**Virologica Sinica**

**Supplementary Data**

**Phospho-proteomics identifies a critical role of ATF2 in pseudorabies virus replication**

**Fang-Fang Jianga #, Ren-Qi Wanga #, Chao-Yue Guoa, Ke Zhenga, HaiLong-Liub, Le Sua, Sheng-Song Xieb, Huan-Chun Chena, Zheng-Fei Liua\***

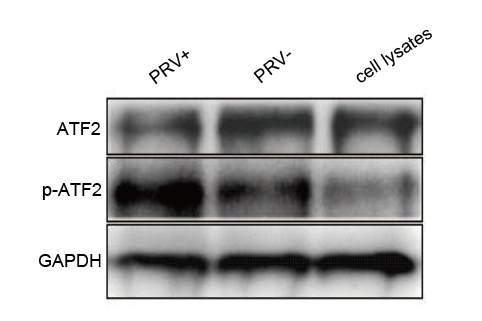
*a State Key Laboratory of Agricultural Microbiology, Hongshan Laboratory and Key laboratory of Preventive Veterinary Medicine in Hubei Province, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China.*

*b Key Lab of Agricultural Animal Genetics, Breeding, and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China.*

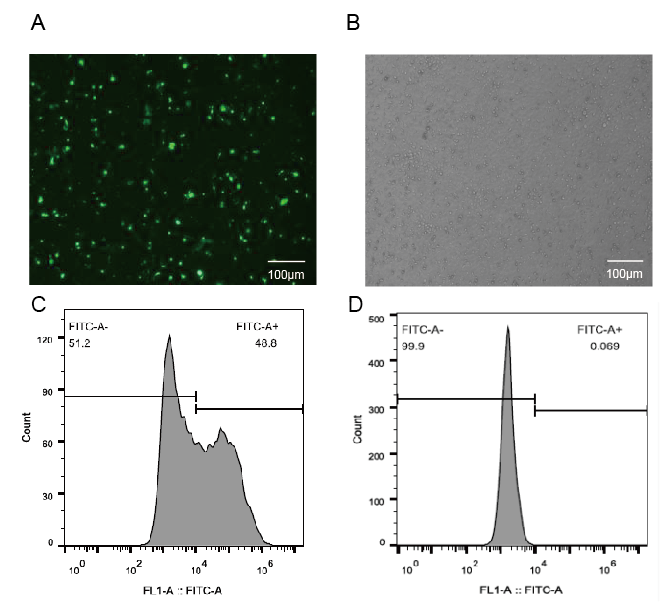
\* Corresponding author.

Email: lzf6789@mail.hzau.edu.cn (Z.F. Liu)

# Fang-Fang Jiang and Ren-Qi Wang contributed equally to this work.



**Supplemental Figure S1.** Confirmation of differentially expressed proteins and phosphoproteins of ATF2 by Western blot. PK-15 cell lysates from total protein was extracted and electrophoresed by SDS-PAGE. ATF2 antibodies were used to detect the expression of ATF2 or phosphorylated ATF2. PRV+: PRV stock (containing PRV and cell lysates). GAPDH serves as the loading control. PRV-: DMEM. Cell lysates: normal PK-15 cell lysates free of PRV.



**Supplemental Figure S2.** Transfection efficiency validation in PK-15 cells. **A** Four microgram of pcDNA3.1-EGFP was transfected into 70%–80% confluent PK-15 cells, and cells were observed under the inverted fluorescent microscope (Nikon ECLIPSE Ti) at 48 hours post transfection (hpt). **B** PK-15 cells from the same field of view in (**A**) were photographed under light microscope. **C** PK-15 cells at 48 hpt were treated with 0.25% trypsin, and the suspended PK-15 cells were detected by flow cytometry (CYTOFLEX). **D** Flow cytometry detection of fluorescence in control PK15 cell. Fluorescence of transfected PK-15 cells was analyzed by CytExport 2.0 CellQuest Pro software. FITC-A+: green fluorescent signal. FITC-A-: no fluorescent signal. Scale bar =100 μm.

**Supplementary Table S1.** siRNA sequences of *NFκB*, *ATF2*, *MAX*, *SOS* and *NC* genes.

|  |  |  |
| --- | --- | --- |
| **Name** | **Sense (5′-3′)** | **Antisense (5′-3′)** |
| siRNA-NFΚB | GGAUCUGCCGAAUCAACAATT | UUGUUGAUUCGGCAGAUCCTT |
| siRNA-ATF2 | GCAGAAGACUUGAGUUCAUTT | AUGAACUCAAGUCUUCUGCTT |
| siRNA-MAX | GCCACAGAGUAUAUCCAGUTT | ACUGGAUAUACUCUGUGGCTT |
| siRNA-SOS | CCAAUUGAUAAGUGGGCAATT | UUGCCCACUUAUCAAUUGGTT |
| siRNA-NC | UUCUCCGAACGUGUCACGACG | ACGUGACACGUUCGGAGAATT |

**Supplementary Table S2.** Primer sequences for RT-qPCR detection (*NFκB, ATF2, MAX*, and *SOS* genes) and plasmid (pEGFP, pATF2, or pmATF2 (T69A, T71A) construction.

|  |  |  |
| --- | --- | --- |
|  | **Name** | **Sequences** |
| RT-qPCR primer | NFΚB | Sense (5′-3′): CAGGAACACGTTCACTGTCACC  Antisense (5′-3′): CATCTCAGTGGTGTTCAGCAGA |
| ATF2 | Sense (5′-3′): GACTGGGAGGAAGGAGCCATAA  Antisense (5′-3′): AGTGCCGAGTAGTCCACATACT |
| MAX | Sense (5′-3′): TCCAGAAGAGCATTCTGCCGCT  Antisense (5′-3′): GCTTTCACAGTTTGCGGGACTCA |
| SOS | Sense (5′-3′): GGTACTAGAAGCACCAGAAGCAG  Antisense (5′-3′): CAGGTACCATGAGACATCCCACA |
| pcDNA3.1-FLAG vector primer | GADPH | Sense (5′-3′): CCATCACTGCCACCCAGAAGACT  Antisense (5′-3′): AGGTCAGATCCACAACCGACACG |
| pcDNA3.1-FLAG-*atf2*(T69A, T71A) point mutations primer | pcDNA3.1-FLAG vector | Sense (5′-3′): CATCTTATCGTCATCGTCTTTGTAATC  Antisense (5′-3′): GCTGATCAGCCTCGACTGTGC |
| ORF primer | pcDNA3.1-FLAG-*atf2*(T69A, T71A) | Sense (5′-3′): GGCTGATCAGGCCCCAGCACCAACACG  Antisense (5′-3′): CGTGTTGGTGCTGGGGCCTGATCAGCC |
|  | ATF2 | Sense (5′-3′): GTAGCTCCGCTTCCACTTCCTGAGGGCTGTGACTG  Antisense (5′-3′): GGAAGCGGAGCTACTAACTTCA |

**Supplementary Table S3.** Quantitative summary of phosphorylated peptides identified by iTRAQ-quantitation.

Column A is listed as Sequence: amino acid Sequence of A peptide; Column B as Protein Group Accessions: Protein login number; Column C as Modifications: modified amino acid, location and modify way; Column D as phosphoRS Site Probabilities: grade potential phosphorylation sites; Column E is listed as phosphoors Binomial Peptide Score. F is listed as IonScore: Mascot score of the peptide; Column G is listed as Charge: the number of charges; H is listed as MH+ [Da] : molecular weight of the peptide; I listed as Δ M (PPM): the error of the measured molecular weight and molecular weight theory; J-L is listed as the ratio of samples to internal parameters.

**Supplementary Table S4.** Significance analysis of phosphopeptides identified in the PRV infection group compared with the blank control group. M is listed as the fold-change, The red mark represents fold change >1.2 (up), and the yellow mark represents fold change <0.83 (down).

**Supplementary Table S5.** All protein informations identified by iTRAQ-quantitation. Column A is listed as protein Accession number; Column B is listed as Description of proteins; Column C is listed as protein Coverage; Column D is listed as the number of Proteins contained in a proteome; Column E as a Unique Peptides: the only number of Peptides; Column F: MW [kDa]: molecular weight of the protein; Column G is listed as calc. pI: isoelectric point of protein.

**Supplementary Table S6. Biological process classification of Go term.**

**Supplementary Table S7. Molecular function classification of Go term.**

**Supplementary Table S8. Cellular components classification of Go term.**

Table S6/S7/S8: Column A represents the hierarchy of Go terms in ontologies. The lower the level number is, the lower the hierarchy is, and the more general the annotation content is.The higher the level number, the more explicit the comment; Column B represents the ID of the Go term; Column C represents the name of the Go term; Column D represents the Go term category; Column E Seqs\_Num represents the number of proteins associated with the Go term; Column F represents the protein ID associated with the Go term.

**Supplementary Table S9. Enriched GO terms.**

Column A represents the ID of the Go term; Column B represents the name of the Go term; Column C represents the Go term category; Column D represents the number of proteins associated with the GO Term in the differential protein sets; Column E represents the number of proteins associated with the GO Term in all qualitative protein sets; Column F represents the number of proteins in the differential protein sets; Column G represents the number of proteins in all qualitative protein sets; Column H represents the proportion of proteins in the differential protein set associated with the GO Term;Column I represents the proportion of proteins associated with the GO Term in all qualitative protein sets; Column J Over.Under means that compared with the proportion of background proteins related to the GO Term, the proportion of target proteins related to the GO Term is either higher (Test\_per > Ref\_per) or lower (Test\_per < Ref\_per), which is usually reserved for over-represented GOterm in the target protein set;Column K represents the ID of the protein associated with the GO Term in the differential protein set; Column L represents the ID of the protein associated with the GO Term in all qualitative protein sets;Column M p-value is the significance index of enrichment analysis,the smaller the p-value is, the more significant the influence of the GO Term is under specific biological treatment, which is calculated by the hypergeometric distribution; Column N FDR (False Discovery Rate) is an error control indicator in multiple hypothesis testing and is the benjamini-hochberg correction for p-value; Column O represents the ratio of the number of proteins associated with a GO Term in the differential protein set to the number of proteins associated with a GO Term in all qualitative protein sets.

**Supplementary Table S10.** **The enriched GO Terms (Top 20)**

Column A represents the ID of the Go term; Column B represents the name of the Go term; Column C represents the Go term category; Column D represents the number of proteins associated with the GO Term in the differential protein sets; Column E represents the number of proteins associated with the GO Term in all qualitative protein sets; Column F represents the number of proteins in the differential protein sets; Column G represents the number of proteins in all qualitative protein sets; Column H represents the proportion of proteins in the differential protein set associated with the GO Term;Column I represents the proportion of proteins associated with the GO Term in all qualitative protein sets; Column J Over.Under means that compared with the proportion of background proteins related to the GO Term, the proportion of target proteins related to the GO Term is either higher (Test\_per > Ref\_per) or lower (Test\_per < Ref\_per), which is usually reserved for over-represented GOterm in the target protein set;Column K represents the ID of the protein associated with the GO Term in the differential protein set; Column L represents the ID of the protein associated with the GO Term in all qualitative protein sets;Column M p-value is the significance index of enrichment analysis,the smaller the p-value is, the more significant the influence of the GO Term is under specific biological treatment, which is calculated by the hypergeometric distribution; Column N FDR (False Discovery Rate) is an error control indicator in multiple hypothesis testing and is the benjamini-hochberg correction for p-value; Column O represents the ratio of the number of proteins associated with a GO Term in the differential protein set to the number of proteins associated with a GO Term in all qualitative protein sets.

**Supplementary Table S11. Query2map**.

Column A represents the target protein ID; Column B represents the protein ID of the target protein in the KEGG database; Column C represents the protein abbreviation corresponding to the target protein in the KEGG database; Column D represents the protein name of the target protein in the KEGG database; Column E represents the ID of the pathway in which the target protein may be involved; Column F indicates the name of the pathway in which the target protein may be involved; Column G URL represents the pathway diagram that can be linked directly to the target protein in the KEGG database, and all target proteins that participate in the pathway are identified in red boxes and fonts.

**Supplementary Table S12. Map2query.**

Column A represents the pathway ids in which the target protein may be involved; Column B indicates the name of the pathway in which the target protein may be involved; Column C represents the ID of the target protein involved in the pathway; Column D represents the number of target proteins involved in the pathway; Column E URL represents the pathway diagram that can be linked directly to the target protein in the KEGG database, and all target proteins that participate in the pathway are identified in red boxes and fonts.

**Supplementary Table S13. Enriched KEGG pathway.**

Column A represents the ID of the annotated KEGG path; Column B represents the name of the annotated KEGG path; Column C represents the number of proteins associated with the pathway in the differential protein set; Column D represents the total number of proteins in the differential protein sets; Column E represents the number of proteins associated with the pathway in all qualitative protein sets; Column F represents the total number of proteins in all qualitative protein collections; Column G represents the proportion of proteins associated with the pathway in the differential protein sets; Column H represents the proportion of proteins associated with this pathway in all qualitative protein sets; Column I represents the protein ids associated with the pathway in the differential protein sets; Column J represents the protein ids associated with this pathway in all qualitative protein sets; Column K represents the ratio of background proteins relative to the KEGG term. The ratio of target proteins related to the KEGG term is either higher (Test\_per > Ref\_per) or lower (Test\_per < Ref\_per), usually reserved for the over-represented KEGG term in the target protein sets; Column L represents the significance index of enrichment analysis, and the smaller the p-value is, the more significant the influence of KEGG term is under specific biological treatment, which is calculated by hypergeometric distribution; Column M FDR (False Discovery Rate) is an error control indicator in multiple hypothesis testing and is the benjamini-hochberg correction for p-value; Columns N represents the ratio of the number of proteins associated with a pathway in the differential protein set to the number of proteins associated with that pathway in all qualitative protein sets.