



Databases and ontologies

PTMint database of experimentally verified PTM regulation on protein–protein interaction

Xiaokun Hong¹, Ningshan Li², Jiyang Lv¹, Yan Zhang ¹, Jing Li ^{1*}, Jian Zhang ^{3*} and Hai-Feng Chen ^{1*}

¹State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China, ²Department of Bioinformatics and Biostatistics, SJTU-Yale Joint Center for Biostatistics and Data Science, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China and ³Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Department of Pathophysiology, Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai 200025, China

*To whom correspondence should be addressed.

Associate Editor: Peter Robinson

Received on May 25, 2022; revised on December 18, 2022; editorial decision on December 19, 2022; accepted on December 21, 2022

Abstract

Motivation: Post-translational modification (PTM) is an important biochemical process. which includes six most well-studied types: phosphorylation, acetylation, methylation, sumoylation, ubiquitylation and glycosylation. PTM is involved in various cell signaling pathways and biological processes. Abnormal PTM status is closely associated with severe diseases (such as cancer and neurologic diseases) by regulating protein functions, such as protein–protein interactions (PPIs). A set of databases was constructed separately for PTM sites and PPI; however, the resource of regulation for PTM on PPI is still unsolved.

Results: Here, we firstly constructed a public accessible database of PTMint (PTMs that are associated with PPIs) (<https://ptmint.sjtu.edu.cn/>) that contains manually curated complete experimental evidence of the PTM regulation on PPIs in multiple organisms, including *Homo sapiens*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Currently, the first version of PTMint encompassed 2477 non-redundant PTM sites in 1169 proteins affecting 2371 protein–protein pairs involving 357 diseases. Various annotations were systematically integrated, such as protein sequence, structure properties and protein complex analysis. PTMint database can help to insight into disease mechanism, disease diagnosis and drug discovery associated with PTM and PPI.

Availability and implementation: PTMint is freely available at: <https://ptmint.sjtu.edu.cn/>.

Contact: haifengchen@sjtu.edu.cn or jian.zhang@sjtu.edu.cn or jing.li@sjtu.edu.cn

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Post-translational modification (PTM) is a biochemical process that involves the modification of proteins after they have been synthesized. PTM is a common and essential process in cells, and it plays a crucial role in regulating protein function, stability, and localization. PTM is involved in various biological processes, including cell signaling, protein–protein interactions (PPIs), and protein–DNA interactions. PTM is also involved in disease pathogenesis, and abnormal PTM status is closely associated with severe diseases (such as cancer and neurologic diseases) by regulating protein functions, such as PPIs. A set of databases was constructed separately for PTM sites and PPI; however, the resource of regulation for PTM on PPI is still unsolved.

PTM is a biochemical process that involves the modification of proteins after they have been synthesized. PTM is a common and essential process in cells, and it plays a crucial role in regulating protein function, stability, and localization. PTM is involved in various biological processes, including cell signaling, protein–protein interactions (PPIs), and protein–DNA interactions. PTM is also involved in disease pathogenesis, and abnormal PTM status is closely associated with severe diseases (such as cancer and neurologic diseases) by regulating protein functions, such as PPIs. A set of databases was constructed separately for PTM sites and PPI; however, the resource of regulation for PTM on PPI is still unsolved.

U da aba e ffe e 'PTM/P ce ' ec ec d e PTM e a d/ ce e e .H e e , e a e a d f e e e ced e e a c a d b e , a d a ac e f a f PTM e 3D c e . A e da aba e , P S eP (H bec et al., 2015) de PTM effec PPI ba ed e a e L a a c f a e (Ba d et al., 2009), c e ab c de c fa e e e . T da e, e e a a d e ce a e e e d ed c e PTM f c .PTMf c ed c e PTM effec ba ed ec e a - ed a (Be a et al., 2012). PTM c de ded a d ed c ed f c a a ca be ee PTM ba ed c - e e e (M e et al., 2015). A e d ffe e e d , e Mec eb e e ba ed e face a e - a (Be et al., 2015) a d e e F dX f a e ba ed e ca f ce f ed (Sc et al., 2005), c e e - fac a PTM e ac ce, e .

W a e ab e c de a , e e e ed e PTM da aba e, a c e e e e a e f ed PTM effec PPI, c a PTM e a d e , e ac e , de ec e d , a c a ed d ea e a d c - ca a .M e e , de fac a e e e a f PTM e , ec b ed ee e - e a e de ce e e ce a d c ea a .T da a - ba e be e f f e a ce e e e e a a .

a PTM, PPI a d d ea e e e ce a d c ea ec .

2 Materials and methods

2.1 Data sources

T e f f e PTM da aba e c c a F e 1, c d da ac ec a da a .We def ed e e a e f PTM PPI (Be et al., 2015, 2017; L et al., 2012; L et al., 2021; See et al., 2006; S e a d R ce 2009; Wa et al., 2022): () E a ce: I ce a e aff a d () I b : Dec ea e aff .We e ac ed ef c a PTM e a da c - a ed ea e f U (U P , 2012), PTMD (X et al., 2018), PTMf c (Be a et al., 2012) a d P S eP (H bec et al., 2015) da aba e .We a d aded e e e a e a e P bMed da aba e b ea c ef e - d a d ec b a : *Homo sapiens, Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster, Saccharomyces cerevisiae, Schizosaccharomyces pombe*, e , b d , a ca e, e abe, e ac , e ac , b , d abe, e e , d ca e, e, PTM, P , Ac, Me, S , Ub a d G c .T e , ec ec ed ef e f e ab e ea 3600 a e ca ef ba c e e e e e a e de ce, c c ded e a PTM e a d e , e ac e , de ec e d , a c a ed d ea e a d c - ca a .B ef , e e a ed e e a be ee PTM a d d ea e ba ed ce a a a d ea e d - e e ac e a e .T e , ba ed e de ec e d f PPI , e e a ed e e e ac affec ed b PTM ef e .B e a , e e ab ed e e a a PTM, - e e ac a d d ea e .

2.2 Protein sequence analysis

We ea c ed e U da aba e (U P , 2012) ba a ca - ca e e e ce .A d e e e ce d (ea a d d ea f e e d e a d e PTM e) a a e ac ed .T e d de e c e e e ca c a ed b IUP ed 2A (Me a et al., 2018). P e e e ce e e a a ed Pfa da aba e ba f c a d a f a (M et al., 2021).

2.3 Protein structure

F d da e , e d aded e f -e e c e b f A aF d P e S c e Da aba e (A aF d DB) (Va ad et al., 2022). If e e e a - e a 2700 a ac d , e ed A aF d (e 2.1.2) (J e et al., 2021) ed c ed a c e f -e e , e ec e .F e c e e , a e a ed e e e ce e e a ed e PDB da aba e (Be a et al., 2000) ba a a dbaa e - a e f 10⁻⁴. T e PDB e e e e e ec ed acc d ef c e a : () T e PTM e e ed e c e . () T e e a e f ac ed e e ce a a e a e e . () T e a c ed c e e c a e e efe ed .D e e ed c a c e , a a e ca e f e e d c a ef ed b ec a d c - f a e . ZDOCK a fa F e a f -ba ed d c ced ef d e a ea ce f a be b d de e a a a d a a ace be ee - e a de a a e eac e a e e -ba ed c f c (P e ce et al., 2014). HDOCK e b d d c a f e a e de a d f ee d c ba ed ed c a a da e e e de be e e b d e ef a d e e d c (Ya et al., 2020). P MOL ef ec a d c ba ed e a e a e , c a e e e e f ed c de f e a e ae c e (DeLa , 2010). F de , e ed P MOL (e 2.4.1) ba c e c e ba ed PDB e ae c e .F ec a d c , e ed ZDOCK (e 3.0.2) a d HDOCK (e 1.1) b XL-MS da a (c -) e a ec a f b ed ea e a d ed c ed d a d a eac ded b IN c da aba e (Me e et al., 2013) a ed c c a .F e e e d c e e ded b a f ae , e - f ed e c e e e c e e f a e a ef a d c c e c e .

2.4 Interaction analysis

I e ac a e a a d ed - e f ae (C e a d L , 2007; Wa et al., 2014). T e d b c e ac (HP) def ed e e a ce e f de c a f d bc e d e a ec ed 6.5A. T e c a e c a e e ac 11A a a e e / a d - b d f ee e e (Q et al., 2010). T , ed a ce be ee e a ce e f c a e e de e a 11A, c a c de ed a e ec - ac e ac (ELE). A d e b d (HB) ec e e def ed e ed a ce f a ea a e a 3.5A a d e b d a e a e a 120°. We ed e I e face Re d e . P c c e a ed b Ve ee J (:// . /) f c e e face Re d e a a .B ef , P c ec e e e e f e e ac c a a d e c a c a e e d ffe e ce be ee ec e -ba ed acce be face a ea . If e a e ea e a c ff (e defa 1.0 A2), e e de a ed a e face a d e .T e a e ce a a d ed f PTM e abe e face a PTM e .

2.5 Secondary structure analysis

T ba e e f e ec da c e f c e c - e a d PTM e , e ec da c e c e a c a c a ed b Dc a f P e Sec da S c e a (Kab c a d Sa de , 1983) acc d e e d e - ec f c HB e

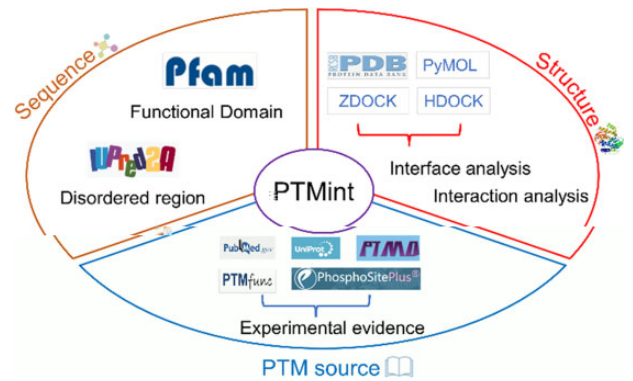


Fig. 1. T e e a de a d c c f PTM da aba e

ec da c e (π -e, 3, (10)-e, α -e, β -b de, β -ee, be da dc). F fca, e a e ca fed f e: () He : π -e, 3, (10)-e ad α -e; () S ee: β -b de ad β -ee; () T : ; () L : be da dc.

2.6 Score calculation

A e 20 ba ca acd e, ea acd ae f e PTM- dfed, c a Se e (S), T e e (T), T e (Y), L e (K) ad A e (R). Se e ad T e e ca be dfed b P ad G c. A d L e ca be dfed b - e dfca, c a Ac, Me, S ad Ub. T a e e - a e a e f PTM PPI e *in vivo*, e d ced e a ce c e f PTM e. T a ce c e a e acc f e be f PTM e ad e ac e a dc a e d E a (1).

$$Sc e = 2 \sum (PTM N). \tag{1}$$

I c N e be f e c ec fc PTM e e a e.

A d e Sc e ca be a ed E a (2).

$$\frac{N a ed_c e}{N a ed_c e} = 1 - 1/Sc e \in [0.5, 1) \tag{2}$$

T e a ed_c e efec e ea e a ce f ec fc e a PTM e, c ce a e a e be f PTM e e ac e ce a e.

2.7 Database and web interface implementation

T e e e face e e e ed H e Te Ma La e e (HTML), Ja aSc (JS) a d Ca cad S e S ee. A d e eb fa e a ed b B a 4 fa e. F e e, 3D. a e ed a e e 3D c e (Re a d K e, 2015). A d e PPI e a a a ed a d a ed b EC a (L et al., 2018). Be de, a f e ad abe e eb e e e f ed P.

3 Results

3.1 Database and content

T e c e e f PTM c a 2477 -ed da PTM e 1169 e affec 2371 e e a 357 d ea e. U da ba e de e 425 ec d f PTM efec PPI a d 322 f c a PTM e. PTMD da aba e de e 45 ec d f PTM effec PPI a d 34 f c a PTM e. A d P eP de ea c e efe e ce e a e ba ed e. I e e a e (E a ce a d I b), 'E a ce' a ab e, e a PTM e d c ea e PPI (Tab e 1). I e, e a f a PTM e e a e be P (87%) (F . 2A). T e a e a PTM e a e Se e (S), T e e (T) a d T e (Y), c a e a c a e f 87.20% (F . 2B).

A S e e a Tab e S1, fa e e e e a e fed e a PTM e, 15.82% ca e e e e - e e face, e a a e f - e fa ca e

Table 1. Summary of PTM effects on PPIs

PTM e	PTM e	E a ce	I b
P a	S, T, Y	2157	968
Ace a	K	170	86
Me a	K, R	81	19
S a	K	53	26
Ub a	K	60	11
G c a	S, T	5	0

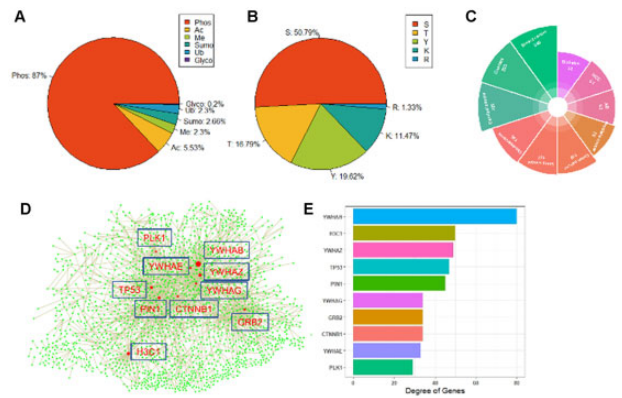


Fig. 2. T e da e PTM da aba e. S a ca a a e e e f PTM e (A) a d PTM e (B). (C) T 10 d ea e affec ed b PTM. (D) PTM-a c - a ed PPI e. T 10 e e e de ee e e ed PPI (D) a d a ed ba (E)

ca a e a e ec a e ac ; 36.61% ca be f d ef c ad a, ca e e a b ca e f PTM da e f c. I e e f ec da c - e, PTM e ed ca e e a e a c - ed e (He, S ee a d T) (Tab e S2). Ba ed e f c e ca c a, e c e f PTM e (f ac : 74.54%) 0.5, c d ca e e dfed b e e f PTM, ca d a e e PPI e c ec ed da a. A d e K10 e f H3C1 e e e c e f 0.97. Ac f K10 ca e a e f ed f e e PPI, c d BAZ1B, BRD7, CHD4, CRH a d TRIM33. I add, e Me f K10 ca e a e 14 d ffe e PPI, c d AGO3, CBX1, CBX3, CDYL, CDYL2, CHAMP1, CHD4, DCAF8, HSFY1, KAT5, MAD2L2, POGZ, UHRF1 a d ZNF470.

T f e de ad e c ca ce c f PTM- dfed e ad e ac e, e e a e e e ed e ca fca acc d eb ca f c (S e e a Tab e S3), a e e ad a - e, e e e e PTM a c a e e eb - ca ce e ad a a a. I e da aba e, e e a a 2960 c e c e c 360 c e (f ac : 12.16%) c e f PDB e e e a c e, 203 c e (f ac : 6.86%) f de (P MOL) a d 2397 (f ac : 80.98%) c e f ec a d c (ZDOCK a d HDOCK). Acc d e d c ed e (XL-MS a d d a d a e ac), eac c e a a ed a c f - de ce a e (H, Med L), 32.26% f c e e 'H' 'Med'.

T e 10 d ea e affec ed b PTM e e c ed a d F e 2C, c a c ded f ee e: () ca ce : b ea ca ce (be : 340), ca ce (be : 323), ce ca ca ce (be : 302), e a c a (be : 136), ca ce (be : 127), c ca ce (be : 119), a e ca ce (be : 76) a d e a ce a ca c a (HCC) (be : 61); () A e e' d ea e (AD) (be : 63); a d () d abe e (be : 52). We a c c ed a PTM-a c a ed PPI e (F . 2D). N de de ee PPI e e e e e c e de. T e e ed a e 10 ee e e de ee ee: YWHAB (de ee: 80), H3C1 (de ee: 50), YWHAZ (de ee: 49), TP53 (de ee: 47), PIN1 (de ee: 45), GRB2 (de ee: 34), YWHAG (de ee: 34), CTNNB1 (de ee: 34), YWHAZ (de ee: 33) a d PLK1 (de ee: 29) (F . 2D a d E), e a e e e e e d ea e - ce be ad e ad a e.

3.2 Web search function

Q c ea c a d ad a ced ea c ee e e ed e e - a e a d 'Sea c' a e, e ec e. O e e a e, e e ca d ec ea c e da aba e b e d (c a Ge e, U, PTM, Effec a d O a) (F . 3B). S e e f - e c d, c a Ge e/U, O a a d PTM e ca

3.4 Web download and help function

After the PTM database is created, the 'Download' and 'Help' functions are available. The 'Download' function allows users to download the database in various formats (e.g., CSV, JSON, XML). The 'Help' function provides detailed information about the database structure and usage.

4 Discussion

The PTM database is a comprehensive resource for researchers studying protein modifications. It provides a central repository for PTM data, including detailed annotations and cross-references to other databases. The database is designed to be user-friendly and accessible, with a variety of search and download options.

PTM databases are essential for understanding the complex signaling pathways and cellular processes regulated by PTMs. The database includes information on the enzymes responsible for PTMs, the substrates, and the specific modifications. This information is crucial for identifying the functional roles of PTMs in various biological contexts. The database also provides a platform for data integration and analysis, allowing researchers to explore the relationships between PTMs and other biological data. The database is continuously updated to reflect the latest research findings in the field of PTMs.

be accessed via the 'Search' page (Fig. 3B). The 'CTDP1' entry is available in the database, providing detailed information on its PTM status, including the enzyme, substrate, and modification type (Fig. 3B). The 'Help' page provides information on the database structure and usage.

The database is a comprehensive resource for researchers studying protein modifications. It provides a central repository for PTM data, including detailed annotations and cross-references to other databases. The database is designed to be user-friendly and accessible, with a variety of search and download options. The database includes information on the enzymes responsible for PTMs, the substrates, and the specific modifications. This information is crucial for identifying the functional roles of PTMs in various biological contexts. The database also provides a platform for data integration and analysis, allowing researchers to explore the relationships between PTMs and other biological data. The database is continuously updated to reflect the latest research findings in the field of PTMs.

3.3 Web browse function

The PTM database is a comprehensive resource for researchers studying protein modifications. It provides a central repository for PTM data, including detailed annotations and cross-references to other databases. The database is designed to be user-friendly and accessible, with a variety of search and download options.

e d e a d e PTM e) c d be a a d c a e e

D e e e d c a c e , e e a f a e a d eb-
e e ffe e b e c a d c , c a ZDOCK (P e c c e
et al., 2014), HDOCK (Y a *et al.*, 2020), C P (K a *et al.*,
2017) a d HADDOCK (a Z d e *et al.*, 2016). I d e
e e e a c c a c f d c e , a a e c a e f e c a
d c a e f e d c b e d e e e a XL-MS da a
(c -) a d e d c e d d a d a e a c .
F e e , e a a a e d e e a c (HB, HP a d ELE)
e e e a c e b e e d e a d e PTM e d e a e.
C a e f e a c (c a HB, HP a d ELE) a d c e
(c a a e , d d e - - d e , e e c a c e a a d d -
a c c e a e) d c e d b PTM c a e a PTM
c d e a e PPI a d d e a e e (D e a a d *et al.*, 2018;
L a d J - B , 2017). T e e f e , d a a b a e d e a
a a b e c a b a f f e e e a , c a e c a
d a c a (MD) a d M a a e d e (MSM).
F e e , d e e e f a e c f c f c e f e d f e a
f e e c e (P a a *et al.*, 2020) d f e d b e
PTM e c d b e a a a b e e e a c f e d . A e a c -
a e e e b a e d b a e d c e c e d d a a (T a b e 1,
S e e a T a b e S1 S3), b b e e a b a e d d e
e e d PTM e a d e - d e d e .
PTM d a a b a e c a b e f e e d e f
a e c . F , e c e e f e d a a b a e c a a
d e a a d PTM e . M e a a d PTM e
b e a d d e d . S e c d , a d d a c a , c a PTM - a e e d
d , PTM e e a a , PTM - a c a e d a a a
a d a c a b e e e PTM a d a b e e a e d .
T d , e e a c e e a e d c e d c e c e
e e - e c e c e a e e e a e d e
PDB d a a b a e . I e f e , e c a a a a d -
d a e e PTM d a a b a e , e e e a PTM e a e
e e d e e a e .
I c c , e d e e e d PTM , a c e e e d a a b a e

- M ,J. et al. (2021) Pfa : e e fa e da aba e 2021. *Nucleic Acids Res.*, **49**, D412 D419.
- Pa a,S. et al. (2020) De e e f a f ce ed f e a f e-c a e a d e - e c e e . *J. Chem. Theory Comput.*, **16**, 2494 2507.
- Pe ce,B.G. et al. (2014) ZDOCK e e : e ac ed c ed c f e - e c e e a d e c e . *Bioinformatics*, **30**, 1771 1773.
- P e ,L. et al. (2009) Ide ca f e a e f e a a c e BAD: e-f ac fBAD e - a ed b a . *J. Biol. Chem.*, **284**, 28004 28020.
- Q ,F. et al. (2010) I d ced c f a a eec f RNA/U1A f d . *RNA*, **16**, 1053 1061.
- Ra ,G. et al. (2006) A e f 14-3-3 - a ed GLUT4 a - ca e ac e RabGAP AS160. *J. Biol. Chem.*, **281**, 29174 29180.
- Re ,N. a d K e ,D. (2015) 3D . : ec a a a WebGL. *Bioinformatics*, **31**, 1322 1324.
- Sc ,J. et al. (2005) T e F dX eb e e : a e f ce ed. *Nucleic Acids Res.*, **33**, W382 W388.
- See ,B.T. et al. (2006) Read e d ca e ac d a . *Nat. Rev. Mol. Cell Biol.*, **7**, 473 483.
- S ,W.X. et al. (2001) S c a a a f a d e e b a f e - a e f *Saccharomyces cerevisiae*. *Biochemistry*, **40**, 10800 10809.
- S ,A. et al. (2010) P e a a e 2A eac a e FOXO3a a d a c e a 14-3-3 a d AKT. *Mol. Biol. Cell.*, **21**, 1140 1152.
- S ,C.Y. et al. (2008) Se e 88 a f e 8- Da d e c a l a ec a c f d e a a a d f c . *J. Biol. Chem.*, **283**, 4004 4013.
- S e ,T.M. a d R ce ,J.C. (2009) Ide ca a d c a ace a f - a a a d ca - ec c b d e b a a a e e ed ca a . *Proc. Natl. Acad. Sci. USA*, **106**, 14808 14813.
- S e ,J.J. et al. (2020) D ec e ce a a a a ec e -ba ed e c . *Curr. Opin. Cell Biol.*, **63**, 20 30.
- Ta a a,H. et al. (2008) Be a 2 e a T 758 ac a a ec a c e a e 14-3-3 a d a b d . *Blood*, **112**, 1853 1862.
- U P ,C. (2012) Re a e e ace a e e a e - ce (U P). *Nucleic Acids Res.*, **40**, D71 D75.
- a Z de ,G.C.P. et al. (2016) T e HADDOCK2.2 eb e e : e-F ed e a e de f b ec a c e e . *J. Mol. Biol.*, **428**, 720 725.
- Va ad ,M. et al. (2022) A a F d e c e da aba e : a e e a d e c a c e a e f e - e e ce ace -acc ac de . *Nucleic Acids Res.*, **50**, D439 D444.
- Wa ,C. et al. (2020) GPS 5.0: a da e e ed c f a e- ec c a e e . *Genomics Proteomics Bioinformatics*, **18**, 72 80.
- Wa ,S. et al. (2022) U c e - a a a d ca -a c a ed e - e e ac . *Curr. Opin. Struct. Biol.*, **74**, 102352.
- Wa ,W. et al. (2014) Ne f ce ed de ca d de ed e .) *Chem. Biol. Drug Des.*, **84**, 253 269.
- W ,S. et al. (2012) Ca e a e l a a e a e a MDMX a ec a e ac a e 53 b d . *Mol. Cell. Biol.*, **32**, 4821 4832.
- X ,H. et al. (2018) PTMD: a da aba e f a d ea e-a ca ed - a - a a d ca . *Genomics Proteomics Bioinformatics*, **16**, 244 251.
- X ,H.D. et al. (2017) PLMD: a da ed da a e ce f e e d - ca . *J. Genet. Genomics*, **44**, 243 250.
- X ,P. et al. (2022) E c e a a a e-2 c ba a a d c c d e d ea eb e de e a d ca e e - a . *Cell Metab.*, **34**, 299 316.e6.
- X e ,Y. et al. (2005) GPS: a c e e e e e f a e ed c . *Nucleic Acids Res.*, **33**, W184 W187.
- Ya ,Y.M. et al. (2020) T e HDCK e e f e a ed e - e d c . *Nat. Protoc.*, **15**, 1829 1852.
- Y ,K. et al. (2019) P : a da aba e f e a d a c a . *Nucleic Acids Res.*, **47**, D451 D458.