

Proteomic analysis reveals diverse proline hydroxylation-mediated oxygen-sensing cellular pathways in cancer cells

Tong Zhou^{1,*}, Luke Erber^{1,*}, Bing Liu², Yankun Gao¹, Hai-Bin Ruan², Yue Chen¹

¹Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota at Twin Cities, Minneapolis, MN 55455, USA

²Department of Integrative Biology and Physiology, University of Minnesota Medical School, Minneapolis, MN 55455, USA

*These authors have contributed equally to this work

Correspondence to: Yue Chen email: YueChen@umn.edu

Keywords: proline hydroxylation, oxygen-sensing, posttranslational modification, hypoxia, LCMS

Received: April 27, 2016

Accepted: September 25, 2016

Published: October 13, 2016

ABSTRACT

Proline hydroxylation is a critical cellular mechanism regulating oxygen-response pathways in tumor initiation and progression. Yet, its substrate diversity and functions remain largely unknown. Here, we report a system-wide analysis to characterize proline hydroxylation substrates in cancer cells using an immunoaffinity-purification assisted proteomics strategy. We identified 562 sites from 272 proteins in HeLa cells. Bioinformatic analysis revealed that proline hydroxylation substrates are significantly enriched with mRNA processing and stress-response cellular pathways with canonical and diverse flanking sequence motifs. Structural analysis indicates a significant enrichment of proline hydroxylation participating in the secondary structure of substrate proteins. Our study identified and validated Brd4, a key transcription factor, as a novel proline hydroxylation substrate. Functional analysis showed that the inhibition of proline hydroxylation pathway significantly reduced the proline hydroxylation abundance on Brd4 and affected Brd4-mediated transcriptional activity as well as cell proliferation in AML leukemia cells. Taken together, our study identified a broad regulatory role of proline hydroxylation in cellular oxygen-sensing pathways and revealed potentially new targets that dynamically respond to hypoxia microenvironment in tumor cells.

INTRODUCTION

Proline hydroxylation (Hyp) is a key oxygen-sensing posttranslational modification that is dynamically modulated during tumor cell proliferation and angiogenesis [1-6]. The modification is mediated by the evolutionarily conserved prolyl hydroxylases that add a single oxygen to the carbon on the proline pyrrolidine ring and form an (*S*)-configured hydroxyl group at the gamma position. O₂-dependent nature of hydroxylation reaction renders prolyl hydroxylases as key oxygen sensors that are dynamically responsive to the hypoxia microenvironment of the cells or organisms. The modification subtly affects the protein structures, activities and the properties of protein-protein interactions in the cell and plays a pivotal role in cancer development and disease progression [7-11].

The most well-known proline hydroxylation substrate is collagen, an essential structural molecule to

support cell matrix and skeleton structures [12-15]. The hydroxylation on the conserved PG motif is required for the proper folding of the collagen polypeptide chain and the release of mature collagen proteins from endoplasmic reticulum. Proline hydroxylation is also known to regulate hypoxia-inducible-factor-1 alpha (HIF-1 α) protein, an important transcription factor [11]. Under normoxia condition, hydroxylated HIF-1 α protein is specifically recognized by von Hippel-Lindau (pVHL) E3 ubiquitin ligase, which leads to rapid HIF-1 α protein degradation [2]. Hypoxia microenvironment, on the other hand, reduces the hydroxylation abundance on HIF-1 α and prevents its rapid degradation. Stabilized HIF-1 α leads to transcriptional activation of nearly 100 proteins in the key hypoxia-response cellular pathways, which are critical for cancer cell survival under low O₂ environment [11]. Recent biochemical studies with high resolution mass spectrometry have identified and validated a number of

other proline hydroxylation substrates including FOXO3a, Argonaute 2, ATF-4 and PKM [16-22], which revealed diverse HIF-independent oxygen-sensing activities in RNA interference and transcriptional regulation. These evidence suggest that the hypoxia-response mechanism mediated through proline hydroxylation regulates diverse cellular pathways and signaling processes in cells and may play a much wider role in regulating cellular physiology and protein functions.

Despite these advances, only limited number of proline hydroxylation substrates have been identified and validated in cancer cells, which hinder the effort to fully understand the cellular response to the hypoxia microenvironment. To address this challenge, we have developed an immunoaffinity-purification assisted approach to system-wide identify proline hydroxylation substrates in cancer cells. Our analysis revealed a broad range of Hyp substrate proteins and essential cellular pathways targeted by this important protein modification.

RESULTS

System-wide analysis of proline hydroxylation proteome

We first developed and validated a pan-antibody recognizing peptides containing trans-4-hydroxylated proline (Figure 1A–1B, Supplementary Figure S1). Using this antibody, we performed an initial proteomic analysis to systematically identify proline hydroxylation targets in HeLa cells (Figure 1C, Supplementary Figure S2). The cells were lysed following a standard protocol, and the proteins were digested by trypsin into short peptides. The peptides were subject to immunoaffinity purification using the custom-made antibody to enrich peptides bearing proline hydroxylation. Peptides with or without immunoaffinity enrichment were then fractionated using strong-cation exchange chromatography into 6 fractions and analyzed by nano-HPLC/MS/MS for deep proteomic analysis to identify hydroxyproline-containing peptide substrates. The LCMS data were processed by MaxQuant and Andromeda search engines for peptide and protein identifications. The database search against human UniProt sequences revealed over 1000 candidate proline hydroxylation sites. However, careful manual inspection of peptide-spectrum matches suggested that some of the identifications are ambiguous hits with incomplete peptide backbone fragmentation and questionable site localizations. In such cases, common chemical oxidation artifacts such as Met and Trp oxidation could confuse the search engine and erroneously assign the modification to the nearby Pro, which has been considered as a significant technical challenge in the confident identification of proline hydroxylation sites [22]. To address this issue, we developed an in-house script to analyze the MS/MS fragmentation patterns of peptide-spectrum matches

and required that at least two distinct peptide fragment ions in each spectrum must be annotated to confidently assign the mass shift to Pro comparing to the nearest Met and Trp amino acids at both peptide N-terminal and C-terminal sides (see Materials and Methods). Although such filtering also removed some potentially true proline hydroxylated peptides, it significantly improved the confidence and quality of the dataset, which are essential for the bioinformatics analysis of the proline hydroxylation proteome. After careful screening, we identified 562 proline hydroxylation sites from 272 proteins in HeLa cells (Supplementary Table S1). Nearly 20% Hyp peptides contain at least four hydroxylation sites (Figure 1D). About one-third of the sites were identified only by immunoaffinity purification, suggesting that the application of this strategy contributed to the complementary coverage for Hyp peptide identification. Our results identified many previously well-known proline hydroxylation targets including collagen, actin and FKBP10 proteins [22-25], while at the same time, the majority of proline hydroxylation substrates have not been reported before (Figure 1E). The dataset provides a rich resource to reveal the diverse cellular pathways involved in proline hydroxylation-mediated oxygen sensing mechanisms in cancer cells.

Flanking sequence and secondary structural preference of proline hydroxylation sites

We first performed flanking sequence analysis using all +/- 15 amino acids in protein sequences surrounding the proline hydroxylation sites. The data showed that proline is one of the most frequently appearing amino acids surrounding hydroxylated proline, suggesting that many of the Hyp modifications occur in proline-rich sequence regions. Interestingly, we found 59% of Hyp peptides identified in our dataset bearing more than two Hyp modifications per peptide (Figure 1D) and such percentage is much higher than we typically observe in lysine acetylation proteomics analysis. We notice that amino acids with short side chains such as Gly and Ala appear very frequently in the immediate vicinity of hydroxylated proline. This observation agrees well with the knowledge that proline hydroxylation sites on collagen are often followed by Gly to form PG motifs. Our dataset revealed 65 sites sharing common PG motifs, suggesting that the enzyme, prolyl-4-hydroxylases that regulate collagen proline hydroxylation, may regulate the hydroxylation on diverse protein substrates. Motif enrichment analysis using Motif-x program [26] identified PP, PxP and PGxP as highly enriched motifs among proline hydroxylation sites (Supplementary Figure S3).

Proline hydroxylation is known to be essential for stabilizing collagen secondary structures. We want to ask whether the localization of identified proline hydroxylation sites have strong association with substrate

protein's secondary structures. Towards this goal, we extracted the manually-curated UniProt secondary structural information for each identified proline hydroxylation sites and for all the proline sites in each Hyp substrate proteins. We found 217 Hyp sites whose proteins have UniProt annotated structural information. Surprisingly, among these sites, 134 proline hydroxylation sites (62%) are located in one of the annotated secondary structures (STRAND, HELIX and TURN), while in the corresponding proteins, only 6% of total proline residues locate in one of these secondary structures. When considering all amino acids, only 18% were involved in the secondary structure among these proteins. The lower

percentage of Pro sites participated in the secondary protein structural is expected and agree well with previous knowledge on the characters of proline imino peptide backbone [27, 28]. The significant enrichment of Hyp sites locating in the secondary structure comparing to the distribution of Pro (HyperG test $P < 1.6 \times 10^{-108}$) and total amino acids (HyperG test $P < 7.5 \times 10^{-47}$), suggest that proline hydroxylation likely specifically targets those Pro sites involved in the protein secondary structure and may have important contribution to the protein folding and structural stability for diverse substrate proteins.

Carbamoyl phosphate synthase (CPS1) is a rate-limiting enzyme in the urea biosynthesis and ammonia

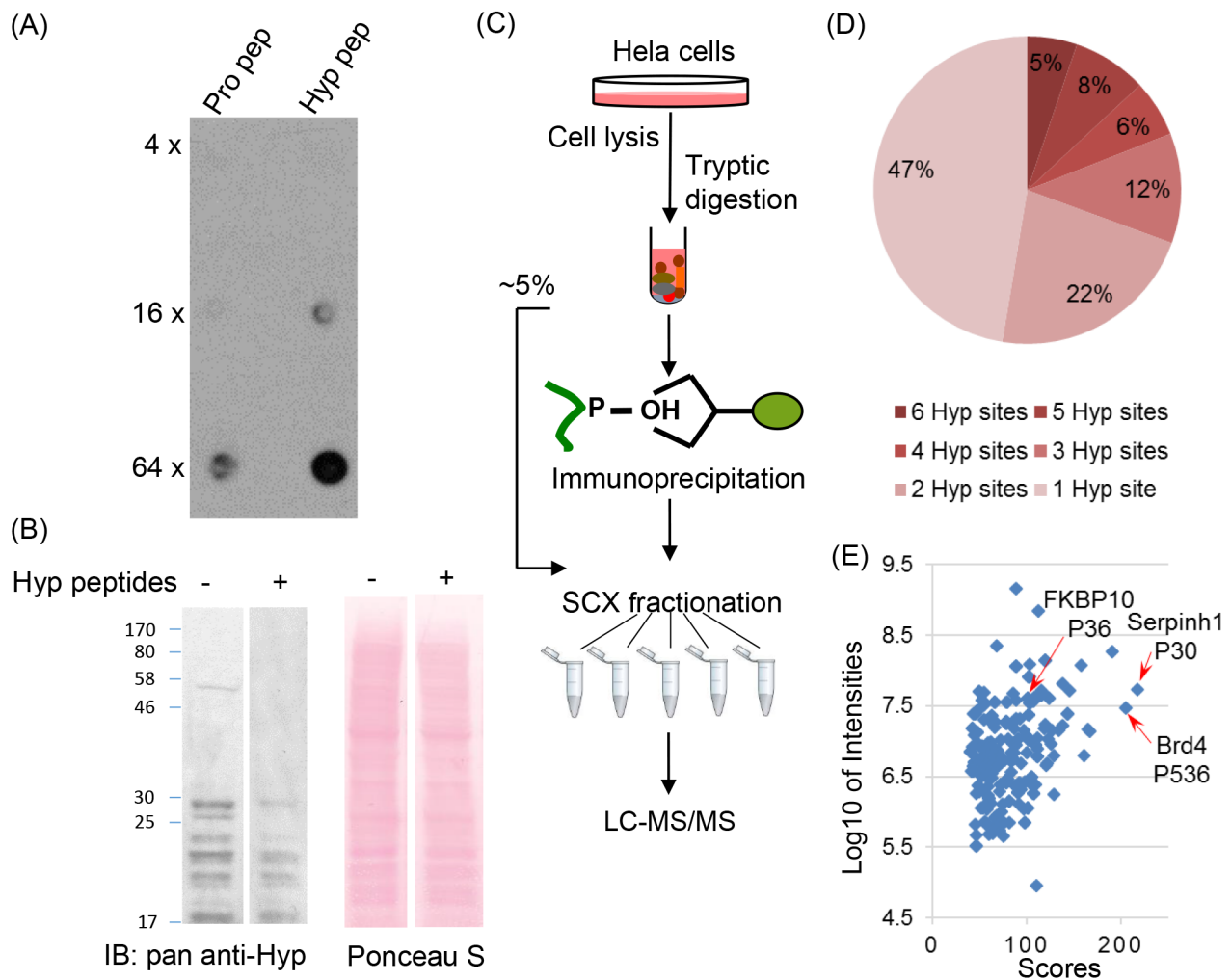


Figure 1: Systematic analysis of proline hydroxylation proteome with immunoaffinity purification and exhaustive LC-MS/MS analysis. **A.** Dot-blot assay for antibody specificity validation. Pro- and Hyp-containing peptide libraries were dotted on the nitrocellulose membrane with 0.125 μg , 0.5 μg and 2 μg on each dot from top to bottom, respectively. Pan-anti Hyp antibody was incubated with the membrane for two hours at room temperature prior to detection. **B.** Competition assay. Pan-anti Hyp antibody was incubated with Hyp-containing peptide library for 16 hours at 4 $^{\circ}\text{C}$. Western blotting of two lanes with equal amounts of HeLa whole cell lysate was performed with pan-anti Hyp antibody with or without prior incubations with Hyp peptides. **C.** Schematic diagram of the proteomics workflow. HeLa cells were lysed and proteins were digested by trypsin. Peptides were subject to immunoaffinity purification and exhaustive LC-MS/MS analysis with fractionation for identification. **D.** Distribution of the number of Hyp sites identified per peptide. Over 30% of the peptides were identified with more than two Hyp sites. **E.** The scatterplot of Hyp peptide identifications with Maxquant scores and intensities.

metabolic pathways. In our dataset, we identified CPS1 as a novel proline hydroxylation substrates with hydroxylation on Pro265 that locates within the glutamine amidotransferase domain of CPS1. Using structural analysis, we showed that hydroxylation on Pro265 may form a strong hydrogen bond between the Pro265 hydroxyl group and the alpha carboxyl group of Pro87 at the neighboring peptide backbone, potentially affecting the structural stability or enzymatic function of the protein (Figure 2D–2E).

Stoichiometric analysis of proline hydroxylation substrates

Site-specific PTM stoichiometry is a key metric to quantitatively understand the physiological importance of individual protein modification sites to the substrate protein's function and activity. To systematically estimate proline hydroxylation site stoichiometry, we extracted all sites identified by the direct LCMS analysis without enrichment and compared the precursor ion intensities of

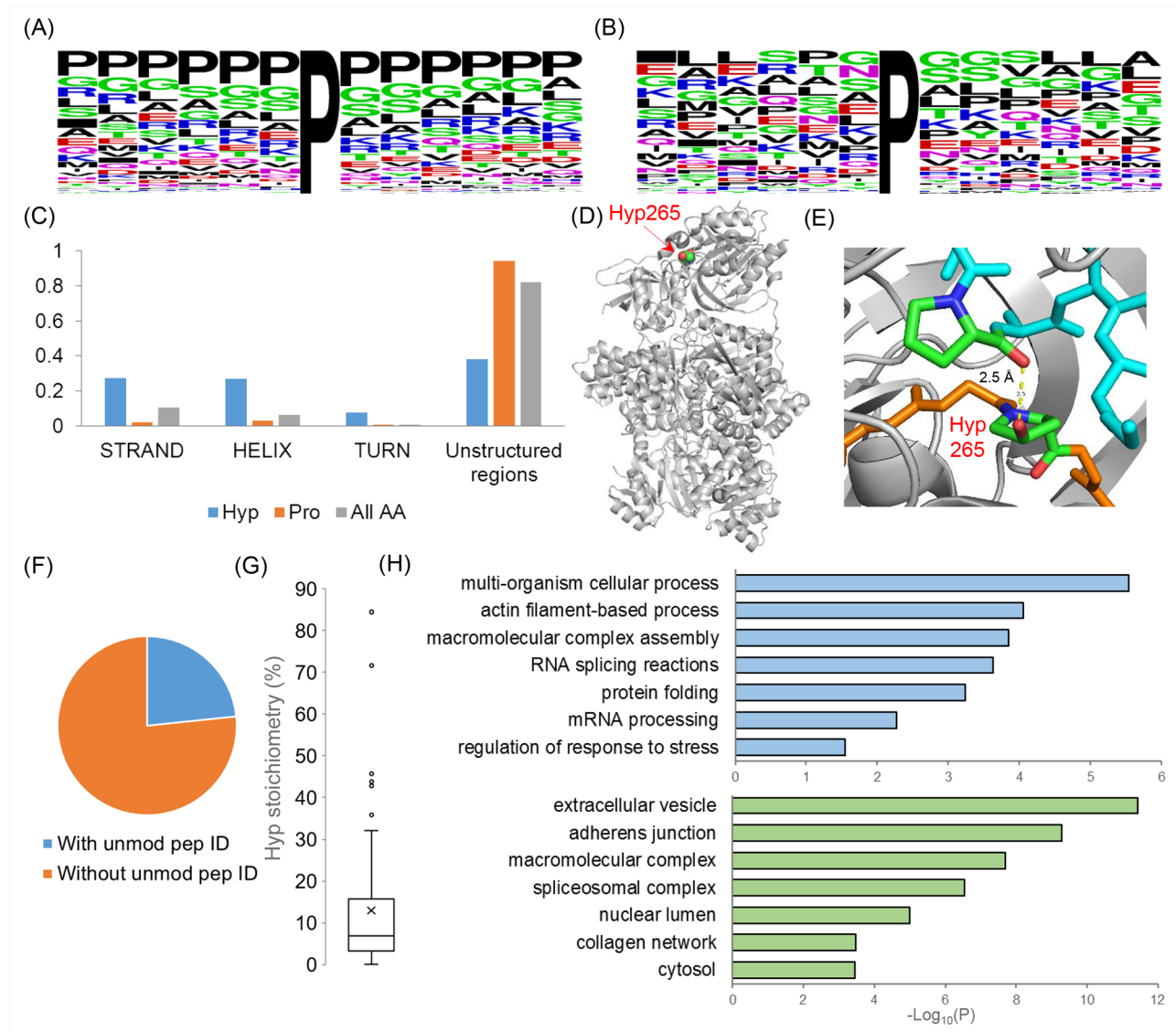


Figure 2: Analysis of the structural and functional characteristics of proline hydroxylation proteome in HeLa cells. **A.** Flanking sequence distributions of proline hydroxylation sites identified in this study. **B.** Flanking sequence distributions of proline hydroxylation sites from peptides identified with less than three Hyp sites. Flanking sequences were visualized with WebLogo [81]. **C.** The distribution of the secondary structures of proline hydroxylation sites (blue bars) comparing to proline (orange bars) and all amino acids (grey bars). **D.** and **E.** structural illustrations showing that the hydroxylation of Pro265 on CPS1 may form a hydrogen bond with the alpha-oxygen of the nearby Pro87 (PDB: 5DOU). **F.** The percentage distribution of Hyp peptides identified with and without the identification of the unmodified counterparts. **G.** The distribution of the calculated stoichiometries of Hyp sites with an average stoichiometry of 13%. **H.** Gene ontology annotation analysis of proline hydroxylation proteome with statistical enrichment in biological processes (upper panel) and cellular compartments (lower panel) (BH corrected $P < 0.05$).

both modified and unmodified peptides. Interestingly, over three-quarters of these proline hydroxylated peptides were identified only in modified form without the identification of the corresponding unmodified counterpart (Figure 2F), which include all the poly-proline hydroxylated peptides (the peptides with more than two sites per peptide). Site-specific stoichiometries were successfully calculated for seventy-two proline hydroxylation sites using the summed precursor ion intensities for the pair of modified and unmodified peptides for each site (Figure 2G and Supplementary Table S2). Our data showed an average stoichiometry of proline hydroxylation about 13%, with nearly 20% percent of sites having more than 20% stoichiometry. Comparing to recently published data, proline hydroxylation has much higher stoichiometry than lysine acetylation [29-31]. However, considering a large portion (>75%) of Hyp sites not quantifiable without the identification of unmodified counterpart including the poly-Hyp peptides from collagen (Figure 2F) and *a priori* knowledge that collagen hydroxylation is essential for its structure and function [8, 12, 32, 33], it is likely that the average of 13% proline hydroxylation stoichiometry based on quantifiable sites significantly under-estimated the overall proline hydroxylation abundance in cells.

Functional annotation analysis of proline hydroxylation proteome

To systematically study the physiological significance of proline hydroxylation substrates, we performed pathway enrichment analysis with Gene Ontology annotation analysis. Our results showed that proline hydroxylation substrates are highly enriched in the biological processes including multi-organism processes (adj $P=2.9\times 10^{-6}$), macromolecular assembly (adj $P=1.4\times 10^{-4}$), RNA splicing (adj $P=2.4\times 10^{-4}$) and regulation of response to stress (adj $P=2.8\times 10^{-2}$) (Figure 2H upper panel). In cellular compartment enrichment analysis, we found that proline hydroxylation substrates are widespread across different compartments but have the most apparent enrichment in extracellular vesicle (adj $P=3.6\times 10^{-12}$), macromolecular complex (adj $P=2.0\times 10^{-8}$), nuclear lumen (adj $P=1.0\times 10^{-5}$) and cytosol (adj $P=3.6\times 10^{-4}$) (Figure 2H lower panel). Gene classification by PANTHER system [34] showed cytosol proteins account for nearly half of the total Hyp proteome (~42%) while ~28% of Hyp proteome is from the cell organelles which mostly consist of cytoskeleton (~43%) and nucleus (~43%) proteins (Supplementary Figure S4). Few mitochondria proteins were identified to be proline hydroxylated, which agrees well with the current knowledge and cellular compartment annotations of prolyl hydroxylases mostly in the cytosol and nucleus [35-37], though it remains likely that mitochondria protein proline hydroxylation may be mediated by the radical-induced mechanism as previously suggested [38].

To investigate how proline hydroxylation substrates involve in the macromolecular assembly, we performed protein complex enrichment analysis with manually-curated CORUM database. We found that proline hydroxylation substrates were highly enriched in over 80 protein complexes, including spliceosome ($P=6.9\times 10^{-10}$), 60s APC containing complex ($P=8.0\times 10^{-5}$), EIF3 core complex ($P=1.1\times 10^{-3}$) and 20S proteasome ($P=1.4\times 10^{-2}$) (Supplementary Table S3). Analysis of proline hydroxylation proteome with protein-protein interaction database STRING showed that Hyp substrates formed vast and highly connected interaction networks (Figure 3A). Using subnetwork connectivity analysis, we identified several representative subnetworks including RNA spliceosome complex, cytoskeleton tubulin and collagen interaction networks (Figure 3B-3G).

Using online disease association database and annotation enrichment tool [39, 40], we performed disease enrichment analysis. Our data revealed the significant enrichment of proline hydroxylation substrates in stress and shock associated cellular pathways (adj $P=1.3\times 10^{-8}$) as well as diseases associated with viral infections such as HIV (adj $P=7.1\times 10^{-6}$). The hydroxylation substrate proteins identified in this pathway include heat shock proteins (HSP90AA1, HSP90AB1, HSPA5), NIAK family kinase (NIAK2), stress-induced phosphoprotein (STIP1) (Supplementary Table S4). The data suggested that proline hydroxylation may directly involve in regulating key protein activities in the stress-response cellular pathways, which are critical to cellular survival and function under stress conditions.

Validation of proline hydroxylation substrates with synthetic peptides

We performed validation experiments for 6 peptides using fragmentation of synthetic peptides bearing the same peptide sequences and modifications. These identifications included Brd4 (UniProt: O60885) Hyp536 (Figure 4 and Supplementary Figure S5), histone H2B 2-E (UniProt: Q99879) Hyp51 (Supplementary Figure S6A), proteasome subunit alpha-5 (UniProt: P28066) Hyp222 (Supplementary Figure S6B) and alpha7 (UniProt: O14818) Hyp149 (Supplementary Figure S6C), Serpin H1 (UniProt: P50454) Hyp30 (Supplementary Figure S6D), and FKBP10 (UniProt: Q96AY3) Hyp36 (Supplementary Figure S6E). Comparison of the MS/MS spectra between the *in vivo* peptides and synthetic peptides showed excellent matches, suggesting the high confidence of the proteomic identification of Hyp targets.

Functional analysis of Brd4 proline hydroxylation

Brd4 is a key transcription factor that has been recently identified to play an important role in leukemia

cell proliferation [41-43]. In our dataset, we identified Brd4 as a novel Hyp substrate with Pro536 hydroxylation at the junction between phosphoserine-rich NPS domain and the lysine-rich BID domain (Figure 4A–4C). Recent studies showed that the interaction between phosphorylated NPS domain and BID domain is essential for full Brd4 transcriptional activity and interaction with chromatin-binding proteins [44].

To further understand the physiological significance of the proline hydroxylation on Brd4, we performed functional analysis. First, we asked whether the proline hydroxylation abundance on Brd4 is enzymatically regulated. We treated the cells with DMOG, a well-established inhibitor of prolyl hydroxylases, and applied parallel reaction monitoring (PRM) in mass spectrometry to monitor the relative abundance of the Brd4 Hyp peptide (containing Hyp536) and its corresponding unmodified counterpart in HeLa cells (Figure 4D). The data showed that DMOG treatment, which inhibits prolyl hydroxylase activities and mimic hypoxia condition, led to significantly decreased prolyl hydroxylation abundance on Brd4 relative to its unmodified peptide (Hyp stoichiometry from 59% to 24%), suggesting that Brd4 proline hydroxylation is enzymatically regulated by prolyl hydroxylases.

Next, we want to ask whether the inhibition of proline hydroxylation impairs Brd4 transcriptional

activity. We selected MV-4,11 cells as a model, which is a leukemia cell line highly dependent on Brd4 transcriptional activity and very sensitive to Brd4 inhibitor JQ1[45]. We performed qRT-PCR to monitor the transcription of several of known Brd4 transcriptional targets, c-Myc, Ran and Rad21[46, 47]. Our data showed that inhibition of prolyl hydroxylase activities using DMOG strongly reduced the abundance and transcriptional expression of Brd4 targets, suggesting that Brd4 transcriptional activity is potentially regulated by proline hydroxylation (Figure 5A–5B). To probe the possible mechanism of DMOG-regulated Brd4 transcription activity, we performed chromatin immunoprecipitation coupled with quantitative PCR (ChIP-qPCR) analysis and evaluated the dynamics of Brd4 chromatin binding under DMOG treatment in MV4;11 cells. Our data showed that DMOG treatment significantly reduced Brd4 binding to the c-Myc promoter in AML leukemia cells (Figure 5C). Since Brd4 transcriptional activity has been shown to be critical for AML cell proliferation and survival, we next asked whether the altered Brd4 transcription activity by prolyl hydroxylase inhibitors will affect leukemia cell growth. Towards this goal, we performed cell proliferation assay under the control and DMOG treatment condition. Our data showed that the growth of MV4;11 cells was significantly inhibited upon DMOG treatment, while the growth of the model cell

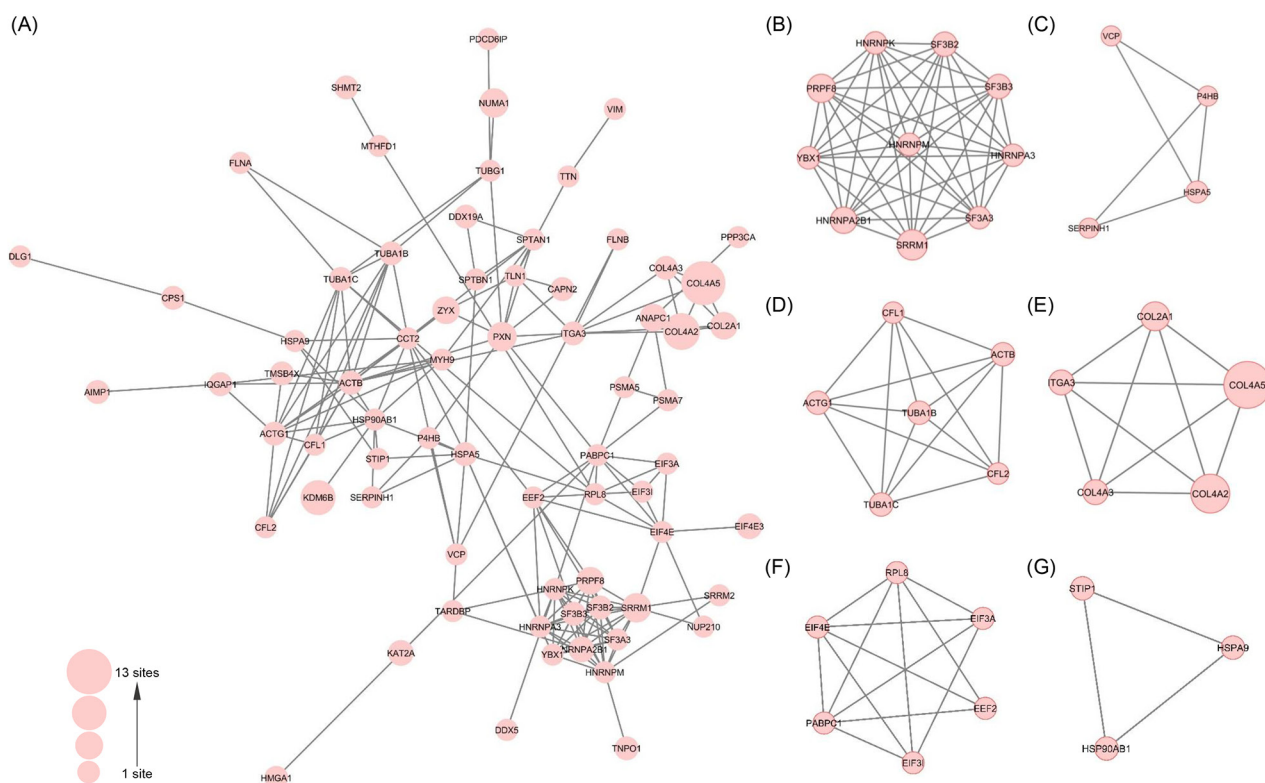


Figure 3: Protein interaction network analysis of proline hydroxylation proteome in HeLa cells. A. The interaction network of proline hydroxylation proteins identified in this study and B-G. an example list of the sub-networks of proline hydroxylation proteins with high interconnectivity.

of prolyl hydroxylases by IOX2 also strongly reduced the expression of c-Myc in MV4;11 cells and inhibited leukemia cell growth (Figure 5D and 5F). These data demonstrated that the enzymatically catalyzed trans-4-hydroxyproline modification is essential for MV4;11 cell proliferation. To further demonstrate the site-specific role of proline hydroxylation on leukemia cell activity, we generated a Brd4 mutant plasmid with P536A through site-directed mutagenesis and identified an shRNA targeting 3'UTR to knock down the endogenous Brd4 (Figure 5G). Knockdown of Brd4 has been previously shown

to significantly inhibit AML cell proliferation [43]. We found that expression of mutant Brd4 with the concurrent knockdown of endogenous Brd4 led to the significantly reduced MV4;11 cell proliferation comparing to the expression of the wild-type Brd4, suggesting that site-specific proline hydroxylation is critical for MV4;11 cell growth and function (Figure 5H).

Since the treatment of prolyl hydroxylase inhibitor is known to stabilize hypoxia-inducible factor1 alpha and previous studies showed that overexpression of HIF1 α may lead to the growth arrest of leukemia cells [51], we

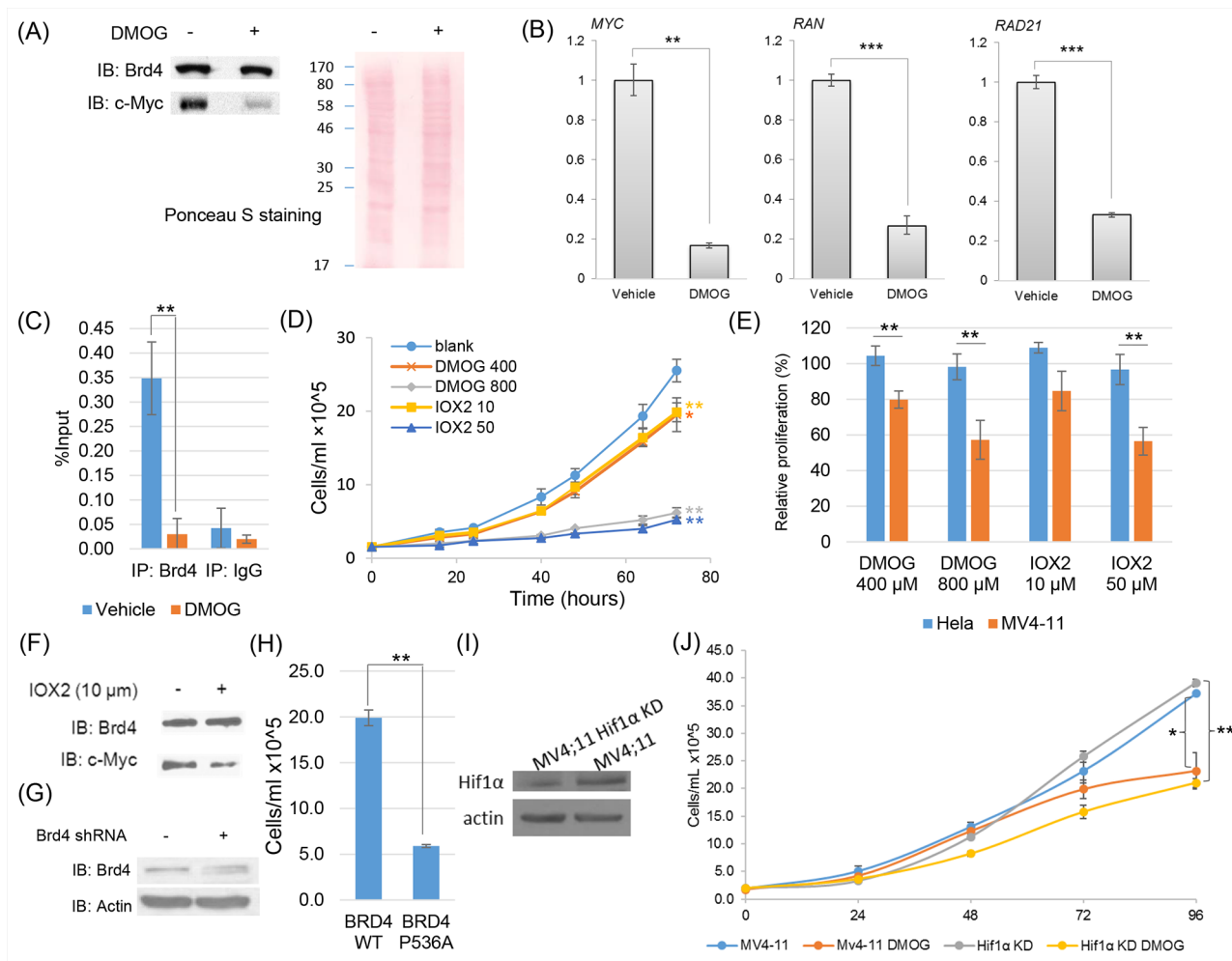


Figure 5: Functional characterization of proline hydroxylation pathway on Brd4 transcriptional activities and cell proliferation in MV4;11 cells. **A.** Western blotting analysis showing that the DMOG treatment of MV4;11 cells (800 μ M, 24 hrs) significantly decreased the protein abundance of c-Myc. **B.** Quantitative real-time PCR analysis showing that the DMOG treatment of MV4;11 cells significantly decreased the transcriptional expression of known Brd4-targeted genes. **C.** Chromatin immunoprecipitation and quantitative PCR (ChIP-qPCR) analysis of Brd4 binding to the c-Myc promoter in MV4;11 cells with or without DMOG treatment. **D.** Cell proliferation assay of MV4;11 cells under control, DMOG and IOX2 treatment. Statistical analysis was performed between treated and control cells at 72 hrs. **E.** The relative proliferation of HeLa and MV4;11 cells under DMOG (400 μ M, 800 μ M) and IOX2 (10 μ M, 50 μ M) treatments for 24 hrs comparing to the vehicle-treated control. **F.** Western blotting analysis showing that IOX2 treatment (10 μ M, 16 hours) significantly decreased the protein abundance of c-Myc. **G.** Cell proliferation assay of MV4;11 cells with shRNA knockdown of endogenous Brd4 and expression of WT Brd4 and Brd4 with P536A mutation. **H.** Western blotting analysis showing that HIF1 α was significantly knocked down in a stable MV4;11 cell line. **I.** Cell proliferation assay demonstrating that the growth of MV4;11 cells with and without HIF1 α knocked down were both strongly inhibited by DMOG treatment (400 μ M). All statistical analysis were performed with three biological replicates (Student's t-test, * $p < 0.05$, ** $p < 0.01$).

wanted to ask whether the growth inhibitory phenotype induced by prolyl hydroxylase inhibitors is dependent on HIF1 α . Towards this goal, we generated a stable MV-4;11 cell line with HIF-1 α significant knockdown using lentivirus-mediated shRNA transfection (Figure 5I). With this cell line, we repeated the cell proliferation assay and our data showed that the growth inhibitory effect of leukemia cells by prolyl hydroxylase inhibitors is HIF-1 α independent (Figure 5J).

DISCUSSION

Proline hydroxylation is an ancient posttranslational modification first discovered in 1902 [52]. However, the enzymatic regulatory pathways of this modification were understood only recently [3-6, 12]. As an irreversible modification, the proline hydroxylation is catalyzed by prolyl hydroxylases which require oxygen as co-substrate and iron, alpha-ketoglutarate and ascorbate as co-factors. Such requirements intimately link the prevalence of proline hydroxylation with key cellular metabolic states including oxygen availability, iron concentration as well as energy homeostasis. In recent years, advances in high resolution mass spectrometry led to the identification of an increasing number of proline hydroxylation targets [16-22]. These recently identified substrates involved in the transcriptional activation or maintenance of diverse hypoxia-response cellular pathways that are critical for tumor initiation and proliferation, including angiogenesis, glycolysis, DNA damage response and inflammation. Despite such advances, the functional and substrate diversity of proline hydroxylation proteome in cancer cells remain largely unknown and the precise identification of proline hydroxylation sites still suffer from ambiguous site-localizations in the large-scale proteomics data analysis[22]. To address these challenges and fill in this important knowledge gap, we performed a systematic study to expand the proline hydroxylation proteome and applied stringent criteria for the identification and validation of candidate Hyp targets. We integrated immunoaffinity enrichment with exhaustive HPLC-mass spectrometry analysis and identified 562 proline hydroxylation sites on 272 protein substrates in HeLa cells. Other than collagen hydroxylation sites, most of the sites identified in our study have not been reported before, giving us a unique opportunity to study the functional diversity of proline hydroxylation proteome and its critical roles in cancer cell activities and survival.

Proline is a nonpolar, aliphatic amino acid with a unique secondary amine and a pyrrolidine side chain. The structural specifications give proline an exceptional conformational rigidity which can play characteristic roles in the secondary structure of a protein [53]. Proline is frequently found in turns and non-repetitive secondary protein structures [27]. It is a structural disruptor in alpha helix and beta sheet and usually exists at the ends

of alpha helices and the edge of beta-sheets [28, 54]. Our analysis of annotated PDB structures showed that only 6% Pro tend to locate within the regular secondary structures, which is in accordance with previous studies, while surprisingly, 62% of proline hydroxylation with structural annotations identified in this study occur in the regular secondary structures, strongly suggesting that hydroxylation preferentially targets proline sites involved in the secondary structure and potentially promote the stabilization of the substrates' protein structure. Mechanistically, hydroxylation of proline may promote the hydrophilic interactions and induce stereo electronic effect, which contribute to the stable helix in collagen. Studies have shown that proline hydroxylation may increase the residue's polarity and serve as a substitution in forming H-bond to compensate for the lack of hydrogen in the backbone nitrogen [33, 55]. In addition, proline hydroxylation induces characteristic stereoelectronic effect due to the electron-withdrawing ability of OH. The *gauche* effect and n- π^* interaction can strengthen X-Hyp dihedral angles and stabilize helical structure [56]. Proline also plays an important role in regulating protein-protein interactions [57, 58]. Hydroxylation of the proline's hydrophobic pyrrolidine ring on the protein surface may fine-tune the surface polarity and potentially alter the properties of protein-protein interactions [59-61].

Analysis of flanking sequence of proline hydroxylation sites showed strong preference of the Hyp modification sites with Gly, Ala and poly-proline as neighboring amino acids. Indeed, nearly one-third of the tryptic peptides with proline hydroxylation (82/262) contain at least three Hyp modification sites. Such percentage is much higher than the typical analysis of phosphorylation and lysine acetylation proteome (data not shown). The preference of small non-polar amino acids like Gly as neighboring residues also agree well with the knowledge on proline hydroxylation sites on collagen, which is considered as a key motif for the prolyl-4-hydroxylase that mediates collagen proline hydroxylation. The prevalence of the peptides containing the PG motif suggests that prolyl-4-hydroxylase likely have more diverse targets than previously thought.

Cellular stress response pathways are strongly enriched with proline hydroxylation substrates, including key structural molecules and chaperone proteins. Such finding suggests that this oxygen-sensing modification could serve as a direct regulatory mechanism that mediates the activities of these pathways under hypoxia microenvironment. In addition, as trans-4-hydroxyproline plays a critical role in stabilizing trans-peptide bond and secondary protein structure, the lack of oxygen could strongly impact the folding and proper maintenance of diverse protein structures and the accumulation of incorrectly folded proteins may induce cellular stress responses. Indeed, unfolded protein response (UPR) in ER is one of the major hypoxia-induced cellular response

pathways [57-60]. The hyperactivated UPR pathways are important for cancer cells to adapt to the hypoxia states and have become novel targets for cancer treatment. The identification of a wide range of structural and chaperone proteins with proline hydroxylation in stress-response pathways demonstrate a potentially critical role of proline hydroxylation in mediating protein folding and ER stress in cancer cells.

Messenger RNA processing and RNA splicing processes are some of the most enriched cellular pathways with proline hydroxylation substrates in HeLa cells, which agrees with previous studies that have linked alternative RNA splicing and microRNA processing with hypoxia and oxygen sensing [20, 21, 66, 67]. Our data suggests a potentially critical role of proline hydroxylation pathways on the RNA processing in cancer cells. The identification and validation of proline hydroxylation site (Hyp50) on histone H2B further expands the inventory of histone marks. The location of the Hyp50 site at the junction between acetyllysine-rich N-terminal tail and the C-terminal histone globular domain indicates a potentially direct oxygen-dependent regulation of chromatin structure and epigenetic regulation. Despite the identification of novel Hyp targets, there are plenty of room to improve the proteomics analysis workflow through strategies with higher sensitivities and specificities. Comparing to the site-specific antibodies, we found that the custom-developed polyclonal pan-specific Hyp antibodies have varied affinities and specificities against sites with different flanking sequences and therefore limit its application in the Western blotting or the affinity enrichment of low abundance Hyp targets such as HIF1 α . In addition, proline hydroxylation is known to promote the degradation of many substrates such as HIF1 α , HIF2 α , Foxo3a and ATF-4 [17-19], which were not identified in our initial proteomics screen. Inhibition of protein degradation pathways prior to the enrichment analysis may allow the identification of these important substrates with rapid turnover. We also acknowledge that confident identification and localization of proline hydroxylation sites through LCMS requires extensive and careful analysis of HPLC tandem mass spectrometry data. Issues like amino acid polymorphism, inefficient backbone fragmentation and co-eluting peptide interference may still lead to false-positive identifications. High-throughput sequencing and complex algorithms for LCMS analysis will be necessary to address these challenges.

Recently, bromodomain-containing proteins have emerged as novel targets for cancer treatments [41-43, 68]. Brd4 is one of the most well-studied bromodomain-containing proteins playing critical roles in c-Myc hyperactivation and leukemia cell proliferation. Our study identified Brd4 as a novel proline hydroxylation substrate with nearly 60% stoichiometry under normoxia

states and its hydroxylation level is enzymatically regulated. Inhibition of prolyl hydroxylase activities significantly diminished the expression of Brd4-targeted genes as well as the proliferation of leukemia cells. Although the detail mechanisms of hydroxylase-regulated Brd4 transcriptional activities require further investigation, it demonstrates that the oncogenic proline-hydroxylation-dependent pathways may potentially serve as new targets for the development of therapeutic strategy in leukemia.

MATERIALS AND METHODS

Cell culture

HeLa cell was cultured in Dulbecco's Modified Eagle Medium (DMEM, Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Sigma), 100 IU penicillin and 100 μ g/ml streptomycin (Fisher Scientific) at 37 °C in a 5% CO₂ incubator. MV 4;11 cell line (a kind gift from Jianjun Chen at the University of Chicago) was maintained in RPMI 1640 media (Fisher scientific) (with 10% FBS and 100 IU penicillin, 100 μ g/ml streptomycin) at 37 °C in a 5% CO₂ incubator.

Generation and validation of pan-hydroxyproline antibody

Randomized hydroxyproline-containing peptide library "CXXXXXX(Hyp)XXXXXX" ("X" refers to any amino acid except for Cys, and "Hyp" indicates trans-4-hydroxyproline) was conjugated to Keyhole Limpet Hemocyanin protein (KLH) and subject to the immunization of rabbit following a standard protocol (SDIX Inc). The antibody was purified with the agarose beads conjugated with hydroxyproline-containing peptide library. For immunoaffinity purification, pan-hydroxyproline antibody was covalently linked to the protein A agarose beads for the immunoaffinity purification using amine-reactive crosslinker as previously described [69].

Cell lysis and proteolytic digestion for proteomics analysis

HeLa cells were lysed on ice in the cell lysis buffer (100 mM NaCl, 20 mM Tris-HCl, 10 mM EDTA, 0.5% (v/v) Nonidet P-40 (NP-40) and 1 mM phenylmethylsulfonyl fluoride (pH 8.0)) followed by sonication. Protein concentration was determined by the Bradford assay (Biorad). For proteolytic digestion, the proteins in the cell lysate were first reduced and alkylated with a final concentration of 5 mM tris(2-carboxyethyl) phosphine (TCEP) and 5 mM iodoacetamide to block cysteine residues. Proteins were then precipitated with 10% trichloroacetic acid (v/v, final concentration), washed three times with cold acetone. The pellet was resuspended in

100 mM NH_4HCO_3 . The protein solution was adjusted to pH 8.0 and digested by trypsin (Promega) at an enzyme-to-substrate ratio of 1:50 (w/w) at 37 °C overnight. The digestion was repeated for the second time for 3 hours at 37 °C with an enzyme-to-substrate ratio of 1:100. The digestion was finally quenched by TFA (final concentration 1%, v/v).

Immunoprecipitation of Hyp-containing peptides

Tryptic peptides were desalted by Sep-Pak C18 column (Waters, 100 mg) following the manufacturer's protocol. The eluted peptides were dried by Speed Vac (ThermoFisher) and then resuspended in 50 mM NH_4HCO_3 buffer. Antibody-bound protein A beads were added to the peptide solution and incubated for overnight at 4 °C with rotation. Then the beads were washed twice with the sample buffer and eluted with 0.1% TFA.

Peptide fractionation and LC-MS/MS analysis

The peptides were fractionated and desalted by Stage-tips packed with Empore membranes (3M) as previously described [70]. Briefly, the peptides were reconstituted in 5% ACN, 0.1% FA (formic acid) in water and loaded onto a Stage-tip packed with Empore Cation Exchange-SR membranes (3M). The Stage-tip was washed with 0.1% FA and sequentially eluted with the following six buffers that contain 0.1% FA (v/v), 20% ACN (v/v) and NH_4OAc with a respective concentration of 50, 75, 125, 200, 300, 500 mM. Peptides in each fraction were subsequently desalted with C18 Stage-tips and dried in Speed-Vac (ThermoFisher). For LC-MS/MS analysis, each sample was resolubilized in 2.5 μl of HPLC buffer A (0.1% formic acid in water, v/v) and loaded onto an in-house packed C18 column (15 cm \times 75 μm , ReproSil-Pur Basic C18, 2.5 μm , Dr. Maisch GmbH) through a Proxeon Easy nLC 1000 Nano-UPLC system connecting on-line to an Orbitrap Fusion mass spectrometer (ThermoFisher). Peptides were eluted with a gradient of 5-30% of HPLC buffer B (0.1% formic acid in acetonitrile, v/v) at a flow rate of 200 nl/min. Full mass spectra were acquired with a resolution of 120,000 at m/z 200. Dynamic exclusion was enabled to allow for one fragmentation in 60 seconds. Isolation window for MS/MS analysis was 1.2 m/z. Precursor ions were fragmented using high-energy collision dissociation (HCD) with 35% collision energy in ion trap.

MaxQuant 1.5.2.8 was used to identify peptides from spectra in the raw data. Hydroxylation on proline (+15.9949 Da), protein N-terminal acetylation and methionine oxidation were specified as variable modifications. Carbamidomethylation of cysteine was specified as a fixed modification. Trypsin was specified as the proteolytic enzyme with a maximum of 2 missing cleavages. *In silico* identification of co-eluting peptides from the same spectra was disabled. The precursor ion tolerance was 4.5 ppm and the fragment ion mass tolerance

was 0.5 Da. The database search was performed against UniProt human database (downloaded at 2014/04/14 with a total of 69078 sequences) with a cutoff threshold at 1% False Discovery Rate (FDR) at peptide, protein and site levels. Modified peptides were required to have a minimum Andromeda score of 40.

To ensure high confidence identification of hydroxyproline peptides and avoid false identification of peptides containing neighboring Met and Trp oxidation due to ambiguous localization of the modification sites, we applied an in-house developed software script to analyze the fragmentation match pattern of each hydroxyproline peptide identification. For confident localization, the software requires a minimum number of two non-redundant ions in MS/MS spectra that represent peptide fragments between Pro and the closest Met or Trp on both N- and C- terminal side of each peptide. The minimum of two non-redundant fragment ions was specified because the closest possible position of Pro with Met and Trp in the same peptide is that they are next to each other, in which case, only two ions (the b and y ions that fragment between them) can specifically determine whether the mass shift can be localized on Pro or the neighboring Met/Trp residues. The specification of the minimum of two non-redundant fragment ions for the site localization enabled the removal of potentially false positive identifications with very high stringency. All peptide identifications that passed through the fragmentation pattern analysis were further evaluated manually to ensure high confidence of the peptide-spectrum matches. LCMS data are publically accessible through ProteomeXchange Consortium and PRIDE partner repository (accession: PXD005018).

To estimate site-specific Hyp stoichiometry from the analysis of Hyp identifications without affinity enrichment, all the peptide forms containing the Hyp sites were analyzed based on Maxquant search results. The total intensities of modified and their corresponding unmodified peptide forms were summed up respectively to calculate site-specific Hyp stoichiometries.

Bioinformatic analysis

The pathway and functional annotation enrichment analysis for proline hydroxylation proteome was performed in GOstats software package [71] in R using hypergeometric test for Gene Ontology annotations [72]. The enrichment with a Benjamini-Hochberg [73] corrected $P < 0.05$ was considered to be statistically significant.

Protein-protein interactions were extracted from STRING database [74] with a cutoff interaction score above 700. Protein interaction network was visualized with Cytoscape [75] and highly interconnected protein clusters were identified by MCODE plugin [76]. Human protein complexes were extracted from CORUM database [77] and significantly enriched proline hydroxylation complexes were identified with hypergeometric test with $P > 0.05$.

Protein secondary structure information were manually curated from the PDB data deposited in the UniProt database [78]. Individual protein structures were visualized by Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) with PyTMs plugin [79].

Synthetic peptide validation

Synthetic peptides were acquired from GL Biochem Ltd (Shanghai, China). The peptides were reconstituted with HPLC buffer A and analyzed by Proxeon nanoLC-Orbitrap Fusion mass spectrometer as described above. Mass-to-charge ratios (m/z ratios) of each peptide based on the endogenous peptide identifications from HeLa cells were specified to acquire targeted HCD MS/MS spectra in ion trap. The data was analyzed by MaxQuant to confirm the peptide identification and evaluated manually for validation.

Targeted quantification of Brd4 proline hydroxylation

MV4;11 cells were plated at 2×10^4 – 2×10^5 /mL and treated for 24 h with 800 μ M DMOG (Cayman Chemical Company). The cells were lysed in urea lysis buffer (2 mM HEPES pH 8.0, 9.0 M Urea, 12.5 mM EDTA, and $1 \times$ HALT protease inhibitor cocktail (Fisher)) followed by sonication. Cysteine was reduced with 5 mM tris(2-carboxyethyl) phosphine (TCEP) (Sigma) and alkylated with 5 mM iodoacetamide (Sigma). The lysate was diluted to 1M urea with 50 mM NH_4HCO_3 and adjusted to pH 8.0. The protein solution was digested by trypsin (1:50) overnight, and by a second batch of trypsin (1:100) for additional two hours. The digestion was finally quenched with TFA (final concentration of 1%, v/v).

The tryptic peptides from both control and DMOG-treated cells were analyzed by LCMS as previously described. The precursor ions with the mass-to-charge ratios of 668.3604 and 663.0288, corresponding to hydroxylated and unmodified Brd4 peptides, respectively, were targeted for fragmentation. The mass spectrometry data were acquired with Orbitrap in high resolution for both MS and MS/MS analysis. Specific HPLC elution peaks for each peptide were determined with Parallel Reaction Monitoring (PRM) of multiple product ions [80]. The peak areas of the precursor ions were calculated to determine the abundance of modified and unmodified peptide isoforms. The stoichiometry was estimated with the following formula: $\text{Hyp}\% = (\text{Peak area of Hyp peptide}) / (\text{Peak area of Hyp peptide} + \text{Peak area of Pro peptide})$.

Cell proliferation assay

MV4;11 cells were cultured in RPMI 1640 media with L-glutamine (Fisher) supplemented with 10% FBS (Sigma), 100IU penicillin and 100 μ g/ml streptomycin (Fisher). Cells were split at an initial density of 2×10^4 –

2×10^5 /mL for proliferation analysis. The chemical treatment began after 24 hours with DMOG or IOX2 (ApexBio). The cell densities were determined every 24 hours in triplicate with a hemacytometer (Spencer). HeLa cell proliferation was measured by MTT assay (Sigma) following the manufacture's instruction.

Western blotting analysis

Proteins were resolved on SDS-PAGE gel and immunoblotted to PVDF membrane (BioRad). Brd4, c-Myc, actin, Hif1 α were detected by anti-Brd4 (GeneTex), anti-c-Myc (Cell Signaling), anti-actin (VWR), anti-Hif1 α (Millipore) antibodies following the manufacturer's instructions.

HeLa cells blank or treated with 200 μ M CoCl_2 for 16 hours were harvested using 2X SDS lysis buffer (125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 5% β -mercaptoethanol). For the competition assay with the pan-anti Hyp antibody, equal amounts of the HeLa whole cell lysate were loaded onto two lanes in gel and transferred onto a PVDF membrane. The membrane was blocked with 3% BSA in TBST (0.1% tween 20, in 1X tris-buffered saline, pH 7.6) for one hour at room temperature. Prior to the incubation with the membrane, the pan-anti Hyp antibody was diluted to a final concentration of 2 μ g/ml in 1.5% (w/v) BSA and incubated with 1 μ g/ml Hyp- or Pro-containing peptide libraries for 16 hours at 4 $^\circ\text{C}$. Then, the antibody solutions were incubated with the membrane at 4 $^\circ\text{C}$ overnight. The secondary antibody (1:20000) was dissolved in 3% BSA and incubated with the membrane at the room temperature for one hour.

Dot blot assay was performed as previously described [81]. Briefly, peptides in the amounts of 0.125 μ g, 0.5 μ g and 2 μ g were dotted onto the nitrocellulose membrane and dried in air. The membrane was blocked with 5% milk at the room temperature for 1 hour. The pan-anti Hyp antibody was diluted in 2% BSA for a final concentration of 2 μ g/ml and incubated with the membrane at the room temperature for 2 hours followed by the secondary antibody incubation. The dot blot was visualized with the VisiGlo Prime HRP kit (Amresco).

Quantitative real-time PCR

MV4;11 cells were treated with 800 μ M DMOG for 24 hours with an initial cell density of at 2×10^6 /mL. Control and DMOG-treated MV4;11 cells were subject to qRT-PCR analysis (ARQGenetics Inc., Bastrop, TX). To quantify the levels of c-Myc, Ran, and Rad21 transcripts in MV4;11, standard curves were generated using serial dilutions of each target transcript. The actin was used as the reference gene for normalization. All reactions were run in triplicate. The primers for qRT-PCR are listed in the Supplementary Information (Supplementary Table S5).

Generation of stable MV-4;11 cell line with HIF-1 α knockdown

The pLKO.1-puro vector encoding shRNA targeting Hif1 α was purchased from Sigma (TRCN 0000003810). Lentivirus for Hif1 α knockdown was packaged by the Minnesota Obesity Center (University of Minnesota, MN). To generate the stable knockdown cell line, MV4;11 cells were infected with the lentivirus for four hours in the presence of polybrene (Sigma). Stable transfectants were selected in an RPMI medium containing 10 μ g/ml puromycin (EMD Millipore), and about 10⁴ puromycin-resistant cells were pooled and expanded. Knockdown of Hif1 α in the stable MV4;11 cells were validated by Western blotting.

Transient transfection of leukemia cells with Brd4 WT, mutant plasmids and shRNA

The pFlag-CMV2-Brd4 (1-1362) plasmid was a gift from Eric Verdin (Addgene plasmid # 22304) [82]. Brd4 plasmid with P536A mutation was obtained using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) following the manufacturer's instructions. The primers are: 5'- GCCCCAGCAGAACAAACGCAAGA AAAAGGAGAAAGAC -3' (forward) and 5'- GTCTT TCTCCTTTTCTTGGCTTTGTTCTGCTGGGGC -3' (reverse). The Brd4 shRNA targeting 3'-UTR (sequence: CCGCCAAATGTCTACACAGTA) was provided by the Genomics Center (University of Minnesota). Transient transfection of MV4;11 cells was performed with the Neon Transfection System (ThermoFisher) following the manufacturer's instructions.

ChIP-qPCR

MV-4;11 cells were washed with PBS and crosslinked in 1% formaldehyde (Sigma) for 10 min, followed by quench with 2.5 M glycine for 5 minutes at room temperature. The cells were washed with PBS, resuspended in cell lysis buffer (10 mM NaCl, 10 mM Tris-HCl (pH 8), 0.4% NP-40, and protease inhibitors (Sigma)) and incubated on ice for 10 minutes. The lysate was centrifuged at 800 g for 10 min. The pellet was washed with PBS and resuspended with 1 mL buffer containing 10 mM EDTA, 0.5 mM EGTA, 10 mM HEPES, and 0.25% Triton X-100 for 10 minutes on ice. After centrifugation, the pellet was resuspended with SDS lysis buffer containing PMSF and incubated on ice for 30 minutes. The lysate was sonicated at 25% power for 10 cycles of 10 sec on and 30 sec off (Sonic Dismembrator Model 500, ThermoFisher). After centrifugation, the supernatant was diluted to less than 0.1% SDS with ChIP buffer (150mM NaCl, 50mM Tris-HCl (pH 7.5), 5mM EDTA, 0.5% NP-40, 1% Triton X-100, and protease inhibitors). Four μ g ChIP grade BRD4 antibody (Abcam) was incubated with the sample and rotated overnight at 4 °C. After clearing the chromatin by centrifugation,

the chromatin was incubated with 40 μ l (25% slurry) of Protein A/G agarose beads (Santa Cruz) by rotation at 4 °C for one hour. The beads were washed three times with cold ChIP buffer. To isolate the input DNA, the input lysate was incubated with 2.5-3x volume of ethanol for precipitation. This sample was washed with 70% ethanol and dried. To isolate input and immunoprecipitated DNA, 10% (wt/vol) Chelex 100 slurry (Bio-Rad) was added to dried pellets or washed beads [83, 84]. The sample was boiled for 10 minutes and then incubated with 20 μ g Proteinase K (Sigma) at 55 °C for 30min. The digested sample was boiled again for 10 minutes and the supernatant was transferred to a new tube. The input DNA, immunoprecipitated DNA and Protein A/G-only immunoprecipitated DNA were subjected to real time PCR (primer forward: TACTCACAGGACAAGGATGCGGTT, primer reverse: TGAATTA ACTACGCGCGCCTACCA [85] (Supplementary Table S5))

ACKNOWLEDGMENTS

We would like to thank Do-Hyung Kim, David Bernlohr, Timothy Griffin and the members of the Chen lab for helpful discussion and suggestions. We also thank Jeongsik Yong for the access of Neon Transfection System for electroporation. We greatly appreciate the Center for Mass Spectrometry and Proteomics for LCMS instrument access and support.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

GRANT SUPPORT

The study was supported by American Cancer Society Pilot Award (IRG-58-001-55) and the start-up fund from the University of Minnesota at Twin Cities.

REFERENCES

1. Semenza GL. Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu Rev Cell Dev Biol.* 1999; 15:551-578.
2. Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, Cockman ME, Wykoff CC, Pugh CW, Maher ER and Ratcliffe PJ. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature.* 1999; 399:271-275.
3. Bruick RK and McKnight SL. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science.* 2001; 294:1337-1340.
4. Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, Mukherji M, Metzen E, Wilson MI, Dhanda A, Tian YM, Masson N, Hamilton DL, Jaakkola P, Barstead R, Hodgkin J, et al. C. elegans EGL-9 and

- mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*. 2001; 107:43-54.
5. Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, Salic A, Asara JM, Lane WS and Kaelin WG, Jr. HIF α targeted for VHL-mediated destruction by proline hydroxylation: implications for O₂ sensing. *Science*. 2001; 292:464-468.
 6. Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A, Hebestreit HF, Mukherji M, Schofield CJ, Maxwell PH, Pugh CW and Ratcliffe PJ. Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science*. 2001; 292:468-472.
 7. Eberhardt ES, Panisik N, Jr. and Raines RT. Inductive Effects on the Energetics of Prolyl Peptide Bond Isomerization: Implications for Collagen Folding and Stability. *J Am Chem Soc*. 1996; 118:12261-12266.
 8. Gorres KL and Raines RT. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol*. 2010; 45:106-124.
 9. Kaelin WG. Proline hydroxylation and gene expression. *Annu Rev Biochem*. 2005; 74:115-128.
 10. Kaelin WG, Jr. The von Hippel-Lindau tumour suppressor protein: O₂ sensing and cancer. *Nat Rev Cancer*. 2008; 8:865-873.
 11. Semenza GL. Targeting HIF-1 for cancer therapy. *Nat Rev Cancer*. 2003; 3:721-732.
 12. Kivirikko KI and Prockop DJ. Enzymatic hydroxylation of proline and lysine in procollagen. *Proc Natl Acad Sci U S A*. 1967; 57:782-789.
 13. Kivirikko KI, Myllyla R and Pihlajaniemi T. Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. *FASEB J*. 1989; 3:1609-1617.
 14. Prockop DJ and Kivirikko KI. Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem*. 1995; 64:403-434.
 15. Holmgren SK, Taylor KM, Bretscher LE and Raines RT. Code for collagen's stability deciphered. *Nature*. 1998; 392:666-667.
 16. Luo W, Hu H, Chang R, Zhong J, Knabel M, O'Meally R, Cole RN, Pandey A and Semenza GL. Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. *Cell*. 2011; 145:732-744.
 17. Wong BW, Kuchnio A, Bruning U and Carmeliet P. Emerging novel functions of the oxygen-sensing prolyl hydroxylase domain enzymes. *Trends Biochem Sci*. 2013; 38:3-11.
 18. Zheng X, Zhai B, Koivunen P, Shin SJ, Lu G, Liu J, Geisen C, Chakraborty AA, Moslehi JJ, Smalley DM, Wei X, Chen X, Chen Z, Beres JM, Zhang J, Tsao JL, et al. Prolyl hydroxylation by EglN2 destabilizes FOXO3a by blocking its interaction with the USP9x deubiquitinase. *Genes Dev*. 2014; 28:1429-1444.
 19. Koditz J, Nesper J, Wottawa M, Stiehl DP, Camenisch G, Franke C, Myllyharju J, Wenger RH and Katschinski DM. Oxygen-dependent ATF-4 stability is mediated by the PHD3 oxygen sensor. *Blood*. 2007; 110:3610-3617.
 20. Qi HH, Ongusaha PP, Myllyharju J, Cheng D, Pakkanen O, Shi Y, Lee SW and Peng J. Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature*. 2008; 455:421-424.
 21. Stoehr A, Yang Y, Patel S, Evangelista AM, Aponte A, Wang G, Liu P, Boylston J, Kloner PH, Lin Y, Gucek M, Zhu J and Murphy E. Prolyl hydroxylation regulates protein degradation, synthesis, and splicing in human induced pluripotent stem cell-derived cardiomyocytes. *Cardiovasc Res*. 2016; 110:346-358.
 22. Arsenault PR, Heaton-Johnson KJ, Li LS, Song D, Ferreira VS, Patel N, Master SR and Lee FS. Identification of prolyl hydroxylation modifications in mammalian cell proteins. *Proteomics*. 2015; 15:1259-1267.
 23. Udenfriend S. Formation of hydroxyproline in collagen. *Science*. 1966; 152:1335-1340.
 24. Gould BS and Woessner JF. Biosynthesis of collagen; the influence of ascorbic acid on the proline, hydroxyproline, glycine, and collagen content of regenerating guinea pig skin. *J Biol Chem*. 1957; 226:289-300.
 25. Luo W, Lin B, Wang Y, Zhong J, O'Meally R, Cole RN, Pandey A, Levchenko A and Semenza GL. PHD3-mediated prolyl hydroxylation of nonmuscle actin impairs polymerization and cell motility. *Mol Biol Cell*. 2014; 25:2788-2796.
 26. Chou MF and Schwartz D. Biological sequence motif discovery using motif-x. *Curr Protoc Bioinformatics*. 2011; Chapter 13:Unit 13 15-24.
 27. Shamsir MS and Dalby AR. Beta-sheet containment by flanking prolines: molecular dynamic simulations of the inhibition of beta-sheet elongation by proline residues in human prion protein. *Biophys J*. 2007; 92:2080-2089.
 28. Kim MK and Kang YK. Positional preference of proline in alpha-helices. *Protein Sci*. 1999; 8:1492-1499.
 29. Baeza J, Dowell JA, Smallegan MJ, Fan J, Amador-Noguez D, Khan Z and Denu JM. Stoichiometry of site-specific lysine acetylation in an entire proteome. *J Biol Chem*. 2014; 289:21326-21338.
 30. Zhou T, Chung YH, Chen J and Chen Y. Site-Specific Identification of Lysine Acetylation Stoichiometries in Mammalian Cells. *J Proteome Res*. 2016; 15:1103-1113.
 31. Nakayasu ES, Wu S, Sydor MA, Shukla AK, Weitz KK, Moore RJ, Hixson KK, Kim JS, Petyuk VA, Monroe ME, Pasa-Tolic L, Qian WJ, Smith RD, Adkins JN and Ansong C. A method to determine lysine acetylation stoichiometries. *Int J Proteomics*. 2014; 2014:730725.
 32. Takeuchi T and Prockop DJ. Biosynthesis of abnormal collagens with amino acid analogues. I. Incorporation of L-azetidino-2-carboxylic acid and cis-4-fluoro-L-proline into procollagen and collagen. *Biochim Biophys Acta*. 1969; 175:142-155.

33. Mizuno K, Hayashi T and Bachinger HP. Hydroxylation-induced stabilization of the collagen triple helix. Further characterization of peptides with 4(R)-hydroxyproline in the Xaa position. *J Biol Chem.* 2003; 278:32373-32379.
34. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, Kitano H and Thomas PD. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 2005; 33:D284-288.
35. Place TL and Domann FE. Prolyl-hydroxylase 3: Evolving Roles for an Ancient Signaling Protein. *Hypoxia (Auckl).* 2013; 2013:13-17.
36. Yasumoto K, Kowata Y, Yoshida A, Torii S and Sogawa K. Role of the intracellular localization of HIF-prolyl hydroxylases. *Biochim Biophys Acta.* 2009; 1793:792-797.
37. Ko MK and Kay EP. Subcellular localization of procollagen I and prolyl 4-hydroxylase in corneal endothelial cells. *Exp Cell Res.* 2001; 264:363-371.
38. Trelstad RL, Lawley KR and Holmes LB. Nonenzymatic hydroxylations of proline and lysine by reduced oxygen derivatives. *Nature.* 1981; 289:310-312.
39. Jourquin J, Duncan D, Shi Z and Zhang B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics.* 2012; 13 Suppl 8:S20.
40. Wang J, Duncan D, Shi Z and Zhang B. WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 2013; 41:W77-83.
41. Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, Jacobs HM, Kastiris E, Gilpatrick T, Paranal RM, Qi J, Chesi M, Schinzel AC, McKeown MR, Heffernan TP, Vakoc CR, Bergsagel PL, et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell.* 2011; 146:904-917.
42. Belkina AC and Denis GV. BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer.* 2012; 12:465-477.
43. Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA, Magoon D, Qi J, Blatt K, Wunderlich M, Taylor MJ, Johns C, Chicas A, Mulloy JC, Kogan SC, Brown P, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature.* 2011; 478:524-528.
44. Wu SY, Lee AY, Lai HT, Zhang H and Chiang CM. Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting. *Mol Cell.* 2013; 49:843-857.
45. Mertz JA, Conery AR, Bryant BM, Sandy P, Balasubramanian S, Mele DA, Bergeron L and Sims RJ, 3rd. Targeting MYC dependence in cancer by inhibiting BET bromodomains. *Proc Natl Acad Sci U S A.* 2011; 108:16669-16674.
46. Dey A, Nishiyama A, Karpova T, McNally J and Ozato K. Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. *Mol Biol Cell.* 2009; 20:4899-4909.
47. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I, Philpott M, Munro S, McKeown MR, Wang Y, Christie AL, West N, et al. Selective inhibition of BET bromodomains. *Nature.* 2010; 468:1067-1073.
48. Murray JK, Balan C, Allgeier AM, Kasparian A, Viswanadhan V, Wilde C, Allen JR, Yoder SC, Biddlecome G, Hungate RW and Miranda LP. Dipeptidyl-quinolone derivatives inhibit hypoxia inducible factor-1alpha prolyl hydroxylases-1, -2, and -3 with altered selectivity. *J Comb Chem.* 2010; 12:676-686.
49. Tian YM, Yeoh KK, Lee MK, Eriksson T, Kessler BM, Kramer HB, Edelman MJ, Willam C, Pugh CW, Schofield CJ and Ratcliffe PJ. Differential sensitivity of hypoxia inducible factor hydroxylation sites to hypoxia and hydroxylase inhibitors. *J Biol Chem.* 2011; 286:13041-13051.
50. Chowdhury R, Candela-Lena JI, Chan MC, Greenald DJ, Yeoh KK, Tian YM, McDonough MA, Tumber A, Rose NR, Conejo-Garcia A, Demetriades M, Mathavan S, Kawamura A, Lee MK, van Eeden F, Pugh CW, et al. Selective small molecule probes for the hypoxia inducible factor (HIF) prolyl hydroxylases. *ACS Chem Biol.* 2013; 8:1488-1496.
51. Song LP, Zhang J, Wu SF, Huang Y, Zhao Q, Cao JP, Wu YL, Wang LS and Chen GQ. Hypoxia-inducible factor-1alpha-induced differentiation of myeloid leukemic cells is its transcriptional activity independent. *Oncogene.* 2008; 27:519-527.
52. Vickery HB and Schmidt CLA. The history of the discovery of the amino acids. *Chem Rev.* 1931; 169-318.
53. Cheng S, Cetinkaya M and Grater F. How sequence determines elasticity of disordered proteins. *Biophys J.* 2010; 99:3863-3869.
54. Savage KN and Gosline JM. The effect of proline on the network structure of major ampullate silks as inferred from their mechanical and optical properties. *J Exp Biol.* 2008; 211:1937-1947.
55. Bella J, Eaton M, Brodsky B and Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science.* 1994; 266:75-81.
56. Kotch FW, Guzei IA and Raines RT. Stabilization of the collagen triple helix by O-methylation of hydroxyproline residues. *J Am Chem Soc.* 2008; 130:2952-2953.
57. Kay BK, Williamson MP and Sudol M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 2000; 14:231-241.
58. Ball LJ, Kuhne R, Schneider-Mergener J and Oschkinat H. Recognition of proline-rich motifs by protein-protein-interaction domains. *Angew Chem Int Ed Engl.* 2005; 44:2852-2869.
59. Ma B, Elkayam T, Wolfson H and Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A.* 2003; 100:5772-5777.

60. Pokkuluri PR, Raffin R, Dieckman L, Boogaard C, Stevens FJ and Schiffer M. Increasing protein stability by polar surface residues: domain-wide consequences of interactions within a loop. *Biophys J.* 2002; 82:391-398.
61. Wong ET, Na D and Gsponer J. On the importance of polar interactions for complexes containing intrinsically disordered proteins. *PLoS Comput Biol.* 2013; 9:e1003192.
62. Feldman DE, Chauhan V and Koong AC. The unfolded protein response: a novel component of the hypoxic stress response in tumors. *Mol Cancer Res.* 2005; 3:597-605.
63. Wouters BG and Koritzinsky M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nat Rev Cancer.* 2008; 8:851-864.
64. Rouschop KM, van den Beucken T, Dubois L, Niessen H, Bussink J, Savelkoul K, Keulers T, Mujcic H, Landuyt W, Voncken JW, Lambin P, van der Kogel AJ, Koritzinsky M and Wouters BG. The unfolded protein response protects human tumor cells during hypoxia through regulation of the autophagy genes MAP1LC3B and ATG5. *J Clin Invest.* 2010; 120:127-141.
65. Mujcic H, Nagelkerke A, Rouschop KM, Chung S, Chaudary N, Span PN, Clarke B, Milosevic M, Sykes J, Hill RP, Koritzinsky M and Wouters BG. Hypoxic activation of the PERK/eIF2alpha arm of the unfolded protein response promotes metastasis through induction of LAMP3. *Clin Cancer Res.* 2013; 19:6126-6137.
66. Sena JA, Wang L, Heasley LE and Hu CJ. Hypoxia regulates alternative splicing of HIF and non-HIF target genes. *Mol Cancer Res.* 2014; 12:1233-1243.
67. Weigand JE, Boeckel JN, Gellert P and Dimmeler S. Hypoxia-induced alternative splicing in endothelial cells. *PLoS One.* 2012; 7:e42697.
68. Filippakopoulos P and Knapp S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov.* 2014; 13:337-356.
69. Udeshi ND, Mertins P, Svinkina T and Carr SA. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat Protoc.* 2013; 8:1950-1960.
70. Rappsilber J, Mann M and Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc.* 2007; 2:1896-1906.
71. Falcon S and Gentleman R. Using GOSTats to test gene lists for GO term association. *Bioinformatics.* 2007; 23:257-258.
72. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25:25-29.
73. Benjamini Y and Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc.* 1995; 289-300.
74. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P and von Mering C. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 2009; 37:D412-416.
75. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498-2504.
76. Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003; 4:2.
77. Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Stransky M, Waegle B, Schmidt T, Doudieu ON, Stumpflen V and Mewes HW. CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* 2008; 36:D646-650.
78. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015; 43:D204-212.
79. Warnecke A, Sandalova T, Achour A and Harris RA. PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. *BMC Bioinformatics.* 2014; 15:370.
80. Peterson AC, Russell JD, Bailey DJ, Westphall MS and Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics.* 2012; 11:1475-1488.
81. Zhang Z, Tan M, Xie Z, Dai L, Chen Y and Zhao Y. Identification of lysine succinylation as a new post-translational modification. *Nat Chem Biol.* 2011; 7:58-63.
82. Bisgrove DA, Mahmoudi T, Henklein P and Verdin E. Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proc Natl Acad Sci U S A.* 2007; 104:13690-13695.
83. Ruan HB, Han X, Li MD, Singh JP, Qian K, Azarhoush S, Zhao L, Bennett AM, Samuel VT, Wu J, Yates JR, 3rd and Yang X. O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability. *Cell Metab.* 2012; 16:226-237.
84. Nelson JD, Denisenko O and Bomsztyk K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc.* 2006; 1:179-185.
85. Fowler T, Ghatak P, Price DH, Conaway R, Conaway J, Chiang CM, Bradner JE, Shilatifard A and Roy AL. Regulation of MYC expression and differential JQ1 sensitivity in cancer cells. *PLoS One.* 2014; 9:e87003.
86. Crooks GE, Hon G, Chandonia JM and Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188-1190.