Identification of SUMO Substrates with Diverse Functions Using the *Xenopus* Egg Extract System

Supplementary materials:

Figure legends:

Figure S1. SUMO orthologs are conserved between *Xenopus laevis* and human.

(A) Sequence comparison of *Xenopus laevis* SUMO3 (NP_001079759) and human SUMO3 (NP_008867). Completely conserved residues are shaded green, identical residues are shaded yellow, similar residues are shaded cyan, and different residues are shaded white.

(B) Sequence comparison of human SUMO1 (NP_003343), SUMO2 (NP_008868) and SUMO3 (NP_008867). Completely conserved residues are shaded green, identical residues are shaded yellow, similar residues are shaded cyan, and different residues are shaded white.

(C) Sequence comparison of *Xenopus laevis* SUMO1b (NP_001090274), SUMO2b (NP_001085595) and SUMO3 (NP_001079759). Completely conserved residues are shaded green, identical residues are shaded yellow, similar residues are shaded cyan, and different residues are shaded white.

Figure S2. Purification of SUMO mutant proteins.

Bacterially expressed and purified His-GFP-tagged human SUMO1 and SUMO2 mutant proteins (SUMO1-WT, AA, HFV, HFV-AA and SUMO2-WT, AA, QFI, QFI-AA) were analyzed by SDS–PAGE followed by Coomassie blue staining.

Figure S3. Further analysis of the isolated SUMO substrates.

(A) Silver staining of His-tagged SUMO1 and SUMO2 pull-down products from interphase sample.

(B) Silver staining of His-tagged SUMO1 and SUMO2 pull-down products from mitotic phase sample.

(C) Venn diagram comparing the different SUMO candidates from all four pull-down conditions.

(D) Venn diagram comparing the different mitotic phase SUMO candidates in SUMO1 and SUMO2 pull-downs.

(E) Venn diagram comparing the different interphase SUMO candidates in SUMO1 and SUMO2 pull-downs.

(F) Test of the *in vivo* sumoylation system by transfecting FLAG-tagged RanGAP1 mutants with GFP-SUMO1 and/or UBC9 in HEK293T cells followed by IP and immunoblotting to detect their sumoylation. The bands highlighted by one red star indicate RanGAP1 sumoylated by endogenous SUMO. The bands highlighted by two red stars indicate RanGAP1 sumoylated by GFP-SUMO1.

(G) Validation of the sumoylation of CKB, ATXN10, BTF3, HABP4 and BZW1. FLAG-tagged CKB, ATXN10, BTF3, HABP4 or BZW was co-transfected with GFP-SUMO1 and UBC9 in HEK293T cells followed by anti-FLAG immunoprecipitation and immunoblotting. The bands showing sumoylated protein are highlighted by red star. The additional GFP-positive bands

above the highlighted bands in the SUMO1 WT lanes may be due to the multiple sumoylation of candidate proteins.

Figure S4. Analysis of the SUMO substrates by PANTHER and DAVID.

(A) Pie chart showing gene ontology analysis of the SUMO candidate proteins for terms associated with molecular function. The largest categories are binding and catalytic activity.

(B) Pie chart for sub-categories of 'binding' in GO molecular function.

(C) Pie chart for sub-categories of 'catalytic activity' in GO molecular function.

(D) Bar chart showing enrichment p-values of SUMO substrates identified from interphase or mitotic phase for GO Cellular Component terms.

(E) Bar chart showing enrichment p-values of SUMO1 and SUMO2 substrates for GO Cellular Component terms.

Figure S5. Functional protein interaction network analysis based on the STRING database.
The image illustrates a network analysis of sumoylated proteins that was created with STRING. Proteins pulled down by SUMO1, SUMO2 or both were highlighted in red, green or white, respectively. The protein complexes and pathways, highlighted by red circle, are involved in: (1) ribosome assembly and translation; (2) proteasomal degradation; (3) DNA replication and repair; (4) chromatin remodeling and transcriptional regulation; (5) cell cycle regulation; (6) the mitochondrial respiratory chain; (7) metabolic regulation; and (8) the sumoylation pathway.

Table S1. A list of all the sumoylated proteins identified from mitotic phase samples. Three independent pull-down experiments were performed by using egg extracts from three *Xenopus*. Peptides were required to be fully tryptic with at least one peptide per protein identification. The protein false discovery rate was held below one percent and all peptide-spectra matches were required to have less than 10 ppm mass error. Peptide count, spectral count, sequence coverage, NSAF value and emPAI value are reported for each protein identification as well as occurrences in different experiments. Candidate sumoylated proteins were those identified in either the SUMO1 or SUMO2 pull-down but not identified in the control pull-down. After subtracting proteins identified in the control pull-downs, there are 197 non-redundant candidate proteins, of which 72 are SUMO1-specific, 50 are SUMO2-specific, and 57 are shared by both.

Table S2. A list of all the sumoylated proteins identified from interphase samples. Three independent pull-down experiments were performed by using egg extracts from three *Xenopus*. Peptides were required to be fully tryptic with at least one peptide per protein identification. The protein false discovery rate was held below one percent and all peptide-spectra matches were required to have less than 10 ppm mass error. Peptide count, spectral count, sequence coverage, NSAF value and emPAI value are reported for each protein identification as well as occurrences in different experiments. Candidate sumoylated proteins were those identified in either the SUMO1 or SUMO2 pull-down but not identified in the control pull-down. After subtracting proteins identified in the control pull-downs, there are 260 non-redundant candidate proteins, of which 3 are SUMO1-specific, 68 are SUMO2-specific, and 189 are shared by both.

Table S3. A list of all the sumoylated proteins identified using the *Xenopus* system. Among all the 346 proteins identified, 34 (9.8%) proteins have been previously validated for their sumoylation and 102 (29.5%) have been identified as SUMO substrates in other high throughput screenings performed in mammalian cells or budding yeast.

Table S4. Annotation of the sumoylated proteins identified in this study for function related to chromatin association and/or a role in cancer.

Table S5. Analysis of SUMO candidate proteins by gene list functional classification analysis performed using PANTHER. Functional classification of candidate sumoylated proteins is provided for the GO category molecular function as well as the sub-categories binding and catalytic activity. Functional classification of the candidate sumoylated proteins is provided for the GO category biological process as well as the sub-category metabolic process. For each GO category and sub-category, the name of the functional classifications, number of genes, percentage of genes and gene names are provided.

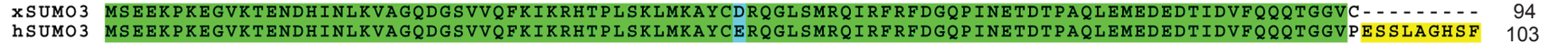
Table S6. Enriched DAVID GO terms for SUMO candidate proteins. Enriched GO terms for the categories biological process, cellular component and molecular function listed for candidate sumoylated proteins identified in mitosis phase, interphase, SUMO1 pull-downs and SUMO2 pull-downs. Enriched terms, count, p-value after Benjamini correction, and gene names are provided. Enriched terms were required to have a p-value < 0.05 after Benjamini correction.

Table S7. Protein interaction pairs among SUMO candidate proteins generated by STRING analysis. Only the highest confidence interactions (confidence score > 0.900) were retained. The number of protein interactions within the STRING analysis is listed for each SUMO candidate protein. The following protein complexes and pathways have been annotated based on protein

interactions: (1) ribosome assembly and translation; (2) proteasomal degradation; (3) DNA replication and repair; (4) chromatin remodeling and transcriptional regulation; (5) cell cycle regulation; (6) the mitochondrial respiratory chain; (7) metabolic regulation; and (8) the sumoylation pathway.

Figure S1

A



B

C

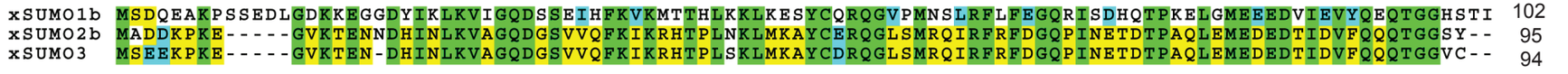


Figure S2

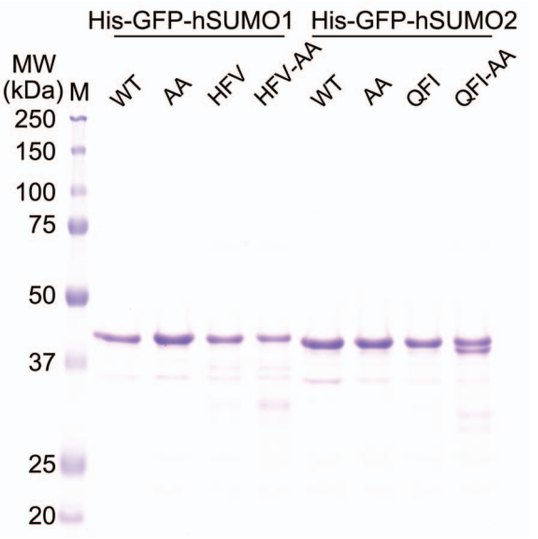

His-GFP-hSUMO1
His-GFP-hSUMO2
MW (kDa)
M
WT
AA
HFV
HFV-AA
WT
AA
QFI
QFI-AA
250
150
100
75
50
37
25
20

Figure S3

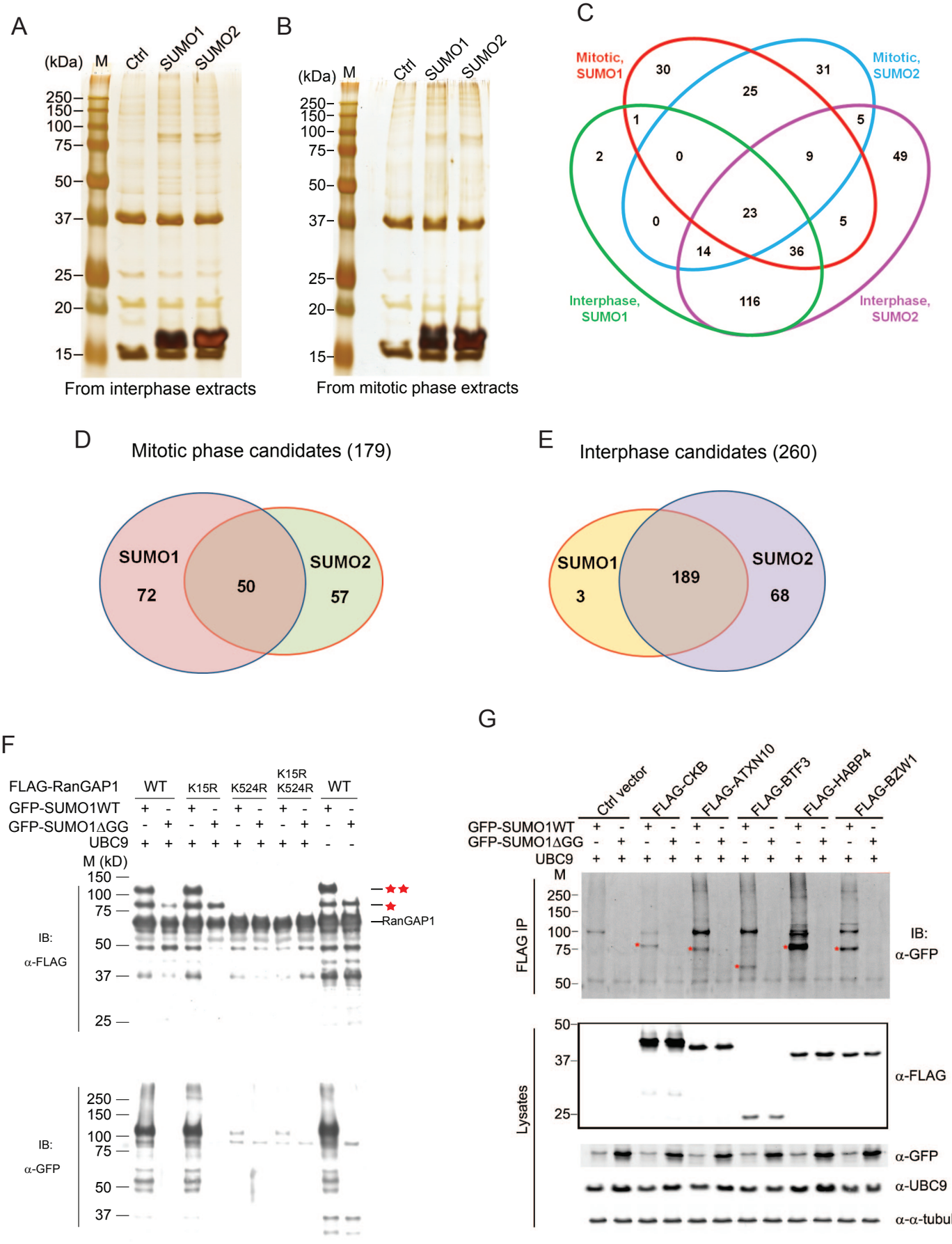
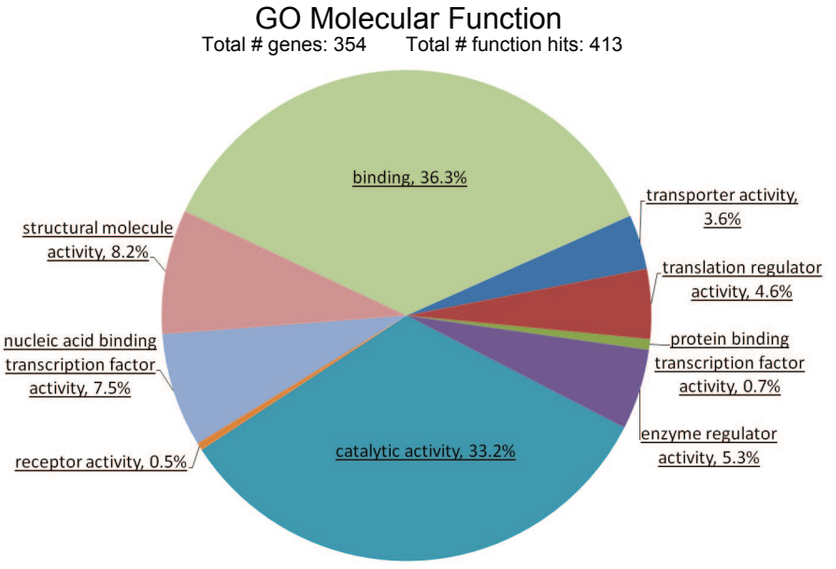
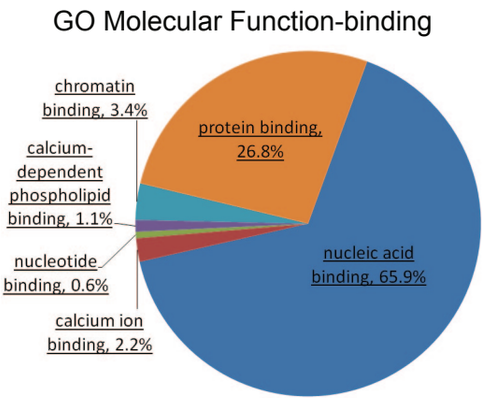

A
(kDa) M Ctrl SUMO1 SUMO2
250–
150–
100–
75–
50–
37–
25–
20–
15–
From interphase extracts
B
(kDa) M Ctrl SUMO1 SUMO2
250–
150–
100–
75–
50–
37–
25–
20–
15–
From mitotic phase extracts
C
Mitotic, SUMO1
Mitotic, SUMO2
Interphase, SUMO1
Interphase, SUMO2
30
31
25
1
5
2
0
9
49
0
23
5
14
36
116
D
Mitotic phase candidates (179)
SUMO1
SUMO2
72
50
57
E
Interphase candidates (260)
SUMO1
SUMO2
3
189
68
F
FLAG-RanGAP1 WT K15R K524R K15R K524R WT
GFP-SUMO1WT + - + - + - + - + -
GFP-SUMO1ΔGG - + - + - + - + - +
UBC9 + + + + + + + + - -
M (kD)
150
100
75
50
37
25
IB: α-FLAG
RanGAP1
250
150
100
75
50
37
IB: α-GFP
G
Ctrl vector
FLAG-CKB
FLAG-ATXN10
FLAG-BTF3
FLAG-HABP4
FLAG-BZW1
GFP-SUMO1WT + - + - + - + - + - + -
GFP-SUMO1ΔGG - + - + - + - + - + - +
UBC9 + + + + + + + + + + + +
M
250
150
100
75
50
FLAG IP
IB: α-GFP
50
37
25
Lysates
α-FLAG
α-GFP
α-UBC9
α-α-tubulin

Figure S4

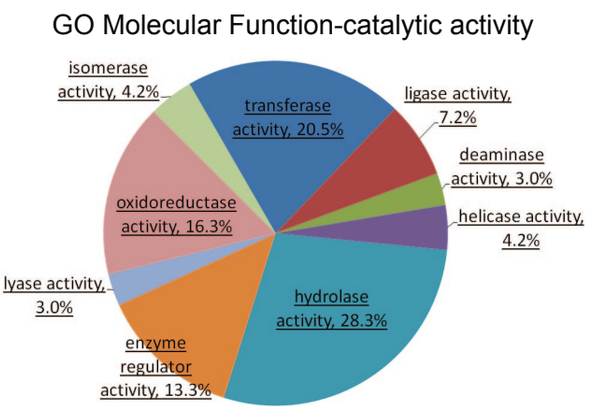
A



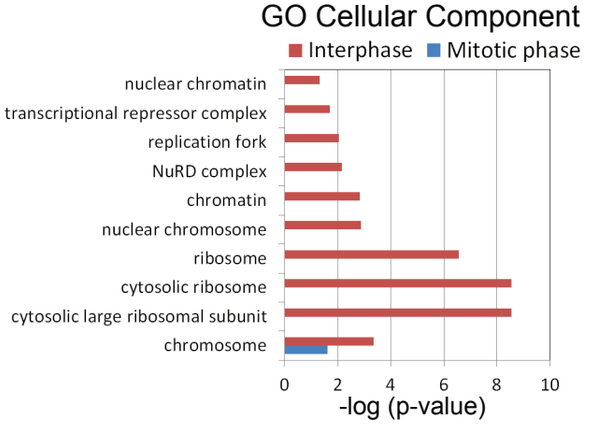
B



C



D



E

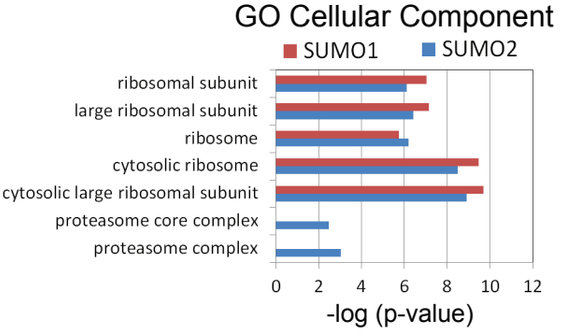
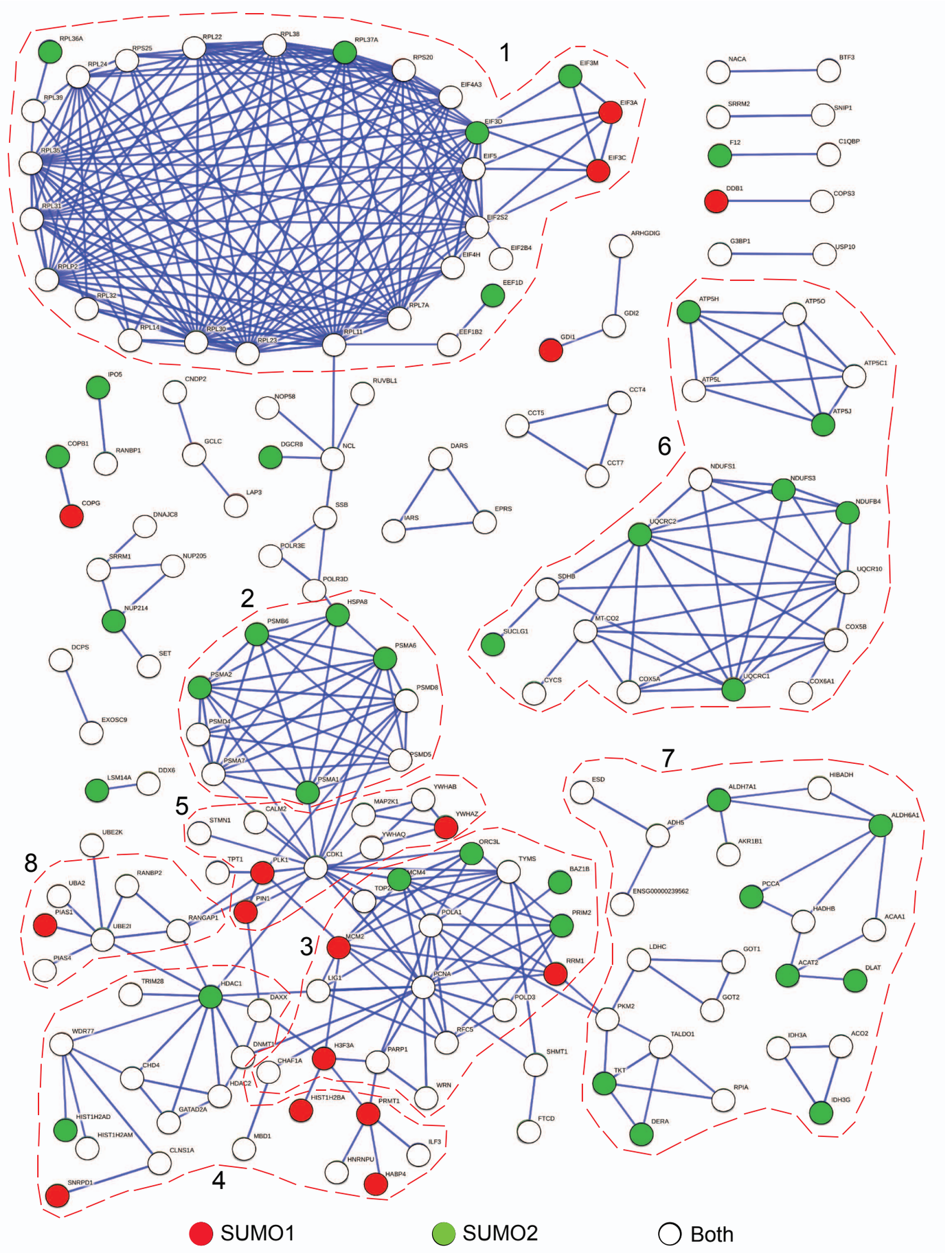


Figure S5


1
2
3
4
5
6
7
8
RPL36A
RPS25
RPL22
RPL38
RPL37A
RPL24
RPS20
EIF4A3
RPL39
EIF3M
EIF3A
EIF3D
RPL35
EIF5
EIF3C
RPL31
EIF2S2
EIF2B4
EIF4H
RPLP2
EEF1D
RPL32
RPL7A
RPL14
RPL30
RPL23
RPL11
EEF1B2
NACA
BTF3
SRRM2
SNIP1
F12
C1QBP
DDB1
COPS3
G3BP1
USP10
ARHGDIG
GDI2
GDI1
ATP5H
ATP5O
ATP5C1
ATP5L
ATP5J
IPO5
CNDP2
RUVBL1
NOP58
NCL
DGCR8
GCLC
COPB1
RANBP1
COPG
LAP3
CCT4
CCT5
CCT7
DARS
IARS
EPRS
DNAJC8
SRRM1
NUP205
NUP214
SET
SSB
POLR3E
POLR3D
DCPS
EXOSC9
HSPA8
PSMB6
PSMA6
PSMA2
PSMD8
PSMD4
PSMD5
PSMA7
PSMA1
LSM14A
DDX6
NDUFS1
NDUFS3
NDUFB4
UQCRC2
UQCR10
SDHB
MT-CO2
COX5B
SUCLG1
CYCS
COX5A
UQCRC1
COX6A1
ESD
HIBADH
ALDH7A1
ALDH6A1
ADH5
AKR1B1
ENSG00000239562
PCCA
HADHB
ACAA1
LDHC
GOT1
ACAT2
DLAT
GOT2
PKM2
TALDO1
IDH3A
ACO2
TKT
RPIA
IDH3G
DERA
YWHAB
MAP2K1
CALM2
YWHAZ
STMN1
YWHAQ
CDK1
ORC3L
TPT1
PLK1
TYMS
MCM4
BAZ1B
TOP2
PIN1
POLA1
PRIM2
MCM2
RRM1
PCNA
LIG1
POLD3
RFC5
SHMT1
PARP1
H3F3A
WRN
PRMT1
HABP4
HNRNPU
ILF3
FTCD
UBE2K
UBA2
RANBP2
PIAS1
UBE2I
RANGAP1
PIAS4
TRIM28
HDAC1
DAXX
WDR77
DNMT1
CHAF1A
CHD4
HDAC2
GATAD2A
HIST1H2BA
HIST1H2AD
HIST1H2AM
MBD1
CLNS1A
SNRPD1
SUMO1
SUMO2
Both