

# **Differential complex formation via paralogs in the human Sin3 protein interaction network.**

Mark K. Adams<sup>1</sup>, Charles A.S. Banks<sup>1</sup>, Janet L. Thornton<sup>1</sup>, Mihaela E. Sardi<sup>1</sup>, Maxime  
Killer<sup>1,3</sup>, Cassandra G. Kempf<sup>1</sup>, Laurence Florens<sup>1</sup>, Michael P. Washburn<sup>1,2,\*</sup>

<sup>1</sup> Stowers Institute for Medical Research, Kansas City, MO 64110

<sup>2</sup> Department of Pathology & Laboratory Medicine, University of Kansas Medical Center,  
Kansas City, KS 66160

<sup>3</sup> Current address: Centre for Structural Systems Biology (CSSB), DESY and European  
Molecular Biology Laboratory Hamburg, Hamburg, Germany

\*Correspondence: [mpw@stowers.org](mailto:mpw@stowers.org)

**Running Title: Sin3 protein domain organization**

## Abbreviations:

AP	affinity purification
APMS	affinity purification mass spectrometry
DMSO	dimethyl sulfoxide
dNSAF	distributed normalized spectral abundance factor
dS	distributed spectral counts
FBS	fetal bovine serum
FDA	Food and Drug Administration
FDR	false discovery rate
HDAC	histone deacetylase
HDACi	HDAC inhibitor
HID	HDAC interaction domain
HPLC	high performance liquid chromatography
KDRI	Kazusa DNA Research Institute
MudPIT	multidimensional protein identification technology
NLS	nuclear localization signal
PAH	paired amphipathic helix
RBBP4_a	RBBP4 isoform a
RBBP4_c	RBBP4 isoform c
RFU	relative fluorescence units
Rpd3L (Sin3L)	Sin3 Large
Rpd3S (Sin3S)	Sin3 Small
SAHA	suberoylanilide hydroxamic acid
SIN3B_1	SIN3B isoform 1
SIN3B_2	SIN3B isoform 2
SIN3B_3	SIN3B isoform 3

## ABSTRACT

Despite the continued analysis of HDAC inhibitor efficacy in clinical trials, the heterogeneous nature of the protein complexes they target limits our understanding of the beneficial and off-target effects associated with their application. Among the many HDAC protein complexes found within the cell, Sin3 complexes are conserved from yeast to humans and likely play important roles as regulators of transcriptional activity. The functional attributes of these protein complexes remain poorly characterized in humans. Contributing to the poor definition of Sin3 complex attributes in higher eukaryotes is the presence of two Sin3 scaffolding proteins, SIN3A and SIN3B. Here we show that paralog switching influences the interaction networks of the Sin3 complexes. While SIN3A and SIN3B do have unique interaction network components, we find that SIN3A and SIN3B interact with a common set of proteins. Additionally, our results suggest that SIN3A and SIN3B may possess the capacity to form hetero-oligomeric complexes. While one principal form of SIN3B exists in humans, the analysis of rare SIN3B proteoforms provides insight into the domain organization of SIN3B. Together, these findings shed light on the shared and divergent properties of human Sin3 proteins and highlight the heterogeneous nature of the complexes they organize.

## SUMMARY

Over 13000 or 70% of protein coding genes within the human genome have at least one paralog (1). The acquisition of additional copies of a gene through duplication events provides opportunities for the development of unique gene products with distinct regulatory mechanisms (2). Functional divergence can result from gene duplications and protein paralog identity can influence the composition of large protein complexes (3). However, the consequences of paralog switching are largely overlooked during the characterization of proteins, protein complexes, and protein interaction networks.

Classically associated with transcriptional repression, the removal of histone lysine acetyl groups by the Sin3 histone deacetylase (HDAC) complexes represents a central mechanism whereby transcriptional status is regulated (4). Named for the scaffolding protein of the complex, Sin3 complexes are well studied in *Saccharomyces cerevisiae* (5, 6). However, the presence of additional components not found in lower eukaryotic forms of the Sin3 complexes likely increases the diversity of complex function in higher eukaryotes. Contributing to this expansion of components is the acquisition of paralogous genes encoding Sin3 proteins. The two Sin3 paralogs present within mammals, SIN3A and SIN3B, have undergone significant divergence and maintain only 63% sequence similarity at the protein level in humans.

Genetic deletion of murine *Sin3a* results in early embryonic lethality whereas deletion of *Sin3b* induces late gestational lethality (7, 8). That SIN3A and SIN3B cannot compensate for the loss of one another provides evidence for paralog-specific functions within mammals and suggests that variations of the Sin3 complex have functional consequences. In addition to having critical roles during mammalian development, SIN3A and SIN3B also have contrasting influences on breast cancer cell metastasis. SIN3A was identified as a suppressor of metastasis,

whereas SIN3B was proposed to be pro-metastatic (9). However, the mechanisms responsible for divergent influences on development as well as cancer cell metastatic potential remain poorly understood.

SIN3A has been identified as one of the 127 most significantly mutated genes across multiple cancer types (10) and is considered a cancer driver gene (11). Not surprisingly, FDA-approved HDAC inhibitors (HDACis) are effective constituents of multicomponent cancer treatment therapies. While these compounds target the activity of the Sin3 complex catalytic subunits HDAC1/2 (12), heterogeneity within the population of HDAC complexes targeted by these compounds is likely responsible for the well-documented off-target effects associated with HDACis (13) and ultimately prevents our complete understanding of the mechanisms through which these compounds function. The evidence that SIN3A and SIN3B differentially influence metastatic potential, along with the existence of FDA-approved chemotherapeutic agents that target Sin3 complex function, warrants further investigation into the unique properties of complexes containing SIN3A and/or SIN3B as components. Here, we characterize human SIN3A and SIN3B, highlighting the unique and shared properties of Sin3 paralogs in humans. Together, our results shed light on the molecular targets of chemotherapeutic HDAC inhibitors and highlight consequences associated with paralog switching on Sin3 complex composition.

## **Results:**

***HDAC1/2 and RBBP4/7 are core complex components common to SIN3A- and SIN3B-containing complexes-*** SIN3A and SIN3B share only 63% sequence similarity at the amino acid level (Fig. S1). Therefore, we sought to determine how paralog identity influences protein complex composition. SIN3B isoform 2 (transcript NM\_001297595.1, NP\_001284524.1)

was chosen for analysis as this isoform most closely resembles SIN3A (Fig. S2) and also has strong support as the principal/main isoform within transcriptome databases (<https://gtexportal.org/home/gene/SIN3B>) (14). To assess differences in SIN3A (NP\_001138829.1) and SIN3B isoform 2 interaction networks, open reading frames encoding these proteins were stably expressed in Flp-In-293 cells with C-terminal HaloTags (Fig. 1A, B). Nuclear localization of both recombinant SIN3A and SIN3B isoform 2 (SIN3B\_2) was confirmed via fluorescence microscopy (Fig. 1A, B).

Recombinant proteins were affinity-purified and analyzed by MudPIT (Tables S1, S2A-C) as previously described (15). Statistically-significant protein enrichment over negative controls was assessed with QSPEC v 1.3.5 (16) (Fig. 1C,D, Tables S2D-E). Halo-tagged SIN3A captured 18 proteins while SIN3B\_2 captured 17 proteins (Fig. 1E, F). While SIN3A and SIN3B\_2 enriched a similar number of proteins, the identities of enriched proteins only partially overlapped. Only 7 proteins were enriched in both SIN3A and SIN3B\_2 purifications (Fig. 1E).

Within *S. cerevisiae*, 2 forms of the Sin3 complex exist, Rpd3L (Sin3L) and Rpd3S (Sin3S) (5, 6). These complexes share a core of proteins consisting of Sin3, Rpd3, and Ume1, which have homology to human SIN3A/B, HDAC1/2, and RBBP4/7, respectively (17, 18). Using dNSAF values as indicators of protein abundance, HDAC1/2 and RBBP4/7 were identified among the most abundant non-bait proteins in both SIN3A and SIN3B\_2 enrichments (Fig. 1F, Table S2F). Thus, SIN3A and SIN3B\_2 interact with a set of proteins that are homologous to shared core complex components present within both forms of yeast Sin3 complexes. Interestingly, SIN3A was enriched by both SIN3A and SIN3B\_2 (Fig. 1E, Tables S2D-E), suggesting that hetero-oligomeric forms of the Sin3 complexes may exist.

***Both Sin3 paralogs enrich homologs of yeast Sin3L-specific components but only SIN3B***

***enriches a Sin3S-specific component-*** While sharing a core of subunits, the two yeast Sin3 complexes are differentiated by the identities of additional subunits. Components specifically found within the Rpd3L complex include Sds3, Sap30, and Pho23. These proteins have homology to human SUDS3/BRMS1/BRMS1L, SAP30/SAP30L, and ING1/2, respectively (19–21). Notably, all of these proteins were enriched by SIN3A (Fig. 1E-F). Among these proteins, only SUDS3 and BRMS1L met statistical criteria for enrichment by SIN3B\_2 (Fig. 1E). While peptides mapping to BRMS1, ING1/2, and SAP30/SAP30L were observed following SIN3B purification (Fig. 1F), these proteins did not meet criteria for enrichment. These data suggest that SIN3B\_2 may interact with at least a subset of proteins with homology to Rpd3L-specific components.

The yeast Rpd3S complex also contains components with homology to human proteins: yeast Rco1 and Eaf3 have homology to human PHF12 and MORF4L1, respectively (22, 23). PHF12 was specifically enriched by SIN3B\_2 but not SIN3A. GATAD1, a PHF12 interaction partner was also specifically enriched by SIN3B\_2. SIN3A-purified samples were devoid of peptides that mapped to PHF12 (Fig. 1F). Together these data suggest that SIN3A and SIN3B\_2 interact with proteins that are homologous to yeast Rpd3L-specific components but only SIN3B\_2 interacts with proteins that are homologous to Rpd3S-specific components.

***SIN3A and SIN3B share a common HDAC Interaction Domain (HID)-*** While isoform 2 likely represents the dominant isoform of SIN3B, the National Center of Biotechnology Information human protein database (June 2016 release) contains 2 additional annotated isoforms. SIN3B isoform 1 (SIN3B\_1, NM\_015260.3, NP\_056075.1) represents the longest isoform and contains

an exon absent within isoform 2 (Fig. 2A, S2). Isoform 3 (SIN3B\_3, NM\_001297597.1, NP\_001284526.1) results from an alternative start codon and lacks the N-terminal regions found within isoforms 1 and 2 (Fig. 2A, S2). Since predicted SIN3B protein domains are either absent or altered within these rare SIN3B isoforms (Fig. 2A), we utilized these SIN3B proteoforms to characterize domain organization within SIN3B.

To assess the characteristics of each isoform, we expressed SIN3B\_1 and SIN3B\_3 fused with HaloTags in Flp-In-293 cells. Analysis of recombinant protein localization patterns revealed that SIN3B\_1 localized within the nucleus (Fig. 2B) whereas SIN3B\_3 resided within the cytoplasm (Fig. 2C). Statistical significance of protein enrichment was calculated using QSPEC 1.3.5 (Fig. 2D-E). Hierarchical clustering of protein dNSAF values for proteins enriched by at least one of the four bait purifications revealed that interaction networks of the four Sin3 forms only partially overlapped (Fig. 3A). A group of four proteins was enriched by all analyzed Sin3 proteins (Fig. 3B), consisting of HDAC1, RBBP4, and 2 isoforms of RBBP7.

Previous studies of mouse SIN3A identified a 327 residue HDAC interaction domain (HID) that is essential and sufficient for interactions with HDAC2 (24). Though a HID within SIN3B has not been experimentally defined, alignment of SIN3A and SIN3B isoforms proteins reveals that a region of SIN3B\_2 has high homology to the experimentally defined SIN3A HID (Fig. 2A, S2). The additional exon found within SIN3B\_1 resides in the region that aligns with SIN3A HID whereas the HID region in SIN3B\_3 is shorter, missing about 1/5<sup>th</sup> of its N-terminus (Fig. 2A, S2). As this region is highly homologous to the SIN3A HID, we examined what influence the additional exon in SIN3B\_1 or the shortened HID in SIN3B\_3 had on interactions with the catalytic HDAC subunits of the complex. While HDAC1/2 were enriched by SIN3B\_1 and SIN3B\_2 (Fig 3A, Table S2E), distributed spectral counts and dNSAF values of these

proteins were consistently lower following SIN3B\_1 purification (Fig. 3C, Table S2E-F). HDAC1 met criteria for enrichment following SIN3B\_3 purification; however, HDAC1 and HDAC2 were less abundant compared to SIN3B\_2-purified samples (Fig. 3C, Table S2E-F). Thus, disruption of the HID by the presence of an additional exon in SIN3B\_1 or due to shortening of SIN3B\_3 N-terminus interferes with, but does not completely inhibit, HDAC1/2 binding.

We next sought to investigate what affect the additional exon present within SIN3B\_1 had on the catalytic properties of SIN3B complexes. SIN3B\_1 and SIN3B\_2 with C-terminal HaloTags were transiently expressed in 293T cells for subsequent protein isolation and HDAC activity assays (Fig. 3D). Activity of the purified protein complexes was assessed with a fluorometric HDAC activity assay. As SIN3B\_1 protein levels were consistently lower than that of SIN3B\_2 (Fig. 3D), activity was normalized to bait protein abundance (Fig. 3E, S3, Table S3). The enzymatic activity of SIN3B\_1-purified samples was consistently lower than purified complexes containing recombinant SIN3B\_2 (Fig. 3E). Of note, the HDAC activity of both SIN3B complexes was almost completely inhibited by suberoylanilide hydroxamic acid (SAHA) (Fig. 3E). These data suggest that, like the HID region in SIN3A, this region influences the construction of catalytically active SIN3B-containing complexes.

***The Sin3 HID influences complex assembly-*** While the HID of Sin3 proteins influences interactions between HDAC1/2 and Sin3 proteins, we next sought to determine what influence alterations to this region in SIN3B isoforms had on other protein-protein interactions.

In addition to HDAC1/2, SIN3A and SIN3B both interact with RBBP4/7. Unlike interactions with HDAC1/2, SIN3B\_1 and SIN3B\_3 did not display a decreased capacity to

interact with RBBP4/7 compared to SIN3B\_2 (Fig. 4A, Table 2E-F). The enrichment of RBBP4/7 by SIN3B\_3 suggests that the C-terminal half of SIN3B is sufficient for association with these proteins. Of note, RBBP4 isoform c (RBBP4\_c, NP\_001128728.1) was consistently observed as an interaction partner of SIN3B\_3 but not of other SIN3B isoforms (Fig. 4A, Table S2E-F). cNLS mapper predicts that a nuclear localization signal with a score of 3.8 is present in RBBP4 isoform a (RBBP4\_a, NP\_005601.1) but is absent in isoform c (Fig. S4). As SIN3B\_3 was observed exclusively within the cytoplasm, this interaction may represent an interaction between cytosolic proteins.

Proteins that have homology to Rpd3S- and Rpd3L-specific components also displayed differential enrichment by the SIN3B isoforms (Fig. 4B). Rpd3L-specific component homologs SUDS3 and BRMS1L were both consistently less abundant following SIN3B\_1 and SIN3B\_3 purifications compared to SIN3B\_2 purified samples (Fig. 4B and Table S2E-F). Further, SIN3A met criteria for enrichment by SIN3B\_2 (Fig. 1E, Table S2E) but not by SIN3B\_1 or SIN3B\_3 (Table S2E), suggesting that Sin3 hetero-dimerization is also influenced by the SIN3B HID.

In addition to Rpd3L-specific component homologs, the Rpd3S-specific component homolog PHF12 and its interaction partner GATAD1 were also less abundant in SIN3B\_1-purified samples and barely detected in the SIN3B\_3 pull-downs compared to SIN3B\_2-purified samples (Fig. 4B, Table S2E-F). These data suggests that the HID region likely influences the organization of both major Sin3 interaction networks but is not required for interactions between SIN3B and RBBP4/7.

***SIN3A and SIN3B have divergent nuclear localization signals-*** While SIN3A, SIN3B\_1, and SIN3B\_2 were observed within the nucleus, SIN3B\_3 was absent from the nucleus. This

localization pattern suggests that SIN3B\_3 may lack domain(s) required for the nuclear import of this protein. Additionally, SIN3B\_1 and SIN3B\_2 enriched karyophorins KPNA2, KPNA3, KPNA4, and KPNB1, which are involved in the import of nuclear proteins (Fig. 1E-F, 4C, Table S2E). Notably, SIN3B\_1 and SIN3B\_2 contain a sequence predicted by cNLS Mapper (25) to be a bipartite nuclear localization signal (NLS) (Fig. 5A, S2). This sequence is absent within SIN3B\_3 (Fig. 5A, S2) and this isoform failed to enrich karyophorin proteins (Fig. 4A-B, Table S2E).

To test the accuracy of the predicted NLS sequence in SIN3B isoforms, basic residues within the predicted bipartite sequence in SIN3B\_2 were mutated to alanine residues. Basic residues found within SIN3A that align with the predicted SIN3B NLS were also mutated. Open reading frames encoding wild-type (Fig. 5B-E) and mutant (Fig. 5F-I) forms of SIN3A and SIN3B proteoforms were transiently expressed in 293T cells as fusions with HaloTag. Mutating residues within either segment of the predicted SIN3B\_2 NLS inhibited the nuclear localization of the recombinant protein (Fig. 5H-I), indicating that this site functions as a bipartite NLS. Surprisingly, the introduction of mutations to homologous residues in SIN3A did not inhibit the nuclear localization of SIN3A (Fig. 5F-G). The observation that KPNA2/3/4 and KPNB1 were not enriched by SIN3A purification (Fig. 1E, 4C, Table S2E) supports the conclusion we derived from the mutational analyses of nuclear localization signals: SIN3A and SIN3B\_1/2 are imported to the nucleus via distinct molecular interactions (Figure 6).

**Discussion:** To understand the therapeutic potential of targeting Sin3 complex function, we must first characterize the heterogeneous population of Sin3 HDAC complexes. Through a comparative analysis of human Sin3 protein forms, we highlight the influence of paralog

switching on complex composition and identify, as well as the shared and unique features of Sin3 protein paralogs.

***SIN3B is a component of two protein interaction networks-*** The single *S. cerevisiae* Sin3 protein is partitioned into 2 distinct protein complexes, known as Sin3S (or Rpd3S) and Sin3L (or Rpd3L) (5, 6). Common to both yeast complexes is a core of proteins consisting of Sin3, Rpd3, and Ume1, which share homology with human SIN3A/B, HDAC1/2, and RBBP4/7, respectively. While the two yeast Sin3 complexes share a common catalytic subunit, the complexes have distinct subunits that fine-tune complex function and target the complexes to different regions of genes (5, 6). There is mounting evidence that two distinct forms of the Sin3 complex exist in humans and that these complexes can be defined based on subunit homology to yeast Sin3 complex components (26–28). Human SUDS3/BRMS1/BRMS1L, ING1/2, and SAP30/SAP30L have homology with yeast Rpd3L-subunits Sds3, Pho23, and Sap30, respectively. Additionally, human PHF12 and MORF4L1 share homology with yeast Rpd3S-specific subunits Rco1 and Eaf3.

Unlike yeast, humans have two gene paralogs, *SIN3A* and *SIN3B*, that encode Sin3 proteins. To determine if the identity of the Sin3 paralog within a complex defines which interaction network is formed, we performed a comparative analysis of SIN3A and SIN3B protein-protein interactions. Our results suggest that both SIN3A and SIN3B can interact with HDAC1/2 and RBBP4/7, proteins with homology to yeast core proteins found within both Rpd3L and Rpd3S complexes. SIN3B was found to interact with SUDS3/BRMS1L and PHF12. These results show that, much like yeast Sin3, human SIN3B can interact with homologs of both Rpd3S and Rpd3L complex-specific components. Though PHF12 was initially identified as an

interaction partner of SIN3A (29, 30), SIN3A-purified samples were devoid of peptides mapping to PHF12. Thus, SIN3A enriched for homologs of Rpd3L-specific but not Rpd3S-specific components. Thus, Sin3 protein identity influences Sin3 complex composition and confirm previous reports that PHF12 interacts with SIN3B but not SIN3A (31).

An interesting finding was that SIN3B\_2 has the capacity to enrich SIN3A, which suggest hetero-oligomeric forms of human Sin3 complexes exist in humans. SUDS3, BRMS1, and BRMS1L have been proposed to induce dimerization of Sin3 complexes (32). It remains to be determined if the formation of SIN3A-SIN3B hetero-oligomers requires the presence of SUDS3/BRMS1/BRMS1L or if SIN3A and SIN3B directly interact with one another.

In addition to proteins with clear homology to yeast Sin3 complex components, several human Sin3 interaction partners exist that have no clear homology to yeast proteins. One such protein GATAD1, a known Sin3 interaction network component (26, 27), was enriched by SIN3B\_2 but not SIN3A (Fig. 1E-F). TNRC18, FAM60A, SAP130, and ARID4A/B are additional Sin3 interaction partners (33, 34) that met criteria for enrichment by SIN3A but not SIN3B\_2. As ARID4A/B have proposed DNA binding capacities (35), such divergent behavior by SIN3A and SIN3B likely has functional consequences. It is important to note that peptides mapping to these proteins were observed following SIN3B\_2 purification but that these proteins did not meet criteria for enrichment (Fig. 1F, Table S2E). Thus, interactions between SIN3B and these proteins may be weak or indirect. As SIN3A was enriched by SIN3B\_2, a possible explanation for the weak identification of these protein following SIN3B\_2 purification is that indirect interactions between SIN3B\_2 and FAM60A, TNRC18, ARID4A/B, and SAP130 could be mediated by SIN3A (Fig. 6).

# ***Protein domain organization is partially conserved between SIN3A and SIN3B-*** SIN3A

contains several experimentally defined domains. A region of mouse SIN3A that is sufficient and necessary for interactions with HDAC1/2 was previously identified and termed the HDAC interaction domain, or HID (24, 32, 36). Additionally, 3 PAH domains are present in the N-terminal half of SIN3A and there is evidence which suggests these domains mediate interaction between SIN3A and a multitude of transcription factors (4, 37). Few studies of SIN3B domain organization have been performed. However, several regions have high homology to defined domains within SIN3A. While a HID has not been experimentally defined within SIN3B, an alignment of the mouse SIN3A and SIN3B protein sequences shows that residues 388 to 651 of SIN3B\_2 have approximately 88% sequence similarity to residues 545 to 808 in SIN3A.

While isoform 2 (NP\_001284524) likely represents the principal SIN3B isoform in humans, alternative splicing produces multiple SIN3B isoforms in mouse (38, 39) and it was reported that ratios of human *SIN3B* gene products are responsive to cellular queues (38). Though alternative isoforms of SIN3B are likely a small portion of the total SIN3B within humans, these alternative proteoforms provide insight into the domain organization of SIN3B. SIN3B\_1 has an additional exon within a region that has high homology to the SIN3A HID (Fig 2A, S2). Therefore, we used this alternative SIN3B proteoform to provide insight into the dependence of SIN3B protein-protein interaction on the SIN3B HID. We demonstrate that isoform 1, which contains a 32-residue sequence not found within isoform 2, has a decreased catalytic potential as well as weaker associations with Sin3 network components (Fig. 2, Table S2). It should be noted that isoform 1 was still capable of enriching HDAC1/2 (Table S2). Thus, the addition of this exon does not completely abolish interactions with HDAC1/2 but does diminish both HDAC1/2 binding and the complex's catalytic capacity (Fig. 3C, 3E).

A region within SIN3B that has sequence homology to a portion of the SIN3A HID is critical for interactions between SIN3B and PHF12 (31). Our results support this observation as the additional exon found within SIN3B\_1 disrupts interactions between SIN3B and PHF12 (Fig. 5F). Additionally, we observed poor interactions between SIN3B\_1 and homologs of yeast Sin3L components SUDS3 and BRMS1L (Fig. 3A, 4B, Table S2). Results obtained by others suggest that the HID regions is critical for interactions between SIN3A and SUDS3 (21). Together, these results suggest that the organizing role of the HID region is conserved between SIN3A and SIN3B. As SUDS3/BRMS1 and PHF12 were less abundant in SIN3B\_1 compared to SIN3B\_2 enrichments, the HID is likely important for the organization of both forms of the Sin3 complex in humans.

Notably, SIN3A was less abundant following SIN3B\_1 enrichment compared to SIN3B\_2 enrichment (Fig. 4B, Table S2E). Thus, the HID appears to be critical for hetero-oligomerization of the Sin3 complex. Previous findings suggest that yeast Sds3 is essential for complex integrity (17) and mammalian BRMS1 (32) and SUDS3 (21, 32) are capable of forming dimers. As SUDS3 and BRMS1L were enriched by both SIN3A and SIN3B\_2, it is possible that these proteins mediate the formation of SIN3A-SIN3B\_2 hetero-oligomeric complexes.

SIN3B\_3 provides us with insight into the mechanisms responsible for SIN3B nuclear import. This isoform, resulting from an alternative start codon, is significantly shorter than SIN3B\_1/SIN3B\_2 and has absent or disrupted PAH and HID domains (Fig. 2A). Interestingly, recombinant SIN3B\_3 failed to localize to the nucleus (Fig. 2C). This isoform lacks a predicted NLS signal and prompted us to investigate the requirement for this sequence for SIN3B nuclear import. Analysis of the SIN3B\_3 interaction network revealed that, unlike other SIN3B isoforms, SIN3B\_3 did not enrich KPNA2/3/4 or KPNB1. The introduction of mutations to the

predicted NLS of SIN3B isoform 2 resulted in cytoplasmic localization; however, mutations to conserved or similar residues within SIN3A had no effect on nuclear localization. Thus, our findings indicate that SIN3A and SIN3B have distinct nuclear localization signals. Consistently, it has recently been shown that nonsense mutations at residue 949 in SIN3A results in a truncated protein with cytoplasmic localization (40), suggesting that the NLS of SIN3A resides within its C-terminus and is distinct from SIN3B NLS. Additionally, SIN3A, unlike SIN3B\_2, did not enrich KPNA2 KPNA3, KPNA4, or KPNB1 (Fig. 4C, Table S2E), further suggesting that SIN3A and SIN3B may experience unique modes of nuclear import (Fig. 6).

Not all identified proteins displayed differential interactions with SIN3B isoforms. In fact, RBBP4 and RBBP7 were consistently identified as the most abundant non-bait proteins in all SIN3B isoform purifications (Fig. 4A). As SIN3B\_1 and SIN3B\_3 enriched RBBP4/7, these results show that the C-terminal half of SIN3B is sufficient for interactions with these proteins. While the HID may act as an organizing domain for much of the complex, RBBP4/7 do not depend on this region for interactions with SIN3B. Thus, while HDAC1/2 and RBBP4/7 likely form a shared core complex with Sin3 proteins in humans, they likely do so via interaction with distinct regions of Sin3 proteins. This is surprising as HDAC1/2 and RBBP4/7 co-exist within multiple HDAC complexes and have been proposed to form a preformed sub-module (41).

In total, these results provide insight into the shared and unique properties of human Sin3 scaffolding proteins. While a conserved HID is present in both proteins, the identity of the Sin3 protein found within complexes influences their interaction networks and, likely, composition. We also show that hetero-oligomeric forms of the Sin3 complex can exist as SIN3B enriches SIN3A. These findings highlight the influence of paralog switching on protein complex composition and outline the need for future studies that further delineate the unique functions of

the distinct classes of Sin3 complexes. Future studies should take into account additional heterogeneity within population of Sin3 complexes that is introduced by other complex subunits that exist as protein paralogs.

## Acknowledgments:

Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at <http://www.stowers.org/research/publications/libpb-1445>. Research reported in this publication was supported by the Stowers Institute for Medical Research and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number F32GM122215 (M.K.A.), F31GM131536 (C.G.K), and R01GM112639 (M.P.W.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions: M.K.A., C.A.S.B., J.L.T., and M.P.W. designed research; M.K.A., C.A.S.B., J.L.T., M.K., and C.G.K. performed research; M.K.A., C.A.S.B., J.L.T., C.G.K., and M.P.W. contributed new reagents/analytic tools; M.K.A., C.A.S.B., J.L.T., M.E.S., M.K., C.G.K., L.F. and M.P.W. analyzed data; M.K.A. and M.P.W. wrote the paper.

## REFERENCES

1. Ibn-Salem, J., Muro, E. M., and Andrade-Navarro, M. A. (2017) Co-regulation of paralog genes in the three-dimensional chromatin architecture. *Nucleic Acids Research*. **45**, 81–91
2. Kondrashov, F. A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*. **279**, 5048–5057
3. Link, S., Spitzer, R. M. M., Sana, M., Torrado, M., Völker-Albert, M. C., Keilhauer, E. C., Burgold, T., Pünzeler, S., Low, J. K. K., Lindström, I., Nist, A., Regnard, C., Stiewe, T., Hendrich, B., Imhof, A., Mann, M., Mackay, J. P., Bartkuhn, M., and Hake, S. B. (2018) PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. *Nature Communications*. **9**, 4300

4. Adams, G. E., Chandru, A., and Cowley, S. M. (2018) Co-repressor, co-activator and general transcription factor: the many faces of the Sin3 histone deacetylase (HDAC) complex. *Biochemical Journal*. **475**, 3921–3932
5. Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*. **123**, 581–592
6. Carrozza, M. J., Florens, L., Swanson, S. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005) Stable incorporation of sequence specific repressors Ash1 and Ume6 into the Rpd3L complex. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. **1731**, 77–87
7. Cowley, S. M., Iritani, B. M., Mendrysa, S. M., Xu, T., Cheng, P. F., Yada, J., Liggitt, H. D., and Eisenman, R. N. (2005) The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. *Molecular and Cellular Biology*. **25**, 6990–7004
8. David, G., Grandinetti, K. B., Finnerty, P. M., Simpson, N., Chu, G. C., and DePinho, R. A. (2008) Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation. *Proceedings of the National Academy of Sciences*. **105**, 4168–4172
9. Lewis, M. J., Liu, J., Libby, E. F., Lee, M., Crawford, N. P., and Hurst, D. R. (2016) SIN3A and SIN3B differentially regulate breast cancer metastasis. *Oncotarget*. **7**, 78713–78725
10. Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J., and Ding, L. (2013) Mutational landscape and significance across 12 major cancer types. *Nature*. **502**, 333–339
11. Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Ng, P. K., Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortés-Ciriano, I., Zhou, D. C., Liang, W. W., Hess, J. M., Yellapantula, V. D., Tamborero, D., Gonzalez-Perez, A., Suphailai, C., Ko, J. Y., Khurana, E., Park, P. J., Van Allen, E. M., Liang, H., MC3 Working Group, Cancer Genome Atlas Research Network, Lawrence, M. S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A. J., Mills, G. B., Karchin, R., and Ding, L. (2018) Comprehensive characterization of cancer driver genes and mutations. *Cell*. **173**, 371–385.e18
12. Lauffer, B. E., Mintzer, R., Fong, R., Mukund, S., Tam, C., Zilberleyb, I., Flicke, B., Ritscher, A., Fedorowicz, G., Vallero, R., Ortwine, D. F., Gunzner, J., Modrusan, Z., Neumann, L., Koth, C. M., Lupardus, P. J., Kaminker, J. S., Heise, C. E., and Steiner, P. (2013) Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. *Journal of Biological Chemistry*. **288**, 26926–26943
13. Suraweera, A., O’Byrne, K. J., and Richard, D. J. (2018) Combination therapy with histone deacetylase inhibitors (HDACi) for the treatment of cancer: achieving the full therapeutic potential of HDACi. *Frontiers in Oncology*. **8**, 92
14. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P.,

- Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N. J., Nicolae, D. L., Gamazon, E. R., Im, H. K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E. T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalov, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J. M., Wilder, E. L., Derr, L. K., Green, E. D., Struwing, J. P., Temple, G., Volpi, S., Boyer, J. T., Thomson, E. J., Guyer, M. S., Ng, C., Abdallah, A., Colantuoni, D., Insel, T. R., Koester, S. E., Little, A. R., Bender, P. K., Lehner, T., Yao, Y., Compton, C. C., Vaught, J. B., Sawyer, S., Lockhart, N. C., Demchok, J., and Moore, H. F. (2013) The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*. **45**, 580–585
15. Banks, C. A. S., Thornton, J. L., Eubanks, C. G., Adams, M. K., Miah, S., Boanca, G., Liu, X., Katt, M. L., Parmely, T. J., Florens, L., and Washburn, M. P. (2018) A structured workflow for mapping human Sin3 histone deacetylase complex interactions using Halo-MudPIT affinity-purification mass spectrometry. *Mol Cell Proteomics*. **17**, 1432–1447
16. Choi, H., Kim, S., Fermin, D., Tsou, C. C., and Nesvizhskii, A. I. (2015) QPROT: Statistical method for testing differential expression using protein-level intensity data in label-free quantitative proteomics. *Journal of Proteomics*. **129**, 121–126
17. Lechner, T., Carrozza, M. J., Yu, Y., Grant, P. A., Eberharder, A., Vannier, D., Brosch, G., Stillman, D. J., Shore, D., and Workman, J. L. (2000) Sds3 (suppressor of defective silencing 3) is an integral component of the yeast Sin3·Rpd3 histone deacetylase complex and is required for histone deacetylase activity. *Journal of Biological Chemistry*. **275**, 40961–40966
18. Chen, X. F., Kuryan, B., Kitada, T., Tran, N., Li, J. Y., Kurdistan, S., Grunstein, M., Li, B., and Carey, M. (2012) The Rpd3 core complex Is a chromatin stabilization module. *Current Biology*. **22**, 56–63
19. Nourani, A., Howe, L., Pray-Grant, M. G., Workman, J. L., Grant, P. A., and Côté, J. (2003) Opposite role of yeast ING family members in p53-dependent transcriptional activation. *J. Biol. Chem*. **278**, 19171–19175
20. Zhang, Y., Sun, Z. W., Iratni, R., Erdjument-Bromage, H., Tempst, P., Hampsey, M., and Reinberg, D. (1998) SAP30, a novel protein conserved between human and yeast, is a component of a histone deacetylase complex. *Molecular Cell*. **1**, 1021–1031
21. Alland, L., David, G., Shen-Li, H., Potes, J., Muhle, R., Lee, H. C., Hou, H., Chen, K., and DePinho, R. A. (2002) Identification of mammalian Sds3 as an integral component of the Sin3/histone deacetylase corepressor complex. *Molecular and Cellular Biology*. **22**, 2743–2750
22. Graveline, R., Marcinkiewicz, K., Choi, S., Paquet, M., Wurst, W., Floss, T., and David, G. (2017) The chromatin-associated Phf12 protein maintains nucleolar integrity and prevents premature cellular senescence. *Molecular and Cellular Biology*. **37**, e00522-16
23. Xu, C., Cui, G., Botuyan, M. V., and Mer, G. (2008) Structural Basis for the Recognition of Methylated Histone H3K36 by the Eaf3 Subunit of Histone Deacetylase Complex Rpd3S. *Structure*. **16**, 1740–1750

24. Laherty, C. D., Yang, W. M., Sun, J. M., Davie, J. R., Seto, E., and Eisenman, R. N. (1997) Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell*. **89**, 349–356
25. Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proceedings of the National Academy of Sciences*. **106**, 10171–10176
26. Varier, R. A., Carrillo de Santa Pau, E., van der Groep, P., Lindeboom, R. G. H., Matarese, F., Mensinga, A., Smits, A. H., Edupuganti, R. R., Baltissen, M. P., Jansen, P. W. T. C., ter Hoeve, N., van Weely, D. R., Poser, I., van Diest, P. J., Stunnenberg, H. G., and Vermeulen, M. (2016) Recruitment of the mammalian histone-modifying EMSY complex to target genes is regulated by ZNF131. *J. Biol. Chem.* **291**, 7313–7324
27. Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., and Mann, M. (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*. **142**, 967–980
28. Nishibuchi, G., Shibata, Y., Hayakawa, T., Hayakawa, N., Ohtani, Y., Sinmyozu, K., Tagami, H., and Nakayama, J. (2014) Physical and functional interactions between the histone H3K4 demethylase KDM5A and the nucleosome remodeling and deacetylase (NuRD) complex. *Journal of Biological Chemistry*. **289**, 28956–28970
29. Yochum, G. S., and Ayer, D. E. (2002) Role for the mortality factors MORF4, MRGX, and MRG15 in transcriptional repression via associations with Pfl, mSin3A, and Transducin-Like Enhancer of Split. *Molecular and Cellular Biology*. **22**, 7868–7876
30. Yochum, G. S., and Ayer, D. E. (2001) Pfl, a novel PHD zinc finger protein that links the TLE corepressor to the mSin3A-histone deacetylase complex. *Molecular and Cellular Biology*. **21**, 4110–4118
31. Jelinic, P., Pellegrino, J., and David, G. (2011) A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Molecular and Cellular Biology*. **31**, 54–62
32. Clark, M. D., Marcum, R., Graveline, R., Chan, C. W., Xie, T., Chen, Z., Ding, Y., Zhang, Y., Mondragón, A., David, G., and Radhakrishnan, I. (2015) Structural insights into the assembly of the histone deacetylase-associated Sin3L/Rpd3L corepressor complex. *Proceedings of the National Academy of Sciences*. **112**, E3669–E3678
33. Muñoz, I. M., MacArtney, T., Sanchez-Pulido, L., Ponting, C. P., Rocha, S., and Rouse, J. (2012) Family with sequence similarity 60A (FAM60A) protein is a cell cycle-fluctuating regulator of the SIN3-HDAC1 histone deacetylase complex. *Journal of Biological Chemistry*. **287**, 32346–32353
34. Sardiú, M. E., Smith, K. T., Groppe, B. D., Gilmore, J. M., Saraf, A., Egidy, R., Peak, A., Seidel, C. W., Florens, L., Workman, J. L., and Washburn, M. P. (2014) Suberoylanilide hydroxamic acid (SAHA)-induced dynamics of a human histone deacetylase protein interaction network. *Molecular & Cellular Proteomics*. **13**, 3114–3125
35. Patsialou, A., Wilsker, D., and Moran, E. (2005) DNA-binding properties of ARID family proteins. *Nucleic Acids Research*. **33**, 66–80
36. Fleischer, T. C., Yun, U. J., and Ayer, D. E. (2003) Identification and characterization of three new components of the mSin3A corepressor complex. *Molecular and Cellular Biology*. **23**, 3456–3467

37. Le Guezennec, X., Vermeulen, M., and Stunnenberg, H. G. (2006) Molecular characterization of Sin3 PAH-domain interactor specificity and identification of PAH partners. *Nucleic Acids Research*. **34**, 3929–3937
38. Faherty, N., Benson, M., Sharma, E., Lee, A., Howarth, A., Lockstone, H., Ebner, D., and Bhattacharya, S. (2016) Negative autoregulation of BMP dependent transcription by SIN3B splicing reveals a role for RBM39. *Scientific Reports*. **6**, 28210
39. Yang, Q., Kong, Y., Rothermel, B., Garry, D. J., Bassel-Duby, R., and Williams, R. S. (2000) The winged-helix/forkhead protein myocyte nuclear factor  $\beta$  (MNF- $\beta$ ) forms a co-repressor complex with mammalian Sin3B. *Biochem J*. **345**, 335–343
40. Watanabe, K., Yamamoto, S., Sakaguti, S., Isayama, K., Oka, M., Nagano, H., and Mizukami, Y. (2018) A novel somatic mutation of SIN3A detected in breast cancer by whole-exome sequencing enhances cell proliferation through ER $\alpha$  expression. *Scientific Reports*. **8**, 16000
41. Zhang, W., Aubert, A., Gomez de Segura, J. M., Karupphasamy, M., Basu, S., Murthy, A. S., Diamante, A., Drury, T. A., Balmer, J., Cramard, J., Watson, A. A., Lando, D., Lee, S. F., Palayret, M., Kloet, S. L., Smits, A. H., Deery, M. J., Vermeulen, M., Hendrich, B., Klenerman, D., Schaffitzel, C., Berger, I., and Laue, E. D. (2016) The nucleosome remodeling and deacetylase complex NuRD is built from preformed catalytically active sub-modules. *Journal of Molecular Biology*. **428**, 2931–2942
42. Swanson, S. K., Florens, L., and Washburn, M. P. (2009) Generation and analysis of multidimensional protein identification technology datasets. in *Mass Spectrometry of Proteins and Peptides: Methods and Protocols* (Lipton, M. S., and Paša-Tolic, L. eds), pp. 1–20, Humana Press, Totowa, NJ, 10.1007/978-1-59745-493-3\_1
43. Banks, C. A. S., Kong, S. E., and Washburn, M. P. (2012) Affinity purification of protein complexes for analysis by multidimensional protein identification technology. *Protein Expression and Purification*. **86**, 105–119
44. Zhang, Y., Wen, Z., Washburn, M. P., and Florens, L. (2011) Improving proteomics mass accuracy by dynamic offline lock mass. *Anal. Chem*. **83**, 9344–9351
45. Xu, T., Park, S. K., Venable, J. D., Wohlschlegel, J. A., Diedrich, J. K., Cociorva, D., Lu, B., Liao, L., Hewel, J., Han, X., Wong, C. C. L., Fonslow, B., Delahunty, C., Gao, Y., Shah, H., and Yates, J. R. (2015) ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *Journal of Proteomics*. **129**, 16–24
46. Tabb, D. L., McDonald, W. H., and Yates, J. R. (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of Proteome Research*. **1**, 21–26
47. Zhang, Y., Wen, Z., Washburn, M. P., and Florens, L. (2010) Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem*. **82**, 2272–2281
48. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Research*. **19**, 1639–1645
49. Adams, M. K., Banks, C. A. S., Miah, S., Killer, M., and Washburn, M. P. (2019) Purification and enzymatic assay of class I histone deacetylase enzymes. in *Methods in Enzymology*, pp. 23–40, Academic Press, Cambridge, MA, 10.1016/bs.mie.2019.07.014
50. Nott, A., Watson, P. M., Robinson, J. D., Crepaldi, L., and Riccio, A. (2008) S-nitrosylation of histone deacetylase 2 induces chromatin remodelling in neurons. *Nature*. **455**, 411–415

51. Wegener, D., Wirsching, F., Riester, D., and Schwienhorst, A. (2003) A fluorogenic histone deacetylase assay well suited for high-throughput activity screening. *Chemistry & Biology*. **10**, 61–68
52. Huerta-Cepas, J., Serra, F., and Bork, P. (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*. **33**, 1635–1638
53. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. **7**, 539
54. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. **47**, W636–W641
55. Grzenda, A., Lomberg, G., Zhang, J. S., and Urrutia, R. (2009) Sin3: Master scaffold and transcriptional corepressor. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. **1789**, 443–450
56. Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardi, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., Lin, Z. Y., Badillo, B. G., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A. J. R., Choi, H., Gstaiger, M., Mohammed, S., Cristea, I. M., Bennett, K. L., Washburn, M. P., Raught, B., Ewing, R. M., Gingras, A. C., and Nesvizhskii, A. I. (2013) The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nature Methods*. **10**, 730–736
57. Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*. **44**, D279–D285

## Experimental Procedures:

*Preparation of expression vectors and expression in Flp-In™-293 cell lines-* To prepare expression constructs for SIN3A expression, the V109A point mutation present in Kazusa DNA Research Institute (KDRI) clone # FHC11647 was corrected using targeted mutagenesis. SIN3A was then excised with AsiSI and PmeI then into cloned into AsiSI and Eco53KI sites of pFC14A (Promega Corporation, Madison, WI) to add a C-terminal HaloTag®. SIN3A with an in-frame C-terminal HaloTag® was excised and cloned into MluI and NotI sites of pcDNA™5/FRT Mammalian Expression Vector (Thermo Fisher Scientific, Waltham, MA). To prepare

expression constructs for SIN3B isoform 2 (SIN3B\_2) expression, KDRI clone # FHC01991 was digested with AsiSI and PmeI, then cloned into AsiSI and Eco53KI sites of pFC14A to add a C-terminal HaloTag®. SIN3B\_2 with an in-frame C-terminal HaloTag® was excised and cloned into MluI and NotI sites of pcDNA™5/FRT Mammalian Expression Vector. To prepare SIN3B isoform 1 (SIN3B\_1) expression constructs, the MiniGene™ (Integrated DNA Technologies, Coralville, Iowa) described in Fig. S5A was digested with NsiI and BsiWI and cloned into these restriction sites of KDRI clone # FHC01991. This construct was then digested with AsiSI and BsiWI then cloned into AsiSI and BsiWI sites of the above described SIN3B\_2 in pFC14A to create SIN3B\_1 with a C-terminal HaloTag®. SIN3B\_1 with an in-frame C-terminal HaloTag® was excised and cloned into MluI and NotI sites of pcDNA™5/FRT Mammalian Expression Vector. To prepare SIN3B isoform 3 (SIN3B\_3) expression constructs, a gBlock® (Integrated DNA Technologies), described in Fig. S5B was digested with AsiSI and BsiWI then cloned into the above described SIN3B\_2-pFC14A. SIN3B\_3 with an in-frame C-terminal HaloTag® was excised and cloned into MluI and NotI sites of pcDNA™5/FRT Mammalian Expression Vector.

Stable cell lines were produced using Flp-In™-293 cells (Thermo Fisher Scientific), authenticated by STR profiling (FTA barcode: STR14169), and tested for mycoplasma using mycoplasma detection kits (American Type Culture Collection, Manassas, VA). The day before transfection, cells were plated at 50% confluency (DMEM/10% FBS) onto a 100 mm tissue culture plate and incubated at 37 °C, 5% CO2 overnight. The following day, the plate was washed two times with Opti-MEM, then incubated until ready to use with 8 mL Opti-MEM with GlutaMAX supplement (Thermo Fisher Scientific). Plasmid DNA (4 µg total; 3.6 µg pOG44 + 0.4 µg DNA of interest) was added to 800 µl of Opti-MEM with GlutaMAX supplement along

with 16  $\mu$ l FuGENE® HD Transfection Reagent (Promega Corporation), incubated for 15-30 minutes in a biosafety cabinet, then added dropwise to the prepared plate. One mL of FBS (Peak Serum, Inc, Wellington CO) was added the next morning. On day three of incubation, cells were split 1:10 and placed into selection media (DMEM/10% FBS/100  $\mu$ g/mL Hygromycin B). Media was changed every three days for a total of three media changes. After two weeks, colonies were visible and picked for screening.

*Fluorescence Microscopy*- Flp-In™-293 cell lines stably expressing HaloTag® fusion proteins were seeded at 40% confluency in 35 mm MatTek glass bottom dishes (MatTek Corporation, Ashland, MA) containing DMEM supplemented with penicillin-streptomycin solution, GlutaMAX supplement, and FBS to a final concentration of 10%. Cell media was supplemented with HaloTag® TMRDirect™ Ligand (Promega Corporation) 16-24 h after seeding to a final concentration of 20 nM and cells were incubated for an additional 16-24 h. Hoechst 33258 solution (Sigma Aldrich Corporation, St. Louis, MO) was added 80 min prior to imaging cells. Cells were washed two times with Opti-MEM media and imaged in Opti-MEM media. Images were captured on a PerkinElmer Life Sciences UltraVIEW VoX spinning disk microscope (PerkinElmer, Inc. Waltham, MA), Axiovert 200M base (Carl Zeiss AG, Oberkochen, Germany), or an inverted LSM-700 point scanning confocal microscope controlled by Zeiss Zen software (Carl Zeiss AG). A 40X plan-apochromat (NA 1.4) oil objective was used to acquire images when operating the LSM-700 microscope. Detection wavelength ranges were 300-483 nm for Hoechst and 570-800 nm for HaloTag® TMRDirect™ Ligand. A SP 490 filter set and LP 490 filter set were employed when imaging Hoechst and HaloTag® TMRDirect™ Ligand, respectively on the LSM-700 microscope.

Media conditions for transient transfection of 293T cells (American Type Culture Collection) with plasmid DNA were as stated for the imaging of the stable cell lines. Cells continued to grow 16-24 h after seeding at 40% in 35 mm MatTek glass bottom dishes before transfection. Cells were transfected with Opti-MEM media containing 2.5 µg of plasmid, 5 µL Lipofectamine LTX along with 2.5 µg PLUS Reagent (Thermo Fisher Scientific), and 20 nM final concentration HaloTag® TMRDirect™ Ligand. After 16-24 h of incubation at 37 °C and 5% CO<sub>2</sub>, Hoechst 33258 solution was added to the dishes and incubation was continued for 1 h. Cells were washed two times with Opti-MEM media before imaging in Opti-MEM media.

*Affinity purification of recombinant proteins from Flp-In™-293 cells and Multidimensional Protein Identification Technology (MudPIT) Analysis of Transiently Expressed Protein-* Cells were lysed and recombinant proteins were isolated using Magne® HaloTag® Beads (Promega) as previously described (15). Affinity purified (AP) proteins were TCA precipitated, digested with Lys-C or rLys-C, Mass Spec Grade (Promega Corporation) then Sequencing Grade Trypsin (Promega Corporation). Peptides were loaded onto triphasic MudPIT microcapillary columns as previously (42). Columns were placed in-line with an 1100 Series HPLC system (Agilent Technologies, Inc., Santa Clara, CA) coupled to a linear ion trap mass spectrometer (Thermo Fisher Scientific) and peptides were resolved using 10-step MudPIT chromatography as previously described (43).

Acquired .RAW files were converted to .ms2 files using RAWDistiller (44). ProLuCID v1.3.5 (45) was used to match spectra against a database containing human protein sequences (National Center of Biotechnology Information, June 2016 release) along with shuffled sequences for false discovery rate (FDR) estimation. The database was searched for fully tryptic

peptides with static modification of +57 daltons for cysteine and a dynamic modification of +16 daltons for methionine residues. DTASelect and Contrast (46) were used to filter results and NSAFv7 (47) was used to calculate label-free quantitative dNSAF values and generate final reports (Tables S2A-B). The spectral FDR mean  $\pm$  s.d. for the 19 MudPIT runs was 0.198%  $\pm$  0.105%, the mean peptide FDR was 0.355%  $\pm$  0.225%, and the mean  $\pm$  s.d protein FDR was 1.30%  $\pm$  0.769%. A DTASelect filter also established a minimum peptide length of 7 amino acids, and proteins that were subsets of others were removed using the parsimony option in Contrast.

Data that has been previously described was included in our analyses and is summarized in Table S1. All mass spectrometry data has been deposited into the MassIVE repository (<http://massive.ucsd.edu>). Data set identifiers are supplied in Table S1.

*Statistical analysis of proteomics data sets-* A minimum of three biological replicates were acquired for each affinity purification mass spectrometry (APMS) analysis. To identify high-confidence interaction partners, QSPEC v1.3.5 (16) was used to calculate Z-statistic, log<sub>2</sub> fold change, and FDR values of identified proteins. Prey proteins that were not present in at least half of at least one bait protein purification (Table S2C) were excluded prior to QSPEC scoring (Tables S2D). QSPEC analysis was performed with a burn in value of 2000 and 10,000 iterations. To identify enriched proteins over negative AP controls, Z-statistic values of  $\geq 3$ , log<sub>2</sub> fold change values  $\geq 2$  were selected as filter values. Only proteins with FDR values  $\leq 0.05$  were considered for analysis. Clustered heatmaps of normalized dNSAF values were generated using the python package Seaborn and the unweighted pair group method with arithmetic mean. Normalized dNSAF values were obtained by subtracting the minimum dNSAF value and

dividing by the maximum dNSAF value. The Circos plot (48) of SIN3A and SIN3B\_2 interactions was generated using average dNSAF values for enriched by SIN3A and/or SIN3B\_2.

*Enzyme activity assays*- HDAC activity assays of transiently produced proteins were performed as described (49). Briefly, approximately  $1 \times 10^7$  293T cells were plated in 150 mm dishes and cultured in 25 mL DMEM + 10% fetal bovine serum + 1x GlutaMAX Supplement. 24 h after seeding, cells were transfected with 7.5  $\mu$ g plasmid DNA, 7.5  $\mu$ L Plus Reagent, and 50  $\mu$ L Lipofectamine LTX diluted in 6.6 mL OptiMEM. Cells were harvested after an additional 48 h of culture. Two mg of whole cell extract was added to 100  $\mu$ L of washed Magne® HaloTag® Beads slurry and incubated at 4 °C for 2 h. Beads were washed 4 times with 1 mL cold TBS pH 7.4 + 0.05% Igepal CA-630 (Sigma Aldrich Corporation). Protein was eluted with 5 units AcTEV™ Protease (Thermo Fisher Scientific) in 100  $\mu$ L of 50 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 mM DTT for 16 h at 4 °C. The HDAC activity assay was based on previously published procedures (50, 51). In brief, 10  $\mu$ L of the 100  $\mu$ L of the purified protein was diluted with 32.5  $\mu$ L TBS (25 mM Tris, 150 mM NaCl, 2mM KCl, pH 7.4). Samples were supplemented with 2.5  $\mu$ L of DMSO or 200  $\mu$ M SAHA (Cayman Chemical Company, Ann Arbor, MI) resuspended in DMSO for a final concentration of 10  $\mu$ M SAHA. 5  $\mu$ L of 1 mM Boc-Lys(Ac)-AMC (APEX-BIO Technology LLC, Houston, TX) was added to each reaction to a final concentration of 100  $\mu$ M. Final reaction volume was 50  $\mu$ L. Reactions were incubated at 37 °C for 1 h. Reactions were quenched with 2.5  $\mu$ L of 200  $\mu$ M SAHA and incubated at 37 °C for 5 min. 6  $\mu$ L of 50 mg/mL trypsin was added to reaction for a final concentration of 5 mg/mL. Reactions were incubated an additional 1 h at 37 °C. Fluorescence was measured with a

SPECTRAmax GEMINI XS (Molecular Devices, San Jose, CA) using an excitation wavelength of 355, emission wavelength of 460 nm, and a cutoff wavelength of 455 nm.

*Sequence Alignment*- Graphic alignment of SIN3A (NP\_001138829.1), SIN3B\_1 (NP\_056075.1), SIN3B\_2 (NP\_001284524.1), and SIN3B\_3 (NP\_001284526.1) was generated using ETE v3 (52) and ClustalO (53). Pairwise alignments were generated using the EMBOSS-Needle algorithm (54).

## Figures:

### **Figure 1. Expression and localization of recombinant Sin3 paralogs and isoforms: (A-B)**

Subcellular localization of stably expressed (A) SIN3A (NP\_001138829.1) and (B) SIN3B\_2 (NP\_001284524.1) recombinant proteins in Flp-In™-293 cells. HaloTag® TMRDirect™ Ligand and Hoechst 33258 solution were used to visualize recombinant protein localization (red) and nuclei (blue), respectively. (C-D) Significance plots of Z-statistic vs log2 fold change (left panels) and FDR values (right panels) for the proteins detected in each APMS analysis of (C) SIN3A and (D) SIN3B\_2 (**Table S2D**). Filter values used for enriched protein identification, Z-statistics  $\geq 3$  and log2 fold change values  $\geq 2$ , are represented as dashed lines. (E) Network of proteins enriched by SIN3A (SIN3A-HaloTag) and/or SIN3B isoform 2 (SIN3B\_2-HaloTag). Recombinant SIN3A and SIN3B are source nodes and proteins enriched by SIN3A (orange), SIN3B\_2 (light blue) or both SIN3A and SIN3B\_2 (white) purifications are displayed. (F) Circos plot of average dNSAF values for proteins enriched by recombinant SIN3A (red) and/or SIN3B\_2 (blue) purifications (Table S2E). For prey proteins with multiple isoforms, dNSAF

values for all isoforms were summed and averaged. Prey species meeting enrichment criteria by SIN3A (orange), SIN3B\_2 (light blue), or both SIN3A and SIN3B\_2 (white) purifications are displayed. With the exception of HDAC1/2 and RBBP4/7, which are known interaction partners of SIN3A/B (55), proteins present in greater than 10% of 411 experiments within the CRAPome database (*Homo sapiens* version 1.1) (56) are designated by black bars.  $n=3$  for SIN3A,  $n=4$  for SIN3B\_2.

**Figure 2: Expression of rare SIN3B proteoforms.** (A) Alignment visualization of human SIN3A (NP\_001138829.1), SIN3B\_1 (NP\_056075.1), SIN3B\_2 (NP\_001284524.1), and SIN3B\_3 (NP\_001284526.1). PAH domains as defined by Pfam (57) are in orange and the experimentally defined mouse SIN3A HID (24) and homologous sequences in SIN3B isoforms are in dark blue. Detailed sequence alignments are provided in **Figure S2**. (B-C) Subcellular localization of stably expressed (B) SIN3B\_1 and (C) SIN3B\_3 recombinant proteins in Flp-In-293 cells. HaloTag® TMRDirect™ Ligand and Hoechst 33258 solution were used to visualize recombinant protein localization (red) and nuclei (blue), respectively. (D-E) Significance plots of Z-statistic vs log2 fold change (left panels) and FDR values (right panels) for the proteins detected in each APMS analysis of Sin3 paralogs and isoforms (**Table S2D**). Filter values used to identify enriched proteins are designated by dotted lines.

**Figure 3. The SIN3A HID is conserved in SIN3B.** (A) Clustered heatmap of normalized dNSAF values for proteins within each bait replicate that were enriched by at least one Sin3 protein bait. Proteins were isolated and analyzed from three (SIN3A) or four (all SIN3B isoforms) independent cell clones stably expressing the recombinant Sin3 protein of interest

(**Table S2A-E**). Proteins for which multiple isoforms are represented are denoted by isoform identifiers after the protein name. (B) Venn diagram of the number of prey proteins identified as interaction partners of human SIN3A (purple), SIN3B\_1 (blue), SIN3B\_2 (green), and SIN3B\_3 (yellow). (C) Average dNSAF values measured for HDAC1 and HDAC2 in SIN3A- (white), SIN3B\_1 (light grey), SIN3B\_2 (dark grey), and SIN3B\_3 (black) affinity-purified samples. *Mean ± S.D.*,  $n=3$  for SIN3A, and  $n=4$  biological replicates for all SIN3B isoforms (**Table S2B**). (D-E) Halo-control (Control), SIN3B\_1-HaloTag (SIN3B\_1), SIN3B\_2-HaloTag (SIN3B\_2) were transiently expressed in 293T cells and purified using Magne® HaloTag® Beads. Purified protein was eluted in 100  $\mu$ L of elution buffer. (D) 10  $\mu$ L of whole cell extract from each transfection replicate was loaded onto 4-15% polyacrylamide gels and blots were probed using a 1:3000 dilution of anti-mSIN3B sc-13145 X antibody (Santa Cruz Biotechnology, Dallas, TX) and a 1:10,000 dilution of IRDye® 800CW secondary antibody (LI-COR, Lincoln, NE). (E) HDAC activity assay of protein complexes purified using SIN3B isoform 1 (SIN3B\_1\_Halo) and SIN3B isoform 2 (SIN3B\_2\_Halo) transiently expressed as baits within 293T cells. Reactions were supplemented with DMSO (grey) or DMSO + SAHA (black). RFU (relative fluorescence unit) values for all biological replicates were normalized to recombinant SIN3B protein abundance in purified samples as measured by Western blot (**Fig. 3D, Table S3**). *Mean ± S.D.*,  $n=3$ . \*:  $p \leq 0.005$ , \*\*:  $p \leq 0.001$ .

**Figure 4. Analysis of interaction partners of SIN3B isoforms.** (A) dNSAF values of RBBP4 and RBBP7 proteins identified as being enriched by at least one SIN3B isoform. (B) dNSAF values of SUDS3 and SIN3A (left panel) as well as BRMS1L, GATAD1, and PHF12 (right panel) following SIN3B\_1- (light grey), SIN3B\_2- (dark grey), and SIN3B\_3- (black)

purifications (Table S2E). (A-B) *Mean ± S.D.*, *n*=4. (C) log2 fold change (left panel) and Z-statistic (right panel) values calculated by QSPEC v1.3.5 for KPNA2, KPNA3, KPNA4, and KPNB1 in SIN3A- (white), SIN3B isoform 1- (light grey), SIN3B isoform 2- (dark grey), and SIN3B isoform 3- (black) purified samples (**Table S2D**).

**Figure 5. Identification of the SIN3B nuclear localization signal.** (A) Alignment of the predicted bipartite nuclear localization sequences in SIN3A and SIN3B. Motifs predicted by cNLS Mapper (25) and residues mutated to test predicted NLS accuracy are highlighted in red. (B-I) Subcellular localization of transiently expressed wild-type (B-E) and mutant (F-I) forms of Halo-tagged recombinant Sin3 proteins in 293T cells. (F-I) Protein names are appended with the identities of mutated residues. HaloTag® TMRDirect™ Ligand and Hoechst 33258 solution to visualize recombinant protein localization (red) and nuclei (blue), respectively.

**Figure 6. Model of the human SIN3A and SIN3B interactions defined in this study.** SIN3A and SIN3B have divergent nuclear localization signals that mediate their nuclear import. SIN3B\_3 is not found within the nucleus and has weak interactions with HDAC1/2 and RBBP4/7. SIN3A and SIN3B\_2 are capable of forming hetero-oligomeric complex.

**Supplementary Figure S1 Related to Figure 1.** Pairwise sequence alignment of SIN3A, (NP\_001138829.1) and SIN3B isoform 2 (NP\_001284524.1) generated using EMBOSS-Needle (54).

**Supplementary Figure S2 Related to Figure 2.** Protein alignment of SIN3A and SIN3B isoforms generated with ClustalO (53). A sequence within SIN3B\_1 and SIN3B\_2 that possesses

a cNLS Mapper score of 12.5 as a bipartite NLS is highlighted in yellow. A sequence within SIN3A possesses a cNLS Mapper score of 4.0 as a bipartite NLS is also highlighted in yellow. Residues within human SIN3A that are homologous to the mouse SIN3A HID (24) are in red.

**Supplementary Figure S3 Related to Figure 3D-E.** Image of Western blot shown in **Fig. 3D** with intensity values displayed.

**Supplementary Figure S4 Related to Figure 4A.** Pairwise sequence alignment of RBBP4 isoform a, (NP\_005601.1) and RBBP4 isoform c (NP\_001128728.1) generated using EMBOSS-Needle (54). A sequence within RBBP4 isoform a that possesses a cNLS Mapper score of 3.8 as a bipartite NLS is highlighted in yellow.

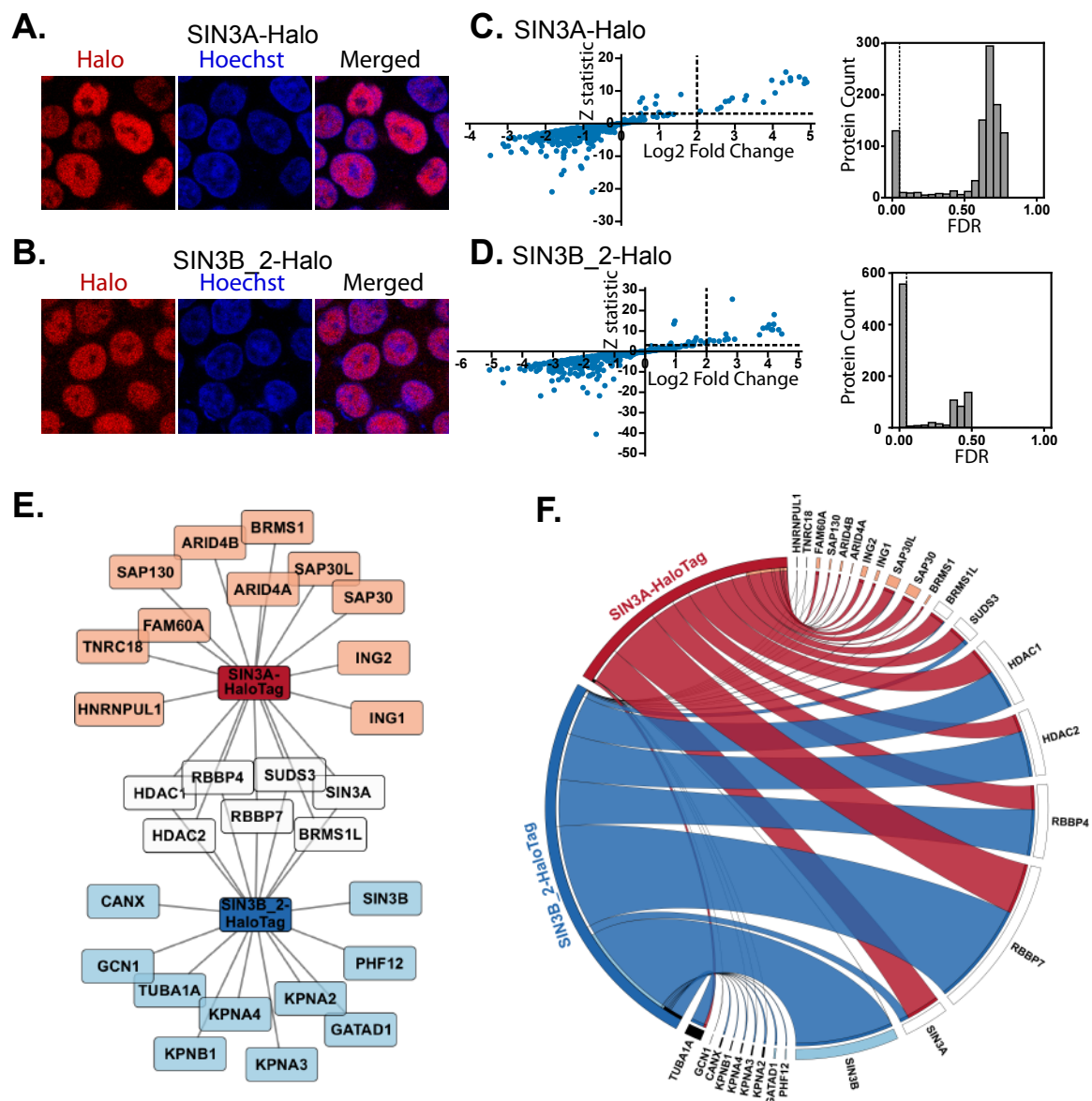
**Supplementary Figure S5 Related to Figures 2, 3, 4, 5.** (A) The MiniGene™ (Integrated DNA Technologies) sequence used to create SIN3B isoform 1. NsiI and BsiWI recognition sequences are underlined. (B) The gBlock® (Integrated DNA Technologies) sequence used to create SIN3B isoform 3. AsiSI and BsiWI recognition sequences are underlined.

**Supplementary Table S1 Related to Figures 1, 2, 3, 4, 5.** Data Availability

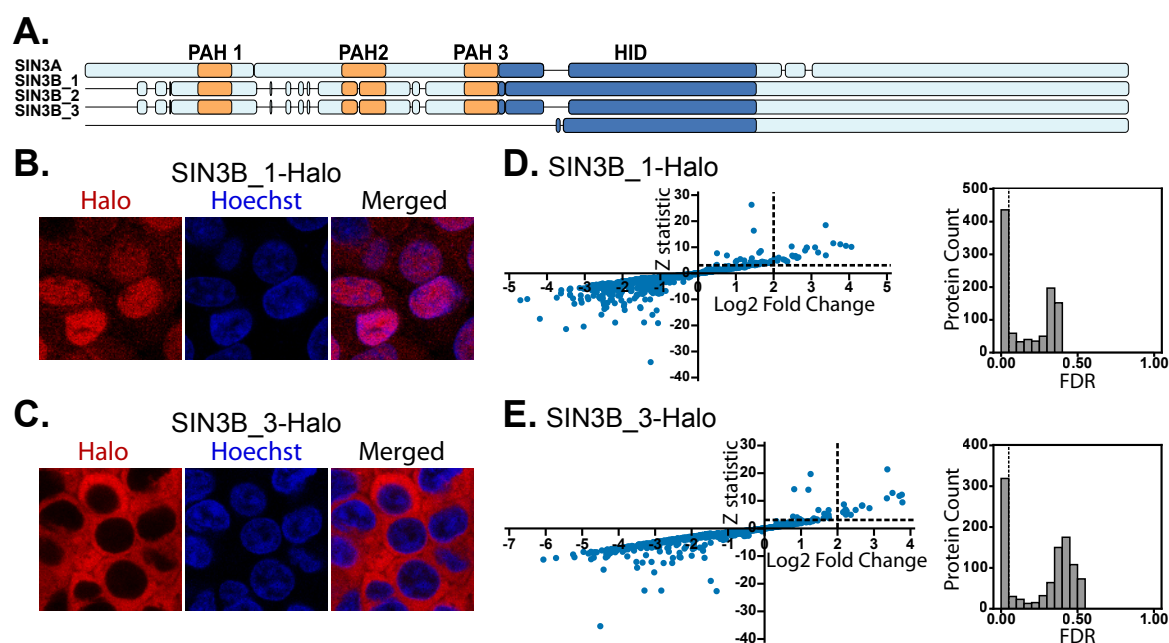
**Supplementary Table S2 Related to Figures 1, 2, 3, and 4.** Identification, label-free quantitation, and statistical analysis of proteins detected in APMS analyses of SIN3 paralogs and isoforms

**Supplementary Table S3 Related to Figures 3.** HDAC activity assay.

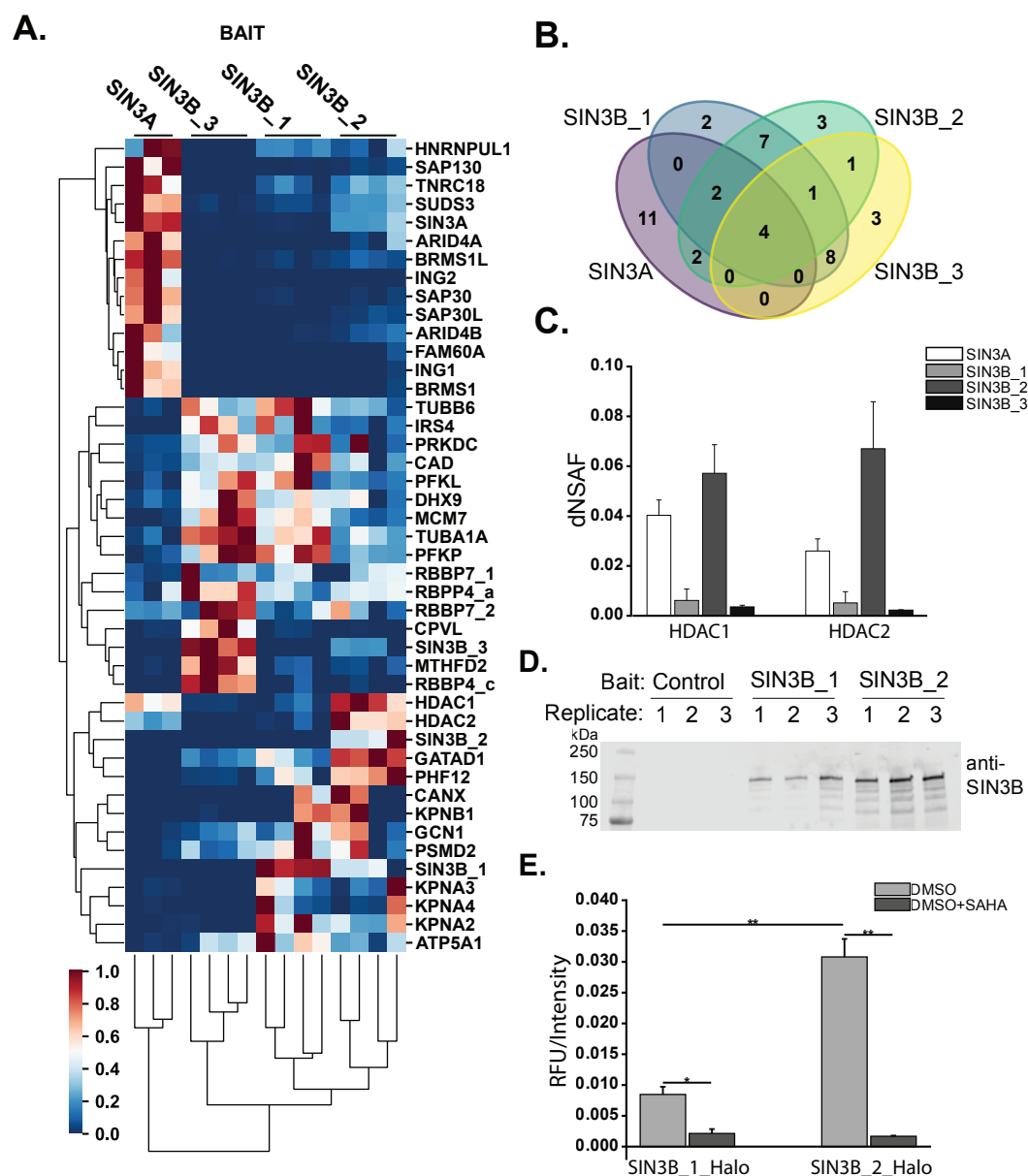
## Figure 1 | Adams et al.



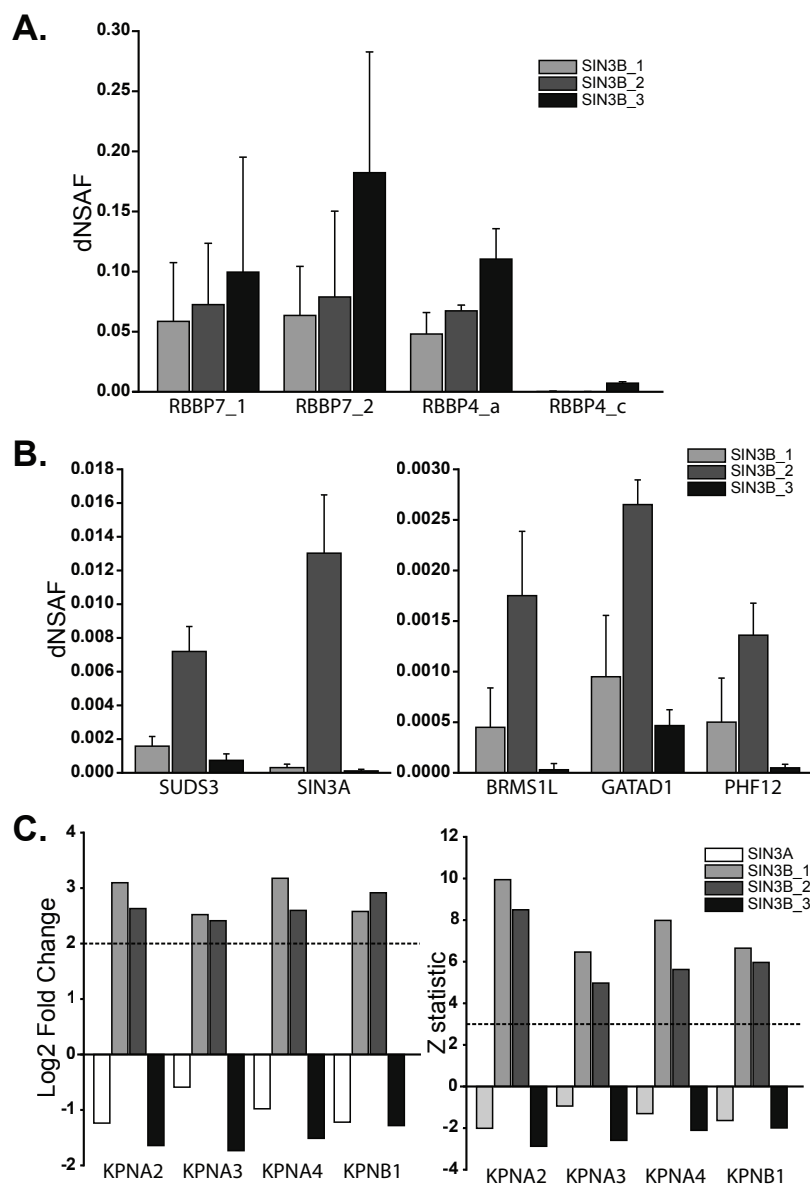
## Figure 2 | Adams et al.



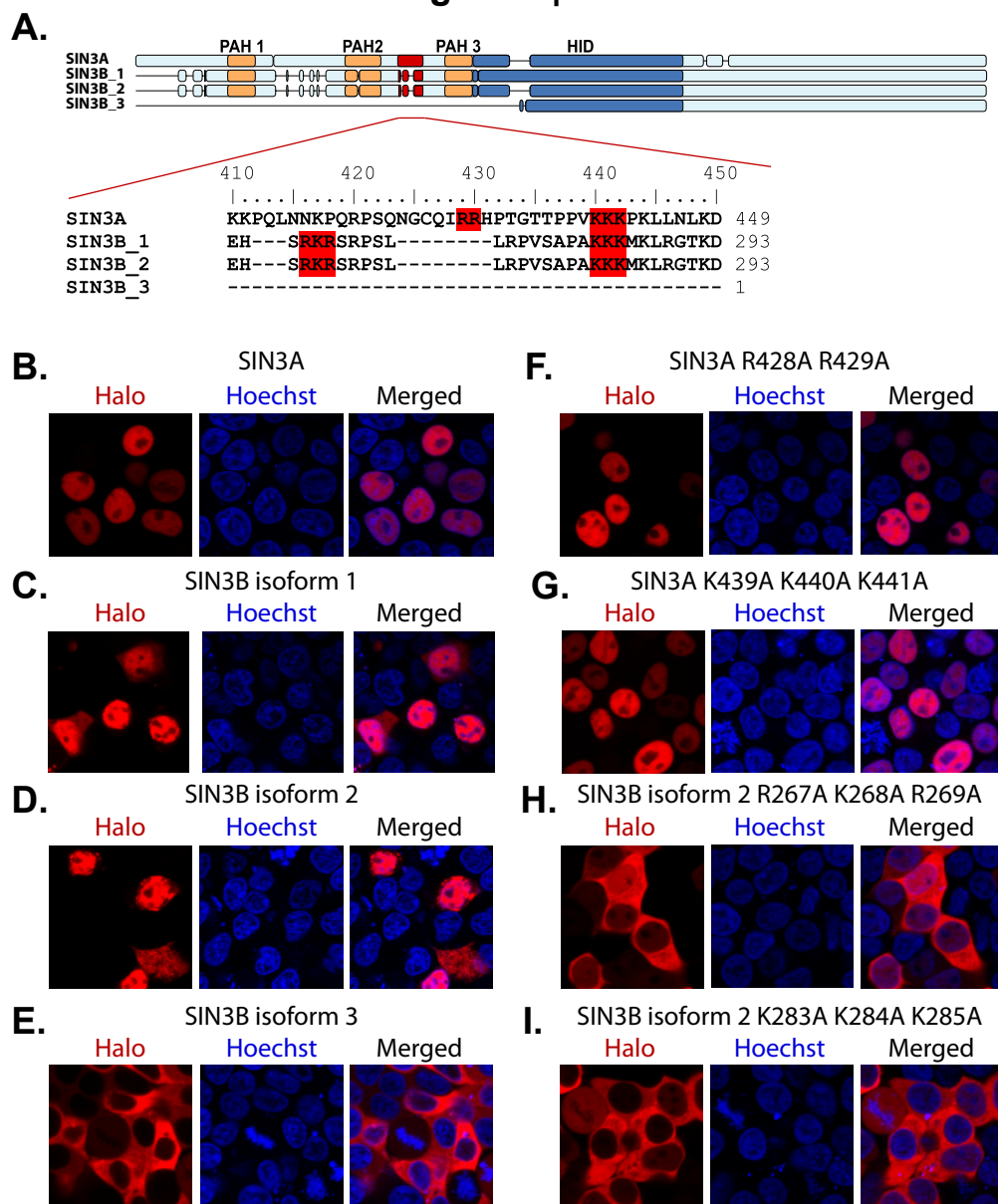
# Figure 3 | Adams et al.



**Figure 4 | Adams et al.**



## Figure 5 | Adams et al.



**Figure 6 | Adams et al.**

