Quantitative profiling of JMJD6-catalysed lysine hydroxylation reveals residue-dependent oxygen sensitivity

Pallavi Kesavan[1, 2*], Sophie M. Stuermer[3*], Katrin Räbel[1], Oliver Popp[1], Philipp Mertins[1], Matthew E. Cockman[3], Yoichiro Sugimoto[1, 4, 5]

[1]Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany

[2]Charité Universitätsmedizin Berlin, Berlin, Germany

[3]The Francis Crick Institute, London, UK

[4]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany

[5]Helmholtz-Institute for Translational AngioCardioScience (HI-TAC) of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) at Heidelberg University, Heidelberg, Germany

* These authors contributed equally to the work

Correspondence to:

Yoichiro Sugimoto, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany

Yoichiro.Sugimoto@mdc-berlin.de

Matthew E. Cockman, The Francis Crick Institute, London, UK

matthew.cockman@crick.ac.uk

Abstract

Lysine hydroxylation is increasingly recognised as a widespread post-translational modification in human cells, with more than 100 modified sites identified to date. Jumonji domain-containing protein 6 (JMJD6) catalyses hydroxylation at multiple lysines within lysine-rich regions and is a major contributor to this modification across the proteome. As JMJD6 requires oxygen as a co-substrate, lysine hydroxylation has been proposed to couple oxygen availability to cellular function. However, the biological significance of this modification remains incompletely understood, in part due to technical challenges associated with the detection of hydroxylation within lysine-rich regions by mass spectrometry. To address limitations of conventional approaches, we systematically evaluated key steps in the analysis of proteomic data from lysine-derivatised samples and developed a workflow for comprehensive, accurate, and quantitative analysis of lysine hydroxylation. Methodological improvements included optimisation of database search strategies to increase peptide coverage in lysine-rich regions and incorporation of immonium ion signatures to substantially improve confidence in hydroxylysine identification. We further demonstrated that stoichiometry derived from peptide precursor ion intensity faithfully captures hypoxia-responsive changes in lysine hydroxylation at amino acid resolution. Application of this workflow to bromodomain (BRD) proteins – epigenetic readers containing lysine-rich regions extensively hydroxylated by JMJD6 – revealed marked heterogeneity in the apparent kinetics of hydroxylation among target lysines, with evidence of interdependence between neighbouring sites. Hypoxia suppressed hydroxylation in a site-dependent manner, with greater suppression observed at sites displaying slower rates of hydroxylation. Together, the development and application of this workflow establish a methodological and biological framework for understanding how oxygen availability regulates protein function through lysine hydroxylation.

Introduction

Protein hydroxylation links cellular oxygen availability to protein function[1]. For instance, prolyl hydroxylase domain (PHD) enzymes hydroxylate hypoxia inducible factor α (HIFα), thereby controlling HIFα stability and mediating oxygen-dependent transcriptional reprogramming[2]. PHDs belong to the 2-oxoglutarate-dependent dioxygenase (2-OGDD) superfamily, whose members catalyse hydroxylation reactions using molecular oxygen and 2-oxoglutarate as co-substrates[3]. Whereas PHD-mediated hydroxylation is restricted to a small set of target sites[2,4], JMJD6 – another 2-OGDD – hydroxylates lysine residues at many more sites. JMJD6 catalyses hydroxylation at the C5 position of at least 150 lysines across 48 substrates, predominantly within lysine-rich regions[5], and a periodate-based proteomic approach that selectively targets C5-hydroxylysine suggests that this modification is even more widespread[6]. The breadth of JMJD6 targets raises the possibility that lysine hydroxylation influences multiple cellular processes according to oxygen availability, given that JMJD6 requires molecular oxygen as a co-substrate. However, despite many studies implicating JMJD6 in development and disease[7], the functional significance of lysine hydroxylation

remains incompletely understood, in part because lysine-rich regions are poorly detected by mass spectrometry, and hydroxylation within them is difficult to measure accurately.

Lysine-rich regions form regulatory interfaces for molecular interaction and post-translational modification. The positively charged lysine side chain can participate in electrostatic interactions and hydrogen bonding, and lysine residues are frequently enriched within intrinsically disordered regions that support multivalent interactions with RNA and other biomolecules. These properties contribute to the assembly of phase-separated molecular condensates[8]. Furthermore, in addition to hydroxylation, lysine-rich regions are subject to a wide range of post-translational modifications (PTMs)[9], including methylation, acetylation, and ubiquitination, enabling dynamic regulation of protein interactions and activity. For example, modification of lysine-rich histone tails plays a central role in transcriptional control[10]. JMJD6-catalysed hydroxylation modifies lysine at the C5 position of the side chain (Fig. 1a). Although the C5 hydroxyl group is distal to the ε-amine targeted by most other PTMs[7], it alters the local chemical environment and provides an additional hydrogen bond donor and acceptor. Consequently, JMJD6-mediated hydroxylation has the potential to modulate molecular interactions, influence condensate formation, and alter substrate recognition by lysine-modifying enzymes in an oxygen-dependent manner.

Bottom-up proteomics using liquid chromatography-tandem mass spectrometry (LC-MS/MS) is the primary method for studying protein hydroxylation[5,11]. In most workflows, proteins are digested with trypsin, which specifically cleaves polypeptide chains at the C-terminus of lysine and arginine residues, to generate peptides of suitable length and charge for efficient ionisation and identification. However, in lysine-rich regions, tryptic digestion produces peptides that are often too small to enable confident protein assignment. Although alternative proteases with different cleavage specificities could, in principle, address this limitation, they generally exhibit lower efficiency than trypsin and frequently generate peptides with suboptimal properties for LC-MS/MS analysis, such as high charge states[12]. Chemical derivatisation of lysine residues using propionic anhydride mitigates these issues by preventing tryptic cleavage at derivatised lysines and neutralising their charge[13,14] (Supplementary Fig. 1). Although this approach enables recovery of MS-amenable peptides from lysine-rich regions, it introduces analytical complexity through incomplete derivatisation and altered digestion patterns. Accurate detection of hydroxylation by LC-MS/MS is also intrinsically challenging. Spontaneous oxidation of residues such as methionine, together with difficulties in precise modification site localisation, can result in false-positive identifications[4,15]. To date, the performance of existing data analysis strategies for studying hydroxylation in lysine-rich regions has not been systematically evaluated.

To address these challenges, we refined key steps of the LC-MS/MS-based analysis of lysine-derivatised samples. Optimisation of database search settings increased the coverage of peptides in lysine-rich regions. In addition, filtering peptides based on diagnostic ions specific to hydroxylysine markedly improved the accuracy of hydroxylation site identification. We further demonstrated that hydroxylation stoichiometry, calculated from software-assigned peptide precursor ion intensity, captures hypoxia-responsive changes in lysine hydroxylation at amino acid resolution. To facilitate this analysis, we developed an R package,

ptm.stoichiometry, which enables direct calculation of PTM stoichiometry from the output of the widely used MS search engine MaxQuant[16].

We applied this workflow to characterise lysine hydroxylation in BRD epigenetic reader proteins, which are extensively hydroxylated by JMJD6[5]. The apparent kinetics of hydroxylation varied substantially between lysine residues, and modifications at neighbouring sites showed interdependence. JMJD6-dependent lysine hydroxylation decreased in a graded manner with increasing severity of hypoxia. At the site level, however, oxygen sensitivity differed markedly across sites, with those exhibiting slower apparent kinetics showing the greatest sensitivity. Collectively, our workflow enables proteome-scale, quantitative analysis of lysine hydroxylation in lysine-rich regions and reveals a principle of site-dependent oxygen sensitivity.

Results

Lysine-rich regions are widespread and linked to nucleic acid binding and biomolecular condensate formation

To characterise lysine-rich regions, we employed the previously defined K-score[5], a measure of local lysine enrichment calculated as the ratio of lysine residues within a sliding 10-amino-acid window. As a reference, histone tails – canonical lysine-rich regions[17] – exhibited a median K-score of 0.3 (Supplementary Fig. 2a). Regions with K-scores comparable to or higher than those of histone tails (K-scores ≥ 0.3) exhibited greater positive charge and were predicted to be intrinsically disordered (Supplementary Fig. 2b–d). Across the human proteome, 38% of proteins contained at least one such region (Fig. 1b), indicating that lysine-rich sequences are prevalent. Lysine-rich regions were enriched in proteins that bind nucleic acids and form RNA-associated biomolecular condensates[18], including Cajal bodies, nucleoli, and nuclear speckles (Fig. 1c). Their positive charge likely promotes nucleic acid binding, whereas intrinsic disorder facilitates multivalent interactions and condensate formation. These findings underscore both the prevalence and functional importance of lysine-rich regions and highlight the need for robust methods to analyse PTMs within them.

Optimisation of database search settings improves peptide recovery from lysine-rich regions

To facilitate comprehensive and robust analysis of lysine hydroxylation in lysine-rich regions, we systematically optimised key steps in the MS data analysis workflow for lysine-derivatised samples. Specifically, we refined the workflow to increase the recovery of peptides derived from lysine-rich regions and improve the accuracy of hydroxylysine identification. We subsequently evaluated the quantitative performance of hydroxylation stoichiometry derived from peptide MS intensities.

We first sought to define database search settings that maximise the recovery of peptides derived from lysine-rich regions. The high density of lysines and incomplete propionic anhydride derivatisation generate heterogeneous peptide populations that standard search

settings with fixed cleavage rules cannot model, motivating tailored search strategies for lysine-derivatised samples. Previous studies of histone modifications typically assumed complete derivatisation[13,14], thereby restricting tryptic cleavage specificity to arginine residues. In contrast, our earlier work on JMJD6-dependent hydroxylation assumed incomplete derivatisation and allowed up to five missed cleavages in tryptic digests[5]. Here, we compared these two strategies. To this end, we reanalysed a dataset from our previous study, in which JMJD6 target proteins were enriched, derivatised with propionic anhydride, digested with trypsin, and analysed by LC-MS/MS[5]. Because this dataset was enriched for lysine-rich proteins and hydroxylated lysines, it was well suited to workflow optimisation. We evaluated the effect of varying two *in silico* digestion parameters: (i) the number of allowed missed cleavages and (ii) whether cleavage was restricted to arginine. The following variable modifications were included in the search: methionine oxidation, N-terminal acetylation, lysine propionylation, lysine hydroxylation, and lysine hydroxylation with propionylation. Lysine propionylation can block tryptic cleavage, so each modified site contributes both a variable modification and a missed cleavage during database searching. Accordingly, the maximum number of variable modifications was matched to the maximum number of allowed missed cleavages per peptide. Database searches were performed using MaxQuant with a 1% FDR.

Allowing additional missed cleavages increased the number of peptide-spectrum matches (PSMs), with saturation at approximately seven missed cleavages (Fig. 2a). Conversely, restricting cleavage to arginine residues yielded the fewest PSMs, likely reflecting incomplete lysine derivatisation in highly lysine-rich regions. Notably, increasing the number of missed cleavages did not increase PSM counts in non-derivatised datasets, indicating that this effect is specific to derivatised samples. We therefore compared settings allowing two (standard workflows), five (previous study[5]), and seven (more permissive) missed cleavages. Increasing the number of permitted missed cleavages improved peptide identification within lysine-rich regions (Fig. 2b). Consistent with this global trend, peptide coverage within lysine-rich regions of BRD2–4 – previously shown to be extensively hydroxylated by JMJD6[5] – was enhanced under more permissive settings (Supplementary Fig. 3). Overall, allowing up to seven missed cleavages led to a higher PSM count in lysine-rich regions, consistent with efficient but incomplete lysine derivatisation.

Screening peptides with diagnostic ions improves the accuracy of hydroxylysine identification

We next evaluated a strategy to improve the accuracy of hydroxylysine identification from MS data analysed at 1% FDR. Specifically, the frequency of false-positive hydroxylation assignments in database search results was assessed using two criteria. First, we analysed the impact of JMJD6 inactivation on the number of identified hydroxylation sites. Since JMJD6 is the lysine hydroxylase with the broadest substrate repertoire[5], its inactivation should substantially reduce lysine hydroxylation levels. However, JMJD6 inactivation had only a modest effect on the number of computationally assigned lysine hydroxylation sites, many of which were detected in both JMJD6-inactivated and wild-type cells (Fig. 3a and Supplementary Fig. 4a). Second, methionine enrichment adjacent to reported hydroxylation sites was analysed. Sequences surrounding reported sites were significantly enriched for methionine (Fig. 3a left

panel, two-sided Fisher's exact test, $P < 10^{-10}$). The prevalence of artefactual methionine oxidation in proteomic datasets makes nearby methionine a strong indicator of spectral misassignment. Collectively, these findings suggest widespread misassignment of lysine hydroxylation, which is at least in part due to methionine oxidation.

We therefore explored the use of immonium ions generated during peptide fragmentation as diagnostic signatures for hydroxylysine-containing peptides. Lysine immonium ions, including stable cyclic species formed by neutral loss of ammonia, can serve as diagnostic markers for certain lysine PTMs[19–21]. Analysis of peptides from lysine-rich regions of BRD4 revealed distinct immonium ion signatures for hydroxylysine and propionylated hydroxylysine. These signatures were observed significantly more frequently in spectra from cells expressing JMJD6 than in JMJD6-inactivated cells, supporting their specificity (Fig. 3b and Supplementary Fig. 4b–d, two-sided Fisher's exact test, $P < 10^{-8}$ and $P < 10^{-10}$ for hydroxylysines and propionylated hydroxylysines, respectively).

We then incorporated these two immonium ions as diagnostic signatures into database searches by MaxQuant to support identification and scoring of peptides containing hydroxylysines. Of 1,902 lysines initially assigned as hydroxylated, 235 sites were identified by peptides with supporting diagnostic immonium ions (Supplementary Table). The immonium ion-supported subset was significantly enriched for previously validated hydroxylation sites compared with the unfiltered set (Fig. 3c, two-sided Fisher's exact test, $P < 10^{-10}$). The incomplete overlap between these sites and those identified in our previous work likely reflects differences in database search settings — including permitting a greater number of missed cleavages — as well as differences in database search algorithms (PEAKS[22] versus MaxQuant). In addition, immonium ion filtering may have excluded genuine sites, as hydroxylysine- and propionylated-hydroxylysine-derived immonium ions are generated with low probability (Fig. 3b).

Within the immonium ion-filtered lysine hydroxylation sites, inactivation of JMJD6 markedly reduced the number of assigned lysine hydroxylation events (143 versus 20 sites, Fig. 3a right panel), consistent with its role as a major lysine hydroxylase. The filtered set also lacked the methionine enrichment in surrounding sequences seen among unfiltered assignments, indicating fewer false-positive assignments.

Some newly identified lysine hydroxylation sites supported by peptides containing diagnostic immonium ions are adjacent to previously reported sites, including those in ARF-like GTPase 6-interacting protein 4 (ARL6IP4) and dyskerin pseudouridine synthase 1 (DKC1) (Supplementary Fig. 5). These sites reside in highly lysine-rich regions, and hydroxylation was detected only in cells with intact JMJD6. Therefore, these residues are likely bona fide JMJD6 targets. In contrast, in a small subset of the immonium ion-filtered set, lysine hydroxylation was identified in both JMJD6-expressing and JMJD6-inactivated cells; these included proteins with previously reported hydroxylysines, such as collagen type V[23], supporting their assignment as genuine JMJD6-independent sites. Overall, these results demonstrate that diagnostic immonium ions enable accurate hydroxylysine identification.

MS intensity-derived stoichiometry resolves dynamic changes in lysine hydroxylation

We next evaluated an approach to quantify lysine hydroxylation stoichiometry from MS data of lysine-derivatised samples. Reliable stoichiometry estimates are important for understanding the functional significance of lysine hydroxylation. We previously estimated stoichiometry using software-assigned peptide precursor ion intensity (referred to as peptide MS intensity) as a proxy for peptide abundance[5], but this approach can be confounded by stochastic peptide identification and variation in ionisation efficiency. We therefore evaluated the quantitative performance of this method in two contexts: comparison of hydroxylation at a given site across conditions, and comparison across sites and conditions.

We first assessed whether stoichiometry captures dynamic changes at individual sites. To this end, we performed a temporal analysis of lysine hydroxylation of BRD2–4 using JQ1-affinity enrichment in JMJD6-inducible HeLa cells, in which JMJD6 was inactivated by CRISPR/Cas9-mediated genome editing and re-expressed under the control of a doxycycline-inducible promoter. Analysis was restricted to sites supported by diagnostic ions and sites assigned in our earlier work[5]. Consistent with previous observations[5], negligible lysine hydroxylation was detected in the absence of JMJD6 induction. Upon re-expression, stoichiometry increased progressively with the duration of JMJD6 induction (Fig. 4a and b, Supplementary Fig. 6). Normalisation of stoichiometry values to those of wild-type HeLa cells at each lysine confirmed a time-dependent increase at individual sites. Importantly, differences between time points were resolvable up to 18 hours, after which hydroxylation patterns closely resembled those of wild-type HeLa cells (Fig. 4c and Supplementary Fig. 6). Accordingly, extended time points (18- and 24-hour induction) were combined for subsequent analyses. Together, these findings indicate that stoichiometry values enable quantitative assessment of lysine hydroxylation at single-residue resolution.

To evaluate whether stoichiometries derived from software-assigned peptide MS intensities support quantitative comparisons across sites and conditions, we compared values from derivatised samples using our pipeline with those obtained by manual integration of extracted ion chromatograms (XICs) of non-derivatised samples, a conventional approach for peptide quantification. This analysis was restricted to hydroxylation sites reliably identifiable in non-derivatised samples; to compensate, stoichiometries were compared across three induction time points and three oxygen levels to increase the number of data points (Supplementary Fig. 7a). The two approaches showed a strong correlation (Pearson's $r = 0.9$, $P < 1 \times 10^{-11}$; Fig. 4d and Supplementary Fig. 7b), supporting the use of derivatised MS intensity-based stoichiometry as a quantitative readout for comparative analysis across sites and conditions.

Site-specific apparent kinetics shape lysine hydroxylation stoichiometry

Having optimised analytical parameters and demonstrated quantitative performance, we applied the workflow to investigate the apparent kinetics and oxygen sensitivity of lysine hydroxylation in BRD2–4. These proteins were selected because they are epigenetic readers that play key roles in physiology and pathology[24] and contain highly lysine-rich regions, in which many lysines are hydroxylated by JMJD6[5].

A characteristic feature of JMJD6-mediated lysine hydroxylation is that modification often occurs at multiple lysines within a lysine-rich region, with variable stoichiometry. To

investigate site-dependent differences in stoichiometry, we analysed the temporal dynamics of lysine hydroxylation at individual sites. Specifically, we estimated the time required to reach half-maximal stoichiometry, hereafter referred to as t_{50}, following JMJD6 induction under normoxia, using the time course data described above and isotonic regression. We note that t_{50} is an apparent, rather than intrinsic, kinetic parameter for two reasons: (i) doxycycline treatment precedes JMJD6 protein induction by an appreciable lag and (ii) hydroxylation at different lysine residues may be interdependent. Nevertheless, the metric provides a useful basis for analysing site-specific kinetic effects in cells. The t_{50} varied substantially across sites. Sites with lower t_{50} reached higher steady-state stoichiometry than those with higher t_{50}, indicating that faster apparent kinetics contribute to higher hydroxylation stoichiometry (Fig. 5a and Supplementary Fig. 6).

Because JMJD6 often hydroxylates multiple lysines within lysine-rich regions, we next investigated whether hydroxylation at neighbouring sites is interdependent. We analysed robustly quantified peptides containing pairs of hydroxylation sites assigned to different t50 categories, and assessed whether hydroxylation at one site was associated with altered hydroxylation stoichiometry at the neighbouring site. Higher-t_{50} sites showed significantly higher stoichiometry when quantified from peptides in which neighbouring lower-t_{50} sites were hydroxylated than when they were unmodified (Supplementary Fig. 8; paired two-sided t-test, $P < 10^{-5}$). This pattern suggests coordinated hydroxylation between neighbouring lysine residues.

We further validated this relationship using a representative pair of sites, K535 and K537 of BRD4, for which hydroxylation was detected in LC-MS/MS data from non-derivatised samples. Assuming independence between sites, the expected abundance of the doubly hydroxylated peptide was calculated from the product of the abundances of the singly hydroxylated forms, scaled by the abundance of the unmodified peptide. Comparison with the observed XIC peak area revealed that the doubly hydroxylated species was substantially enriched relative to this expectation, with an observed-to-expected ratio of 6.3 (Fig. 5b). This marked deviation indicates that hydroxylation at these sites does not occur independently. Together, these analyses demonstrate interdependence between hydroxylation events at neighbouring lysine residues.

Hypoxia represses lysine hydroxylation in a site-dependent manner

JMJD6 requires oxygen for catalysis[5,25,26], raising the possibility that JMJD6-mediated lysine hydroxylation is sensitive to cellular oxygen availability. Consistent with this, we and others have shown, using distinct peptides that share residues 535–541 of the BRD4 lysine-rich region, that hydroxylation within this region is reduced under hypoxia both in cells[5] and *in vitro*[26]. However, these analyses were restricted to individual peptides and reported only the total number of hydroxylation events per peptide, leaving the site-specific and protein-wide effects of hypoxia unresolved.

To pursue this further, we quantified hydroxylation stoichiometries in JMJD6-inducible HeLa cells subjected to concurrent JMJD6 induction and different oxygen concentrations for 18–24 hours. Lysine hydroxylation decreased progressively with increasing hypoxic severity (Fig. 5c

and d). JMJD6 abundance moderately increased under hypoxia, whereas BRD4 abundance was unaffected, suggesting that reduced hydroxylation reflects inhibition of JMJD6 catalytic activity rather than reduced enzyme or substrate abundance (Supplementary Fig. 10). Notably, inhibition of hydroxylation was evident at 4% O_2 (Fig. 5d), indicating that JMJD6-catalysed lysine hydroxylation is sensitive to physiologically relevant changes in oxygen availability.

Oxygen sensitivity varied substantially across sites and was strongly associated with apparent hydroxylation kinetics. Fast sites (t_{50} = 0–4 hours) were repressed only under severe hypoxia (0.1% O_2), while they remained largely unaffected at more moderate hypoxia (1% and 4% O_2). In contrast, intermediate sites (t_{50} = 4–8 hours) showed graded repression across 0.1–4% O_2. Slow sites (t_{50} > 8 hours) were repressed at all tested hypoxic conditions (Fig. 5e). Thus, hypoxia inhibits lysine hydroxylation in a site-dependent manner, with the greatest sensitivity observed at sites with slower apparent kinetics.

Interestingly, hydroxylation patterns observed under hypoxia resembled those seen at earlier time points of JMJD6 induction under normoxia. The stoichiometry profile after 18–24 hours of JMJD6 induction at 0.1% O_2 most closely resembled that observed after 4 hours of induction at normoxia or 4% O_2, and that after 8 hours at more severe hypoxia of 1% O_2 (Fig. 5f, Supplementary Fig. 11). These findings are consistent with the notion that hypoxia represses lysine hydroxylation by effectively slowing the apparent kinetics of JMJD6 activity according to the severity of hypoxia.

Discussion

Lysine-rich regions present a persistent challenge for mass spectrometry-based proteomics. Indeed, this difficulty may explain why the extensive substrate repertoire of JMJD6 has only recently become apparent[5,6]. Here, we systematically refined key steps in the analysis of LC-MS/MS data from lysine-derivatised samples to improve detection and quantification of hydroxylation within these regions. By allowing multiple missed cleavages and variable modifications to account for incomplete lysine derivatisation by propionic anhydride, we increased peptide recovery from lysine-rich regions. Requiring immonium ion support improved confidence in lysine hydroxylation site assignments. Importantly, we demonstrated that stoichiometry estimates derived from software-assigned peptide MS intensities captured condition-dependent changes in lysine hydroxylation. Together, these advances establish a framework for comprehensive and quantitative analysis of lysine hydroxylation within lysine-rich regions.

JMJD6-catalysed lysine hydroxylation is widespread and frequently occurs at multiple lysines within target regions, as observed in BRD proteins[5]. Analysis of these proteins revealed that the temporal rate at which hydroxylation accumulates varies markedly across sites. These differences may reflect local sequence context and interdependence between neighbouring residues. Consistent with this possibility, we found that hydroxylation at neighbouring lysines does not occur independently. Such interdependence may arise from altered JMJD6 affinity after initial hydroxylation, or from local conformational changes induced by lysine

hydroxylation. Importantly, oxygen sensitivity also varied significantly across hydroxylation sites. Sites with slower apparent kinetics were more strongly repressed by hypoxia than sites with faster kinetics, and hydroxylation patterns observed under hypoxia resembled those seen following short JMJD6 induction under normoxia. Collectively, these findings suggest that reduced oxygen availability slows the apparent kinetics of JMJD6-catalysed lysine hydroxylation, resulting in site-dependent repression of modification in proportion to the severity of hypoxia.

Site-dependent oxygen sensitivity has important biological implications. Individual JMJD6 target sites have distinct threshold oxygen concentrations at which hydroxylation is inhibited. Given the widespread occurrence of JMJD6-catalysed lysine hydroxylation across the human proteome, such site-dependent oxygen sensitivity may permit progressive responses to increasing hypoxic severity. Within individual substrates, proteins such as BRD4 contain multiple hydroxylation sites with distinct oxygen sensitivities, suggesting that hydroxylation-dependent functions of these proteins are regulated in a graded, rather than binary, manner as oxygen availability decreases. More broadly, many JMJD6 substrates function within related RNA-processing pathways, including the SR proteins involved in pre-mRNA splicing[5,25]. Differences in oxygen sensitivity across substrates could therefore enable fine-tuning of splicing activity in response to hypoxia. Finally, JMJD6 is overexpressed in several cancers, including neuroblastoma[27–29], raising the possibility that increased JMJD6 abundance may alter apparent hydroxylation kinetics and thereby modulate the oxygen sensitivity of target proteins in pathophysiological contexts.

Although the workflow developed here facilitates accurate identification and quantitative analysis of lysine hydroxylation, several limitations should be considered. While diagnostic immonium ions provide critical support for the presence of hydroxylation, particularly in lower-quality MS/MS data where backbone fragmentation is sparse, they cannot provide positional information. Furthermore, these ions may be confounded by isobaric species; for example, propionylated hydroxylysines and lactyllysine exhibit identical mass shifts relative to unmodified lysine[9]. In lysine-rich peptides, co-eluting hydroxylated positional isomers present a significant analytical challenge. Co-fragmentation of peptides hydroxylated at different lysine residues can produce chimeric spectra, where overlapping site-determining ions may support conflicting localisation assignments and lead to erroneous hydroxylation-site assignment. In addition, although stoichiometry estimates derived from peptide MS intensities are suitable for comparative analyses, absolute quantification requires synthetic peptide standards[30]. However, the use of peptide standards is difficult to scale for proteome-wide analysis, particularly because hydroxylysine-containing peptides are not readily available.

Our biological analyses focused on the bromodomain proteins, which contain well-defined JMJD6-targeted lysine-rich regions. Although the principles identified here – including site-dependent apparent hydroxylation kinetics, differential oxygen sensitivity, and inter-site relationships – are likely to apply to other JMJD6 targets, future studies extending this work across additional substrates and cellular contexts will be needed to establish their generality.

Taken together, this work establishes a data analysis workflow for efficient, accurate, and quantitative assessment of hydroxylation in lysine-rich regions at amino acid resolution, and identifies site-dependent oxygen sensitivity as a defining feature of JMJD6-catalysed lysine hydroxylation. These findings provide a framework for understanding how graded site-resolved lysine hydroxylation may contribute to cellular responses to hypoxia.

Methods

Data availability

LC-MS/MS data generated during this study are available from PRIDE under accession ID: PXD079306, PXD079391, and PXD078942.

Code availability

The R package developed during this study and the computational pipeline used for the data analysis are available on GitHub at https://github.com/YoichiroSugimoto/ptm.stoichiometry and https://github.com/YoichiroSugimoto/20241111_PTMs_in_lysine_rich_domains, respectively.

Cell lines

JMJD6-inducible HeLa cells, in which JMJD6 was inactivated and re-expressed from a doxycycline-inducible promoter, were generated previously[5]. Cell identity was confirmed by STR profiling, and cells were confirmed to be free of mycoplasma contamination. HeLa cells were maintained in DMEM (Thermo Fisher Scientific, no. 31966047) supplemented with 10% tetracycline-free FBS (Pan Biotech, no. P30-3602) at 37 °C in 5% CO_2. JMJD6-inducible HeLa cells were maintained under the same conditions with medium containing 1 µg/mL puromycin (InvivoGen, no. ant-pr-1). JMJD6 expression was induced by treatment with 400–600 ng/mL doxycycline (Sigma-Aldrich, no. D9891-1G). Hypoxic incubation was performed using a Whitley H35 HEPA Hypoxystation or Baker Ruskinn InvivO$_2$ 500 at indicated oxygen levels, 5% CO_2, and 37 °C.

Antibodies

The following antibodies were used for immunoblotting analysis: anti-BRD4 (Abcam, ab128874) and anti-JMJD6 (Santa Cruz Biotechnology, sc-28348).

Affinity purification of BRD proteins and sample preparation for MS analysis

Samples of BRD proteins for LC-MS/MS analysis were prepared as previously described[11], with minor modifications. The procedure involves the following steps:

Preparation of JQ1-coated beads. 10 µL of Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific, no. 65001) was washed twice with native IP buffer (20 mM Tris-HCl, pH 7.5; 100 mM NaCl; 0.5% Igepal CA-630; 5 µM $MgCl_2$). The beads were resuspended in 100 µL of the same buffer. 100 pmol of JQ1-biotin (MedChemExpress, HY-145667) was added, and the samples were incubated at 4 °C for 1 hour with rotation to allow coupling. The beads were

washed twice with native IP buffer to remove unbound JQ1-biotin and resuspended in 10 µL of native IP buffer.

Pulldown of BRD proteins. Cells were grown on 15-cm dishes, washed twice with ice-cold PBS, and lysed with 1 mL of RIPA buffer (Serva Electrophoresis, no. 3924402) supplemented with HALT protease inhibitor (Thermo Fisher Scientific, no. 78445) and 1 mM DMOG (Sigma-Aldrich, no. 40001-50MG). Genomic DNA was digested by adding 1:1,000 (v/v) benzonase (Sigma-Aldrich, no. E1014-25KU) to the lysate. The lysate was clarified by centrifugation at 14,000*g* for 5 min at 4 °C. JQ1-coated beads (10 µL) were added to 1 mL of lysate containing 5 mg total protein, and the mixture was incubated with rotation for 2 h at 4 °C. The beads were washed three times with RIPA buffer. JQ1-bound proteins were eluted and subjected to SDS-PAGE. Gel regions containing BRD2–4 were excised and cut into small pieces. In a subset of experiments, cells were treated with 500 µM DMOG prior to harvesting instead of adding DMOG to the lysis buffer. Differences in conditions between datasets are summarised in Supplementary Table.

Lysine derivatisation. Gel pieces were washed overnight at 4 °C in 50% methanol and 5% acetic acid, then washed for an additional 5 min at room temperature. Gel pieces were dehydrated with acetonitrile, rehydrated in 80 µL of 100 mM ammonium bicarbonate, and treated with 20 µL of 25% propionic anhydride (Sigma-Aldrich, no. 240311-50G) in acetonitrile. The reaction pH was immediately adjusted to 8 by addition of 8 µL of ammonium hydroxide (Sigma-Aldrich, no. 5438300250). Samples were incubated at 51 °C for 20 minutes with shaking at 2,000 rpm. The supernatant was removed, the gel pieces were dehydrated with acetonitrile, and the propionylation reaction was repeated once.

Trypsinolysis. Gel pieces were rehydrated in 100 µL of 2.5 ng/µL Trypsin Platinum (Promega, no. VA9000) in 50 mM ammonium bicarbonate, and samples were incubated at 37 °C overnight with shaking at 1,200 rpm. Tryptic peptides were collected from residual digest buffer and then extracted sequentially with 50% and 80% acetonitrile, each containing 5% formic acid. Extracted peptides were purified using Evotip Pure tips (Evosep Biosystems, no. EV2013) according to the manufacturer's instructions.

Mass spectrometry

LC-MS/MS analyses were performed using two acquisition workflows. For experiments conducted at the Max Delbrück Center for Molecular Medicine, peptides were separated using an EASY-nLC 1200 system (Thermo Fisher Scientific) coupled online to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific). For each run, 2 µL of peptide solution in solvent A (3% acetonitrile, 0.1% formic acid) was injected. Samples were loaded onto a reversed-phase column at 1 µL/min and eluted at a constant flow rate of 250 nL/min using a 110 min gradient from 2% to 90% solvent B (90% acetonitrile, 0.1% formic acid). Full gradient profile: 2%–20% B over 67 min, 20%–30% B over 20 min, 30%–60% B over 10 min, ramping to 90% B for column washing, followed by re-equilibration.

The mass spectrometer was operated in data-dependent acquisition (DDA) mode. Full MS1 scans were acquired at 60,000 resolution (m/z 350–1800), followed by higher-energy

collisional dissociation (HCD) fragmentation of the top precursors with a cycle time of 1 s. MS2 scans were acquired at 15,000 resolution, with an isolation window of 1.3 m/z, and normalised collision energy of 28%. Dynamic exclusion was set to 20 s. Only precursor ions with charge states between 2+ and 6+ and intensities above 50,000 were selected for fragmentation. The maximum injection time for MS2 was 100 ms.

For experiments conducted at The Francis Crick Institute, samples were prepared and analysed by LC-MS/MS on an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific), using methods described previously[5]. The workflow used for each dataset is summarised in Supplementary Table.

Overview of data analyses

Data analyses were performed in R (4.5.1), using the following packages: data.table (1.17.8), dplyr (1.1.4), stringr (1.5.2) and ggplot2 (4.0.0). Reference human protein sequences were downloaded from UniProt (UP000005640_9606). K-score, predicted disorderedness (using IUPred2A[31]), and local charge were calculated in the previous study[5]. In addition to data generated in this study, MS data from HeLa cells reported by Cockman *et al*.[5] (PRIDE ID: PXD031221) were analysed.

Analysis of the association between functional protein classes and maximum K-score

A generalised additive model was used to assess the association between functional protein classes and K-score. A model using thin plate regression spline was constructed to predict the maximum K-score of a protein based on its functional class and length. The gam function of the mgcv package (1.9-1) was used for implementation. The coefficients for protein functional classes were then analysed.

Diagnostic ion identification

To identify immonium-derived diagnostic ions of hydroxylysine, public MS data[5] for the proteins extracted from HeLa cells with or without JMJD6 inactivation and pulled down using JQ1-coated beads were analysed. FragPipe (22.0)[32], MSFragger (4.1), IonQuant (1.10.27), diaTracer (1.1.5), and Philosopher (5.1.1) were used with the following parameters: search_enzyme_cut_1 = K, allowed_missed_cleavage_1 = 5, search_enzyme_cut_2 = R, allowed_missed_cleavage_2 = 2, variable_mod_01 = 15.9949 M 3, variable_mod_02 = 42.0106, variable_mod_10 = 15.9949 K 5, variable_mod_11 = 56.026215 K 5, variable_mod_12 = 72.021126 K 5, and max_variable_mods_per_peptide = 5.

Diagnostic ions for hydroxylysine (100.0762 Da) and propionylated hydroxylysine (156.1025 Da) were confirmed in peptides covering the lysine-rich region of BRD4 (residues 535–555), and subsequently used to configure diagnostic ion-aware database searches in MaxQuant.

Database search

Database searches were performed using MaxQuant (2.6.2.0). The following parameters were used: fixed modifications: carbamidomethylation (C); variable modifications: Oxidation (M), Acetylation (Protein N-Terminus), Propionylation (C_3H_4O at K), Oxidation (O at K) and

Oxidised Propionylation ($C_3H_4O_2$ at K); diagnostic ions: Oxidation at K (C_5H_9ON) and Oxidised propionylation ($C_8H_{13}O_2N$). Unless otherwise specified, database searches allowed up to seven missed cleavages and seven variable modifications per peptide.

To improve analysis speed, the initial search considered only oxidation (M), acetylation (protein N-term), and propionylation as variable modifications; the main search included all modifications listed above, with hydroxylysine and propionylated-hydroxylysine immonium ions as diagnostic signatures.

Lysine hydroxylation sites with diagnostic ion support

Sites identified by diagnostic ion-supported peptides across all analysed datasets were pooled to generate a consensus list of hydroxylation sites, which was used to define diagnostic ion-supported hydroxylation sites throughout this study.

Calculation of the stoichiometry of PTMs

Stoichiometry was calculated as described previously[5], using the following formula:

$$\frac{\sum_{\text{All the peptides with the PTM at the site}} Intensity}{\sum_{\text{All the peptides with the PTM at the site}} Intensity + \sum_{\text{All the peptides without the PTM at the site}} Intensity}$$

For comparisons between conditions, when multiple datasets were available for the same condition, the dataset with the highest coverage was selected as representative. Hydroxylation stoichiometry was calculated for lysines whose hydroxylation was identified by peptides with diagnostic ions and/or by manual inspection of spectra in our previous work. Only sites showing hydroxylation in genetically unmodified HeLa cells were analysed.

Calculation of the time required to reach half-maximal stoichiometry (t_{50}) upon JMJD6 induction

The time required to reach half-maximal stoichiometry (t_{50}) was calculated using isotonic regression. Data from HeLa cells and JMJD6-inducible HeLa cells after 0, 4, 8, 18, and 24 hours of JMJD6 induction were used. For isotonic regression, genetically unmodified HeLa cells were treated as representing prolonged JMJD6 induction (100 hours), as the model requires finite numeric values.

Author contributions

M.E.C. and Y.S. conceived the project. S.M.S., K.R., and O.P. performed experiments. P.K., S.M.S., M.E.C., and Y.S. analysed the data. P.K., S.M.S., O.P., P.M., M.E.C., and Y.S. contributed to the interpretation of the data. P.K., S.M.S., M.E.C., and Y.S. wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary Table

Overview of the datasets and the stoichiometry of lysine hydroxylation calculated from each dataset.

Appendix

Representative MS/MS spectra of non-derivatised peptides quantified by XIC analysis.

Acknowledgements

We thank Sugimoto group members and Peter Ratcliffe (Francis Crick Institute) for support and discussion. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (CC2092), the UK Medical Research Council (CC2092), and the Wellcome Trust (CC2092). We are grateful to Helen Flynn and Mark Skehel from the Proteomics STP at The Francis Crick Institute for their valued contributions to the work. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. Markolovic, S., Wilkins, S. E. & Schofield, C. J. Protein Hydroxylation Catalyzed by 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20712–20722 (2015) doi: 10.1074/jbc.R115.662627.

2. Ratcliffe, P. J. Oxygen sensing and hypoxia signalling pathways in animals: the implications of physiology for cancer. *J. Physiol.* **591**, 2027–2042 (2013) doi: 10.1113/jphysiol.2013.251470.

3. Losman, J.-A., Koivunen, P. & Kaelin, W. G. 2-Oxoglutarate-dependent dioxygenases in cancer. *Nat. Rev. Cancer* **20**, 710–726 (2020) doi: 10.1038/s41568-020-00303-3.

4. Cockman, M. E. *et al.* Lack of activity of recombinant HIF prolyl hydroxylases (PHDs) on reported non-HIF substrates. *eLife* **8**, e46490 (2019) doi: 10.7554/eLife.46490.

5. Cockman, M. E. *et al.* Widespread hydroxylation of unstructured lysine-rich protein domains by JMJD6. *Proc. Natl. Acad. Sci.* **119**, e2201483119 (2022) doi: 10.1073/pnas.2201483119.

6. Sin, Y.-C., Park, M., Griffin, T. J., Yong, J. & Chen, Y. A constitutional isomer selective chemical proteomic strategy for system-wide profiling of protein lysine 5-hydroxylation. *Chem. Sci.* 10.1039.D4SC05397D (2024) doi: 10.1039/D4SC05397D.

7. Böttger, A., Islam, Md. S., Chowdhury, R., Schofield, C. J. & Wolf, A. The oxygenase Jmjd6–a case study in conflicting assignments. *Biochem. J.* **468**, 191–202 (2015) doi: 10.1042/BJ20150278.

8. Ukmar-Godec, T. *et al.* Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat. Commun.* **10**, 2909 (2019) doi: 10.1038/s41467-019-10792-y.

9. Wang, Z. A. & Cole, P. A. The Chemical Biology of Reversible Lysine Post-translational Modifications. *Cell Chem. Biol.* **27**, 953–969 (2020) doi: 10.1016/j.chembiol.2020.07.002.

10. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293**, 1074–1080 (2001) doi: 10.1126/science.1063127.

11. Cockman, M. E., Webb, J. D., Kramer, H. B., Kessler, B. M. & Ratcliffe, P. J. Proteomics-based Identification of Novel Factor Inhibiting Hypoxia-inducible Factor (FIH) Substrates Indicates Widespread Asparaginyl Hydroxylation of Ankyrin Repeat Domain-containing Proteins. *Mol. Cell. Proteomics* **8**, 535–546 (2009) doi: 10.1074/mcp.M800340-MCP200.

12. Dau, T., Bartolomucci, G. & Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Anal. Chem.* **92**, 9523–9527 (2020) doi: 10.1021/acs.analchem.0c00478.

13. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* **2**, 933–938 (2007) doi: 10.1038/nprot.2007.106.

14. Sidoli, S. & Garcia, B. A. Characterization of Individual Histone Posttranslational Modifications and Their Combinatorial Patterns by Mass Spectrometry-Based Proteomics Strategies. in *Histones* (eds Guillemette, B. & Gaudreau, L. R.) vol. 1528 121–148 (Springer New York, 2017). doi: 10.1007/978-1-4939-6630-1_8.

15. Jiang, H. *et al.* Systematic characterization of site-specific proline hydroxylation using hydrophilic interaction chromatography and mass spectrometry. Preprint at https://doi.org/10.1101/2023.07.28.550951 (2023) doi: 10.1101/2023.07.28.550951.

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

17. Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to chromatin through histone modifications. *Cell* **103**, 263–271 (2000) doi: 10.1016/s0092-8674(00)00118-5.

18. Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 196–213 (2021) doi: 10.1038/s41580-020-00326-6.

19. Couttas, T. A., Raftery, M. J., Bernardini, G. & Wilkins, M. R. Immonium Ion Scanning for the Discovery of Post-Translational Modifications and Its Application to Histones. *J. Proteome Res.* **7**, 2632–2641 (2008) doi: 10.1021/pr700644t.

20. Zolg, D. P. *et al.* ProteomeTools: Systematic Characterization of 21 Post-translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Using Synthetic Peptides. *Mol. Cell. Proteomics* **17**, 1850–1863 (2018) doi: 10.1074/mcp.TIR118.000783.

21. Wan, N. *et al.* Cyclic immonium ion of lactyllysine reveals widespread lactylation in the human proteome. *Nat. Methods* **19**, 854–864 (2022) doi: 10.1038/s41592-022-01523-1.

22. Zhang, J. *et al.* PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012) doi: 10.1074/mcp.M111.010587.

23. Ishikawa, Y. *et al.* Type I and type V procollagen triple helix uses different subsets of the molecular ensemble for lysine posttranslational modifications in the rER. *J. Biol. Chem.* **296**, 100453 (2021) doi: 10.1016/j.jbc.2021.100453.

24. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, (2018) doi: 10.1126/science.aar3958.

25. Webby, C. J. *et al.* Jmjd6 Catalyses Lysyl-Hydroxylation of U2AF65, a Protein Associated with RNA Splicing. *Science* **325**, 90–93 (2009) doi: 10.1126/science.1175865.

26. Corner, T. P., Salah, E., Tumber, A., Brewitz, L. & Schofield, C. J. Biochemical investigations using mass spectrometry to monitor JMJD6-catalysed hydroxylation of multi-lysine containing bromodomain-derived substrates. *RSC Chem. Biol.* **6**, 642–656 (2025) doi: 10.1039/D4CB00311J.

27. Yang, J. *et al.* Jumonji domain-containing protein 6 protein and its role in cancer. *Cell Prolif.* **53**, e12747 (2020) doi: 10.1111/cpr.12747.

28. Wong, M. *et al.* JMJD6 is a tumorigenic factor and therapeutic target in neuroblastoma. *Nat. Commun.* **10**, 3319 (2019) doi: 10.1038/s41467-019-11132-w.

29. Jablonowski, C. *et al.* Metabolic reprogramming of cancer cells by JMJD6-mediated pre-mRNA splicing is associated with therapeutic response to splicing inhibitor. Preprint at https://doi.org/10.7554/eLife.90993.2 (2024) doi: 10.7554/eLife.90993.2.

30. Zhang, G. *et al.* Protein Quantitation Using Mass Spectrometry. in *Computational Biology* (ed. Fenyö, D.) vol. 673 211–222 (Humana Press, 2010). doi: 10.1007/978-1-60761-842-3_13.

31. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018) doi: 10.1093/nar/gky384.

32. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nat. Methods* **14**, 513–520 (2017) doi: 10.1038/nmeth.4256.

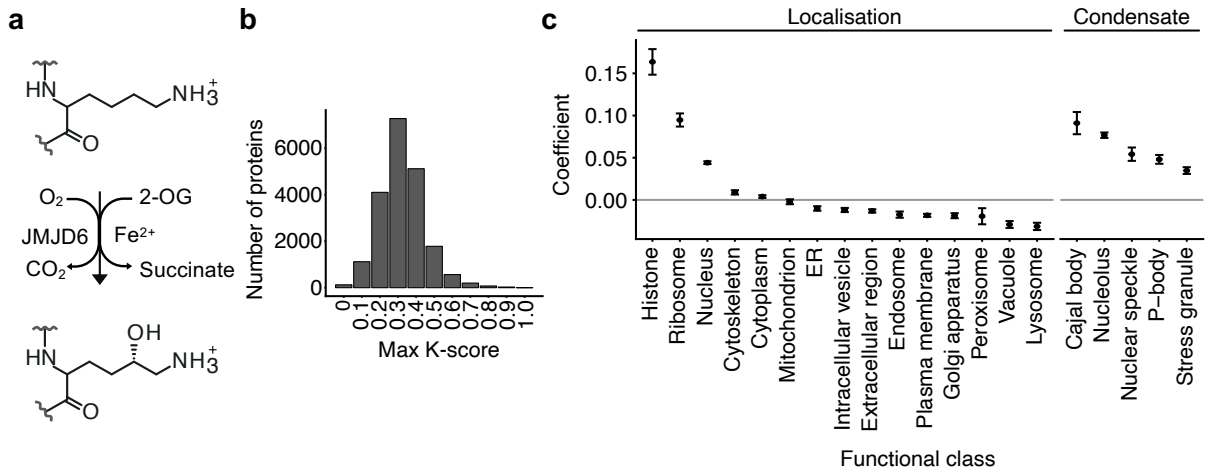


Fig. 1 | Lysine-rich regions are widespread and associated with molecular condensates.

a. Schematic of lysine C5 hydroxylation catalysed by JMJD6. **b.** Number of human proteins as a function of maximum K-score. **c.** Enrichment of lysine-rich regions among protein functional classes, presented as coefficients from a generalised additive model predicting K-score (mean ± s.e.m.). The model accounts for the association between protein length and K-score (see Supplementary Fig. 2e).

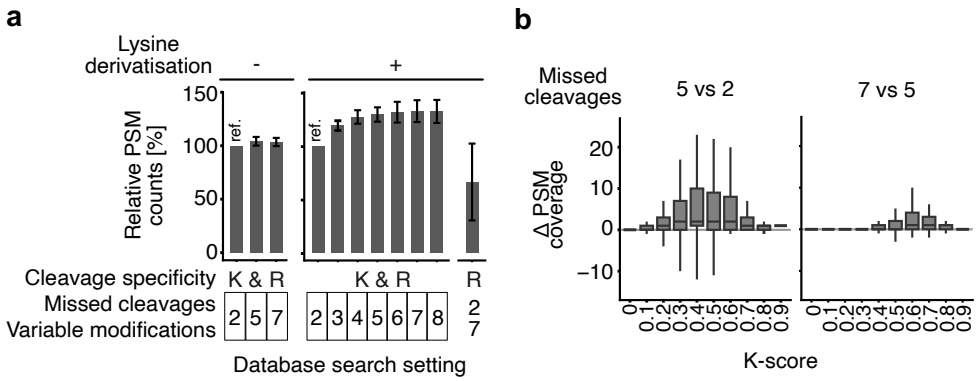


Fig. 2 | Optimisation of database search settings improves peptide recovery from lysine-rich regions.

a. Bar charts showing the effect of varying the number of allowed missed cleavages on peptide-spectrum match (PSM) counts. Data are normalised to the reference setting (ref.) defined as: cleavage at lysine (K) and arginine (R), two missed cleavages, and two variable modifications. Bars show mean ± s.d.. Left and right panels show data from BRD protein-enriched samples — prepared by immunoprecipitation or JQ1 enrichment — without or with lysine derivatisation, respectively. **b.** Box plots showing differences in per-residue PSM coverage (Δ PSM coverage) stratified by K-score, for searches allowing 5 versus 2, and 7 versus 5 missed cleavages. The maximum number of allowed missed cleavages was matched to the maximum number of variable modifications allowed per peptide.

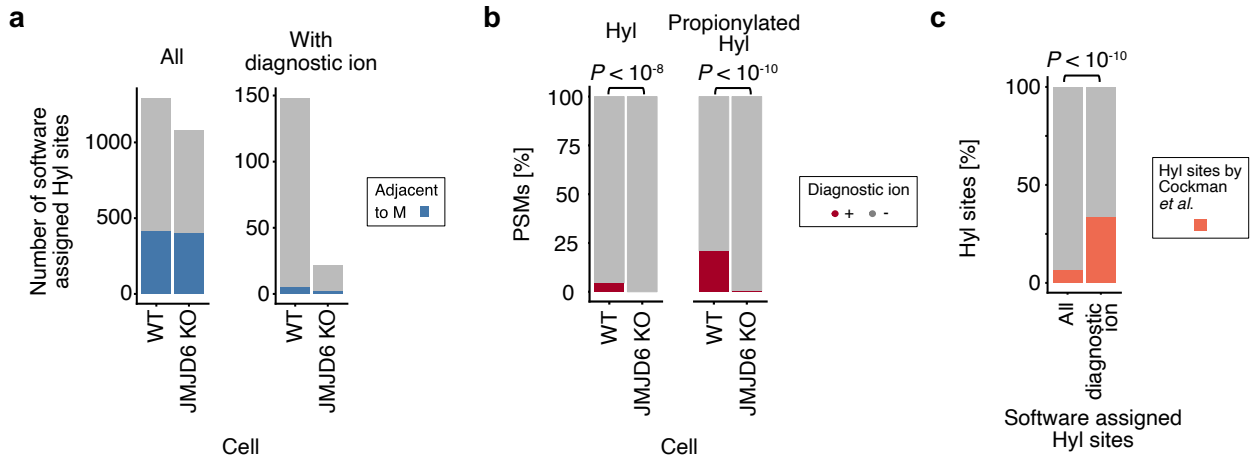


Fig. 3 | Diagnostic ions improve the accuracy of hydroxylysine identification.

a. Stacked bar charts showing the number of lysine residues identified as hydroxylated in wild-type HeLa cells (WT) or JMJD6 knockout HeLa cells (JMJD6 KO); blue indicates the subset of assignments in which a methionine residue lies within 2 residues of the target lysine. The left panel shows all hydroxylysine (Hyl) sites reported by the software; the right panel shows sites supported by diagnostic immonium ions for hydroxylysine or propionylated hydroxylysine. Only sites identified by at least one PSM in both WT and JMJD6 KO datasets were analysed. **b.** Stacked bar charts showing the proportion of PSMs containing diagnostic ions (red) among peptides spanning the lysine-rich region of BRD4 (residues 535–555) in WT and JMJD6 KO HeLa datasets. The left and right panels show data for the diagnostic ion of hydroxylysine (Hyl) and propionylated hydroxylysine (Propionylated Hyl), respectively. Proportions were compared using a two-sided Fisher's exact test. n = 752 and 586 PSMs for WT and JMJD6 KO, respectively. **c.** Stacked bar charts showing the proportion of software-reported lysine hydroxylation sites previously identified by manual inspection of spectra in our previous work[5]. All software-reported sites were compared with sites supported by diagnostic ions using a two-sided Fisher's exact test. n = 1,902 and 235, respectively.

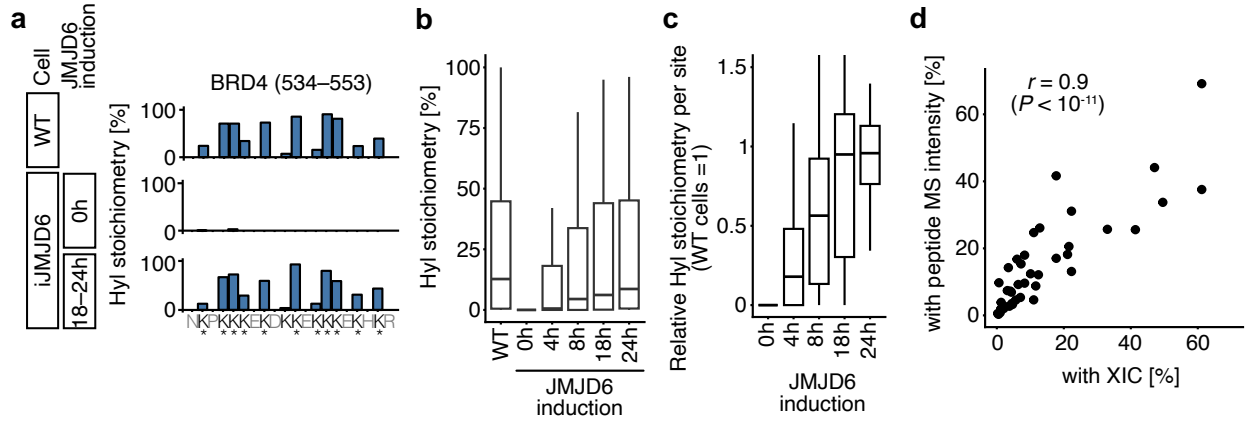


Fig. 4 | Peptide MS intensity-derived stoichiometry captures residue-dependent changes in lysine hydroxylation.

a. Stoichiometry of hydroxylysine (Hyl) in the lysine-rich region of BRD4. Data for wild-type HeLa cells (WT) are compared with those for JMJD6-inducible (iJMJD6) HeLa cells with or without 18–24h re-expression of JMJD6. Asterisks indicate lysine residues identified as hydroxylated by peptides with diagnostic ions and/or manual inspection of spectra in our previous work[5]. **b.** Changes in Hyl stoichiometry upon re-expression of JMJD6 in JMJD6-inducible HeLa cells for the indicated duration. Data for wild-type HeLa cells are also shown. **c.** Similar to **b**, but showing Hyl stoichiometry per site, normalised to that of wild-type HeLa cells. **d.** Scatter plot comparing BRD2–4 hydroxylysine stoichiometry calculated from XICs of non-derivatised samples with stoichiometry calculated from MS peptide intensities of derivatised samples. Each data point represents one site–oxygen level–time point combination (n = 48; see Supplementary Fig. 7b for individual data points). **b** and **c**: n = 42.

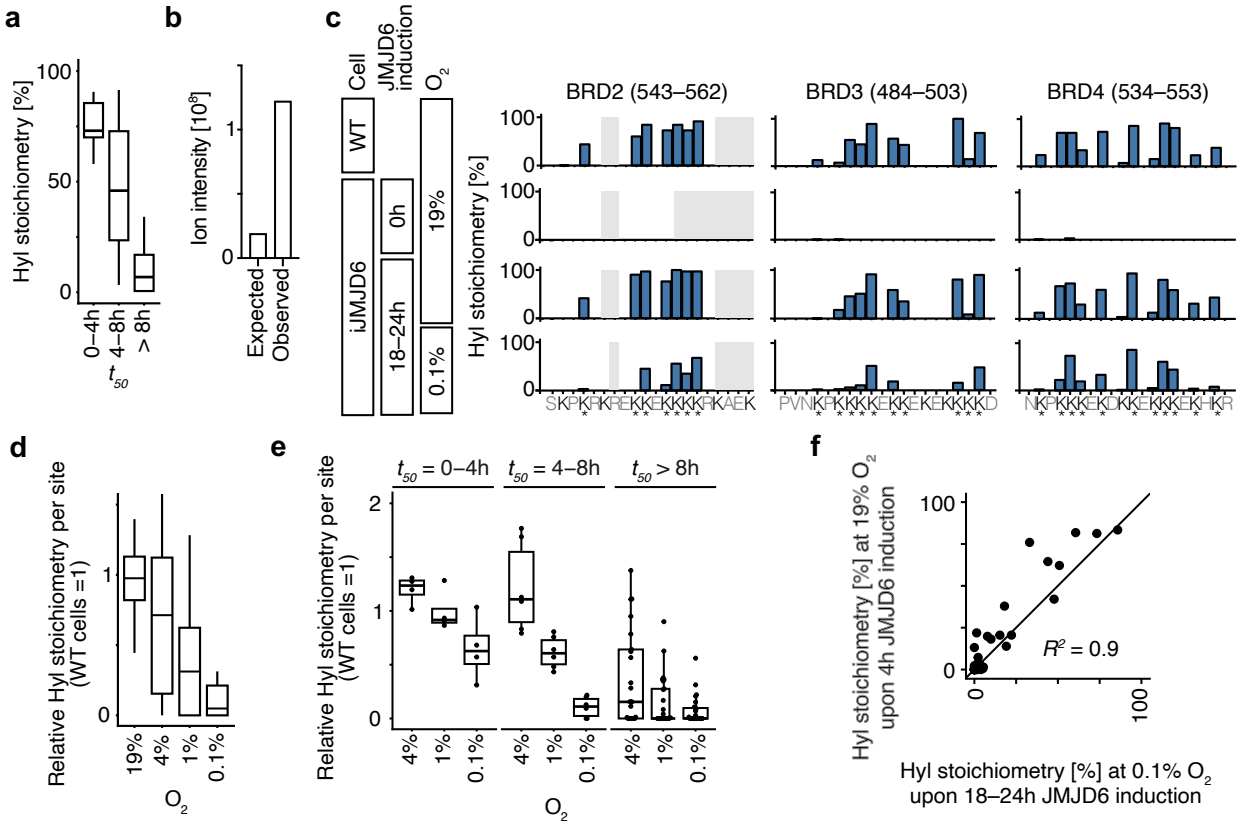
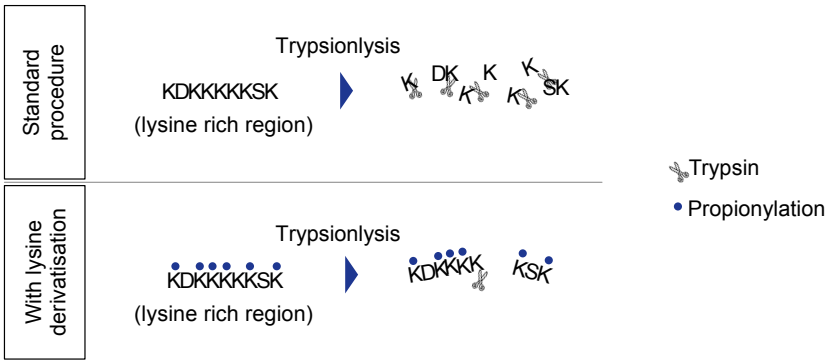
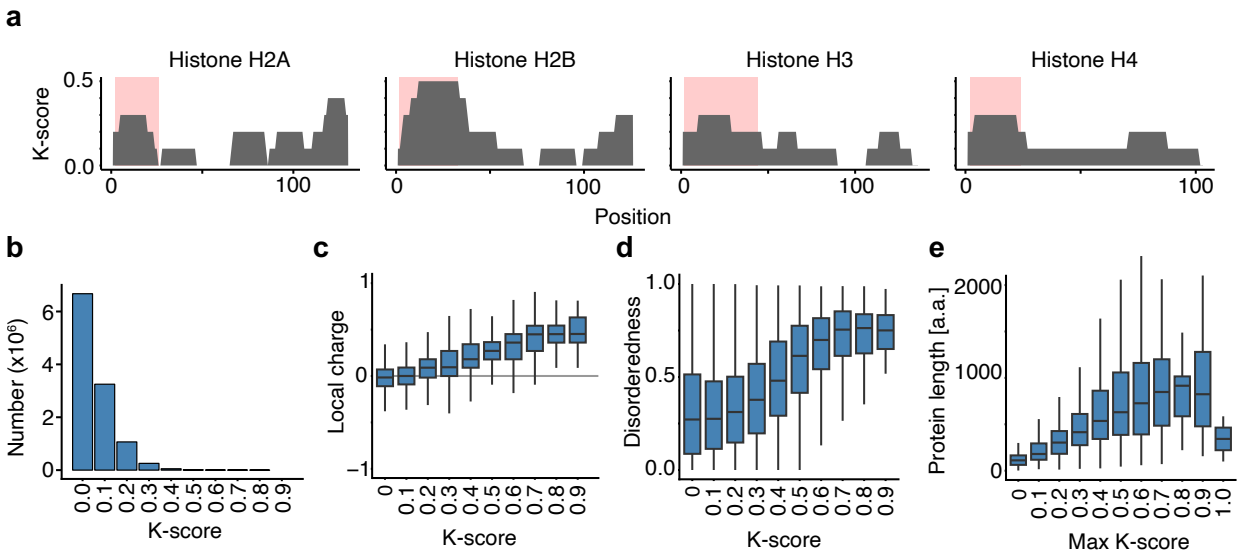


Fig. 5 | Hypoxia inhibits lysine hydroxylation in a site-dependent manner.

a. Hydroxylysine (Hyl) stoichiometry in wild-type (WT) HeLa cells as a function of the time required to reach half-maximal stoichiometry (t_{50}) under normoxia. n = 9, 13, and 26 for t_{50} = 0–4h, 4–8h, and > 8h, respectively. **b.** Bar plot comparing the expected XIC peak area of the doubly hydroxylated BRD4 (520–537) peptide if hydroxylation at K535 and K537 occurred independently with the observed values. **c.** Hyl stoichiometry in the lysine-rich regions of BRD2–4. WT HeLa cells were compared with JMJD6-inducible (iJMJD6) HeLa cells with or without JMJD6 re-expression. Grey boxes indicate positions with insufficient data coverage (PSMs < 3). Asterisks indicate lysine residues identified as hydroxylated by peptides with diagnostic ions and/or by manual inspection of spectra in our previous work[5]. **d.** Hyl stoichiometry per site in iJMJD6 HeLa cells after 18–24h of JMJD6 induction under the indicated oxygen level, normalised to Hyl stoichiometry in WT HeLa cells at 19% O_2. n = 48. **e.** As in **d**, stratified by t_{50} under normoxia. n = 4, 6, and 25 for t_{50} = 0–4h, 4–8h, and > 8h, respectively. **f.** Scatter plot showing Hyl stoichiometry in iJMJD6 HeLa cells after 18–24h of JMJD6 induction with concurrent exposure to 0.1% O_2 versus 4h of JMJD6 induction under normoxia (19% O_2).

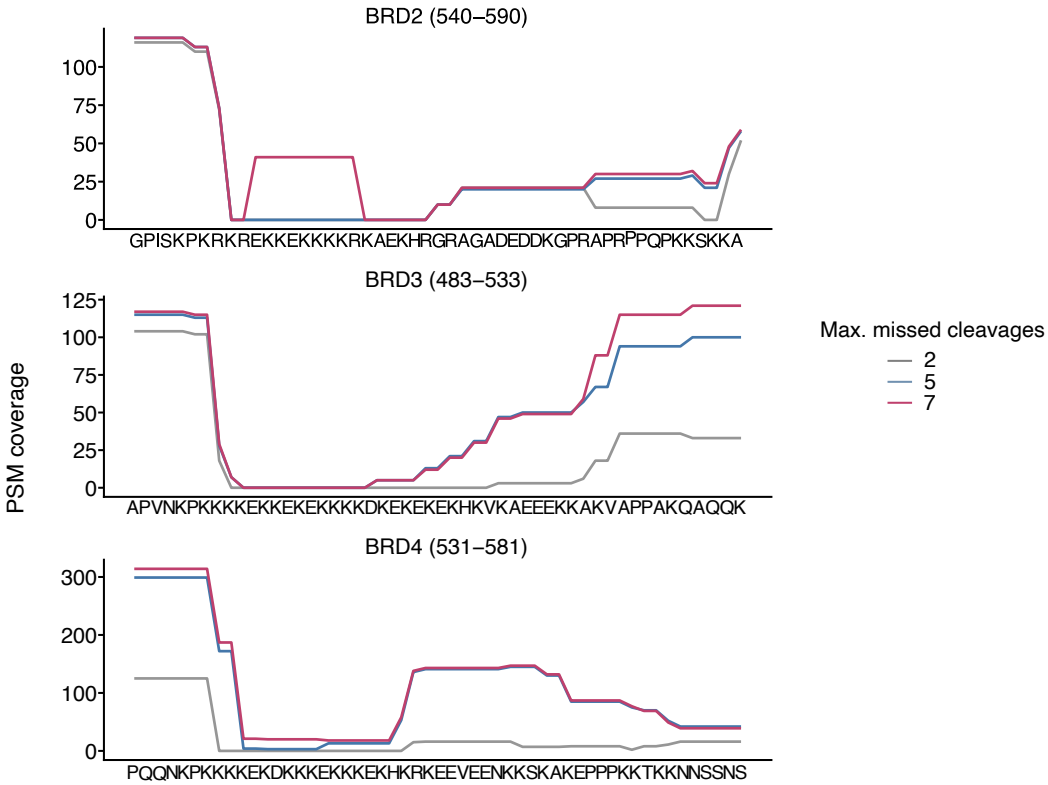


Supplementary Fig. 1 | Schematic of the effect of lysine derivatisation on trypsinolysis of lysine-rich regions.



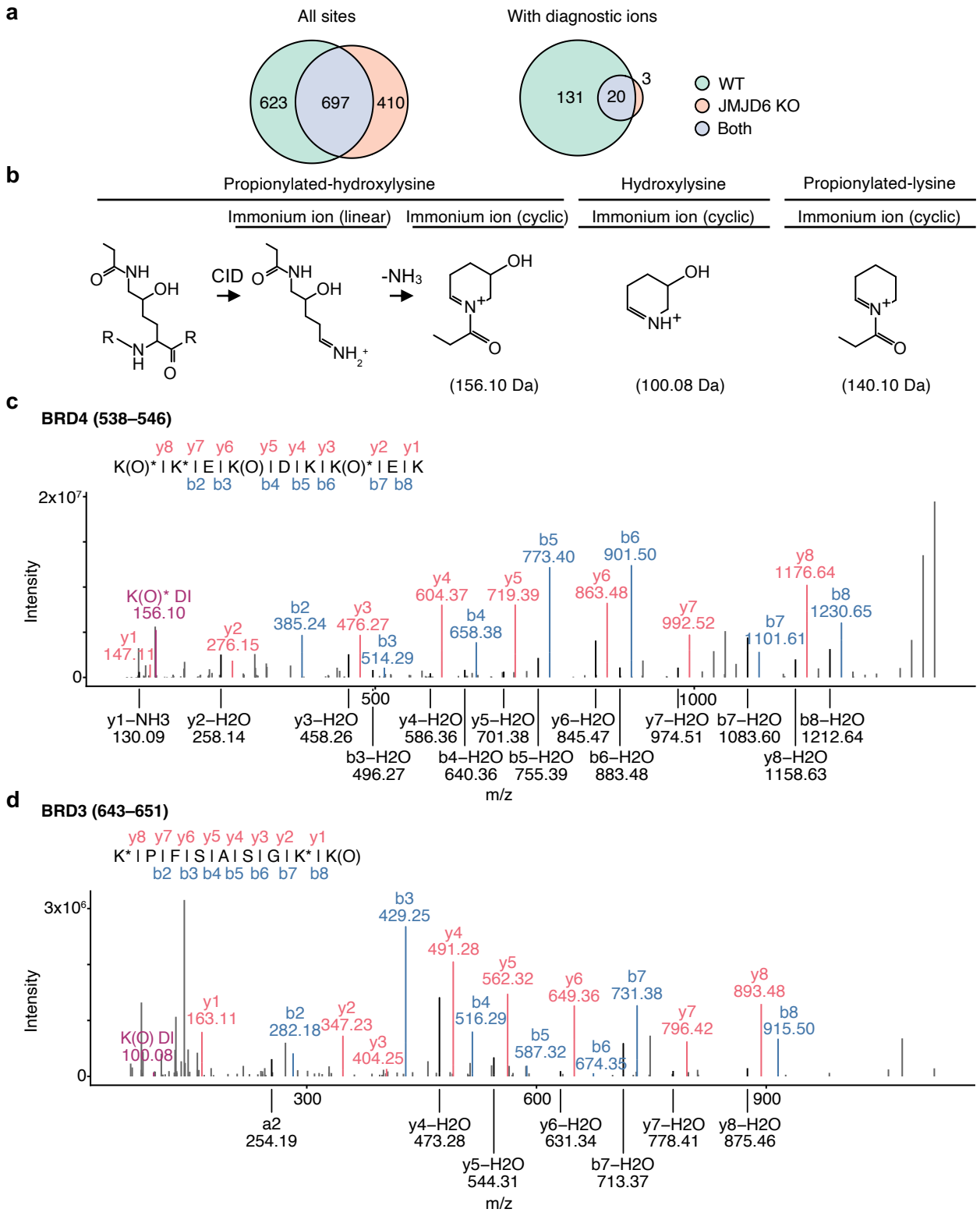
Supplementary Fig. 2 | Characteristics of lysine-rich regions.

a. Area plots showing K-score profiles across histone proteins. N-terminal histone tails are highlighted in red. **b.** Bar chart showing the distribution of the K-score calculated for each residue across the human proteome. **c.** Distribution of local charge as a function of K-score. Local charge was calculated over a window of ±5 residues. All sliding windows across the human proteome were analysed. **d.** As in **c**, but showing disorderedness as a function of K-score. Disorderedness was predicted using IUPred2A[31]. **e.** Distribution of protein length (amino acids; a.a.) as a function of maximum K-score.



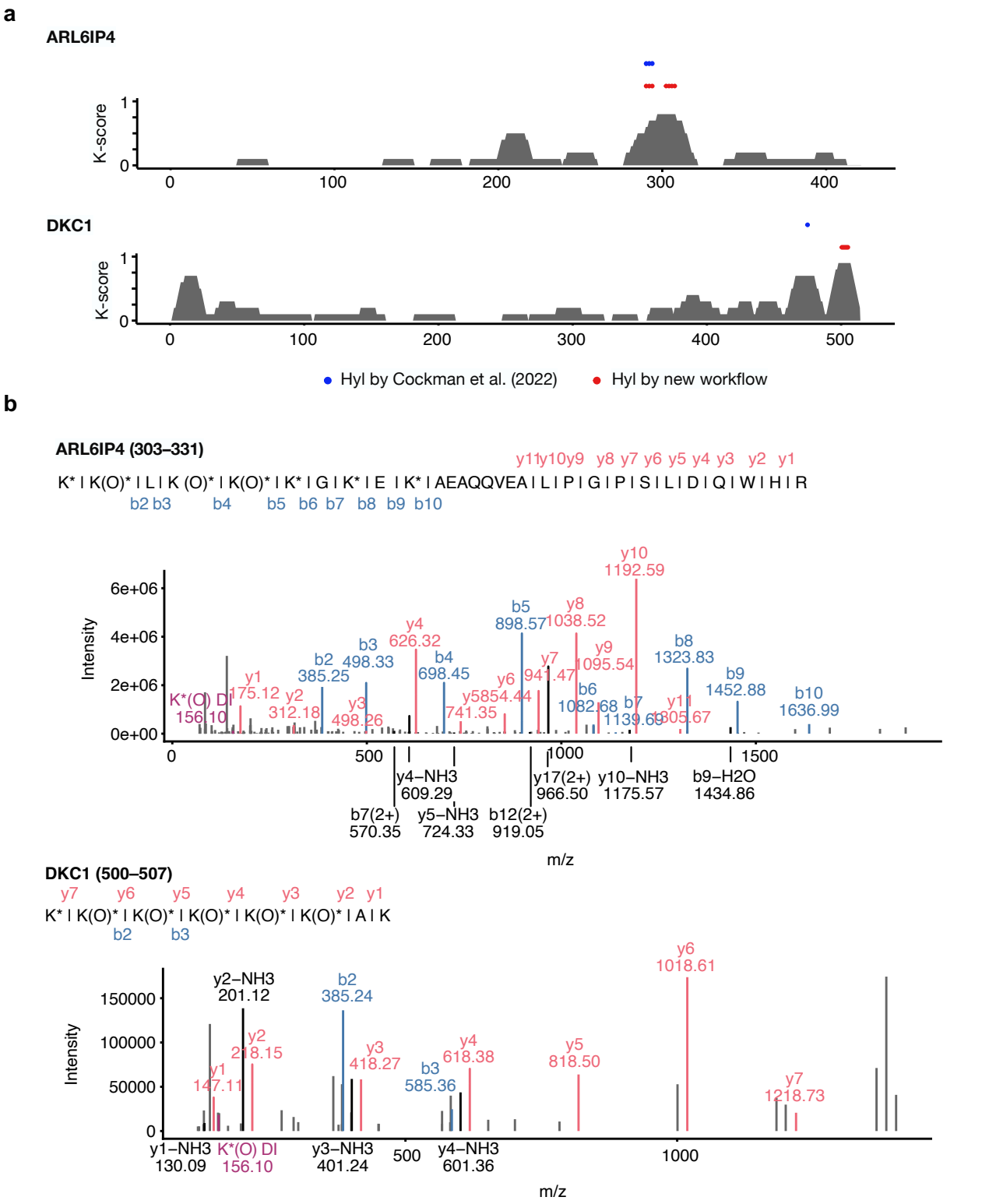
Supplementary Fig. 3 | Optimisation of database search settings.

Peptide-spectrum match (PSM) coverage across lysine-rich regions of BRD2–4 under database search settings allowing 2, 5, or 7 missed cleavages. In each search, the maximum number of allowed missed cleavages was matched to the maximum number of variable modifications per peptide.



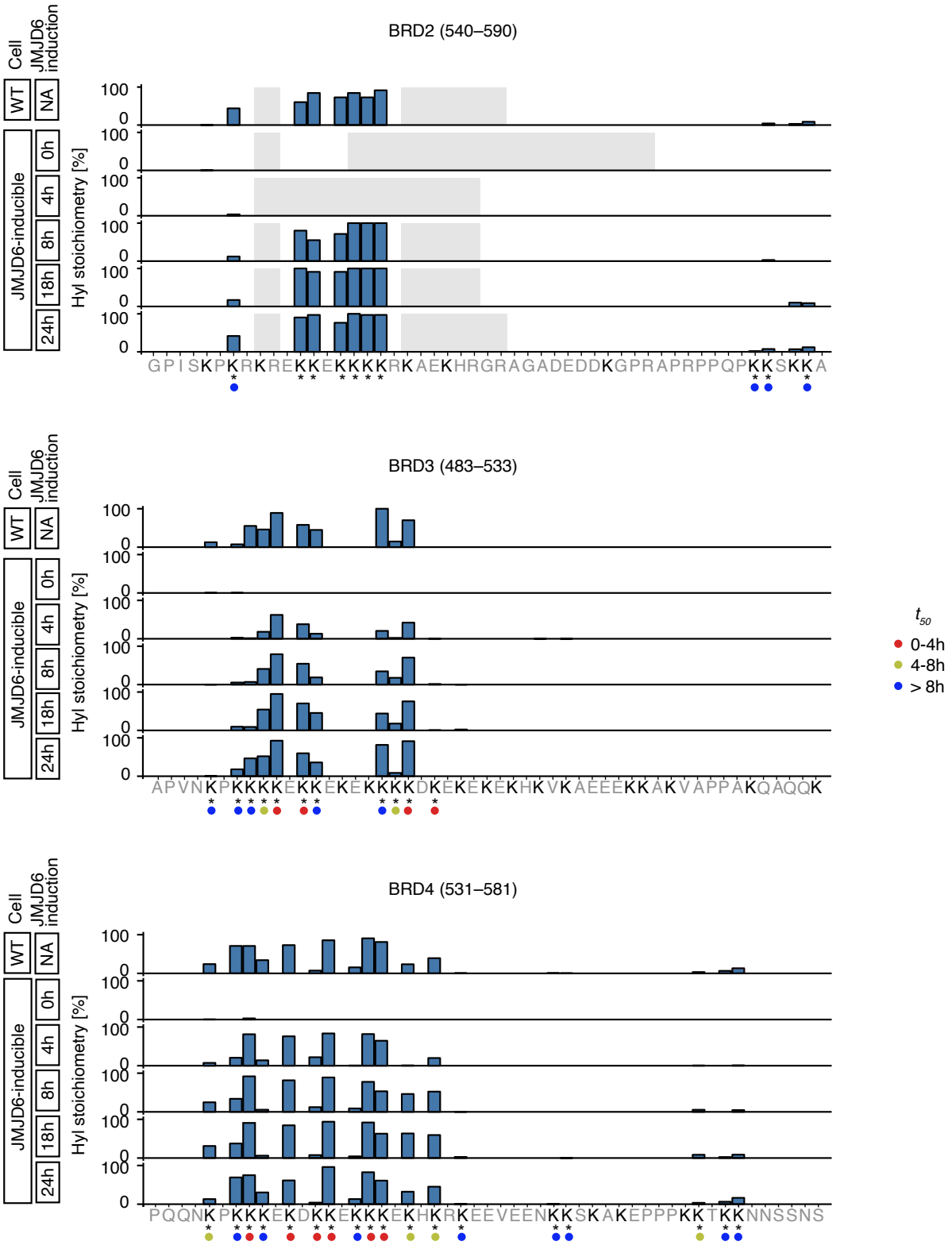
Supplementary Fig. 4 | Diagnostic ions of hydroxylysine and propionylated hydroxylysine.

a. Number of lysine hydroxylation sites identified in wild-type (WT) and JMJD6-inactivated (JMJD6 KO) HeLa cells. Data are shown for all sites and for sites supported by diagnostic ions. **b.** Schematic of diagnostic ions for propionylated hydroxylysine, hydroxylysine, and propionylated lysine. For propionylated hydroxylysine, collision-induced dissociation (CID) generates a linear immonium ion that can undergo ammonia loss to produce the cyclic form. **c.** MS/MS spectrum of a BRD4 peptide (residues 538–546), showing the diagnostic ion peak for propionylated hydroxylysine. Assigned b- and y-ions are annotated. Asterisks indicate propionylation, O indicates hydroxylation, and DI indicates the diagnostic ion. **d.** As in **c**, but for a BRD3 peptide (residues 643–651), showing the diagnostic ion peak for hydroxylysine.



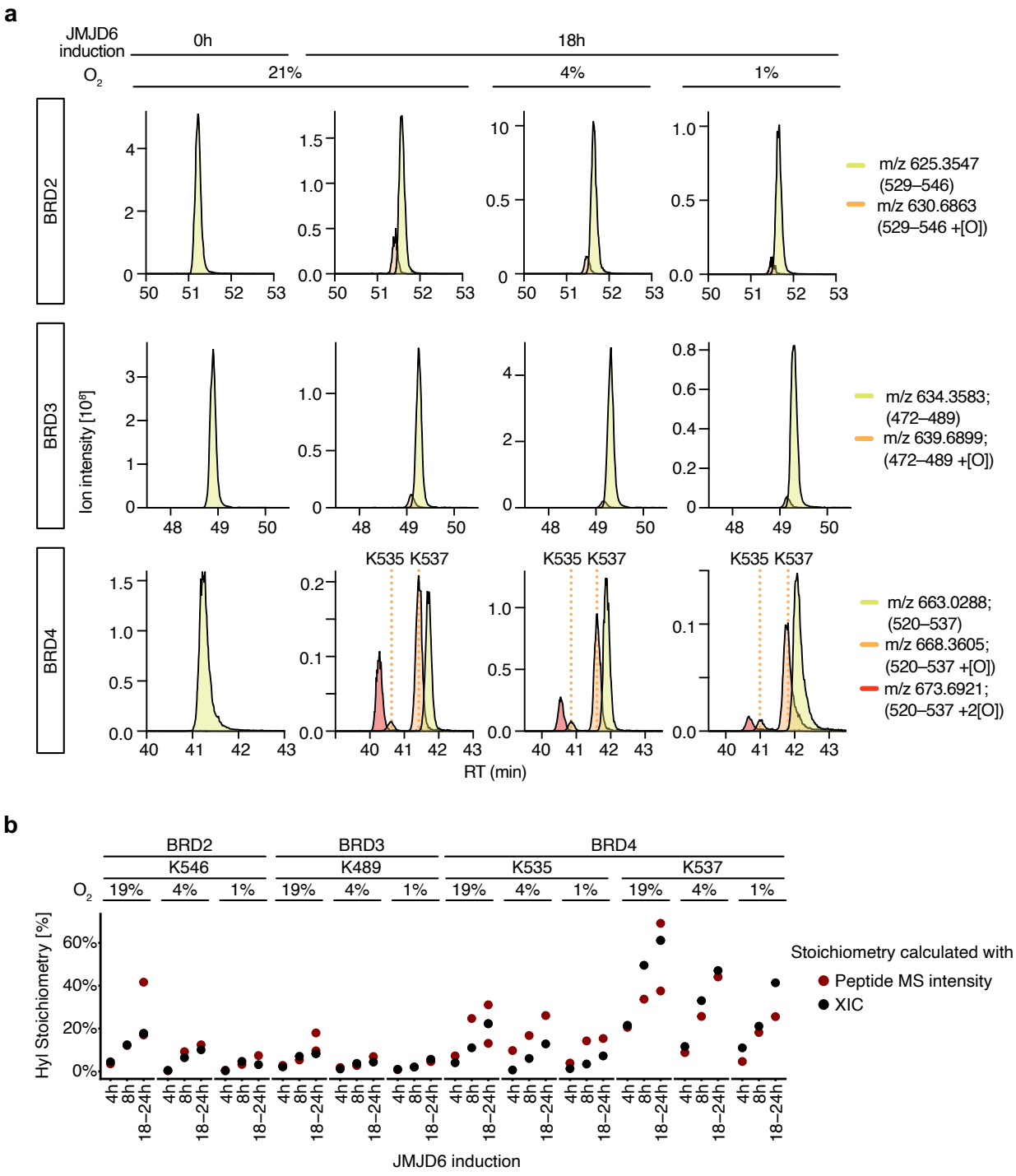
Supplementary Fig. 5 | Novel JMJD6-dependent hydroxylation sites.

a. K-score profiles across ARL6IP4 and DKC1. Lysine hydroxylation sites identified in our previous work[5] and by the new workflow developed in this study are indicated by blue and red, respectively. **b.** MS/MS spectra of the ARL6IP4 peptide (residues 303–331) and the DKC1 peptide (residues 500–507). Assigned b- and y-ions are annotated. Asterisks indicate propionylation, O indicates hydroxylation, and DI indicates the diagnostic immonium ion.



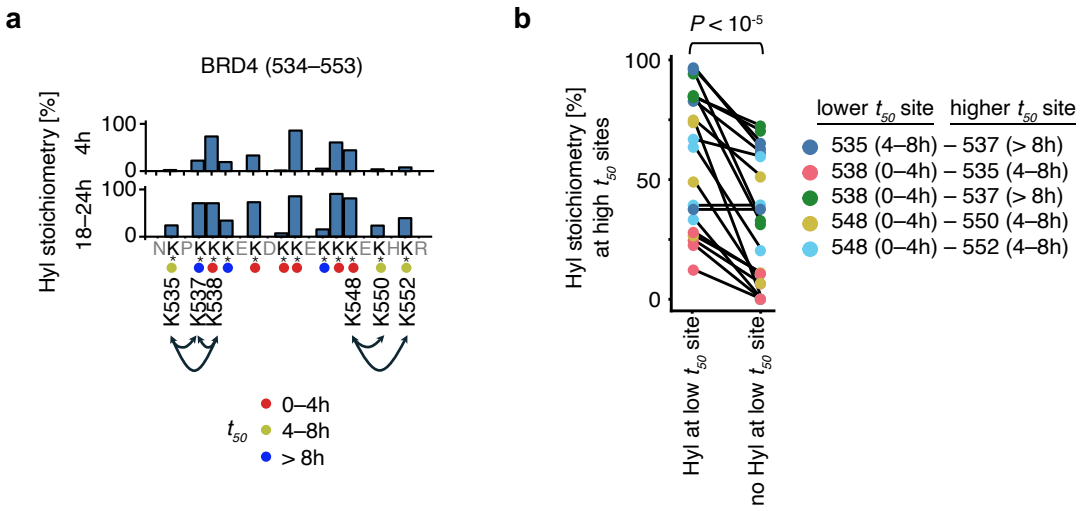
Supplementary Fig. 6 | Stoichiometry of lysine hydroxylation in BRD proteins.

Hydroxylysine (Hyl) stoichiometry in the lysine-rich regions of BRD2–4. Data are shown for wild-type (WT) HeLa cells and JMJD6-inducible HeLa cells following increasing durations of JMJD6 re-expression. Dots below each lysine residue indicate the time required to reach half-maximal stoichiometry (t_{50}) under normoxia. Asterisks indicate lysine residues identified as hydroxylated by peptides with diagnostic ions and/or by manual inspection of spectra in our previous work[5].



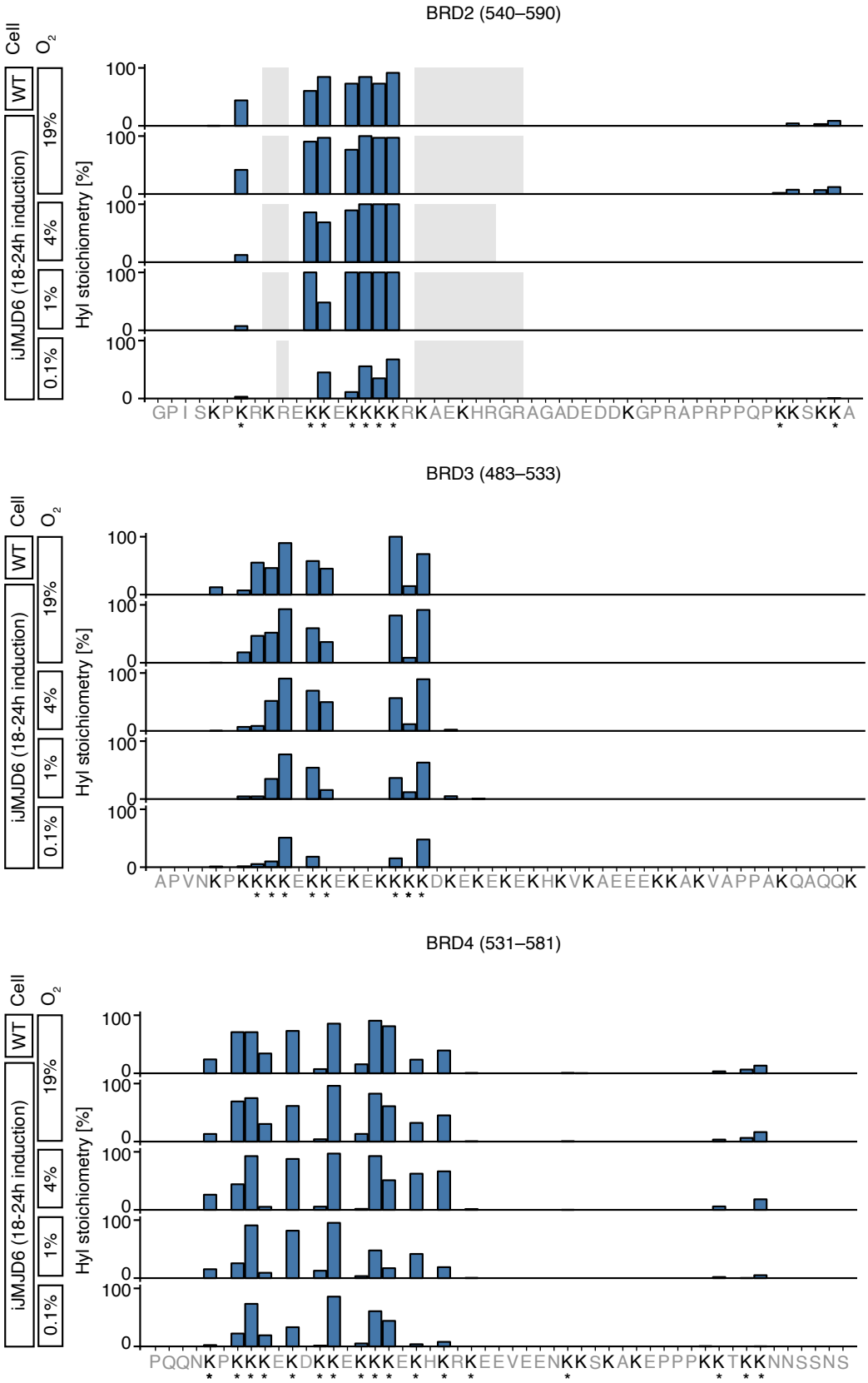
Supplementary Fig. 7 | Consistent hydroxylysine stoichiometry quantification by XIC- and peptide MS intensity-based methods.

a. Extracted ion chromatograms (XICs) of the unmodified (yellow), singly hydroxylated (orange), and doubly hydroxylated (red) triply charged BRD2 (529–546), BRD3 (472–489), and BRD4 (520–537) peptides. JMJD6-inducible HeLa cells were cultured with doxycycline at 19%, 4%, or 1% O_2 for 18–24 h, or without doxycycline at 19% O_2. A 2h gradient resolved BRD4 K535 and K537 as distinct chromatographic peaks (vertical dotted lines). **b.** Scatter plots comparing hydroxylysine (Hyl) stoichiometry at specific residues in BRD2–4, calculated from XIC analysis of non-derivatised samples (black) and from MS peptide intensities of derivatised samples (red). Data comprise 48 data points in total (4 sites × 3 oxygen levels × 3 induction durations, with biological duplicate measurements for the 18–24 h time point in derivatised samples). **a** and **b**: Only residues with sufficient data coverage in non-derivatised datasets were analysed.

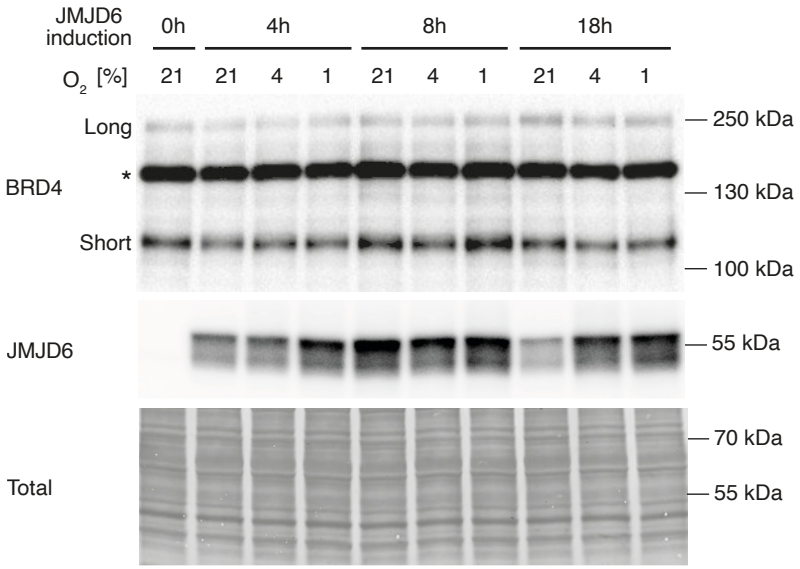


Supplementary Fig. 8 | Lysine hydroxylation sites exhibit interdependence.

a. Schematic of a lysine hydroxylation site cluster in BRD4. Hydroxylysine (Hyl) stoichiometry is shown for JMJD6-inducible HeLa cells following 4 or 18–24h of JMJD6 re-expression. Dots below each lysine residue indicate the time required to reach half-maximal stoichiometry (t_{50}) under normoxia. Arrows indicate lysine pairs spanning different t_{50} categories that were analysed in **b**. Asterisks indicate lysine residues identified as hydroxylated by peptides with diagnostic ions and/or by manual inspection of spectra in our previous work[5]. **b.** Hyl stoichiometry calculated from peptides in which low-t_{50} sites were hydroxylated, compared with peptides in which the low-t_{50} site was not hydroxylated. Groups were compared using a two-sided paired t-test.

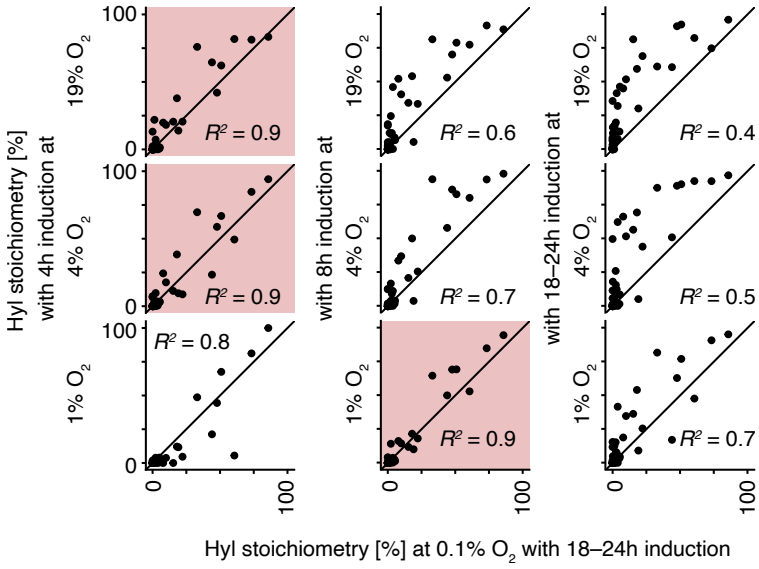


Supplementary Fig. 9 | Hypoxic inhibition of lysine hydroxylation in BRD proteins.
Hydroxylysine (Hyl) stoichiometry in lysine-rich regions of BRD2–4. Wild-type (WT) HeLa cells were compared with JMJD6-inducible (iJMJD6) HeLa cells following 18–24 h of JMJD6 re-expression under the indicated oxygen levels. Asterisks indicate lysine residues identified as hydroxylated by peptides with diagnostic ions and/or by manual inspection of spectra in our previous work[5].



Supplementary Fig. 10 | Effect of hypoxia on JMJD6 and BRD4 abundance.

Immunoblotting analysis of JMJD6-inducible HeLa cells following varying durations of JMJD6 induction under different oxygen levels. Cells were treated with doxycycline for 4, 8, or 18h at 19%, 4%, or 1% O_2, or cultured without doxycycline at 19% O_2. The short and long BRD4 isoforms were detected, consistent with previous studies. Non-specific species are indicated by an asterisk. Equivalent protein loading was confirmed by Coomassie staining of total protein.



Supplementary Fig. 11 | Hypoxia appears to reduce the apparent kinetics of JMJD6-catalysed lysine hydroxylation.

Scatter plots comparing hydroxylysine (Hyl) stoichiometry in JMJD6-inducible HeLa cells after 18–24h of JMJD6 re-expression at 0.1% O_2 with Hyl stoichiometry measured after the indicated duration of JMJD6 re-expression at higher oxygen concentrations. For each oxygen concentration, the condition with the highest R^2 value is highlighted in red.