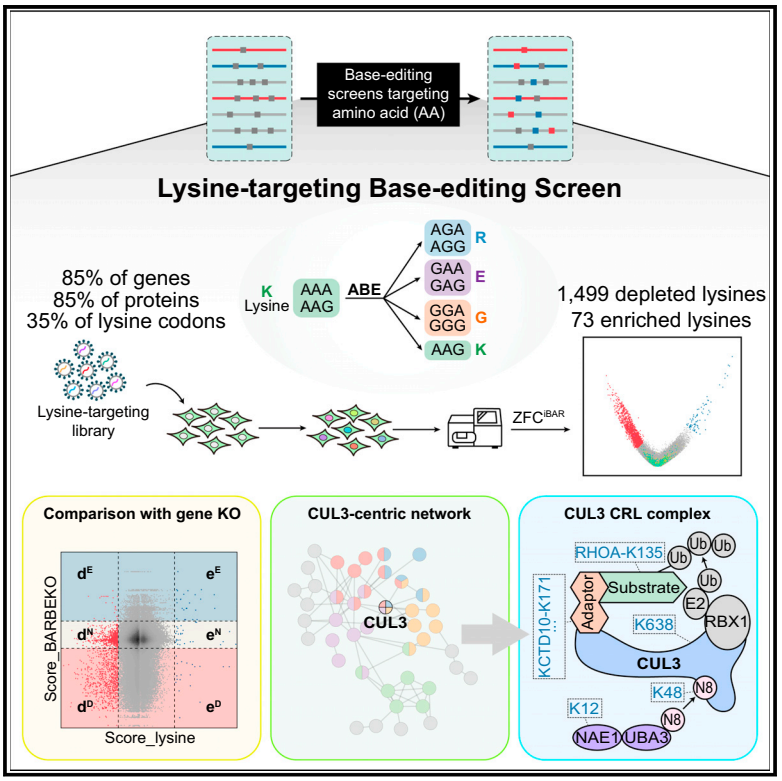
Molecular Cell

Resource

Unbiased interrogation of functional lysine residues in human proteome

Graphical abstract



Highlights

- A base-editing screen reveals functional lysine residues at the genome scale
- Identification of 1,572 lysine codons whose mutations affect cell fitness
- Functional lysine residues frequently undergo post-translational modifications
- Key lysine residues in the CUL3 CRL complex proteins regulate the cell cycle

Authors

Ying Bao, Qian Pan, Ping Xu, ..., Ying Yu, Zhuo Zhou, Wensheng Wei

Correspondence

zhouzhuo@gmail.com (Z.Z.), wswei@pku.edu.cn (W.W.)

In brief

Bao et al. establish a CRISPR strategy for unbiasedly probing functional amino acid residues at the genome scale. They systematically survey the functional lysine residues that affect cell fitness. Novel mechanisms that govern the cell cycle via lysine residues are unraveled, and clinically significant lysine mutations are identified.

Bao et al., 2023, Molecular Cell *83*, 1–19
December 21, 2023 © 2023 Elsevier Inc.
https://doi.org/10.1016/j.molcel.2023.10.033

CellPress

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

Molecular Cell

CellPress

Resource

Unbiased interrogation of functional lysine residues in human proteome

Ying Bao,[1,3,4] Qian Pan,[1,4] Ping Xu,[1,4] Zhiheng Liu,[1,4] Zhixuan Zhang,[1] Yongshuo Liu,[1] Yiyuan Xu,[1] Ying Yu,[1] Zhuo Zhou,[1,2,*] and Wensheng Wei[1,3,5,*]

[1]Biomedical Pioneering Innovation Center, Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, Peking University Genome Editing Research Center, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China

[2]State Key Laboratory of Common Mechanism Research for Major Diseases, Suzhou Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Suzhou 215123, Jiangsu, China

[3]Changping Laboratory, Beijing 102206, China

[4]These authors contributed equally

[5]Lead contact

*Correspondence: zhouzhuo@gmail.com (Z.Z.), wswei@pku.edu.cn (W.W.)

https://doi.org/10.1016/j.molcel.2023.10.033

SUMMARY

CRISPR screens have empowered the high-throughput dissection of gene functions; however, more explicit genetic elements, such as codons of amino acids, require thorough interrogation. Here, we establish a CRISPR strategy for unbiasedly probing functional amino acid residues at the genome scale. By coupling adenine base editors and barcoded sgRNAs, we target 215,689 out of 611,267 (35%) lysine codons, involving 85% of the total protein-coding genes. We identify 1,572 lysine codons whose mutations perturb human cell fitness, with many of them implicated in cancer. These codons are then mirrored to gene knockout screen data to provide functional insights into the role of lysine residues in cellular fitness. Mining these data, we uncover a CUL3-centric regulatory network in which lysine residues of CUL3 CRL complex proteins control cell fitness by specifying protein-protein interactions. Our study offers a general strategy for interrogating genetic elements and provides functional insights into the human proteome.

INTRODUCTION

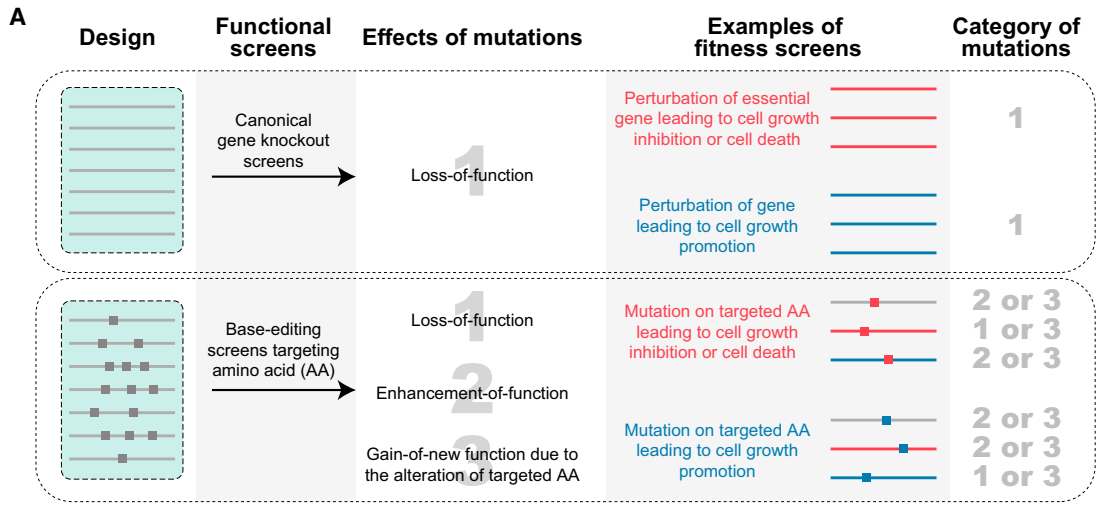
Amino acid (AA) residues are protein building blocks that underpin protein structure and activity. Genetic substitutions of AAs are widely associated with physiological changes or diseases and take up a large proportion among all the known pathogenic mutations.[1] Thus, it is valuable to systematically interrogate AA residues in certain physiological or pathological contexts. Currently, this is primarily accomplished through low-throughput fashion.

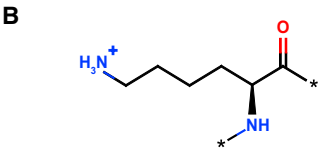
The recently developed CRISPR-Cas-mediated base-editing technology (BE) has enabled single-base substitutions at the DNA level. Base editors include adenine base editors (ABEs) and cytosine base editors (CBEs), which convert an A•T base pair into a G•C base pair[2,3] and a C•G base pair into a T•A base pair,[4,5] respectively. Because BEs enable the efficient installation of point mutations in various organisms,[6] they are applicable to high-throughput genetic screens at a high resolution. Recently, BEs have been leveraged in pooled screens to survey single nucleotide variants (SNVs) or map functional residues of specific proteins in mammalian cells, identifying many functional variants involved in diverse pathways.[7–11] Nevertheless, only specific genes or limited SNVs are included in these studies.

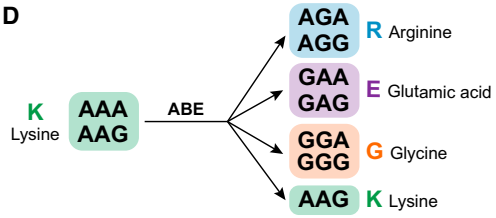
Here, we aim to unbiasedly probe functional AA residues at the genome scale by BE screens. Because eukaryotic proteins consist of a median length of 361 AA,[12] the total number of genetic codons for targeting is hundreds of times larger than the number of genes, posing challenges for handling genome-wide libraries.[13,14] Thus, we focus on one specific AA residue. Lysine residues are attractive targets for such a survey because they serve as the primary acceptors for multiple post-translational modifications (PTMs) such as ubiquitination and acetylation, which are critical for modulating protein function and stability.

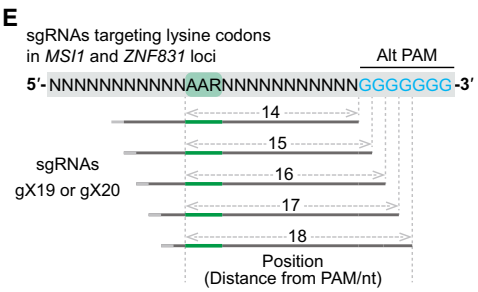
To accomplish this, we leverage ABEs to edit lysine codons across the genome, resulting in mutations of lysine residues to glycine, arginine, or glutamic acid residues. To improve screening efficiency and accuracy, we integrate the lysine-targeting sgRNA library with iBARs, a strategy we previously devised to enhance screening accuracy and reduce the number of starting cells by coupling sgRNAs with internal barcodes.[15,16] With this platform, we conduct a screen for functional lysine residues that regulate cell fitness, a fundamental cellular process. In addition to known lysine residues, the screen reveals several hundred lysine residues involved in cell fitness, many of which are enriched in specific structural domains, signaling pathways, or protein complexes. Our study provides functional annotations

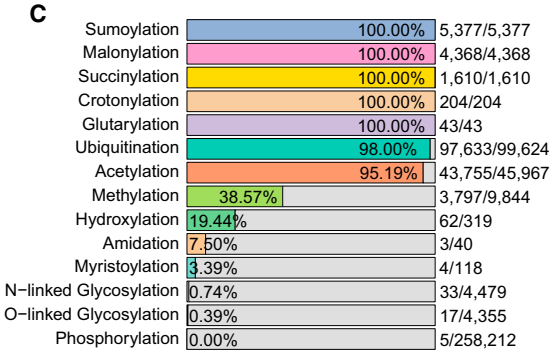
Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

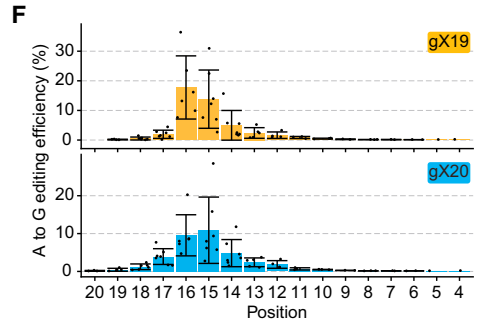












G

Position		22	21	20	19	18	17	16	15	14	13	12	11	10	…	PAM
Spacer length (nt)	21	g	N	N	**A**	**A**	R	N	N	N	N	N	N	N	…	
	20		g	N	N	**A**	**A**	R	N	N	N	N	N	N	…	
	19			g	N	N	**A**	**A**	R	N	N	N	N	N	…	
	19			g	N	N	N	**A**	**A**	R	N	N	N	N	…	
	19			g	N	N	N	N	**A**	**A**	R	N	N	N	…	
	20		g	N	N	N	N	N	**A**	**A**	**A**	R	N	N	…	
	20		g	N	N	N	N	N	N	**A**	**A**	R	N	N	…	

Figure 1. Rational design for efficient mutagenesis of lysine codons by base editing

(A) Schematics for interpreting functional screens at the gene level (upper) versus the AA level (lower).

(B) A skeletal formula of L-lysine residue.

(C) Bar plot showing the proportions of indicated PTMs at lysine residues (dbPTM, May 2020).

(D) Possible editing consequences at lysine codons by ABE.

(legend continued on next page)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

of lysine residues in cell fitness and offers a general strategy for mapping functional genetic elements at the genome scale.

RESULTS

Interpreting functional screens at the amino acid level

We first deduced how to interpret the consequences of perturbations of AA residues. Canonical CRISPR knockout (KO) screens interrogate gene functions by introducing insertions or deletions (indels) to cause loss-of-function mutations. In the context of cell fitness screens, essential genes whose loss-of-function cause cell growth inhibition or cell death can be negatively selected, whereas genes whose loss-of-function promote cell growth can be positively selected (Figure 1A, upper).

When leveraging BEs for screening functional AA residues, this paradigm becomes more complicated because perturbation of functional AA may either disrupts or enhances protein function, or even confers the protein a new function (Figure 1A, lower). In a given screen, loss-of-function mutation on targeted AA may give rise to the same phenotype as the gene KO; however, AA mutation that endows a protein with an enhanced or novel function may produce a distinct phenotype from the gene KO. For example, a mutation at a specific AA of an essential gene, whose KO causes cell death, potentially promotes cell growth. These analyses are helpful to interpret the results from functional screens at the AA level.

Leveraging ABEmax for site-directed mutagenesis of lysine codons

As one of the proteinogenic AA, L-lysine contains an ε-amino group (Figure 1B), which is positively charged at physiological pH and plays numerous essential roles in protein functions.[17] Moreover, lysine residues undergo a wide range of reversible PTMs (Figure 1C). Among them, ubiquitination is well known for its role in regulating protein activity and stability.[18] More recently, lysine acetylation and methylation in histone proteins have revolutionized our understanding of the epigenetic regulation of cellular functions.[19,20] Other PTMs at lysine residues, such as sumoylation, malonylation, and succinylation, have also been widely reported.[21,22]

To gain a systematic insight into functions of lysine residues, we attempted to establish an approach for forward genetic screens at the genome scale. Lysine residues are encoded by the genetic codes of 5′-AAA and 5′-AAG, which can be mutated by ABEs to the codons of arginine (5′-AGA or 5′-AGG), glutamic acid (5′-GAA or 5′-GAG), or glycine (5′-GGA or 5′-GGG) (Figure 1D). By coupling with CRISPR sgRNA library, this ABE-based site-directed mutagenesis strategy allows for functional screens of lysine residues in a high-throughput fashion.

First, we tested lysine mutagenesis efficiency with the ABEmax system.[3] We designed multiple sgRNAs targeting the codons of lysine 109 of MSI1 (MSI1-K109) and lysine 1018 of ZNF831 (ZNF831-K1018) (Figure 1E; Table S1). Both codons are located upstream of a sequence containing 7 or 8 consecutive guanosines, which serve as sliding regions for alternative protospacer-adjacent motif (PAM). The editing outcomes of ABEmax are affected by the relative position of targeted adenine and the spacer length of sgRNA.[2,23] To maximize the coverage and mutagenesis efficiency, we tested gX19 and gX20 types of sgRNAs ("g" indicates the guanine at the 5′ end, whereas "X19/20" indicates a 19/20-nt spacer sequence). It turned out that gX19 sgRNAs showed higher A-to-G editing efficiency at positions 15 and 16, while gX20 sgRNAs had higher efficiency at positions 12–14 and 17–18 (Figure 1F).

Then, we analyzed the distribution of AA mutations generated by ABEmax editing. For both gX19 and gX20 sgRNAs, lysine codons were more likely to be mutated to arginine at positions 17 and 18, and to glycine at position 16. At positions 11–15, glutamic acid took up the highest proportion (Figure S1).

Based on the above observations and related publications,[3,23] we established the principle of designing an sgRNA library for lysine mutagenesis screen. To maximize the editing efficiency of both the 1st and 2nd adenosines in lysine codons, gX19 sgRNAs were employed at positions 15–17; meanwhile, gX20 sgRNAs were selected at positions 13, 14, and 18. To further increase the coverage, gX21 sgRNAs were included at position 19. All qualified sgRNAs targeting the same lysine codon, which may yield diverse editing outcomes, were included (Figure 1G).

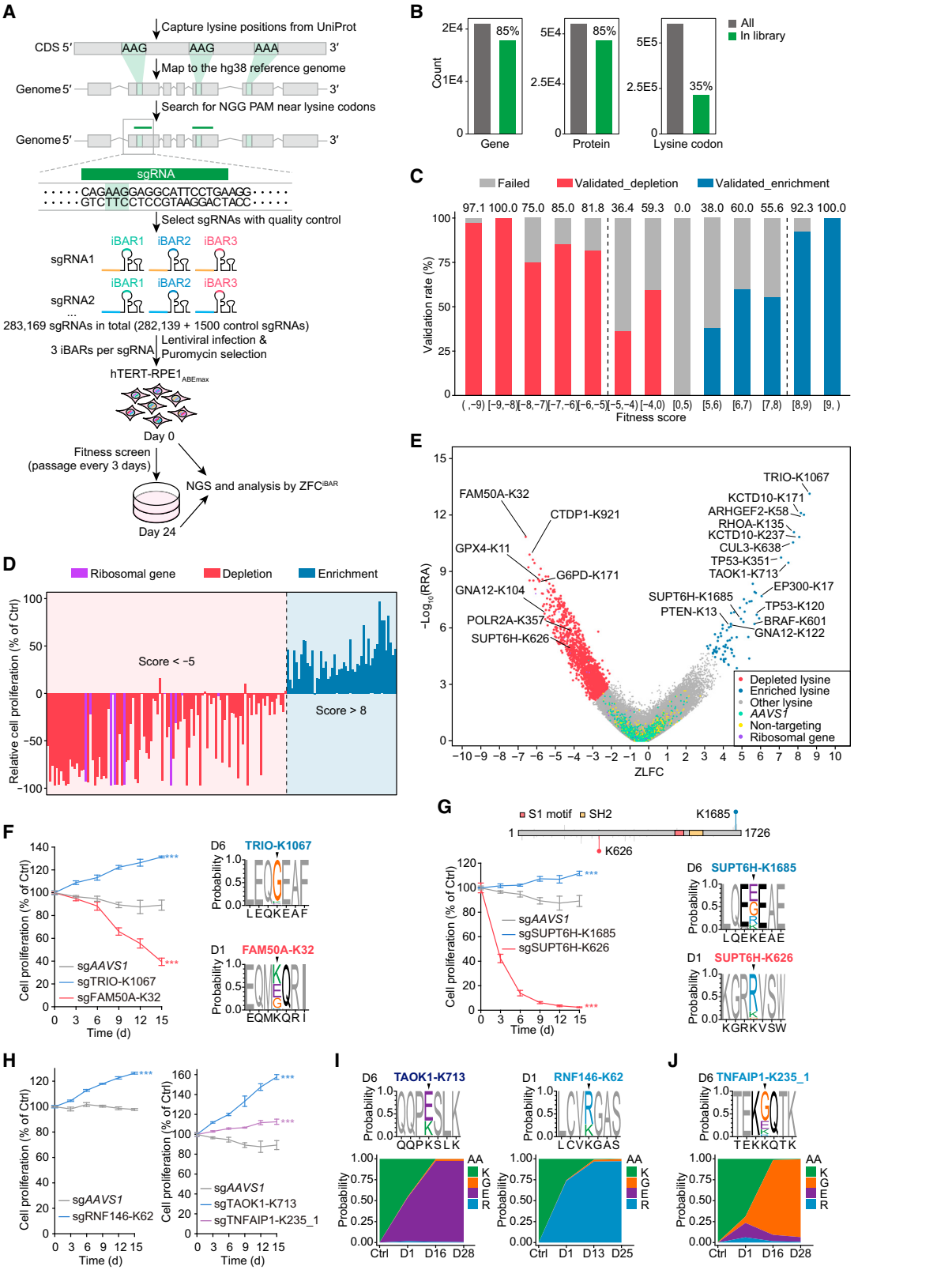
Genome-wide screen of functional lysine residues in cell fitness

To design a genome-wide lysine-targeting sgRNA library, we acquired the relative position of every lysine residue in the human reference proteome (ID: UP000005640, UniProt database, release 2018_07) (Figure 2A). After mapping lysine residues to their corresponding genetic codons in the hg38 reference genome, we searched for all available PAM sequences (5′-NGG) to design sgRNAs using the principle summarized in Figure 1G. After quality control, the number of sgRNAs targeting lysine codons reached 282,139 (see STAR Methods; Table S1). This library targeted 35% of lysine codons, encompassing 85% of protein-coding genes or proteins across human genome or proteome (Figure 2B). Afterward, 500 non-targeting sgRNAs and 500 sgRNAs targeting *AAVS1* loci were included as negative controls, with 30 sgRNAs targeting splice sites or start codons of 10 ribosomal genes serving as essential gene KO controls.

It would be onerous to manage a cell library containing nearly 300,000 sgRNAs using the traditional CRISPR screen method.[13] To address this, we employed the previously established iBAR method that enables high-throughput screens using libraries made from lentiviral transduction at high multiplicity of infection (MOI),[15,16] which reduces the number of starting cells while increases screening accuracy. Briefly, iBARs refer to several different 6-nt barcodes assigned to one sgRNA serving as internal

(E) Design for testing the editing efficiency of ABEmax.
(F) The A-to-G editing efficiency of ABEmax at different positions with gX19 (upper) and gX20 (lower) sgRNAs. HEK293T cells were harvested for NGS analysis 3 days post-transfection. Data are averaged over 10 sgRNAs targeting *MSI1* and *ZNF831* loci and are presented as mean ± SD.
(G) Rules of spacer length selection for sgRNA library design.
See also Figure S1.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033



(legend on next page)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

replicates within the same screen process, and the effect of each sgRNA can be evaluated according to the significance and consistency of the abundance change of different sgRNAs[iBAR]. For the current screen, we employed three verified iBARs.

To investigate functional lysine residues that regulate cell fitness, we chose hTERT-RPE1 (referred to as RPE1), a model cell line of normal cell types,[24] for the screen. The sgRNA library was transduced into ABEmax-expressing RPE1 cells at an MOI of 3. Cells from the reference (day 0) and experimental (day 24) groups were subjected to next-generation sequencing (NGS) followed by analysis with ZFC[iBAR] algorithm.[16] ZLFC (*z* score of log fold-change) and RRA (robust rank aggregation)[25] were calculated for 3 sgRNAs[iBAR] per sgRNA. The Fitness Score ($FS = sign(zLFC) \times (-log_{10}(RRA) + |zLFC|)$) was used to rank sgRNAs, synthesizing the effect strength and statistical confidence of each sgRNA (Table S2).

Screen quality evaluation and analysis on targeted sites

After sgRNA ranking, we determined the threshold of the screen hits. To do this, we performed extensive validation by cell proliferation assay (CPA) using a randomly selected subset of sgRNAs spanning various FS regions, together with ribosomal gene-targeting sgRNAs (Figure S2; Table S3). The validation rate declined sharply when the FS region was within the range of −5 to 8 (Figure 2C). Therefore, sgRNA with an FS value less than −5 or greater than 8 was categorized as negatively or positively selected, respectively (Table S4). A thorough evaluation of sgRNAs meeting the selection criteria indicated a validation rate of 91.0% (142 out of 156; Figures 2D and S2; Table S3). Under the selection threshold, 1,605 depleted sgRNAs targeting 1,499 lysine codons and 76 enriched sgRNAs targeting 73 lysine codons were identified as candidates involved in cell-fitness regulation (Figure 2E; Table S4).

We then evaluated the screen quality by statistical analysis. Out of the assayed 349 sgRNAs, 245 that validated to influence cell proliferation are regarded as "actual positives" (Figure S2; Table S3). By generating an "FS threshold-true positive rate (TPR) curve" based on the performance of these 349 sgRNAs, the TPR of our screen was estimated to be 58.4% and 51.2% for the depletion and enrichment direction, respectively (Figure S3A). The area under the receiver operating characteristic curve (AUC), a measurement of screening performance, was 0.86 and 0.93 for validated sgRNAs with $FS < 0$ and $FS \geq 0$, respectively (Figure S3B). The AUC and TPR of the ribosomal gene-targeting sgRNAs were 0.79 and 33.3%, respectively (Figures S3B and S3C). Further analysis revealed that our screen performed comparably to previously published BE screens[7,8] (Table S5).

ABE-mediated editing could be more efficient at optimal positions (positions 14–17) compared with neighboring suboptimal positions (positions 12, 13, 18, and 19). This variability could potentially introduce bias to the screen results. By examining the distribution of sgRNAs before and after the screen, as represented by the positions of targeted lysine codons, no evidence of a skewed distribution of sgRNAs after selection was observed (Figure S3D).

Notably, genes targeted by either negatively or positively selected sgRNAs exhibited higher expression levels than others (Figure S3E), suggesting an enrichment of selected lysine codons in genes that are more functional in cell-fitness regulation. We then analyzed potential PTMs on selected lysine residues using dbPTM database.[26] Ubiquitination and acetylation were significantly enriched in both directions, while sumoylation and methylation were only enriched in lysine residues with depleted sgRNAs (Figure S3F). These results suggest that the selected lysine residues are susceptible to PTMs, coinciding with the fact that PTMs could confer functionalities to lysine residues. Besides, selected lysine residues are significantly enriched in specific protein domains annotated in PROSITE database (Figure S3G).

Then, we performed NGS to examine the editing outcomes at targeted sites (see STAR Methods). The NGS analysis of the most enriched (sgTRIO-K1067) and depleted sgRNA (sgFAM50A-K32) showed precise edits at the targeted sites (Figure 2F). Of note, bystander editing, a widespread effect in BEs, was also observed in NGS analysis. Nevertheless, 30 out of 40 sgRNAs in the enrichment direction exhibited dominant editing at the targeted lysine codons compared with the surrounding codons (Figures 2F, 2G, 2I, 2J, 4D, 4E, 4G, 5C, 6B, and S3H), indicating that the majority of selected sgRNAs function on target.

Intriguingly, mutagenesis of two lysine residues within the transcription elongation factor SPT6 (SUPT6H) had contrasting effects on cell proliferation. *SUPT6H* is a common essential gene across numerous cell lines,[27] whose loss-of-function impedes cell proliferation. Although sgSUPT6H-K626 inhibited cell proliferation, sgSUPT6H-K1685 significantly promoted cell

Figure 2. Genome-wide screen and sgRNA validation

(A) Schematics illustrating the sgRNA[iBAR] library design and cell fitness screen.

(B) The library coverage ratios of targeted protein-coding genes, proteins, and lysine codons.

(C) Validation rate of sgRNAs within each FS subrange according to cell proliferation assay (CPA).

(D) Overview of CPA results for high-ranking sgRNAs. Bar plot shows the effects of indicated sgRNAs on cell proliferation in RPE1 cells. Mean ± SD, n = 3.

(E) Volcano plot showing the screen results of functional lysine residues in cell fitness. Red and blue dots represent negatively (FS < −5) and positively selected (FS > 8) sgRNAs, respectively.

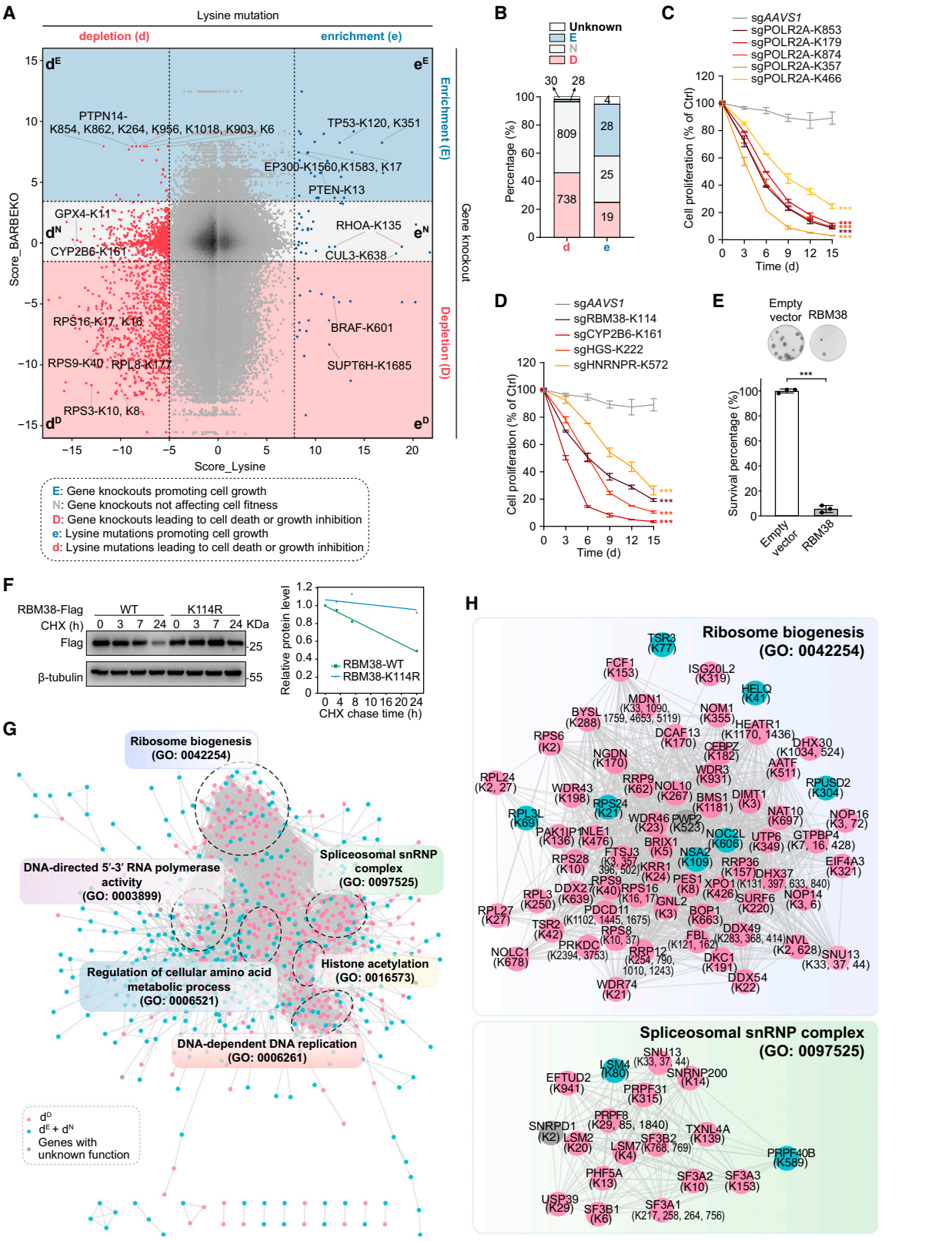
(F) Validation of sgFAM50A-K32 and sgTRIO-K1067 by CPA (left) and NGS analysis (right). NGS results are visualized as sequence logos. X-axis: reference sequences from cells expressing sg*AAVS1*. Targeted lysine residues are centered within the sequences. Residues in gray: no edits; residues in black: unintended edits. The time points for NGS are indicated in the upper-left corner. Mean ± SD, n = 3.

(G) Validation of sgRNAs targeting *SUPT6H* by CPA (lower-left) and NGS analysis (lower-right). Upper: *SUPT6H* gene structure with the FS values of sgRNAs. The length of each bar represents the absolute FS value. Bars positioned above or below the rectangle indicate positive or negative FS, respectively. Mean ± SD, n = 3.

(H–J) Validation of three sgRNAs with varied mutation distributions. Validation by CPA (H) and NGS analysis (I and J) are shown. Area plots in the lower panels of (I) and (J) display changing proportions of K/G/R/E at targeted sites over time analyzed by NGS tracking. Mean ± SD, n = 3.

See also Figures S2 and S3.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033



(legend on next page)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

growth (Figure 2G), most likely due to a gain-of-function mutation at this site. Similarly, two validated sgRNAs targeting distinct lysine codons of G protein subunit alpha 12 (*GNA12*), a reportedly potential oncogene,[28] caused opposite phenotypes in RPE1 cells (Figure S2B [panel 1]. These data well illustrate that BE screens at the AA level yield more in-depth information than conventional loss-of-function screens.

As aforementioned, ABEmax may result in the conversion of lysine (K) to arginine (R), glutamic acid (E), or glycine (G), as well as bystander editing. In specific scenarios, it becomes crucial to pinpoint the precise mutation(s) responsible for a phenotypic change. To address this, we employed NGS to track the mutation distributions across multiple hits at three time points after ABE editing (Figure S3I). All three possible mutations, i.e., K to R, K to E, and K to G, could be dominant editing outcomes (Figures 2H–2J). In some sites, the mutation distribution remained consistent throughout (Figure 2I), whereas in others, a minor mutation at the beginning became dominant at the end (Figure 2J). These dynamic changes in mutations could be unpredictable as it is confounded by editing efficiency and the selective pressure of screens. Hence, it is valuable to perform NGS at later time points to confirm the mutations that are actually functional.

Annotating functional lysine sites in the gene context

Previously, to avoid DNA double-strand break-associated side effects, we developed the BARBEKO method, which combines CBEs with iBARs, to achieve a gene KO screen for cell-fitness-associated genes.[16] This endeavor identified thousands of essential genes and hundreds of genes whose KOs promote cell growth, creating a high-quality atlas of fitness genes. By comparing lysine screen data with the BARBEKO data, both of which were obtained in RPE1 cells by BE screens, we were able to understand the functions of selected lysine residues in the gene context. Lysine residues were mirrored to their corresponding genes, which have also been assigned FS values in the BARBEKO screen (Figure 3A). For ease of identification, the uppercase letters "E," "D," and "N" denote genes whose KOs enhance, decrease, or have no effect on cell fitness, respectively. The lowercase letters "e" and "d" represent lysine mutations that enhance or decrease cell fitness, respectively.

In the e^E sector (where both lysine mutations and the corresponding gene KOs promote cell fitness), we identified TP53-K120 and PTEN-K13. This reinforced the robustness of our approaches because TP53 and PTEN are well-characterized tumor suppressors whose gene KOs unleash cell proliferation, and TP53-K120 and PTEN-K13 are known critical residues for gene function.[29,30] The d^D sector (where both lysine mutations and the corresponding gene KOs decrease cell fitness) comprised multiple lysine residues localized in ribosomal proteins, indicating their critical roles in ribosome biogenesis. Interestingly, numerous lysine residues fell into d^E, e^D, d^N, and e^N sectors, where lysine mutations and the corresponding gene KOs function divergently. Collectively, approximately half of lysine mutations cause consistent phenotypic alterations to gene KOs, while others display varied or even opposite effects (Figure 3B).

Depleted lysine residue mutations provide insights into therapeutic targets

We then focused on the depleted lysine residues. A subset of depleted sgRNAs was mirrored to 461 essential genes identified by the BARBEKO screen (d^D, 738 sgRNAs) (Figure 3B; Table S4), suggesting that they lead to loss-of-function mutations. Among them, sgFAM50A-K32 was top-ranking (Figure 2E). FAM50A is a poorly characterized protein. Recent discoveries indicate that its homolog FAM50B is underexpressed across various tumor types[31]; meanwhile, *FAM50A*/*FAM50B* is identified as a synthetic lethal pair,[31,32] making FAM50A a promising target for cancer therapeutics. Lysine residues are appealing drug targets due to their nucleophilic ε-amine that could react with irreversible inhibitors.[33,34] Moreover, K32 of FAM50A could be accessible to drugs as it is predicted to be localized to protein surface by AlphaFold (Figure S4). Therefore, the identification of FAM50A-K32 may offer valuable clues for the rational design of FAM50A-specific drugs.

Besides, the pooled screen identified five *POLR2A*-targeting sgRNAs, each of which was individually validated (Figure 3C), and one of these targeted sites, K853, is found to be ubiquitinated during transcription-coupled nucleotide excision repair.[35] Other high-ranking sgRNAs in the d^D sector, including those targeting DNA damage repair genes *RAD51C* and *RAD9A*,[36,37] were also verified (Figure S2A [panel 1]).

Another group of depleted sgRNAs was linked to 693 non-essential genes (d^E + d^N, 837 sgRNAs) (Figure 3B; Table S4), implying that the corresponding mutations confer functionalities to genes. Four high-ranking sgRNAs in this category, i.e., sgRBM38-K114, sgCYP2B6-K161, sgHGS-K222, and sgHNRNPR-K572, were verified to significantly inhibit cell proliferation (Figure 3D). PTMs such as ubiquitination at lysine residues can promote protein degradation,[18] whereas missense mutations at lysine residues could block degradative PTMs, resulting in protein accumulation. Therefore, the disparate phenotype observed between lysine mutations and gene KOs could, at least in part, results from protein accumulation caused by a PTM defect. To

Figure 3. Annotating selected lysine residues in the gene context

(A) Scatter plot showing the distribution of FS values of lysine residues and their corresponding genes. Dashed lines indicate the FS threshold of each screen. The threshold of BARBEKO screen is FS < −2 or FS > 3.

(B) Bar plot showing the distribution of sgRNAs targeting different lysine mutation categories. The uppercase letter "U" indicates genes with unavailable information in the BARBEKO screen.

(C) Validation of indicated sgRNAs targeting lysine codons of *POLR2A* (d^D). Mean ± SD, n = 3.

(D) Validation of indicated sgRNAs targeting lysine codons of four non-essential genes (d^N). Mean ± SD, n = 3.

(E) Clonogenic assay of RPE1 cells upon RBM38 overexpression. Mean ± SD, n = 3 (see STAR Methods).

(F) Degradation kinetics of RBM38 proteins in HEK293T cells (see STAR Methods).

(G–H) PPI network of genes with lysine mutations decreasing cell fitness (G). Two subgroups of genes enriched in GO analysis are shown in detail (H).

See also Figures S4 and S5.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

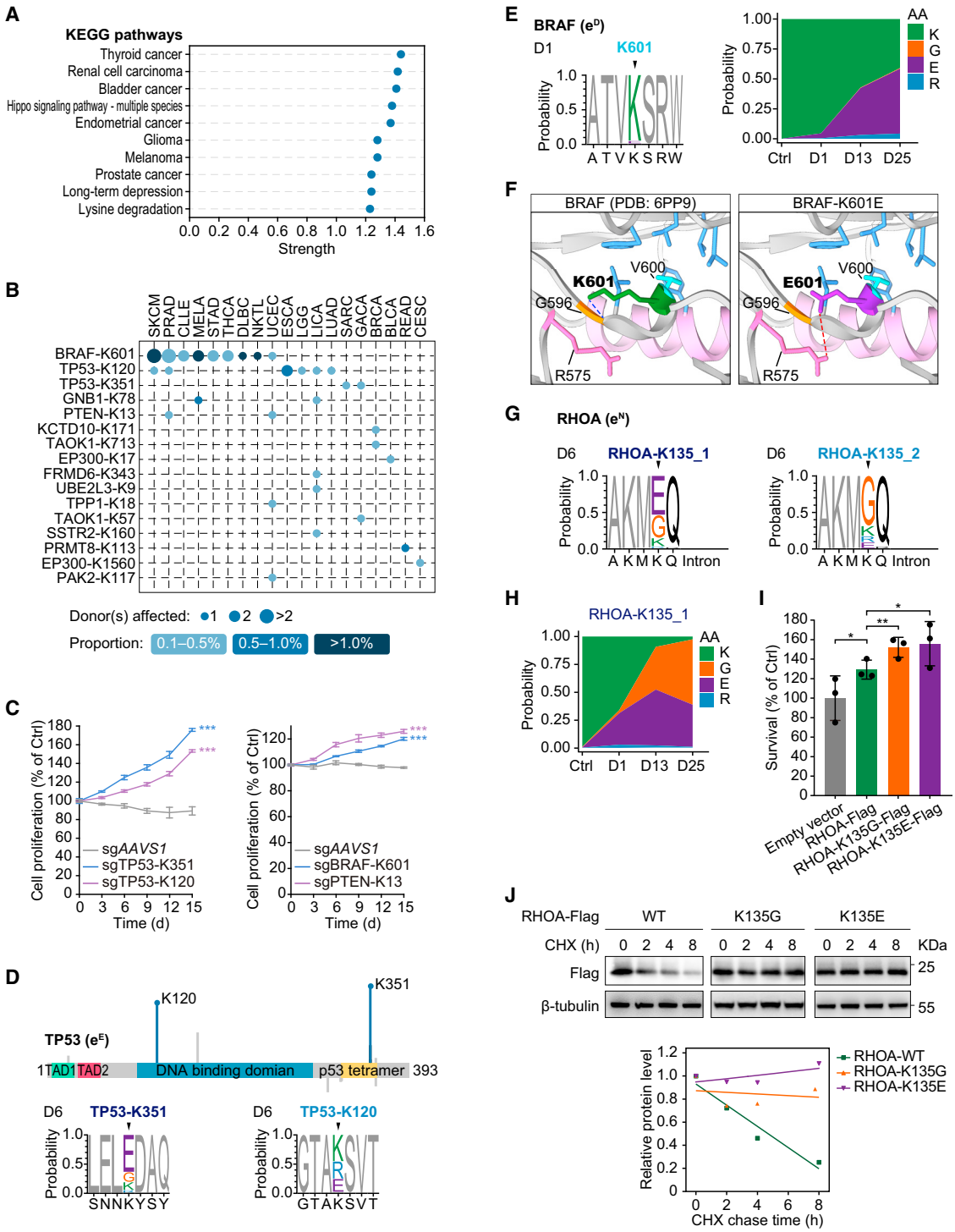


Figure 4. Enriched lysine residues are associated with cancer-related pathways

(A) GO analysis of genes with enriched lysine residues.

(B) Representation of lysine mutations in cancer cases from ICGC data portal.

(C) Validation of indicated sgRNAs targeting lysine residues whose mutations are detected in cancer cases. Mean ± SD, n = 3.

(D) Upper: *TP53* gene structure with FS values of sgRNAs. Lower: sequence logos showing the editing outcomes.

(E) The editing outcomes of sgBRAF-K601. Left: sequence logo of the editing outcomes. Right: NGS tracking of BRAF-K601 mutation distribution.

(legend continued on next page)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

test this postulation, we overexpressed RBM38 in RPE1 cells and performed clonogenic assay. RBM38 overexpression significantly impeded cell proliferation, phenocopying the effect caused by the RBM38-K114 mutation (Figures 3E and S5), and K114R mutation was verified to stabilize RBM38 protein (Figure 3F). Moreover, ubiquitination at RBM38-K114 is detected by independent high-throughput quantitative proteomics.[38,39] These results collectively suggest that the K114 mutation may prevent ubiquitination-mediated degradation of RBM38, leading to RBM38 accumulation and consequent cell death. Additionally, RBM38 reportedly plays a tumor-suppressive role in diverse human cancer types,[40] and its expression is reduced in human breast cancer.[41] Our data provide mechanistic clues for understanding how this critical protein is regulated.

Subsequently, we performed a global bioinformatics analysis to gain a functional overview of depleted mutations. Protein-protein interaction (PPI) analysis by STRING[42] revealed that corresponding proteins of these mutations formed a network comprising several densely connected subnetworks (Figure 3G). Functional analysis suggested that these subnetworks were mainly composed of proteins attributed to several gene ontology (GO) categories, such as ''ribosome biogenesis'' and ''spliceosomal snRNP complex'' (Figures 3G and 3H), both of which are key for cell-fitness regulation. Notably, most of these lysine mutations have not been identified before. Thus, our data provide a rich resource for studying cell-fitness regulation and a repertoire of lethal lysine mutations for drug targeting.

Enriched lysine residue mutations are associated with cancer-related pathways

We then investigated the enriched lysine residues. GO analysis revealed enrichment of cancer-related pathways in genes corresponding to these sites (Figure 4A). After searching International Cancer Genome Consortium (ICGC) database, we identified 16 positively selected lysine residues involved in 19 cancer types (Figure 4B). Some of these sites are found in well-studied oncogenic or tumor suppressor genes. For example, K120 and K351 of the cellular tumor antigen p53, a tumor suppressor whose inactivation promotes cell growth,[43] were high-ranking among the positively selected sites and both were verified (Figures 4C and 4D). Consistently, K120 of p53 is a known acetylation site regulating p53's apoptotic function,[44,45] and K120R mutation inactivates p53[46]; K351 can also be acetylated and plays a role in cell-cycle arrest.[47] Mutations of K120 or K351, detected in multiple clinical tumor samples (Figure 4B), could potentially induce tumorigenesis in humans because they promoted cell growth consistent with *TP53* KO.

Notably, the screen identified K601 as a critical site in the serine/threonine-protein kinase B-raf (BRAF), a MEK activator.[48] Mutations at K601 of BRAF are detected in various cancer types (Figure 4B). Among these mutations, K601E is the third most frequent BRAF mutation, occurring in 5% of melanoma patients,[49] and is reported to disrupt BRAF kinase activity.[50] Consistently, NGS tracking revealed that K601E was the dominant mutation at all time points, suggesting a significant growth advantage for cells carrying K601E mutation compared with K601G or K601R (Figure 4E). Protein structure analysis illustrated that K601, located in the BRAF kinase activation loop, forms a hydrogen bond with G596 to stabilize the autoinhibited conformation of BRAF (Figure 4F, left).[51] Structure simulation of the K601E mutant indicated that E601 potentially interacts with R575, resulting in a minor conformational change (Figure 4F, right). We infer that this confirmation change could disrupt the inactive conformation and constitutively activate BRAF.

Moreover, mutations of K57 and K713 in TAOK1, a key Hippo pathway regulator, are detected in gastric cancer and breast cancer, respectively (Figure 4B). Because Hippo pathway is pivotal for cell proliferation and cancer development,[52–54] we curated all reported components with selected lysines and built a regulatory network (Figure S6A). Mutations of functional sites in major Hippo regulators, such as NF2-K322, MAP4K4-K168, TAOK1-K57, TAOK1-K484, TAOK1-K713, TAOK3-K481, FRMD6-K343, and RHOA-K135, were verified to significantly promote cell growth (Figure S2B [panels 4 and 5], 2H, and S6B). Reportedly, NF2, TAOK1, and TAOK3 positively regulate YAP/TAZ phosphorylation, leading to the cytoplasmic retention of YAP/TAZ and restraint of cell proliferation, whereas RHOA has the opposite effect.[55] Thus, it can be speculated that sgRNAs targeting NF2-K322, TAOK1-K57, TAOK1-K484, TAOK1-K713, and TAOK3-K481 cause loss-of-function mutations, while sgRNAs targeting RHOA-K135 lead to gain-of-function mutations. Next, we investigated mutations at RHOA-K135. Two individual sgRNAs targeting RHOA-K135 predominantly induced K135E and K135G edits, respectively, and both significantly promoted cell proliferation (Figures 4G, 4H, and S6B). Moreover, overexpressing wild-type RHOA promoted cell proliferation, and K135G/E mutation potentiated this effect, implying an increase in RHOA protein level or activity (Figure 4I). Consistently, in the cycloheximide (CHX) chase assay, RHOA carrying K135G/E mutation showed slower degradation than wild type, indicating that these mutations lead to the stabilization and accumulation of RHOA proteins (Figure 4J). Mechanistically, K135G/E mutation substantially reduced K48-linked, but not K63-linked polyubiquitination of RHOA, suggesting that K135 is a critical acceptor for K48-linked ubiquitination, a common PTM-mediating protein degradation (Figure S6C). Thus, our lysine screen identifies K135 as a critical site for tuning the

(F) Schematics representing structure and intramolecular interactions around BRAF-K601 and BRAF-K601E. Left: green stick, K601; blue dashed line, hydrogen bond between K601 and G596; sky-blue sticks, a cluster of hydrophobic residues; bright blue, V600. Right: purple stick, E601; red dashed line, salt bridge between E601 and R575.
(G) Sequence logos showing NGS analysis of sgRHOA-K135_1.
(H) NGS tracking of RHOA-K135 mutation distribution.
(I) Clonogenic assay of RPE1 cells stably overexpressing RHOA or indicated RHOA-K135 mutants. Mean ± SD, n = 3 (see STAR Methods).
(J) Degradation kinetics of RHOA proteins in HEK293T cells (see STAR Methods).
See also Figure S6.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

protein level of RHOA, a regulator of diverse cellular processes with mutations significantly associated with cancer.[56]

To sum up, our screen identifies 76 lysine-targeting sgRNAs that promote cell fitness. Moreover, we postulate potential mechanisms of these mutations by integrating lysine screen data with gene KO screen data, as demonstrated by the identification of loss-of-function mutations in *TP53* (e^E), and gain-of-function mutations in *BRAF* (e^D) and *RHOA* (e^N) (Figure 3A). Because uncontrolled cell growth underlies tumorigenesis, we envision that these positively selected lysine mutations could serve as potential biomarkers in cancer diagnosis, prognosis, and therapy.

Functional analysis revealed a CUL3-centric regulatory network

Next, we mined the screen data to obtain a global functional view of enriched lysine residues. To do this, we built a PPI network followed by GO analysis. A dominant network was created comprising five functional subnetworks, namely, "regulation of small GTPase-mediated signal transduction," "actin cytoskeleton organization," "ubiquitin-dependent protein catabolic process," "regulation of cell cycle," and "histone modification" (Figure 5A).

Notably, CUL3, a core component of cullin-RING ligases (CRLs),[57] was found to be a node connecting multiple subnetworks. CRLs are multisubunit complexes composed of cullin proteins as the scaffold, a RING-containing protein for E2s docking, and multiple adaptors for substrates binding[57] (Figure 5B). Functional lysine residues were identified in CUL3, its BTB domain-containing adaptors (KCTD5, KCTD10, and TNFAIP1),[58] NEDD8 (a ubiquitin-like protein conjugated to CUL3 to regulate its activity),[59] and NAE1 (a NEDD8 activating protein)[60] (Figure 5B).

By CPA, we observed that CUL3-K638-mutant cells exhibited an overwhelming proliferation advantage over wild-type cells (Figure 5C). NGS analysis revealed a predominant K to E mutation at CUL3-K638 site over time (Figures 5C and S3I), indicating the crucial role of K638E mutation. Consistently, CUL3-K638E mutation is identified in one colorectal carcinoma patient,[61] highlighting its physiological significance in carcinogenesis.

We then sought to understand the underlying mechanism by which CUL3-K638E regulates cell proliferation. Given CUL3's role as the scaffold for complex assembly, we speculated that K638E mutation may disrupt CUL3 CRL complex. To investigate this, we conducted affinity purification-mass spectrometry (AP-MS) to detect proteins interacting differently with CUL3 and CUL3-K638E (Figure 5D). With CompPASS algorithm (comparative proteomic analysis software suite),[62] we identified 43 proteins with stronger binding to CUL3 than CUL3-K638E (Z score > 0.70, weighted D (WD) score > 0.70; Figure 5E; Table S6). Among them were Kelch-like (KLHL) proteins, known as putative adaptors for CUL3,[58] and components of COP9 signalosome (CSN), including GPS1, COPS2, COPS3, COPS4, COPS6, and COPS8. NEDD8 conjugation (neddylation) is vital for cullin activation, while CSN removes NEDD8 from activated cullins, rendering them inactive but more stable.[63] We then investigated whether the dissociation of CSN from CUL3-K638E affects its neddylation and/or stability with MLN4924 treatment, an inhibitor of NEDD8 conjugation.[64] In the absence or presence of MLN4924, wild-type CUL3 remained in the unmodified form; nonetheless, a portion of the K638E mutant appeared as a higher band on the immunoblot, which diminished following MLN4924 treatment, indicating that CUL3-K638E is susceptible to neddylation (Figures 5F and S7A). Further, the CHX chase assay and protein stability reporter assay revealed that CUL3-K638E protein had a lower basal level and a faster degradation rate than wild-type CUL3 (Figures 5G and S7B), supporting the notion that neddylated cullins are unstable.[63]

Collectively, K638 of CUL3 is a critical lysine residue for the organization, assembly, and recycling of CUL3-centric CRL complex, and thus regulates CRL-dependent proteostasis (Figure 5H). However, as K638E mutation both increases neddylated CUL3, i.e., an active form, and destabilizes CUL3 CRL complex, it remains to be determined how this mutation regulates substrates stability, downstream pathways, and eventually, uncontrolled cell proliferation.

Cognate lysine residues potentially determine the functional divergence of KCTD10 and TNFAIP1, homologous proteins serving as CUL3 adaptors

In the CUL3 CRL complex, functional lysine residues were enriched in the KCTD family members that serve as substrate adaptors for CUL3, including KCTD10-K171, KCTD10-K237, KCTD10-K241, TNFAIP1-K168, TNFAIP1-K234, TNFAIP1-K235, and KCTD5-K84 (Figure 5B). The KCTD family consists of 26 members, the majority of which are thought to be involved in numerous pathological processes, including tumorigenesis.[65] The phylogenetic tree of the KCTD family reveals that KCTD10 and TNFAIP1 are clustered in the same clade, suggesting their close evolutionarily relationship (Figure 6A). Intriguingly, two pairs of cognate lysine sites of these two genes were identified in the screen: KCTD10-K171 vs. TNFAIP1-K168 and KCTD10-K237 vs. TNFAIP1-K234. As homologous genes usually possess high sequence similarity, possible off-target editing by ABEmax could cause mutations in the non-targeted gene of the gene pair. However, NGS analysis revealed predominant K to E mutations at both KCTD10-K171 and TNFAIP1-K168, with no evidence of cross-editing in either gene (Figure 6B).

Further, we identified multiple pairs of cognate lysine residues in homologous proteins, which were validated by individual sgRNAs (Figures 6C and S2B [panels 4, 6, 17, and 7]). Thus, it is tempting to speculate that these cognate lysine residues underpin divergent functionalities of paralogs. To test this, we investigated the functional divergence between KCTD10-K171 and TNFAIP1-K168 in regulating protein stability (Figure 6D) since CUL3 CRLs commonly target substrates for degradation.[66] By comparing the proteomes of KCTD10-K171E and TNFAIP1-K168E cells to control cells, we identified 451 and 104 differentially regulated proteins, respectively, implying that KCTD10-K171E induced more profound changes in the cellular proteome than TNFAIP1-K168E (Figure 6E; Table S7). Intersection analysis revealed a minor portion of proteins regulated by both mutants (Figure 6F). Gene set enrichment analysis (GSEA) revealed that upregulated proteins in KCTD10-K171E cells were enriched in cell-cycle regulation (Figure 6G); however, GSEA failed to detect significant functional enrichment for proteins upregulated in TNFAIP1-K168E cells. GO analysis showed that proteins upregulated in TNFAIP1-K168E cells were involved

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

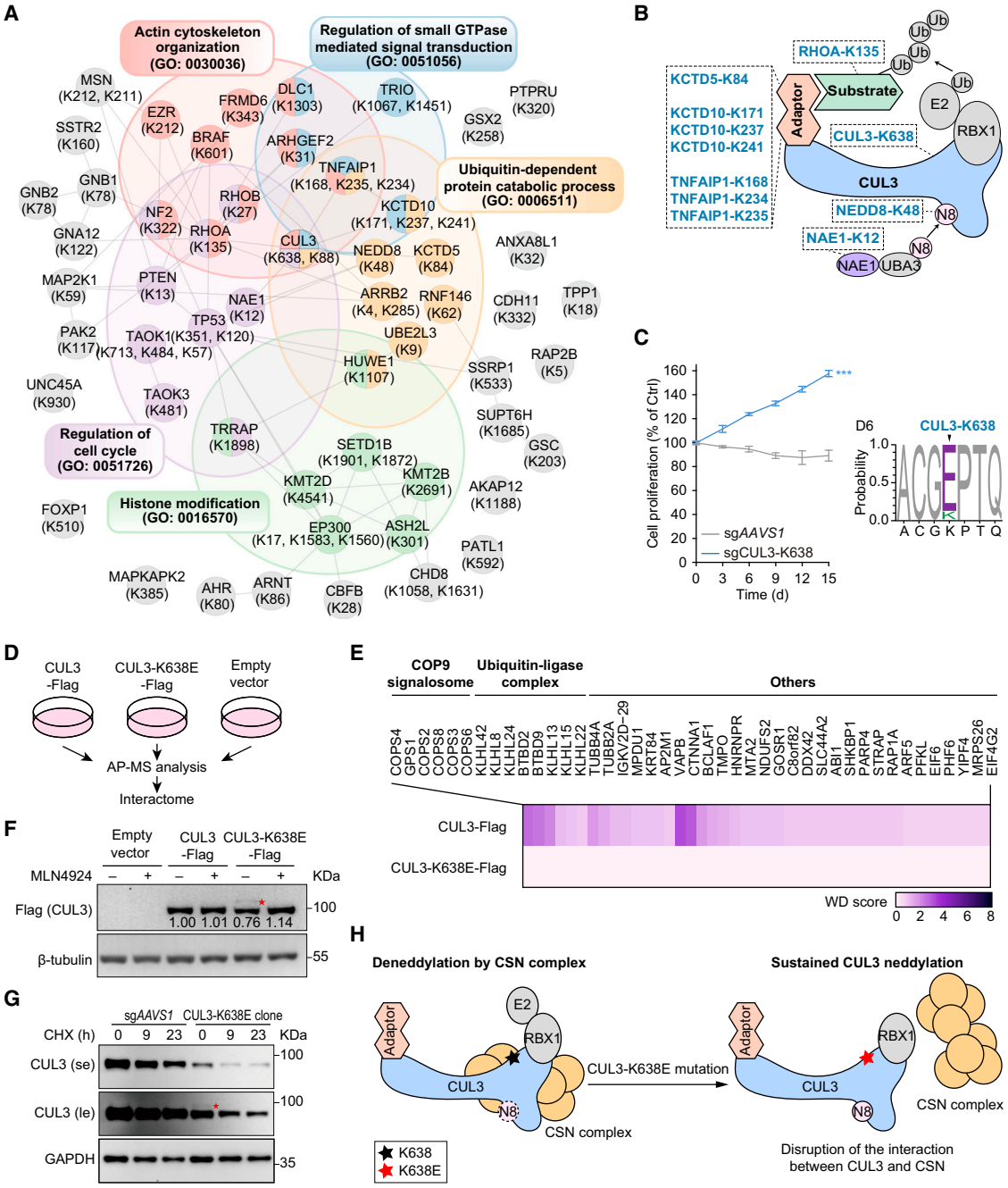


Figure 5. CUL3-centric regulatory network in the enrichment direction

(A) PPI network of genes with enriched lysine residues. Genes are grouped and colored based on the terms of Biological Process domain in the GO KnowledgeBase. Genes in gray circles are not enriched in any GO terms.

(B) Schematic showing the main components of CUL3 CRL complex.

(C) Validation of sgCUL3-K638 by CPA (left) and NGS analysis of the editing outcomes (right). Mean ± SD, n = 3.

(D) Workflow of AP-MS analysis of CUL3 and CUL3-K638E (see STAR Methods).

(E) Heatmap showing WD scores of indicated proteins in the AP-MS analysis indicated in (D).

(F) Immunoblot analysis of CUL3 level in RPE1 cells stably expressing CUL3-FLAG or CUL3-K638E-FLAG. The red star indicates neddylated CUL3.

(G) Immunoblot analysis of the degradation kinetics of CUL3 proteins in RPE1 cells. The red star indicates neddylated CUL3. Se: short-time exposure; le: long-time exposure.

(H) Proposed model illustrating the impact of CUL3-K638E mutation on CUL3 CRL complex turnover.

See also Figure S7.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

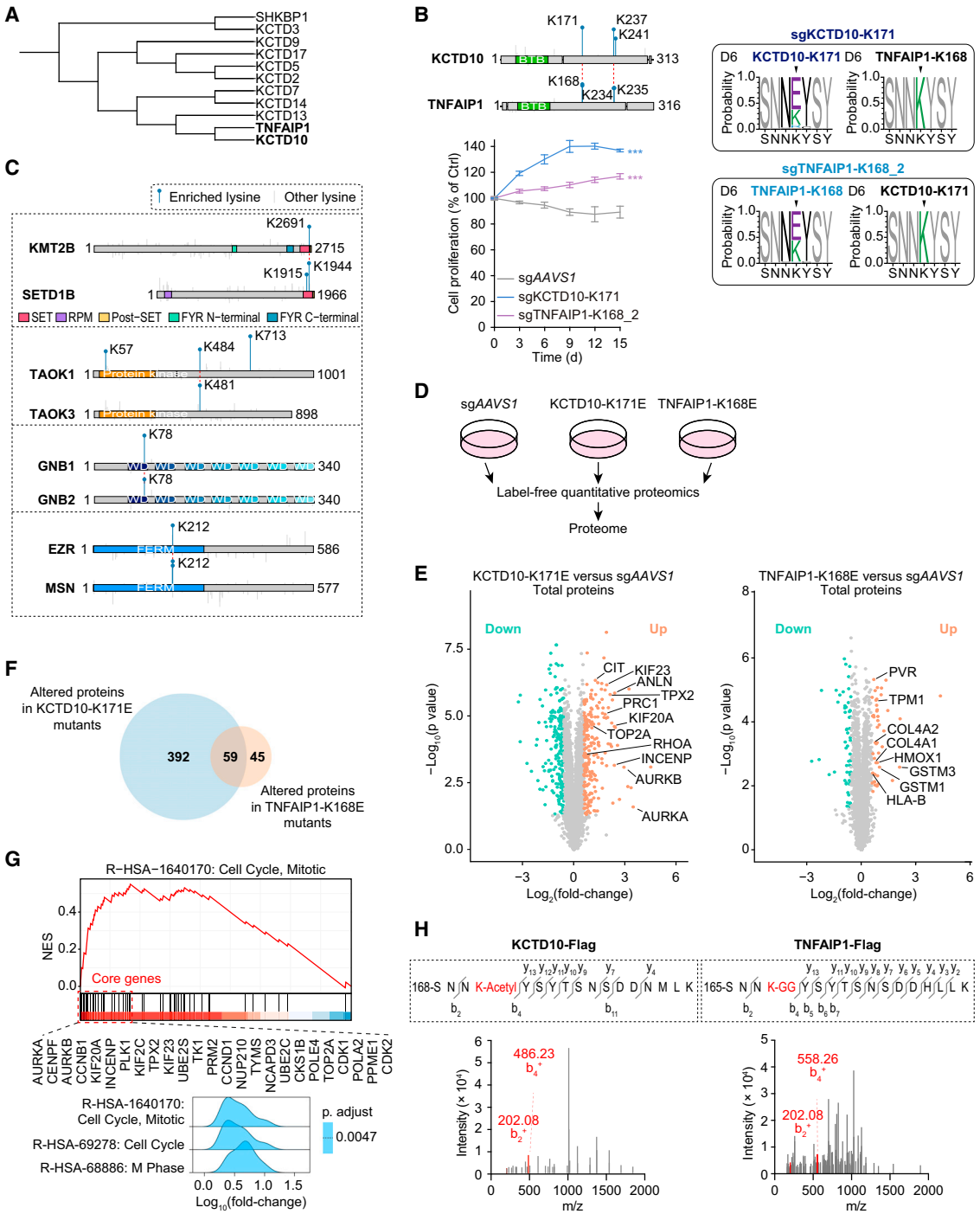


Figure 6. Critical lysine residues underlie functional divergence of KCTD10 and TNFAIP1

(A) A phylogenetic tree of proteins closely related to KCTD10 and TNFAIP1 within the KCTD family.

(B) Validation of paired sgRNAs targeting cognate lysine codons in *KCTD10* and *TNFAIP1*. Upper-left: *KCTD10* and *TNFAIP1* gene structures with FS values of sgRNAs. Lower-left: CPA plots. Right: editing outcomes of sgKCTD10-K171 and sgTNFAIP1-K168_2 at both loci. Mean ± SD, n = 3.

(C) Bar plots with diagrams of gene structures showing FS values of sgRNAs targeting cognate lysine codons of homologous genes.

(D) Workflow of label-free mass spectrometry for proteome analysis (see STAR Methods).

(E) Volcano plots showing protein abundance changes in clonal cell lines with homozygous KCTD10-K171E mutation (left) or TNFAIP1-K168E mutation (right) compared with control cells (see STAR Methods). N = 3.

(legend continued on next page)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

in cellular detoxification processes, disparate from those observed in KCTD10-K171E cells (Figure S8). These results collectively suggest that KCTD10-K171 and TNFAIP1-K168 regulate cell proliferation by regulating the abundance of proteins in different pathways.

Because PTMs could mediate PPIs,[67] we further investigated whether KCTD10-K171 or TNFAIP1-K168 undergoes PTMs. According to mass spectrometry analysis, KCTD10-K171 was found to be acetylated, whereas TNFAIP1-K168 was found to be ubiquitinated (Figure 6H; Table S8). This suggests that KCTD10-K171 or TNFAIP1-K168 may employ different PTMs to determine substrates binding. Collectively, we reason that cognate lysine residues may play crucial roles in the functional divergence of homologous proteins during evolution.

KCTD10-K171 is critical for CUL3 CRL-mediated degradation of mitotic proteins

Because mutations at KCTD10-K171 are clinically relevant (Figure 4B), we further scrutinized the mechanism by which KCTD10-K171 regulates cell proliferation. The proteomics study revealed that KCTD10-K171E mutation causes the accumulation of proteins involved in cell-cycle regulation (Figures 6G and S9A); therefore, we examined whether cell cycle is dysregulated in KCTD10-K171E cells. To do this, we generated five individual RPE1 cell clones harboring homozygous KCTD10-K171E mutation and determined their cell-cycle profiles. All five clones showed significantly higher percentages of cells within S and G2/M phases than control cells (Figures 7A and S9B). Moreover, KCTD10-K171E cells showed markedly higher level of CCNB1 than controls, which is predominantly expressed in G2/M phase[68] (Figures S9C and S9D). These data suggest that KCTD10-K171E mutation increases the number of cycling cells and/or the rate of cell-cycle progression.

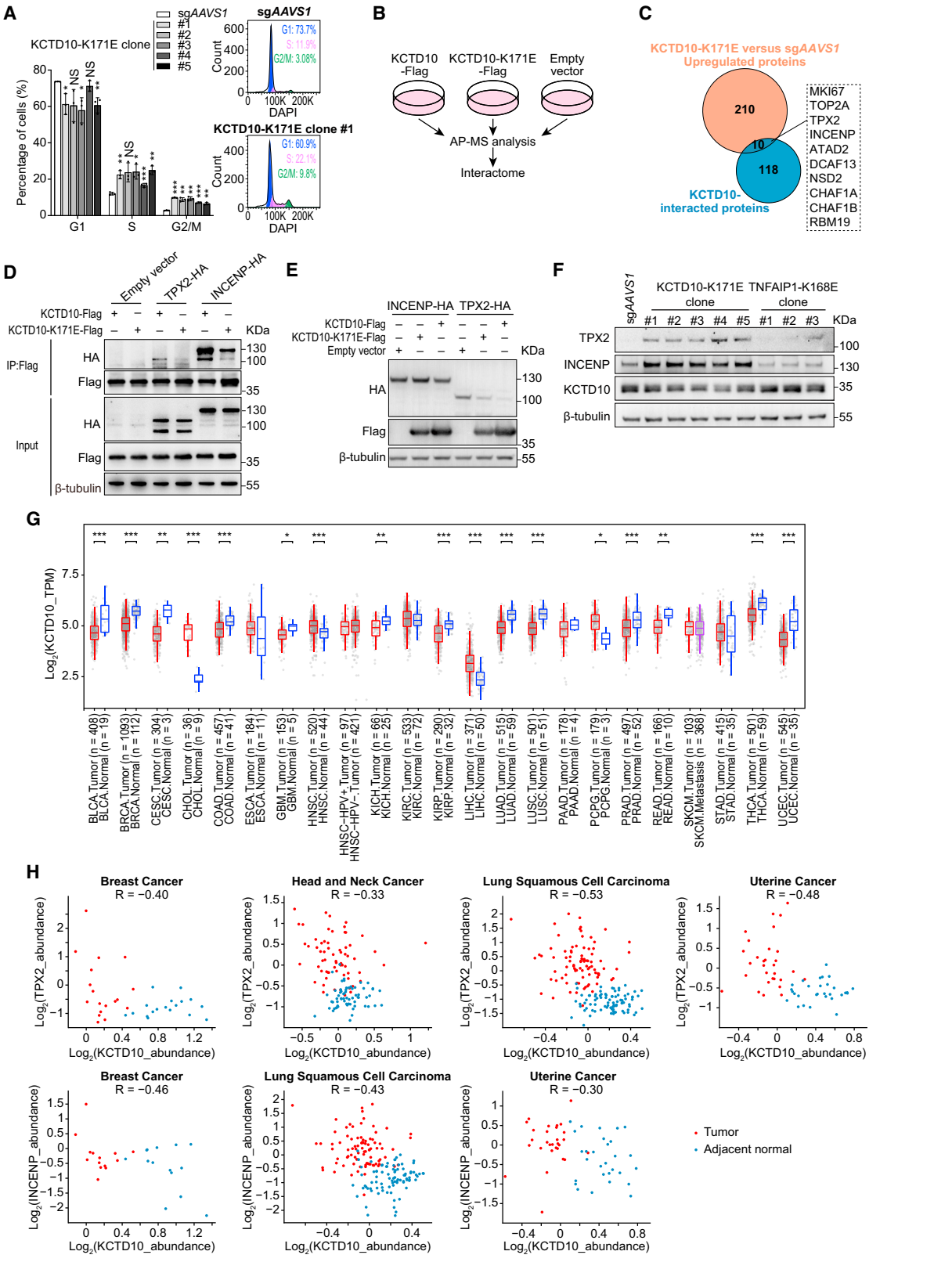
Next, we systematically investigated proteins whose binding to KCTD10 depend on K171. Through AP-MS (Figure 7B) and subsequent CompPASS analysis, we identified 128 proteins that preferentially interacted with KCTD10 over KCTD10-K171E mutant (Z score > 0.75, WD score > 0.70; Table S9). We then intersected these proteins with the 220 proteins upregulated in KCTD10-K171E cells to identify bona fide CUL3 CRL substrates whose degradation depend on KCTD10-K171. This intersection identified ten proteins, namely, MKI67, TOP2A, TPX2, INCENP, ATAD2, DCAF13, NSD2, CHAF1A, CHAF1B, and RBM19 (Figure 7C). Notably, nine of these proteins are implicated in cancer development,[69–77] and among them, MKI67, TOP2A, TPX2, INCENP, ATAD2, NSD2, CHAF1A, and CHAF1B are known to regulate cell cycle.[70,71,78–82] Particularly, TPX2 and INCENP are well-known interaction partners of AURKA and AURKB, respectively, which are key mitotic kinases.[83,84] Immunoprecipitation assay confirmed the interaction between KCTD10 and TPX2 or INCENP, whereas the K171E mutation disrupted both interactions (Figure 7D). Further, TPX2 could be ubiquitinated by CUL3 in the presence of KCTD10 but not the K171E mutant (Figure S10A). Consistently, overexpression of KCTD10 decreased TPX2 and INCENP protein levels, while the K171E mutant had no or reduced effects (Figure 7E). In line with this, all five KCTD10-K171E clones, but not the TNFAIP1-K168E clones, exhibited higher levels of TPX2 and INCENP than control cells (Figure 7F). These data collectively suggest that KCTD10 functions as an adaptor, bringing TPX2 and INCENP to CUL3 CRL for degradation, with K171 mediating this process.

Then, we examined the impact of TPX2 or INCENP upregulation on cell proliferation. ATP measurement-based quantitation of viable cells revealed that TPX2 overexpression significantly promoted cell proliferation. INCENP also exhibited pro-proliferative potential, albeit to a lesser extent, highlighting their oncogenic potential (Figure S10B). Differential expression analysis between tumor and adjacent normal tissues revealed significant overexpression of TPX2 in 21 cancer types[85] (Figure S10C). In EAC, KIRC, KIRP, LIHC, LUAD, PDAC, and UCEC, high TPX2 expression is associated with poor prognosis[86] (Figures S10C and S10D). Heightened expression of INCENP is also linked to 13 cancer types (Figure S10E). Intriguingly, RNA levels of KCTD10 are lower in tumors than in adjacent normal tissues across 13 cancer types (Figure 7G). Moreover, at the protein level, tumor tissues exhibit lower KCTD10 abundance but higher TPX2/INCENP abundance, whereas adjacent normal tissues show the opposite trend, suggesting that KCTD10 may act as a safeguard against tumorigenesis by preventing aberrantly high levels of TPX2 and INCENP[87] (Figure 7H). Overall, KCTD10 plays a critical role in controlling normal cell proliferation by targeting TPX2 and INCENP for degradation.

We subsequently explored downstream effectors of the CUL3-KCTD10-TPX2/INCENP axis controlling cell proliferation. TPX2 and INCENP could activate and stabilize AURKA and AURKB, which are inherently unstable.[88–90] Intriguingly, AURKA and AURKB were upregulated in KCTD10-K171E mutants according to mass spectrometry analysis (Figure S9A). Consistently, all five KCTD10-K171E clones exhibited higher abundance of AURKA and AURKB than controls (Figure S10F). Thus, it is speculated that KCTD10-K171E mutation caused the accumulation of TPX2 and INCENP, preventing the rapid degradation of AURKA and AURKB. Indeed, overexpression of TPX2 and INCENP markedly increased the abundance of AURKA and AURKB, respectively (Figure S10G). *AUKRA* and *AUKRB* are well-known oncogenes with high expression in various cancer types.[90] Consistently, AURKA or AURKB overexpression significantly promoted cell proliferation (Figure S10H). Therefore, by regulating TPX2 and INCENP stability, KCTD10-K171 can indirectly modulate AURKA and AURKB kinase activity and protein abundance, thereby fine-tuning cell proliferation to prevent tumorigenesis.

(F) Venn diagram showing the intersection of proteins with altered abundance in cell lines with homozygous KCTD10-K171E mutation and TNFAIP1-K168E mutation.
(G) GSEA analysis of upregulated proteins in cell lines with homozygous KCTD10-K171E mutation.
(H) Identification of acetylation at KCTD10-K171 residue (left) and ubiquitination at TNFAIP1-K168 residue (right).
See also Figure S8.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033



(legend on next page)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

DISCUSSION

Here, we present a strategy for the genome-wide screen of functional AA residues. Unlike conventional CRISPR KO screens, our strategy could pinpoint functional residues of proteins, offering direct indications for mechanistic studies. Besides, because AA substitutions could either decrease or increase protein abundance, and inactivate or activate protein function, our strategy allows for both gain-of-function and loss-of-function perturbations in a single screen. Moreover, our approach could detect mutations that bestow unique activities to proteins. These mutations may be implicated in pathogenic processes, while their corresponding genes could not be identified by either CRISPR KO or CRISPRa screens. Thus, this approach provides unprecedented opportunities to gain functional insights into human proteomes.

Protein PTMs add an important layer of complexity to the proteome.[91] Advances in mass spectrometry technology have now enabled the discovery of thousands of PTMs[92]; however, it remains to be determined which PTMs are functional at the system level. This study focuses on one particular AA—lysine, primarily because lysine residues are the major receptors for numerous important PTMs. Indeed, we identify that K135 of RHOA receives degradative ubiquitination, whereas K171 of KCTD10 undergoes acetylation that might mediate PPI. Meanwhile, we identify multiple lysine residues that operate on protein structure; for incidence, K601 of BRAF functions through interacting with adjacent residues. Through this screening, we are able to generate informative annotations regarding the functionality of lysine residues in cellular fitness. Moving forward, this method could be extended to investigate other residues using ABEs or CBEs, or it could be carried out in a specific biological setting.

The constellation of functional lysine residues could provide a rich resource for biomedical research. Positively selected lysine residues are enriched in cancer-associated pathways, and many of their mutations have been detected in cancer patient samples, implying their functional associations with cancer. Notably, a plethora of positively selected residues are found in CUL3-centric complex. Recent research points to a crucial function for CUL3 in tumorigenesis by demonstrating that CUL3 loss could promote persistent proliferation in TP53-deficient cells.[93] Consistently, our data demonstrate that CUL3-K638E significantly promotes cell proliferation, possibly by destabilizing CUL3, and this mutation is detected in one colorectal carcinoma patient.[61] KCTD10 is an adaptor for CUL3 that promotes the ubiquitination of multiple proteins, including RHOB, CEP97, EIF3D, and TRIF. Here, we identify TPX2 and INCENP as substrates for KCTD10, whose binding depends on KCTD10-K171. Interestingly, pathogenic mutations of another CUL3 adaptor, KBTBD4, could enable CoREST complex recruitment, thereby driving epigenetic reprogramming in medulloblastoma.[94] Therefore, specific residues in adaptor proteins may govern substrates specificity, and their mutations potentially cause diseases. Indeed, KCTD10-K171 mutations are identified in samples from breast cancer patients (Figure 4B).

We also identify an assembly of negatively selected lysine residues that warrant further investigation. These lysine mutations are of particular interest because they cause cell growth arrest or cell death, making them possible targets for cancer drugs. Covalent drugs have been developed to target non-catalytic residues in functional sites of proteins.[95] More recently, Abbasov et al. mapped the interactions of small covalent molecules with lysine residues, creating a proteome-wide atlas of lysine-reactive chemistry.[96] Thus, we envision that combining screens of functional residues with covalent chemistry technologies will open new avenues for identifying cancer drug targets, particularly those deemed "undruggable." Moreover, these lysine residues could be targeted by genome editing with CRISPR-Cas9 or "LEAPER,"[97,98] a therapeutic RNA editing tool that converts A to I (G) in gene transcripts.

Limitations of the study

BE-based screens may come with certain limitations compared with conventional CRISPR KO screens. To begin with, in gene KO screens, multiple sgRNAs are usually employed to target a single gene, and a functional score for the gene can be calculated based on these sgRNAs. However, in BE screens, the availability of only one or two sgRNAs for each target site, limited by PAM constraints and editing windows, can make the assessment of sgRNA perturbation effect and the functionality of the targeted site more challenging. In our current study, we employ iBAR technology along with the specifically devised ZFC^{iBAR} algorithm to evaluate the effect of each sgRNA. This strategy could help improve BE screen quality even when working with a limited number of sgRNAs for a given target site.

Second, due to the PAM restrictions, our current screen addresses only 35% of lysine codons within the coding genome. Therefore, it is important to acknowledge that lysine residues with no observed phenotypes might still possess functionality. In the future, untargeted lysines could be explored using "NG" PAM targetable BEs,[99] "PAMless" BEs,[100] or Prime editors,[101,102] allowing for a more comprehensive screening approach.

Figure 7. KCTD10-K171 regulates cell proliferation by controlling mitotic protein stability

(A) Bar plot showing the cell-cycle profiles of homozygous KCTD10-K171E mutants. Representative histograms are shown in the right panel. Mean ± SD, n = 3. P values: unpaired one-tailed Student's t test with Benjamini-Hochberg adjustment.

(B) Workflow of AP-MS analysis of KCTD10 and KCTD10-K171E (see STAR Methods).

(C) Venn diagram showing the intersection of proteins with elevated abundance and reduced interaction with KCTD10 upon K171E mutation.

(D) Co-immunoprecipitation assay showing interactions between INCENP or TPX2 and KCTD10 or KCTD10-K171E in HEK293T cells.

(E) Immunoblot analysis of INCENP and TPX2 levels in HEK293T cells transfected with KCTD10-Flag or KCTD10-K171E-FLAG plasmids.

(F) Immunoblot analysis of TPX2 and INCENP levels in clonal cell lines with homozygous KCTD10-K171E or TNFAIP1-K168E mutation.

(G) Differential expression analysis of human KCTD10 between tumor and adjacent normal tissues using TIMER2.0. P value is calculated by Wilcoxon test. TPM: transcripts per million.

(H) Correlation analysis of protein abundance between KCTD10 and TPX2 or INCENP in cancer patients using cProSite. The correlation coefficient (R value) is calculated using Pearson correlation coefficient. TMT: tandem mass tags.

See also Figures S9 and S10.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

Third, we observe bystander edits adjacent to the targeted lysine codons due to the susceptibility of all editable adenines within the "activity window" to editing, a common limitation of BE-based screens.[7–10] This could potentially lead to misattribution of targeted lysine residues, resulting in false positives. To address this issue, some studies have incorporated "sensor sequences" (sequences matching the sgRNA-targeting sites) into their sgRNA constructs for BE screens, enabling the validation of edits in a high-throughput manner.[9,103] This approach helps filter out false positives caused by bystander edits, especially when the edits affect the surrounding codons rather than the targeted lysine codons. Additionally, NGS tracking could be employed to accurately identify functional mutations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines and culture
- METHOD DETAILS
  - Plasmid construction
  - SgRNA library design
  - Library production
  - Functional lysine screening in RPE1 cells
  - Cell proliferation assay
  - Editing efficiency detection by NGS analysis
  - Construction of clonal cells with homozygous lysine mutations
  - Immunoblotting
  - Clonogenic assay
  - ATP measurement-base cell proliferation assay
  - CHX chase assay
  - Immunoprecipitation
  - Ubiquitination assay
  - Label-free quantitative proteomics
  - AP-MS analysis
  - Cell cycle profiling and synchronization
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data analysis
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.10.033.

ACKNOWLEDGMENTS

This project was funded by the the National Key R&D Program of China (2020YFA0707600 to Z. Zhou), National Natural Science Foundation of China (grant nos. NSFC31870893 and NSFC82221004 to Z. Zhou, NSFC31930016 to W.W.), the Beijing Advanced Innovation Center for Genomics at Peking University (PKU), the Peking-Tsinghua Center for Life Sciences (W.W.), the Non-profit Central Research Institute Fund of the Chinese Academy of Medical Sciences (2023-RC310-01), CAMS Innovation Fund for Medical Sciences (2022-I2M-2-004, 2022-I2M-1-024), the State Key Laboratory Special Fund 2060204, the NCTIB Fund for R&D platform for Cell and Gene Therapy, and the Suzhou Municipal Key Laboratory (SZS2022005 to Z. Zhou). We thank Y. Sun and C. Zhang (PKU) for cell lines, the BIOPIC High-Throughput Sequencing Center for NGS analysis, the National Center for Protein Sciences (PKU) for flow cytometry and mass-spectrometry support (particularly H. Yang, H. Lyu, and D. Liu), and PKU's High-performance Computing Platform for technical support.

AUTHOR CONTRIBUTIONS

Conceptualization, W.W. and Z. Zhou; methodology, P.X., Z.L., Y.B., and Q.P.; software, Z.L.; validation, Y.B., Q.P., and P.X.; formal analysis, Z.L., Y.B., Q.P., and P.X.; investigation, Y.B., Q.P., P.X., Z. Zhang, Y.L., Y.X., and Y.Y.; writing – original draft, Z. Zhou, P.X., Y.B., Q.P., and Z.L.; writing – review & editing, Z. Zhou, Y.B., Q.P., and P.X.; visualization, Q.P., Y.B., P.X., and Z.L.; supervision, W.W. and Z. Zhou.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 6, 2022
Revised: September 6, 2023
Accepted: October 23, 2023
Published: November 22, 2023

REFERENCES

1. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755. https://doi.org/10.1038/nrg3031.
2. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471. https://doi.org/10.1038/nature24644.
3. Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., and Liu, D.R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. *36*, 843–846. https://doi.org/10.1038/nbt.4172.
4. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424. https://doi.org/10.1038/nature17946.
5. Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods *13*, 1029–1035. https://doi.org/10.1038/nmeth.4027.
6. Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788. https://doi.org/10.1038/s41576-018-0059-1.
7. Cuella-Martin, R., Hayward, S.B., Fan, X., Chen, X., Huang, J.W., Taglialatela, A., Leuzzi, G., Zhao, J., Rabadan, R., Lu, C., et al. (2021). Functional interrogation of DNA damage response variants with base editing screens. Cell *184*, 1081–1097.e19. https://doi.org/10.1016/j.cell.2021.01.041.
8. Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

screens. Cell *184*, 1064–1080.e20. https://doi.org/10.1016/j.cell.2021.01.012.

9. Kim, Y., Lee, S., Cho, S., Park, J., Chae, D., Park, T., Minna, J.D., and Kim, H.H. (2022). High-throughput functional evaluation of human cancer-associated mutations using base editors. Nat. Biotechnol. *40*, 874–884. https://doi.org/10.1038/s41587-022-01276-4.
10. Lue, N.Z., Garcia, E.M., Ngan, K.C., Lee, C., Doench, J.G., and Liau, B.B. (2023). Base editor scanning charts the DNMT3A activity landscape. Nat. Chem. Biol. *19*, 176–186. https://doi.org/10.1038/s41589-022-01167-4.
11. Perner, F., Stein, E.M., Wenge, D.V., Singh, S., Kim, J., Apazidis, A., Rahnamoun, H., Anand, D., Marinaccio, C., Hatton, C., et al. (2023). MEN1 mutations mediate clinical resistance to menin inhibition. Nature *615*, 913–919. https://doi.org/10.1038/s41586-023-05755-9.
12. Brocchieri, L., and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res. *33*, 3390–3400. https://doi.org/10.1093/nar/gki615.
13. Doench, J.G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. Nat. Rev. Genet. *19*, 67–80. https://doi.org/10.1038/nrg.2017.97.
14. Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. Nat. Rev. Genet. *16*, 299–311. https://doi.org/10.1038/nrg3899.
15. Zhu, S., Cao, Z., Liu, Z., He, Y., Wang, Y., Yuan, P., Li, W., Tian, F., Bao, Y., and Wei, W. (2019). Guide RNAs with embedded barcodes boost CRISPR-pooled screens. Genome Biol. *20*, 20. https://doi.org/10.1186/s13059-019-1628-0.
16. Xu, P., Liu, Z., Liu, Y., Ma, H., Xu, Y., Bao, Y., Zhu, S., Cao, Z., Wu, Z., Zhou, Z., et al. (2021). Genome-wide interrogation of gene functions through base editor screens empowered by barcoded sgRNAs. Nat. Biotechnol. *39*, 1403–1413. https://doi.org/10.1038/s41587-021-00944-1.
17. Flodin, N.W. (1997). The metabolic roles, pharmacology, and toxicology of lysine. J. Am. Coll. Nutr. *16*, 7–21. https://doi.org/10.1080/07315724.1997.10718644.
18. Clague, M.J., and Urbé, S. (2010). Ubiquitin: same molecule, different degradation pathways. Cell *143*, 682–685. https://doi.org/10.1016/j.cell.2010.11.012.
19. Ali, I., Conrad, R.J., Verdin, E., and Ott, M. (2018). Lysine acetylation goes global: from epigenetics to metabolism and therapeutics. Chem. Rev. *118*, 1216–1252. https://doi.org/10.1021/acs.chemrev.7b00181.
20. Luo, M. (2018). Chemical and biochemical perspectives of protein lysine methylation. Chem. Rev. *118*, 6656–6705. https://doi.org/10.1021/acs.chemrev.8b00008.
21. Johnson, E.S. (2004). Protein modification by SUMO. Annu. Rev. Biochem. *73*, 355–382. https://doi.org/10.1146/annurev.biochem.73.011303.074118.
22. Hirschey, M.D., and Zhao, Y. (2015). Metabolic regulation by lysine malonylation, succinylation, and glutarylation. Mol. Cell. Proteomics *14*, 2308–2315. https://doi.org/10.1074/mcp.R114.046664.
23. Ryu, S.M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.T., Kim, H.S., Kim, D.E., Lee, H., Chung, E., et al. (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat. Biotechnol. *36*, 536–539. https://doi.org/10.1038/nbt.4148.
24. Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. Science *279*, 349–352. https://doi.org/10.1126/science.279.5349.349.
25. Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics *28*, 573–580. https://doi.org/10.1093/bioinformatics/btr709.
26. Lee, T.Y., Huang, H.D., Hung, J.H., Huang, H.Y., Yang, Y.S., and Wang, T.H. (2006). dbPTM: an information repository of protein post-translational modification. Nucleic Acids Res. *34*, D622–D627. https://doi.org/10.1093/nar/gkj083.
27. Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784. https://doi.org/10.1038/ng.3984.
28. O'Hayre, M., Vázquez-Prado, J., Kufareva, I., Stawiski, E.W., Handel, T.M., Seshagiri, S., and Gutkind, J.S. (2013). The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. Nat. Rev. Cancer *13*, 412–424. https://doi.org/10.1038/nrc3521.
29. Sykes, S.M., Mellert, H.S., Holbert, M.A., Li, K., Marmorstein, R., Lane, W.S., and McMahon, S.B. (2006). Acetylation of the p53 DNA-binding domain regulates apoptosis induction. Mol. Cell *24*, 841–851. https://doi.org/10.1016/j.molcel.2006.11.026.
30. Trotman, L.C., Wang, X., Alimonti, A., Chen, Z., Teruya-Feldstein, J., Yang, H., Pavletich, N.P., Carver, B.S., Cordon-Cardo, C., Erdjument-Bromage, H., et al. (2007). Ubiquitination regulates PTEN nuclear import and tumor suppression. Cell *128*, 141–156. https://doi.org/10.1016/j.cell.2006.11.040.
31. Thompson, N.A., Ranzani, M., van der Weyden, L., Iyer, V., Offord, V., Droop, A., Behan, F., Gonçalves, E., Speak, A., Iorio, F., et al. (2021). Combinatorial CRISPR screen identifies fitness effects of gene paralogues. Nat. Commun. *12*, 1302. https://doi.org/10.1038/s41467-021-21478-9.
32. Dede, M., McLaughlin, M., Kim, E., and Hart, T. (2020). Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. Genome Biol. *21*, 262. https://doi.org/10.1186/s13059-020-02173-2.
33. Pettinger, J., Jones, K., and Cheeseman, M.D. (2017). Lysine-targeting covalent inhibitors. Angew. Chem. Int. Ed. Engl. *56*, 15200–15209. https://doi.org/10.1002/anie.201707630.
34. Hacker, S.M., Backus, K.M., Lazear, M.R., Forli, S., Correia, B.E., and Cravatt, B.F. (2017). Global profiling of lysine reactivity and ligandability in the human proteome. Nat. Chem. *9*, 1181–1190. https://doi.org/10.1038/nchem.2826.
35. Nakazawa, Y., Hara, Y., Oka, Y., Komine, O., van den Heuvel, D., Guo, C., Daigaku, Y., Isono, M., He, Y., Shimada, M., et al. (2020). Ubiquitination of DNA damage-stalled RNAPII promotes transcription-coupled repair. Cell *180*, 1228–1244.e24. https://doi.org/10.1016/j.cell.2020.02.010.
36. Rein, H.L., Bernstein, K.A., and Baldock, R.A. (2021). RAD51 paralog function in replicative DNA damage and tolerance. Curr. Opin. Genet. Dev. *71*, 86–91. https://doi.org/10.1016/j.gde.2021.06.010.
37. Vasileva, A., Hopkins, K.M., Wang, X., Weisbach, M.M., Friedman, R.A., Wolgemuth, D.J., and Lieberman, H.B. (2013). The DNA damage checkpoint protein RAD9A is essential for male meiosis in the mouse. J. Cell Sci. *126*, 3927–3938. https://doi.org/10.1242/jcs.126763.
38. Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J., et al. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol. Cell *44*, 325–340. https://doi.org/10.1016/j.molcel.2011.08.025.
39. Boeing, S., Williamson, L., Encheva, V., Gori, I., Saunders, R.E., Instrell, R., Aygün, O., Rodriguez-Martinez, M., Weems, J.C., Kelly, G.P., et al. (2016). Multiomic analysis of the UV-induced DNA damage response. Cell Rep. *15*, 1597–1610. https://doi.org/10.1016/j.celrep.2016.04.047.
40. Zou, C., Wan, Y., He, L., Zheng, J.H., Mei, Y., Shi, J., Zhang, M., Dong, Z., and Zhang, D. (2021). RBM38 in cancer: role and mechanism. Cell. Mol. Life Sci. *78*, 117–128. https://doi.org/10.1007/s00018-020-03593-w.
41. Léveillé, N., Elkon, R., Davalos, V., Manoharan, V., Hollingworth, D., Oude Vrielink, J.O., le Sage, C., Melo, C.A., Horlings, H.M., Wesseling, J., et al. (2011). Selective inhibition of microRNA accessibility by RBM38 is required for p53 activity. Nat. Commun. *2*, 513. https://doi.org/10.1038/ncomms1519.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

42. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. *49*, D605–D612. https://doi.org/10.1093/nar/gkaa1074.

43. Levine, A.J. (1997). p53, the cellular gatekeeper for growth and division. Cell *88*, 323–331. https://doi.org/10.1016/S0092-8674(00)81871-1.

44. Tang, Y., Luo, J., Zhang, W., and Gu, W. (2006). Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. Mol. Cell *24*, 827–839. https://doi.org/10.1016/j.molcel.2006.11.021.

45. Tang, Y., Zhao, W., Chen, Y., Zhao, Y., and Gu, W. (2008). Acetylation is indispensable for p53 activation. Cell *133*, 612–626. https://doi.org/10.1016/j.cell.2008.03.025.

46. Lee, S.Y., Park, J.H., Jeong, S., Kim, B.Y., Kang, Y.K., Xu, Y., and Chung, S.K. (2019). K120R mutation inactivates p53 by creating an aberrant splice site leading to nonsense-mediated mRNA decay. Oncogene *38*, 1597–1610. https://doi.org/10.1038/s41388-018-0542-3.

47. Beckerman, R., Yoh, K., Mattia-Sansobrino, M., Zupnick, A., Laptenko, O., Karni-Schmidt, O., Ahn, J., Byeon, I.J., Keezer, S., and Prives, C. (2016). Lysines in the tetramerization domain of p53 selectively modulate G1 arrest. Cell Cycle *15*, 1425–1438. https://doi.org/10.1080/15384101.2016.1170270.

48. Köhler, M., and Brummer, T. (2016). B-Raf activation loop phosphorylation revisited. Cell Cycle *15*, 1171–1173. https://doi.org/10.1080/15384101.2016.1159111.

49. Marconcini, R., Galli, L., Antonuzzo, A., Bursi, S., Roncella, C., Fontanini, G., Sensi, E., and Falcone, A. (2017). Metastatic BRAF K601E-mutated melanoma reaches complete response to MEK inhibitor trametinib administered for over 36 months. Exp. Hematol. Oncol. *6*, 6. https://doi.org/10.1186/s40164-017-0067-4.

50. Dai, X., Zhang, X., Yin, Q., Hu, J., Guo, J., Gao, Y., Snell, A.H., Inuzuka, H., Wan, L., and Wei, W. (2022). Acetylation-dependent regulation of BRAF oncogenic function. Cell Rep. *38*, 110250. https://doi.org/10.1016/j.celrep.2021.110250.

51. Park, E., Rawson, S., Li, K., Kim, B.W., Ficarro, S.B., Pino, G.G.-D., Sharif, H., Marto, J.A., Jeon, H., and Eck, M.J. (2019). Architecture of autoinhibited and active BRAF–MEK1–14-3-3 complexes. Nature *575*, 545–550. https://doi.org/10.1038/s41586-019-1660-y.

52. Gobbi, G., Donati, B., Do Valle, I.F., Reggiani, F., Torricelli, F., Remondini, D., Castellani, G., Ambrosetti, D.C., Ciarrocchi, A., and Sancisi, V. (2019). The Hippo pathway modulates resistance to BET proteins inhibitors in lung cancer cells. Oncogene *38*, 6801–6817. https://doi.org/10.1038/s41388-019-0924-1.

53. Yu, F.X., Zhao, B., and Guan, K.L. (2015). Hippo pathway in organ size control, tissue homeostasis, and cancer. Cell *163*, 811–828. https://doi.org/10.1016/j.cell.2015.10.044.

54. Yu, F.X., Meng, Z., Plouffe, S.W., and Guan, K.L. (2015). Hippo pathway regulation of gastrointestinal tissues. Annu. Rev. Physiol. *77*, 201–227. https://doi.org/10.1146/annurev-physiol-021014-071733.

55. Plouffe, S.W., Meng, Z., Lin, K.C., Lin, B., Hong, A.W., Chun, J.V., and Guan, K.L. (2016). Characterization of hippo pathway components by gene inactivation. Mol. Cell *64*, 993–1008. https://doi.org/10.1016/j.molcel.2016.10.034.

56. Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y., et al. (2014). Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nat. Genet. *46*, 583–587. https://doi.org/10.1038/ng.2984.

57. Hua, Z., and Vierstra, R.D. (2011). The cullin-RING ubiquitin-protein ligases. Annu. Rev. Plant Biol. *62*, 299–334. https://doi.org/10.1146/annurev-arplant-042809-112256.

58. Bennett, E.J., Rush, J., Gygi, S.P., and Harper, J.W. (2010). Dynamics of cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics. Cell *143*, 951–965. https://doi.org/10.1016/j.cell.2010.11.017.

59. Pan, Z.Q., Kentsis, A., Dias, D.C., Yamoah, K., and Wu, K. (2004). Nedd8 on cullin: building an expressway to protein destruction. Oncogene *23*, 1985–1997. https://doi.org/10.1038/sj.onc.1207414.

60. Enchev, R.I., Schulman, B.A., and Peter, M. (2015). Protein Neddylation: beyond cullin–RING ligases. Nat. Rev. Mol. Cell Biol. *16*, 30–44. https://doi.org/10.1038/nrm3919.

61. Giannakis, M., Mu, X.J., Shukla, S.A., Qian, Z.R., Cohen, O., Nishihara, R., Bahl, S., Cao, Y., Amin-Mansour, A., Yamauchi, M., et al. (2016). Genomic correlates of immune-cell infiltrates in colorectal carcinoma. Cell Rep. *15*, 857–865. https://doi.org/10.1016/j.celrep.2016.03.075.

62. Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389–403. https://doi.org/10.1016/j.cell.2009.04.042.

63. Wu, J.T., Lin, H.C., Hu, Y.C., and Chien, C.T. (2005). Neddylation and deneddylation regulate Cul1 and Cul3 protein accumulation. Nat. Cell Biol. *7*, 1014–1020. https://doi.org/10.1038/ncb1301.

64. Soucy, T.A., Smith, P.G., Milhollen, M.A., Berger, A.J., Gavin, J.M., Adhikari, S., Brownell, J.E., Burke, K.E., Cardin, D.P., Critchley, S., et al. (2009). An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. Nature *458*, 732–736. https://doi.org/10.1038/nature07884.

65. Angrisani, A., Di Fiore, A., De Smaele, E., and Moretti, M. (2021). The emerging role of the KCTD proteins in cancer. Cell Commun. Signal. *19*, 56. https://doi.org/10.1186/s12964-021-00737-8.

66. Genschik, P., Sumara, I., and Lechner, E. (2013). The emerging family of CULLIN3-RING ubiquitin ligases (CRL3s): cellular functions and disease implications. EMBO J. *32*, 2307–2320. https://doi.org/10.1038/emboj.2013.173.

67. Duan, G., and Walther, D. (2015). The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput. Biol. *11*, e1004049. https://doi.org/10.1371/journal.pcbi.1004049.

68. Vermeulen, K., Van Bockstaele, D.R., and Berneman, Z.N. (2003). The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell Prolif. *36*, 131–149. https://doi.org/10.1046/j.1365-2184.2003.00266.x.

69. Uxa, S., Castillo-Binder, P., Kohler, R., Stangner, K., Müller, G.A., and Engeland, K. (2021). Ki-67 gene expression. Cell Death Differ. *28*, 3357–3370. https://doi.org/10.1038/s41418-021-00823-x.

70. Chen, T., Sun, Y., Ji, P., Kopetz, S., and Zhang, W. (2015). Topoisomerase IIα in chromosome instability and personalized cancer therapy. Oncogene *34*, 4019–4031. https://doi.org/10.1038/onc.2014.332.

71. Neumayer, G., Belzil, C., Gruss, O.J., and Nguyen, M.D. (2014). TPX2: of spindle assembly, DNA damage response, and cancer. Cell. Mol. Life Sci. *71*, 3027–3047. https://doi.org/10.1007/s00018-014-1582-7.

72. Sun, M., Veschi, V., Bagchi, S., Xu, M., Mendoza, A., Liu, Z., and Thiele, C.J. (2019). Targeting the chromosomal passenger complex subunit INCENP induces polyploidization, apoptosis, and senescence in neuroblastoma. Cancer Res. *79*, 4937–4950. https://doi.org/10.1158/0008-5472.CAN-19-0695.

73. Caron, C., Lestrat, C., Marsal, S., Escoffier, E., Curtet, S., Virolle, V., Barbry, P., Debernardi, A., Brambilla, C., Brambilla, E., et al. (2010). Functional characterization of ATAD2 as a new cancer/testis factor and a predictor of poor prognosis in breast and lung cancers. Oncogene *29*, 5171–5181. https://doi.org/10.1038/onc.2010.259.

74. Cao, J., Hou, P., Chen, J., Wang, P., Wang, W., Liu, W., Liu, C., and He, X. (2017). The overexpression and prognostic role of DCAF13 in hepatocellular carcinoma. Tumour Biol. *39*. 1010428317705753.

75. Zhao, L.H., Li, Q., Huang, Z.J., Sun, M.X., Lu, J.J., Zhang, X.H., Li, G., and Wu, F. (2021). Identification of histone methyltransferase NSD2 as an important oncogenic gene in colorectal cancer. Cell Death Dis. *12*, 974. https://doi.org/10.1038/s41419-021-04267-6.

76. Camilo, V., and Henrique, R. (2018). Oncogenic potential of CHAF1A in gastric cancer: A novel link with Helicobacter pylori-driven carcinogenesis? EBiomedicine *38*, 3–4. https://doi.org/10.1016/j.ebiom.2018.11.033.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

77. Volk, A., Liang, K., Suraneni, P., Li, X., Zhao, J., Bulic, M., Marshall, S., Pulakanti, K., Malinge, S., Taub, J., et al. (2018). A CHAF1B-dependent molecular switch in hematopoiesis and leukemia pathogenesis. Cancer Cell *34*, 707–723.e7. https://doi.org/10.1016/j.ccell.2018.10.004.

78. Cuylen, S., Blaukopf, C., Politi, A.Z., Müller-Reichert, T., Neumann, B., Poser, I., Ellenberg, J., Hyman, A.A., and Gerlich, D.W. (2016). Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. Nature *535*, 308–312. https://doi.org/10.1038/nature18610.

79. Ruchaud, S., Carmena, M., and Earnshaw, W.C. (2007). Chromosomal passengers: conducting cell division. Nat. Rev. Mol. Cell Biol. *8*, 798–812. https://doi.org/10.1038/nrm2257.

80. Revenko, A.S., Kalashnikova, E.V., Gemo, A.T., Zou, J.X., and Chen, H.W. (2010). Chromatin loading of E2F-MLL complex by cancer-associated coregulator ANCCA via reading a specific histone mark. Mol. Cell. Biol. *30*, 5260–5272. https://doi.org/10.1128/MCB.00484-10.

81. Tanaka, H., Igata, T., Etoh, K., Koga, T., Takebayashi, S.I., and Nakao, M. (2020). The NSD2/WHSC1/MMSET methyltransferase prevents cellular senescence-associated epigenomic remodeling. Aging Cell *19*, e13173. https://doi.org/10.1111/acel.13173.

82. Ye, X., Franco, A.A., Santos, H., Nelson, D.M., Kaufman, P.D., and Adams, P.D. (2003). Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. Mol. Cell *11*, 341–351. https://doi.org/10.1016/S1097-2765(03)00037-6.

83. Kufer, T.A., Silljé, H.H.W., Körner, R., Gruss, O.J., Meraldi, P., and Nigg, E.A. (2002). Human TPX2 is required for targeting aurora-A kinase to the spindle. J. Cell Biol. *158*, 617–623. https://doi.org/10.1083/jcb.200204155.

84. Xu, Z., Ogawa, H., Vagnarelli, P., Bergmann, J.H., Hudson, D.F., Ruchaud, S., Fukagawa, T., Earnshaw, W.C., and Samejima, K. (2009). INCENP–aurora B interactions modulate kinase activity and chromosome passenger complex localization. J. Cell Biol. *187*, 637–653. https://doi.org/10.1083/jcb.200906053.

85. Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., Li, B., and Liu, X.S. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. *48*, W509–W514. https://doi.org/10.1093/nar/gkaa407.

86. Nagy, Á., Munkácsy, G., and Győrffy, B. (2021). Pancancer survival analysis of cancer hallmark genes. Sci. Rep. *11*, 6047. https://doi.org/10.1038/s41598-021-84787-5.

87. Wang, D., Qian, X., Du, Y.-C.N., Sanchez-Solana, B., Chen, K., Kanigicherla, M., Jenkins, L.M., Luo, J., Eng, S., Park, B., et al. (2023). cProSite: A web based interactive platform for online proteomics, phosphoproteomics, and genomics data analysis. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.10.543932.

88. Hümmer, S., and Mayer, T.U. (2009). Cdk1 negatively regulates midzone localization of the mitotic kinesin Mklp2 and the chromosomal passenger complex. Curr. Biol. *19*, 607–612. https://doi.org/10.1016/j.cub.2009.02.046.

89. Giubettini, M., Asteriti, I.A., Scrofani, J., De Luca, M., Lindon, C., Lavia, P., and Guarguaglini, G. (2011). Control of aurora-A stability through interaction with TPX2. J. Cell Sci. *124*, 113–122. https://doi.org/10.1242/jcs.075457.

90. Lindon, C., Grant, R., and Min, M. (2015). Ubiquitin-mediated degradation of Aurora kinases. Front. Oncol. *5*, 307. https://doi.org/10.3389/fonc.2015.00307.

91. Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip. Rev. Syst. Biol. Med. *4*, 565–583. https://doi.org/10.1002/wsbm.1185.

92. Olsen, J.V., and Mann, M. (2013). Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol. Cell. Proteomics *12*, 3444–3452. https://doi.org/10.1074/mcp.O113.034181.

93. Drainas, A.P., Lambuta, R.A., Ivanova, I., Serçin, Ö., Sarropoulos, I., Smith, M.L., Efthymiopoulos, T., Raeder, B., Stütz, A.M., Waszak, S.M., et al. (2020). Genome-wide screens implicate loss of cullin Ring ligase 3 in persistent proliferation and genome instability in TP53-deficient cells. Cell Rep. *31*, 107465. https://doi.org/10.1016/j.celrep.2020.03.029.

94. Chen, Z., Ioris, R.M., Richardson, S., Van Ess, A.N., Vendrell, I., Kessler, B.M., Buffa, F.M., Busino, L., Clifford, S.C., Bullock, A.N., et al. (2022). Disease-associated KBTBD4 mutations in medulloblastoma elicit neomorphic ubiquitylation activity to promote CoREST degradation. Cell Death Differ. *29*, 1955–1969. https://doi.org/10.1038/s41418-022-00983-4.

95. De Vita, E.D. (2021). 10 years into the resurgence of covalent drugs. Future Med. Chem. *13*, 193–210. https://doi.org/10.4155/fmc-2020-0236.

96. Abbasov, M.E., Kavanagh, M.E., Ichu, T.A., Lazear, M.R., Tao, Y., Crowley, V.M., am Ende, C.W., Hacker, S.M., Ho, J., Dix, M.M., et al. (2021). A proteome-wide atlas of lysine-reactive chemistry. Nat. Chem. *13*, 1081–1092. https://doi.org/10.1038/s41557-021-00765-4.

97. Qu, L., Yi, Z., Zhu, S., Wang, C., Cao, Z., Zhou, Z., Yuan, P., Yu, Y., Tian, F., Liu, Z., et al. (2019). Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. Nat. Biotechnol. *37*, 1059–1069. https://doi.org/10.1038/s41587-019-0178-z.

98. Yi, Z., Qu, L., Tang, H., Liu, Z., Liu, Y., Tian, F., Wang, C., Zhang, X., Feng, Z., Yu, Y., et al. (2022). Engineered circular ADAR-recruiting RNAs increase the efficiency and fidelity of RNA editing in vitro and in vivo. Nat. Biotechnol. *40*, 946–955. https://doi.org/10.1038/s41587-021-01180-3.

99. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259–1262. https://doi.org/10.1126/science.aas9129.

100. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science *368*, 290–296. https://doi.org/10.1126/science.aba8853.

101. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157. https://doi.org/10.1038/s41586-019-1711-4.

102. Anzalone, A.V., Gao, X.D., Podracky, C.J., Nelson, A.T., Koblan, L.W., Raguram, A., Levy, J.M., Mercer, J.A.M., and Liu, D.R. (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat. Biotechnol. *40*, 731–740. https://doi.org/10.1038/s41587-021-01133-w.

103. Sánchez-Rivera, F.J., Diaz, B.J., Kastenhuber, E.R., Schmidt, H., Katti, A., Kennedy, M., Tem, V., Ho, Y.J., Leibold, J., Paffenholz, S.V., et al. (2022). Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. Nat. Biotechnol. *40*, 862–873. https://doi.org/10.1038/s41587-021-01172-3.

104. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. *14*, 1188–1190. https://doi.org/10.1101/gr.849004.

105. Scott, S.J., Suvarna, K.S., and D'Avino, P.P. (2020). Synchronization of human retinal pigment epithelial-1 cells in mitosis. J. Cell Sci. *133*, jcs247940. https://doi.org/10.1242/jcs.247940.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-FLAG	Sigma-Aldrich	Cat# F7425; RRID:AB_439687
Mouse monoclonal anti-FLAG	Sigma-Aldrich	Cat# F1804; RRID:AB_262044
Rabbit polyclonal anti-HA	Sigma-Aldrich	Cat# H6908; RRID:AB_260070 Cat# SAB4300603; RRID:AB_10620829
Mouse monoclonal anti-β-Tubulin	CWBIO	Cat# CW0098; RRID:AB_2814800
Rabbit polyclonal anti-GAPDH	EASYBIO	Cat# BE3407; RRID: AB_3075319
Rabbit monoclonal anti-RHOA	Cell Signaling	Cat# 2117; RRID:AB_10693922
Rabbit monoclonal anti-CUL3	Cell Signaling	Cat# 10450; RRID:AB_2943632
Rabbit monoclonal anti-INCENP	ORIGENE	Cat# TA890129; RRID: AB_3075321
Rabbit monoclonal anti-CCNB1	Cell Signaling	Cat# 12231; RRID:AB_2783553
Rabbit monoclonal anti-AURKA	Cell Signaling	Cat# 14475; RRID:AB_2665504
B/AIM1 anti-AURKB	Cell Signaling	Cat# 3094; RRID:AB_10695307
Rabbit polyclonal anti-KCTD10	Proteintech	Cat# 27279-1-AP; RRID:AB_2880827
Rabbit monoclonal anti-TPX2	Cell Signaling	Cat# 12833; RRID:AB_2798039
Rabbit monoclonal anti-TOP2A	Cell Signaling	Cat# 12286; RRID:AB_2797871
Rabbit polyclonal anti-CRISPR-Cas9	Abcam	Cat# ab204448; RRID:AB_2893352
Goat anti-rabbit IgG-HRP secondary antibody	Jackson Immunoresearch	Cat# 111-035-003; RRID:AB_2313567
Goat anti-mouse IgG-HRP secondary antibody	Jackson Immunoresearch	Cat# 115-035-003; RRID:AB_10015289
Bacterial and virus strains		
E. coli HST08 premium electro-cells	TaKaRa	Cat# 9028
*Trans*1-T1 phage resistant chemically competent cell	TransGen Biotech	Cat# CD501
Chemicals, peptides, and recombinant proteins		
Gibson Assembly® Master Mix	New England Biolabs	Cat# E2611
T4 DNA Ligase	New England Biolabs	Cat# M0202
X-tremeGENE HP DNA transfection reagent	Roche	Cat# 06366236001
Puromycin	Solaribio	Cat# P8230
Passive lysis buffer	Promega	Cat# E1941
Phosphatase inhibitor cocktail	Thermo Fisher Scientific	Cat# 78441
Crystal Violet Stain solution, 0.1%	Solarbio	Cat# G1063
CellTiter-Glo® 2.0	Promega	Cat# G9242
Cycloheximide	Sigma-Aldrich	Cat# 5087390001
cOmplete™ protease inhibitor cocktail	Roche	Cat# 4693116001
ANTI-FLAG® M2 affinity gel	Sigma-Aldrich	Cat# A2220
Reducing SDS loading buffer	CWBIO	Cat# CW0027
Non-reducing SDS loading buffer	CWBIO	Cat# CW0028
MG132	Sigma-Aldrich	Cat# C2211
Paraformaldehyde	Solarbio	Cat# P1110
FxCycle Violet	Invitrogen	Cat# F10347
Palbociclib	Selleck	Cat# S4482
Nocodazole	Sigma-Aldrich	Cat# M1404
Critical Commercial Assays		
DNeasy Blood and Tissue Kit	Qiagen	Cat# 69506
KAPA HiFi HotStart ReadyMix PCR Kit	Roche	Cat# KK2602

(*Continued on next page*)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DNA Clean & Concentrator-5	Zymo Research Corporation	Cat# D4013
NEBNext Ultra II Q5 Master Mix	New England Biolabs	Cat# M0544
Clarity™ Western ECL Substrate Kit	Bio-Rad	Cat# 1705060
Pierce™ BCA Protein Assay Kit	Thermo Fisher Scientific	Cat# 23225
TGX™ FastCast™ Acrylamide Starter Kit	Bio-Rad	Cat# 1610172
Trans-Blot Turbo RTA Midi 0.2 μm PVDF Transfer Kit	Bio-Rad	Cat# 1704273
Deposited data		
NGS data	This study	NCBI BioProject accession: No. PRJNA811956
Original western blots	This study	Mendeley: https://doi.org/10.17632/3nfzxn4svd.1
Original code	This study	Zenodo: https://doi.org/10.5281/zenodo.8320606 and https://doi.org/10.5281/zenodo.8320602
Experimental models: Cell lines		
HEK293T	C. Zhang's laboratory, Peking University	N/A
hTERT-RPE1	Y. Sun's laboratory, Peking University	N/A
hTERT-RPE1-ABEmax-SC	This study	N/A
hTERT-RPE1-CUL3-K638E-SC	This study	N/A
hTERT-RPE1-KCTD10-K171E-SC	This study	N/A
hTERT-RPE1-TNFAIP1-K168E-SC	This study	N/A
hTERT-RPE1-RHOA-OE	This study	N/A
hTERT-RPE1-RHOA-K135E-OE	This study	N/A
hTERT-RPE1-RHOA-K135G-OE	This study	N/A
hTERT-RPE1-CUL3-OE	This study	N/A
hTERT-RPE1-CUL3-K638E-OE	This study	N/A
hTERT-RPE1-TPX2-OE	This study	N/A
hTERT-RPE1-INCENP-OE	This study	N/A
hTERT-RPE1-AURKA-OE	This study	N/A
hTERT-RPE1-AURKB-OE	This study	N/A
Oligonucleotides		
Oligonucleotides are listed in Table S1	Synbio Technologies, Ruibiotech and TsingkeBiotechnology	Table S1
Recombinant DNA		
pCMV-ABEmax-P2A-GFP	Addgene	Cat# 112101
pLenti-ABEmax-EGFP	This study	N/A
pCG-2.0-CMV-mCherry	This study	N/A
pLenti-cDNA-Flag-mCherry	This study	N/A
pLenti-cDNA-HA-EGFP	This study	N/A
pRK5-HA-Ubiquitin-WT	Addgene	Cat# 17608
pRK5-HA-Ubiquitin-K63	Addgene	Cat# 17606
pRK5-HA-Ubiquitin-K48	Addgene	Cat# 17605
pVSV-G	Addgene	N/A
pR8.74	Addgene	N/A
pCG-2.0-SV40-Puro-iBAR-1	This study	N/A
pCG-2.0-SV40-Puro-iBAR-2	This study	N/A
pCG-2.0-SV40-Puro-iBAR-3	This study	N/A

(*Continued on next page*)

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
ImageJ	National Institutes of Health	https://imagej.nih.gov/ij/
FlowJo	BD Biosciences	https://www.bdbiosciences.com/en-us/products/software/flowjo-v10-software
STRING	Global Biodata Coalition and ELIXIR	https://string-db.org/
Cytoscape	Cytoscape	https://cytoscape.org/
Comparative Proteome Analysis Software Suite, CompPASS	Sowa et al.[62]	http://bioplex.hms.harvard.edu/comppass/
Prism 9	Graphpad	https://www.graphpad.com/features
R (v3.4.1, v3.5.0)	R	https://www.r-project.org
Python (v2.7, v3.8)	Python	https://www.python.org
Chimera	Resource for Biocomputing, Visualization, and Informatics	https://www.cgl.ucsf.edu/chimera/
Adobe Illustrator	Adobe	https://www.adobe.com/products/illustrator.html
Adobe Photoshop	Adobe	https://www.adobe.com/products/photoshop.html
Image Lab	Bio-Rad	https://www.bio-rad.com/zh-cn/product/image-lab-software?ID=KRE6P5E8Z
Alphafold2	DeepMind and EMBL-EBI	https://alphafold.ebi.ac.uk/
Timer2.0	Li et al.[85]	http://timer.cistrome.org
cProSite	Wang et al.[87]	https://cprosite.ccr.cancer.gov

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Wensheng Wei (wswei@pku.edu.cn).

Materials availability

All unique materials generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- Western blots have been deposited at Mendeley: https://doi.org/10.17632/3nfzxn4svd.1 and is publicly available as of the date of publication. DOI is listed in the key resources table. NGS data are available in the NCBI BioProject accession: No. PRJNA811956. Data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at Zenodo: https://doi.org/10.5281/zenodo.8320606 and https://doi.org/10.5281/zenodo.8320602, and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and culture

The HEK293T cell line was obtained from C. Zhang's laboratory (Peking University). HEK293T cells were cultured in high-glucose DMEM (Gibco, Thermo Fisher Scientific), supplemented with 2 mM L-glutamine, 10% (vol/vol) fetal bovine growth serum (Biological Industries), and 1% (vol/vol) penicillin-streptomycin (Thermo Fisher Scientific). The hTERT-RPE1 cell line was obtained from Y. Sun's laboratory (Peking University). HTERT-RPE1 cells were cultured in high-glucose DMEM/F-12 (Gibco, Thermo Fisher Scientific), supplemented with 2 mM L-glutamine, 10% (vol/vol) fetal bovine serum (Biological Industries or CellMax), and 1% (vol/vol) penicillin-streptomycin (Thermo Fisher Scientific). All cells were cultured in a humidified atmosphere at 37°C with 5% CO_2, and were confirmed to be mycoplasma-negative. To inhibit protein neddylation, cells were treated with 1 μM MLN4924 for 20 h.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

METHOD DETAILS

Plasmid construction

The pCMV-ABEmax-P2A-GFP was obtained from Addgene. The coding sequence of ABEmax was cloned into the pLenti-P2A-EGFP vector through restriction enzyme double digestion (New England Biolabs) and T4 DNA ligase ligation. The pLenti-ABEmax-EGFP plasmid was used for packing the lentivirus expressing ABEmax. SgRNA plasmids used for individual validation were generated by cloning the spacer sequences into the pCG-2.0-CMV-mCherry vector through Golden-Gate assembly. Protein-coding sequences of RBM38, RHOA, CUL3, KCTD10, TNFAIP1, TPX2, INCENP, AURKA, and AURKB were cloned into pLenti-cDNA-Flag-mCherry or pLenti-cDNA-HA-EGFP through Gibson assembly, with mutations introduced through PCR.

SgRNA library design

The human reference sequences of proteins and genome were obtained from Uniprot (UP000005640, release 2018_07) and Illumina iGenome (hg38 assembly), respectively. The sgRNA library was designed as follows:

1. All positions of lysine residues in protein sequences from the Uniprot reference Fasta file were captured using a Python script.
2. According to the positions of lysine residues in protein sequences, all positions of lysine codons were captured from the reference genome. Regarding codons separated by introns, lysine codons were classified as three types: unspliced, "AA|R" type of spliced, and "A|AR" type of spliced.
3. Searching for PAM sequences:
   a) For unspliced and "AA|R" type of spliced lysine codons, "NGG" PAM sequences were scanned at a distance of 12-nt to 18-nt downstream from the first adenosine.
   b) For "A|AR" type of spliced lysine codons, "NGG" PAM sequences were scanned at a distance of 12-nt to 17-nt downstream from the first adenosine, as well as a distance of 12-nt to 17-nt downstream from the second adenosine.
4. For lysine codons with available PAM sequences, the positions of each sgRNA were defined as the distance from the first adenosine in lysine codons to the first nucleoside of "NGG" PAMs. Exceptions were made for targets located in the second adenosine of "A|AR" type, where the sgRNA positions were defined as the distance from the second adenosine in these lysine codons to the first nucleoside of "NGG" PAMs.
5. Extracting sgRNA sequences:
   a) Regarding unspliced and "AA|R" type of spliced lysine codons, if the first adenosine was at position 19, a 21-nt sequence upstream of the PAM was captured as the sgRNA sequence; if the first adenosine was at position 13, 14, or 18, a 20-nt sequence upstream of the PAM was captured; if the first adenosine was at position 15, 16, or 17, a 19-nt sequence upstream of the PAM was captured.
   b) Regarding "A|AR" type of spliced lysine codons, if the targeted adenosine was at position 18, a 21-nt sequence upstream of PAM was captured as the sgRNA sequence; if the targeted adenosine was at position 13 or 17, a 20-nt sequence upstream of PAM was captured; if the targeted adenosine was at the position 14, 15, or 16, a 19-nt sequence upstream of PAM was captured.
6. To analyze the on-target and 1-bp mismatch off-target of all captured sgRNAs, the Bowtie software with the parameter "bowtie -a -v 1" was used to align the spacer sequences along with "NGG" sequences to the reference genome.

Library production

The oligonucleotide pool was synthesized by Synbio Technologies. SgRNA sequences were PCR amplified with a primer pair targeting the flanking sequences of the oligos (Table S1). Then, sgRNA sequences were cloned into three types of pLenti-sgRNA^{iBAR} vectors through Golden-Gate assembly. Three verified iBARs were used in this library: CTCGCT, GATGGT, and GCACTG (5'–3'). The Golden-Gate products were purified and then electroporated into competent cells to produce the plasmid library. Next, the plasmid library was co-transfected into HEK293T cells with lentiviral packaging plasmids pVSV-G and pR8.74 (Addgene) using the X-tremeGENE HP DNA transfection reagent to produce lentiviral library. The lentiviral library was harvested 72 h post-transfection. For the titration of the lentivirus library, RPE1 cells were seeded onto 6-well plates and infected with lentiviruses at volumes of 0, 2, 4, 8, 16, and 32 μL, respectively. Seventy-two hours post-infection, these RPE1 cells were replated and cultured with or without 15 μg/mL puromycin for an additional 48 h. Viable cells in each group were counted, and the virus titer was determined based on the cell viability ratios.

Functional lysine screening in RPE1 cells

ABEmax-expressing RPE1 cells were infected with the lentiviral library at a high MOI (≈ 3). Seventy-two hours post-infection, RPE1 cells were treated with 15 μg/mL puromycin for 2 d. Then, a subset of viable cells was collected as the reference group, and this time point was designated as Day 0. The remaining library cells were cultured and maintained at a 1,000-fold coverage for fitness screening. Library cells were passaged every 3 d and harvested on Day 24 as the experimental group. Genomic DNA was extracted from the cell pellets of both the reference and experimental groups using the DNeasy Blood and Tissue Kit. The sgRNA sequences within the cellular genome were amplified by PCR using the KAPA HiFi HotStart ReadyMix PCR Kit with five pairs of primers

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

(Table S1). The PCR condition was as follows: 30 s at 95°C for DNA melting; 10 s at 95°C, 30 s at 60°C, and 15 s at 72°C for 26 cycles; 15 s at 72°C for an extension. The PCR products were purified by DNA Clean & Concentrator-5, followed by NGS analysis.

Cell proliferation assay

ABEmax-expressing RPE1 cells were infected with lentiviruses carrying individual sgRNAs (Table S1) at an infection efficiency of 40–60%. Sg*AAVS1* served as negative controls. The lentiviral plasmids expressing sgRNA contained an mCherry marker gene. The percentage of mCherry-positive cells was assessed through flow cytometry (LSRFortessa, Becton Dickinson Inc.). The flow cytometry analysis started on the fifth or sixth day post-infection, designated as Day 0, serving as the baseline for data normalization. Then the percentage of mCherry-positive cells was analyzed every 3 d, extending to Day 15 or Day 18.

Editing efficiency detection by NGS analysis

Genomic regions flanking the targeted lysine codons were PCR amplified with NEBNext Ultra II Q5 Master Mix and sequenced (Table S1). The FASTQ files were analyzed using an in-house script. The NGS reads were mapped to the human genome, and then sequences around the targeted lysine codons were extracted and translated into protein sequences. The mutation frequency was calculated as the percentage of mutated reads in total matched reads. The seqlogo graphs of the protein sequences were generated by WebLogo 3.[104]

Construction of clonal cells with homozygous lysine mutations

ABEmax-expressing RPE1 cells used in the screen were infected with lentiviruses carrying selected sgRNAs at an infection efficiency of 40%–60%. Cells with moderate mCherry intensity were sorted and isolated into 96-wells, with one cell per well. Homozygous mutants were identified and selected by sequencing of the targeted region. At least three clonal cells were selected for each mutation.

Immunoblotting

Cells were lysed in Passive Lysis Buffer supplemented with protease and phosphatase inhibitor cocktail when approaching 90% confluency. Total protein quantitation was performed using Pierce™ BCA Protein Assay Kit. The lysates were denatured with reducing SDS loading buffer followed by boiling and fractionated in SDS-PAGE (prepared using TGX™ FastCast™ Acrylamide Starter Kit), and then transferred to PVDF membranes. Membranes were blocked with 5% skim milk in PBST for 1 h at room temperature and incubated with primary antibodies (refer to the key resources table) at 4°C overnight. Next, PVDF membranes were washed 5 times with PBST. Subsequently, the PVDF membranes were incubated with secondary antibodies (refer to the key resources table) for 1 h at room temperature, followed by an additional 5 washes. Bands were detected using Clarity™ Western ECL Substrate Kit and Chemidoc™ system.

Clonogenic assay

RPE1 cells (200 cells per well) were seeded onto 6-well plates and cultured for 10–14 d. Then the plates were rinsed by PBS, fixed by methanol, and stained with 0.1% crystal violet. The number of colonies in each well was counted manually. Survival fractions of experimental groups are normalized by colony numbers of controls. Data are presented as the mean ± SD of three independent experiments, and p values are calculated using an unpaired two-tailed Student's t test.

ATP measurement-base cell proliferation assay

RPE1 cells (1000 cells per well) were seeded onto opaque-walled 96-well plates. To measure cell viability, 35 μL of CellTiter-Glo® 2.0 at room temperature was added to each well, followed by shaking for 2 min and incubation for 10 min. The luminescence was then recorded with microplate reader (TECAN Infinite M200).

CHX chase assay

HEK293T cells (4×10^5 cells per well) were seeded onto 12-well plates and transfected with plasmids encoding wild-type or mutant proteins. About 36 h post-transfection, cells were treated with 100 μg/mL of CHX. Then, cell lysates were collected at different time points and subjected to immunoblot analysis. For CUL3, cells were harvested at 0, 9, and 23 h after CHX treatment; for RHOA, cells were harvested at 0, 2, 4, and 8 h after CHX treatment. Quantification was performed using ImageJ software and normalized by β-tubulin level of each sample.

Immunoprecipitation

Cells expressing Flag-tagged cDNAs were lysed in mammalian cell lysis buffer (composed of 50 mM Tris·HCl (pH 7.4), 150 mM NaCl, 0.25% Na-deoxycholate, 1 mM EDTA, 1% Nonidet P-40 and cOmplete™ protease inhibitor cocktail) when approaching 90% confluency. The lysates were incubated with ANTI-FLAG® M2 Affinity Gels at 4°C with rotation overnight. Protein-bound beads were washed with lysis buffer or TSBT for 3 times, and the proteins were eluted with non-reducing SDS loading buffer followed by boiling.

Please cite this article in press as: Bao et al., Unbiased interrogation of functional lysine residues in human proteome, Molecular Cell (2023), https://doi.org/10.1016/j.molcel.2023.10.033

Ubiquitination assay

HEK293T cells (6–8 × 10^5 cells per well) cells were seeded onto 6-well plates and co-transfected with plasmids encoding substrate proteins, enzymatic proteins, together with HA-tagged ubiquitin or ubiquitin mutants. Fourteen hours post-transfection, cells were treated with 10 μM MG132 for 12 h. Subsequently, cells were lysed and subjected to immunoprecipitation and immunoblotting with indicated antibodies.

Label-free quantitative proteomics

Label-free quantitative proteomics of RPE1 clonal cell lines with homozygous KCTD10-K171E or TNFAIP1-K168E mutation, and sg*AAVS1*-expressing cells as negative controls were conducted by PTM Bio. Briefly, cells were lysed in lysis buffer (composed of 8 M urea, 1% protease inhibitor cocktail, and 50 μM PR-619). Proteins were precipitated in 20% vol/vol TCA and digested with trypsin overnight. The samples were then reduced with 5 mM dithiothreitol and alkylated with 11 mM iodoacetamide. The tryptic peptides were separated and analyzed by timsTOF Pro (Bruker Daltonics) mass spectrometry. The resulting MS/MS data were processed using MaxQuant search engine (v.1.6.15.0). Tandem mass spectra were searched against the human SwissProt database (Homo_sapiens_9606_SP_20201214.fasta, 20395 entries) concatenated with reverse decoy database.

AP-MS analysis

RPE1 cells infected with lentiviruses expressing Flag-tagged wild-type or mutant CUL3/KCTD10 were subjected to AP-MS analysis. For AP-MS of KCTD10, cells were treated with MLN4924 to inhibit degradation of substrates. Briefly, samples were prepared by immunoprecipitation with ANTI-FLAG® M2 Affinity Gels, followed by SDS-PAGE separation. The gels were stained with Coomassie blue, and total aggregated proteins of each sample were cut out of the gel and destained with a solution of 100 mM ammonium bicarbonate in 50% acetonitrile. After dithiothreitol reduction and iodoacetamide alkylation, the proteins were digested with porcine trypsin (Sequencing grade modified; Promega, Madison, WI) overnight. The tryptic peptides were separated and analyzed by Thermo Orbitrap Fusion Lumos mass spectrometry. The MS data were alignment with homo sapiens Reviewed Swiss-Port database by Proteome Discoverer 2.2 software. *Z* score and WD score were calculated according to CompPASS algorithm (http://bioplex.hms.harvard.edu/comppass/) based on the PSM number. For CUL3, preys with $Z > 0.70$ and WD > 0.70 were selected as candidates, while for KCTD10, the threshold was set as $Z > 0.75$ and WD > 0.70. Flag-tagged empty vectors served as negative controls. Three replicates were conducted for each bait.

Cell cycle profiling and synchronization

RPE1 cells were harvested and washed with PBS followed by fixation with 4% paraformaldehyde for 30 min at room temperature. After washing with PBS for 2 times, cells were stained with FxCycle Violet, and the cell cycle profiles were analyzed using cell cytometry. For cell cycle synchronization to the M phase, RPE1 cells were treated with 1 μM palbociclib for 18 h. After washing with PBS for 3 times, cells were cultured in a fresh medium for 8 h. Then, cells were treated with 50 ng/mL nocodazole for 12 h to enter into G2/M phase.[105]

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

Quantification of protein abundance in immunoblotting was performed using ImageJ, and normalized by controls as indicated in the corresponding figure legends.

Fluorescence-activated cell sorting data were quantified using FlowJo, and normalized by controls as indicated in the corresponding figure legends.

Protein-protein interaction networks were analyzed by STRING (confidence = 0.7) and created by Cytoscape.

Statistical analysis

Comparisons between experimental sgRNAs and sg*AAVS1* in cell proliferation assay (CPA) were calculated using the unpaired one-tailed Student's t test and adjusted using the Benjamini-Hochberg method. Other statistical analyses are described in the corresponding figure legends. Significance levels are denoted as $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. "NS" indicates non-significant results. Bar plots were generated by GraphPad Prism 9 and R.