

Cite this: *Chem. Sci.*, 2024, 15, 18395 All publication charges for this article have been paid for by the Royal Society of Chemistry

# A constitutional isomer selective chemical proteomic strategy for system-wide profiling of protein lysine 5-hydroxylation†

Yi-Cheng Sin,<sup>ab</sup> Meeyeon Park,<sup>a</sup> Timothy J. Griffin,<sup>a</sup> Jeongsik Yong<sup>\*a</sup> and Yue Chen<sup>ib</sup> <sup>\*a</sup>

Lysine 5-hydroxylation (5-Hyl) has been well recognized as an essential protein post-translational modification regulating cellular structural stability, RNA alternative splicing and epigenetic gene expression. System-wide enrichment and quantification of 5-Hyl targets have been challenging due to their chemical inert nature and difficulties in differentiating structural isomers in a complex biological sample. Here, we report the development of an efficient chemical proteomic workflow for affinity enrichment and constitutional isomer specific profiling of endogenous 5-Hyl substrates based on highly selective periodate chemistry. Our study confidently identified over 1600 5-Hyl sites on 630 proteins in human 293T cells, revealing functional significance of the modification in protein structure, transcription and chromatin regulation. Analysis of histone 5-Hyl sites showed that histones H2B and H1 are major targets of the 5-hydroxylysine epigenetic mark. Quantitative proteomic analysis through our chemical enrichment workflow identified specific 5-Hyl substrate proteins mediated by the overexpression of Jumonji-domain containing protein 6 (JMJD6). Our study uncovered two cancer-relevant alternative splice isoforms of JMJD6 that regulate 5-Hyl proteins in distinct cellular pathways, providing unique insights into the functional roles of JMJD6 alternative splicing in transcriptional regulation and cellular development.

Received 12th August 2024  
Accepted 4th October 2024

DOI: 10.1039/d4sc05397d

rsc.li/chemical-science

## Introduction

As an ancient and evolutionarily conserved protein post-translational modification (PTM), lysine hydroxylation (Hyl) is broadly involved in protein structural maintenance, RNA processing and epigenetics regulations.<sup>1–5</sup> Pioneering studies showed that lysine hydroxylation products can be further classified into 5*R*-hydroxylysine, 5*S*-hydroxylysine, 4-hydroxylysine and 3*S*-hydroxylysine based on structural characteristics, mediated by several families of enzymes including PLOD proteins (procollagen-lysine and 2-oxoglutarate 5-dioxygenases), JMJD4 (Jumonji domain containing protein 4), JMJD6 (Jumonji domain containing protein 6) and JMJD7 (Jumonji domain containing protein 7).<sup>2–8</sup> Among all the constitutional isomers, lysine 5-hydroxylation (5-Hyl) was first discovered and known to be mediated by the PLOD family and JMJD6 protein.<sup>2,3</sup>

JMJD6 is a JmjC family nuclear protein with both arginine demethylation and lysine hydroxylation activities regulating RNA alternative splicing processes, cancer progression and embryo development.<sup>3,4,9–12</sup> According to the TCGA SpliceSeq database, tumors compared to normal tissues may have different preference in the alternative termination of JMJD6.<sup>13</sup> For example, tumor samples in glioblastoma (GBM) showed 86.4 PSI (Percent Spliced-In) at exon 7.2 and 13.4 PSI at exon 8. In contrast, normal tissue samples in the same dataset exhibit approximately 47 PSI at exon 7.2 and 53 PSI at exon 8.<sup>13</sup> The different alternative termination of the JMJD6 gene results in different splicing isoforms but the functional difference of these splicing isoforms is not known.

Due to the lack of affinity enrichment strategies and tools to differentiate constitutional isomers of hydroxylysine, the classic approach to identify 5-Hyl typically involved mass spectrometry (MS)-based identification of the candidate 5-Hyl protein, followed by *in vitro* enzymatic reactions, peptide hydrolysis and HPLC separation of Hyl isomers for comparisons with synthetic standards.<sup>4,6,8</sup> This strategy provided confident identification of 5-Hyl *in vitro* but cannot directly identify endogenous 5-Hyl in a site-specific manner. Recently, a chemical labeling strategy was developed by applying imidate to label 5-Hyl followed by selective hydrolysis with ammonia.<sup>14</sup> Though the strategy could label 5-Hyl, it also

<sup>a</sup>Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota at Twin Cities, Minneapolis, Minnesota, USA. E-mail: YueChen@umn.edu; jyong@umn.edu

<sup>b</sup>Bioinformatics and Computational Biology Program, University of Minnesota at Twin Cities, Minneapolis, Minnesota, USA

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4sc05397d>



readily reacts with unmodified lysine and cysteine. Moreover, their chemistry cannot differentiate 5-, 4- or 3-hydroxylysine modification through the ring structure by the imidate chemistry. In addition, the selectivity of the strategy requires the hydrolysis of the labeling on unmodified lysine in ammonia for one hour at 90 °C, a condition that may induce unwanted side reactions in complex protein mixtures. A more recent study applied JMJD6 affinity enrichment followed by a pan-propionylation chemical proteomic approach to identify hydroxylysine targets based on interaction with JMJD6.<sup>9</sup> However, the exact isomer structure of the modified hydroxylysine was still uncertain.

In this study, we applied periodate oxidation with a highly efficient capturing and releasing strategy for constitutional isomer-selective enrichment and site-specific identification of protein 5-Hyl targets in complex biological samples. Our study reported 1618 human 5-Hyl sites with high confidence and revealed 5-Hyl modifications on RNA binding proteins, histones and transcription factors. Also, we introduced a strategy to enable label-free quantitation of 5-Hyl induced by JMJD6 splicing isoforms and revealed the isoform-specific enrichment of 5-Hyl substrates in diverse biological pathways. Our chemical proteomic strategy and the large coverage of 5-Hyl targets provide unique insights into the mechanisms and regulations of 5-Hyl pathways in cells.

## Results

### Constitutional isomer selective chemical reaction of the 5-hydroxylysine modification

Periodate-cleavage of vicinal diol groups has been widely applied in glycobiology for global and site-specific analysis of protein glycosylation.<sup>15–18</sup> Given the chemical similarity between vicinal diol and the lysine side chain with 5-hydroxylation, we reasoned that periodate-based chemistry can selectively oxidize the side chain of 5-Hyl on proteins to aldehyde but not for other constitutional isomers of hydroxylysine, which would enable constitutional isomer selective identification of protein 5-Hyl (Fig. 1A). The oxidized 5-Hyl-containing peptides or proteins with a reactive aldehyde functional group on the side chain can be further conjugated to hydrazide beads for affinity enrichment and released by Schiff-base formation with methoxyamine for detection with mass spectrometry or western blotting (Fig. 1B).

To validate the specificity of this chemical strategy, we obtained commercially available chemicals, 1-aminopentan-2-ol, 1-aminopentan-3-ol, and 5-amino-2-pentanol that are structurally identical to the side chains of 5-hydroxylysine, 4-hydroxylysine, and 3-hydroxylysine, respectively (Fig. 2A) and performed periodate oxidation followed by methoxyamine labeling. Analysis of these reaction products showed that only 1-aminopentan-2-ol which is structurally similar to the side chain of 5-hydroxylysine was readily oxidized and conjugated with methoxyamine, resulting in a change of the precursor ion  $m/z$  from 104.107  $m/z$  to 102.091, confirming the high selectivity of the periodate reaction for 5-hydroxylysine (Fig. S1†).

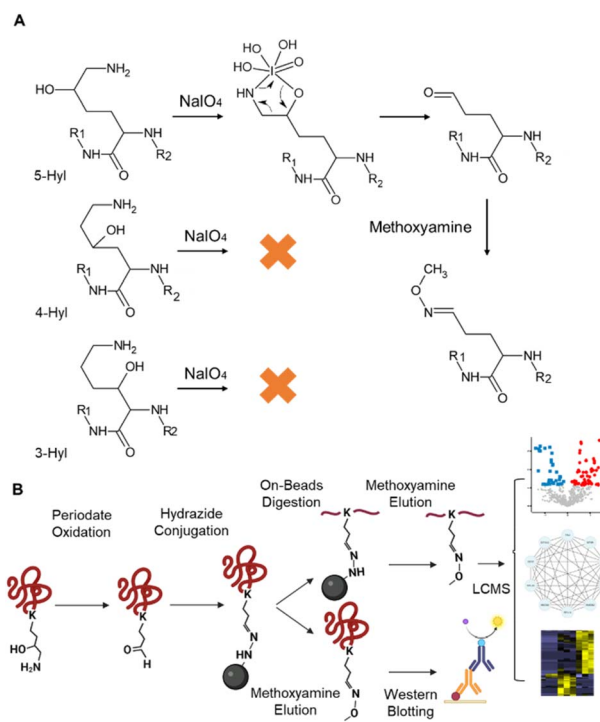


Fig. 1 Constitutional isomer selective enrichment and analysis of protein 5-Hyl modification. (A) Chemistry of the constitutional isomer selective reaction of protein lysine 5-hydroxylation. (B) Enrichment and analysis workflow to systematically profile protein 5-Hyl modification.

### Efficient labeling and detection of lysine 5-hydroxylation in peptides and proteins

To test if the chemical strategy is applicable in a mixture, we applied this chemical workflow to analyze calf collagen digests, which are known to contain 5-Hyl modification.<sup>19</sup> To this end, tryptic peptides from collagen were oxidized with periodate and conjugated with methoxyamine, followed by LC-MS/MS analysis for site-specific identification of modified peptides (Fig. S2A†). From the analysis, we identified the COL1A2 K175 site, a known 5-Hyl site on the peptide “GFPGTPGLPGFKGIR” for targeted quantification (Fig. S2B†).<sup>20</sup> Comparing the percentages of chromatographic peak areas of the unreacted, oxidized and methoxyamine-conjugated peptide forms in LCMS analysis after reactions showed that the efficiency of oxidation was 94% (sum of the peak area percentages of oxidized and conjugated peptide forms) and the efficiency of conjugation was over 70% (Fig. 2B). Based on the potential gas-phase chemical reaction, we further deduced the neutral loss signature for peptides with lysine side chains modified with either aldehyde or methoxyamine conjugation that served as a diagnostic indicator of the modification (Fig. S3†). This neutral loss feature was further validated by comparing the spectra of the D<sub>3</sub>- and H<sub>3</sub>-methoxyamine conjugated peptides with that of *in vivo* 5-Hyl (Fig. S4†).

Using biotin-hydrazide instead of methoxyamine for conjugation, we performed streptavidin blot to validate the selectivity and efficiency of our labeling strategy. We prepared nuclear







beads. Following incubation, the beads were extensively washed. Conjugated proteins were digested on beads with trypsin followed by a second wash to remove non-conjugated peptides. Finally, conjugated peptides were eluted off the beads with competitive methoxyamine elution and desalted for LCMS analysis (Fig. 3A). For specific characterization of histone 5-Hyl sites, we performed acid extraction of nuclear pellets from 293T cells expressing short-form JMJD6. Extracted proteins were subjected to a similar procedure for the enrichment and identification of 5-Hyl targets as the analysis of nuclear lysates (Fig. 3A).

Our systematic profiling confidently identified a total of 1618 5-Hyl sites on 630 proteins, representing the largest coverage of the 5-Hyl proteome to date with constitutional isomer specificity (Fig. 3B, C, S6, Tables S1 and S2†). Nearly 60% of the proteins were identified with more than one site and nearly 10% were identified with  $\geq 6$  sites, suggesting the preference of 5-Hyl to target multiple lysines on the same substrate (Fig. 3D).

Histones were found to be extensively modified with 5-Hyl in our analysis with 40 site identifications, which agreed with the observations in the streptavidin blot (Fig. 3E, Tables S1 and S2†). Interestingly, most of the histone 5-Hyl sites were identified on variants of histones H2B and H1, which is in distinct contrast to other well-known histone epigenetic marks such as acetylation and methylation that are abundant on histones H3 and H4 (Fig. 3E and S7†). Even including lysine acetylation as a variable modification in data analysis did not boost the identification of 5-Hyl sites on histones H3 and H4. Notably, the 5-Hyl sites on histones H3 and H4 were only identified in the histone-enriched acid extraction samples with low spectral counts. These data suggested that histones H2B and H1 are major targets of histone lysine 5-hydroxylation. In this study, we also identified known JMJD6 targets U2AF2 and BRD4 (ref. 3 and 9) with K70 sites on U2AF2 and nine sites on BRD4 (Tables S1 and S2†).

To evaluate if our workflow could also identify O-GlcNAc-containing proteins as previously suggested,<sup>21</sup> we analyzed our data by specifying methoxyamine-conjugated O-GlcNAc modification as variable modifications for Ser and Thr. Indeed, we identified over 1200 O-GlcNAc modification sites from the nuclear lysate dataset, suggesting that our chemical proteomics workflow can be efficiently applied to simultaneously study multiple PTMs in a streamlined manner (Table S3†).

### Bioinformatics analysis revealed structural characteristics and functional enrichment of protein 5-Hyl targets in human cells

To determine the enrichment of structure features and biological pathways, we performed bioinformatic analysis of 5-Hyl protein targets identified in the total nuclear lysate analysis. Flanking sequence motif analysis showed the strong preference of 5-Hyl on lysine-rich regions with some preferences for glutamic acid, serine, alanine and proline in the close vicinity of the modified lysine (Fig. 4A). Aligning identified 5-Hyl sites with experimentally validated secondary structures in the Uniprot human database showed that only 10% of 5-Hyl sites were

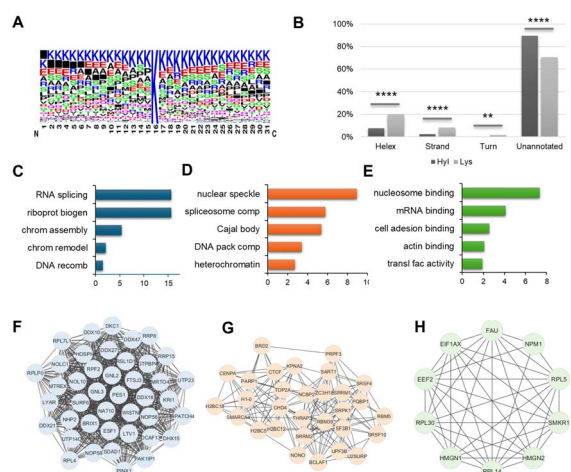


Fig. 4 Bioinformatic characterization of 5-Hyl proteins from the system-wide profiling analysis. (A) Flanking sequence analysis of protein 5-Hyl modification sites identified in this study. (B) Secondary structure distribution analysis of protein 5-Hyl modification sites (dark grey bars) compared to all lysines among identified peptides in this study (light grey bars). \*\*\*\* indicates  $p < 0.0001$  in the hypergeometric test. (C–E) Gene ontology annotation enrichment analysis of 5-Hyl proteins in this study on (C) biological processes, (D) cellular compartment and (E) molecular function. (F–H) Network representation of highly connected clusters of 5-Hyl proteins with (F) a subnetwork related to ribonucleoprotein biogenesis, (G) a subnetwork related to transcription and DNA/RNA binding and (H) a subnetwork related to the ribonucleoprotein complex.

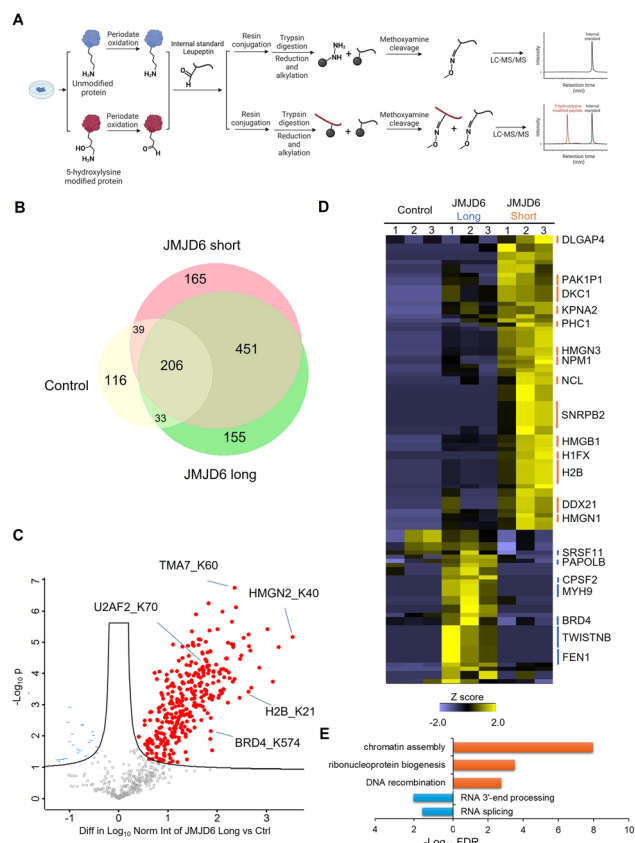
annotated in the secondary structures with 8% localized to alpha-helix and 2% localized to the beta strand, while 30% of all lysines in tryptic peptides were annotated in secondary structures with 20% localized to alpha-helix, 8% localized to the beta strand and 2% localized to turns (Fig. 4B). These data indicated a significant depletion of 5-Hyl on protein regions with secondary structures.

Pathway and gene ontology enrichment analysis showed that 5-Hyl modified proteins were significantly enriched in pathways and processes related to RNA processing and chromatin remodeling with cellular localization enrichment in nuclear speckle, spliceosome and heterochromatin (Fig. 4C–E and Table S3A–C†). Interaction network analysis identified multiple highly connected interaction subnetworks including the ribonucleoprotein biogenesis network, the transcription network and ribonucleoprotein complexes (Fig. 4F–H). Hence, our analysis of 5-Hyl modified proteins suggested the critical role of modification in RNA processing and chromatin-related activities in human cells.

### Differential analysis of JMJD6 long-form and short-form regulated 5-Hyl sites

For quantitative analysis of JMJD6-regulated 5-Hyl sites, we applied a label-free quantification strategy with an aldehyde-containing peptide, leupeptin, as a spike-in internal standard after periodate oxidation (Fig. 5A). By normalizing the identified 5-Hyl peptide peak intensity to the methoxyamine conjugated leupeptin, this strategy eliminated quantification variations in





**Fig. 5** Quantitative analysis of 5-Hyl modification sites in nuclear lysate analysis mediated by JMJD6 expression. (A) A scheme for the label-free quantitative of site-specific 5-Hyl with a spike-in internal standard. (B) A Venn diagram showing the overlap of protein 5-Hyl identification with or without the expression of JMJD6 splicing isoforms. (C) A volcano plot showing the dynamic changes in 5-Hyl abundances upon the expression of N-terminally flag-tagged JMJD6 long isoform (two-sample Student's *T*-test with FDR < 0.05). (D) A heatmap and clustering analysis of significantly changed 5-Hyl sites induced by the overexpression of either the long or short form of JMJD6. Normalized label-free quantitative of significantly changed 5-Hyl targets ( $p < 0.05$ , one-tail Student's *T*-test analysis assuming unequal variance) were Z-score converted and clustered with hierarchical clustering. (E) Gene ontology biological process enrichment analysis of 5-Hyl proteins differentially induced by the expression of specific JMJD6 isoforms (orange bars: short-form and blue bars: long-form).

the enrichment process such as variations in the efficiencies of Schiff-base formation and cleavage as well as inconsistent loss during sample processing.

Our systematic profiling showed that overexpression of either the long or short form of JMJD6 led to an apparent increase in the total number of 5-Hyl site identifications in nuclear lysate analysis with 845 sites identified from long form expression and 861 sites identified from short form expression, compared to 394 sites identified from the control plasmid expression (Fig. 5B). The large overlap between these analyses demonstrated excellent reproducibility in identifying common 5-Hyl modification sites. We further performed the label-free quantitative analysis with internal standard normalization. The

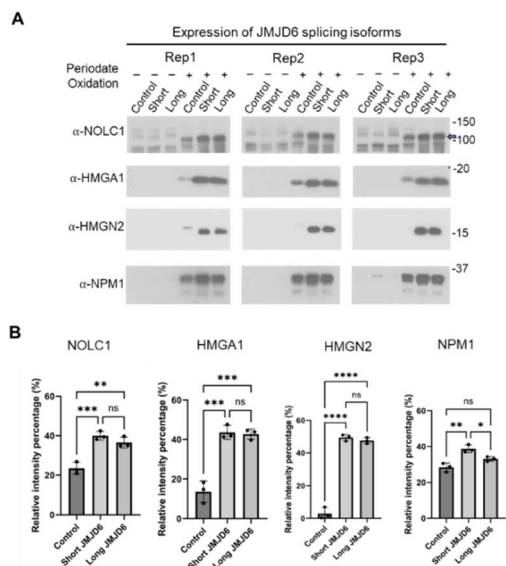
volcano plot analysis identified 283 sites and 303 sites that were significantly upregulated with the overexpression of the JMJD6 short form and long form, respectively (two-sample Student's *T*-test with an FDR threshold of 0.05) (Fig. 5C and S8†). Given the well-recognized role of JMJD6 in transcription and RNA processing, it is important to note that the upregulation of specific 5-Hyl site abundance upon the expression of JMJD6 isoforms could not only be due to direct JMJD6 hydroxylase activities, but also result from JMJD6-mediated transcriptional activities or potentially other 5-Hyl enzymes regulated by JMJD6.

Quantitative comparison of 5-Hyl abundance from the expression of long or short JMJD6 isoform identified a group of 108 sites, among which 71 sites showed a stronger increase in abundance with the expression of the short-form JMJD6, while 37 sites showed a stronger increase in abundance with the expression of the long-form JMJD6 (Fig. 5D and Table S5†). Pathway and functional annotation enrichment analysis showed that 5-Hyl proteins with a higher abundance by the short-form JMJD6 expression were significantly enriched in chromatin assembly (adj. FDR <  $1.3 \times 10^{-8}$ ) and ribonucleoprotein complex biogenesis (adj. FDR <  $3.2 \times 10^{-4}$ ) processes (Fig. 5E). On the other hand, 5-Hyl proteins with a higher abundance upon the long-form JMJD6 expression were significantly enriched in RNA 3'-end processing (adj. FDR <  $0.3 \times 10^{-3}$ ) and RNA splicing (adj. FDR <  $2.7 \times 10^{-2}$ ) (Fig. 5E). Interestingly, motif enrichment analysis showed that 5-Hyl sites identified by the overexpression of the short-form JMJD6 harbored a unique "PK" motif which likely correlates with structural characteristics and substrate interaction of the short-form JMJD6 (Fig. S9†). Taken together, these data clearly demonstrated that splicing isoforms of JMJD6 have differential preferences for regulated 5-Hyl sites and aberrant expression of JMJD6 splicing isoform under certain pathophysiological conditions may affect the 5-Hyl substrates and activities in specific cellular pathways.

### Chemical pulldown assay and western blotting analysis of 5-Hyl modification proteins regulated by JMJD6

To further confirm these findings, we performed a chemical pulldown assay with periodate oxidation of 5-Hyl proteins, hydrazide-bead conjugation and methoxyamine elution followed by western blotting detection (Fig. S10†). From the list of our identifications, we selected NOLC1, HMGN2, HMGA1 and NPM1 as examples where each of these proteins was identified with high quality MS/MS spectra and confident site localizations (Fig. S11A–D†). Our western blots showed that these proteins showed minimal signals from the pulldown assay without oxidation regardless of the overexpression conditions. Upon oxidation and pulldown, each of these proteins was clearly captured, enriched and detected. Expression of either the long or short form of JMJD6 led to a significant increase among each pulled-down target, consistent with our label-free quantitative analysis of NOLC1, HMGN2 and HMGA1 (Fig. 6). Interestingly, NPM1 showed a consistently stronger intensity in the chemical pulldown assay upon the expression of the short-form JMJD6 compared to the expression of the long-form JMJD6 across multiple biological replicates, confirming that the





**Fig. 6** Chemical pull-down assay to validate JMJD6-regulated 5-Hyl proteins. 293T cells were transfected with either a control plasmid or a flag-tagged JMJD6 short or long form plasmid. Proteins were pulled down with hydrazide beads following periodate oxidation and eluted using the methoxyamine buffer. (A) Western blotting analysis of selected 5-Hyl proteins, NOLC1, HMGA1, HMGN2 and NPM1. (B) Quantification of the western blotting analysis from the chemical pull-down assay with one-way ANOVA. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ .

expression of short and long forms of JMJD6 have differential regulations towards the overall NPM1 5-Hyl abundance *in vivo* (Fig. 6).

## Conclusions and discussion

In this study, we developed an efficient chemical proteomics workflow for global enrichment and constitutional isomer specific profiling of 5-Hyl protein modification substrates based on highly selective periodate chemistry. Confident identification of 1618 5-Hyl sites on 630 proteins represented the largest coverage of the modification in human cells with constitutional isomer specificity and suggested that lysine 5-hydroxylation is a widespread PTM broadly involved in chromatin and RNA processing activities in addition to the canonical view of its important functions related to collagen crosslinking and glycosylation. Interestingly, we observed significant depletion of 5-Hyl on the secondary structure indicating a potentially important role of the modification in protein folding and structural stability. Application of label-free quantitative analysis demonstrated the *in vivo* activity of JMJD6 and allowed us to confidently identify over 300 potential JMJD6 regulatory sites. Importantly, with our strategy, we discovered that long and short alternative splicing forms of JMJD6 exhibit differential activities towards specific proteins or signaling pathways in cells. A technical limit of the current study was that we could not determine whether the 5-Hyl sites with increased abundance with JMJD6 overexpression were JMJD6 direct hydroxylation targets. Further technology development by integrating this

chemical proteomics workflow with isotopic-labeling strategies such as SILAC and TMT multiplexing<sup>22,23</sup> and quantitative JMJD6 interactome analysis<sup>9</sup> will reveal novel insights into direct JMJD6 targets in mammalian cells.

Our study revealed many novel targets as 5-Hyl substrates. Histones are highly abundant nuclear proteins and decorated with over 10 types of PTMs for epigenetic regulation of gene expression and the chromatin structure.<sup>24</sup> We identified histone lysine 5-hydroxylation with both streptavidin blot and MS analysis. Surprisingly, we found that histones H2B and H1 are major targets of 5-Hyl modification, suggesting the potentially distinct role of 5-Hyl as an epigenetic mark compared to other well-studied histone modifications on histones H3 and H4 such as acetylation and methylation. From our analysis, we also identified eukaryotic peptide chain release factor subunit 1 (ETF1) K63 as a 5-Hyl target (Fig. S11E†). Interestingly, a previous study showed with *in vitro* enzymatic reactions that ETF1 K63 is targeted by JMJD4 for 4-hydroxylysine modification to regulate translational termination efficiency.<sup>6</sup> Our analysis demonstrated that ETF1 can be targeted for 5-Hyl *in vivo* as a potentially alternative mechanism to regulate translational termination.

JMJD6 is increasingly recognized for its multifaceted role in the pathogenesis of various diseases, particularly cancer, due to its enzymatic functions as a lysyl hydroxylase and arginine demethylase.<sup>25–29</sup> These activities are crucial for modulating RNA splicing, transcription, and signal transduction pathways. Our study, highlighting differential activities among JMJD6 isoforms, introduces a novel aspect to our understanding of its pathogenic roles. This discovery not only enhances our grasp of JMJD6-driven mechanisms but also sharpens our focus on identifying precise molecular targets, potentially unveiling novel therapeutic strategies and elucidating previously unrecognized pathogenic pathways. Furthermore, our chemical MS-based proteomic workflow should be applicable to characterization of other enzyme mediators of this important protein PTM.

## Experimental section

### Chemicals and reagents

Anti-B23/NPM1 polyclonal antibody (10306-1-AP), anti-HMGN2 antibody (10953-1-AP), anti-HMGA1 polyclonal antibody (29895-1-AP), anti-JMJD6 antibody (16476-1-AP) and anti-NOLC1 polyclonal antibody (11815-1-AP) were purchased from Proteintech (Billerica, MA). Anti-rabbit IgG (HRP-linked) antibody and anti-DYKDDDDK Tag antibody (D6W5B) were from Cell Signaling (Danvers, MA). Ultralink hydrazide resins (53149) and sodium *meta*-periodate (PI20504) were from ThermoFisher Scientific (Waltham, MA). Methoxyamine hydrochloride (SIAL-226904), (+)-biotin hydrazide (B7639) and benzonase (70746-4) were from Sigma-Aldrich (St. Louis, MO). Methoxyl-d3-amine HCl (D-1272) was from CDN Isotopes (Pointe-Claire, Quebec, Canada). Leupeptin hemisulfate (L22035-0.005) was from RPI (Mount Prospect, IL). HRP-labeled streptavidin (405210) was from BioLegend (San Diego, CA). The FLAG-HA-pcDNA3.1 plasmid was a kind gift from Adam Antebi (Addgene # 52535). 5-Amino-





2-pentanol (ST0059AI) was from 1st Scientific (San Diego, CA). 1-Aminopentan-3-ol (CSSB00000748770) and 1-aminopentan-2-ol (CSSB00000755064) were from ChemSpace (Monmouth Junction, NJ).

### Plasmid preparation for JMJD6 splicing variants

Human JMJD6 isoforms were cloned into pCMV10 (addgene plasmid #47334) for N-terminal flag-epitope tagging by PCR using the following primer sets. For the JMJD6 forward primer, 5'-GATAAAGCTTAACCACAAGAGCAAGAAGCGCATCCG-3' was used. For the short form JMJD6 isoform construction (ENST00000397625.9), a reverse primer 5'-ATTATCTAGATCACCTGGAGGAGCTGCGCTCTTTGCTGAC-3' was used. For the long form JMJD6 isoform construction (ENST00000445478.6), the 5'-ATTATCTAGATCAGGGGTGAGCCCGGCCTCCA-CAAGTGTCC-3' reverse primer was used. Total RNAs from HEK293 cells were isolated and oligo d(T) primed cDNA synthesis was done. Then, PCR amplification using the above primers was conducted for cloning into pCMV10.

### Cell culture and transient transfection

293T cells (ATCC #CRL-3216) were cultured at 37 °C with 5% CO<sub>2</sub> and maintained in Dulbecco's Modified Eagle Medium (DMEM) (ThermoFisher Scientific, Waltham, MA), supplemented with 10% fetal bovine serum (Sigma, St. Louis, MO), 100 IU penicillin, and 100 µg ml<sup>-1</sup> streptomycin (Genesee Scientific, Morrisville, NC). Cells were seeded in a 15 cm plate overnight and were then transfected with 1 ml of premixed medium containing 45 µg polyethylenimine (PEI) (Sigma, St. Louis, MO) and 20 µg plasmid for 48 hours.

### Reactivity analysis with chemical compounds

Aminopentanol at a final concentration of 25 mM was mixed with 75 mM sodium periodate in 0.25 M sodium acetate solution pH 4.5 and incubated in the dark for 5 min. Then, an equal volume of 200 mM methoxyamine was added to the sample and incubated at room temp. for 5 minutes. Finally, 20% volume of acetonitrile was added to the sample and the sample was infused to an Exactive Orbitrap mass spectrometer (ThermoFisher Scientific, Waltham, MA) with a positive mode full scan (90–110 *m/z*) and 25 000 resolution (FWHM) at *m/z* 400.

### Hydroxylysine reactivity analysis with collagen

Calf collagen (Sigma, St. Louis, MO) was dissolved in 9 M urea with 100 mM TEAB and centrifuged to remove the pellet. The dissolved collagen solution was added to TCEP and IAA to a final concentration of 10 mM for each with rotation at room temp in the dark for 30 min for reduction and alkylation. Then, samples were added to 20 mM beta mercaptoethanol to quench the reaction. The solution was diluted with water to 1.5 M Urea and subjected to trypsin digestion. The digested protein solution was further desalted using an Oasis C18 cartridge (Waters, Milford, MA) and reconstituted into 200 mM sodium acetate pH 4.5 solution. The samples were centrifuged to remove the pellet and incubated with 50 mM sodium periodate in 200 mM

sodium acetate pH 4.5 in the dark at room temp with rotation for 30 min. The oxidized peptides were desalted using Pierce™ C18 desalting tips (ThermoFisher Scientific, Waltham, MA). Finally, the desalted oxidized peptides were incubated with 200 mM methoxyamine 200 mM sodium acetate pH 4.5 at room temp for 30 min with rotation and desalted using home packed StageTips for LCMS analysis.

### Nuclear lysate preparation for chemical labelling

Cells were briefly rinsed and scraped down with PBS and were lysed in the hypotonic buffer (Active motif, Carlsbad, CA) with 0.5% NP40 and 1 mM PMSF (ThermoFisher Scientific, Waltham, MA) and incubated on ice with manual rotation from time to time for 10 minutes followed by centrifugation at 16 000×*g* to obtain the nuclear pellets. The nuclear pellets were then lysed with the urea lysis buffer (9 M urea with 1 mM PMSF) and subjected to sonication (45%, 10 cycles, one second on and one second off for each cycle), and centrifugation at 21 000×*g* to obtain soluble nuclear extracts in the supernatant. To remove the oligonucleotide in the sample, the lysate was diluted with water to 3 M urea and treated with 250–290 units of benzonase and 1 mM MgCl<sub>2</sub> at the final concentration for incubation with rotation at 37 °C for 30 min. Then, the reaction was quenched with 0.5% SDS (w/w) and 50 mM sodium acetate pH 4.5 and the sample was subjected to buffer exchange with water through an Amicon® Ultra-15 centrifugal filter unit (3 kDa) (Sigma, St. Louis, MO). After the buffer exchange, the sample was subjected to oxidation in the periodate oxidation buffer (100 mM sodium periodate, 0.1% SDS (w/w) and 50 mM sodium acetate at pH 4.5) at room temperature with rotation in the dark for an hour. Then, the protein solution was subjected to a second buffer exchange. The protein concentration was measured with the Bradford assay (ThermoFisher Scientific, Waltham, MA).

### Acid extraction for chemical labelling

Cells expressing short form JMJD6-flag were pelleted and lysed in a lysis buffer containing 1× PBS with 0.5% Triton X-100 and 1 mM PMSF (ThermoFisher Scientific, Waltham, MA) for 30 min at 4 °C and centrifuged to obtain the pellet. Then, the pellet was subjected to acid extraction in 0.4 N H<sub>2</sub>SO<sub>4</sub> overnight. The extracted proteins were precipitated using 20% trichloroacetic acid (TCA) and washed with cold acetone once. The precipitated proteins were dissolved in 0.1% SDS with 100 mM ammonium bicarbonate and subjected to oxidation with 100 mM sodium periodate at room temperature with rotation in the dark for an hour. The oxidized protein was again precipitated using 20% TCA and washed with cold acetone once. Finally, the oxidized protein precipitates were dissolved in 0.1% SDS with 100 mM ammonium bicarbonate and 50 mM sodium acetate (pH 4.5) and incubated with the hydrazide resins for enrichment, followed by on-bead tryptic digestion and elution for LCMS analysis similar to the processing of the nuclear lysate samples.

### Streptavidin blot detection of 5-Hyl proteins

Nuclear protein lysates with or without periodate oxidation were added to biotin hydrazide to a final concentration of 10 mM.



Then, samples were further reconstituted in LAEMMLI SDS sample buffer with pH neutralization and proteins were labeled at room temperature with rotation for 30 min. After the incubation, the samples were boiled at 95 °C and subjected to SDS-PAGE analysis for western blot and Imperial™ Protein Stain (ThermoFisher Scientific, Waltham, MA).

### Western blot detection of 5-Hyl proteins with the protein chemical pulldown assay

Nuclear protein lysates with or without periodate oxidation were incubated with the hydrazide resins in acetate buffer (0.1% SDS and 50 mM sodium acetate at pH 4.5) at room temperature overnight with rotation. After the overnight incubation, the supernatant was removed and the beads were subjected to four times washing with 9 M urea, one time washing with 9 M urea in 0.4 M ammonium bicarbonate, one time washing with PBS and one time washing with water. After the washing, the protein conjugated beads were treated with a final concentration of 100 mM methoxyamine in the LAEMMLI SDS sample buffer and incubated at room temperature for 30 min. After the incubation, the sample's pH was neutralized and the sample was then boiled at 95 °C for 5 minutes for SDS-PAGE analysis and western blotting.

### Analysis of 5-Hyl peptides with the protein chemical pulldown assay

One milligram of nuclear proteins from lysates with periodate oxidation was added to 0.5 ng of leupeptin hemisulfate as an internal standard. Proteins were incubated with the hydrazide resins in acetate buffer (0.1% SDS and 50 mM sodium acetate at pH 4.5) with rotation at room temperature overnight. After the supernatant removal, the beads were washed with 9 M urea four times, 9 M urea with 0.4 M ammonium bicarbonate once and PBS once. Then, beads were re-suspended in 100 mM ammonium bicarbonate buffer and the proteins were digested using trypsin with rotation at 37 °C overnight. Following digestion, the beads were washed with PBS four times. Then, we performed the reduction and alkylation on beads with 10 mM TCEP and 10 mM iodoacetamide at the final concentrations in 100 mM ammonium bicarbonate for 30 min with rotation at room temperature in the dark, followed by four times washing with PBS. Finally, the peptides conjugated on beads were eluted using the methoxyamine elution buffer (0.2 M methoxyamine, 0.2 M sodium acetate pH 4.5 and 0.1 M aniline) three times. The eluted fractions were pooled and peptides were desalted for LCMS analysis. Biological triplicate analysis was performed for each treatment condition in quantification.

### NanoHPLC – Orbitrap MS analysis

Peptides were analyzed with a Dionex Ultimate 3000 RSLCnano HPLC system coupled to an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (ThermoFisher Scientific, Waltham, MA). Digested peptides dissolved in HPLC buffer A (0.1% formic acid in water (v/v)) were separated on an in-house packed capillary HPLC column (length 20 cm and 75 μm inner diameter) from CoAnn Technologies with Luna C18 beads (5 μm

particle size, 100 Å pores) from Phenomenex. The peptides were separated with a 30-min (for collagen peptide analysis), 60-min (for 5-Hyl site analysis in histone acid extract analysis) or 90-min gradient (for 5-Hyl site analysis in nuclear lysate) from 1% to 95% HPLC buffer B (0.1% formic acid in acetonitrile (v/v)) in HPLC buffer A. Full scan of precursor ions (MS1) was performed in a positive mode and with a resolution of 120 K (at 200 *m/z*) in a mass range of 375–1600 *m/z*. The linear ion trap with an isolation window of 1.6 *m/z* using a quadrupole and a 35% of normalized high-energy collision dissociation (HCD) energy was set for the tandem mass spectral (MS/MS) analysis. Dynamic exclusion for 45 seconds with a mass tolerance of ±10 ppm was enabled for the MS/MS acquisition.

### MaxQuant data analysis for the identification of peptides and proteins

LCMS data was analyzed with the MaxQuant search engine (ver 2.2.0.0) against the Uniprot human reference protein database (UP000005640\_9606) concatenated with reversed decoy sequences and common contaminant sequences.<sup>30</sup> Carbamidomethylation on cysteine was specified as a fixed modification, trypsin was specified as the protease and match between runs was allowed. By default, precursor mass tolerance was set at 4.5 ppm and MS/MS mass tolerance was set at 0.5 Da. A stringent 1% false discovery rate was specified for the identification of proteins, peptides and modification sites with an additional requirement of a minimum score of 40 for the modification site identification. To analyze 5-Hyl modification, the maximum missing cleavage was specified as 6 and the maximum number of modifications allowed per peptides was specified as 6. Variable modifications included oxidized 5-Hyl with methoxyamine conjugation modification (defined as the composition of OH(-2) and a neutral loss of CH(5)NO, and not at the C terminal), protein N-terminal acetylation and methionine oxidation. For the identification of histone 5-Hyl sites that may be co-modified by acetylation, lysine acetylation was included as an additional variable modification. To analyze O-GlcNAc modification, the maximum missing cleavage was specified as 1 and the maximum number of modifications allowed per peptides was specified as 2. Variable modifications included oxidized O-GlcNAc with methoxyamine conjugation modification (defined as the composition of C(10)H(17)N(3)O(10) with a neutral loss of C(10)H(17)N(3)O(10)), protein N-terminal acetylation and methionine oxidation.

### Perseus analysis for label-free quantification

All peptide and protein identifications from the reversed decoy sequences or contaminant sequences or identifications with an Uniprot ID not matching the gene name were removed for label-free quantitative analysis. The total intensity of each modification site was normalized by the chromatographic peak area of methoxyamine-conjugated leupeptin (the spike-in internal standard) of each sample. Then, the data were analyzed with the Perseus bioinformatics platform (version 1.6.15.0)<sup>31</sup> and statistical analysis performed with two sample Student's *T*-tests with an FDR threshold of 0.05.





## Bioinformatics analysis and software

A flanking sequence motif was generated with Weblogo.<sup>32</sup> Pathway and gene ontology enrichment analysis were performed with Webgestalt<sup>33</sup> and statistical significance was specified as an FDR of 0.05 with Benjamini–Hochberg adjustment. The protein–protein interaction network was extracted based on the String database<sup>34</sup> and visualized with Cytoscape.<sup>35</sup> Protein clusters with significant connectivity were identified with MCODE.<sup>36</sup> Heat map and hierarchical clustering analysis was performed with Genesis.<sup>37</sup> Flanking sequence motif analysis was performed with R package RmotifX.<sup>27</sup> Schematic figures for workflows were created with <https://Biorender.com>. The molecular structure was drawn with Chemsketch (2022.2.2, Advanced Chemistry Development, Inc. (ACD/Labs), Toronto, ON, Canada). Bar graph representation with statistical analysis was prepared with GraphPad Prism (GraphPad Software, Boston, Massachusetts, USA). Quantitative image analysis of western blotting was performed with Image J.<sup>38</sup> Other data analysis and figure presentation was performed with in-house developed R using the libraries (readxl, ggplot2, dplyr, openxlsx, and eulerr) and Python scripts.

## Data availability

The data supporting this article have been included as part of the ESI.† The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE<sup>39</sup> partner repository with the dataset identifier PXD054608.

## Author contributions

YS performed all biochemical, chemical and proteomics studies, MP performed cloning and plasmid preparation, TG, JY and YC supervised the study, and YS and YC wrote the manuscript with inputs from all authors.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We are very thankful for the discussion and suggestions from the members of the Chen lab and helpful advice from Drs David Bernlohr and Doug Mashek. We greatly appreciate Drs Yingchun Zhao and Peter Villalta from the mass spectrometry resource center in the Masonic Cancer Center at the University of Minnesota for LCMS setup and access. YS and YC were supported by funding from the University of Minnesota and the National Institute of Health (GM124896 to YC) and JY was supported by the National Institute of Health (GM113952 to JY).

## References

- 1 D. D. Van Slyke and A. Hiller, An Unidentified Base among the Hydrolytic Products of Gelatin, *Proc. Natl. Acad. Sci. U. S. A.*, 1921, 7, 185–186.
- 2 E. Hausmann, Cofactor requirements for the enzymatic hydroxylation of lysine in a polypeptide precursor of collagen, *Biochim. Biophys. Acta*, 1967, 133, 591–593.
- 3 C. J. Webby, A. Wolf, N. Gromak, M. Dreger, H. Kramer, B. Kessler, M. L. Nielsen, C. Schmitz, D. S. Butler, J. R. Yates 3rd, C. M. Delahunty, P. Hahn, A. Lengeling, M. Mann, N. J. Proudfoot, C. J. Schofield and A. Bottger, Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing, *Science*, 2009, 325, 90–93.
- 4 M. Unoki, A. Masuda, N. Dohmae, K. Arita, M. Yoshimatsu, Y. Iwai, Y. Fukui, K. Ueda, R. Hamamoto, M. Shirakawa, H. Sasaki and Y. Nakamura, Lysyl 5-hydroxylation, a novel histone modification, by Jumonji domain containing 6 (JMJD6), *J. Biol. Chem.*, 2013, 288, 6053–6062.
- 5 W. Manni, X. Jianxin, H. Weiqi, C. Siyuan and S. Huashan, JMJD family proteins in cancer and inflammation, *Signal Transduction Targeted Ther.*, 2022, 7, 304.
- 6 T. Feng, A. Yamamoto, S. E. Wilkins, E. Sokolova, L. A. Yates, M. Munzel, P. Singh, R. J. Hopkinson, R. Fischer, M. E. Cockman, J. Shelley, D. C. Trudgian, J. Schodel, J. S. McCullagh, W. Ge, B. M. Kessler, R. J. Gilbert, L. Y. Frolova, E. Alkalaeva, P. J. Ratcliffe, C. J. Schofield and M. L. Coleman, Optimal translational termination requires C4 lysyl hydroxylation of eRF1, *Mol. Cell*, 2014, 53, 645–654.
- 7 Y. Qi and R. Xu, Roles of PLODs in Collagen Synthesis and Cancer Progression, *Front. Cell Dev. Biol.*, 2018, 6, 66.
- 8 S. Markolovic, Q. Zhuang, S. E. Wilkins, C. D. Eaton, M. I. Abboud, M. J. Katz, H. E. McNeil, R. K. Lesniak, C. Hall, W. B. Struwe, R. Konietzny, S. Davis, M. Yang, W. Ge, J. L. P. Benesch, B. M. Kessler, P. J. Ratcliffe, M. E. Cockman, R. Fischer, P. Wappner, R. Chowdhury, M. L. Coleman and C. J. Schofield, The Jumonji-C oxygenase JMJD7 catalyzes (3S)-lysyl hydroxylation of TRAFAC GTPases, *Nat. Chem. Biol.*, 2018, 14, 688–695.
- 9 M. E. Cockman, Y. Sugimoto, H. B. Pegg, N. Masson, E. Salah, A. Tumber, H. R. Flynn, J. M. Kirkpatrick, C. J. Schofield and P. J. Ratcliffe, Widespread hydroxylation of unstructured lysine-rich protein domains by JMJD6, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, 119, e2201483119.
- 10 B. Chang, Y. Chen, Y. Zhao and R. K. Bruick, JMJD6 is a histone arginine demethylase, *Science*, 2007, 318, 444–447.
- 11 N. Raguz, A. Heim, E. Engal, J. Wesche, J. Merl-Pham, S. M. Hauck, S. Erkelenz, H. Schaal, O. Bensaude, A. Wolf, M. Salton and A. Bottger, JMJD6 Regulates Splicing of Its Own Gene Resulting in Alternatively Spliced Isoforms with Different Nuclear Targets, *Int. J. Mol. Sci.*, 2020, 21, 6618.
- 12 C. M. Jablonowski, W. Quarni, S. Singh, H. Tan, D. H. Bostanthirige, H. Jin, J. Fang, T. C. Chang, D. Finkelstein, J. H. Cho, D. Hu, V. Pagala, S. M. Sakurada, S. M. Pruett-Miller, R. Wang, A. Murphy, K. Freeman, J. Peng, A. M. Davidoff, G. Wu and J. Yang, Metabolic reprogramming of cancer cells by JMJD6-mediated pre-mRNA splicing associated with therapeutic response to splicing inhibitor, *eLife*, 2024, 12, RP90993.
- 13 M. Ryan, W. C. Wong, R. Brown, R. Akbani, X. Su, B. Broom, J. Melott and J. Weinstein, TCGASpliceSeq a compendium of



- alternative mRNA splicing in cancer, *Nucleic Acids Res.*, 2016, **44**, D1018–D1022.
- 14 G. Hayashi, R. Sakamoto and A. Okamoto, 2-Oxazoline formation for selective chemical labeling of 5-hydroxylysine, *Chem. - Asian J.*, 2015, **10**, 1138–1141.
- 15 H. Zhang, X. J. Li, D. B. Martin and R. Aebersold, Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry, *Nat. Biotechnol.*, 2003, **21**, 660–666.
- 16 S. Pan, R. Chen, R. Aebersold and T. A. Brentnall, Mass spectrometry based glycoproteomics-from a proteomics perspective, *Mol. Cell. Proteomics*, 2011, **10**, R110003251.
- 17 T. N. Ramya, E. Weerapana, B. F. Cravatt and J. C. Paulson, Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans, *Glycobiology*, 2013, **23**, 211–221.
- 18 S. Xu, J. Zheng, H. Xiao and R. Wu, Simultaneously Identifying and Distinguishing Glycoproteins with O-GlcNAc and O-GalNAc (the Tn Antigen) in Human Cancer Cells, *Anal. Chem.*, 2022, **94**, 3343–3351.
- 19 P. H. Morgan, H. G. Jacobs, J. P. Segrest and L. W. Cunningham, A comparative study of glycopeptides derived from selected vertebrate collagens. A possible role of the carbohydrate in fibril formation, *J. Biol. Chem.*, 1970, **245**, 5042–5048.
- 20 P. P. Fietzek and F. W. Rexrodt, The covalent structure of collagen. The amino-acid sequence of alpha2-CB4 from calf-skin collagen, *Eur. J. Biochem.*, 1975, **59**, 113–118.
- 21 E. Klement, Z. Lipinszki, Z. Kupihar, A. Udvardy and K. F. Medzihradzky, Enrichment of O-GlcNAc modified proteins by the periodate oxidation-hydrazide resin capture approach, *J. Proteome Res.*, 2010, **9**, 2200–2206.
- 22 M. Stepath, B. Zulch, A. Maghnouj, K. Schork, M. Turewicz, M. Eisenacher, S. Hahn, B. Sitek and T. Bracht, Systematic Comparison of Label-Free, SILAC, and TMT Techniques to Study Early Adaption toward Inhibition of EGFR Signaling in the Colorectal Cancer Cell Line DiFi, *J. Proteome Res.*, 2020, **19**, 926–937.
- 23 Z. Li, R. M. Adams, K. Chourey, G. B. Hurst, R. L. Hettich and C. Pan, Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos, *J. Proteome Res.*, 2012, **11**, 1582–1590.
- 24 Y. Zhao and B. A. Garcia, Comprehensive Catalog of Currently Documented Histone Modifications, *Cold Spring Harbor Perspect. Biol.*, 2015, **7**, a025064.
- 25 W. Liu, Q. Ma, K. Wong, W. Li, K. Ohgi, J. Zhang, A. Aggarwal and M. G. Rosenfeld, Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release, *Cell*, 2013, **155**, 1581–1595.
- 26 F. Wang, L. He, P. Huangyang, J. Liang, W. Si, R. Yan, X. Han, S. Liu, B. Gui, W. Li, D. Miao, C. Jing, Z. Liu, F. Pei, L. Sun and Y. Shang, JMJD6 promotes colon carcinogenesis through negative regulation of p53 by hydroxylation, *PLoS Biol.*, 2014, **12**, e1001819.
- 27 O. Wagih, N. Sugiyama, Y. Ishihama and P. Beltrao, Uncovering Phosphorylation-Based Specificities through Functional Interaction Networks, *Mol. Cell. Proteomics*, 2016, **15**, 236–245.
- 28 M. Wong, Y. Sun, Z. Xi, G. Milazzo, R. C. Poulos, C. Bartenhagen, J. L. Bell, C. Mayoh, N. Ho, A. E. Tee, X. Chen, Y. Li, R. Ciaccio, P. Y. Liu, C. C. Jiang, Q. Lan, N. Jayatilke, B. B. Cheung, M. Haber, M. D. Norris, X. D. Zhang, G. M. Marshall, J. Y. Wang, S. Huttelmaier, M. Fischer, J. W. H. Wong, H. Xu, G. Perini, Q. Dong, R. E. George and T. Liu, JMJD6 is a tumorigenic factor and therapeutic target in neuroblastoma, *Nat. Commun.*, 2019, **10**, 3319.
- 29 J. Yang, S. Chen, Y. Yang, X. Ma, B. Shao, S. Yang, Y. Wei and X. Wei, Jumonji domain-containing protein 6 protein and its role in cancer, *Cell Proliferation*, 2020, **53**, e12747.
- 30 J. Cox and M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.*, 2008, **26**, 1367–1372.
- 31 S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann and J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nat. Methods*, 2016, **13**, 731–740.
- 32 G. E. Crooks, G. Hon, J. M. Chandonia and S. E. Brenner, WebLogo: a sequence logo generator, *Genome Res.*, 2004, **14**, 1188–1190.
- 33 Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi and B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, *Nucleic Acids Res.*, 2019, **47**, W199–W205.
- 34 D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen and C. von Mering, 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.
- 35 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, 2003, **13**, 2498–2504.
- 36 G. D. Bader and C. W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinf.*, 2003, **4**, 2.
- 37 A. Sturn, J. Quackenbush and Z. Trajanoski, Genesis: cluster analysis of microarray data, *Bioinformatics*, 2002, **18**, 207–208.
- 38 C. A. Schneider, W. S. Rasband and K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods*, 2012, **9**, 671–675.
- 39 Y. Perez-Riverol, J. Bai, C. Bandla, D. Garcia-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D. J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma and J. A. Vizcaino, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, *Nucleic Acids Res.*, 2022, **50**, D543–D552.

