulation, and white indicating $\log_2$ FC of 0. **(C)** For each activity-dependent gene, Bic versus TTX $\log_2$ FC is plotted against $-\log_{10}$ of adjusted P value from PDCD4 WT samples (black) and PDCD4 S71A samples (red). Gene names for the top five activity-dependent genes by adjusted P value are labeled. For both PDCD4 WT and PDCD4 S71A samples, *Npas4* Bic versus TTX adjusted P value was 0 ($-\log_{10}$ of 0 is not defined); therefore, for display, the $-\log_{10}$ adjusted P value for *Npas4* was set to 250 for both samples. **(D)** Activity-dependent up-regulated genes (left bar) and activity-dependent down-regulated genes (right bar) were categorized by the activity-dependent differential expression in PDCD4 S71A samples. The colors in each bar show the percentage of activity-dependent genes showing activity-dependent up-regulation (red), no change (gray), or down-regulation (blue) in PDCD4 S71A samples.

with putative PDCD4 target genes. GO analysis of the putative PDCD4 target genes showed enrichment for neuronal signaling terms, such as "nervous system development" (GO:0007399; 28 genes; false discovery rate [FDR] = 9.28E-04) and "synapse" (GO:0045202; 18 genes; FDR = 6.54E-03), whereas, for comparison,

other CHX-insensitive activity-dependent genes showed enrichment for terms related to transcription, such as "regulation of gene expression" (GO:0010468; 169 genes; FDR = 3.11E-25) and "nuclear chromosome" (GO:0000228; 51 genes; FDR = 2.88E-09; Fig. 6 B and Table S4). Putative PDCD4 targets included genes
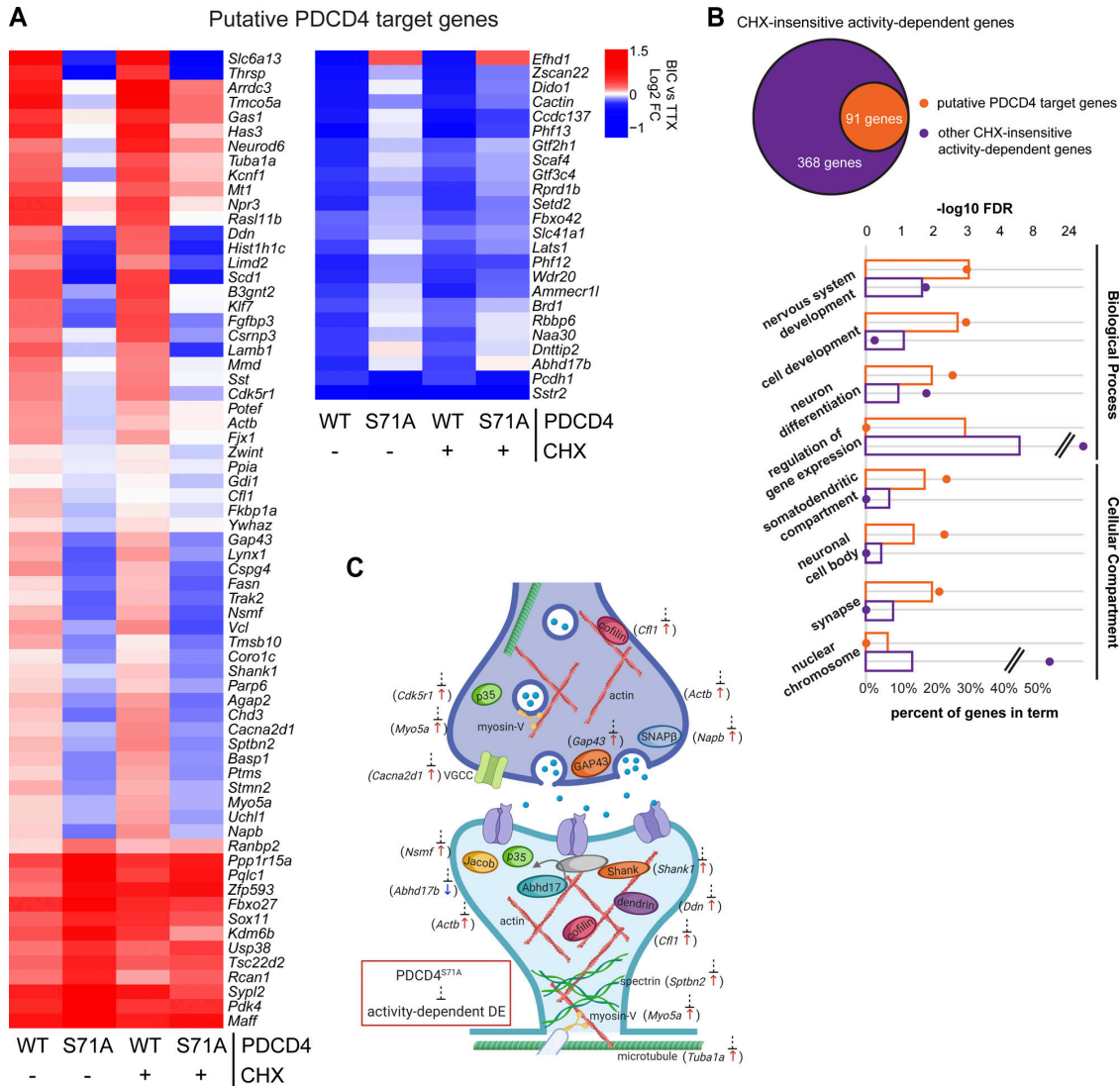
75

Figure 6. **Degradation-resistant PDCD4 suppresses activity-dependent changes in expression of synaptic genes. (A)** Bic versus TTX log$_2$ FC for all 91 putative PDCD4 transcriptional regulation target genes, clustered by FC across sample type. Each row represents a gene, and each column represents a sample type. The color legend represents Bic versus TTX log$_2$ FC, with red representing up-regulation, blue representing down-regulation, and white indicating a log$_2$ FC of 0. **(B)** GO analysis −log$_{10}$ FDR (circles) and percentage of genes (bars) in terms from biological process (top four terms) and cellular compartment (bottom four terms) analyses (Ashburner et al., 2000; Mi et al., 2019, The Gene Ontology Consortium, 2019). Data from putative PDCD4 target genes (91 genes) are shown in orange and, for comparison, data from other CHX-insensitive activity-dependent genes (368 genes) are shown in purple. Select GO terms are shown for simplicity (see Table S4 for top 15 GO terms by FDR for both groups of genes). **(C)** Diagram depicting a generic synapse and synaptic proteins. The labeled synaptic proteins are encoded by putative PDCD4 target genes (gene name indicated in parentheses alongside protein). The activity-dependent changes in expression of these genes are inhibited by degradation-resistant PDCD4. The presynaptic terminal is shown above with neurotransmitter-loaded synaptic vesicles, and the post-synaptic terminal is shown below with neurotransmitter receptors in the post-synaptic membrane (one receptor is shown anchored to an unlabeled gray PSD-95 protein). Arrow next to gene name illustrates the direction of activity-dependent differential expression, and dashed line with bar illustrates the suppression of this activity-dependent change in the PDCD4 S71A samples. DE, differential expression.

encoding proteins critical for synapse formation, remodeling, and transmission such as Shank1, p35, Abhd17b, Gap43, cofilin, spectrin-β2, myosin-Va, dendrin, Jacob, SNAP-β, voltage-dependent calcium channel-α2/δ1, α-tubulin, and β-actin (Fig. 6 C). Together, these results suggest that PDCD4 functions in the nucleus to regulate the expression of a subset of genes and that inhibiting the stimulation-induced degradation of nuclear PDCD4 results in suppression of the transcription of many activity-dependent genes important for neuronal synaptic function (Fig. 7).
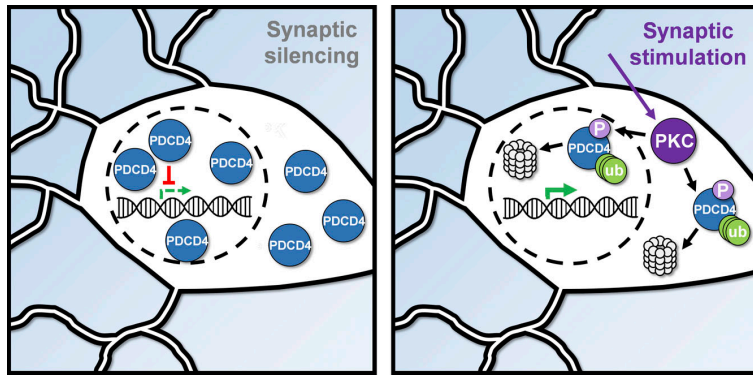
Figure 7. **Summary diagram of the activity-dependent proteasome-mediated degradation of PDCD4.** In silenced neurons (left), PDCD4 is highly expressed and suppresses the expression of specific genes. In stimulated neurons (right), PDCD4 is phosphorylated by PKC and undergoes proteasome-mediated degradation, thereby facilitating the expression of specific genes important for neuron synaptic function. ub, ubiquitin.

Discussion

In this study, we sought to test how neuronal activity at the synapse changes the nuclear proteome to regulate activity-dependent transcription. While advances in transcriptomic technologies have enabled the identification of genes that undergo activity-dependent changes in expression (Brigidi et al., 2019; Chen et al., 2017; Fernandez-Albert et al., 2019; Tyssowski et al., 2018), the systematic identification of proteins that undergo changes in subcellular localization and/or stability has been more challenging. We implemented a new proximity ligation approach to characterize changes in the nuclear proteome triggered by neuronal stimulation. Our results provide the first, to our knowledge, unbiased characterization of the population of proteins that undergo changes in nuclear abundance following neuronal silencing and/or glutamatergic stimulation, and they do so in a manner that is independent of translation or transcription. We detected activity-dependent changes in the known nucleocytoplasmic shuttling proteins, CRTC1 and HDAC4/5 (Ch'ng et al., 2012; Chawla et al., 2003), demonstrating the validity of this neuron-specific, subcellular compartment–specific assay. In addition, our results highlight an understudied mechanism of transcriptional regulation in neurons: activity-dependent nuclear proteasome-mediated degradation.

We found that synaptic activity led to proteasome-mediated degradation of the tumor suppressor protein PDCD4 in the nucleus and, to a lesser extent, in the cytoplasm. Given that the nuclear export inhibitor LMB did not block the activity-dependent decrease of nuclear PDCD4, we demonstrated that nuclear PDCD4 is degraded without leaving the nucleus. Many examples of activity-dependent degradation of proteins within the cytoplasm have been reported in neurons (Hegde et al., 1993; Banerjee et al., 2009; Jarome et al., 2011), but fewer cases of activity-dependent degradation of proteins within the nucleus have been described (Upadhya et al., 2004; Bayraktar et al., 2020; Kravchick et al., 2016). Nonetheless, the nucleus contains machinery for proteasome-mediated degradation, and there are numerous examples of proteins that are degraded by the nuclear proteasome in nonneuronal cells, including transcriptional regulators and cell-cycle proteins (von Mikecz, 2006). Neuronal nuclei have also been shown to contain machinery for proteasome-mediated degradation (Mengual et al., 1996) and exhibit proteasomal activity, albeit with less activity

than is present in the cytoplasm (Upadhya et al., 2006; Tydlacka et al., 2008). Activity-dependent nuclear proteasome–mediated degradation of transcriptional regulators represents an important mechanism by which synaptic activity regulates gene expression.

In neurons, synaptic activity leads to an increase in the activity of many kinases, including S6K and PKC (Callender and Newton, 2017; Biever et al., 2015). We found that a phospho-incompetent serine-to-alanine mutation at Ser71 prevented the activity-dependent degradation of PDCD4 and that phosphorylation by PKC, but not S6K, was required for this activity-dependent degradation. This result is consistent with previous studies in other cell types, demonstrating that either S6K or PKC is required for PDCD4 phosphorylation, depending on the signaling pathway (Matsuhashi et al., 2019, 2014; Dorrello et al., 2006; Nakashima et al., 2010; Schmid et al., 2008). The stimulus-specific requirement of either S6K or PKC for PDCD4 degradation raises the interesting possibility that different types of neuronal stimulation could trigger PDCD4 degradation via distinct signaling pathways. Supporting this idea, two studies in neurons have suggested that PDCD4 degradation may be regulated by S6K in response to injury and stress (Li et al., 2021; Di Paolo et al., 2020), whereas our study demonstrated that PDCD4 degradation was mediated by PKC. PKC is typically activated at the cell surface (Gould and Newton, 2008); however, it can translocate from the cytoplasm to the nucleus after activation and can be activated directly in the nucleus (Lim et al., 2015; Martelli et al., 2006), where it phosphorylates nuclear targets, including histones and transcription factors (Lim et al., 2015; Martelli et al., 2006).

PDCD4 has been well characterized as a translational repressor in the cytoplasm of cancer cells (Wang et al., 2017; Wedeken et al., 2011; Yang et al., 2004). Low concentrations of PDCD4 have been reported to correlate with invasion, proliferation, and metastasis of many types of cancers (Allgayer, 2010; Chen et al., 2003; Wang and Yang, 2018; Wei et al., 2012). However, the role of PDCD4 in the nucleus is less well characterized, even though the protein is predominantly localized in the nucleus of many cells (Böhm et al., 2003). Despite being highly expressed in neurons, few studies have examined the neuronal function of PDCD4 (Di Paolo et al., 2020; Li et al., 2021; Narasimhan et al., 2013), and, as far as we are aware, no

previous study has identified a role for PDCD4 in activity-dependent transcription in neurons. In our study, we detected a larger activity-dependent decrease of PDCD4 in the nucleus than in the cytoplasm, and we found that blocking PDCD4 degradation suppressed activity-dependent gene expression. These findings suggest that the proteasome-mediated degradation of PDCD4 is important for regulating activity-dependent transcription following neuronal stimulation. Previous studies in nonneuronal cells have indicated that PDCD4 can inhibit AP-1–dependent transcription, although it is unclear whether this is a direct role in the nucleus (Bitomsky et al., 2004) or an indirect role regulating the translation of signaling proteins in the cytoplasm (Yang et al., 2006). PDCD4 has also been shown to bind to the transcription factors CSL (Jo et al., 2016) and TWIST1 (Shiota et al., 2009) and inhibit their transcriptional activity. Although activity-dependent degradation of cytoplasmic PDCD4 was weaker and less consistent than that of nuclear PDCD4, we cannot rule out the possibility that PDCD4 has a translation-independent role in the cytoplasm (such as binding to signaling proteins) that could indirectly regulate transcription. However, we propose that PDCD4 has a direct role in regulating activity-dependent transcription in the nucleus, in addition to its well-characterized role as a translational repressor in the cytoplasm.

We identified 91 genes that are putative activity-dependent targets of PDCD4, including genes encoding proteins that are important for synaptic function. The activity-dependent downregulation of PDCD4 in neurons is reminiscent of the concept of "memory suppressor genes" (Abel and Kandel, 1998), genes that act as inhibitory constraints on activity-dependent neuronal plasticity. By analogy to PDCD4 function during cancer metastases, decreases in PDCD4 in neurons would function to enable experience-dependent neuronal growth and remodeling. Dysregulated PDCD4 concentrations have also been reported to underlie a variety of metabolic disorders, including polycystic ovary syndrome, obesity, diabetes, and atherosclerosis, highlighting the critical role PDCD4 plays in regulating gene expression in multiple cell types (Lu et al., 2020). Our study provides the first transcriptomic profile of PDCD4 that is independent of PDCD4's role in translation. These results provides insight into the transcriptional targets of PDCD4, which is of relevance not only to neuroscience but also to the study of PDCD4 in cancer.

Taken together, our findings illustrate the utility of proximity ligation assays in identifying activity-dependent changes in the proteome of subcellular neuronal compartments and point to the array of cell biological mechanisms by which activity can regulate the neuronal proteome. They also focus attention on the tumor suppressor protein PDCD4 as a critical regulator of activity-dependent gene expression in neurons, highlighting a role for PDCD4 in regulating the transcription of genes involved in synapse formation, remodeling, and transmission. This new role is in addition to PDCD4's well-characterized role as a translational inhibitor (Wang and Yang, 2018), and future investigation of the mechanisms by which PDCD4 regulates transcription of these genes will provide further insight into the understudied role of PDCD4 as a transcriptional regulator. Such studies also promise to deepen our understanding of the specific cell and molecular biological mechanisms by which experience alters gene expression in neurons to enable the formation and function of neural circuits.

Materials and methods

Experimental model details

Primary neuronal cultures

All experiments were performed using approaches approved by the University of California, Los Angeles, Animal Research Committee. Forebrains from postnatal day 0 Sprague-Dawley rats (Charles River) were dissected in cold HBSS (Thermo Fisher Scientific) supplemented with 10 mM Hepes buffer and 1 mM sodium pyruvate. Sex was not determined, and tissues from male and female pups were pooled. The tissue was chopped finely and digested in 1× trypsin solution (Thermo Fisher Scientific) in HBSS (supplemented with 120 µg/ml DNase and 1.2 mM $CaCl_2$) for 15 min at 37°C. The tissue was washed and triturated in DMEM (Thermo Fisher Scientific) + 10% FBS (Omega Scientific) before plating on polyDL-lysine (PDLL)-coated (0.1 mg/ml; Sigma) 10-cm dishes or 24-well plates containing acid-etched PDLL-coated coverslips (Carolina Biologicals). Neurons were plated at a density of one forebrain per 10-cm dish (for MS experiments) or one-half forebrain per entire 24-well plate (for ICC, RNA-seq, and Western blot experiments). Neurons were cultured in Neurobasal-A medium (Thermo Fisher Scientific) supplemented with 1× B-27 (Thermo Fisher Scientific), 0.5 mM GlutaMAX (Thermo Fisher Scientific), 25 µM monosodium glutamate (Sigma), and 25 µM β-mercaptoethanol (Sigma) and incubated at 37°C, 5% CO_2. When applicable, neurons were transfected with plasmids using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions at day in vitro (DIV) 2, transduced with AAV at DIV 13, or treated with 1 µM control Accell siRNA (D-001910-10-20; Horizon Discovery) or 1 µM PDCD4 Accell siRNA (E-097927-00-0020; Horizon Discovery) at DIV 15. All experiments were performed at DIV 20.

Detailed methods

Generation of plasmids and AAV

To create human synapsin promoter (hSyn) NLS-APEX2-EGFP-NLS, APEX2 was amplified from the pcDNA3 APEX2-NES plasmid (gift from Alice Ting, Stanford University, Stanford, CA; Addgene plasmid 49386) with three sequential sets of primers to add SV40 NLS to both the N-terminus and C-terminus of APEX2 (primer sets 1–3; Table S5). The design of using NLS on both sides of APEX2 was based on the design of Cas9-NLS (Swiech et al., 2015). NLS-APEX2-NLS was then inserted into the pAAV-hSyn-EGFP plasmid (gift from Bryan Roth, University of North Carolina, Chapel Hill, NC; Addgene plasmid 50465) between the BamHI and EcoRI sites, replacing the EGFP insert. The final hSyn NLS-APEX2-EGFP-NLS construct was created by amplifying hSyn NLS-APEX2-NLS (primer set 4; Table S5) and EGFP (primer set 5; Table S5) and joining the two products at NheI and SacI.

To create hSyn PDCD4-HA, rat PDCD4 was amplified from cultured neuron cDNA (primer set 6; Table S5) with a C-terminal

Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology   12 of 19
https://doi.org/10.1083/jcb.202103087

HA tag and then inserted into the pAAV-hSyn-EGFP plasmid between NcoI and EcoRI, replacing the EGFP insert. The S71A mutation was created using site-directed mutagenesis (services by GENEWIZ) to mutate serine 71 (TCT) to alanine (GCT). The hSyn V5-PDCD4-HA construct was created by adding a V5 tag to the N-terminus of PDCD4-HA using PCR-based mutagenesis (services by GENEWIZ).

AAV9 was generated for APEX2-NLS, PDCD4-HA WT, and PDCD4-HA S71A at the Penn Vector Core.

Pharmacological treatments
Neurons were preincubated with CHX (60 µM; Sigma) for 15 min; LMB (10 nM; Sigma) for 30 min; LY2584702 (1 µM; Cayman Chemical) or Ro-31-8425 (5 µM; Sigma) for 1 h; or Epox (5 µM; Enzo Life Sciences), Bort (10 µM; APExBIO), or MLN (50 nM; APExBIO) for 2 h. Inhibitors remained in the media throughout the duration of each respective experiment, incubated at 37°C. For neurons treated with inhibitors dissolved in DMSO (Epox, Bort, MLN, LY2584702, and Ro-31-8425), the final DMSO concentration in the media was 0.1% or less. For neurons treated with LMB, the final methanol concentration in the media was 0.08%. All control groups received an equivalent concentration of vehicle (DMSO or methanol). To silence the neurons, 1 µM TTX (Tocris Bioscience) was applied to the neurons for 1 h. To stimulate the neurons, 40 µM (–)-Bic methiodide (Tocris Bioscience) was applied to the neurons for 1 h unless otherwise stated.

For KCl stimulations, neurons were preincubated with standard Tyrode's solution (140 mM NaCl, 10 mM Hepes, 5 mM KCl, 3 mM CaCl$_2$, 1 mM MgCl$_2$, and 20 mM glucose, pH 7.4) containing 1 µM TTX for 15 min at RT and then stimulated for 5 min with 40 mM KCl isotonic Tyrode's solution containing TTX. Control cells remained in the standard Tyrode's solution containing TTX throughout the experiment.

Control data from LMB (Fig. 3 A and Fig. S3 D) and CHX (Fig. S3, A and B) experiments were combined to generate the data shown in Fig. 2, B and C. Two of the Bort experiments were performed concurrently with two of the MLN experiments, so these experiments partially share control data in Fig. 3, C and D; and Fig. S3, F and G. One of the LY2584702 experiments was performed concurrently with one of the Bort experiments, so Fig. 4 E and Fig. S3 I partially share control data with Fig. 3 C and Fig. S3 F.

APEX2 proximity biotinylation, streptavidin pulldown, and on-bead tryptic digestion
For APEX2 MS experiments, three biological replicates (sets of cultures) were prepared with three samples in each replicate (APEX + Bic, APEX + TTX, and no APEX). Neurons were TTX silenced or Bic stimulated for 1 h in the presence of CHX (as above). During the final 30 min of the treatment, neurons were incubated with 500 µM biotin-phenol (APExBIO) at 37°C. During the final 1 min, labeling was performed by adding H$_2$O$_2$ to a final concentration of 1 mM. To stop the labeling reaction, neurons were washed three times in large volumes of quencher solution (PBS with 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox [6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid]).

Neurons were lysed with radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, and 1% Triton X-100, pH 7.5) containing protease inhibitor cocktail (Sigma), phosphatase inhibitor cocktail (Sigma), and quenchers. Lysates were treated with benzonase (200 U/mg protein; Sigma) for 5 min and then clarified by centrifugation at 15,000 g for 10 min. Samples were concentrated with Amicon centrifugal filter tubes (3K NMWL; Millipore) to at least 1.5 mg/ml protein and quantified using a Pierce 660-nm protein assay kit (Thermo Fisher Scientific).

For each sample, 2 mg lysate was incubated with 220 µl Pierce streptavidin magnetic beads (Thermo Fisher Scientific) for 60 min at RT. Samples were washed twice with RIPA buffer; once with 1 M KCl; once with 0.1 M sodium carbonate; once with 2 M urea, and 10 mM Tris-HCl, pH 8.0; and twice with RIPA buffer according to Hung et al. (2016).

The streptavidin beads bound by biotinylated proteins were then washed three times with 8 M urea and 100 mM Tris-HCl, pH 8.5, and three times with pure water, and then the samples were resuspended in 100 µl 50 mM tetraethylammonium bromide. Samples were reduced and alkylated by sequentially incubating them with 5 mM tris(2-carboxyethyl)phosphine and 10 mM iodoacetamide for 30 min at RT in the dark on a shaker set to 1,000 rpm. The samples were incubated overnight with 0.4 µg Lys-C and 0.8 µg trypsin protease at 37°C on a shaker set to 1,000 rpm. Streptavidin beads were removed from peptide digests, and peptide digests were desalted using Pierce C18 tips (100-µl bead volume), dried, and then reconstituted in water.

TMT labeling
The desalted peptide digests were labeled by the TMT10plex Isobaric Label Reagent Set (Thermo Fisher Scientific) according to the manufacturer's instructions. In brief, peptides were incubated with acetonitrile (ACN)-reconstituted TMT labeling reagent for 1 h and then quenched by adding hydroxylamine. Sample-label matches were no-APEX replicate 1 labeled with TMT126, APEX+Bic replicate 1 labeled with TMT127N, APEX+TTX replicate 1 labeled with TMT127C, APEX+TTX replicate 2 labeled with TMT128N, no-APEX replicate 2 labeled with TMT128C, no-APEX replicate 3 labeled with TMT129N, APEX+Bic replicate 2 labeled with TMT129C, APEX+Bic replicate 3 labeled with TMT130N, and APEX+TTX replicate 3 labeled with TMT130C. Labeled samples were then combined, dried, and reconstituted in 0.1% trifluoroacetic acid for high pH reversed-phase fractionation.

High pH reversed-phase fractionation
High pH reversed-phase fractionation was performed using the Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. In brief, peptides were bound to the resin in the spin column and then eluted by stepwise incubations with 300 µl of increasing ACN concentrations. The eight fractions were combined into four fractions (fractions 1 and 5, 2 and 6, 3 and 7, and 4 and 8). Fractions were then dried by vacuum centrifugation and reconstituted in 5% formic acid for MS analysis.

Liquid chromatography–MS data acquisition

A 75-µm × 25-cm custom-made C18 column was connected to a nanoflow Dionex Ultimate 3000 ultra-high-pressure liquid chromatography system. A 140-min gradient of increasing ACN was delivered at a 200 nl/min flow rate as follows: 1–5.5% ACN phase from minutes 0 to 5, 5.5–27.5% ACN from minutes 5 to 128, 27.5–35% ACN from minutes 128 to 135, 35–80% ACN from minutes 135 to 136, 80% ACN hold from minutes 136 to 138, and then down to 1% ACN from minutes 138 to 140. An Orbitrap Fusion Lumos Tribrid mass spectrometer TMT-MS3-SPS method was used for data acquisition. Full MS scans were acquired at 120K resolution in the Orbitrap device with the automatic gain control target set to 2e5 and a maximum injection time set to 50 ms. MS2 scans were collected in ion trap with turbo scan rate after isolating precursors with an isolation window of 0.7 mass-to-charge ratio and collision-induced dissociation fragmentation using 35% collision energy. MS3 scans were acquired in the Orbitrap spectrometer at 50K resolution, and 10 synchronized selected precursor ions were pooled for each scan using 65% higher-energy C-trap dissociation energy for fragmentation. For data-dependent acquisition, a 3-s cycle time was used to acquire MS/MS spectra corresponding to peptide targets from the preceding full MS scan. Dynamic exclusion was set to 30 s.

MS/MS database search

MS/MS database searching was performed using MaxQuant (1.6.10.43; Cox and Mann, 2008) against the rat reference proteome from the European Molecular Biology Laboratory (UP000002494-10116 RAT, *Rattus norvegicus*, 21,649 entries). The search included carbamidomethylation as a fixed modification and methionine oxidation and N-terminal acetylation as variable modifications. The digestion mode was set to trypsin and allowed a maximum of two missed cleavages. The precursor mass tolerances were set to 20 and 4.5 ppm for the first and second searches, respectively, whereas a 20-ppm mass tolerance was used for fragment ions. Datasets were filtered at 1% FDR at both the peptide spectrum match and protein levels. Quantification type was set to reporter ion MS3 with 10-plex TMT option.

Statistical inference in MS data

MSStatsTMT (1.4.1; Huang et al., 2020) was used to analyze the MaxQuant TMT-MS3 data in the APEX2 proximity labeling experiment to statistically assess protein enrichment. TTX channels were used for MS run-level normalization. The "msstats" method was then used for protein summarization. P values for t tests were corrected for multiple hypothesis testing using the Benjamini-Hochberg adjustment. We identified proteins that were enriched above the no-APEX negative control using a $\log_2$ FC >3 and adjusted P value <0.05 cutoff above the no-APEX condition. Using this protein list, we then identified proteins that were differentially expressed when comparing between Bic and TTX conditions using $\log_2$ FC >0.5 or $\log_2$ FC less than –0.5 and P value <0.05. It is important to note that we used a nonadjusted P value cutoff when identifying candidate proteins that were differentially expressed between the TTX-

silenced and Bic-stimulated conditions, because only HDAC4 and six other proteins had a significant adjusted P value when using these cutoffs. Even CRTC1, a protein that has been shown to undergo activity-dependent changes in multiple studies (Ch'ng et al., 2012, 2015; Nonaka et al., 2014) and confirmed again in the present study, did not reach adjusted P value significance, suggesting that we did not have the statistical power to detect certain activity-dependent changes. Because we used nonadjusted P values to identify candidate proteins, it is especially important to experimentally validate any potential candidate protein.

Protein extraction and Western blot analysis

Neurons were washed in Tyrode's solution (140 mM NaCl, 10 mM Hepes, 5 mM KCl, 3 mM CaCl$_2$, 1 mM MgCl$_2$, and 20 mM glucose, pH 7.4) and lysed with RIPA buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, and 1% Triton X-100, pH 7.5) containing protease and phosphatase inhibitor cocktails (Sigma). Samples were clarified by centrifugation at 10,000 *g* for 10 min. Protein concentration was determined using the Pierce bicinchoninic acid protein assay kit (Thermo Fisher Scientific).

Protein lysates were boiled in loading buffer (10% glycerol, 1% SDS, 60 mM Tris-HCl, pH 7.0, 0.1 M DTT, and 0.02% bromophenol blue) for 10 min at 95°C and run on an 8% polyacrylamide gel for 90 min at 120 V. Samples were wet transferred onto a 0.2-µm nitrocellulose membrane for 16 h at 40 mA. The membrane was blocked with Odyssey Blocking Buffer (LI-COR Biosciences) and incubated with the following primary antibodies: mouse HA (901513, 1:1,000; BioLegend), mouse TUJ1 (801201, 1:1,000; BioLegend), mouse S6 (2317, 1:1,000; Cell Signaling Technology), and rabbit phospho-S6 Ser235/236 (4858, 1:2,000; Cell Signaling Technology) for 3–4 h at RT or overnight at 4°C. The membrane was washed with TBS with Tween 20 and incubated with the following secondary antibodies: anti-rabbit IRDye 800CW (1:10,000), anti-mouse IRDye 800CW (1:10,000), anti-mouse IRDye 680CW (1:10,000), and IRDye 800CW streptavidin (1:1,000) for 1 h at RT. The membrane was imaged using the Odyssey infrared imaging system (LI-COR Biosciences). Western blots were quantified using the Image Studio (LI-COR Biosciences) rectangle tool. The relative intensity of each band was calculated by normalizing to a loading control (TUJ1). Within each experiment, all values were normalized to the control (basal) sample.

ICC

Neurons were fixed with 4% PFA in PBS for 10 min, permeabilized in 0.1% Triton X-100 in PBS for 5 min, and blocked in 10% goat serum in PBS for 1 h. Neurons were incubated with the following primary antibodies: chicken microtubule-associated protein 2 (MAP2; 1100-MAP2, 1:1,000; PhosphoSolutions), rabbit PDCD4 (9535, 1:600; Cell Signaling Technology), mouse HA (901513, 1:1,000; BioLegend), mouse V5 (R960-25, 1:250; Thermo Fisher Scientific), rabbit CRTC1 (A300-769, 1:1,000; Bethyl Laboratories), rabbit HDAC4 (7628, 1:100; Cell Signaling Technology), and rabbit FOS (2250, 1:500; Cell Signaling Technology) for 3–4 h at RT or overnight at 4°C. Neurons were washed with

80

PBS and incubated at 1:1,000 with the following secondary antibodies: anti-chicken Alexa Fluor 647, anti-rabbit Alexa Fluor 555, anti-mouse Alexa Fluor 555, streptavidin Alexa Fluor 555, and Hoechst 33342 stain for 1 h at RT. Neurons were washed with PBS and mounted on slides with Aqua-Poly/Mount (Polysciences) for confocal imaging.

Protein synthesis assay (AHA labeling)

Neurons were washed once in warm Tyrode's solution and then incubated in methionine-free Hibernate A (BrainBits) supplemented with 1× B-27 for 30 min at 37°C, 0% CO_2 (due to the buffering conditions of Hibernate A). Neurons were then incubated in 4 mM AHA (Thermo Fisher Scientific) or the non-labeling control 4 mM methionine (Sigma) for 2 h in Hibernate A at 37°C, 0% CO_2. When applicable, 60 µm CHX was preincubated for 15 min before the start of the AHA labeling and remained in the media throughout the duration of the experiment. Neurons were washed twice in cold Tyrode's solution, fixed with 4% PFA in PBS for 10 min, permeabilized in 0.1% Triton X-100 in PBS for 5 min, and washed three times in 3% BSA in PBS. The Click-IT reaction was performed using the Click-iT Cell Reaction Buffer Kit (Thermo Fisher Scientific) and Alexa Fluor 488 alkyne (Thermo Fisher Scientific) according to the manufacturer's instructions, with a 30-min reaction at RT using 5 µM Alexa Fluor 488 alkyne. Neurons were washed three times in 3% BSA in PBS and then proceeded to the normal ICC protocol above, starting at the goat serum blocking step.

Confocal imaging

Samples were imaged using a Zeiss LSM 700 confocal microscope with a 40×/1.3 NA oil objective and a 63×/1.4 NA oil objective at RT and 405-nm, 488-nm, 555-nm, and 639-nm lasers, using ZEN microscopy software. All images were acquired using the 40× objective, except for the AHA incorporation protein synthesis experiments, which were imaged using the 63× objective. Identical image acquisition settings were used for all images within an experiment. For each image acquisition, the experimenter viewed the MAP2 and Hoechst channels to select a field of view and was blind to the experimental channel (e.g., HA, PDCD4). For each coverslip, images were taken at multiple regions throughout the coverslip, and two or three coverslips were imaged per condition. Images were collected from at least three experimental replicates (sets of cultures), unless otherwise stated.

Image analysis

ICC images were processed using ImageJ (Schindelin et al., 2012). An ImageJ macro was used to create regions of interest (ROIs) for neuronal nuclei. In brief, the Hoechst signal was used to outline the nucleus, and the MAP2 signal was used to select neurons and exclude nonneuronal cells. To create ROIs for neuronal cytoplasm, the cell body of each neuron was manually outlined using the MAP2 signal, and then the nuclear ROI was subtracted from the total cell body ROI. The ROIs were used to calculate the mean intensity in the channel of interest (e.g., HA, PDCD4) for the nucleus and cytoplasm of each neuron. Within each ICC experimental replicate, the measured values from all ROIs were normalized to the median value of the control condition (basal). For experiments using transfected cells (V5 experiments), the measured intensity of each ROI was normalized to the cotransfection marker (nuclear GFP intensity) in order to normalize for differences in transfection efficiency between cells.

RNA extraction, library preparation, RNA-seq, and data analysis

Samples were prepared from three biological replicates (sets of cultures), with eight samples in each replicate (WT Bic, WT TTX, S71A Bic, S71A TTX, CHX WT Bic, CHX WT TTX, CHX S71A Bic, CHX S71A TTX). RNA was extracted from neuronal cultures using the RNeasy Micro Kit (Qiagen) according to the manufacturer's instructions. Libraries for RNA-seq were prepared with the NuGEN Universal Plus mRNA-Seq Kit (NuGEN) to generate strand-specific RNA-seq libraries. Samples were multiplexed, and sequencing was performed on an Illumina HiSeq 3000 system to a depth of 25 million reads per sample with single-end 65-bp reads. Demultiplexing was performed using Illumina Bcl2fastq v2.19.1.403 software. The RNA-seq data discussed in this publication have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (Edgar et al., 2002) and are accessible through Gene Expression Omnibus accession no. GSE163127. Reads were aligned to *Rattus norvegicus* reference genome version Rnor_6.0 (rn6), and reads per gene were quantified by STAR 2.27a (Dobin et al., 2013) using an Rnor_6.0 gtf file. We used DESeq2 (Love et al., 2014) to obtain normalized read counts and perform differential expression analysis, including batch correction for replicate number (Table S3). Putative PDCD4 target genes were identified by first focusing on genes that showed activity-dependent differential expression in both the presence and absence of CHX (CHX-insensitive activity-dependent genes; 459 genes after excluding 3 genes that showed differential expression in different directions with or without CHX). We then calculated the PDCD4 activity-dependent change by taking the difference between the activity-dependent FC of PDCD4 WT and PDCD4 S71A samples and normalizing:

$$\text{PDCD4 change index} = \text{abs}(S - W)/\text{abs}(W),$$

where S is the PDCD4 S71A no-CHX Bic versus TTX $\log_2$ FC and W is the PDCD4 WT no-CHX Bic versus TTX $\log_2$ FC. We defined putative PDCD4 target genes as those with a PDCD4 change index >0.75. Motif analysis was performed using the findMotifsGenome command in HOMER (Heinz et al., 2010), using sequences from the transcription start site and upstream 500 bp as the promoter sequences for each gene. GO analysis was performed using the GO resource (Ashburner et al., 2000, The Gene Ontology Consortium, 2019) and PANTHER enrichment tools (Mi et al., 2019). A cartoon of putative PDCD4 targets was generated using BioRender.com.

RT-qPCR

As above, RNA was extracted from neuronal cultures using the RNeasy Micro Kit (Qiagen). cDNA was synthesized from RNA using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) with random hexamers. A no–reverse transcriptase sample was also prepared as a negative

Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology 15 of 19
https://doi.org/10.1083/jcb.202103087

81

control. RT-qPCR was performed on the CFX Connect Real-Time System (Bio-Rad Laboratories) using PowerUp SYBR Green Master Mix (Applied Biosystems). Primer pairs were designed for two housekeeping genes (*Hprt*, *Gapdh*), two candidate genes (*Scd1*, *Thrsp*), and *Pdcd4* using Primer3Plus (Untergasser et al., 2012) and National Center for Biotechnology Information Primer-BLAST (Ye et al., 2012; Table S5). RT-qPCR was performed on six or seven sets of cultures, with technical triplicate reactions for each sample. For each gene, relative quantity was calculated using the formula: $E^{\Delta Ct}$, where E was calculated from the primer efficiencies ($E \approx 2$). Relative gene expression was calculated by normalizing the relative quantity of the gene of interest to the relative quantity of the housekeeping genes *Hprt* and *Gapdh*: $(E_{gene})^{\Delta Ct\,gene}/\text{GeoMean}[(E_{HPRT})^{\Delta Ct\,HPRT}, (E_{GAPDH})^{\Delta Ct\,GAPDH}]$ (for Fig. S4 A) or *Gapdh* only (for Fig. S4, B and C; $E_{gene})^{\Delta Ct\,gene}/(E_{GAPDH})^{\Delta Ct\,GAPDH}$.

Quantification and statistical analysis

For ICC experiments, the quantification of signal intensity is displayed in violin plots created using GraphPad Prism. The medians are indicated with thick lines, and the quartiles are indicated with thin lines. *n* refers to the number of neurons in each condition, and all individual data points were plotted on the graphs. Our sample sizes were not predetermined. A nonparametric statistical test (Mann-Whitney *U* test) was used to calculate statistical significance because our data were not normally distributed, as indicated by the violin plots. A Bonferroni correction was used to adjust for multiple hypothesis testing. The sample sizes, statistical tests, medians, and P values are indicated in the figure legends.

For RT-qPCR and Western blot experiments, all data points were displayed using GraphPad Prism, with solid lines indicating the median values. *n* refers to the biological replicates (sets of cultures), and all data points were plotted on the graphs. The Mann-Whitney *U* test (Prism) was used to calculate statistical significance. The sample size, statistical tests, medians, and P values are indicated in the figure legends.

Online supplemental material

Fig. S1 accompanies Fig. 1 and provides verification of the protein synthesis inhibitor CHX, examines the nuclear localization of APEX2-NLS, and provides GO analysis results. Fig. S2 accompanies Fig. 2 and provides ICC data of PDCD4 using antibodies against the endogenous protein and epitope-tagged PDCD4. Fig. S3 accompanies Figs. 3 and 4 and provides verification of the nuclear export inhibitor LMB and S6K inhibitor Ly2584702, as well as quantification of cytoplasmic PDCD4. Fig. S4 accompanies Fig. 6 and provides RT-qPCR validation of candidate genes from RNA-seq. Fig. S5 accompanies Fig. 6 and provides motif analysis of promoters of putative PDCD4 target genes. Table S1 contains the MS data from the nuclear proteomes of silenced and stimulated neurons. Table S2 lists the GO analysis of the APEX2-NLS MS. Table S3 contains the RNA-seq data from silenced and stimulated neurons, with PDCD4 WT or S71A. Table S4 lists the GO analysis for putative PDCD4 target genes. Table S5 lists the primer sequences used for cloning and RT-qPCR.

Data availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact author, K.C. Martin (kcmartin@mednet.ucla.edu). Plasmids generated in this study are available upon request from the lead contact author. This published article includes the MS data generated during this study. The full RNA-seq data are available online (Gene Expression Omnibus accession no. GSE163127).

References

Abel, T., and E. Kandel. 1998. Positive and negative regulatory mechanisms that mediate long-term memory storage. *Brain Res. Brain Res. Rev.* 26: 360–378.

Alberini, C.M. 2009. Transcription factors in long-term memory and synaptic plasticity. *Physiol. Rev.* 89:121–145. https://doi.org/10.1152/physrev.00017.2008

Allgayer, H. 2010. Pdcd4, a colon cancer prognostic that is regulated by a microRNA. *Crit. Rev. Oncol. Hematol.* 73:185–191. https://doi.org/10.1016/j.critrevonc.2009.09.001

Asangani, I.A., S.A.K. Rasheed, D.A. Nikolova, J.H. Leupold, N.H. Colburn, S. Post, and H. Allgayer. 2008. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene.* 27:2128–2136. https://doi.org/10.1038/sj.onc.1210856

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, et al. The Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29. https://doi.org/10.1038/75556

Banerjee, S., P. Neveu, and K.S. Kosik. 2009. A coordinated local translational control point at the synapse involving relief from silencing and MOV10 degradation. *Neuron.* 64:871–884. https://doi.org/10.1016/j.neuron.2009.11.023

Bayraktar, G., P. Yuanxiang, A.D. Confettura, G.M. Gomes, S.A. Raza, O. Stork, S. Tajima, I. Suetake, A. Karpova, F. Yildirim, and M.R. Kreutz. 2020. Synaptic control of DNA methylation involves activity-dependent degradation of DNMT3A1 in the nucleus. *Neuropsychopharmacology.* 45: 2120–2130. https://doi.org/10.1038/s41386-020-0780-2

Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology 16 of 19
https://doi.org/10.1083/jcb.202103087

82

Biever, A., E. Valjent, and E. Puighermanal. 2015. Ribosomal protein S6 phosphorylation in the nervous system: from regulation to function. *Front. Mol. Neurosci.* 8:75. https://doi.org/10.3389/fnmol.2015.00075

Bitomsky, N., M. Böhm, and K.H. Klempnauer. 2004. Transformation suppressor protein Pdcd4 interferes with JNK-mediated phosphorylation of c-Jun and recruitment of the coactivator p300 by c-Jun. *Oncogene.* 23:7484–7493. https://doi.org/10.1038/sj.onc.1208064

Bloodgood, B.L., N. Sharma, H.A. Browne, A.Z. Trepman, and M.E. Greenberg. 2013. The activity-dependent transcription factor NPAS4 regulates domain-specific inhibition. *Nature.* 503:121–125. https://doi.org/10.1038/nature12743

Böhm, M., K. Sawicka, J.P. Siebrasse, A. Brehmer-Fastnacht, R. Peters, and K.H. Klempnauer. 2003. The transformation suppressor protein Pdcd4 shuttles between nucleus and cytoplasm and binds RNA. *Oncogene.* 22:4905–4910. https://doi.org/10.1038/sj.onc.1206710

Brigidi, G.S., M.G.B. Hayes, N.P. Delos Santos, A.L. Hartzell, L. Texari, P.A. Lin, A. Bartlett, J.R. Ecker, C. Benner, S. Heinz, and B.L. Bloodgood. 2019. Genomic decoding of neuronal depolarization by stimulus-specific NPAS4 heterodimers. *Cell.* 179:373–391.e27. https://doi.org/10.1016/j.cell.2019.09.004

Callender, J.A., and A.C. Newton. 2017. Conventional protein kinase C in the brain: 40 years later. *Neuronal Signal.* 1:NS20160005. https://doi.org/10.1042/NS20160005

Ch'ng, T.H., B. Uzgil, P. Lin, N.K. Avliyakulov, T.J. O'Dell, and K.C. Martin. 2012. Activity-dependent transport of the transcriptional coactivator CRTC1 from synapse to nucleus. *Cell.* 150:207–221. https://doi.org/10.1016/j.cell.2012.05.027

Ch'ng, T.H., M. DeSalvo, P. Lin, A. Vashisht, J.A. Wohlschlegel, and K.C. Martin. 2015. Cell biological mechanisms of activity-dependent synapse to nucleus translocation of CRTC1 in neurons. *Front. Mol. Neurosci.* 8:48. https://doi.org/10.3389/fnmol.2015.00048

Chawla, S., P. Vanhoutte, F.J.L. Arnold, C.L.H. Huang, and H. Bading. 2003. Neuronal activity-dependent nucleocytoplasmic shuttling of HDAC4 and HDAC5. *J. Neurochem.* 85:151–159. https://doi.org/10.1046/j.1471-4159.2003.01648.x

Chen, Y., T. Knösel, G. Kristiansen, A. Pietas, M.E. Garber, S. Matsuhashi, I. Ozaki, and I. Petersen. 2003. Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis. *J. Pathol.* 200:640–646. https://doi.org/10.1002/path.1378

Chen, P.B., R. Kawaguchi, C. Blum, J.M. Achiro, G. Coppola, T.J. O'Dell, and K.C. Martin. 2017. Mapping gene expression in excitatory neurons during hippocampal late-phase long-term potentiation. *Front. Mol. Neurosci.* 10:39. https://doi.org/10.3389/fnmol.2017.00039

Cox, J., and M. Mann. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26:1367–1372. https://doi.org/10.1038/nbt.1511

Di Paolo, A., G. Eastman, R. Mesquita-Ribeiro, J. Farias, A. Macklin, T. Kislinger, N. Colburn, D. Munroe, J.R. Sotelo Sosa, F. Dajas-Bailador, and J.R. Sotelo-Silveira. 2020. PDCD4 regulates axonal growth by translational repression of neurite growth-related genes and is modulated during nerve injury responses. *RNA.* 26:1637–1653. https://doi.org/10.1261/rna.075424.120

Dieterich, D.C., A. Karpova, M. Mikhaylova, I. Zdobnova, I. König, M. Landwehr, M. Kreutz, K.H. Smalla, K. Richter, P. Landgraf, et al. 2008. Caldendrin-Jacob: a protein liaison that couples NMDA receptor signalling to the nucleus. *PLoS Biol.* 6:e34. https://doi.org/10.1371/journal.pbio.0060034

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dörrbaum, A.R., B. Alvarez-Castelao, B. Nassim-Assir, J.D. Langer, and E.M. Schuman. 2020. Proteome dynamics during homeostatic scaling in cultured neurons. *eLife.* 9:e52939. https://doi.org/10.7554/eLife.52939

Dorrello, N.V., A. Peschiaroli, D. Guardavaccaro, N.H. Colburn, N.E. Sherman, and M. Pagano. 2006. S6K1- and betaTRCP-mediated degradation of PDCD4 promotes protein translation and cell growth. *Science.* 314:467–471. https://doi.org/10.1126/science.1130276

Edgar, R., M. Domrachev, and A.E. Lash. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30:207–210. https://doi.org/10.1093/nar/30.1.207

Fernandez-Albert, J., M. Lipinski, M.T. Lopez-Cascales, M.J. Rowley, A.M. Martin-Gonzalez, B. Del Blanco, V.G. Corces, and A. Barco. 2019. Immediate and deferred epigenomic signatures of in vivo neuronal

activation in mouse hippocampus. *Nat. Neurosci.* 22:1718–1730. https://doi.org/10.1038/s41593-019-0476-2

Flavell, S.W., C.W. Cowan, T.K. Kim, P.L. Greer, Y. Lin, S. Paradis, E.C. Griffith, L.S. Hu, C. Chen, and M.E. Greenberg. 2006. Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science.* 311:1008–1012. https://doi.org/10.1126/science.1122511

Frankel, L.B., N.R. Christoffersen, A. Jacobsen, M. Lindow, A. Krogh, and A.H. Lund. 2008. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J. Biol. Chem.* 283:1026–1033. https://doi.org/10.1074/jbc.M707224200

Gallegos, D.A., U. Chan, L.F. Chen, and A.E. West. 2018. Chromatin regulation of neuronal maturation and plasticity. *Trends Neurosci.* 41:311–324. https://doi.org/10.1016/j.tins.2018.02.009

Gould, C.M., and A.C. Newton. 2008. The life and death of protein kinase C. *Curr. Drug Targets.* 9:614–625. https://doi.org/10.2174/138945008785132411

Hegde, A.N., A.L. Goldberg, and J.H. Schwartz. 1993. Regulatory subunits of cAMP-dependent protein kinases are degraded after conjugation to ubiquitin: a molecular mechanism underlying long-term synaptic plasticity. *Proc. Natl. Acad. Sci. USA.* 90:7436–7440. https://doi.org/10.1073/pnas.90.16.7436

Heinz, D.A., and B.L. Bloodgood. 2020. Mechanisms that communicate features of neuronal activity to the genome. *Curr. Opin. Neurobiol.* 63:131–136. https://doi.org/10.1016/j.conb.2020.03.002

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* 38:576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Holt, C.E., K.C. Martin, and E.M. Schuman. 2019. Local translation in neurons: visualization and function. *Nat. Struct. Mol. Biol.* 26:557–566. https://doi.org/10.1038/s41594-019-0263-5

Hrvatin, S., D.R. Hochbaum, M.A. Nagy, M. Cicconet, K. Robertson, L. Cheadle, R. Zilionis, A. Ratner, R. Borges-Monroy, A.M. Klein, et al. 2018. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* 21:120–129. https://doi.org/10.1038/s41593-017-0029-5

Huang, T., M. Choi, M. Tzouros, S. Golling, N.J. Pandya, B. Banfai, T. Dunkley, and O. Vitek. 2020. MSstatsTMT: statistical detection of differentially abundant proteins in experiments with isobaric labeling and multiple mixtures. *Mol. Cell. Proteomics.* 19:1706–1723. https://doi.org/10.1074/mcp.RA120.002105

Hung, V., N.D. Udeshi, S.S. Lam, K.H. Loh, K.J. Cox, K. Pedram, S.A. Carr, and A.Y. Ting. 2016. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* 11:456–475. https://doi.org/10.1038/nprot.2016.018

Jarome, T.J., C.T. Werner, J.L. Kwapis, and F.J. Helmstetter. 2011. Activity dependent protein degradation is critical for the formation and stability of fear memory in the amygdala. *PLoS One.* 6:e24349. https://doi.org/10.1371/journal.pone.0024349

Jiang, Y., S. Zhao, Y. Ding, L. Nong, H. Li, G. Gao, D. Zhou, and N. Xu. 2017. MicroRNA-21 promotes neurite outgrowth by regulating PDCD4 in a rat model of spinal cord injury. *Mol. Med. Rep.* 16:2522–2528. https://doi.org/10.3892/mmr.2017.6862

Jo, S.H., D.E. Kim, A. Clocchiatti, and G.P. Dotto. 2016. PDCD4 is a CSL associated protein with a transcription repressive function in cancer associated fibroblast activation. *Oncotarget.* 7:58717–58727. https://doi.org/10.18632/oncotarget.11227

Kalderon, D., B.L. Roberts, W.D. Richardson, and A.E. Smith. 1984. A short amino acid sequence able to specify nuclear location. *Cell.* 39:499–509. https://doi.org/10.1016/0092-8674(84)90457-4

Kravchick, D.O., A. Karpova, M. Hrdinka, J. Lopez-Rojas, S. Iacobas, A.U. Carbonell, D.A. Iacobas, M.R. Kreutz, and B.A. Jordan. 2016. Synapto-nuclear messenger PRR7 inhibits c-Jun ubiquitination and regulates NMDA-mediated excitotoxicity. *EMBO J.* 35:1923–1934. https://doi.org/10.15252/embj.201593070

Lacar, B., S.B. Linker, B.N. Jaeger, S.R. Krishnaswami, J.J. Barron, M.J.E. Kelder, S.L. Parylak, A.C.M. Paquola, M. Venepally, M. Novotny, et al. 2016. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* 7:11022. https://doi.org/10.1038/ncomms11022

Lein, E.S., M.J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A.F. Boe, M.S. Boguski, K.S. Brockway, E.J. Byrnes, et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature.* 445:168–176. https://doi.org/10.1038/nature05453

Li, Y., Y. Jia, D. Wang, X. Zhuang, Y. Li, C. Guo, H. Chu, F. Zhu, J. Wang, X. Wang, et al. 2021. Programmed cell death 4 as an endogenous

suppressor of BDNF translation is involved in stress-induced depression. *Mol. Psychiatry.* 26:2316–2333. https://doi.org/10.1038/s41380-020 -0692-x

Lim, P.S., C.R. Sutton, and S. Rao. 2015. Protein kinase C in the immune system: from signalling to chromatin regulation. *Immunology.* 146: 508–522. https://doi.org/10.1111/imm.12510

Lin, Y., B.L. Bloodgood, J.L. Hauser, A.D. Lapan, A.C. Koon, T.K. Kim, L.S. Hu, A.N. Malik, and M.E. Greenberg. 2008. Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature.* 455:1198–1204. https://doi.org/10.1038/nature07319

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15: 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, K., Q. Chen, M. Li, L. He, F. Riaz, T. Zhang, and D. Li. 2020. Programmed cell death factor 4 (PDCD4), a novel therapy target for metabolic diseases besides cancer. *Free Radic. Biol. Med.* 159:150–163. https://doi.org/ 10.1016/j.freeradbiomed.2020.06.016

Ma, H., R.D. Groth, S.M. Cohen, J.F. Emery, B. Li, E. Hoedt, G. Zhang, T.A. Neubert, and R.W. Tsien. 2014. γCaMKII shuttles Ca²⁺/CaM to the nucleus to trigger CREB phosphorylation and gene expression. *Cell.* 159: 281–294. https://doi.org/10.1016/j.cell.2014.09.019

Martelli, A.M., C. Evangelisti, M. Nyakern, and F.A. Manzoli. 2006. Nuclear protein kinase C. *Biochim. Biophys. Acta.* 1761:542–551. https://doi.org/10 .1016/j.bbalip.2006.02.009

Martin, K.C., and A. Ephrussi. 2009. mRNA localization: gene expression in the spatial dimension. *Cell.* 136:719–730. https://doi.org/10.1016/j.cell .2009.01.044

Matsuhashi, S., H. Hamajima, J. Xia, H. Zhang, T. Mizuta, K. Anzai, and I. Ozaki. 2014. Control of a tumor suppressor PDCD4: Degradation mechanisms of the protein in hepatocellular carcinoma cells. *Cell. Signal.* 26:603–610. https://doi.org/10.1016/j.cellsig.2013.11.038

Matsuhashi, S., M. Manirujjaman, H. Hamajima, and I. Ozaki. 2019. Control mechanisms of the tumor suppressor PDCD4: expression and functions. *Int. J. Mol. Sci.* 20:2304. https://doi.org/10.3390/ijms20092304

Mengual, E., P. Arizti, J. Rodrigo, J.M. Giménez-Amaya, and J.G. Castaño. 1996. Immunohistochemical distribution and electron microscopic subcellular localization of the proteasome in the rat CNS. *J. Neurosci.* 16: 6331–6341. https://doi.org/10.1523/JNEUROSCI.16-20-06331.1996

Merlet, J., J. Burger, J.E. Gomes, and L. Pintard. 2009. Regulation of cullin-RING E3 ubiquitin-ligases by neddylation and dimerization. *Cell. Mol. Life Sci.* 66:1924–1938. https://doi.org/10.1007/s00018-009-8712-7

Mi, H., A. Muruganujan, D. Ebert, X. Huang, and P.D. Thomas. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47(D1): D419–D426. https://doi.org/10.1093/nar/gky1038

Nakashima, M., H. Hamajima, J. Xia, S. Iwane, Y. Kwaguchi, Y. Eguchi, T. Mizuta, K. Fujimoto, I. Ozaki, and S. Matsuhashi. 2010. Regulation of tumor suppressor PDCD4 by novel protein kinase C isoforms. *Biochim. Biophys. Acta.* 1803:1020–1027. https://doi.org/10.1016/j.bbamcr.2010 .05.002

Narasimhan, M., M. Rathinam, A. Riar, D. Patel, S. Mummidi, H.S. Yang, N.H. Colburn, G.I. Henderson, and L. Mahimainathan. 2013. Programmed cell death 4 (PDCD4): a novel player in ethanol-mediated suppression of protein translation in primary cortical neurons and developing cerebral cortex. *Alcohol. Clin. Exp. Res.* 37:96–109. https://doi.org/10.1111/j.1530 -0277.2012.01850.x

Ning, F.L., F. Wang, M.L. Li, Z.S. Yu, Y.Z. Hao, and S.S. Chen. 2014. MicroRNA-182 modulates chemosensitivity of human non-small cell lung cancer to cisplatin by targeting PDCD4. *Diagn. Pathol.* 9:143. https://doi.org/10.1186/1746-1596-9-143

Nonaka, M., R. Kim, H. Fukushima, K. Sasaki, K. Suzuki, M. Okamura, Y. Ishii, T. Kawashima, S. Kamijo, S. Takemoto-Kimura, et al. 2014. Region-specific activation of CRTC1-CREB signaling mediates long-term fear memory. *Neuron.* 84:92–106. https://doi.org/10.1016/j.neuron.2014.08.049

Polleux, F., G. Ince-Dunn, and A. Ghosh. 2007. Transcriptional regulation of vertebrate axon guidance and synapse formation. *Nat. Rev. Neurosci.* 8: 331–340. https://doi.org/10.1038/nrn2118

Ramanan, N., Y. Shen, S. Sarsfield, T. Lemberger, G. Schütz, D.J. Linden, and D.D. Ginty. 2005. SRF mediates activity-induced gene expression and synaptic plasticity but not neuronal viability. *Nat. Neurosci.* 8:759–767. https://doi.org/10.1038/nn1462

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods.* 9: 676–682. https://doi.org/10.1038/nmeth.2019

Schlumm, F., D. Mauceri, H.E. Freitag, and H. Bading. 2013. Nuclear calcium signaling regulates nuclear export of a subset of class IIa histone deacetylases following synaptic activity. *J. Biol. Chem.* 288:8074–8084. https://doi.org/10.1074/jbc.M112.432773

Schmid, T., A.P. Jansen, A.R. Baker, G. Hegamyer, J.P. Hagan, and N.H. Colburn. 2008. Translation inhibitor Pdcd4 is targeted for degradation during tumor promotion. *Cancer Res.* 68:1254–1260. https://doi.org/10 .1158/0008-5472.CAN-07-1719

Sekeres, M.J., V. Mercaldo, B. Richards, D. Sargin, V. Mahadevan, M.A. Woodin, P.W. Frankland, and S.A. Josselyn. 2012. Increasing CRTC1 function in the dentate gyrus during memory formation or reactivation increases memory strength without compromising memory quality. *J. Neurosci.* 32: 17857–17868. https://doi.org/10.1523/JNEUROSCI.1419-12.2012

Shiota, M., H. Izumi, A. Tanimoto, M. Takahashi, N. Miyamoto, E. Kashiwagi, A. Kidani, G. Hirano, D. Masubuchi, Y. Fukunaka, et al. 2009. Programmed cell death protein 4 down-regulates Y-box binding protein-1 expression via a direct interaction with Twist1 to suppress cancer cell growth. *Cancer Res.* 69:3148–3156. https://doi.org/10.1158/0008-5472 .CAN-08-2334

Swiech, L., M. Heidenreich, A. Banerjee, N. Habib, Y. Li, J. Trombetta, M. Sur, and F. Zhang. 2015. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* 33:102–106. https://doi.org/10.1038/nbt.3055

The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47(D1):D330–D338. https://doi .org/10.1093/nar/gky1055

Ting, L., R. Rad, S.P. Gygi, and W. Haas. 2011. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods.* 8: 937–940. https://doi.org/10.1038/nmeth.1714

Tydlacka, S., C.E. Wang, X. Wang, S. Li, and X.J. Li. 2008. Differential activities of the ubiquitin-proteasome system in neurons versus glia may account for the preferential accumulation of misfolded proteins in neurons. *J. Neurosci.* 28:13285–13295. https://doi.org/10.1523/JNEUROSCI.4393-08.2008

Tyssowski, K.M., N.R. DeStefino, J.H. Cho, C.J. Dunn, R.G. Poston, C.E. Carty, R.D. Jones, S.M. Chang, P. Romeo, M.K. Wurzelmann, et al. 2018. Different neuronal activity patterns induce different gene expression programs. *Neuron.* 98:530–546.e11. https://doi.org/10.1016/j.neuron .2018.04.001

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. https://doi.org/10.1093/nar/gks596

Upadhya, S.C., T.K. Smith, and A.N. Hegde. 2004. Ubiquitin-proteasome-mediated CREB repressor degradation during induction of long-term facilitation. *J. Neurochem.* 91:210–219. https://doi.org/10.1111/j.1471-4159 .2004.02707.x

Upadhya, S.C., L. Ding, T.K. Smith, and A.N. Hegde. 2006. Differential regulation of proteasome activity in the nucleus and the synaptic terminals. *Neurochem. Int.* 48:296–305. https://doi.org/10.1016/j.neuint.2005.11.003

von Mikecz, A. 2006. The nuclear ubiquitin-proteasome system. *J. Cell Sci.* 119:1977–1984. https://doi.org/10.1242/jcs.03008

Wang, Q., and H.S. Yang. 2018. The role of Pdcd4 in tumour suppression and protein translation. *Biol. Cell.* 110:169–177. https://doi.org/10.1111/boc .201800014

Wang, Q., J. Zhu, Y.W. Wang, Y. Dai, Y.L. Wang, C. Wang, J. Liu, A. Baker, N.H. Colburn, and H.S. Yang. 2017. Tumor suppressor Pdcd4 attenuates Sin1 translation to inhibit invasion in colon carcinoma. *Oncogene.* 36: 6225–6234. https://doi.org/10.1038/onc.2017.228

Wayman, G.A., S. Impey, D. Marks, T. Saneyoshi, W.F. Grant, V. Derkach, and T.R. Soderling. 2006. Activity-dependent dendritic arborization mediated by CaM-kinase I activation and enhanced CREB-dependent transcription of Wnt-2. *Neuron.* 50:897–909. https://doi.org/10.1016/j .neuron.2006.05.008

Wedeken, L., P. Singh, and K.H. Klempnauer. 2011. Tumor suppressor protein Pdcd4 inhibits translation of p53 mRNA. *J. Biol. Chem.* 286:42855–42862. https://doi.org/10.1074/jbc.M111.269456

Wei, N., S.S. Liu, K.K.L. Chan, and H.Y.S. Ngan. 2012. Tumour suppressive function and modulation of programmed cell death 4 (PDCD4) in ovarian cancer. *PLoS One.* 7:e30311. https://doi.org/10.1371/journal.pone .0030311

West, A.E., and M.E. Greenberg. 2011. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* 3:a005744. https://doi.org/10.1101/cshperspect .a005744

Yang, H.-S., A.P. Jansen, A.A. Komar, X. Zheng, W.C. Merrick, S. Costes, S.J. Lockett, N. Sonenberg, and N.H. Colburn. 2003. The transformation

suppressor Pdcd4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation. *Mol. Cell. Biol.* 23:26–37. https://doi.org/10.1128/MCB.23.1.26-37.2003

Yang, H.-S., M.-H. Cho, H. Zakowicz, G. Hegamyer, N. Sonenberg, and N.H. Colburn. 2004. A novel function of the MA-3 domains in transformation and translation suppressor Pdcd4 is essential for its binding to eukaryotic translation initiation factor 4A. *Mol. Cell. Biol.* 24:3894–3906. https://doi.org/10.1128/MCB.24.9.3894-3906.2004

Yang, H.-S., C.P. Matthews, T. Clair, Q. Wang, A.R. Baker, C.-C.H. Li, T.-H. Tan, and N.H. Colburn. 2006. Tumorigenesis suppressor Pdcd4 down-regulates mitogen-activated protein kinase kinase kinase kinase 1 expression to suppress colon carcinoma cell invasion. *Mol. Cell. Biol.* 26:1297–1306. https://doi.org/10.1128/MCB.26.4.1297-1306.2006

Yap, E.L., and M.E. Greenberg. 2018. Activity-regulated transcription: bridging the gap between neural activity and behavior. *Neuron.* 100:330–348. https://doi.org/10.1016/j.neuron.2018.10.013

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T.L. Madden. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* 13:134. https://doi.org/10.1186/1471-2105-13-134

GO Cellular Component	# Terms in Dataset	# Terms Expected	Fold Enrichment	FDR
Supraspliceosomal Complex	4	0.49	8.17	3.75E-02
Proteasome Activator Complex	4	0.49	8.17	3.74E-02
Dendritic Filopodium	5	0.61	8.17	1.50E-02
Barr Body	4	0.49	8.17	3.73E-02
NURF Complex	6	0.73	8.17	5.94E-03
Paraspeckles	6	0.73	8.17	5.92E-03
Presynaptic Endocytic Zone	12	1.59	7.54	4.16E-05
Proteasome Regulatory Particle, Base Subcomplex	11	1.47	7.49	1.06E-04
Presynaptic Endocytic Zone Membrane	8	1.10	7.26	1.56E-03
PTW/PP1 Phosphatase Complex	6	0.86	7.00	9.56E-03
Transcription Factor TFIIIC Complex	5	0.73	6.81	2.32E-02
pICln-Sm Protein Complex	5	0.73	6.81	2.32E-02
Nuclear Stress Granule	5	0.73	6.81	2.31E-02
NuRD Complex	13	1.96	6.64	4.24E-05
CHD-type Complex	13	1.96	6.64	4.21E-05
Cyclin/CDK Positive Transcription Elongation Factor Complex	8	1.22	6.53	2.39E-03
Nuclear Lamina	8	1.22	6.53	2.38E-03
nBAF Complex	12	1.84	6.53	1.06E-04

Figure S1. **Identification of the nuclear proteomes from silenced and stimulated neurons using APEX2 proximity biotinylation.** Related to Fig. 1. **(A)** Metabolic labeling and ICC of control, CHX-treated, and nonlabeled cells. Protein synthesis was measured by AHA incorporation and Click-IT Alexa Fluor 488 alkyne reaction. The Alexa Fluor 488 signal intensity is displayed using a lookup table. Scale bar, 10 µm. **(B)** Violin plots of normalized somatic Click-IT intensity. Control, n = 90 cells; CHX, n = 63 cells; and no label, n = 7 cells from two sets of cultures. Control median = 1.00; CHX median = 0.32; no-label median = 0.05. Control versus CHX, P < 0.0001. **(C)** FOS ICC of basal and Bic-stimulated neurons in the presence or absence of CHX. Scale bar, 10 µm. **(D)** Violin plots of normalized nuclear FOS ICC intensity. Basal, n = 28; Bic, n = 40 cells; CHX-basal, n = 32; and CHX-Bic, n = 26 cells from one set of cultures. Basal median = 1.00; Bic median = 4.484; CHX-basal median = 0.8941; CHX-Bic median = 0.8307. Basal versus Bic, P < 0.0001; CHX-basal versus CHX-Bic, P = 0.3064. **(E)** Localization of APEX2-NLS GFP in TTX-silenced and Bic-stimulated neurons in the presence of CHX. Scale bar, 30 µm. **(F)** Violin plots of APEX2-NLS GFP nucleocytoplasmic localization (nuclear [Nuc]/cytoplasmic [Cyto] ratio) in TTX-silenced and Bic-stimulated neurons in the presence of CHX. TTX, n = 116 cells; and Bic, n = 120 cells from nine sets of cultures. TTX median = 4.31; Bic median = 4.00. TTX versus Bic, P = 0.0562. **(G)** Bar graph displaying the percentage of proteins containing the GO term nucleus (GO:0005634) from the list of proteins detected by APEX2-NLS MS. The list of proteins detected by MS was ranked according to MS/MS count abundance, and separate bars are displayed for the top 50 proteins, top 100 proteins, and so forth. **(H)** Cellular component GO analysis of the APEX2-NLS MS, with the top 18 terms listed in order of fold enrichment. Nuclear cellular components are listed in blue. ****, P < 0.0001; Mann-Whitney U test.
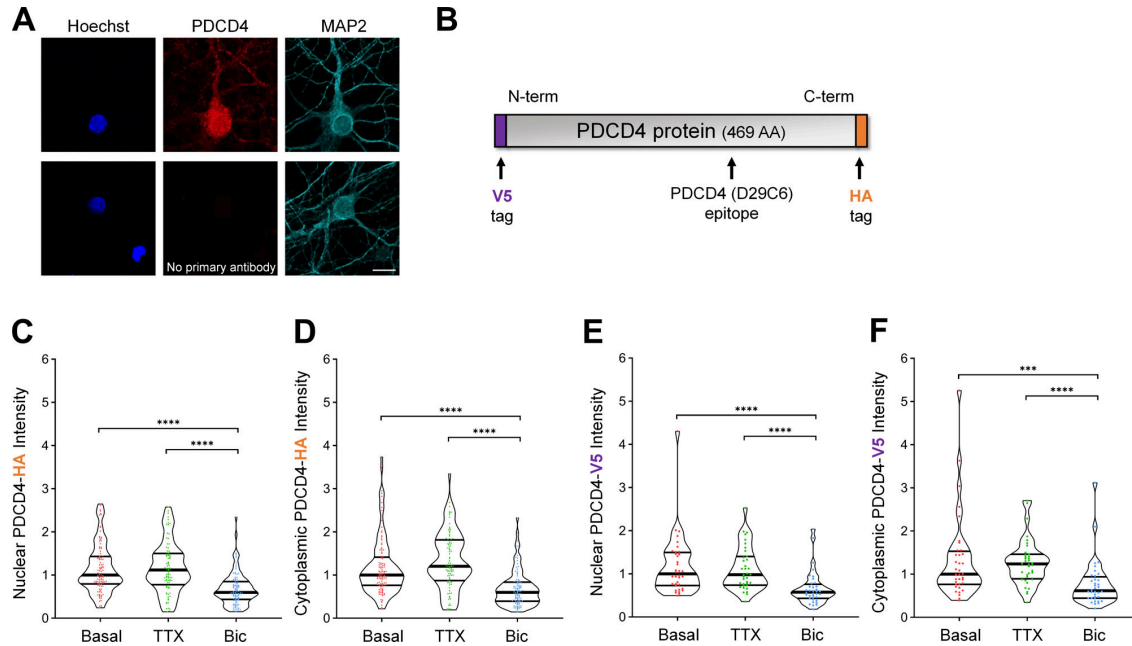
Figure S2. **Neuronal stimulation decreases PDCD4 protein concentration in the nucleus and cytoplasm of neurons.** Related to Fig. 2. **(A)** Top: ICC of endogenous PDCD4 protein. Bottom: Negative control (no primary PDCD4 antibody). Scale bar, 10 μm. **(B)** Schematic of PDCD4 protein, with locations of V5 tag, HA tag, and PDCD4 (D29C6) epitope (recognized by the PDCD4 Cell Signaling Technology antibody used in this study). **(C)** Violin plots of normalized nuclear HA ICC intensity in neurons transduced with PDCD4-HA AAV. Basal, n = 107 cells; TTX, n = 88 cells; Bic, n = 102 cells from three sets of cultures. Basal median = 1.00; TTX median = 1.116; Bic median = 0.5972. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001. **(D)** Violin plots of normalized cytoplasmic HA ICC intensity in the same cells as in C. Basal median = 1.00; TTX median = 1.203; Bic median = 0.5983. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001. **(E)** Violin plots of normalized nuclear V5 ICC intensity in neurons transfected with V5-PDCD4 plasmid and cotransfected with GFP plasmid as a transfection marker. Basal, n = 36 cells; TTX, n = 36 cells; and Bic, n = 36 cells from two sets of cultures. For each cell, the nuclear V5 intensity was normalized to the nuclear GFP intensity in order to normalize for differences in transfection efficiency between cells. Basal median = 1.00; TTX median = 0.9810; Bic median = 0.5760. Basal versus Bic, P < 0.0001; TTX versus Bic, P < 0.0001. **(F)** Violin plots of normalized cytoplasmic V5 ICC intensity in the same cells as in E. For each cell, the cytoplasmic V5 intensity was normalized to the nuclear GFP intensity in order to normalize for differences in transfection efficiency between cells. Basal median = 1.00; TTX median = 1.237; Bic median = 0.6167. Basal versus Bic, P = 0.0002; TTX versus Bic, P < 0.0001. ***, P < 0.001; ****, P < 0.0001; Mann-Whitney U test with Bonferroni correction.
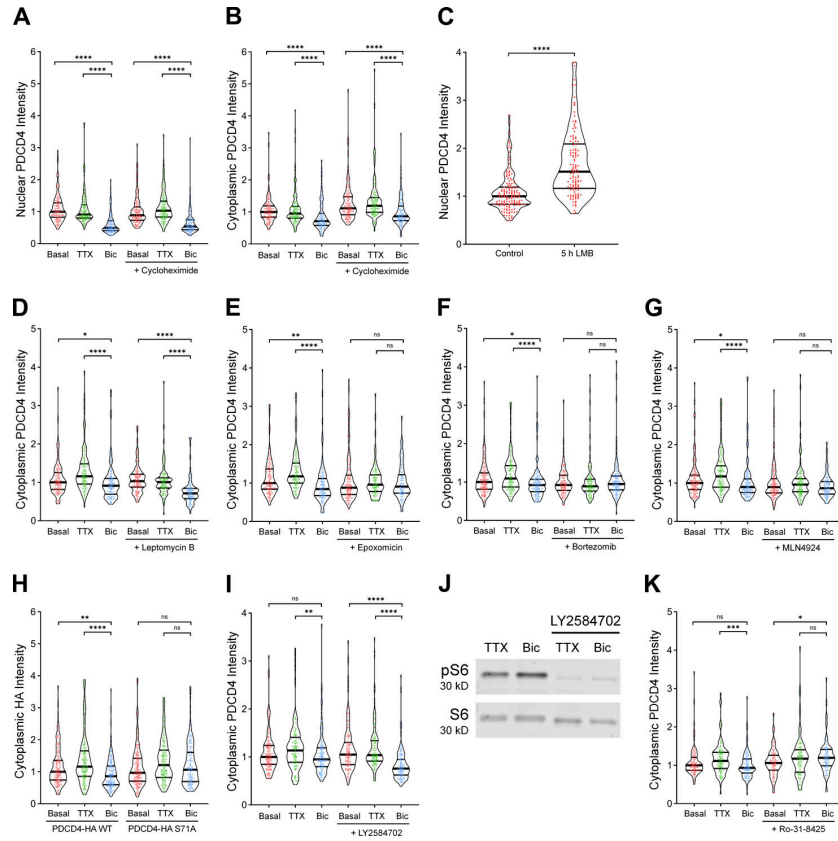
Figure S3. **Verification of nuclear export inhibitor LMB and S6K inhibitor Ly2584702 and quantification of cytoplasmic PDCD4.** Related to Figs. 3 and 4. **(A)** Violin plots of normalized nuclear PDCD4 ICC intensity. Basal, $n = 117$ cells; TTX, $n = 118$ cells; Bic, $n = 109$ cells; CHX-basal, $n = 123$ cells; CHX-TTX, $n = 120$ cells; and CHX-Bic, $n = 104$ cells from 3 sets of cultures. Basal median = 1.00; TTX median = 0.9119; Bic median = 0.4924; CHX-basal median = 0.8890; CHX-TTX median = 1.033; CHX-Bic median = 0.5375. Basal versus Bic, $P < 0.0001$; TTX versus Bic, $P < 0.0001$; CHX-basal versus CHX-Bic, $P < 0.0001$; CHX-TTX versus CHX-Bic, $P < 0.0001$. **(B)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in A. Basal median = 1.00; TTX median = 0.9501; Bic median = 0.7138; CHX-basal median = 1.112; CHX-TTX median = 1.191; CHX-Bic median = 0.8626. Basal versus Bic, $P < 0.0001$; TTX versus Bic, $P < 0.0001$; CHX-basal versus CHX-Bic, $P < 0.0001$; CHX-TTX versus CHX-Bic, $P < 0.0001$. **(C)** Violin plots of normalized nuclear PDCD4 ICC intensity. Control, $n = 137$ cells; and LMB, $n = 122$ cells from three sets of cultures. Control median = 1.00; LMB median = 1.513. Control versus LMB, $P < 0.0001$. **(D)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in Fig. 3 A. Basal median = 1.00; TTX median = 1.158; Bic median = 0.9127; LMB-basal median = 1.031; LMB-TTX median = 1.005; LMB-Bic median = 0.7122. Basal versus Bic, $P = 0.034$; TTX versus Bic, $P < 0.0001$; LMB-basal versus LMB-Bic, $P < 0.0001$; LMB-TTX versus LMB-Bic, $P < 0.0001$. **(E)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in Fig. 3 B. Basal median = 1.00; TTX median = 1.174; Bic median = 0.8439; Epox-basal median = 0.8789; Epox-TTX median = 0.9596; Epox-Bic median = 0.9077. Basal versus Bic, $P = 0.001$; TTX versus Bic, $P < 0.0001$; Epox-basal versus Epox-Bic, $P = 1$; Epox-TTX versus Epox-Bic, $P = 0.8258$. **(F)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in Fig. 3 C. Basal median = 1.00; TTX median = 1.093; Bic median = 0.9226; Bort-basal median = 0.9229; Bort-TTX median = 0.8904; Bort-Bic median = 0.9472. Basal versus Bic, $P = 0.0156$; TTX versus Bic, $P < 0.0001$; Bort-basal versus Bort-Bic, $P = 1$; Bort-TTX versus Bort-Bic, $P = 0.3544$. **(G)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in Fig. 3 D. Basal median = 1.00; TTX median = 1.173; Bic median = 0.8955; MLN-basal median = 0.8940; MLN-TTX median = 0.9603; MLN-Bic median = 0.8633. Basal versus Bic, $P = 0.0324$; TTX versus Bic, $P < 0.0001$; MLN-basal versus MLN-Bic, $P = 0.6294$; MLN-TTX versus MLN-Bic, $P = 0.11$. **(H)** Violin plots of normalized cytoplasmic HA ICC intensity in the same cells as in Fig. 4 C. WT-basal median = 1.00; WT-TTX median = 1.161; WT-Bic median = 0.8621; S71A-basal median = 0.9676; S71A-TTX median = 1.209; S71A-Bic median = 1.063. WT-basal versus WT-Bic, $P = 0.0012$; WT-TTX versus WT-Bic, $P < 0.0001$; S71A-basal versus S71A-Bic, $P = 0.6704$; S71A-TTX versus S71A-Bic, $P = 0.2012$. **(I)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in Fig. 4 E. Basal median = 1.00; TTX median = 1.14; Bic median = 0.95; LY-basal median = 1.05; LY-TTX median = 1.04; LY-Bic median = 0.76. Basal versus Bic, $P = 0.4742$; TTX versus Bic, $P = 0.0088$; LY-basal versus LY-Bic, $P < 0.0001$; LY-TTX versus LY-Bic, $P < 0.0001$. Note that, in these experiments, the basal versus Bic-induced decrease in cytoplasmic PDCD4 is not statistically significant for the untreated (no inhibitor) cells. The Bic-induced cytoplasmic decrease is smaller and less consistent than the nuclear decrease, with statistical significance in the control cells of Fig. S3, E–G, but not Fig. S3, I and K. Nonetheless, there is a significant Bic-induced decrease of cytoplasmic PDCD4 in the treated (LY2584702) cells, demonstrating that S6K is not required for the Bic-induced decrease of PDCD4. **(J)** Western blot of protein lysates from neurons treated with or without LY2584702. Western blot was stained with antibodies for phospho-S6 (Ser235/236) and total S6 to confirm that LY2584702 inhibits S6K activity. **(K)** Violin plots of normalized cytoplasmic PDCD4 ICC intensity in the same cells as in Fig. 4 F. Basal median = 1.00; TTX median = 1.114; Bic median = 0.9326; Ro-31-8425 (Ro)-basal median = 1.059; Ro-TTX median = 1.172; Ro-Bic median = 1.195. Basal versus Bic, $P = 0.2592$; TTX versus Bic, $P = 0.0008$; Ro-basal versus Ro-Bic, $P = 0.0118$; Ro-TTX versus Ro-Bic, $P = 0.9892$. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; and ****, $P < 0.0001$; Mann-Whitney U test with Bonferroni correction.
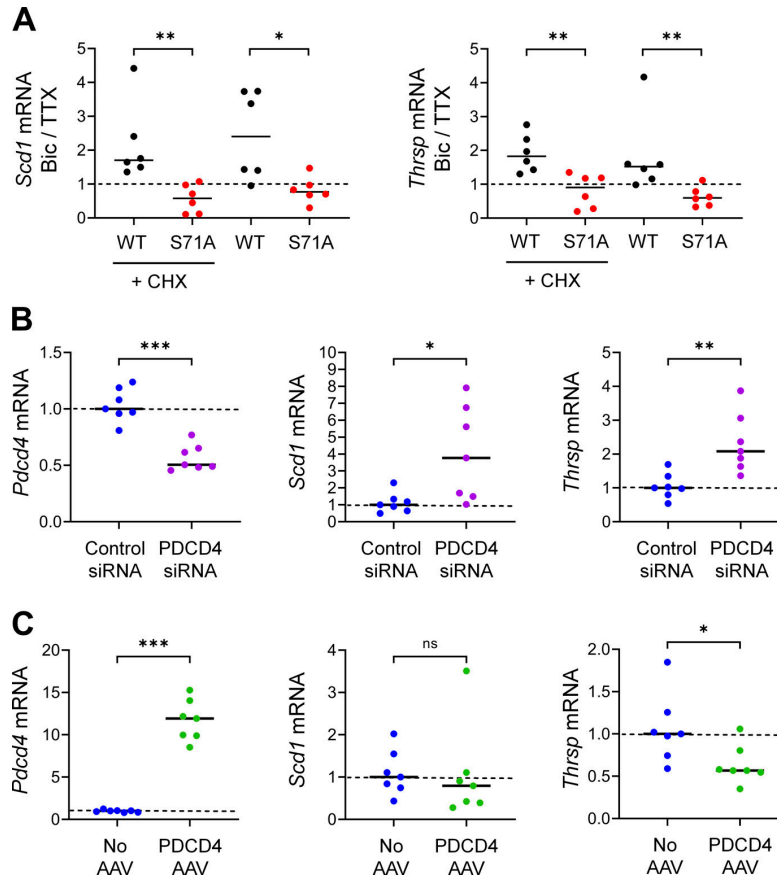
Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology S4
https://doi.org/10.1083/jcb.202103087

88

Figure S4. **RT-qPCR of putative PDCD4 target genes.** Relating to Fig. 6. **(A)** RT-qPCR of putative PDCD4 target genes, *Scd1* and *Thrsp*, from TTX-silenced and Bic-stimulated neurons that were transduced with PDCD4 WT or S71A from six sets of cultures. Samples were normalized using two housekeeping genes, *Hprt* and *Gapdh*. The abundance of the target gene in each Bic sample was normalized to its respective TTX sample. *Scd1* WT CHX median = 1.702; *Scd1* S71A CHX median = 0.5776; *Scd1* WT median = 2.401; *Scd1* S71A median = 0.7672; *Thrsp* WT CHX median = 1.826; *Thrsp* S71A CHX median = 0.9080; *Thrsp* WT median = 1.522; *Thrsp* S71A median = 0.5995. *Scd1* CHX WT versus CHX S71A, P = 0.0022; *Scd1* WT versus S71A, P = 0.0260; *Thrsp* CHX WT versus CHX S71A, P = 0.0043; *Thrsp* WT versus S71A, P = 0.0043. **(B)** RT-qPCR of *Pdcd4*, *Scd1*, and *Thrsp* from PDCD4 knockdown experiments in TTX-silenced neurons from seven sets of cultures. Samples were normalized to the housekeeping gene *Gapdh*. All values were normalized to the median value of the control condition (control siRNA). *Pdcd4* control siRNA median = 1.00; *Pdcd4* PDCD4 siRNA median = 0.51. *Pdcd4* control versus PDCD4 siRNA, P = 0.0006. *Scd1* control siRNA median = 1.00; *Scd1* PDCD4 siRNA median = 3.77. *Scd1* control versus PDCD4 siRNA, P = 0.0111. *Thrsp* control siRNA median = 1.00; *Thrsp* PDCD4 siRNA median = 2.08. *Thrsp* control versus PDCD4 siRNA, P = 0.0023. **(C)** RT-qPCR of *Pdcd4*, *Scd1*, and *Thrsp* from PDCD4 overexpression experiments in TTX-silenced neurons from seven sets of cultures. Samples were normalized to the housekeeping gene *Gapdh*. All values were normalized to the median value of the control condition (no AAV). *Pdcd4* no AAV median = 1.00; *Pdcd4* PDCD4 AAV median = 11.93. *Pdcd4* no AAV versus PDCD4 AAV, P = 0.0006. *Scd1* no AAV median = 1.00; *Scd1* PDCD4 AAV median = 0.80. *Scd1* No AAV versus PDCD4 AAV, P = 0.3176. *Thrsp* no AAV median = 1.00; *Thrsp* PDCD4 AAV median = 0.57. *Thrsp* no AAV versus PDCD4 AAV, P = 0.0262. *, P < 0.05; **, P < 0.01; ***, P < 0.001; Mann-Whitney *U* test.

Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology   S5
https://doi.org/10.1083/jcb.202103087

89

Activity-dependent upregulated genes

Promoters of putative PDCD4 target genes

Motif		% in targets: background	P-value
	JunD	12:2	1E-4
	TATA-Box	42:22	1E-3
	Klf4	30:13	1E-3
	Klf5	66:46	1E-3
	Sp1	36:20	1E-2
	cJun	15:5	1E-2
	CRE	15:5	1E-2
	NFY	31:18	1E-2
	Isl1	45:29	1E-2
	Atf2	15:6	1E-2

Promoters of other CHX-insensitive genes

Motif		% in targets: background	P-value
	CRE	23:6	1E-15
	JunD	13:2	1E-15
	Atf1	30:12	1E-13
	cJun	16:5	1E-9
	Atf7	21:8	1E-8
	Atf2	17:6	1E-8
	NFY	33:17	1E-8
	Sp1	36:22	1E-6
	Klf5	61:46	1E-5
	TATA-Box	36:22	1E-5

Activity-dependent downregulated genes

Promoters of putative PDCD4 target genes

no significant motifs
(too few genes)

Promoters of other CHX-insensitive genes

Motif		% in targets: background	P-value
	Nrf	23:8	1E-6
	Nrf1	20:8	1E-5
	Sp1	34:18	1E-4
	Gfy-Staf	8:2	1E-3
	Klf4	24:13	1E-3
	Klf5	55:42	1E-2
	Gfy	6:2	1E-2
	Elk4	30:20	1E-2

Figure S5. **Motif analysis of promoters of putative PDCD4 target genes.** Related to Fig. 6. Motif analyses of promoters of putative PDCD4 target genes (left column) and, for comparison, other CHX-insensitive activity-dependent genes (right column) using HOMER software. The top panel shows the motif image logos, enrichment, and P values for the top 10 motifs by P value for activity-dependent up-regulated genes, and the bottom panel shows the same but for activity-dependent down-regulated genes (only 8 motifs were significant for down-regulated genes).

Provided online are five tables in Excel files. Table S1 contains the MS data from the nuclear proteomes of silenced and stimulated neurons. Table S2 lists the GO analysis of the APEX2-NLS MS. Table S3 contains the RNA-seq data from silenced and stimulated neurons, with PDCD4 WT or S71A. Table S4 lists the GO analysis for putative PDCD4 target genes. Table S5 lists the primer sequences used for cloning and RT-qPCR.

Herbst et al.
Neuronal activity regulates the nuclear proteome

Journal of Cell Biology   S6
https://doi.org/10.1083/jcb.202103087

90

Chapter 5

High-field asymmetric waveform ion mobility spectrometry interface enhances parallel reaction monitoring on an Orbitrap mass spectrometer

High-field asymmetric waveform ion mobility spectrometry interface enhances parallel reaction monitoring on an Orbitrap mass spectrometer

Weixian Deng[1,2,*], Jihui Sha[1,*], Fanglei Xue[3], Yasaman Jami-Alahmadi[1], Kathrin Plath[1], James Wohlschlegel[1]

1. Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA 90095, USA

2. Molecular Biology Interdepartmental Graduate Program, University of California Los Angeles, Los Angeles, CA 90095, USA

3. University of Technology Sydney, Ultimo NSW 2007, Australia

Abstract. (80–250 words)

High-field asymmetric waveform ion mobility spectrometry (FAIMS) enables gas phase separations on a chromatographic timescale and has become a useful tool for proteomic applications. Despite its emerging utility, however, the molecular determinants underlying peptide separation by FAIMS have not been systematically investigated. Here, we characterize peptide transmission in a FAIMS device across a broad range of compensation voltages (CV) and used machine learning to identify charge state and 3D electrostatic peptide potential as major contributors to peptide intensity at a given CV. We also demonstrate that the machine learning model can be used to predict optimized CV values for peptides which significantly improves parallel reaction monitoring workflows. Together these data provide insight into peptide separation by FAIMS and highlight its utility in targeted proteomic applications.

Introduction

FAIMS differential ion mobility devices typically function as an ion filter placed between the electrospray ionization (ESI) ion source and the mass spectrometer (MS)[1–3]. As ions

move through high and low electric fields of opposite polarity generated by an asymmetric waveform, a small compensation voltage (CV) can be applied to the waveform that enables a subset of ions to travel through the device based on their physicochemical properties while effectively filtering out other ions[4]. By applying different CVs, FAIMS devices are thereby able to fractionate analytes such as peptides. Gas phase fractionation by FAIMS is dictated primarily by the ion charge state of the analyte and allows it to effectively remove singly charged interference ions from the desired peptide analytes and simplify the overall analyte composition entering the mass analyzer[5]. The current commercial FAIMSpro interface being marketed by Thermo Fisher can change CV values in as little as 50 ms enabling its use on a chromatographic timescale but also requires the development of optimized strategies for alternating CVs and maximizing peptide identification[5–7].

Despite the advantages of FAIMS, alternating between CVs slows the duty cycle; therefore, only a limited number of CVs can be used in a given experiment without negatively impacting the number of peptide identifications[8]. This limitation on the number of CVs that can be effectively used in an experiment can lead to peptides being characterized using suboptimal CV settings resulting in compromised sensitivity or a complete failure to detect peptides of interest. A detailed understanding of the determinants that dictate peptide detectability as a function of CV in a FAIMS-based fractionation is still largely lacking but considered essential for maximizing the utility of these analyses. In this study, we analyzed peptide intensity distributions derived from proteomic characterization of the human proteome carried out over a range of different

CVs. These analyses identified 3D electrostatic descriptors as determinants of peptide isoform detectability on FAIMS instruments.

Targeted proteomic methods such as parallel reaction monitoring (PRM) also depend on the selection of optimal CV settings to maximize sensitivity and the lower limit of quantitation (LLOQ) for each target. However, in many cases the optimal CV for a given peptide target is not empirically known as its selection was based on datasets collected on instruments using generally applied CV settings or not equipped with FAIMS. This uncertainty regarding the ideal CV settings for a given panel of peptides has the potential to greatly diminish the effectiveness of the analysis. In this study, we built a machine learning model to predict a peptide's peak intensity given its sequence and charge state as a function of CV values and demonstrate its utility in a series of proteomics analyses.

Experimental Section

Sample preparation

HEK293 cells were cultured in high glucose and glycine DMEM containing 10% FBS and 1% penicillin-streptomycin and then harvested by trypsinization. Cells were lysed by incubation in lysis buffer (8M Urea, 0.1M Tris-HCl pH 8.0) at 4°C for 30 minutes followed by centrifugation to clarify the sample. Two milligrams of protein were reduced and alkylated by sequentially incubating with 5mM TCEP and 10mM iodoacetamide for 30 minutes at room temperature in the dark. The protein sample was then diluted 4-fold with 0.1M Tris-HCl pH8.0 to reduce the final urea concentration to 2M before incubating

overnight at 37° C with 50 μg trypsin protease. Peptide digests were desalted using Pierce C18 tips (100ul bed volume), dried, and then reconstituted in water.

LC-MS/MS

Liquid Chromatography Settings

Short gradient. 800 ng of a tryptic peptide digest prepared from whole cell lysates of HEK293 cells was loaded onto a 25-cm long, 75 μm inner diameter fused-silica capillary, packed in-house with bulk 1.9 μM ReproSil-Pur C18 beads. Peptides were delivered by using a Thermo Scientific™ EASY-nLC™ 1200 HPLC system. The mobile phase buffers are buffer A (water solution with 3% DMSO and 0.1% formic acid) and buffer B (80% acetonitrile solution with 3% DMSO and 0.1% formic acid). The 70-min short gradient is delivered as follows: 1-6% buffer B from minutes 0-6 at a flowrate of 300 nL/min, 6-28% buffer B from minutes 6-54 at a flowrate of 220 nL/min, 28-32% buffer B from minutes 54-62 at a flowrate of 220 nL/min, and 85% buffer B from 62-70 minutes at a flowrate of 220 nL/min.

Long gradient. Same amount of lysate digest was loaded to the same column. Then peptide analyte was eluted using a 140-min gradient of increasing acetonitrile (ACN) as follows: 350 nL/min 2% buffer B from 0 min, 2-6% buffer B 300 nL/min from minutes 0 to 2, 2-35% buffer B from minutes 2 to 116, 35-42% buffer B from minutes 116 to 136, 42-85% buffer B from minutes 136-138, 85-95% buffer B from minutes 138-140 and switch the flow rate to 350nL/min.

Data Acquisition Settings

Data-dependent Acquisition. The eluted peptides were ionized using a Thermo Scientific™ Nanospray Flex™ ion source and injected into a Thermo Scientific™ Orbitrap Eclipse™ Tribrid™ mass spectrometer operated in positive mode equipped with a FAIMSpro interface. Spectra were acquired using data-dependent acquisition mode where a 120k resolution full MS1 scan was followed by sequential MS2 scans at a resolution of 15,000 to utilize the remainder of the 3 second cycle time. MS/MS spectra were collected using a 1.5 m/z window for precursor ion quadrupole isolation and normalized HCD collision energy of 30% with dynamic exclusion of 10s and monoisotopic peak determination set to peptide. The "auto" maximum injection time was selected to allow the orbitrap to calculate the maximum injection time available to maximize sensitivity while maintaining maximum scan rate. FAIMS separations were performed using FAIMS mode on standard resolution set to static gas mode with nitrogen carrier gas flow of 0 L/min and inner and outer electrodes temperature of 100°C with asymmetric dispersion voltage (DV) of −5000 V. In order to selectively filter ions that enter the mass spectrometer, individual compensation voltages (CVs) between the range of −25 V and −70 V were applied to sequential survey scans and MS/MS cycles.

Parallel Reaction Monitoring. LC method is set the same as DDA experiments. Target precursor ions for different experiments are shown in Supplemental Table 1. Precursors are fragmented by HCD at 30% collision energy and then analyzed using the targeted MS2 mode in Xcalibur at 15000 resolution on a Orbitrap mass analyzer. PRM experiments were analyzed with Skyline using mProphet models[9] trained using the second best peak.

Data Analysis

Machine learning. Features for building machine learning models were obtained from the Peptides package. Machine learning models including the linear and stacked ensemble models are built through H2O-AI (3.36.0.1). The empirical constraints used for the stacked ensemble model 2 are generated by +/- 5V centered by the prediction reported from stacked ensemble model 1.

Database searching. MS/MS database searching was performed using MaxQuant (1.6.10.43) against the human reference proteome from EMBL (UP000005640_9606 HUMAN Homo sapiens, 20874 entries). The search included carbamidomethylation on cysteine as a fixed modification and methionine oxidation and N-terminal acetylation as variable modifications. The digestion mode was set to trypsin and allowed a maximum of 2 missed cleavages. The precursor mass tolerances were to 20 and 4.5 ppm for the first and second searches, respectively, while a 20-ppm mass tolerance was used for fragment ions. Datasets were filtered at 1% FDR at both the PSM and protein-level. Peptide quantitation was performed using MaxQuant's LFQ mode.

Results and Discussion

Systematic investigation of peptide detectability across different CV values

Previous studies have shown that peptide identification can vary widely based on the CV values employed in the analysis[5–7]. However, peptide detectability has not been systematically investigated as a function of CV. To address this gap, we examined peptide detectability across a range of CVs using two commonly used LC-MS settings.

We analyzed HEK293 whole cell lysate peptide digests by LC-MS/MS using data-dependent analysis and label free quantification with (1) 70 minute LC gradients across 9 CV values (-20V to -70V in -5V increment) with a single CV per injection and (2) 140 minute LC gradients across 18 CV values (-20V to -76V in -3V increments) with three alternating CVs per injection (Table 1). To examine the peptide intensity distributions, we first filtered the two datasets to only include peptides detected in at least 3 different CV values which included 56.02% of the peptides identified in the short gradient dataset and 57.43% of peptides in the long gradient dataset. Peptide intensities were standardized by transforming to the proportion of the corresponding highest intensities across the CV range, classified by the CV value where maximum intensity was observed, and then plotted (Figure 1A). For this analysis, we considered different charge state isoforms of a given peptide e states as separate peptides. Excluding both ends of the CV range (-25V and -76V), the peptide intensities of individual CV values appear as relatively narrow, well-defined bell-shaped distributions.

From these data, we observe several key features related to how peptide detectability is influenced by CV. First, we find that almost all peptides are detectable in an approximately 20V span of CV values centered around the CV value that corresponds to their maximum intensity. (94.26% of peptides under 20V in the short gradient data and 95.74% peptides in the long gradient data) (Figure 1B). Second, we explored the correlation between a peptide's maximum intensity and its detectable CV span (Figure 1C) and observed a clear trend in which peptides with a higher detectable intensity exhibit wider CV spans. Finally, we asked how necessary it is to know a peptide's ideal CV to detect it at its highest intensity. Therefore, we measured the percentage of the intensity for each peptide found

in its most intense CV bin relative to its total intensity across all CV values. As shown in Figure 1D, peptide intensity in the most intense CV bin accounts for over 50% of total detectable intensity for a large fraction of identified peptides (47.1% for the long gradient, 66.1% for short gradient) suggesting that maximizing the sensitivity of detection for any given peptide will require sampling of that peptide at its optimized CV value. From these data, we conclude that many of the peptides observed in a standard proteomic analysis are detectable only in a relatively narrow CV range and that approaches to maximizing sensitivity will require integration of this knowledge into data acquisition strategies.

Linear model suggests main determination factors for peptide's peak intensity CV

Previous reports[5] have shown that a peptide isoform's ability to pass through the FAIMS device at a given CV value can be weakly correlated with charge state. However, other physicochemical properties displaying a clear correlation with FAIMS transmission have not been reported. Using a machine learning approach, we attempted to identify additional determinants that govern peptide transmission by FAIMS. We built a linear regression model to identify factors that impact peptide detectability and then used Lasso regression to minimize the number of factors in the model.

The goal of the model was to use peptide sequence and charge state to predict the CV value where a given peptide can be detected at maximum intensity. We first built a pool of parameters for each peptide isoform. In addition to the observed charge state from MS, we calculated 76 different parameters for each peptide comprised of 18 peptide amino acid composition features, 7 basic physicochemical properties[10–12], 10 Quantitative Structure-Activity Relationship (QSAR)[13–19] descriptors based on50 peptide sequence

features for a total of 76 parameters for each peptide (Figure 2A). Since the data used to build the model was acquired under discrete CV values and could potentially miss the CV corresponding to maximal intensity using our approach, we calculated the weighted average CV values for the peak in addition to the observed CV value of the peak intensity. For peptides detected at only a single CV value, we tested the effects of excluding or including them in the model. We then trained the Lasso regression using either the weighted average CV or the observed CV corresponding to highest peak intensity with or without inclusion of the single observation data using the two datasets quantified under different LC gradients (Figure 2A) for a total of 8 different trained linear models (Table 2).

As shown in Figure 2B, models trained from both short and long gradient datasets show a linear correlation between peptide parameters and weighted average CV of peak intensity/observed peak intensity CV with the short gradient data showing overall better linearity. For the mean absolute error (MAE) of training sets (randomly selected 80% of data), the short gradient showed lower error than the long gradient data. Additionally, filtering out single observation data resulted in slightly better prediction accuracy across training sets. The model showing the highest R^2 (0.56) and the lowest MAE (5.89) was trained with the short gradient dataset and lacking the single observation data (Figure 2C). From this model, we observed that charge state is the highest contributor across all used parameters consistent with previous studies. Interestingly, we see MS-WHIM1 and MS-WHIM3, which are both descriptors for the 3D molecule surface electrostatic potential[20,21], ranked second and third suggesting that charge distribution across the 3D structure of peptides significantly contributes to efficient FAIMS transmission. In addition, PP3, peptide predicted charge when pH=7, the Cruciani property H-bonding descriptor,

and other QSAR features also contributed to the model. The top contributors found across all 8 models are shown in Supplemental Table 1.

Next, we evaluated the quality of the model's prediction by examining peak peptide intensities observed at different CV values. As shown in Figure 2D, we determined the predicted weighted average peak intensity CV values for all peptides from the short gradient dataset lacking the single observation data and then calculated the absolute difference between the CV values corresponding to the observed and predicted weighted average peak intensity from the model. These data were used to build an accumulation curve to determine the proportion of peptides that exist below a given difference in CV values. Peak intensities determined for CV settings higher than -55V are generally of high quality with ~90% of peptide isoforms having an absoluteerror of less than 10V. However, when the peak intensity CV was below -55V, the linear model prediction quality diminished quickly, suggesting that other as of yet unaccounted for factors influence FAIMS transmission at the low end of the CV range.

Deep learning model predicts peptide's best CV value for improving PRM performance

Targeted proteomic techniques such as parallel reaction monitoring have emerged as powerful analytical tools for monitoring the abundance of discrete sets of peptides in a high throughput and sensitive manner. Although the utility of FAIMS in targeted proteomic workflows may offer significant advantages with respect to sensitivity, the integration of these technologies has not been widely reported. Based on our observation that peptide intensity peaks within a relatively narrow CV range, we hypothesized that targeted

proteomic assays would require accurate prediction of a peptide's optimal CV value in order to maximize the benefit from FAIMS-based analyses. However, the 5.89V MAE obtained from linear model built above may not be sufficient. Therefore, we trained two additional machine learning models that can provide higher accuracy at the expense of knowing the relative contribution from different determinants.

We first trained a stacked ensemble model (model 1) with the same 76 features mentioned above which resulting in an overall MAE of 3.81V n the test set data (Figure 3B, left). However, we noticed that for peptides with a weighted average peak intensity CV lower than -55V, the prediction error increased significantly (Figure 3C, left). Therefore, we examined whether prediction could be improved by the inclusion of an additional constraint where we included an empirical determined CV value at which the given peptide was detected irrespective of whether this was the CV where the maximum intensity is observed. Essentially, based on the prediction error in model 1, we take one of possible 3 CV values (the predicted CV from model 1 and 5V above and below it) corresponding to the CV closest to the observed peak intensity and use it as an empirical constraint to train model 2. As shown in Figure 3C right, this approach generated a more accurate model with lower prediction errors but similar performance in all observed peak intensity CV bins.

To optimize CV usage in the context of PRM experiments, we propose to integrate these two machine learning models into a single workflow (Figure 3A). Researchers commonly select peptides for inclusion in a PRM target list based on experimental data (DDA or PRM/SRM) collected on instruments not equipped with FAIMS or predicted peptide

sequences lacking any empirical data supporting their identification. For these peptides, we utilize the peptide sequence and charge state (either measured or predicted) to model 1. Model 1 will generate a low precision CV prediction that can either (1) be used directly in a targeted assay or (2) used as a starting point for additional experiments to narrow down its optimal CV. In the latter case, a 3-alternating-CV run (DDA with inclusion list or PRM) using the predicted CV value as well CV values shifted 5V in either direction to determine the detectability of the peptide of interest. If the peptide is detected, then its empirically determined CV value (even if not optimized) can be used as a constraint in model 2 to predict higher precision CV values for future targeted assays. We simulated the process on the short gradient data where over 92% of peptides can be covered by the 3 CV values used in the pilot experiment and the overall MAE for the high precision prediction is 1.48V. The precision for each predicting in a given observed peak intensity CV bin is shown in Figure 3C while Figure 3D shows the prediction error for the whole short gradient dataset.

FAIMS improves PRM performance

We next examined the significance of FAIMS-based separations on targeted PRM experiments as well as test the utility of our CV prediction models. First, we tested the performance of PRMs in the presence and absence of a FAIMS device using a Orbitrap (OT) mass analyzer as implemented on the Thermofisher Fusion Eclipse mass spectrometer to explore the extent to which FAIMS might improve PRM analyses. We selected 85 peptide isoforms whose maximal peak intensity occurred at a range of different CV values as targets for the experiment. Representative extracted ion

chromatograms are shown in Figure 4A and demonstrate that FAIMS can significantly improve PRM data quality. To quantify the performance improvement provided by FAIMS, we built mProphet models and compared q-values for target peptides between FAIMS and no FAIMS datasets. The use of a FAIMS device significantly boosted q-values highlighting the improved data quality (p-value = 0.018) (Figure 4B). To further characterize this improvement, we examined individual mProphet features at the fragment ion transition level and found that peak shape and signal to noise ratio are significantly higher for FAIMS vs no FAIMS with p-values of 0.0019 and 2.2e-16, respectively (Figure 4C). These data suggest that the use of a FAIMS device significantly improves PRM analyses.

We next explored the importance of selecting accurate CVs in a PRM analysis. To this end, we utilized our machine learning model to build a predictor for peptide CV based on its CV and applied it to the peptide panel described in Figure 3. Figure 4D shows a comparison between the empirically determined and predicted CV values for these peptides. These data validate the predictor and demonstrate its ability to accurately predict CV values based on peptide sequence. Next, we performed a parallel reaction monitoring experiment targeting this peptide panel using FAIMS with both the predicted CV and a CV value shifted by 10V. We reasoned that if accurate CV settings were necessary for an optimized PRM experiment, then the dataset with the predicted CV values should be significantly better than the dataset with the shifted less optimal CV values. The mProphet Q value distribution for both the predicted CV and shifted CV datasets are shown in Figure 4E and shows that the data quality of target peptides is significantly increased when data is acquired using the predicted CV values. This

observation is also clearly observed in scatterplots comparing the intensity of individual peptide fragment ion transitions between predicted and shifted CV experiments (Fig. 4E). Together these data emphasize the importance of acquiring data using optimized CV values and that this process can be facilitated using the CV predictor developed based on our machine learning model.

Conclusions

In this study, we systematically profiled the intensity distribution of peptides during FAIMS separation across different CV settings and found that the majority of peptides is efficiently transmitted through relatively narrow CV windows. Using machine learning approaches, we showed that the charge state together with 3D descriptors such as MS-WHIM scores are the top determinants for dictating peptide isoform transmission across CV space and that this information can be used to accurately predict optimal peptide CV from a peptide's sequence. Furthermore, we find that the proper selection of CV values is essential for PRM assay data quality with optimized CV setting leading to significant improvements in detection and quantitation.

Contributions

W. D. and J. A. W. designed the study and analyzed the data. W. D., J. S. and J. Y. conducted the experiments. W. D. and F.X. conducted the machine learning analysis. W.D., K.P., and J.A.W. wrote the manuscript.

Figure Legends

Figure 1: Systematic investigation of peptides' detectability across different CV values. (A) Normalized intensity distributions for peptides detected at least 3 different CV values. Peptide intensities are normalized to the maximum observed intensity to ensure all intensity values are between 0 to 1. Peptides are binned by observed peak intensity CV with with a red line connecting the median intensities of the detected peptides across CV values. **(B)** Accumulation curve of the percentage of peptides identified in a given CV span for both the short gradient (red) and long gradient (blue) datasets. **(C)** Log2-transformed Intensity distribution of peptides identified at different CV values. The green line indicates the trend. **(D)** Histogram of the number of peptides binned by the fraction of the peptide intensity of the most intense CV bin compared to the total intensity of that peptide across all bins for the long gradient (blue) and short gradient (red) datasets.

Figure 2: Linear model suggests main determination factors for peptide's peak intensity CV. (A) Schematic of the machine learning approach for building the linear model. **(B)** R^2 of the 8 different linear models built with (1) different datasets, (2) filtered by only observed under one CV value or not, and (3) either the observed intensity or weighted average peak intensity CV. **(C)** Standardized coefficients for the top 15 parameters of the model built by short gradient dataset after filtering out single observations and using the weighted average peak intensity CV. **(D)** Accumulation curves of the percentage of peptides with increasing absolute prediction error. Each line is binned by peptide's observed peak intensity CV.

Figure 3: Higher order machine learning models predict CV corresponding to maximum peptide peak intensity. (A) Scheme for using stacked ensemble models to predict peptide peak intensity at different CV values. **(B)** Correlation of model 1 and 2's predicted peak intensity CV with the observed weighted average peak intensity CV. **(C)** Accumulation curves of the percentage of peptide with increasing absolute prediction error for weighted average peak intensity CV values predicted by model 1 (left) and model 2 (right). Each line is binned by observed peptide peak intensity CV. **(D)** Prediction error histogram for all peptides in the short gradient dataset.

Figure 4: FAIMS and suitable CV settings improve PRM performance. (A) XIC of the peptide: TLFPGTDHIDQLK acquired with and without FAIMS enabled. **(B)** mProphet Q value distribution (-Log10 transformed) for all of the target peptides analyzed using parallel reaction monitoring without and without FAIMS. **(C)** Distribution of peak shape scores (left) and signal to noise ratio score (right) for all extracted fragment ion transitions from PRM data acquired with (blue) and without (red) FAIMS. **(D)** Observed CV vs. predicted weighted average peak intensity CV for PRM target peptides. **(E)** Distribution of the -log10 (mProphet q-values) for PRM target peptides using the CV values either predicted by machine learning model (red) or offset by +10V from the predicted CV (blue). **(F)** Scatter plot of the mean area of fragment ion transitions from PRM target peptides where data was acquired using the predicted CV (x-axis) and the predicted CV offset by +10V (y-axis).

References

(1) Guevremont, R. High-Field Asymmetric Waveform Ion Mobility Spectrometry: A New Tool for Mass Spectrometry. *J Chromatogr A* 2004, *1058* (1–2), 3–19. https://doi.org/10.1016/j.chroma.2004.08.119.

(2) Purves, R. W.; Guevremont, R. Electrospray Ionization High-Field Asymmetric Waveform Ion Mobility Spectrometry−Mass Spectrometry. *Anal Chem* 1999, *71* (13), 2346–2357. https://doi.org/10.1021/ac981380y.

(3) Dodds, J. N.; Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J Am Soc Mass Spectr* 2019, *30* (11), 2185–2195. https://doi.org/10.1007/s13361-019-02288-2.

(4) Barnett, D. A.; Ells, B.; Guevremont, R.; Purves, R. W. Application of ESI-FAIMS-MS to the Analysis of Tryptic Peptides. *J Am Soc Mass Spectr* 2002, *13* (11), 1282–1291. https://doi.org/10.1016/s1044-0305(02)00527-5.

(5) Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J. J.; Thibault, P. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics* 2018, *17* (10), 2051–2067. https://doi.org/10.1074/mcp.tir118.000862.

(6) Bekker-Jensen, D. B.; Martínez-Val, A.; Steigerwald, S.; Rüther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol Cell Proteomics* 2020, *19* (4), 716–729. https://doi.org/10.1074/mcp.tir119.001906.

(7) Schweppe, D. K.; Prasad, S.; Belford, M. W.; Navarrete-Perea, J.; Bailey, D. J.; Huguet, R.; Jedrychowski, M. P.; Rad, R.; McAlister, G.; Abbatiello, S. E.; Woulters, E. R.; Zabrouskov, V.; Dunyach, J. J.; Paulo, J. A.; Gygi, S. P. Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Anal Chem* 2019, *91* (6), 4010–4016. https://doi.org/10.1021/acs.analchem.8b05399.

(8) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* 2018, *90* (15), 9529–9537. https://doi.org/10.1021/acs.analchem.8b02233.

(9) Reiter, L.; Rinner, O.; Picotti, P.; Hüttenhain, R.; Beck, M.; Brusniak, M.-Y.; Hengartner, M. O.; Aebersold, R. MProphet: Automated Data Processing and Statistical Validation for Large-Scale SRM Experiments. *Nat Methods* 2011, *8* (5), 430–435. https://doi.org/10.1038/nmeth.1584.

(10) Bjellqvist, B.; Hughes, G. J.; Pasquali, C.; Paquet, N.; Ravier, F.; Sanchez, J.; Frutiger, S.; Hochstrasser, D. The Focusing Positions of Polypeptides in Immobilized PH Gradients Can Be Predicted from Their Amino Acid Sequences. *Electrophoresis* 1993, *14* (1), 1023–1031. https://doi.org/10.1002/elps.11501401163.

(11) Juretić, D.; Lučić, B.; Zucić, D.; Trinajstić, N. Protein Transmembrane Structure: Recognition and Prediction by Using Hydrophobicity Scales through Preference Functions. *Theor Comput Chem* 1998, *5*, 405–445. https://doi.org/10.1016/s1380-7323(98)80015-0.

(12) Guruprasad, K.; Reddy, B. V. B.; Pandit, M. W. Correlation between Stability of a Protein and Its Dipeptide Composition: A Novel Approach for Predicting in Vivo Stability of a Protein from Its Primary Sequence. *Protein Eng Des Sel* 1990, *4* (2), 155–161. https://doi.org/10.1093/protein/4.2.155.

(13) Cruciani, G.; Baroni, M.; Carosati, E.; Clementi, M.; Valigi, R.; Clementi, S. Peptide Studies by Means of Principal Properties of Amino Acids Derived from MIF Descriptors. *J Chemometr* 2004, *18* (3-4), 146–155. https://doi.org/10.1002/cem.856.

(14) Liang, G.; Li, Z. Factor Analysis Scale of Generalized Amino Acid Information as the Source of a New Set of Descriptors for Elucidating the Structure and Activity Relationships of Cationic Antimicrobial Peptides. *Qsar Comb Sci* 2007, *26* (6), 754–763. https://doi.org/10.1002/qsar.200630145.

(15) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. The Hydrophobic Moment Detects Periodicity in Protein Hydrophobicity. *Proc National Acad Sci* 1984, *81* (1), 140–144. https://doi.org/10.1073/pnas.81.1.140.

(16) Yang, L.; Shu, M.; Ma, K.; Mei, H.; Jiang, Y.; Li, Z. ST-Scale as a Novel Amino Acid Descriptor and Its Application in QSAM of Peptides and Analogues. *Amino Acids* 2010, *38* (3), 805–816. https://doi.org/10.1007/s00726-009-0287-y.

(17) Tian, F.; Zhou, P.; Li, Z. T-Scale as a Novel Vector of Topological Descriptors for Amino Acids and Its Application in QSARs of Peptides. *J Mol Struct* 2007, *830* (1–3), 106–115. https://doi.org/10.1016/j.molstruc.2006.07.004.

(18) Mei, H.; Liao, Z. H.; Zhou, Y.; Li, S. Z. A New Set of Amino Acid Descriptors and Its Application in Peptide QSARs. *Peptide Sci* 2005, *80* (6), 775–786. https://doi.org/10.1002/bip.20296.

(19) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A Multivariate Characterization of 87 Amino Acids. *J Med Chem* 1998, *41* (14), 2481–2491. https://doi.org/10.1021/jm9700575.

(20) Zaliani, A.; Gancia, E. MS-WHIM Scores for Amino Acids: A New 3D-Description for Peptide QSAR and QSPR Studies. *J Chem Inf Comp Sci* 1999, *39* (3), 525–533. https://doi.org/10.1021/ci980211b.

(21) Zaliani, A.; Gancia, E. QSPR Studies. *J. Chem. Inf. Comput. Sci.* 1999, 525–533.

Table 1

Injection No.	Single CV, 70 min/run (V)	3 CV, 140 min/run (V)
1	-25	-25, -28, -31
2	-30	-34, -37, -40
3	-35	-43, -46, -49
4	-40	-53, -55, -58
5	-45	-61, -64, -67
6	-50	-70, -73, -76
7	-55	
8	-65	
9	-70	

Table 2

Model	Dataset	Single observation filtered	Weighted average	Measured
1	Short gradient	+	+	
2	Short gradient	+		+
3	Short gradient		+	
4	Short gradient	+		+
5	Long gradient	+	+	
6	Long gradient	+		+
7	Long gradient		+	
8	Long gradient	+		+

1A

1B

1C

1D

Figure 1: Systematic investigation of peptides' detectability across different CV values.
(A) Normalized intensity distributions for peptides detected at least 3 different CV values.
Peptide intensities are normalized to the maximum observed intensity to ensure all intensity
values are between 0 to 1. Peptides are binned by observed peak intensity CV with with a red
line connecting the median intensities of the detected peptides across CV values. **(B)**
Accumulation curve of the percentage of peptides identified in a given CV span for both the
short gradient (red) and long gradient (blue) datasets. **(C)** Log2-transformed Intensity
distribution of peptides identified at different CV values. The green line indicates the trend. **(D)**
Histogram of the number of peptides binned by the fraction of the peptide intensity of the most
intense CV bin compared to the total intensity of that peptide across all bins for the long gradient
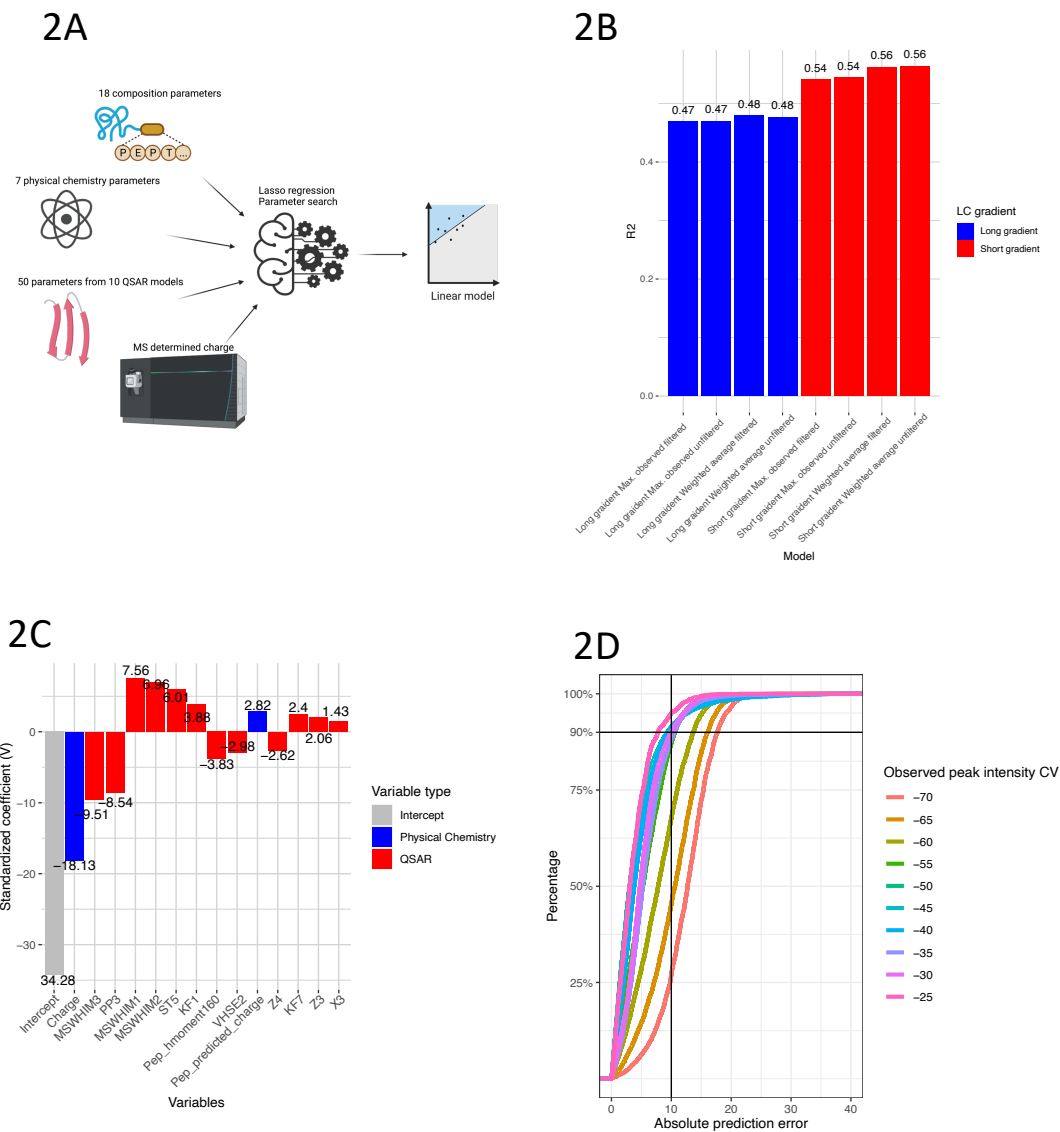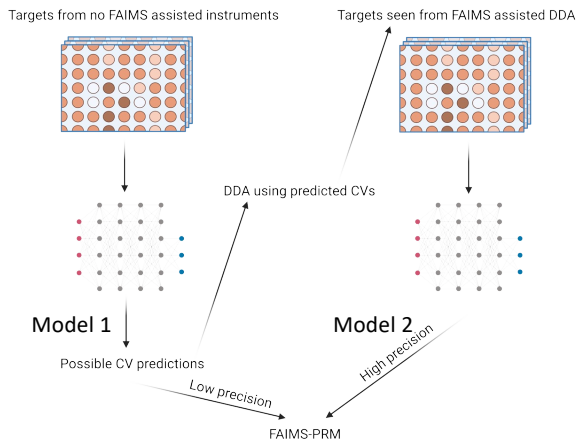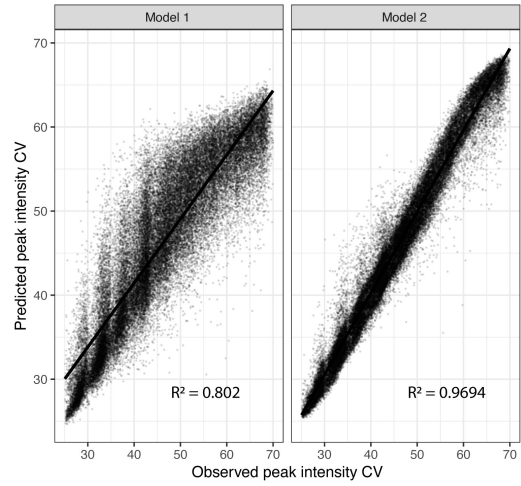(blue) and short gradient (red) datasets.

Figure 1

Figure 2: Linear model suggests main determination factors for peptide's peak intensity CV. (A) Schematic of the machine learning approach for building the linear model. **(B)** R^2 of the 8 different linear models built with (1) different datasets, (2) filtered by only observed under one CV value or not, and (3) either the observed intensity or weighted average peak intensity CV. **(C)** Standardized coefficients for the top 15 parameters of the model built by short gradient dataset after filtering out single observations and using the weighted average peak intensity CV. **(D)** Accumulation curves of the percentage of peptides with increasing absolute prediction error. Each line is binned by peptide's observed peak intensity CV.
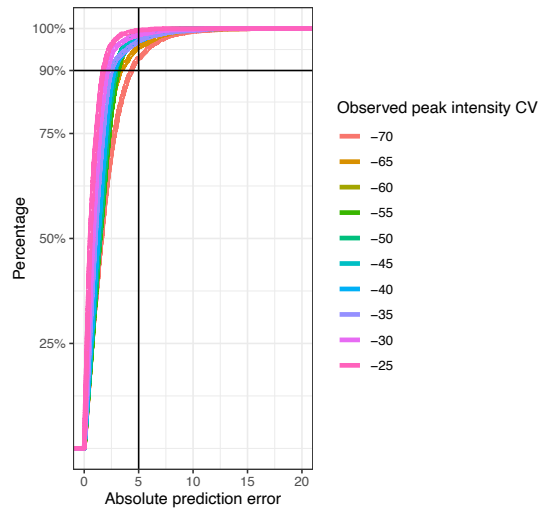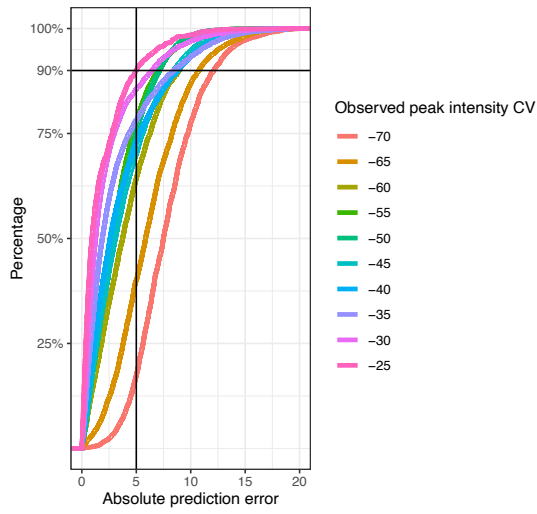
Figure 2

113

3A
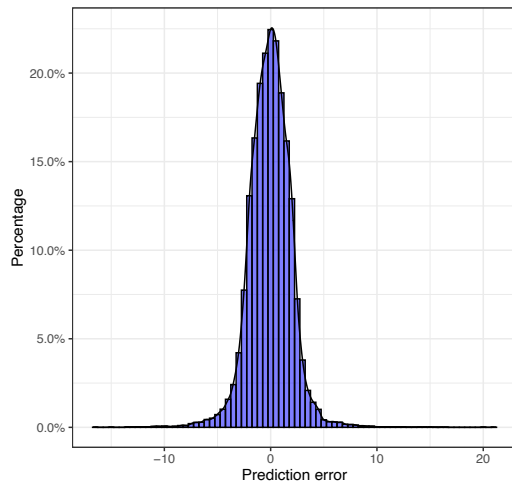
Targets from no FAIMS assisted instruments

Targets seen from FAIMS assisted DDA

DDA using predicted CVs

Model 1

Model 2

Possible CV predictions

High precision

Low precision

FAIMS-PRM

3B

Model 1   Model 2

Predicted peak intensity CV

Observed peak intensity CV

R² = 0.802   R² = 0.9694

3C

Observed peak intensity CV

Percentage

Absolute prediction error

−70, −65, −60, −55, −50, −45, −40, −35, −30, −25

3D

Percentage

Prediction error

Figure 3

Figure 3: Higher order machine learning models predict CV corresponding to maximum peptide peak intensity. (A) Scheme for using stacked ensemble models to predict peptide peak intensity at different CV values. **(B)** Correlation of model 1 and 2's predicted peak intensity CV with the observed weighted average peak intensity CV. **(C)** Accumulation curves of the percentage of peptide with increasing absolute prediction error for weighted average peak intensity CV values predicted by model 1 (left) and model 2 (right). Each line is binned by observed peptide peak intensity CV. **(D)** Prediction error histogram for all peptides in the short gradient dataset.
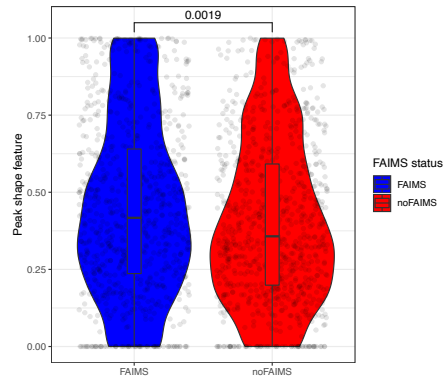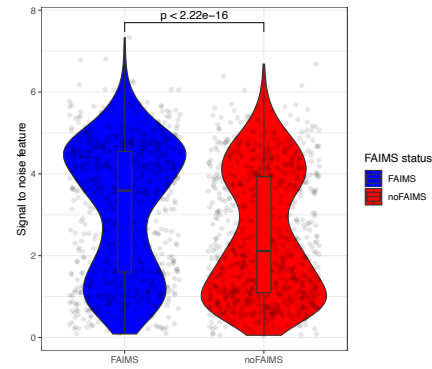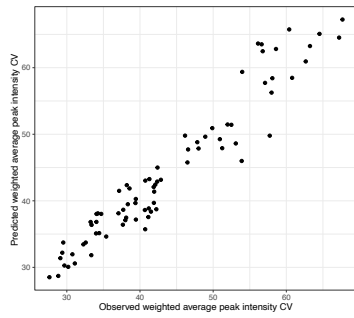
Figure 3

4A TLFPGTDHIDQLK

FAIMS

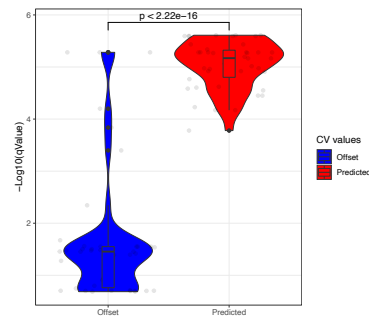noFAIMS
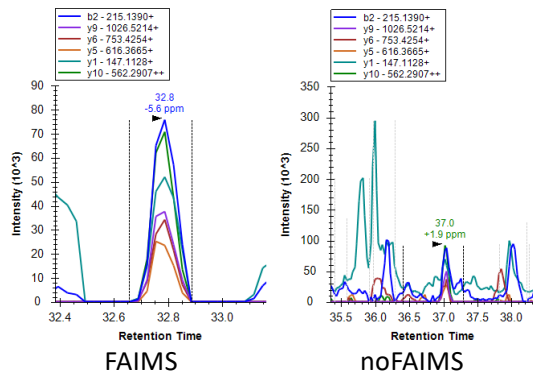
4B
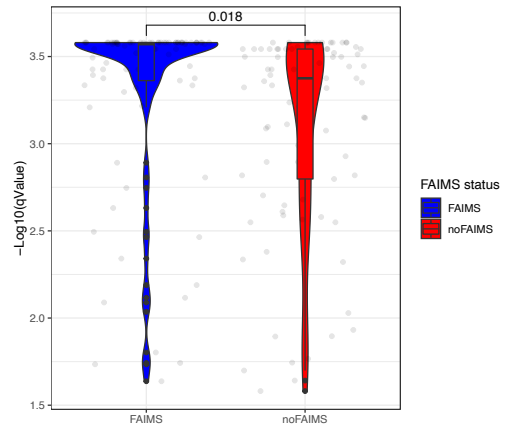
4C

Peak shape

Signal to noise ratio
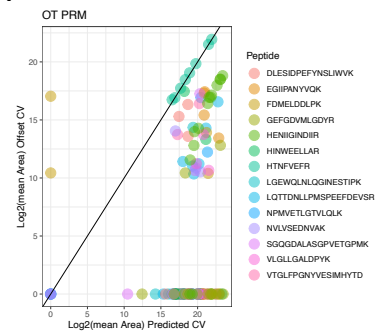
4D

4E

4F

Figure 4

Figure 4: FAIMS and suitable CV settings improve PRM performance. (A) XIC of the peptide: TLFPGTDHIDQLK acquired with and without FAIMS enabled. **(B)** mProphet Q value distribution (-Log10 transformed) for all of the target peptides analyzed using parallel reaction monitoring without and without FAIMS. **(C)** Distribution of peak shape scores (left) and signal to noise ratio score (right) for all extracted fragment ion transitions from PRM data acquired with (blue) and without (red) FAIMS. **(D)** Observed CV vs. predicted weighted average peak intensity CV for PRM target peptides. **(E)** Distribution of the -log10 (mProphet q-values) for PRM target peptides using the CV values either predicted by machine learning model (red) or offset by +10V from the predicted CV (blue). **(F)** Scatter plot of the mean area of fragment ion transitions from PRM target peptides where data was acquired using the predicted CV (x-axis) and the predicted CV offset by +10V (y-axis).

Figure 4

Chapter 6

Conclusion

The mechanism of cell fate determination has long been a question being studied by developmental biologists. Since 2006[1,2], Yamanaka and his colleagues published the first iPSC paper inducing terminally differentiated cells to pluripotency stage by expressing a cocktail of transcriptional factors (TFs), as a reverse process to differentiation led by developmental clues, the mechanism of how OSKM, just four TFs, can make this artificial cell fate transition happen has raised a lot of attention[3–6]. By combining genomics, proteomics, cell biology and molecular biology approaches, we are able to create some new insights to answer this question. Another aspect of my thesis work is towards developing and optimizing bottom-up proteomics approaches to address biological question by increasing sensitivity and quantitation accuracy of existing workflows.

Essential Reprogramming Factors Redistribute Somatic Transcription Factors in Early Reprogramming through Multiple Molecular Mechanisms

In Chapter 2, we first leveraged the optimized ChIP-SICAP workflow [7] confirmed that O S and K co-bind with many somatic TFs on the chromatin in the early stage of reprogramming. Next, an *in vivo* validation is conducted by inducing expression of OSK combo under a Bl6 and PWK crossed genetic background. Under this background, SNPs between both alleles disrupt TF binding motifs and cause loss of binding of TFs from one of the alleles but not the other. Then, through testing O S and K binding sites difference between alleles, we can correlate somatic TF motif disruption with the OSK binding difference between alleles, and we found OSK's binding skewness towards one of the two alleles is depending on the motif disruption of a subset of somatic TFs including AP-1, CEBP and TEAD family members. Furthermore, to test if such binding dependency is

caused by direct protein-protein interactions (PPIs) between OSK and somatic TFs, we used purified recombinant proteins and performed pairwise *in vitro* pull-down assay between O and somatic TFs obtained from the allelic specific binding analysis and confirmed that Oct4 can directly interact with Jun, Cebpa/b but not TEAD3 and Fosl1, suggesting that Oct4's binding motif dependency on somatic TF motifs is determined by both direct and indirect PPIs.

With the direct PPIs between Oct4 and Cebpa/b proved by *in vitro* pull-down assay, it doesn't rule out the possibility of having both Oct4 and Cebps synergistically bind to their motifs on the opened chromatin. To test it, we designed PAQMAN[8] assay to test the binding affinity ($Kd_{apparent}$) in the presence and absence of Oct4's expression as well as Cebps' motif. Having this result will help us to answer if Oct4-Cebps PPI or Oct4's co-binding to Cebps' motif increases Cebps' binding affinity to chromatin.

Moving forward, previously, Chronis et.al. [4]suggested that Oct4 and Klf4 bind to MEF specific enhancers (MEs) and remove the active histone mark H3K27ac by recruiting HDAC1 to MEs. We overexpressed an Oct4 nucleosome binding mutant Oct4^{del79} [9] in MEFs and found Oct4^{del79} cannot open new sites but maintain majority of its binding sites on MEs. By checking H3K27ac binding signal on these ME sites, intriguingly, the decreasing signal caused by wild-type Oct4 on MEs is not reproduced by Oct4^{del79}. This is suggesting that without opening new sites Oct4^{del79} cannot remove H3K27ac from MEs and we further hypothesized that it is because ME-bound somatic TFs cannot be redistributed to Oct4 opened new sites and hence H3K27ac is maintained. This hypothesis still requires functional experiment to test it.

Lastly, to test how essential Oct4's direct PPI for redistributing somatic TFs in the reprogramming process for silencing MEs, we designed an Oct4 mutant screen experiment to find a Oct4 mutant lacking PPI ability to somatic TFs including Cebps and AP-1 TFs, and we will overexpress the Oct4$_{\Delta PPI}$ alone or along with wild-type SKM, to test H3K27ac signal on MEs and reprogramming efficiency difference.

CMMB (Carboxylate Modified Magnetic Bead) -based isopropanol gradient peptide fractionation (CIF) enables rapid and robust off-line peptide mixture fractionation in bottom-up proteomics

In Chapter 3, we describe a CMMB-based peptide fractionation method that offers several features that provide significant utility in proteomics applications. First, desalting and fractionation are done in the same tube which minimizes sample loss and facilitates potential automation. Second, the binding capacity of CMMB is high enabling small beads volumes and hence small elution volumes which also enhances sensitivity and limits losses. Third, CIF fractionation is orthogonal to SCX and high pH reversed phase allowing it to be easily integrated into multidimensional chromatographic schemes.

One major advantage to CIF is its orthogonality to the acidic online reversed phase separations which are standard in the majority of LC-MS/MS workflows. Considering how evenly peptides are distributed across the LC gradient determines how efficiently data-dependent MS/MS acquisition occurs, the orthogonality of offline separations becomes a determining factor for the effectiveness of the analysis. Based on data in Fig. 3A and Fig. 3C, we demonstrate that CIF displays excellent orthogonality to reversed phase chromatography in LC-MS/MS applications and is likely the reason for improved peptide identification and quantitation in CIF fractions. Additionally, we take advantage of this

orthogonality by demonstrating that CIF and high pH RP can be used in 2-D fractionation experiments to sequentially fractionate peptides offline before LC-MS/MS analysis to further increase peptide coverage.

Another major advantage of CIF is its scalability in terms of peptide input. Standard spin column-based high pH reversed phase kits typically used for offline fractionation separate input peptides ranging in amount from 10-100ug. However, in the two applications we reported here, CIF is compatible with the fractionation of low input affinity purified samples. Specifically, for low input samples that are particularly sensitive to material loss during processing and which limits fractionation options, we demonstrate CIF retains the ability to efficiency fractionate samples and can improve data quality at those peptide concentration regimes. Based on our experience, fractionation of affinity purified samples at the level of either cell compartment or peptide significantly improves acquisition of reproducible and biologically meaningful data (data not shown here).

Lastly, we built a linear model that predicts the elution properties of a peptide based on its sequence. This model not only sheds light on the mechanism of underlying CMMB peptide-protein binding but also provides a tool for enriching peptides with particular properties. Since the current model assigns very high weights to the number of charged amino acid residuals and peptides that are charged under pH 7, we speculate that CIF might have utility for fractionation of phospho-peptides.

Neuronal activity regulates the nuclear proteome to promote activity-dependent transcription

In Chapter 4, by using optimized APEX2-based proximity labeling technology[10], we identified 91 genes that are putative activity-dependent targets of PDCD4, including

genes encoding proteins that are important for synaptic function. The activity-dependent downregulation of PDCD4 in neurons is reminiscent of the concept of "memory suppressor genes" [11] genes that act as inhibitory constraints on activity-dependent neuronal plasticity. By analogy to PDCD4 function during cancer metastases, decreases in PDCD4 in neurons would function to enable experience-dependent neuronal growth and remodeling. Dysregulated PDCD4 concentrations have also been reported to underlie a variety of metabolic disorders, including polycystic ovary syndrome, obesity, diabetes, and atherosclerosis, highlighting the critical role PDCD4 plays in regulating gene expression in multiple cell types[12]. Our study provides the first transcriptomic profile of PDCD4 that is independent of PDCD4's role in translation. These results provides insight into the transcriptional targets of PDCD4, which is of relevance not only to neuroscience but also to the study of PDCD4 in cancer. Taken together, our findings illustrate the utility of proximity ligation assays in identifying activity-dependent changes in the proteome of subcellular neuronal compartments and point to the array of cell biological mechanisms by which activity can regulate the neuronal proteome. They also focus attention on the tumor suppressor protein PDCD4 as a critical regulator of activity-dependent gene expression in neurons, highlighting a role for PDCD4 in regulating the transcription of genes involved in synapse formation, remodeling, and transmission. This new role is in addition to PDCD4's well-characterized role as a translational inhibitor [13] and future investigation of the mechanisms by which PDCD4 regulates transcription of these genes will provide further insight into the understudied role of PDCD4 as a transcriptional regulator. Such studies also promise to deepen our understanding of the

specific cell and molecular biological mechanisms by which experience alters gene expression in neurons to enable the formation and function of neural circuits.

High-field asymmetric waveform ion mobility spectrometry interface enhances parallel reaction monitoring on an Orbitrap mass spectrometer

In Chapter 5, we systematically profiled the intensity distribution of peptides during FAIMS separation across different CV settings and found that the majority of peptides is efficiently transmitted through relatively narrow CV windows. Using machine learning approaches, we showed that the charge state together with 3D descriptors such as MS-WHIM scores are the top determinants for dictating peptide iso-form transmission across CV space and that this information can be used to accurately predict optimal peptide CV from a peptide's sequence. Furthermore, we find that the proper selection of CV values is essential for PRM assay data quality with optimized CV setting leading to significant improvements in detection and quantitation.

Reference

1. Takahashi K, Yamanaka S: Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006, 126:663–676.

2. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007, 131:861–872.

3. Takahashi K, Yamanaka S: A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Bio* 2016, 17:183–193.

4. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K: Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* 2017, 168:442-459 e20.

5. Papp B, Plath K: Epigenetics of reprogramming to induced pluripotency. *Cell* 2013, 152:1324–1343.

6. Deng W, Jacobson EC, Collier AJ, Plath K: ScienceDirect The transcription factor code in iPSC reprogramming. *Curr Opin Genet Dev* 2021, 70:89–96.

7. Rafiee MR, Girardot C, Sigismondo G, Krijgsveld J: Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Mol Cell* 2016, 64:624–635.

8. Makowski MM, Gräwe C, Foster BM, Nguyen NV, Bartke T, Vermeulen M: Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. *Nat Commun* 2018, 9:1653.

9. Roberts GA, Ozkan B, Gachulincová I, O'Dwyer MR, Hall-Ponsele E, Saxena M, Robinson PJ, Soufi A: Dissecting OCT4 defines the role of nucleosome binding in pluripotency. *Nat Cell Biol* 2021, 23:834–845.

10. Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, Ting AY: Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* 2015, 12:51–54.

11. Abel T, Kandel E: Positive and negative regulatory mechanisms that mediate long-term memory storage. *Brain Res Brain Res Rev* 1998, 26:360–78.

12. Lu K, Chen Q, Li M, He L, Riaz F, Zhang T, Li D: Programmed cell death factor 4 (PDCD4), a novel therapy target for metabolic diseases besides cancer. *Free Radical Bio Med* 2020, 159:150–163.

13. Wang Q, Yang H-S: The role of Pdcd4 in tumour suppression and protein translation: Pdcd4 in tumour suppression and protein translation. *Biol Cell* 2018, 110:169–177.