

Supplementary Data:

iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations

Jiaqi Zhou^{1,†}, Yang Xu^{1,†}, Shaofeng Lin^{1,†}, Yaping Guo¹, Wankun Deng¹, Ying Zhang¹, Anyuan Guo¹
and Yu Xue^{1,*}

¹Key Laboratory of Molecular Biophysics of Ministry of Education, College of Life Science and Technology and the Collaborative Innovation Center for Biomedical Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Running title: An updated database of UB/UBL regulators

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

*To whom correspondence should be addressed. Tel:+86 27 87793903; Fax:+86 27 87793172;
Email: xueyu@hust.edu.cn

Supplementary Figure S1. The performance evaluation of the HMM-based HMM identifications. The ROC curves were drawn and AUC values calculated for the self-consistency (curves and diamonds in blue) and LOO (curves and triangles in red) validations. Due to the page limitation, the results for 9 families are shown.

Supplementary Figure S2. The heatmap of classifications and numbers of UB/UBL regulators for all 74 families. The evolutionary relations among 148 species are shown at the bottom. The numbers of proteins in the same family can be dramatically different across species. For example, ULD proteins in the RAWUL family were extensively detected in animals and plants but were rarely found in fungi. Additionally, UBD proteins in the UBZ family were mostly identified in animals but not in plants or fungi.

Supplementary Table S1. The statistics of experimentally characterized E1s, E2s, E3s, DUBs, UBDs and ULDs manually collected from the literature. The classification information and corresponding references are also provided.

Supplementary Table S2. The comparison of HMM-based identifications of human UB/UBL regulators in iUUCD 2.0 to InterPro (1) and Pfam (2) databases. *a.* Number, the number of proteins identified in each family. *b.* N/A, not available.

Supplementary Table S3. The detailed classifications and data statistics of E1s, E2s, E3s, DUBs, UBDs and ULDs in iUUCD 2.0.

Supplementary Table S4. For UB/UBL enzymes in 148 eukaryotes, we took the annotations of active sites from the UniProt database (3) if available. It should be noted that the active sites of a large proportion of UB/UBL enzymes were difficult to be exactly pinpointed. For example, the mutagenesis of C449A, H452A, H457S, C461S, C464A, C475G or C478S in human MDM2 (UniProt ID: Q00987) results in the loss of its E3 activity. Also, the UniProt annotations were not integrative for a number of species. *a.* The number of proteins annotated in iUUCD 2.0 for each family. *b.* The number of proteins with UniProt annotations. *c.* The number of proteins annotated

with active sites.

Supplementary Table S5. The 67 additional public databases that cover 11 aspects of the data. We mapped all regulators in 8 model organisms to these databases to provide a multi-layer annotation in iUUCD 2.0. The PTM sites not annotated as 'By similarity', 'Potential' or 'Probable' in UniProt (3) were also integrated.

Supplementary Table S6. The comparison of UUCD 1.0 and iUUCD 2.0.

Supplementary Figure S1. The performance evaluation of the HMM-based HMM identifications. The ROC curves were drawn and AUC values calculated for the self-consistency (curves and diamonds in blue) and LOO (curves and triangles in red) validations. Due to the page limitation, the results for 9 families are shown.



