From Beginning to End: Expanding the SERINC3 Interactome Through an *in silico* Analysis

Mckenzie Tu and Sarah Saputo [iD]

Department of Chemistry and Biochemistry, SUNY Brockport, Brockport, NY, USA.

ABSTRACT: The serine incorporator (SERINC) family of proteins are a family of multipass transmembrane proteins associated with biosynthesis of serine-containing phospholipids and sphingolipids. Humans have 5 paralogs, SERINC1-5, which have been linked to disease including variable expression in tumor lines and possessing activity as restriction factors against HIV-1. Despite recent studies, the cellular function of SERINC proteins have yet to be fully elucidated. The goal of this study as to investigate the role of SERINC3 by expanding upon its interactome. We used a variety of bioinformatic tools to identify cellular factors that interact with SERINC3 and assessed how sequence variation might alter these interactions. Analysis of the promoter region indicates that SERINC3 is putatively regulated by transcription factors involved in tissue-specific development. Analysis of the unique 3′-untranslated region of one variant of *Hs*SERINC3 revealed that this region serves as a conserved site of regulation by both RNA binding proteins and miRNA. In addition, SERINC3 is putatively regulated at the protein level by several posttranslational modifications. Our results show that extra-membrane portions of SERINC3 are subject to variation in the coding sequence as well as areas of relatively low conservation. Overall, our data suggest that regions of low homology as well as presence of variations in the nucleotide and protein sequences of *Hs*SERINC3 suggest that these variations may lead to aberrant function and alternative regulatory mechanisms in homologs. The functional consequences of these sequence and structural variations need to be explored systematically to fully appreciate the role of SERINC3 in both health and disease.

KEYWORDS: Serine incorporator, SERINC3, *in silico* characterization, homology modeling

Introduction

Serine incorporator (SERINC) proteins constitute a unique protein family that show minimal amino acid homology to other proteins but are highly conserved among eukaryotes.[1,2] Yeast possess a similar membrane protein, dubbed TMS1, which localizes to the vacuolar membrane and exhibits modest homology to mammalian homologs.[3,4] Humans encode five paralogs that contain between 8 and 11 transmembrane domains that are characteristic to SERINC-family proteins. These SERINCs were originally named for their proposed ability to incorporate serine into membranes as phosphatidyl-serine or sphingolipids.[2] Localization studies have revealed that SERINC3 and SERINC5 are present in the perinuclear region, Golgi apparatus (SERINC3) as well as the plasma membrane.[5]

In multiple model systems, SERINC-family proteins have been linked to membrane trafficking. Both SERINC1 and SERINC3 in *Homo sapiens* were found to be cargo proteins that act in trafficking to exchange intermediates between cellular compartments.[6] Cells deficient in the adaptor complex 4 (AP-4) possessed aberrant localization of SERINC1 and SERINC3. Both SERINC proteins colocalize with the autophagy-related protein 9A (ATG9A) and interact with AP-4 complex factors. The five adaptor protein complexes observed in humans act in distinct pathways to regulate transport of vesicles to distinct cellular localizations. The clathrin-independent complexes have specifically been associated with the trans-golgi network, facilitating transport from the Golgi apparatus to the early endosome and plasma membrane.[7] Disruption of AP-4-associated transport has been linked to several forms of spastic paraplegia, a disease associated with weakness and abnormal gait.[8]

The functions of SERINC-family proteins have been expanded to include functioning as restriction factors against gamma-retroviruses in *Mus musculus* and lentiviruses in *Homo sapiens*.[5,9] Members of this protein family have been demonstrated to function by impairing the penetration of the viral particle into the cytoplasm through a mechanism dependent on Nef, a HIV1 accessory protein. Viral accessory proteins, like Nef, play a significant part in viral replication and infection.[10] Protein systems responsible for host cell trafficking are hijacked and can function in immune cell circumnavigation.[11] Of the SERINC proteins encoded in the human genome, SERINC3 and SERINC5 possess the greatest activity to inhibit Nef-defective virus infectivity upon ectopic expression in "low Nef-responsive" cells.[9] In the absence of Nef, SERINC3 is successfully incorporated into viral particles preventing delivery of the viral core by inhibiting the expansion of the fusion pore.

The structures of *Hs*SERINC5 and the ortholog from *Drosophila melanogaster* were elucidated confirming the presence of a multipass helical structure as well as a well-defined lipid binding groove.[12] Our analysis of the SERINC3 structure confirms the presence of the 11 transmembrane helices that are

conserved in most SERINC family proteins. Other studies have taken a step further to assign cellular roles to specific structures and posttranslational modifications. For example, mutational analysis revealed the key amino acids that are associated with the ability of SERINC5 to localize to the plasma membrane and possess the ability to restrict HIV-1 infection, an activity consistent with its association with AP-4.[6] Another study linked the cellular functions of SERINC5 in HIV-1 restriction have been linked to posttranslational modification and proteasomal degradtion.[13] Based on our findings, it is likely that SERINC3 undergoes a similar mechanism of regulation as SERINC5. However, these findings need to be validated using *in vitro* and *in vivo* model systems.

Although the SERINC-family proteins were first described in relation to their differential expression in tumor cell lines,[14,15] disruption of homologs have observed in other diseases. For example, variants of other serine-family proteins have been identified with links to alcohol dependence.[16] Allelic variations may also be associated with differential ability to interact with or restrict HIV-1 infection.[12] Despite these recent studies, the cellular roles and functional network associated with SERINC proteins have yet to be fully characterized. Therefore, we chose to further investigate the function, structure, and regulation of the *Hs*SERINC3. The goal of this study was to use an *in silico* approach to conduct an in-depth analysis of the SERINC3 genomic loci, protein structure, and functional networks.

Materials and Methods

SERINC3 sequences

The data on the human SERINC3 gene, including sequences and single nucleotide polymorphisms, were collected from Entrez Gene on the National Center for Biological Information (NCBI) website.

Promoter analysis

Pairwise alignment of promoter sequences was conducted using the EMBOSS Sequence Alignment tool (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) with the BLOSUM62 matrix and a default gap open penalty of 14.[17] Prediction and conservation of transcription factor binding sites within the SERINC3 promoter was completed with Ciiider (http://ciiider.com/)[18] using a matrix of transcription factor binding profiles[19] and a deficit of 0.1. The promoter was defined as the 1000 nucleotides preceding the SERINC3 start codon as done previously.[20] Conservation of the promoter region (chr20: 43150589-43151592) was analyzed using the Dcode.org tool developed by the Ovcharenko lab.[21] To do this, the Evolutionary Conserved Regions (ECR) browser was used with the following parameters: graph (smooth), ECR length (100), ECR similarity (70), layer height (55), and coordinate system (relative).

Gene ontology analysis

The proteins that interact with SERINC3 were analyzed with the GeneOntology tool (http://geneontology.org/),[22,23] which allows for the categorization of proteins based on annotated biological function. The analysis type was PANTHER Overrepresentation Test (Released 20210224) using GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18. The embedded Fisher exact test was used to calculate the *P* value and a false discovery rate cut-off of .01 was used.

3′ UTR characterization

Analysis of the 3′ untranslated region was done using RBPSuite (http://www.csbio.sjtu.edu.cn/bioinf/RBPsuite/)[24] and miRDB (http://mirdb.org/).[25] The RBPSuite and miRDB provides the probability that the analyzed sequence has a binding site for a RNA binding protein or microRNA, respectively. The GeneOntology tool allows for the categorization of proteins based on the annotated biological function.

Posttranslational modification prediction

Posttranslational modification prediction was performed using the following tools and cut-offs: NetPhos (https://services.healthtech.dtu.dk/service.php?NetPhos-3.1)[26] (threshold: 0.90), NetGlyc (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0)[27] (threshold: 0.50), ASEB (http://cmbi.bjmu.edu.cn/huac)[28] (*P* value cut-off .05), and Biocuckoo (http://pail.biocuckoo.org/)[29] (balanced cut-off option). These tools predict the presence of sites of posttranslational modification in the by comparing with a database of previously characterized sequences. When possible, redundant analyses were performed on these tools to confirm results.

Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) were downloaded from the NCBI Variation Viewer (https://www.ncbi.nlm.nih.gov/variation/view). The effect of select SNPs on protein stability was analyzed with I-Mutant2.0 (https://folding.biofold.org/i-mutant/i-mutant2.0.html), an online support vector machine.[30] The ΔΔG value was calculated at 25°C from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the wild type.[30]

Structure analysis

Two-dimensional (2D) projections of SERINC3 were prepared using the web-accessible Protter software (http://wlab.ethz.ch/protter/start/).[31] Colors and red lines were added using Adobe Illustrator (Adobe Systems, San Jose, CA, USA). Analysis of the relative conservation of the SERINC3 compared with homolog sequences was performed using the

ConSurf Server (https://consurf.tau.ac.il/).[32] ConSurf uses a multiple protein alignment data to predict the relative conservation of amino acids. The search was done using the HMMER homolog search algorithm, a E-value cut-off of 0.0001, and the UNIREF-90 protein database.

Results

Study design

The activity of proteins is controlled at different levels in a hierarchy that is suited to its cellular role. One way to characterize the function of a protein is to identify the set of factors that interact with a protein of interest. As the cellular role of SERINC3 is not completely understood, we reasoned that identification of cellular factors that interact with SERINC3 would further elucidate its function. Starting with analysis of the SERINC3 promoter, we used an *in silico* approach to expand on the interactome of SERINC3 at each step of the central dogma. Regulation at the RNA level has the potential to occur through several different mechanisms;[33] however, as the 3′-untranslated region (UTR) of SERINC3 is unique among SERINC paralogs, we analyzed this region for sequences that matched binding sites for characterized microRNAs (miRNAs) or RNA binding proteins. Finally, we analyzed the protein structure of SERINC3 to search for amino acids that be modified by posttranslational modifications (PTMs). At each point, DNA, RNA, and protein, conservation of sequences was also assessed.

Regulation of SERINC3 expression

The expression of SERINC3 has been reported to be dysregulated in tumor cell lines.[34] Although the presence of a 5' enhancer has been documented,[15,35] little is known about the proteins that control the expression of SERINC3. To investigate the regulation of the SERINC3 gene, we used an *in silico* prediction tool to locate putative sites for the binding of transcription factors. Use of the Ciiider toolkit allowed for the analysis and visualization of putative transcription factor binding sites based on data contained in position frequency matrices.[18] Using the matrices developed by Khan et al,[19] we analyzed the promoter of SERINC3, as defined as the 1000 base pairs prior to the ATG-site. Our results indicated the presence of 579 putative binding sites that matched the consensus sequences of 238 regulatory proteins with a deficit of 0.1 set by the Ciiider software (Supplemental Table S1). Gene ontology (GO) analysis of the transcription factors with putative binding sites revealed a majority involved in differentiation (110), tissue-specific morphogenesis (104), or development (203) (Supplemental Table S2). For example, 15 transcription factors that have been previously documented for their roles in kidney development have a total of 41 sites in the *Hs*SERINC3 that match the respective consensus sequences (Figure 1). There are also 16 putative binding sites for the Lhx1 regulator in the *Hs*SERINC3 promoter, both on the coding and noncoding strands, which have been omitted from Figure 1 for clarity. Of the 15 transcription factors involved in kidney development that bind the *Hs*SERINC3 promoter, only the consensus sites of GATA3, Lhx1, Pax2, OSR2, and Smad4 are found in the coding strand of *Hs*SERINC3.

Next, we considered the conservation of the SERINC3 promoter and asked if transcriptional regulation would be altered in other model organisms. Pairwise analyses of the SERINC3 promoter sequences from human and select model organisms revealed conservation of the promoter region in primates. Compared with the human SERINC3 promoter, the percent identity was the highest with *Pan troglodytes* (98.8%), *Gorilla gorilla* (98.8%), and *Maacaca mulatta* (87.7%) (Figure 1). The SERINC3 promoter sequence was somewhat less conserved in other mammals, including *Sus scrofa* (pig, 64.4%), *Cricetulus griseus* (Chinese hamster, 47.7%), and *Mus musculus* (mouse, 44.5%), for example.

The presence of putative transcription factor binding sites in the SERINC3 promoter region led us to next ask if the respective binding regions within the promoter were conserved. As before, the SERINC3 promoter sequences were retrieved from NCBI and the Ciiider software tool[18] was used to both align the sequences and predict putative sites of transcription factor binding. The conservation of the putative sites transcriptional regulator binding associated with the GO term "kidney development" was mapped on the SERINC3 of *Homo sapiens* and 13 model organisms (Figure 1). Of the 15 transcription factors involved in kidney development with conserved consensus sites in the *Hs*SERINC3 promoter, many of the consensus sites were conserved. Among primates SERINC3 promoters, regulation appears to be highly conserved among the set of transcription factors that include FOXC1, SOX8, Pax2, and others. In other mammals, a reduced level of sequence conservation meant that the putative binding sites for regulators were less conserved. For example, the SERINC3 promoter of *Felis catus* had 58.7% identity to the human promoter. Our results demonstrated that only 4 (SOX4, Nhx3-1, Pax2, and Smad4) regulators are predicted to bind the *Fc*SERINC3 promoter of the 15 regulatory proteins with the *Hs*SERINC3 promoter.

Taken together, these results suggest that the transcription of SERINC3 is controlled by a variety of transcriptional regulators and may contribute to tissue-specific development. In addition, our analysis suggests that SERINC3 expression may be regulated in a similar manner in primates to that of humans, whereas it may be subject to different transcriptional regulation in other mammals.

The 3′ UTR controls expression of SERINC3 at the RNA level

The role of both coding and noncoding RNA has expanded in recent decades to include an additional layer of control over
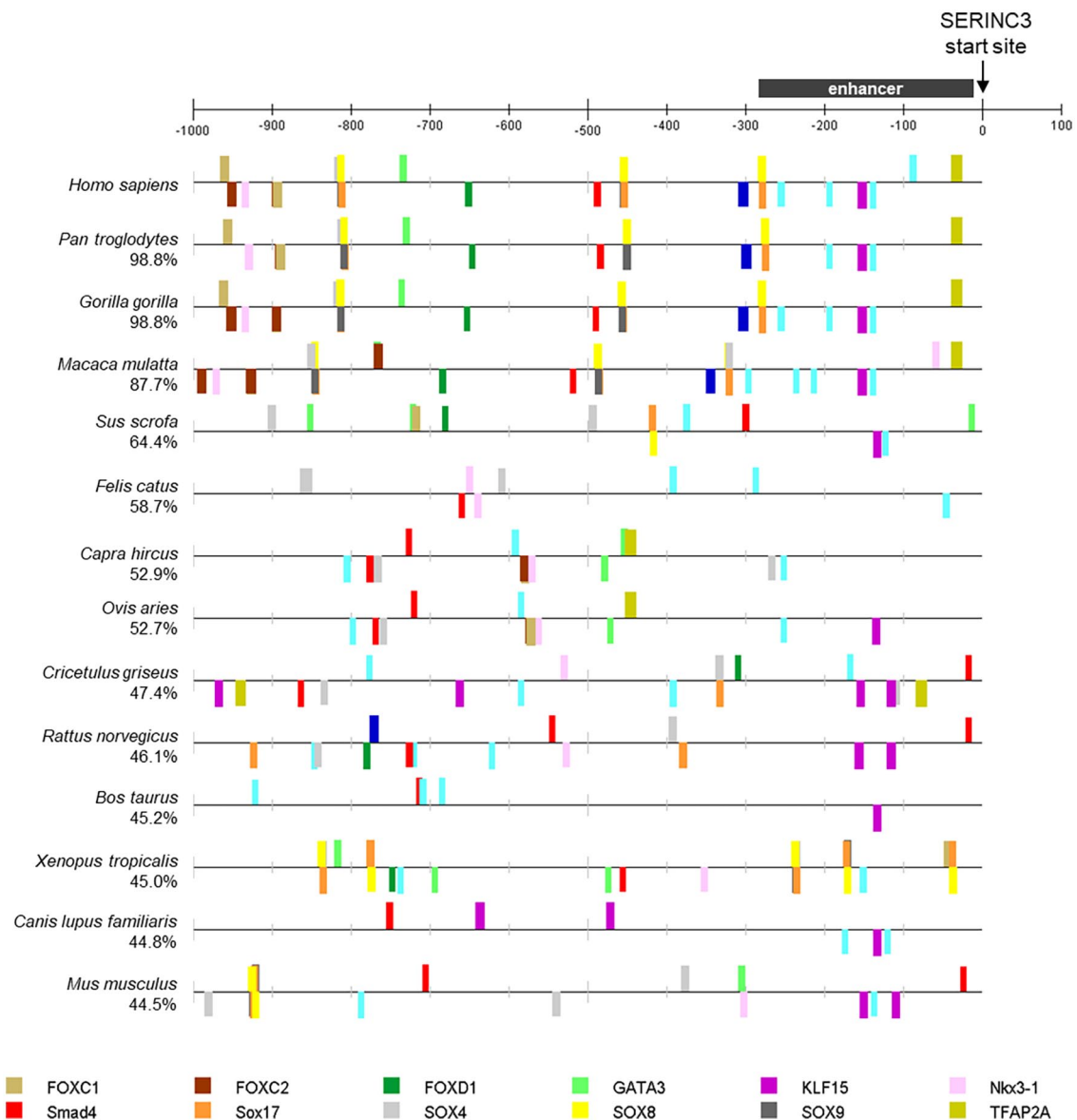
Figure 1. SERINC3 promoter analysis. Putative transcription factor binding sites were mapped to promoters using the Ciiider Software and the JASPAR 2018 core vertebrae set of matrices. Promoters were defined as the 1000 nucleotides preceding the SERINC3 start codon. SERINC3 promoters from model organisms were analyzed for percent identity compared to the *Hs*SERINC3 promoter using EMBOSS matcher (noted in italics). Conservation of select putative transcription factor binding sites associated with transcriptional regulation of human kidney development are shown, with the exception of Lhx1 which was omitted for clarity.

eukaryotic gene expression. The presence of a ~2.8 kb untranslated region in the 3′ region variant 1 of SERINC3 suggested a method of alternative control that is unique among human SERINC paralogs. To gain insight on the function of the SERINC3 3′UTR, we examined the 3′UTR sequence for putative binding sites for regulatory proteins and microRNA (miRNA). The sequence corresponding to the SERINC3 3′UTR, according to NCBI (ch20: 44500295-44497441), was used as input for analysis with the RBPSuite[24] and miRDB[25] to identify putative sites for regulation by proteins and miRNA, respectively.

First, putative sites for protein binding were detected in segments of 101 nonoverlapping segments of the SERINC3 3′UTR using a score threshold of 0.90. A total of 144 proteins binding 807 sites were identified (Figure 2, top and Supplemental Table S3). This analysis revealed the presence of potential hotspots of regulation by RNA binding proteins, most notably in segments 3, 13, 21 with 103, 108, 83 RBPs binding each respective region. The poly-A region, corresponding to segment 29, was also identified as a potential hotspot of protein binding with a predicted 98 binding sites. Using the gene ontology tool, we were able to further classify the proteins with predicted sites in the SERINC3 3′UTR. GO analysis[23] indicated that the proteins predicted to bind the SERINC3 3′UTR have roles in mRNA stability, transport, processing, and others (Table 1).
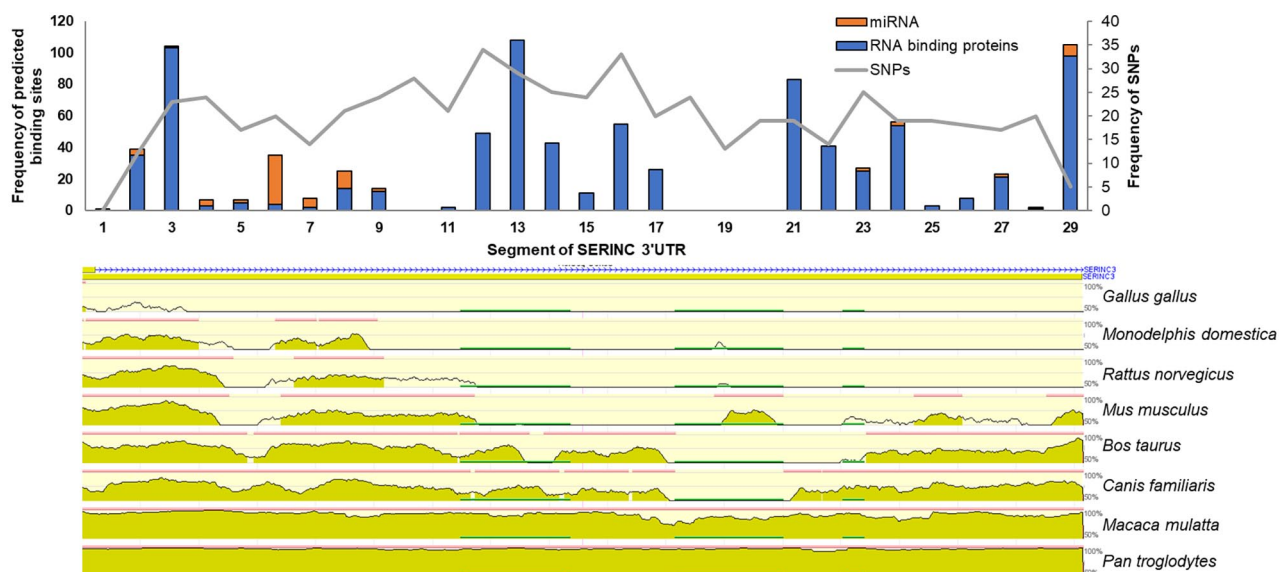
Figure 2. Analysis of SERINC3 3′UTR. *Top*: Sites of predicted binding by miRNA and RNA binding proteins as predicted by miRDB and RBPSuite, respectively. The 3′UTR of SERINC3 was retrieved from NCBI and was the 2854 nucleotides after the stop codon. Plot shows the number of predicting binding sites (left vertical axis) for RNA binding proteins (blue) and miRNA (orange) and the frequency of SNPs (gray line, right axis) in nonoverlapping segments of 101 nucleotides. *Bottom*: Axis for each organism corresponds to the percent homology relative to HsSERINC3 3′UTR. Colored regions are indicated as yellow for untranslated regions, green for simple repeats.

Table 1. Gene ontology terms associated with the proteins predicted to bind the 3′UTR.

GO BIOLOGICAL PROCESS	GO ACCESSION	*P* VALUE
establishment of RNA localization	51 236	2.58E−09
gene expression	10 467	1.01E−33
gene silencing by RNA	31 047	9.12E−09
IRES-dependent viral translational initiation	75 522	2.89E−05
mRNA 3′-end processing	31 124	5.88E−13
mRNA cleavage involved in mRNA processing	98 787	5.36E−05
ncRNA metabolic process	34 660	1.66E−11
negative regulation of RNA metabolic process	51 253	1.25E−04
negative regulation of translation	17 148	9.48E−10
protein export from nucleus	6611	5.30E−07
regulation of gene expression	10 468	3.49E−15
regulation of mRNA processing	50 684	9.56E−21
regulation of mRNA stability	43 488	2.81E−05
regulation of translation	6417	6.92E−14
RNA export from nucleus	6405	2.24E−07
RNA transport	50 658	2.33E−09
viral process	16 032	9.89E−05

Next, the SERINC3 3′UTR was analyzed with the miRDB tool to detect putative sites of miRNA binding. A total of 25 putative regulatory sites was identified and 61 miRNAs with a target prediction score greater than 80 (Supplemental Table S4). We used a lower threshold for the miRDB tool based on the likelihood of true-positive hits as determined by the creators. Similar to the RBPs, the miRNAs appeared to bind in hotspots in the SERINC3 3′UTR (Figure 2, top). These hotspots correspond to segments 6, 7, 8, and 29, each possessing 31, 6, 11, and 7 sites, respectively. With the exception of the poly-A tail, located in segment 29, these miRNA binding hotspots do not correspond to segments with a high frequency of putative RBP sites.

The pattern of putative regulatory sites in the 3′UTR led us to next ask if these sites were conserved. An alignment of the region following the stop codon was performed the web-based Dcode.org tool[21] (Figure 2, bottom). The species that exhibited the highest homology to the 3′UTR of *Hs*SERINC3 were *P. troglodytes, M. mulatta,* and *C. familiaris* supporting the previous finding that the SERINC3 is highly conserved in mammals.[2] The conserved regions roughly lined up to the hotspot regions of predicted sites for RNA binding proteins and miRNA binding (Figure 2). For example, our analysis suggested that the miRNA binding hotspots seen in segments 6 to 8 of *Hs*SERINC3 appear to be conserved in *Monodelphis domestica, Rattus norvegicus,* and *M. musculus.* In addition, the sequences corresponding to the poly-adenine tail, which also were predicted hotspot regions for RNA binding regions for RBPs and miRNAs, were also conserved. In contrast, segments 21 to 24 that contained 203 predicted binding sites for 103 RBPs, there was a noticeable lack of homology among the analyzed sequences. Interestingly, although the 3′UTR of *M. musculus* displayed relatively low homology relative to the primate sequences, there was a select region between segments 19 and 20 that possessed higher similarity relative to the surrounding sequences. Although no binding sites were identified in this region, other structural features may have a role in SERINC3 regulation.

Variation in the sequence of SERINC3 3′UTR can also be observed through SNPs. Regions with an elevated frequency of SNPs in this region may have altered mechanisms of regulation of homologs. SNP data were retrieved from the NCBI Variation Viewer and we found a total of 580 documented nucleotide variations was observed in this region (Supplemental Table S5). The frequency of documented sequence variation was plotted against the segments of the 3′UTR of SERINC3 (Figure 2). We observed that a range of 0 to 34 SNPs per 101 nucleotide segments throughout the 3′UTR of SERINC3. There was no obvious correlation between SNP frequency and putative binding sites of RBPs or miRNAs. The regions of the lowest SNP frequency were in segments 1 and 29, the beginning and end of the 3′ region. The central region possessed the highest frequency of SNPs with the highest being 34 SNPs throughout the span of the 101 nucleotides of segment 12 (Figure 2). It is possible that the higher frequency of SNPs in this region may result in altered regulatory sequences and differential control. Overall, our analysis of the 3′UTR of SERINC3 reveals that this region serves as a conserved site of regulation for mRNA that may be altered by relative conservation and variation in sequence. In addition, the presence of numerous putative sites of regulation suggests that the 3′UTR of SERINC3 is highly regulated to alter properties such as RNA half-life and localization.

Posttranslational modification of the SERINC3 protein

The two variants of SERINC3 encoded by *Homo sapiens* are predicted to yield the same protein product. However, due to the difficulty associated with purification of membrane proteins, the structure of *Hs*SERINC3 has not been elucidated. To gain insight on the cellular roles of SERINC3, the protein sequence was examined for putative sites of posttranslational modification.

The web-based tool Protter was used to determine the amino acids that are amendable to modification by predicting the membrane topology of SERINC3. Consistent with other SERINC family proteins, the topology of SERINC3 protein contained 11 transmembrane domains[1,2] (Figure 3). This tool also verified the presence of 3 glycosylation sites, N33 and N187 on the Golgi side and N314 that is exposed to the cytoplasm.

Next, we used a series of web-based tools to predict the presence of posttranslational modifications. We limited our search for residues that were predicted to be exposed to the cytoplasm or Golgi side of the membrane (Figures 3). According to our findings, SERINC3 has several putative modifications on both the Golgi- and cytoplasmic-facing regions. The modification included ubiquitination sites (8), phosphorylation (18), acetylation (1), and *N*-glycosylation (3). The sequence of SERINC3-contained sites that matched the consensus phosphorylation sites of PKC, PKA, CKI, DNAPK, and cdk5.[26] One predicted acetylation site on the cytoplasmic side of the membrane was detected at K266 (Figure 3). Based on prediction of consensus sites, the SIRT1 enzyme was predicted to modify K266 with a *P* value of .0298.

A similar analysis revealed putative ubiquitination sites using the web-based tool, Biocuckoo[29](Figure 3). A total of eight sites were predicted, 3 on the cytoplasmic side of the membrane and 5 on the Golgi side. Interestingly, the lysine residues at positions 33, 118, 120, 123, and 328 were located proximal to other sites of predicted PTM. This site was near a site of predicted phosphorylation suggesting the potential for competing modifications.
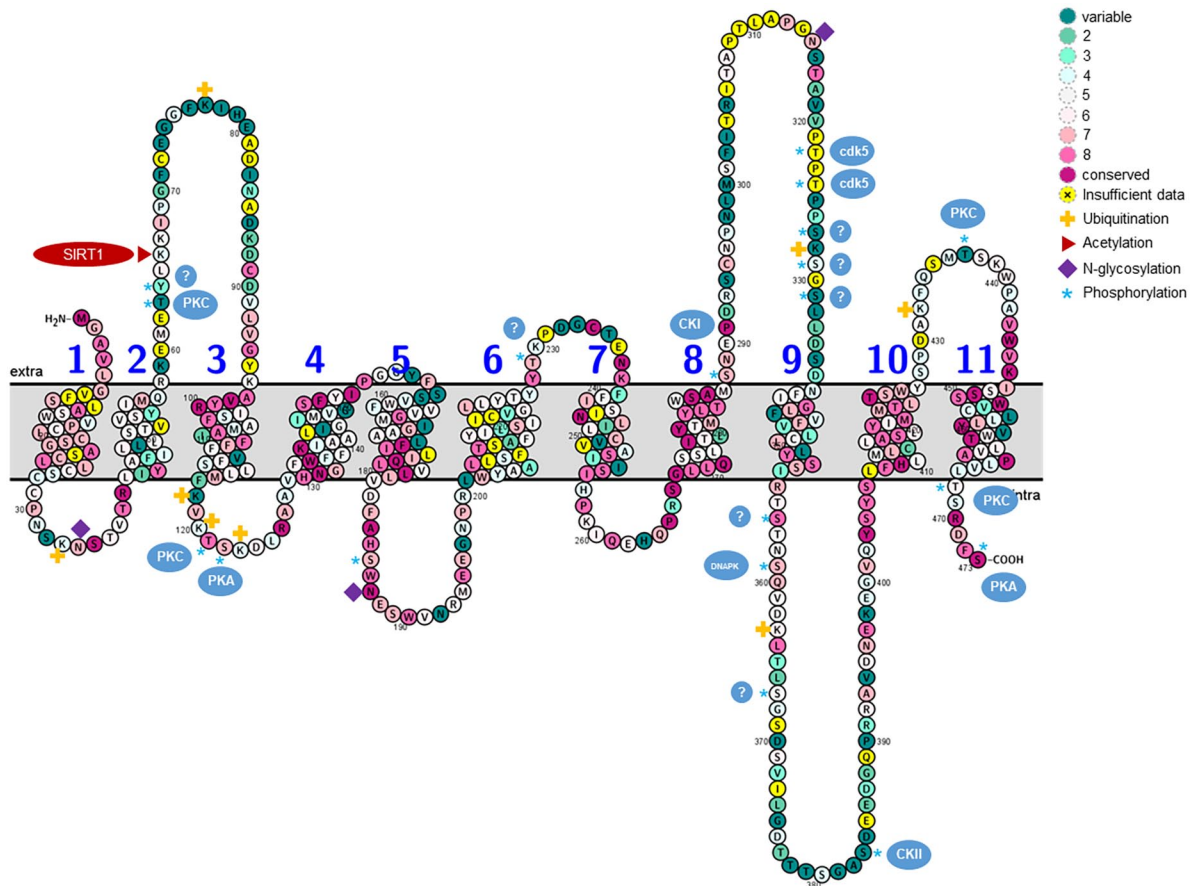
Figure 3. Prediction of the structure, modification, and relative sequence conservation of *Homo sapiens* SERINC3. Amino acid residues were colored according to the relative conservation when compared with >50 homologous sequences. Predicted PTM including sites of ubiquitination (+), acetylation (triangle), glycosylation (diamond), and phosphorylation (*) are labeled with putative enzyme, if known.

Variation in sequence may alter SERINC3 regulation and structure

The level of variation in protein regions is strongly dependent on its structural and functional importance within a protein. Therefore, we next asked if these sites of predicted PTM were conserved or susceptible to variability through SNPs. A previous study revealed that the presence of point mutations in SERINC5 glycosylation sites resulted in mislocalization as well as failure to successfully incorporate into the HIV-1 viron.[12] Although SERINC family proteins are well-conserved in mammals, we asked if the sites of putative regulation were susceptible to variation. To answer this question, we evaluated the sequence for conservation among homologs and for the presence of single nucleotide polymorphisms.

Using the ConSurf tool[32] to align select homologous sequences, we analyzed the relative conservation of each amino acid in the SERINC3 sequence (Figure 3 and Supplemental Table S5). Based on a multiple sequence alignment of more than 50 homologous sequences, each amino acid in SERINC3 was scored for the relative conservation on a scale of 1 (indicating a variable residue) to 9 (indicating a conserved residue) (Figure 3). Our analysis showed that the exposed loops corresponding to the regions between helices 2 and 3, 8 and 9, and 9

and 10 showed the greatest variability in sequences relative to homologs. A majority of the putative ubiquitination sites (7/8) were ranked with a score of 5 or less according to the analysis performed by ConSurf indicating that these sites are not highly conserved. Of the predicted phosphorylation sites outside of the membrane, 8 were conserved, having ConSurf-associated homology scores of 6 or above. Overall, the considerable variation in amino acid sequence in homologs suggests that regulatory mechanisms may not be conserved in other model organisms.

Next, we considered variation in SERINC3 that occurs through documented SNPs. Missense point mutations in the coding sequence of SERINC3 were retrieved using the NCBI Variation Viewer. In total, SERINC3 protein exhibited 321 documented SNPs resulting in a missense mutation, where 173 of those were located outside of the membrane regions (Supplemental Table S4). Our analysis of SNPs revealed that the loop regions were susceptible to increased variability relative to the transmembrane regions. Specifically, we observed the presence of a ratio of 0.82 SNPs: residues in the exposed regions, compared with the transmembrane regions which only had a ratio of 0.64. These data suggest that the regions exposed to the cytoplasm or Golgi apparatus are susceptible to more

Table 2. Frequency of single nucleotide polymorphisms (SNPs) in cytoplasmic or Golgi-exposed regions of SERINC3.

	# SNPS	# RESIDUES	# SNP/REGION
N-term	4	5	0.80
loop 1-2	12	11	1.09
loop 2-3	22	39	0.56
loop 3-4	8	12	0.67
loop 4-5	4	5	0.80
loop 5-6	6	21	0.29
loop 6-7	11	11	1.00
loop 7-8	9	11	0.82
loop 8-9	38	49	0.78
loop 9-10	34	53	0.64
loop 10-11	16	19	0.84
C-term	9	6	1.50

Table 3. SERINC3 structural stability based on free energy change.

# VARIANT ID	RESIDUE CHANGE	ΔΔG (KCAL/MOL)
rs768124514	Lys33Asn	−1.02
rs1555830242	Ser122Thr	0.31
rs1555830242	Ser122Asn	0.04
rs748350977	Ser122Arg	−0.28
rs185462055	Ser328Asn	−1.88
rs762874477	Ser331Thr	−1.51
rs762106624	Ser359Arg	−1.42
rs1254122914	Lys432Glu	−0.45
rs936670439	Thr468Ile	−0.01
rs201460770	Thr468Ala	−0.93
rs760368420	Ser473Cys	−1.91

variation than the transmembrane regions. Further analysis revealed that the 5-6 loop region exhibited the lowest number of SNPs as expected with its high number of conserved residues (Table 2). In contrast, loops 1-2, 6-7, as well as the C-terminal tail, exhibited the highest ratio of SNPs: amino acids.

The presence of SNPs can alter both the identity of the residue susceptible to PTM and the stability of the structure. The change in free energy of protein folding associated with the new amino acid was determined using the iMutant2.0 online tool[30] (Table 3 and S6). Of the SNPs we analyzed, 19 point mutations resulted in a ΔΔG that was greater than zero, indicating that the new amino acid would increase the stability of the SERINC3 tertiary structure. Our analysis revealed that the majority of SNPs resulted in a structure that is less stable (168 with ΔΔG > 0, Supplemental Table S6). Several missense mutations were also found to occur at a putative site of post-translational modification. These sites included the predicted site of acetylation (K66), phosphorylation (S327, S331, S359, T468, S473), and ubiquitination (K33, K328, K432). In contrast, the SNPs S122T and S122N, with ΔΔG of 0.31 and 0.04, respectively, would result in an increase in stability of the folded SERINC3 protein. Overall, these results suggest that sequence variation, as a result of relative conservation of the amino acids as well as individual nucleotide variation, have the potential to alter the structure and regulatory mechanisms of the SERINC3 protein.

Discussion

The SERINC proteins have emerged as proteins of interest because their cellular functions are not well characterized.

Since the initial observation of variable expression in tumor lines,[1] the function of SERINC family of proteins has been expand to include serine-containing phospholipid biosynthesis, cellular trafficking, as well as ability to restrict lentiviruses, such as HIV-1.[2,6] Herein, we used several bioinformatics tools to gain insight on the cellular factors that interact with SERINC3 at the level of DNA, RNA, and protein (Figure 4).

Our analysis of the SERINC3 promoter indicated the presence of putative transcription factor binding sites with roles in development of organs, including the generation of neurons (Figure 1 and Supplemental Table S2). Gene expression of SERINC3 in humans is also regulated by the enhancer present 16 nucleotides upstream of the start site.[35] Conservation of regions with predicted transcription factor binding sites suggests that regulation of SERINC3 at the transcriptional level might be conserved in mammals.

More than half of human genes use alternative cleavage and polyadenylation to generate alternative 3′UTR isoforms. Untranslated regions contain sequence-specific binding sites for proteins and regulatory RNA that can alter splicing, cellular location, and mRNA stability. The effects of variants on protein structure can vary dramatically depending on the type of protein and the extent of variation. Examination of the SERINC3 3′UTR revealed that the transcripts are differentially regulated. Based on the GO terms associated with the putative RBPs, the variant 1 transcript of SERINC3 may have an alternate location or relative stability in the cell. In support of this finding, roles for SERINC3 have been indicated in the plasma membrane as well as the Golgi apparatus.[6] In support of this, the SERINC proteins appear to be associated with different adaptor proteins associated that start in the Golgi and are destined for various organelles.[7] The roles of SERINC3 and paralogs in membrane trafficking directly links to the ini-
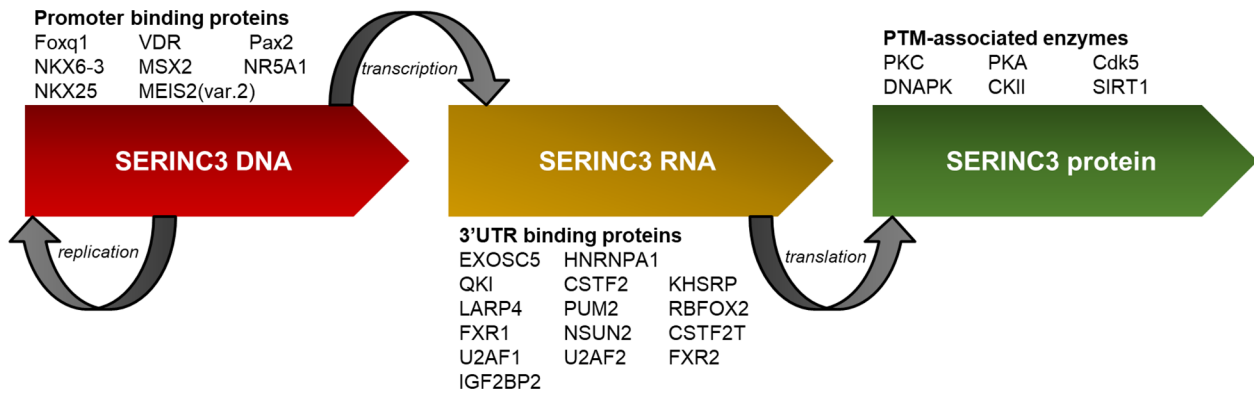
Figure 4. Model of the predicted SERINC3 interactome.

tial studies of variable expression in tumors, as vesicular trafficking can be highly up regulated in tumors.

The cellular activities of SERINC paralogs appear to be regulated by posttranslational modification. For example, SERINC4 protein is subject to degradation by the proteasome, which contributes to its activity in restricting HIV-1 replication.[36] In addition, *N*-glycosylation of SERINC5 by has been observed to be preferentially incorporated into HIV-1 virions.[13] Our search for putative sites of posttranslational modifications in SERINC3 extra-membrane regions yielded the prediction of sites of putative phosphorylation, glycosylation, acetylation, and ubiquitination. Addition of these functional groups to the primary structure of SERINC3 may be a mechanism to modulate protein functions and dynamically coordinate a signaling network.

Other studies that used high throughput techniques have detected the presence of sites of ubiquitination[37-39] and phosphorylation in SERINC3. Proximity of the PTMs sites to suggest that SERINC3 might be coordinately regulated. For example, sites of predicted phosphorylation by PKC or PKA are in close proximity on the 3-4 loop and C-terminal of SERINC3. Many of the residues with predicted PTM are also susceptible to missense mutations that could result in changes to the tertiary structure as well as regulation of SERINC3.

Prediction of an acetylation site at K266 in the 2-3 loop of SERINC3 by SIRT1 might add another layer of regulation. The enzyme sirtuin 1 (SIRT1) is a conserved enzyme that has been demonstrated to have roles in oxidative stress and other metabolic activities.[40] Although most commonly associated with modulation of histone activity, acetylases and deacetylases can target other proteins.

The role of SERINC3 in the cell as well as association to disease is unclear. According to the Cancer Genome Atlas, somatic mutations in SERINC3 have been observed in bladder, endometrial, and other cancers. Another large-scale study found SNPs in the sequence of SERINC3 associated with breast cancer[41] and progressive supranuclear palsy.[42] Similarly, our data of SNPs in *Hs*SERINC3 coding sequence have the potential alter its regulation as well as the stability of its folded structure that may contribute to abnormal cellular phenotypes. These findings are in agreement with a recent study of SERINC5 that found that variability within a cytoplasmic exposed region alters the ability to restrict HIV.[43] The exposed regions of SERINC3 be susceptible to SNPS in humans and have a considerable range of conservation at the level of amino acids. The variability in the SERINC3 protein sequence that we observed in our studies have the potential to alter the structure and function of SERINC3 at the cellular level. It is possible that these variations in SERINC3 sequence can generate cellular phenotypes that can contribute to disease, which is a future area of investigation.

In this study, we used an *in silico* approach to predict and characterize the functional network of SERINC3; as such, additional studies are required to validate these findings. Throughout the study, statistical significance was kept stringent to limit the number of predicted false-positive results (see "Methods" section). When possible, redundant analyses were performed using separate tools to confirm hits and reduce false-positives. *In vivo* analysis might reveal tissue-specific interactions or regulation. To our knowledge, our study is the first to investigate the SERINC3 interactome as well as a predictive analysis of variation and regulators at the level of DNA, RNA, and protein. Our data suggest that SERINC3 is regulated at the transcriptional level by several transcription factors and at the RNA level by RBPs and miRNAs. Several sites of predicted protein modification were also identified. The functional and structural impact of SNPs were also investigated using computational prediction tools. The results found here suggest that SERINC3 is coordinately regulated, and sequence variation have the potential to alter both protein structure and cellular function.

Author Contributions

MT and SS conceptualized the project and conducted the research. MT contributed to the preparation of the manuscript. SS made critical revisions. All authors reviewed and approved the final manuscript.

ORCID iD

Sarah Saputo https://orcid.org/0000-0001-8344-8748

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Bossolasco M, Lebel M, Lemieux N, Mes-Masson AM. The human TDE gene homologue: localization to 20q13.1–13.3 and variable expression in human tumor cell lines and tissue. *Mol Carcinog*. 1999;26:189-200.
2. Inuzuka M, Hayakawa M, Ingi T. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *J Biol Chem*. 2005;280:35776-35783.
3. De Hertogh B, Carvajal E, Talla E, Dujon B, Baret P, Goffeau A. Phylogenetic classification of transporters and other membrane proteins from Saccharomyces cerevisiae. *Funct Integr Genomics*. 2002;2:154-170.
4. Huh WK, Falvo JV, Gerke LC, et al. Global analysis of protein localization in budding yeast. *Nature*. 2003;425:686-691.
5. Rosa A, Chande A, Ziglio S, et al. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature*. 2015;526:212-217.
6. Davies AK, Itzhak DN, Edgar JR, et al. AP-4 vesicles contribute to spatial control of autophagy via RUSC-dependent peripheral delivery of ATG9A. *Nat Commun*. 2018;9:3958.
7. Park SY, Guo X. Adaptor protein complexes and intracellular transport. *Biosci Rep*. 2014;34:e00123.
8. Sanger A, Hirst J, Davies AK, Robinson MS. Adaptor protein complexes and disease at a glance. *J Cell Sci*. 2019;132:jcs2229992.
9. Usami Y, Wu Y, Göttlinger HG. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature*. 2015;526:218-223.
10. Kestler HW, Ringler DJ, Mori K, et al. Importance of the nef gene for maintenance of high virus loads and for development of AIDS. *Cell*. 1991;65:651-662.
11. Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature*. 1998;391:397-401.
12. Pye VE, Rosa A, Bertelli C, et al. A bipartite structural organization defines the SERINC family of HIV-1 restriction factors. *Nat Struct Mol Biol*. 2020;27:78-83.
13. Sharma S, Lewinski MK, Guatelli J. An N-glycosylated form of SERINC5 is specifically incorporated into HIV-1 virions. *J Virol*. 2018;92:e00753-18.
14. Liang P, Averboukh L, Pardee AB. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic Acids Res*. 1993;21:3269-3275.
15. Lebel M, Mes-Masson AM. Sequence analysis of a novel cDNA which is overexpressed in testicular tumors from polyomavirus large T-antigen transgenic mice. *DNA Seq*. 1994;5:31-39.
16. Zuo L, Wang KS, Zhang XY, et al. Rare SERINC2 variants are specific for alcohol dependence in subjects of European descent. *Pharmacogenet Genomics*. 2013;23:395-402.
17. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47:W636-W641.
18. Gearing LJ, Cumming HE, Chapman R, et al. CiiiDER: a tool for predicting and analysing transcription factor binding sites. *PLoS ONE*. 2019;14:e0215495.
19. Khan A, Fornes O, Stigliani A, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*. 2018;46:D260-D266.
20. Yang X, Vingron M. Classifying human promoters by occupancy patterns identifies recurring sequence elements, combinatorial binding, and spatial interactions. *BMC Biol*. 2018;16:138.
21. Loots GG, Ovcharenko I. Dcode.org anthology of comparative genomic tools. *Nucleic Acids Res*. 2005;33:W56-W64.
22. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25-29.
23. The Gene Ontology Consortium Carbon S, Douglass E, et al. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res*. 2021;49:D325-D334.
24. Pan X, Fang Y, Li X, Yang Y, Shen HB. RBPsuite: RNA-protein binding sites prediction suite based on deep learning. *BMC Genomics*. 2020;21:884.
25. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48:D127-D131.
26. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4:1633-1649.
27. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*. 2002:310-322.
28. Zhai Z, Tang M, Yang Y, Lu M, Zhu WG, Li T. Identifying human SIRT1 substrates by integrating heterogeneous information from various sources. *Sci Rep*. 2017;7:4614.
29. Zhou J, Xu Y, Lin S, et al. iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. *Nucleic Acids Res*. 2018;46:D447-D453.
30. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 2005;33:W306-W310.
31. Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*. 2014;30:884-886.
32. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44:W344-W350.
33. Hawkins P, Morris KV. RNA and transcriptional modulation of gene expression. *Cell Cycle Georget Tex*. 2008;7:602-607.
34. Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13:397-406.
35. Ernst J, Melnikov A, Zhang X, et al. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol*. 2016;34:1180-1190.
36. Qiu X, Eke IE, Johnson SF, Ding C, Zheng YH. Proteasomal degradation of human SERINC4: a potent host anti-HIV-1 factor that is antagonized by nef. *Curr Res Virol Sci*. 2020;1.
37. Wagner SA, Beli P, Weinert BT, et al. Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues. *Mol Cell Proteomics*. 2012;11:1578-1585.
38. Akimov V, Barrio-Hernandez I, Hansen SVF, et al. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat Struct Mol Biol*. 2018;25:631-640.
39. Udeshi ND, Svinkina T, Mertins P, et al. Refined preparation and use of anti-diglycine remnant (K-ε-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol Cell Proteomics*. 2013;12:825-831.
40. Alqarni MH, Foudah AI, Muharram MM, Labrou NE. The pleiotropic function of human sirtuins as modulators of metabolic pathways and viral infections. *Cells*. 2021;10:460.
41. Mertins P, Mani DR, Ruggles KV, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*. 2016;534:55-62.
42. Melquist S, Craig DW, Huentelman MJ, et al. Identification of a novel risk locus for progressive supranuclear palsy by a pooled genomewide scan of 500,288 Single-nucleotide polymorphisms. *Am J Hum Genet*. 2007;80:769-778.
43. Dai W, Usami Y, Wu Y, Göttlinger H. A long cytoplasmic loop governs the sensitivity of the anti-viral host protein SERINC5 to HIV-1 Nef. *Cell Rep*. 2018;22:869-875.