

TITLE

Structural basis of the interaction between SETD2 methyltransferase and hnRNP L paralogs for governing co-transcriptional splicing

AUTHORS

Saikat Bhattacharya^{1*}, Suman Wang^{2,3*}, Divya Reddy¹, Siyuan Shen^{2,3}, Ying Zhang¹, Ning Zhang¹, Hua Li¹, Michael P. Washburn⁴, Laurence Florens¹, Yunyu Shi^{2,3}, Fudong Li^{2,3#} and Jerry L. Workman^{1#}

¹ Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA

² Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230026, China

³ Ministry of Education Key Laboratory for Membraneless Organelles and Cellular Dynamics, University of Science and Technology of China, Hefei, China

⁴ Department of Cancer Biology, University of Kansas Medical Center, Kansas City, KS, 66160, USA

#email: jlw@stowers.org, lifudong@ustc.edu.cn

*These authors contributed equally to this work

ABSTRACT

The RNA recognition motif (RRM) binds to nucleic acids as well as proteins. More than one such domain is found in the pre-mRNA processing hnRNP proteins. While the mode of RNA recognition by RRMs is known, the molecular basis of their protein interaction remains obscure. Here we describe the mode of interaction between hnRNP L and LL with the methyltransferase SETD2. We demonstrate that for the interaction to occur, a leucine pair within a highly conserved stretch of SETD2 insert their side chains in hydrophobic pockets formed by hnRNP L RRM2. Notably, the structure also highlights that RRM2 can form a ternary complex with SETD2 and RNA. Remarkably, mutating the leucine pair in SETD2 also results in its reduced interaction with other hnRNPs. Importantly, the similarity that the mode of SETD2-hnRNP L interaction shares with other related protein-protein interactions reveals a conserved design by which splicing regulators interact with one another.

INTRODUCTION

Alternate splicing (AS) is a deviation from the more prevalent process of splicing in which certain exons are skipped resulting in various forms of mature mRNA¹. AS is a vital process that enables cells to synthesize multiple protein isoforms from the same gene. 95% of human genes are estimated to undergo AS and it gives rise to the protein diversity needed for the varied cell types and functions from a limited set of genes^{2,3}. AS functions in critical biological processes including cell growth, cell death, cell differentiation pluripotency, and development^{4,5}. Defects in AS cause neurodegenerative diseases and cancer⁶⁻¹⁰.

The exons that end up in the mature mRNA during the process of AS are defined by the interaction between cis-acting elements and trans-acting factors. Cis-acting elements include exonic- and intronic-splicing enhancers (ESEs/ISEs) that are bound by positive trans-acting factors, such as SR (serine/arginine-rich) proteins, and exonic- and intronic-splicing silencers (ESSs/ISSs) are bound by negative trans-acting factors, such as the heterogeneous nuclear ribonucleoproteins (hnRNPs)^{1,11,12}. The collaboration between these elements results in the promotion or inhibition of spliceosome assembly on weak splice sites. HnRNPs are highly conserved from nematodes to mammals and have several critical roles in mRNA maturation^{13,14}. Their function during AS is to bind to the ESS and exclude the SR proteins.

Studies over the years have largely focused on the nucleic acid binding aspect of hnRNPs to understand their function and the mechanism of AS. However, analysis of the RNA binding motif of hnRNPs has revealed that their binding regions are widespread in mRNAs. For example, studies aimed to find the RNA binding motif of hnRNP L revealed that it binds to CA-rich regions¹⁵⁻¹⁹. Such sequences, however, occur in the human genome at a frequency of 19.4 CA repeats per megabase²⁰, representing the most common simple sequence repeat motif. Combined with the fact that hnRNPs are very abundant and ubiquitous proteins, this makes it unclear how the hnRNPs engage their target transcripts specifically²¹. Also, while a context-

dependent regulation of splicing by hnRNP L has been noted, it is unknown what factors determine this process²².

One possibility may be that hnRNPs rely on their interacting protein partners to engage their specific target pre-mRNA in a context-dependent and cell line-specific manner. The evidence to back this possibility comes from the study of the interactome of hnRNP L. We and others have shown that it specifically interacts with the methyltransferase SETD2 and the mediator complex component Med23^{23,24}. Med23 brings hnRNP L to the promoter of target genes from where it might be handed over to SETD2 to co-regulate a common subset of AS events. Notably, SETD2 binds to RNA Pol II during transcription elongation. Therefore, the hnRNPs interactome not only regulates its function but also couples transcription and alternative splicing, which permits the sequential recognition of emerging splicing signals by the splicing machinery.

The RNA Recognition Motifs (RRMs), specifically the RRM2 of hnRNP L, mediate its interaction with other proteins. We recently showed that the RRM2 of hnRNP L binds to a novel SETD2-hnRNP Interaction (SHI) domain in SETD2²³. Consistent with the focus on the RNA binding aspect of hnRNPs, numerous crystal structures of RRM-RNA complexes are available²⁵⁻²⁷. However, the molecular basis for the specific binding of protein interactors by hnRNPs remains elusive. Furthermore, it is not clear why the RRM2 is specifically able to interact with SETD2 but the other RRM2s of hnRNP L cannot.

In this work, we report the crystal structure of the hnRNP L-RRM2 in complex with the SETD2-SHI domain at 1.80 Å resolution and show that a leucine pair in the disordered region of SETD2 form hydrophobic interactions with RRM2 and is crucial for binding. Notably, the SETD2 binding region in RRM2 is distinct from its RNA-binding interface. Furthermore, we demonstrate that the hnRNP L paralog, hnRNP LL, also interacts with SETD2. SETD2 co-purifies numerous RNA binding proteins besides hnRNP L/LL. Strikingly, mutating the two conserved leucines in

the SETD2 SHI domain results in loss of its interaction with most RNA-binding proteins.

Moreover, our findings reveal that the mode of SETD2-hnRNP L interaction shares similarity with RAVR1-PTB interaction, pointing towards a possible common design behind protein binding by RRM.

RESULTS

A conserved region within the SETD2 SHI domain mediates hnRNP L binding

Previously, we performed a detailed characterization of SETD2-hnRNP L interaction and demonstrated that the two proteins co-regulate the AS of a subset of genes²³. This characterization revealed a novel SHI domain in SETD2 that engages the hnRNPs. However, the underlying mechanism of the interaction between them has not yet been characterized. The SHI domain is located in a predicted disordered region in SETD2 and is expected to assume a random coil structure based on *ab initio* structure modeling²³. We performed sequence analysis of the SHI domain to look for functional regions using ConSurf²⁸. Multiple sequence alignment followed by HMMER homolog search algorithm revealed that the residues 2167-2192 within the 50 residue SHI domain are highly conserved across different organisms [Figure 1a]. Notably, all these species also code for hnRNP L. We wondered whether this conserved stretch is responsible for hnRNP L binding.

To gain insights into this we assayed the binding affinities of SETD2 fragments to hnRNP L by performing isothermal titration calorimetry (ITC). The results showed that SETD2²¹⁶⁷⁻²¹⁹² binds to hnRNP L RRM2 with a dissociation constant (KD) of $2.02 \pm 0.21 \mu\text{M}$ and an N value of ~ 1 (1.02) [Figure 1b, Supplementary Table 1]. Also, we could not detect any interaction between an adjacent fragment SETD2²¹¹³⁻²¹⁴⁰ and hnRNP L RRM2 as expected [Figure 1c, Supplementary Table 1].

Next, we created deletion mutants of SETD2C (1404-2564) in which 10 amino acid bins were deleted in the 50 amino acid SHI domain [Figure 1d]. Previously we have shown that full-length SETD2 protein is robustly degraded²⁹ and hence, the C-terminal region was used for performing purifications. Subsequently, Halo-tagged WT SETD2C and its mutants were affinity-purified using Halo ligand-conjugated magnetic resin from 293T extracts. The purified complexes were analyzed by silver staining and western blotting [Figure 1e, f]. Consistent with our previous report, hnRNP L was co-purified with SETD2C and the interaction was lost upon deletion of the SHI domain (SETD2C Δ SHI) [Figure 1e, f]. Notably, the SETD2C Δ SHI 4 and 5 did not affect co-purification of hnRNP L whereas any one of C Δ SHI 1-3 (spanning 2164-2193) completely abolished interaction with hnRNP L demonstrating that the stretch 2167-2192 is indeed responsible for hnRNP L binding [Figure 1e, f]. RNA Pol II which binds to the distinct SRI (Set2-Rpb1 Interaction) domain of SETD2 was co-purified with both WT SETD2C and its SHI domain mutants as expected [Figure 1e].

To conclude, we identified a conserved 30 amino acid long stretch (2167-2192) in SETD2 that mediates interaction with hnRNP L.

The crystal structure reveals the details of hnRNP L-SETD2 interaction

The conserved 30 amino acid sequence of the SHI domain did not reveal any clues as to how it might engage hnRNP L. Hence, to further explore the molecular basis of this interaction, we determined the crystal structure of hnRNP L RRM2 in complex with the SETD2²¹⁶⁷⁻²¹⁹² peptide at 1.80 Å resolution. There are two complexes contained in a crystallographic asymmetric unit, with hnRNP L RRM2 chains A and C associating with SETD2²¹⁶⁷⁻²¹⁹² chains B and D, respectively [Supplementary Figure 1a, Supplementary Table 2]. The two complex pairs are highly similar to each other, with root-mean-square deviation (RMSD) of 0.357 Å of overall 101 aligned C α atoms. The hnRNP L RRM2 in the complex adopts a β 1 α 1 β 2 β 3 α 2 β 4 β 5 conformation that forms a five-stranded β sheet packed against two α -helices. Further, three-

dimensional structure superimposition (RMSD = 0.672 Å over 92 aligned C α atoms) reveals no significant structural changes in RRM2 upon SETD2²¹⁶⁷⁻²¹⁹² peptide binding [Supplementary Figure 1b].

When bound to hnRNP L RRM2, SETD2²¹⁶⁷⁻²¹⁹² adopts a U-shaped conformation contacting the RRM dorsal helical face [Figure 2a]. Although all residues of the SETD2²¹⁶⁷⁻²¹⁹² peptide could be clearly traced, only a fraction at the C-terminus of this peptide (²¹⁸³NAGKVLLPTP²¹⁹²) was found to directly contact the RRM, burying ~477 Å² of the solvent-exposed area of hnRNP L RRM2. In detail, the bulk side chain of SETD2^{Leu2188} is accommodated by a hydrophobic pocket which is lined by hnRNP L^{Leu251} from α 2, hnRNP L^{Ile214} from α 1, hnRNP L^{Ile256} from the α 2- β 4 loop, and hnRNP L^{Leu263} from β 4 [Figure 2b-d]. In addition, the side chain of SETD2^{Leu2189} stretches into a neighboring shallow apolar depression lined by hnRNP L Ile214, Val210, Ile256, and Tyr257 [Figure 2b-d]. Furthermore, the residues downstream of the Leu pair, including 2190PTP2192, together with SETD2^{Leu2189}, wrap around the hnRNP L^{Tyr257} and hnRNP L^{Tyr204}, providing further binding affinity and specificity [Figure 2e]. Notably, hnRNP L^{Tyr257} makes van der Waals contacts with both SETD2^{Leu2189} and SETD2^{Pro2192}, playing an important role in hnRNP L-SETD2 interaction. Besides apolar contacts, there are also substantial hydrogen bonding interactions between SETD2 peptide and hnRNP L RRM2. The amide nitrogen of SETD2^{Leu2188} and SETD2^{Leu2189} form hydrogen bonds with the carbonyl oxygen of hnRNP L^{Asp255} and hnRNP L^{Ile256}, respectively [Figure 2f]. In addition, the carbonyl oxygen and N ζ atom of SETD2^{Lys2186} make hydrogen bonding interactions with hnRNP L^{Asp255} and hnRNP L^{Ser250}, respectively [Figure 2f]. Furthermore, the side chain of SETD2^{Asn2183} forms bifurcated hydrogen bonds with the main chain of hnRNP L^{Ala249} and hnRNP L^{Asn252} [Figure 2f]. Notably, the structure also revealed that hnRNP L RRM2 can form a ternary complex with SETD2 and RNA by binding them simultaneously as their binding sites are non-overlapping [Supplementary Figure 1c].

hnRNP LL and SETD2 interact *in vivo*

Previously, we showed that ingenuity pathway analysis (IPA) of proteins co-purified with SETD2 revealed an enrichment of RNA processing proteins, including numerous hnRNPs²³. HnRNP LL has a very similar amino acid sequence with 68% sequence identity and domain organization to its paralog hnRNP L [Figure 3a]. Despite being paralogous, hnRNP L and LL have different RNA-binding constraints and have been shown to have a non-redundant role in regulating AS^{30,31}. Notably, unlike hnRNP L, hnRNP LL did not interact with Med23 *in vitro* and did not interact much in coimmunoprecipitation experiment²⁴. We wanted to test whether SETD2 and hnRNP LL also interact *in vivo*.

Multi-dimensional protein identification technology (MudPIT) mass spectrometry analysis of purified complexes with SETD2C indeed revealed hnRNP LL as an interactor [Figure 3b]. Moreover, MudPIT analysis of purified SETD2C complexes from 293T cell extracts depleted of hnRNP L revealed enrichment of hnRNP LL suggesting that it binds to SETD2 independent of hnRNP L [Figure 3c]. In line with our previous finding that SETD2-hnRNP L interaction occurs irrespective of SETD2-Pol II interaction, mass spectrometry analysis revealed that SETD2-hnRNP LL interaction persisted even upon the deletion of the SRI domain from SETD2C [Figure 3b, d]. To confirm the mass spectrometry results, western blotting with an antibody specific for hnRNP LL was performed with affinity-purified complexes using Halo-SETD2CΔSRI as bait from 293T extracts with and without RNase treatment. In this technique, elution of the proteins purified involves cleaving off the Halo-tag with TEV protease, thus, resulting in a difference in molecular weight of bait between input and eluted samples. The results confirmed that SETD2 interaction with hnRNP LL persisted even without the Pol II interaction domain and upon RNase treatment [Figure 3e]. Furthermore, mass spectrometry analysis of SETD2CΔSHI mutant revealed loss of interaction with hnRNP LL strongly suggesting that it binds to the same region in SETD2 as hnRNP L [Figure 3b, d]. To validate the mass spectrometry results, SETD2

deletion mutant SETD2CΔ2164-2213, was affinity-purified from 293T cells and analyzed by immunoblotting. As anticipated, immunoblotting for RNA Pol II and anti-hnRNP LL revealed that upon deletion of the stretch 2164-2213 from SETD2, the SETD2-hnRNP LL interaction was abolished without affecting the SETD2-Pol II interaction [Figure 3f]. Further, the mass spectrometry analysis of purification of SETD2 2164-2213 revealed hnRNP LL as an interactor demonstrating that hnRNP LL binds to the SETD2 SHI domain [Figure 3g].

Mass spectrometry analysis of yeast SETD2 homolog, Set2 (ySet2) purified from 293T cells revealed that it can interact with Pol II even in human cells and this interaction was lost upon the deletion of the SRI domain as expected [Figure 3h, i]. However, an interaction between ySet2 and hnRNP LL was not observed consistent with the fact that although ySet2 and SETD2 share the conserved AWS, SET, Post-SET, WW, and SRI domains, ySet2 lacks the SHI domain [Figure 3h]. Remarkably, the addition of the 2164-2213 stretch of SETD2 to ySet2 (ySet2 2164-2213), and ySet2ΔSRI resulted in gain of interaction with hnRNP LL [Figure 3i]. These mass spectrometry findings were confirmed by immunoblotting the Set2 purified complexes with an anti-hnRNP LL antibody [Figure 3j].

To conclude, hnRNP LL binds to the SHI domain of SETD2 *in vivo*.

hnRNP LL RRM2 domain directly interacts with SETD2

Based on sequence homology between hnRNP L and LL, RRM2 of LL is expected to interact with SETD2 [Figure 4a]. To test this, Halo-SETD2C and mCherry-HA-hnRNP L/LL RRM2 constructs were co-expressed in 293T cells and protein complexes were purified using Halo affinity-purification. Immunoblotting of the purified complexes with anti-SETD2 and anti-HA antibodies demonstrated that similar to hnRNP L, the RRM2 of hnRNP LL interacts with SETD2 [Figure 4b]. These findings were confirmed by reciprocal co-purification in which Halo-HA-hnRNP LL RRM2 successfully co-purified GFP-FLAG-SETD2C [Figure 4c].

To confirm the specificity of SETD2-hnRNP LL RRM2 interaction, hnRNP LL FL (1-542) and 272-542 were tagged with mCherry-HA [Figure 4d]. Next, Halo-SETD2C and mCherry-HA-hnRNP LL constructs were co-expressed in 293T cells and protein complexes were purified using Halo affinity-purification. Immunoblotting with anti-HA antibody demonstrated that hnRNP LL 1-542 interacts with SETD2 as expected, whereas hnRNP LL 272-542 does not [Figure 4e]. Immunoblotting with an anti-Pol II antibody confirmed that Pol II was co-purified [Figure 4e]. These results demonstrate that the RRM2 of hnRNP LL interacts with SETD2 *in vivo*.

Next, to confirm direct physical interaction, we performed ITC of hnRNP LL RRM2 with SETD2. The ITC result showed that hnRNP LL binds to SETD2²¹⁶⁷⁻²¹⁹² with a K_D value of $5.75 \pm 0.896 \mu\text{M}$, which is slightly weaker than that of hnRNP L (higher K_D value suggests lower binding affinity) [Figure 4f, Supplementary Table 1]. In the sequence alignment of RRM2 of hnRNP L vs LL, we noted that the residues belonging to the SETD2 binding surface are highly conserved, except that I214 of hnRNP L corresponds to a Val in hnRNP LL [Figure 4a]. Next, we generated a hnRNP LL^{V188I} mutant and performed ITC and found that it shows nearly the same binding affinity with SETD2²¹⁶⁷⁻²¹⁹² as that of hnRNP L [Supplementary Figure 2a, Supplementary Table 1].

We next tried to crystallize the hnRNP LL RRM2-SETD2²¹⁶⁷⁻²¹⁹² complex but did not succeed. However, we got a high-resolution structure when a truncated peptide encompassing the core interacting motif of SETD2²¹⁸⁰SNPNAGKVLLPTP²¹⁹² was used for crystallization. We determined the structure in the P1 space group at a 1.60 Å resolution, with two almost similar hnRNP LL RRM2-SETD2²¹⁸¹⁻²¹⁹² complexes in each crystallographic asymmetric unit. HnRNP LL-SETD2²¹⁸¹⁻²¹⁹² structure shows high similarity to the structure of hnRNP L- SETD2²¹⁶⁷⁻²¹⁹² complex (RMSD = 0.672 Å over 92 aligned Ca atoms) [Figure 4g, Supplementary Table 2].

Mutating key residues in hnRNP L and SETD2 abolishes their interaction

The crystal structures revealed the important residues in RRM2 that mediate interaction with SETD2. To confirm that those residues are indeed important for binding SETD2, ITC was performed using a series of hnRNP L mutants. Mutating the residues I256 and Y257 of hnRNP L had the biggest impact on reducing SETD2 binding [Figure 5a, Supplementary Table 1]. The I256A mutation resulted in a nine-fold decrease in SETD2 binding whereas the Y257A mutation completely abolished SETD2 binding [Figure 5b, Supplementary Table 1].

To test whether these mutations also resulted in the loss of hnRNP L-SETD2 binding *in vivo*, Halo-SETD2C and mCherry-HA-hnRNP L constructs were co-expressed in 293T cells and protein complexes were purified using Halo affinity-purification. Immunoblotting with anti-SETD2 and anti-hnRNP L revealed successful purification of endogenous hnRNP L with SETD2 as expected [Figure 5c]. Also, immunoblotting with an anti-HA antibody confirmed the co-purification of ectopically expressed WT hnRNP L FL and the lack of SETD2 binding to hnRNP L 322-589 as we have previously shown²³ [Figure 5c]. Importantly, the hnRNP L FL mutant in which both I256 and Y257 were mutated to alanine did not interact with SETD2, confirming our ITC data [Figure 5c].

Only the RRM2 out of the four RRMs of hnRNP L binds SETD2. To understand this, we analyzed the sequences of hnRNP L RRMs. The residues V210, I214, L251, I256, Y257, L263 of hnRNP L RRM2 form two hydrophobic pockets [Figure 2c]. Sequence alignment shows that the other three RRMs do not contain these six conserved hydrophobic residues, therefore, SETD2 cannot recognize these three RRMs [Supplementary Figure 2b].

The crystal structure suggests that the leucine pair in the SETD2 SHI domain is important for binding hnRNP L/LL. Our ITC confirmed this observation as the interaction of SETD2 L2188A and L2189A peptides with hnRNP L RRM2 was completely abolished [Figure 5d]. To test whether these mutations also result in loss of hnRNP L-SETD2 binding *in vivo*,

Halo-SETD2C constructs were expressed in 293T cells and protein complexes were purified using Halo affinity-purification followed by silver staining and western blotting [Figure 5e, f]. Immunoblotting with anti-SETD2 and anti-hnRNP L revealed that the endogenous hnRNP L co-purified with SETD2C but not with SETDCASHI as expected. Importantly, the SETD2 mutant in which both L2188 and L2189 were mutated to alanine lost interaction with hnRNPL, confirming our ITC data [Figure 5e, f].

Residues besides the LL pair in SETD2 are critical for hnRNP L binding

The mode of interaction between hnRNP L RRM2 and SETD2²¹⁶⁷⁻²¹⁹² peptide shares resemblance to the mechanism by which RRM2 of PTB, a splicing suppressor, binds to the PRI3 peptide of its co-repressor RAVR1³². The two peptides both contain a pair of Leu residues and bind to nearly the same position on the dorsal surface of the RRM [Supplementary Figure 3a, b]. We then asked whether the SETD2²¹⁶⁷⁻²¹⁹² peptide could also interact with PTB RRM2. The ITC results showed that there is no detectable affinity between SETD2²¹⁶⁷⁻²¹⁹² and PTB RRM2, suggesting that the amino acid sequence flanking the LL pair is important for the recognition to occur [Supplementary Figure 3c]. This is consistent with our data in which, besides SETD2CASHI 3 that has the LL pair crucial for interaction with hnRNP L, SETD2CASHI 1 and 2 also lost interaction with hnRNP L [Figure 1e, f].

To further ascertain these observations, we performed ITC of hnRNP L RRM2 with a truncated peptide, ²¹⁸⁰SNPNAGKVLLPTP²¹⁹², belonging to the SETD2 SHI domain. ITC revealed a K_D value of $12.4 \pm 1.47 \mu\text{M}$, indicating a weaker binding as compared to SETD2²¹⁶⁷⁻²¹⁹² and confirming that indeed the N-terminus region of SETD2²¹⁶⁷⁻²¹⁹² (containing the $\Delta 1$ and $\Delta 2$ deletions) also plays important role in the binding [Figure 5g, Supplementary Table 1]. Similar results were observed when the binding of hnRNP LL RRM2 with SETD2²¹⁸⁰⁻²¹⁹² was tested and found to be much weaker as compared to binding with SETD2²¹⁶⁷⁻²¹⁹² [Supplementary Table 1].

The functions of hnRNP L and LL are partially redundant

Previously, we have shown that SETD2 and hnRNP L co-regulate the AS of an overlapping set of events²³. To test whether SETD2 and hnRNP LL depletion result in overlapping transcriptome changes, RNA-seq was performed post depleting SETD2 and hnRNP LL in 293T cells. Depletion of SETD2 and hnRNP LL resulted in significant differential gene expression (FDR<0.05, fold change>1.5); 203 differentially expressed genes (57 upregulated, 146 downregulated) were observed upon SETD2 depletion while the expression levels of 514 genes (219 upregulated and 295 downregulated) were altered upon hnRNP LL depletion [Supplementary Figure 4a, Supplementary Table 3]. Furthermore, the analysis of differential AS events revealed that SETD2 and hnRNP LL depletion results in significant changes in 1225 and 1379 AS events, respectively [Supplementary Figure 4b, Supplementary Table 4]. However, the overlap between SETD2 and hnRNP LL regulated gene expression and AS events was very small.

One possible reason behind this observation might be the redundancy between the paralogs hnRNP L and LL. To test the possibility of redundancy in the regulation of transcripts by hnRNP L and LL, RNA-seq was performed post specific depletion of hnRNP L and LL in 293T cells. The depletion of the targets at the protein level was confirmed by western blotting with antibodies specific for hnRNP L and LL [Supplementary Figure 5a]. Also, hnRNP L depletion did not alter the transcript level of hnRNP LL and vice-versa [Figure 6a]. The RNA-seq data revealed a global perturbation in terms of transcription and AS changes upon hnRNP L and LL depletion [Supplementary Table 3 and 4]. Strikingly, the overlap of gene expression and AS changes brought about by hnRNP L and LL depletion were very few [Figure 6b, c, Supplementary Figure 5b]. Moreover, out of the 114 overlapping AS events, almost half of them (43.85%) showed an opposite trend of splicing [Figure 6d].

Although these results argue for a non-redundant function between hnRNP L and LL, the 10 fold higher expression of hnRNP L as compared to hnRNP LL has to be taken into consideration while interpreting these results¹³. This was also reflected in the distributed Normalized Spectral Abundance Factor (dNSAF) of these proteins in the MudPIT analysis of SETD2C purification, where hnRNP LL (dNSAF 0.002072) was 28-fold less enriched than hnRNP L (dNSAF 0.058612) [Supplementary Table 5]. Therefore, we decided to perform rescue experiments to test whether the differential AS events observed can be specifically rescued by one paralog but not the other. In order to perform rescue experiments, we first looked for suitable candidate genes. The genes *tjp1* and *bptf* have been reported previously to exhibit differential splicing upon hnRNP L depletion but not hnRNP LL¹³. Our RNA-Seq data confirmed these results as clear retention of exon 20 in *tjp1*- and exon 18a in *bptf* could be seen in genome browser tracks upon hnRNP L depletion but not in control and hnRNP LL depleted cells [Figure 6e, f]. To validate these results, individual AS events were measured by quantitative PCR (qPCR) and represented by the ratios of different exons. Again, the depletion of one paralog did not significantly alter the expression of the other, consistent with our RNA-Seq results [Figure 6g]. Indeed, hnRNP L depletion led to the increase in retention of exon 20 in *tjp1*, and exon 18a in *bptf* [Figure 6h, i]. Also, no significant change in the ratio of exons was observed upon hnRNP LL depletion consistent with our RNA-Seq data [Figure 6h, i]. Hence, *tjp1* and *bptf* were chosen as candidate genes to test rescue.

For rescue experiments, empty vector, hnRNP L, or hnRNP LL was introduced in 293T cells depleted of hnRNP L. qPCR revealed specific upregulation of *hnrnpl* and *hnrnp1l* as expected [Figure 6j]. Analysis of AS of *tjp1* and *bptf* revealed that the introduction of mCherry-HA-hnRNP L in hnRNP L depleted cells indeed resulted in the decrease in the ratio of exons, indicating rescue. The expression of mCherry alone did not have that effect despite having a much higher expression [Figure 6k, l, Supplementary Figure 6]. Strikingly, rescue with mCherry-

HA-hnRNP LL also resulted in the decrease in the ratio of exons, suggesting that it could rescue the AS changes brought about by hnRNP L depletion [Figure 6k, l].

These results show that the function of hnRNP L and hnRNP LL are at least partially redundant and the non-overlapping transcriptome changes observed upon their depletion might be due to differences in their expression level.

Mutating LL pair in SETD2 decreases its binding with splicing proteins

Previously, we have shown that the SHI domain mediates the interaction of SETD2 with numerous RNA processing proteins²³. We wanted to test whether mutating the LL pair within the SHI domain affects SETD2's ability to bind with other proteins besides hnRNP L and LL. For this, affinity-purified complexes of SETD2C mutants were subjected to MudPIT. A GO-term analysis of the proteins co-purified with the SETD2C revealed enrichment in RNA processing pathways [Figure 7a]. Notably, such enrichment was not observed on the GO-term analysis of co-purified proteins with the SETD2C-LL mutant [Figure 7b].

Next, IPA of the proteins identified through MudPIT revealed that similar to deleting the SHI domain, mutating the LL pair in the SHI domain led to a significant reduction in the enrichment of protein groups belonging to RNA processing [Figure 7c]. This was not observed upon the deletion of the SRI domain. A closer inspection of the specific proteins associated with such pathways revealed that mutating the LL pair in the SHI domain not only led to the loss of hnRNP L and LL interactions but also resulted in the loss of interaction with other RNA binding proteins [Figure 7d].

Collectively, the analysis suggests that the two consecutive leucines in the SHI domain are important for the interaction between SETD2 and proteins related to RNA processing.

DISCUSSION

It is now clear that AS is coupled to transcription which permits the sequential recognition of emerging splicing signals by the splicing machinery³³. The finding that hnRNPs such as hnRNP L and LL can directly bind to transcription regulators such as Med23 and SETD2 not only shows that they play an important role in such coupling but also signifies the importance of the protein-binding ability of RRMs. Importantly, the non-overlapping binding interfaces for engaging RNA and proteins in hnRNP L RRM2 as revealed by our crystal structure shows that hnRNP L/LL can form a ternary complex and strongly supports the possibility that the hnRNPs interacting partners aid in their specific recruitment to the target pre-mRNA that we have proposed before²³.

The canonical RRMs are characterized by a $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ - $\beta 4$ structure and the presence of two highly degenerate RNP consensus sequences, RNP-1 and RNP-2^{34,35}. The RRM contacts RNA using the RNP-1 and RNP-2 consensus sequences, which are present on the $\beta 3$ and $\beta 1$ strands, respectively. This involves primarily hydrophobic interactions between four conserved aromatic protein side chains and two bases, resulting in the binding of RNA to the β -sheet surface. Our crystal structure revealed that the binding of SETD2 occurs in a very different fashion where its LL pair inserts its side chains into the two hydrophobic cavities formed by the RRM2 in the opposite face of its RNA binding surface. This mode of interaction is very similar to the way RAVR1 binds PTB. This also shares resemblance with the way the peptide motif from SF1 binds to U2AF65 RRM3³⁶. In this case, the binding of SF1 occurs through the insertion of a tryptophan side chain into a hydrophobic groove on the dorsal face of the RRM. The similarities in the mode of SETD2-hnRNP L, RAVR1-PTB, and SF1-U2AF65 interaction suggests the possibility of a conserved design by which splicing regulators interact with one another.

Besides the LL pair, the N-terminus residues of SETD2 are also important for the binding to occur. Although the N-terminus of the SETD2²¹⁶⁷⁻²¹⁹² does not interact with the RRM2 directly,

we found that the residues form substantial van der Waals and hydrogen contacts with the C-terminus of the SETD2 peptide, possibly helping to fix the conformation of the peptide and thus, enhance the binding. Interestingly, a similar phenomenon was found in PTB-RAVER1 interactions, in which a long peptide bound more strongly as compared to the core interacting motif³². Also, the flanking region of the core interacting motif belongs to a predicted disordered region in both SETD2 and RAVR1 as per IUPRED2³⁷ [Supplementary Figure 3d]. This might provide the required flexibility needed to position the core interacting motif for engaging RRM.

Our crystal structure revealed the residues that are important in mediating the interaction between SETD2 and hnRNP L. Notably, some of them are mutated in cancer including the critical residue Y257 of hnRNP L, mutating which led to the abolishment of SETD2-hnRNP L interaction, and P2192 which is part of the LLPTP motif which hnRNP L seems to prefer [Supplementary Figure 7]. We already tested that one such mutation in cancer I214V, a substitution also found in hnRNP LL, results in weaker binding of SETD2. Changes in the expression of both SETD2 and hnRNP L are known to be associated with cancer progression³⁸⁻⁴⁰. It remains to be seen whether SETD2-hnRNP L interaction or the lack of thereof also affects transformation.

The paralogs hnRNP L and LL have a high sequence identity and RNA-binding specificity. Here we demonstrated that they both engage SETD2. Both are believed to exert their downstream effects in AS by recruiting hnRNPA1³⁰. Hence, it is not surprising that they are functionally at least partly redundant. The high expression level of hnRNP L could explain the existence of events that were exclusive to hnRNP L such as *tjp1* and *bptf*, which we confirmed could be rescued by ectopically expressing hnRNP LL also. Moreover, the GO-term analysis of genes showing decreased splicing upon hnRNP L depletion showed enrichment in RNA splicing and mRNA processing pathways [Supplementary Figure 8a]. This was not the case with hnRNP LL [Supplementary Figure 8b]. Moreover, the GO-term analysis of genes showing decreased

splicing uniquely upon hnRNP L depletion again showed enrichment in mRNA processing pathways. This makes the comparison between hnRNP L and LL regulated AS events more complicated.

It also must be considered that non-redundancy between these paralogs might exist, especially considering that differences in RNA binding constraints of these proteins have been noted³⁰. Moreover, hnRNP LL shows a tissue-specific expression, for instance, high levels in testes, which might suggest a possible tissue-specific role¹³. Regulation of hnRNPs by modulating expression of their tissue-specific paralogs is known. For instance, the replacement of PTB (hnRNP I) by its paralog, nPTB, which is less repressive for SRC N1 exon splicing than PTB, promotes assembly of an enhancer complex downstream of the exon⁴¹. Besides, the expression of hnRNP LL increases significantly during B cell to plasma cells differentiation and T cell activation pointing to a context-dependent function^{42,43}. It is possible that in such scenarios where hnRNP LL is more abundant, it might engage SETD2 more preferably than hnRNP L does. It will be interesting to examine in future studies what other proteins the hnRNPs interact with to govern co-transcriptional splicing and the context behind it.

METHODS

Plasmids

hnRNP L, hnRNP LL, and SETD2 human ORF were procured from Promega. Deletion mutants of hnRNP L, LL, and SETD2 were constructed by PCR (Phusion polymerase, NEB) using a full-length version of the constructs, respectively as a template, and individual fragments were cloned. All constructs generated were confirmed by sequencing. pCDNA3-ySet2 were procured from Addgene. siRNA for *setd2*, *hnrnp1l*, and *hnrnp1* as well as scramble siRNA sequences were procured from Dharmacon.

Cell line maintenance and drug treatment

293T cells were procured from ATCC and maintained in DMEM supplemented with 10% FBS and 2 mM L-glutamine at 37 °C with 5% CO₂. Transfections of plasmids were performed using Fugene HD (Promega) and that of siRNAs was performed using Lipofectamine RNAi Max (Thermofisher) at 40% cell confluency.

Affinity purification

293T cells expressing the protein of interest were harvested in 1xPBS and collected by centrifugation. The cells were lysed by resuspending in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton-X 100, 0.1% Na-deoxycholate, and a protease inhibitor cocktail). The lysed cells were centrifuged at 16000g for 20 min. The supernatant was collected and diluted 1:3 by adding dilution buffer (1x PBS, pH 7.5 with 1mM DTT and 0.005% NP-40). The diluted lysate was added to pre-equilibrated Magne® HaloTag® Beads (Promega, G7282) and incubated overnight on a rotator at 4 °C. The beads were then washed with wash buffer (50 mM Tris-HCL, pH 7.5, 300 mM NaCl, 0.005% NP40, and 1 mM DTT. AcTEV (ThermoFisher, 12575015) protease was used for elution.

Mass spectrometry

TCA precipitated protein samples were analyzed independently by Multidimensional Protein Identification Technology (MudPIT)^{44,45}. Briefly, precipitated protein samples were resuspended in 100mM Tris pH 8.5, 8M urea to denature the proteins. Proteins were reduced and alkylated prior to digestion with recombinant LysC (Promega) and trypsin (Promega). Reactions were quenched by the addition of formic acid (FA) to a final concentration of 5%. Peptide samples were pressure-loaded onto 100 µm fused silica microcapillary columns packed first with 9 cm of reverse phase material (Aqua; Phenomenex), followed by 3 cm of 5-µm Strong Cation Exchange material (Luna; Phenomenex), followed by 1 cm of 5-µm C18 RP. The loaded microcapillary columns were placed in-line with a 1260 Quaternary HPLC (Agilent). The

application of a 2.5 kV distal voltage electrosprayed the eluting peptides directly into LTQ linear ion trap mass spectrometers (Thermo Scientific) equipped with a custom-made nano-LC electrospray ionization source. Full MS spectra were recorded on the eluting peptides over a 400 to 1600 m/z range, followed by fragmentation in the ion trap (at 35% collision energy) on the first to fifth most intense ions selected from the full MS spectrum. Dynamic exclusion was enabled for 120 sec⁴⁶. Mass spectrometer scan functions and HPLC solvent gradients were controlled by the XCalibur data system (Thermo Scientific).

RAW files were extracted into .ms2 file format⁴⁷ using RawDistiller v. 1.0, in-house developed software⁴⁸. RawDistiller D(g, 6) settings were used to abstract MS1 scan profiles by Gaussian fitting and to implement dynamic offline lock mass using six background polydimethylcyclsiloxane ions as internal calibrants⁴⁸. MS/MS spectra were first searched using ProLuCID⁴⁹ with a 500 ppm mass tolerance for peptide and fragment ions. Trypsin specificity was imposed on both ends of candidate peptides during the search against a protein database combining 44,080 human proteins (NCBI 2019-11-03 release), as well as 426 common contaminants such as human keratins, IgGs, and proteolytic enzymes. To estimate false discovery rates (FDR), each protein sequence was randomized (keeping the same amino acid composition and length) and the resulting "shuffled" sequences were added to the database, for a total search space of 89,038 amino acid sequences. A mass of 57.0125 Da was added as a static modification to cysteine residues and 15.9949 Da was differentially added to methionine residues.

DTASelect v.1.9⁵⁰ was used in combination with our in-house script, swallow v. 0.0.1 (<https://github.com/tzwwen/kite>) to control FDRs to less than 1%. Results from each sample were merged and compared using CONTRAST⁵⁰. Combining all replicates, proteins had to be detected by at least 2 peptides and/or 2 spectral counts. Proteins that were subsets of others were removed using the parsimony option in DTASelect on the proteins detected after merging

all runs. Proteins that were identified by the same set of peptides (including at least one peptide unique to such protein group to distinguish between isoforms) were grouped together, and one accession number was arbitrarily considered as representative of each protein group.

NSAF⁵¹ was used to create the final reports on all detected peptides and non-redundant proteins identified across the different runs. Spectral and peptide level FDRs were, on average, $0.52 \pm 0.41\%$ and $0.39 \pm 0.1\%$, respectively. QPROT⁵² was used to calculate a log fold change and Z-score for the samples compared to the mock control.

For instances where there was more than one replicate analyzed by MudPIT, proteins with log fold change >1 and Z-score > 2 were further analyzed in Ingenuity Pathway Analysis (IPA, Qiagen) to determine pathways enriched by the bait proteins. For proteins with only one replicate, a ratio was calculated of dNSAF values between sample and mock. For those to be further analyzed in IPA, the dNSAF ratio had to be > 2 compared to mock. Pathways were considered significantly enriched with p-value < 0.05 ($-\log_{10}(\text{p-value}) > 1.3$).

Protein expression and purification

The hnRNP L_RRM2 fragment (residues 189-286) was amplified from the human brain cDNA library and cloned into a modified pET28a (Novagen) vector without a thrombin protease cleavage site (termed p28a). All the mutants were generated using a MutantBEST kit (Takara) and verified via DNA sequencing. All the proteins were expressed in *Escherichia coli* BL21 (DE3) cells. The cells were cultured at 37 °C in LB medium, until the OD₆₀₀ reached about 0.8. Then the proteins were induced with 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). After induction at 16°C for 24h, the cells were harvested and lysed by sonication in buffer A (20 mM Tris-HCl, pH 7.8, 1 M NaCl). Proteins were first purified by Ni-NTA agarose beads and further purified by size-exclusion chromatography on a Hiload 16/60 Superdex 75 column (GE Healthcare) in buffer A. Purified proteins were dialyzed with Buffer B (20 mM Tris-HCl pH 7.8, 150 mM NaCl) and concentrated for subsequent analysis.

Protein complex preparation

The synthetic SETD2 (residues 2167-2192 and 2180-2192) peptides were dissolved into buffer B, and the peptide was mixed with the purified protein at a 1.5:1 molar ratio and incubated at 16 °C overnight to form a complex. The complex was condensed to 1.0 mM in preparation for crystallization.

Crystallography

The crystals were grown at 20 °C via the sitting-drop vapor diffusion method. The crystals of hnRNP L in complex with SETD2²¹⁶⁷⁻²¹⁹² were grown by mixing 1 µL of the protein complex and 1 µL of reservoir buffer (0.8 M Potassium/Sodium phosphate, pH 7.5). The crystals of hnRNP LL in complex with SETD2²¹⁸⁰⁻²¹⁹² were grown by mixing 1 µL of the protein complex and 1 µL of reservoir buffer (0.3 M Ammonium Sulfate, 20% PEG 4000). All the crystals were harvested in their corresponding reservoir buffers supplemented with 25% (v/v) glycerol and frozen in liquid nitrogen.

X-ray diffraction data sets of the crystals were collected at beamline 19U1 at the Shanghai Synchrotron Radiation Facility (SSRF) with a diffraction wavelength of 0.979 Å. The two data sets, including hnRNP L-SETD2²¹⁶⁷⁻²¹⁹² and hnRNP LL-SETD2²¹⁸⁰⁻²¹⁹², were indexed, integrated, and scaled by the HKL-2000 program suite⁵³. The structure of the hnRNP L-SETD2²¹⁶⁷⁻²¹⁹² complex was determined by molecular replacement with the MOLREP program using the structure of PTB1-PRI3 (PDB ID: 3ZZY) as the search model. The structure of the hnRNP LL-SETD2²¹⁸⁰⁻²¹⁹² complex was determined by molecular replacement with the MOLREP program^{54,55} using the structure of hnRNP L-SETD2²¹⁶⁷⁻²¹⁹² as the search model. The model was further built and refined using Coot⁵⁶ and Phenix.refine⁵⁷, respectively. Crystal diffraction data and refinement statistics are shown in the Supplementary Table. All the structures in the figures were generated using PyMOL (DeLano Scientific LLC).

Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) assays were performed at 20 °C by using a Microcal PEAQ-ITC instrument (Malvern). We conducted the ITC experiments using SETD2 peptides for titration into hnRNP L and hnRNP LL proteins. SETD2 peptides are added to the syringe at a concentration of about 600 μ M and proteins are added to the sample pool at a concentration of about 50 μ M. A typical ITC experiment is consisted of 19 drops, with one injection of 1 μ L followed by 18 injections of 2 μ L of protein sample. The integrated heat data were analyzed using a one-site binding model by MicroCal PEAQ-ITC Analysis Software provided by the manufacturer.

Isolation of total RNA and PCR

Total RNA was extracted from cells as per the manufacturer's (Qiagen) instructions. It was further treated with DNaseI (NEB) for 30 min at 72 °C to degrade any possible DNA contamination. RNA (2 μ g) was subjected to reverse transcription using QScript cDNA synthesis mix according to the manufacturer's instructions. cDNAs were then amplified with the corresponding gene-specific primer sets. For RTPCR, PCR was conducted for 24 cycles using the condition of 30 s at 94 °C, 30 s at 60 °C and 30 s at 72 °C. The PCR products were analyzed on 1% agarose gels containing 0.5 μ g/ml ethidium bromide. The sequence of oligos is in Supplementary Table 1.

Antibodies

hnRNP L (CST 37562, dilution 1:3000), FLAG (Sigma-Aldrich A8592, dilution 1:10000), Pol II (Abcam ab5095, dilution 1:5000), Halo (Promega G9211, dilution 1:10000), SETD2 (Abclonal A3194, dilution 1:6000 and Aviva OAEB00589, dilution 1:3000), HA (Sigma 04-902, dilution 1:10000), His (Abcam ab18184, dilution 1:3000), H3K36me3 (CST 4909S, dilution 1:1000) H3 (CST 9715S, dilution 1:3000), β -actin (Abcam ab8224, dilution 1:2500). The scanned films from western blotting experiments are provided as Source Data file.

High throughput sequencing

Sequencing libraries were prepared using High Throughput Library Prep Kit (KAPA Biosystems) following the manufacturer's instructions. The library was sequenced on an Illumina HiSeq platform with paired reads of 75 bp for RNA-seq.

RNA-seq analysis

Raw reads were demultiplexed into FASTQ format allowing up to one mismatch using Illumina bcl2fastq v2.18. Reads were aligned to the human genome (hg38 and Ensembl 102 gene models) using STAR (version STAR_2.7.3a)⁵⁸. TPM expression values were generated using RSEM (version v1.3.0). edgeR (version 3.24.3 with R 3.5.2) was applied to perform differential expression analysis, using only protein-coding and lncRNA genes⁵⁹. To perform differential splicing analysis, we used rMATs (version 4.0.2) with default parameters starting from FASTQ files⁶⁰. FDR cutoff of 0.05 was used to determine statistical significance.

DATA AVAILABILITY

All relevant data are available. The RNA-data sets are available in the Gene Expression Omnibus (GEO) database under the accession number GSE174426 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174426>). The mass spectrometry proteomics data is available at the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD019376 (<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD019376>) and PXD022946 (<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD022946>). Also, raw mass spectrometry data and search results files have been deposited to the Proteome Xchange (accession: PXD025942) via the MassIVE repository and may be accessed at <ftp://MSV000087403@massive.ucsd.edu> with password stb-0511-2021. The hnRNP L-SETD2 and the hnRNP LL-SETD2 complexes have been deposited to PDB with entry ID: 7EVR and 7EVS, respectively.

REFERENCES

1. Wang, Y. et al. Mechanism of alternative splicing and its regulation. *Biomedical reports* **3**, 152-158 (2015).
2. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**, 1413-5 (2008).
3. Wang, E.T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-6 (2008).
4. Barbosa-Morais, N.L. et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587-93 (2012).
5. Kalsotra, A. & Cooper, T.A. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12**, 715-29 (2011).
6. Han, N., Li, W. & Zhang, M. The function of the RNA-binding protein hnRNP in cancer metastasis. *J Cancer Res Ther* **9 Suppl**, S129-34 (2013).
7. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851-67 (2016).
8. Hutten, S. & Dormann, D. hnRNPA2/B1 Function in Neurodegeneration: It's a Gain, Not a Loss. *Neuron* **92**, 672-674 (2016).
9. Kattuah, W. et al. Heterogeneous Nuclear Ribonucleoprotein E2 (hnRNP E2) Is a Component of TDP-43 Aggregates Specifically in the A and C Pathological Subtypes of Frontotemporal Lobar Degeneration. *Front Neurosci* **13**, 551 (2019).
10. Douglas, J.N. et al. Antibodies to the RNA-binding protein hnRNP A1 contribute to neurodegeneration in a model of central nervous system autoimmune inflammatory disease. *J Neuroinflammation* **13**, 178 (2016).
11. Heiner, M., Hui, J., Schreiner, S., Hung, L.-H. & Bindereif, A. HnRNP L-mediated regulation of mammalian alternative splicing by interference with splice site recognition. *RNA Biology* **7**, 56-64 (2010).
12. Rothrock, C.R., House, A.E. & Lynch, K.W. HnRNP L represses exon splicing via a regulated exonic splicing silencer. *The EMBO Journal* **24**, 2792-2802 (2005).
13. Hung, L.-H. et al. Diverse roles of hnRNP L in mammalian mRNA processing: A combined microarray and RNAi analysis. *RNA* **14**, 284-296 (2008).
14. Weighardt, F., Biamonti, G. & Riva, S. The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *BioEssays* **18**, 747-756 (1996).
15. McClory, S.P., Lynch, K.W. & Ling, J.P. HnRNP L represses cryptic exons. *RNA (New York, N.Y.)* **24**, 761-768 (2018).
16. HUI, J., REITHER, G. & BINDEREIF, A. Novel functional role of CA repeats and hnRNP L in RNA stability. *RNA* **9**, 931-936 (2003).
17. Jafarifar, F., Yao, P., Eswarappa, S.M. & Fox, P.L. Repression of VEGFA by CA-rich element-binding microRNAs is modulated by hnRNP L. *The EMBO Journal* **30**, 1324-1334 (2011).
18. Lee, D.-H. et al. hnRNP L binds to CA repeats in the 3'UTR of bcl-2 mRNA. *Biochemical and Biophysical Research Communications* **382**, 583-587 (2009).
19. Hui, J. et al. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *The EMBO Journal* **24**, 1988-1998 (2005).
20. Waterston, R.H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).

21. Chaudhury, A., Chander, P. & Howe, P.H. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA (New York, N.Y.)* **16**, 1449-1462 (2010).
22. Motta-Mena, L.B., Heyd, F. & Lynch, K.W. Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Mol Cell* **37**, 223-34 (2010).
23. Bhattacharya, S. et al. The methyltransferase SETD2 couples transcription and splicing by engaging mRNA processing factors through its SHI domain. *Nature Communications* **12**, 1443 (2021).
24. Huang, Y. et al. Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Molecular cell* **45**, 459-469 (2012).
25. Blatter, M. et al. The Signature of the Five-Stranded vRRM Fold Defined by Functional, Structural and Computational Analysis of the hnRNP L Protein. *J Mol Biol* **427**, 3001-22 (2015).
26. Skrisovska, L. & Allain, F.H.T. Improved Segmental Isotope Labeling Methods for the NMR Study of Multidomain or Large Proteins: Application to the RRM of Npl3p and hnRNP L. *Journal of Molecular Biology* **375**, 151-164 (2008).
27. Zhang, W. et al. Crystal Structures and RNA-binding Properties of the RNA Recognition Motifs of Heterogeneous Nuclear Ribonucleoprotein L*: INSIGHTS INTO ITS ROLES IN ALTERNATIVE SPLICING REGULATION. *Journal of Biological Chemistry* **288**, 22636-22649 (2013).
28. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research* **44**, W344-W350 (2016).
29. Bhattacharya, S. & Workman, J.L. Regulation of SETD2 stability is important for the fidelity of H3K36me3 deposition. *Epigenetics & Chromatin* **13**, 40 (2020).
30. Shankarling, G. & Lynch, K.W. Minimal functional domains of paralogues hnRNP L and hnRNP LL exhibit mechanistic differences in exonic splicing repression. *The Biochemical journal* **453**, 271-279 (2013).
31. Smith, S.A. et al. Paralogs hnRNP L and hnRNP LL Exhibit Overlapping but Distinct RNA Binding Constraints. *PLOS ONE* **8**, e80701 (2013).
32. Rideau, A.P. et al. A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. *Nat Struct Mol Biol* **13**, 839-48 (2006).
33. Carrillo Oesterreich, F., Bieberstein, N. & Neugebauer, K.M. Pause locally, splice globally. *Trends Cell Biol* **21**, 328-35 (2011).
34. Cléry, A., Blatter, M. & Allain, F.H.T. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology* **18**, 290-298 (2008).
35. Han, S.P., Tang, Y.H. & Smith, R. Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* **430**, 379-92 (2010).
36. Selenko, P. et al. Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF65 and SF1/mBBP. *Molecular Cell* **11**, 965-976 (2003).
37. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research* **46**, W329-W337 (2018).
38. Fahey, C.C. & Davis, I.J. SETting the Stage for Cancer Development: SETD2 and the Consequences of Lost Methylation. *Cold Spring Harbor perspectives in medicine* **7**, a026468 (2017).
39. Li, J. et al. SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget* **7**, 50719-50734 (2016).
40. Gu, J., Chen, Z., Chen, X. & Wang, Z. Heterogeneous nuclear ribonucleoprotein (hnRNPL) in cancer. *Clinica Chimica Acta* **507**, 286-294 (2020).
41. Markovtsov, V. et al. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol Cell Biol* **20**, 7463-79 (2000).

42. Chang, X., Li, B. & Rao, A. RNA-binding protein hnRNPLL regulates mRNA splicing and stability during B-cell to plasma-cell differentiation. *Proceedings of the National Academy of Sciences* **112**, E1888 (2015).
43. Benson, M.J. et al. Heterogeneous nuclear ribonucleoprotein L-like (hnRNPLL) and elongation factor, RNA polymerase II, 2 (ELL2) are regulators of mRNA processing in plasma cells. *Proceedings of the National Academy of Sciences* **109**, 16252 (2012).
44. Florens, L. & Washburn, M.P. Proteomic analysis by multidimensional protein identification technology. *Methods Mol Biol* **328**, 159-75 (2006).
45. Washburn, M.P., Wolters, D. & Yates, J.R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19**, 242-247 (2001).
46. Zhang, Y., Wen, Z., Washburn, M.P. & Florens, L. Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Analytical Chemistry* **81**, 6317-6326 (2009).
47. McDonald, W.H. et al. MS1, MS2, and SQT—three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Communications in Mass Spectrometry* **18**, 2162-2168 (2004).
48. Zhang, Y., Wen, Z., Washburn, M.P. & Florens, L. Improving Proteomics Mass Accuracy by Dynamic Offline Lock Mass. *Analytical Chemistry* **83**, 9344-9351 (2011).
49. Xu, T. et al. ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *Journal of proteomics* **129**, 16-24 (2015).
50. Tabb, D.L., McDonald, W.H. & Yates, J.R., 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of proteome research* **1**, 21-26 (2002).
51. Zhang, Y., Wen, Z., Washburn, M.P. & Florens, L. Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Analytical Chemistry* **82**, 2272-2281 (2010).
52. Choi, H., Kim, S., Fermin, D., Tsou, C.-C. & Nesvizhskii, A.I. QPROT: Statistical method for testing differential expression using protein-level intensity data in label-free quantitative proteomics. *Journal of proteomics* **129**, 121-126 (2015).
53. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276**, 307-326 (1997).
54. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
55. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22-5 (2010).
56. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
57. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
58. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2012).
59. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139-140 (2010).
60. Shen, S. et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences* **111**, E5593 (2014).

ACKNOWLEDGEMENTS

This work was supported by funding from the National Institute of General Medical Sciences (grant no. R35GM118068) and the Stowers Institute for Medical Research to Jerry L Workman. This work was also financially supported by the grants from Ministry of Science and Technology of China [2016YFA0500700, 2019YFA0508403]; Strategic Priority Research Program of the Chinese Academy of Sciences [XDB39000000]; Chinese National Natural Science Foundation [32090040, 31870760, U1932122]. The authors would like to thank the members of the Workman lab for their critical suggestions to improve the manuscript. We thank the staff of BL17U1, BL18U1 and BL19U1 beamlines at the Shanghai Synchrotron Radiation Facility for assistance during x-ray diffraction data collection.

AUTHOR CONTRIBUTIONS

S.B. and S.W. contributed equally to the manuscript. S.B. conceptualized the work, designed, and performed the experiments. S.B. wrote the manuscript. S.W. performed ITC and crystallization experiments. S.W., S.S., Y.S., and F.L. analyzed the ITC and crystallography data and wrote the structural aspect of the work. D.R. generated the mutants and performed qPCR. Y. Z. and L.F. conducted mass spectrometry. Y.Z., L.F., and M.P.W. analyzed mass spectrometry data. N.Z. and H.L. analyzed the high-throughput sequencing data. F.L., Y.S. and J.L.W conceived the idea of the work, provided supervision, acquired funding, and revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

Figure 1. **A conserved stretch in SETD2 is responsible for its interaction with hnRNP L.** (a) Multiple sequence alignment of previously identified SH1 domain in SETD2. (b) The raw ITC titration data of hnRNP L RRM2 with SETD2²¹⁶⁷⁻²¹⁹² and its fitting curve are shown. KD, dissociation constant; DP, differential power; N, binding stoichiometry. (c) ITC fitting curves of hnRNP L RRM2

with SETD2²¹⁶⁷⁻²¹⁹² (black) and SETD2²¹¹³⁻²¹⁴⁰ (red) are shown. **(d)** Illustration showing the deletions in SETD2 SHI domain that were made to perform affinity purification. **(e)** Silver staining and **(f)** western blotting of affinity purified complexes of SETD2C and its mutants.

Figure 2. Crystal structure reveals the molecular basis of hnRNP L-SETD2 interaction. **(a)** Ribbon representations of hnRNP L RRM2 bound to the SETD2²¹⁶⁷⁻²¹⁹² peptide. hnRNP L is colored in purple and the bound SETD2²¹⁶⁷⁻²¹⁹² peptide is colored in orange. **(b)** The SETD2²¹⁶⁷⁻²¹⁹² peptide is represented as sticks on the molecular face of hnRNP L RRM2. Red and blue colors denote negative and positive surface charge, respectively. **(c-f)** Closeup views of the interactions between hnRNP L and the SETD2²¹⁶⁷⁻²¹⁹² peptide. (Left) The van der Waals surface views of hnRNP L-SETD2²¹⁶⁷⁻²¹⁹². hnRNP L (purple) and SETD2²¹⁶⁷⁻²¹⁹² (orange) are shown as ribbons with selected sidechains as sticks. The van der Waals surface of the hnRNP L is depicted as a semitransparent skin. The SETD2 peptide is represented as a stick diagram (orange). (Right) Hydrogen bonds are shown as black dashed lines.

Figure 3. SETD2 binds to hnRNP LL through its SHI Domain. **(a, d, h)** Cartoon illustrating the domain organization of hnRNP L, hnRNP LL, SETD2 and ySet2. **(b, c)** Heat maps showing the enrichment of proteins normalized to the bait (SETD2C) in MudPIT analysis. **(e, f, j)** Western blotting and silver staining of affinity-purified complexes of Halo-SETD2C or Halo-ySet2 and its mutants from 293T extracts. **(g, i)** Table showing the dNSAFs of the listed proteins post mass spectrometry analysis of purified complexes obtained by affinity purification of Halo-SETD2 or ySet2 from 293T extracts. AWS-Associated with SET, SET-Su(var)3-9, Enhancer-of-zeste and Trithorax, SRI-Set2-Rpb1 Interaction, SHI- SETD2-hnRNP Interaction, dNSAF-distributed Normalized Spectral Abundance Factor, NLS-Nuclear Localization Signal.

Figure 4. hnRNP LL RRM2 -SETD2 interaction occurs in a similar fashion as hnRNP L-SETD2 binding. **(a)** Sequence alignment of hnRNP L RRM2 and hnRNP LL RRM2 proteins. The alignment was generated by ESPript3 with CLUSTALW. The secondary structures of hnRNP L RRM2, as determined by DSSP, are shown above the sequences. Red squares denote identical residues

whereas black triangles highlight the key residues involved in the interaction with SETD2. **(b, e)** Halo purification was performed from extracts of 293T cells co-expressing Halo-tagged SETD2C and mCherry-HA-hnRNP L/LL constructs. Input and eluted samples were resolved on gel followed by western blotting. The experiment was repeated at least 2 times all yielding similar results. **(c)** Halo purification was performed from extracts of 293T cells co-expressing Halo-HA-tagged hnRNP LL and GFP-FLAG-SETD2C constructs. Input and eluted samples were resolved on gel followed by western blotting. The experiment was repeated at least 2 times all yielding similar results. **(d)** Schematic representation of hnRNP LL segments used in purification experiments. **(f)** The raw ITC titration data of hnRNP LL with SETD2²¹⁶⁷⁻²¹⁹² and its fitting curve are shown. **(g)** (Left) Structure comparison of hnRNP L_RRM2 - SETD2²¹⁶⁷⁻²¹⁹² complex (orange) and hnRNP LL_RRM2 - SETD2²¹⁶⁷⁻²¹⁹² complex (purple). (Right) Two complexes are shown as ribbons with selected sidechains as sticks. KD, dissociation constant; DP, differential power; N, binding stoichiometry. RRM-RNA Recognition Motif, NLS-Nuclear Localization Signal.

Figure 5. Mutating key residues in hnRNP L and SETD2 disrupts their *in vitro* and *in vivo* binding. **(a, b)** ITC fitting curves of hnRNP L RRM2 WT and mutants with SETD2²¹⁶⁷⁻²¹⁹² peptide are shown. **(c, e, f)** Western blotting and silver staining of affinity-purified complexes of Halo-SETD2C WT and its mutants. In (c) mCherry-HA-hnRNP L proteins were used as prey. * denotes non-specific band. **(d, g)** ITC fitting curves of hnRNP L RRM2 with SETD2 peptides.

Figure 6. Transcriptome-wide and gene-specific analysis of the regulatory effect of hnRNP L paralogs. **(a)** Chart showing the decrease in expression of the genes depicted based on RNA-seq analysis post siRNA treatment. **(a)** RNA was isolated from 293T cells transfected with siRNA and RT-PCR was performed to check transcript levels. *gapdh* was used as a normalization control. Western blot of whole-cell lysates was performed with the depicted antibodies. The experiment was repeated at least 7 times all yielding similar results. **(b, c)** Venn diagrams showing the overlap of differentially expressed genes and AS events upon *hnrnp1* and *hnrnp11* depletion as compares to scramble siRNA treated cells. **(d)** Pie chart showing the fractions of differentially AS events that

occur in both *hnrnp1* and *hnrnp11* depletion. **(e, f)** Genome browser view showing retention of introns in *tjp1* and *bptf* genes upon *hnrnp1* depletion. **(g-i)** RNA was isolated from 293T cells 72 hours post-transfection with scramble siRNA, or siRNA against *hnrnp1* or *hnrnp11*. Also, **(j-l)** RNA was isolated from sihRNPL expressing 293T cells rescued with vector control (VC), hnRNP L or hnRNP LL constructs. Specific primers were designed to detect the indicated genes and exons and individual alternative splicing events were measured by quantitative PCR and represented by the ratios of depicted exons. For each sample n = 3 independent biological samples were examined in the same sequencing run. Data are presented as mean values with Standard Error of Mean. Unpaired t test (two-tailed) was performed. p-value <0.05 was considered significant. p-values are depicted on the top of the respective graphs.

Figure 7. Mutating LL pair in SH1 domain results in decreased interaction of SETD2 with RNA binding proteins. **(a, b)** GO-term analysis of proteins using ShinyGO (<http://bioinformatics.sdstate.edu/go/>) identified by MudPIT in the affinity-purification of SETD2C and SETD2C-LL mutant. **(c, d)** Heat maps showing the enrichment of pathways in the IPA (Ingenuity Pathway Analysis) and proteins in MudPIT analysis. Enrichment of proteins is normalized to the bait (SETD2C).

Figure 1

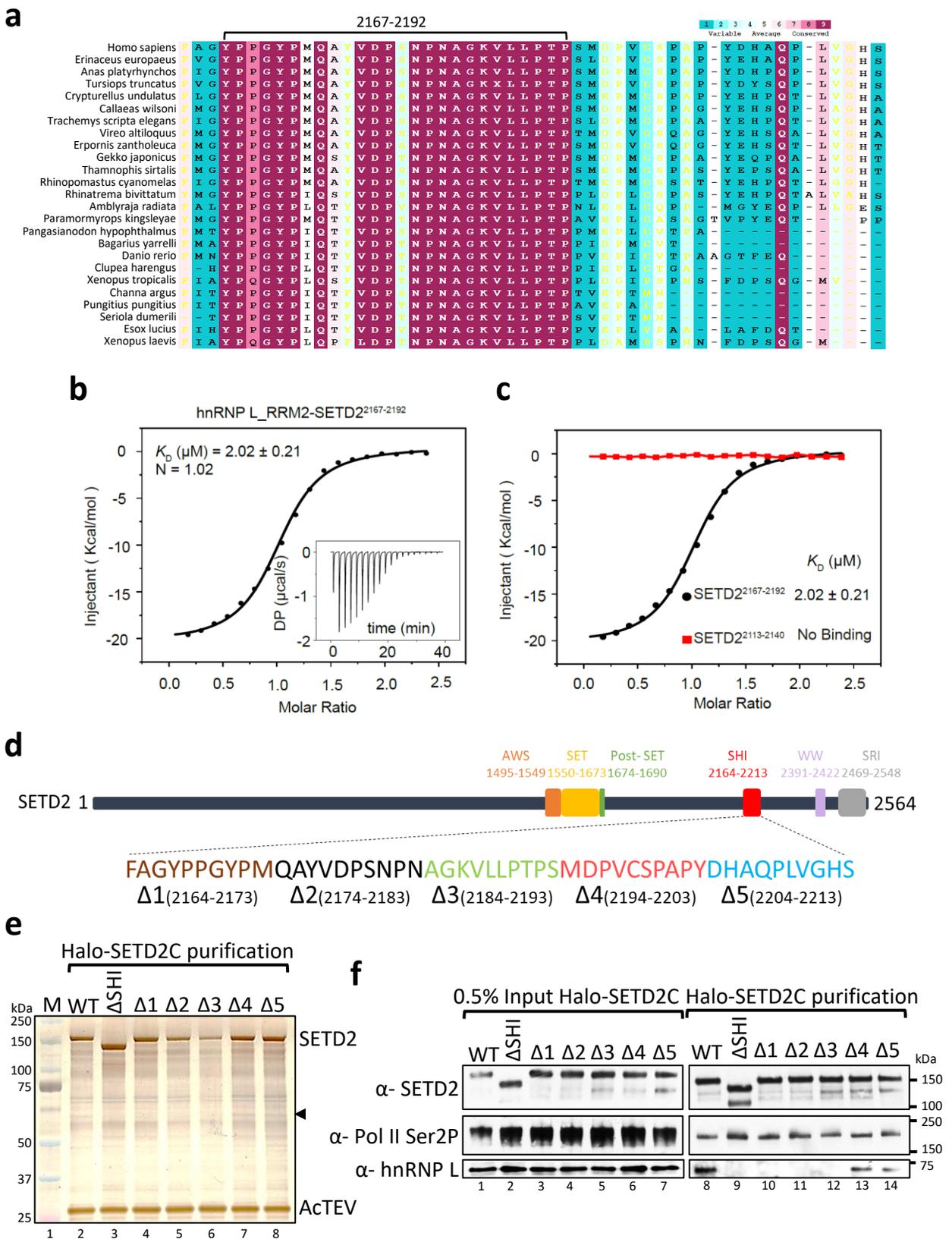


Figure 2

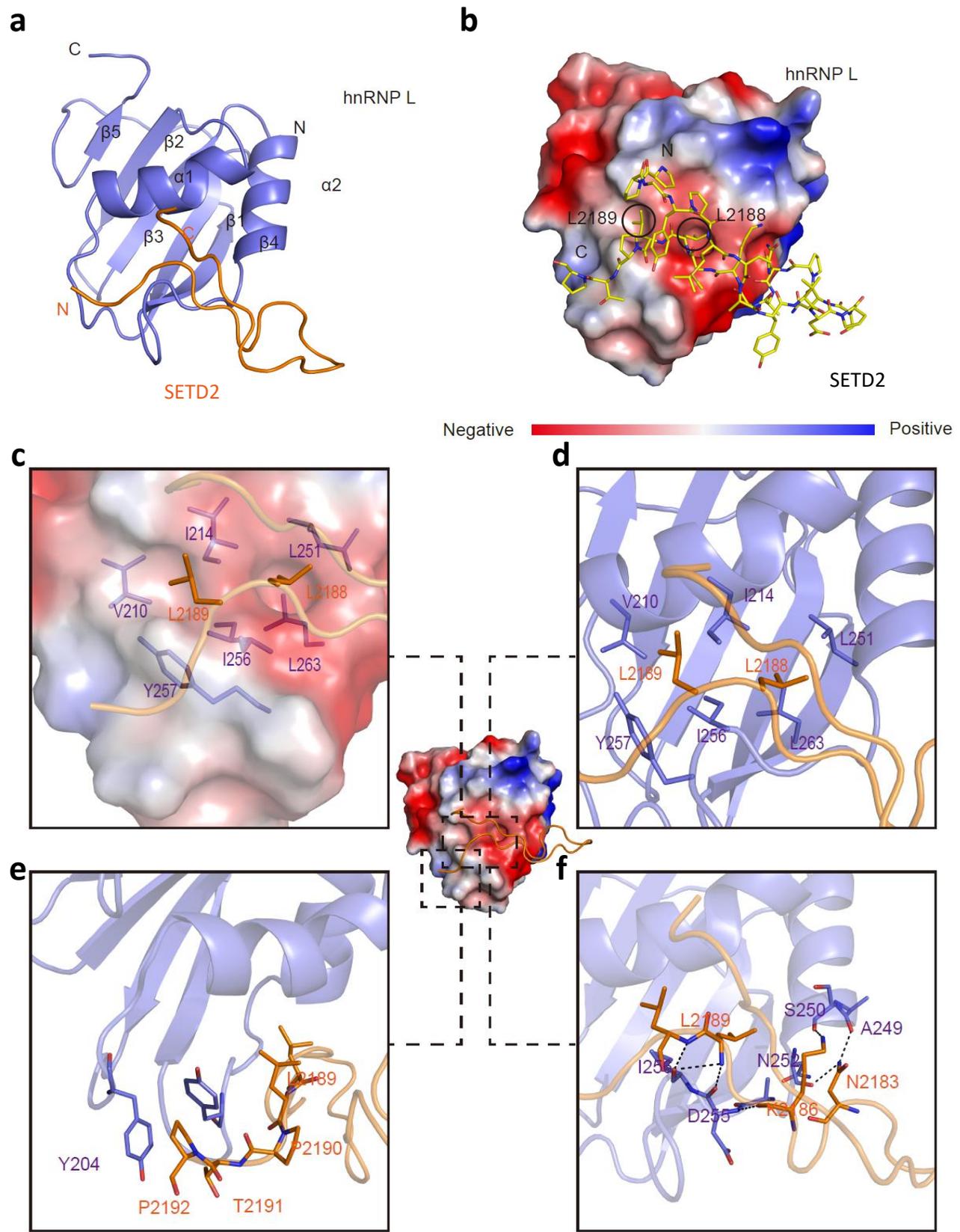


Figure 3

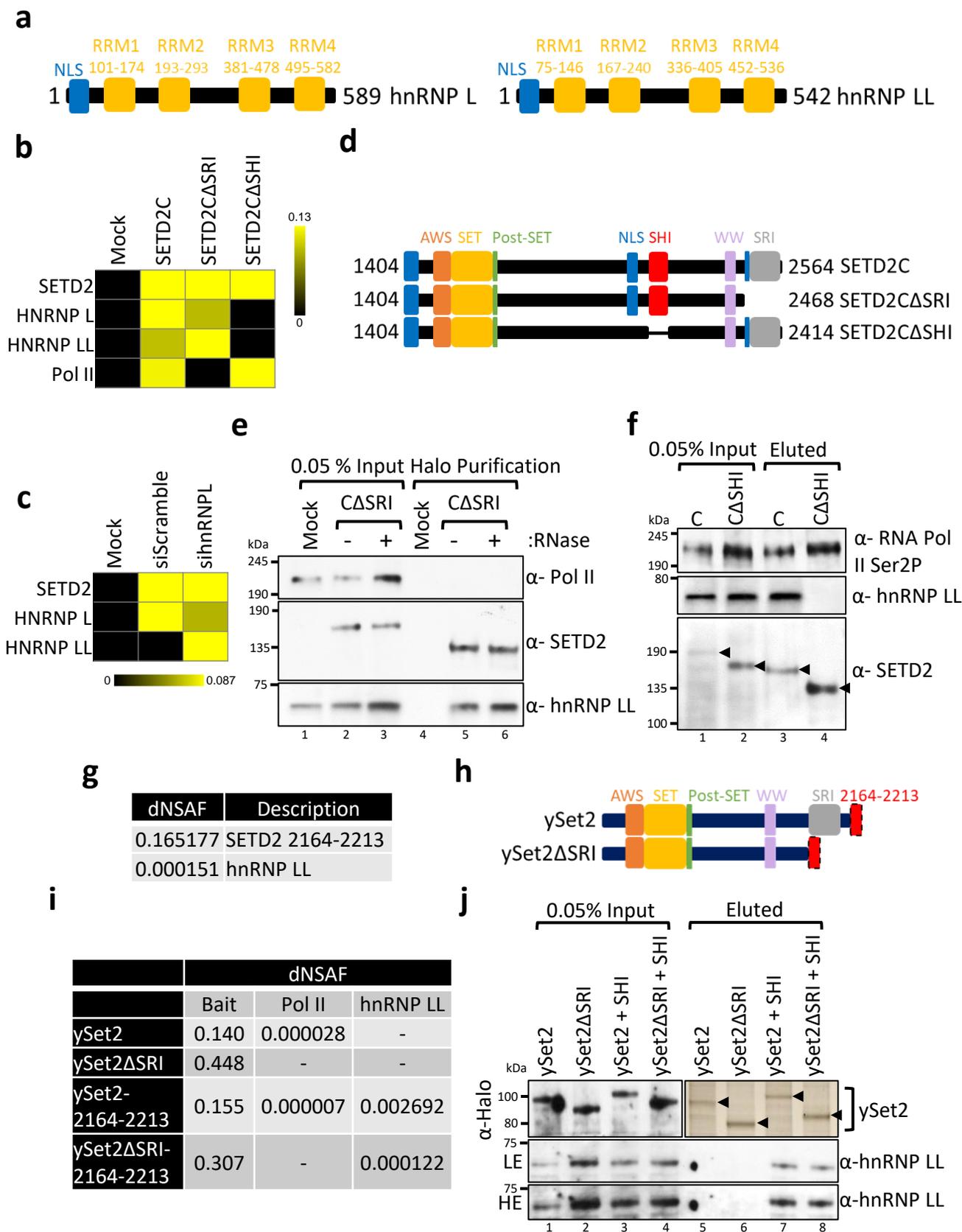
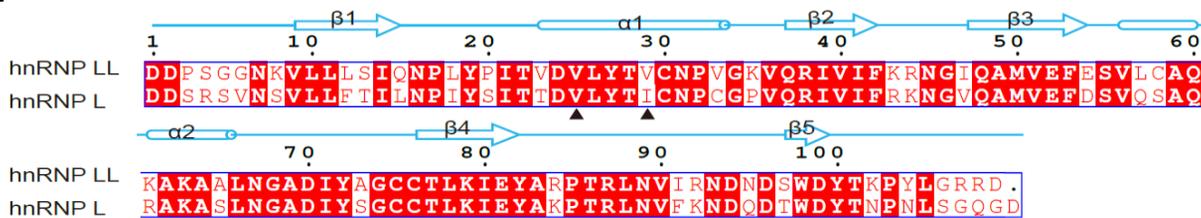
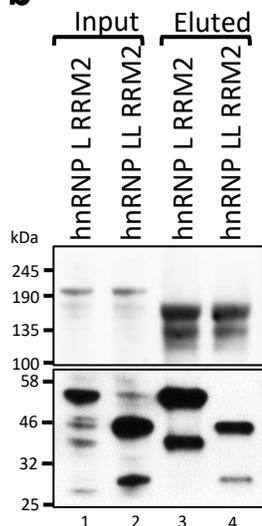


Figure 4

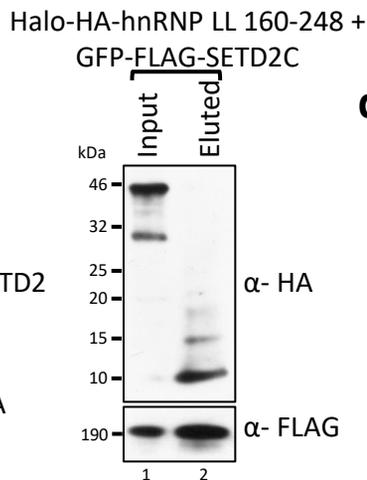
a



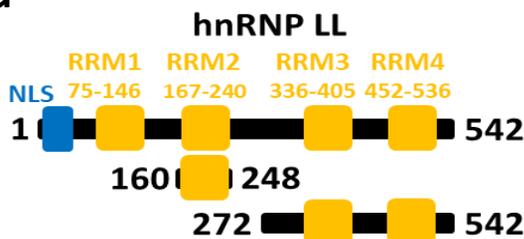
b



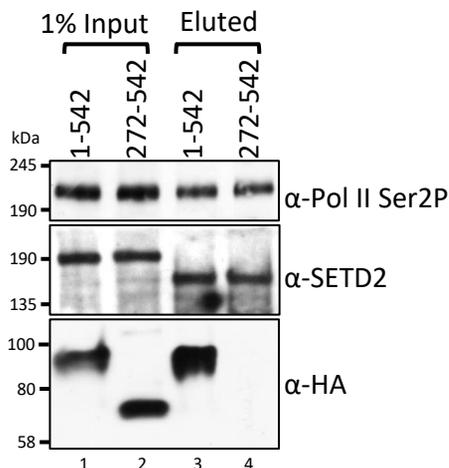
c



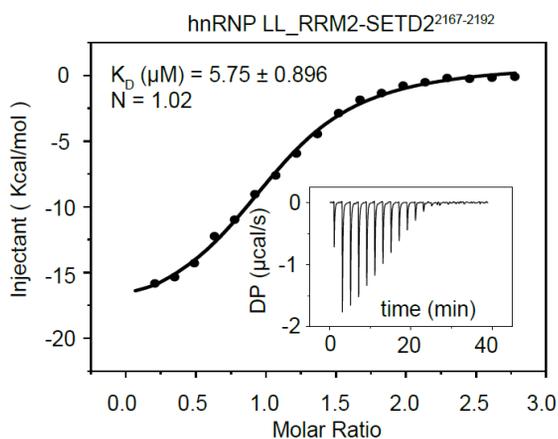
d



e



f



g

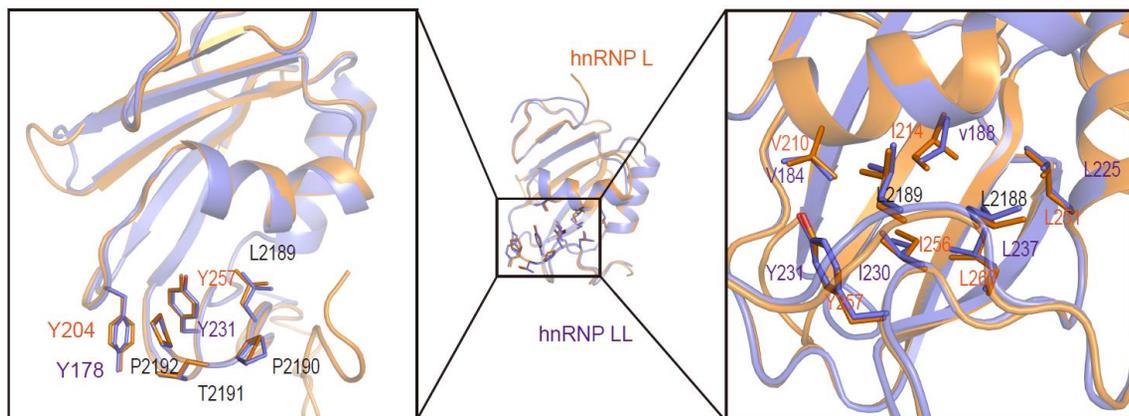


Figure 5

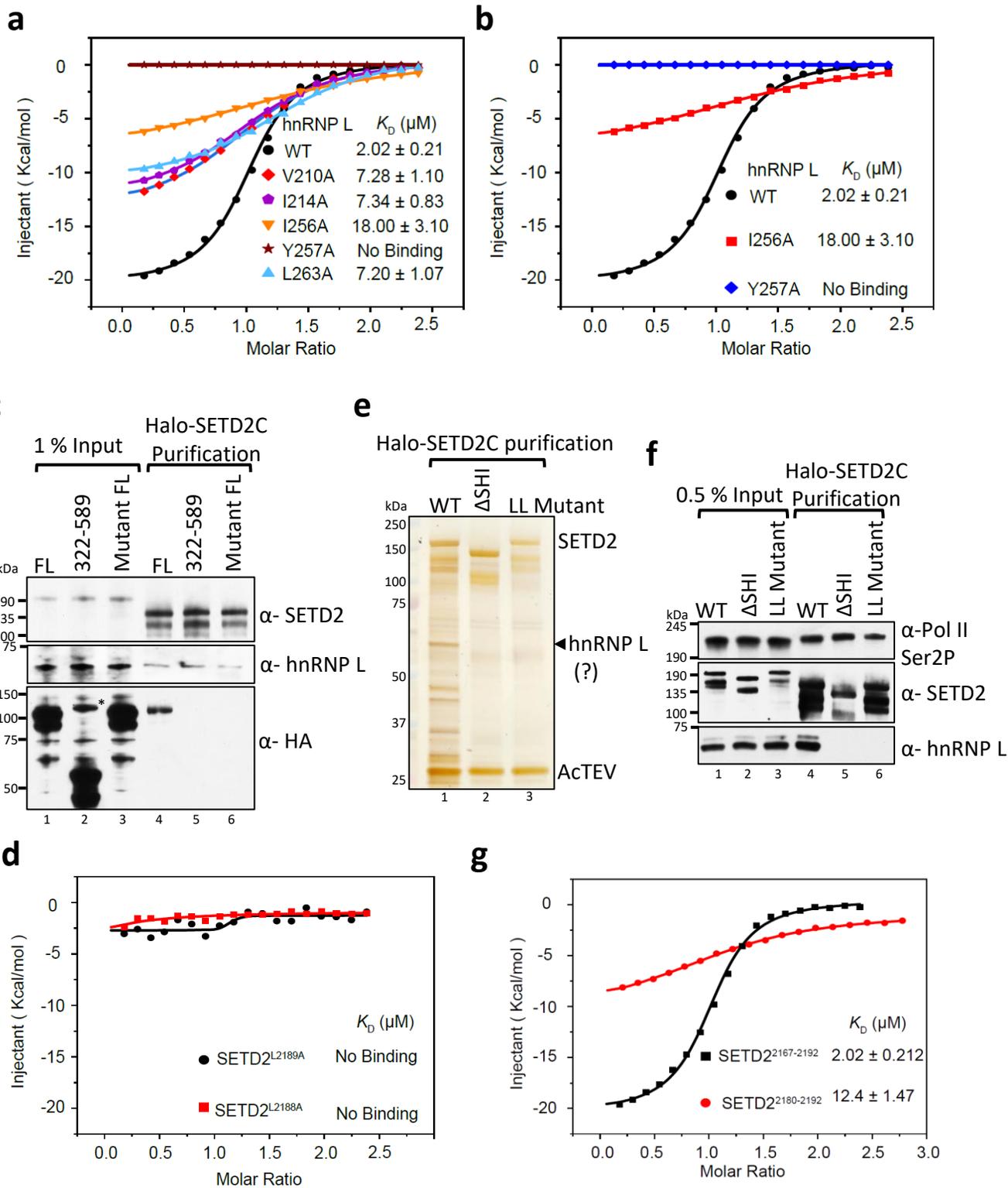


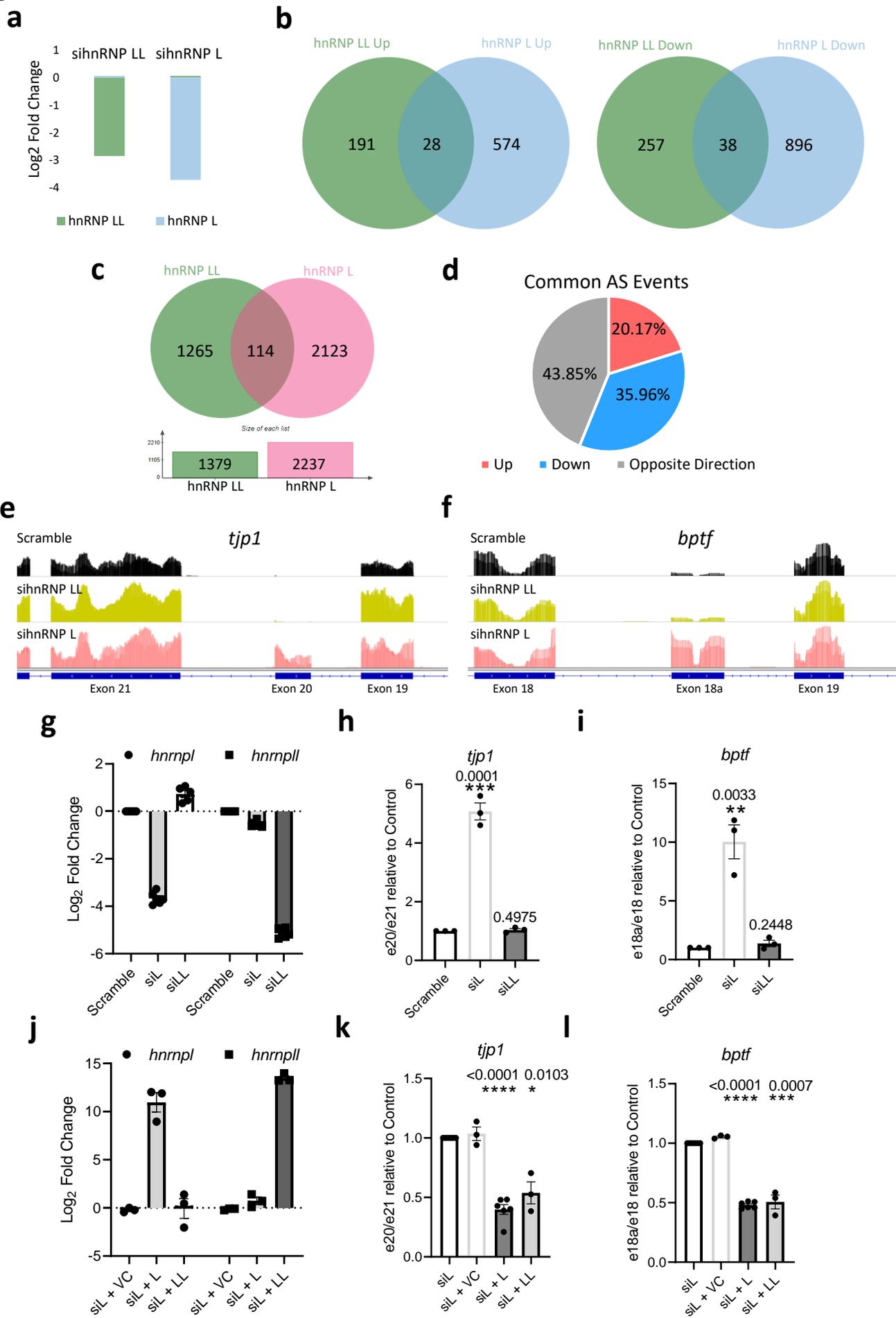
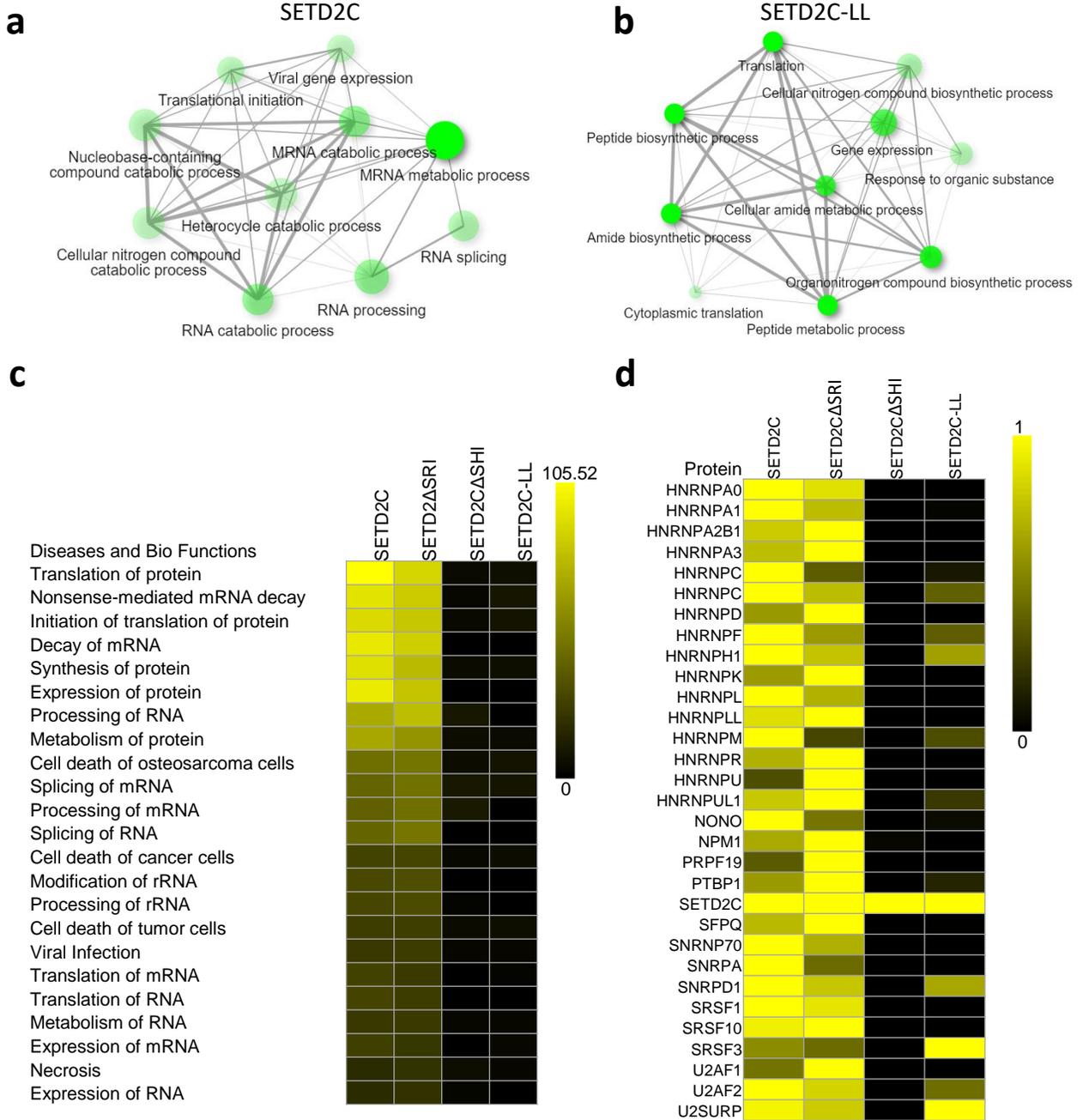
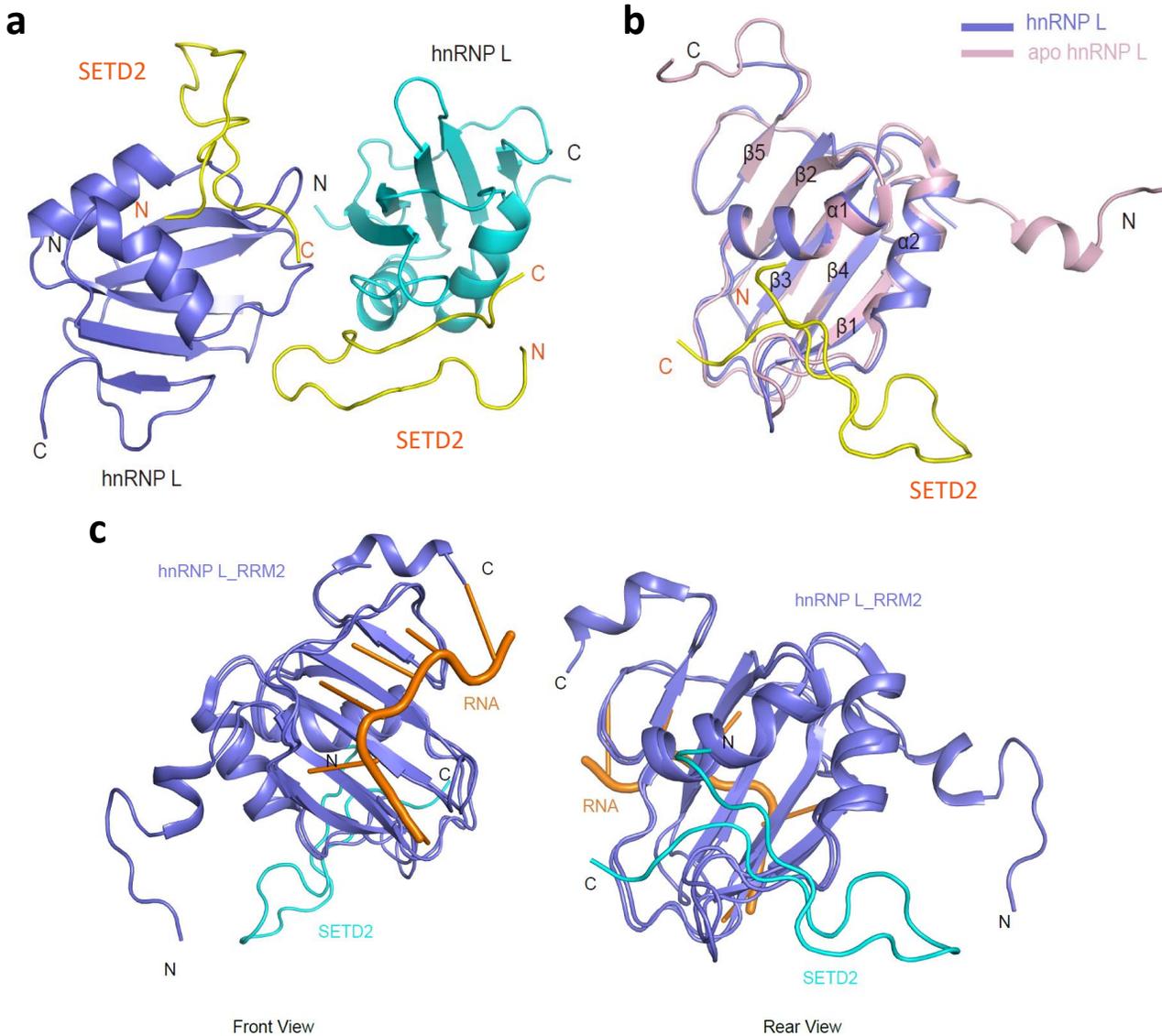
Figure 6

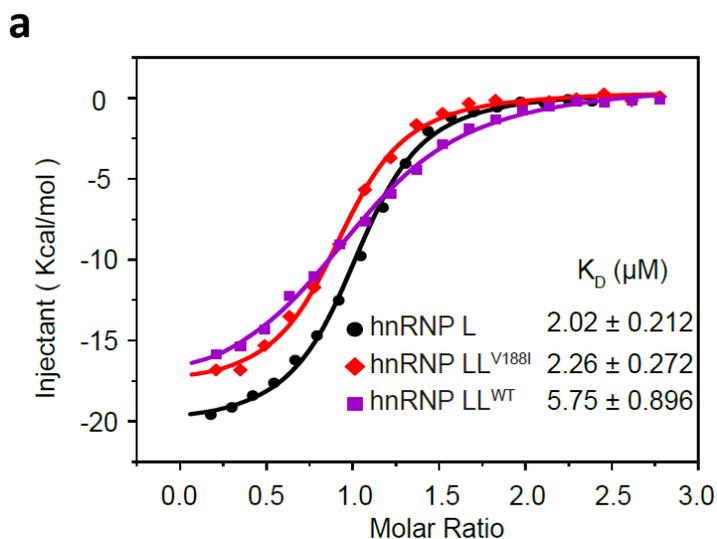
Figure 7



Structural basis of the interaction between SETD2 methyltransferase
and hnRNP L paralogs for governing co-transcriptional splicing



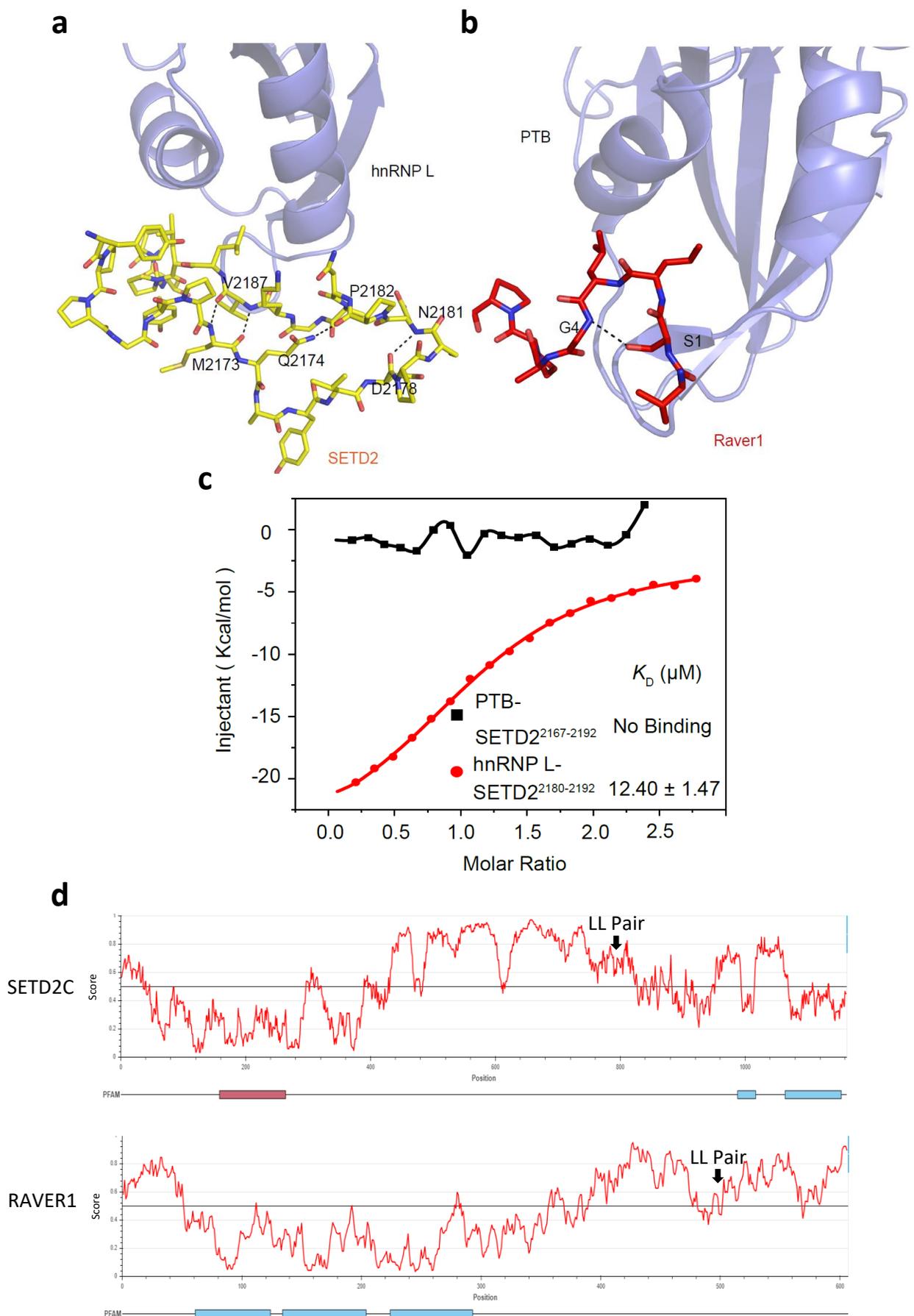
Supplementary Figure 1: (a) Two highly similar hnRNP L RRM2 chains A (purple) and C (Cyan) in complex with SETD2²¹⁶⁷⁻²¹⁹² chains B and D (yellow), respectively. (b) Alignment of hnRNP L to apo hnRNP L reveals no significant structural changes upon SETD2²¹⁶⁷⁻²¹⁹² binding. (c) hnRNP L RRM2 in complex with RNA and SETD2 SETD2²¹⁶⁷⁻²¹⁹².



b

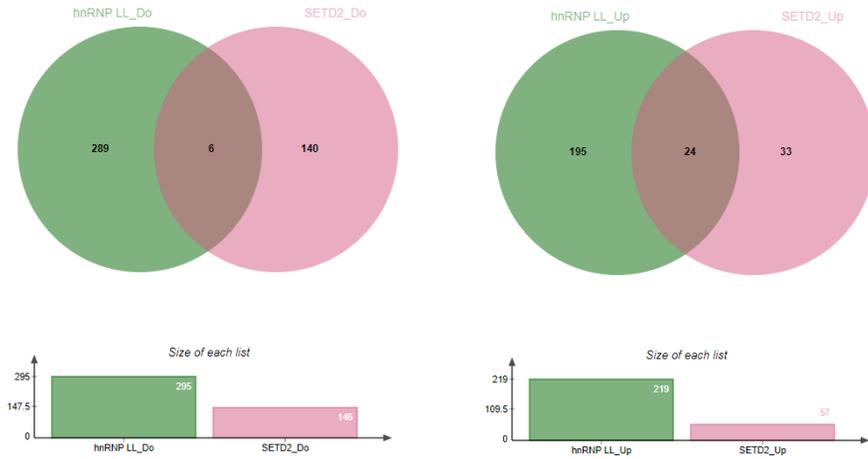
| | | |
|------|---|-----|
| RRM4 | -----RIQHPSNVLHFFN-APLEVTEENFFE I CDELGVKRPSSVK | 39 |
| RRM2 | -----RSVNSVLLFTILNPI Y SITTD V LYT I CNPCGPVQR--IV | 37 |
| RRM1 | -----SPVVHIRGLIDGVVEADLV-EALQEFGPISY--VV | 32 |
| RRM3 | PSRYGPQYGHPPPPPPPEYGPADS PVL ^M VYGLDQSKMNC ^D R ^V FN ^V FCLYGNVEK--VK | 58 |
| | * :: : : . * : | |
| RRM4 | VFSGKSERSSSGLEWESKSDALET L GF L NHYQMKNP NGPY TL L CF----- | 88 |
| RRM2 | I FRKN---GVQAMVEFDSVQSAQRAK A S L NGA----- D I Y SGCCT L KIE---YAKPTRL | 85 |
| RRM1 | VMPKK---RQALVEFEDVLGACNAVNYAADN-----QIYIA----- | 65 |
| RRM3 | FMKSK---PGAAMVEMADGYAVDRAI T LNNN-----FMFGQKLNVCVSKQPAIMPGQS | 109 |
| | .: : .:.* . .: : . | |
| RRM4 | ----- | 88 |
| RRM2 | NVEK-NDQDTWDYTNP--- | 100 |
| RRM1 | -----GHPAFVNYSTSQKI | 79 |
| RRM3 | YGLEDGSCSYKDFSES--- | 125 |

Supplementary Figure 2: (a) ITC fitting curves of hnRNP LWT (black), hnRNP LLWT (purple), hnRNP LLV188I (red), titrated with SETD2²¹⁶⁷⁻²¹⁹² peptide. (b) Sequence alignment of the RRM of hnRNP L. Residues contributing to hydrophobic interactions and hydrogen bonding with SETD2 are highlighted in red and green, respectively.



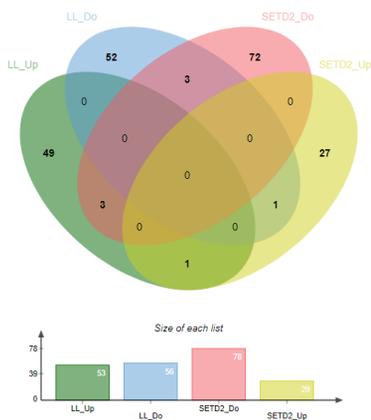
Supplementary Figure 3: (a, b) Intramolecular interaction comparison of hnRNP L-SETD2²¹⁶⁷⁻²¹⁹² and PTB RRM2-PR13. Hydrogen bonds are shown as black dashes. (c) ITC fitting curves of PTB RRM2 with SETD2²¹⁶⁷⁻²¹⁹² (black), hnRNP L with SETD2²¹⁸⁰⁻²¹⁹² (red). (d) IUPRED2 analysis showing disordered regions of SETD2C (1404-2564) and Raver1.

a Differential Expression

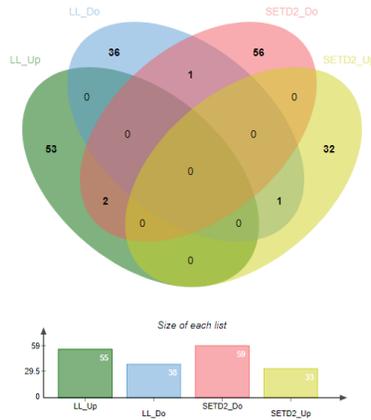


b Differential AS

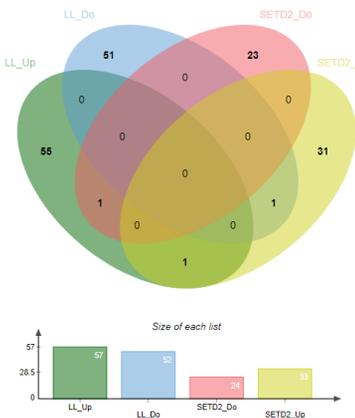
A3SS



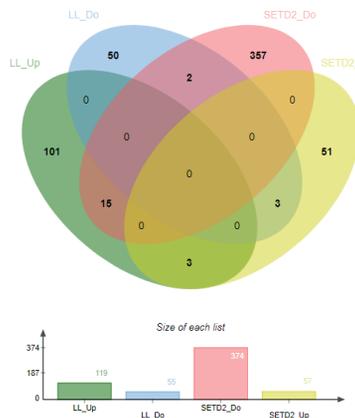
A5SS



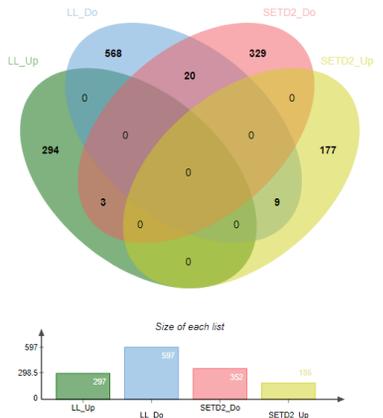
MXE



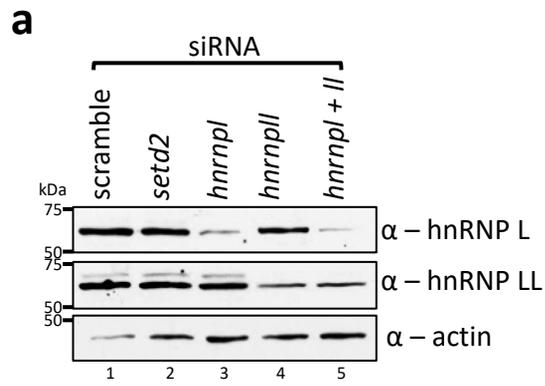
RI



SE

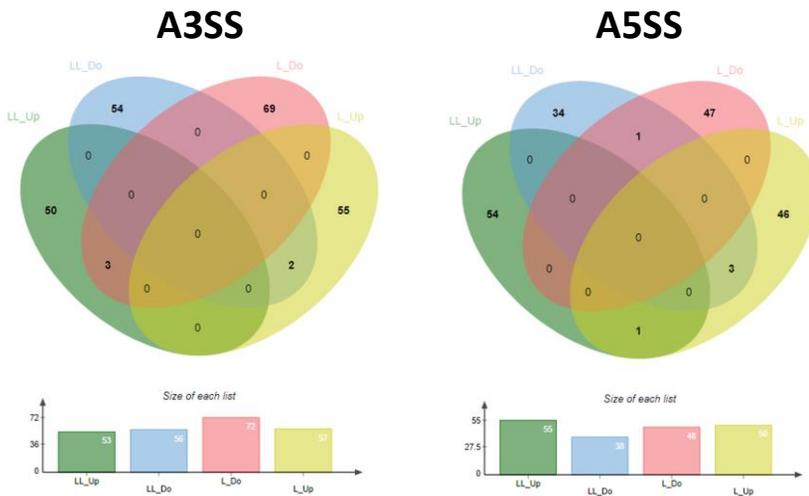


Supplementary Figure 4: (a) Venn diagram showing the overlap between differential gene expression and (b) AS events upon SETD2 and hnRNP LL depletion as compared to scramble siRNA treated cells. The overlap between different type of AS events are shown. A3SS-Alternate 3' Splice Site, A5SS-Alternate 5' Splice Site, MXE-Mutually Exclusive Exons, RI-Retained Intron, SE-Skipped Exon.



b

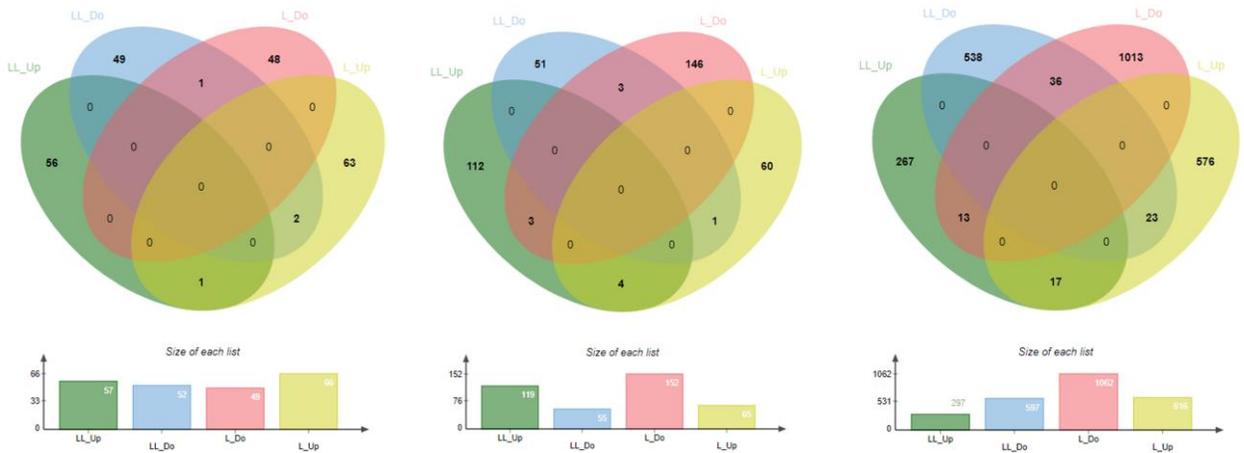
Differential AS



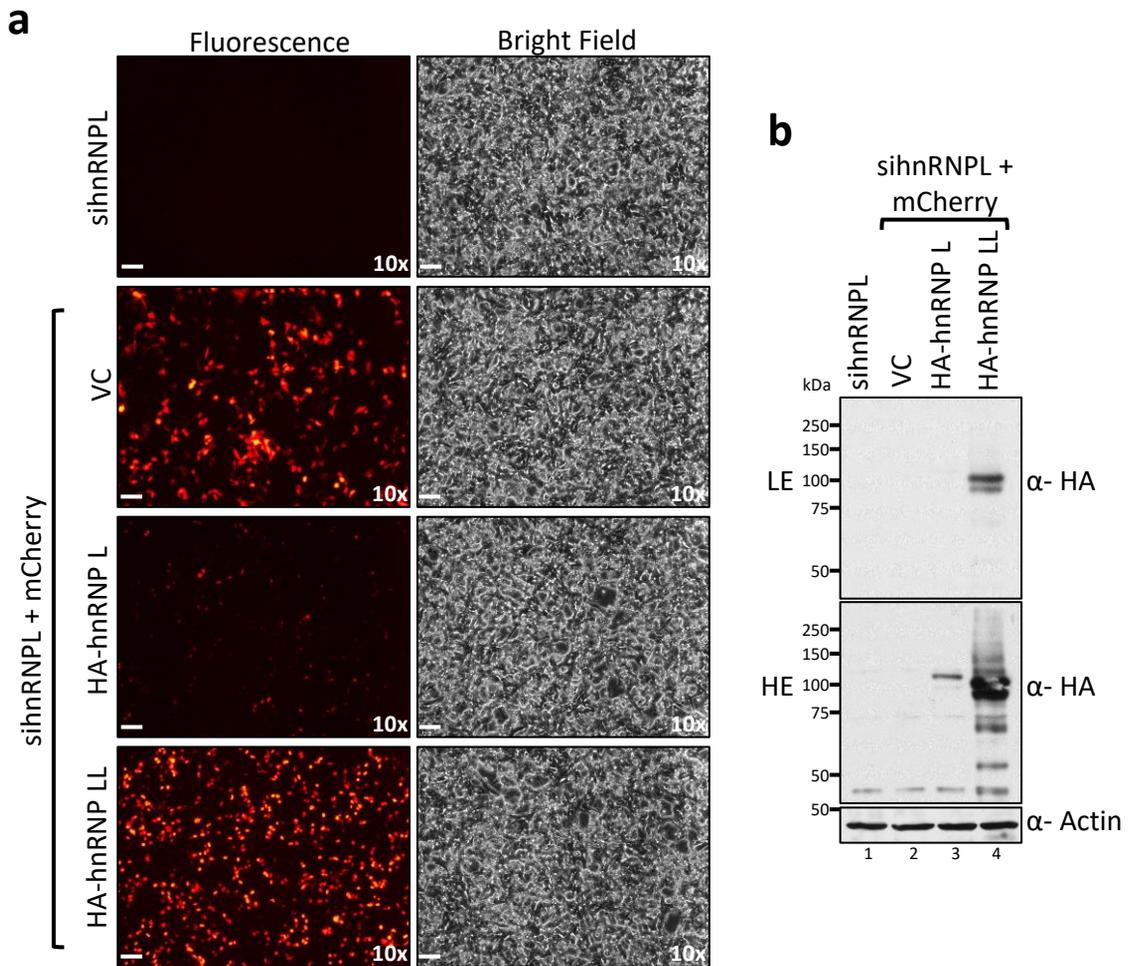
MXE

RI

SE



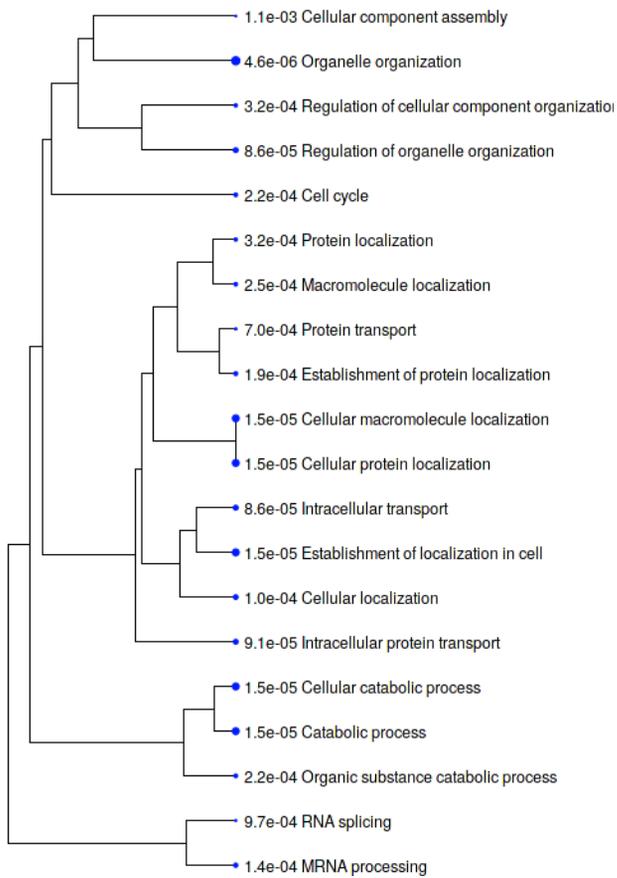
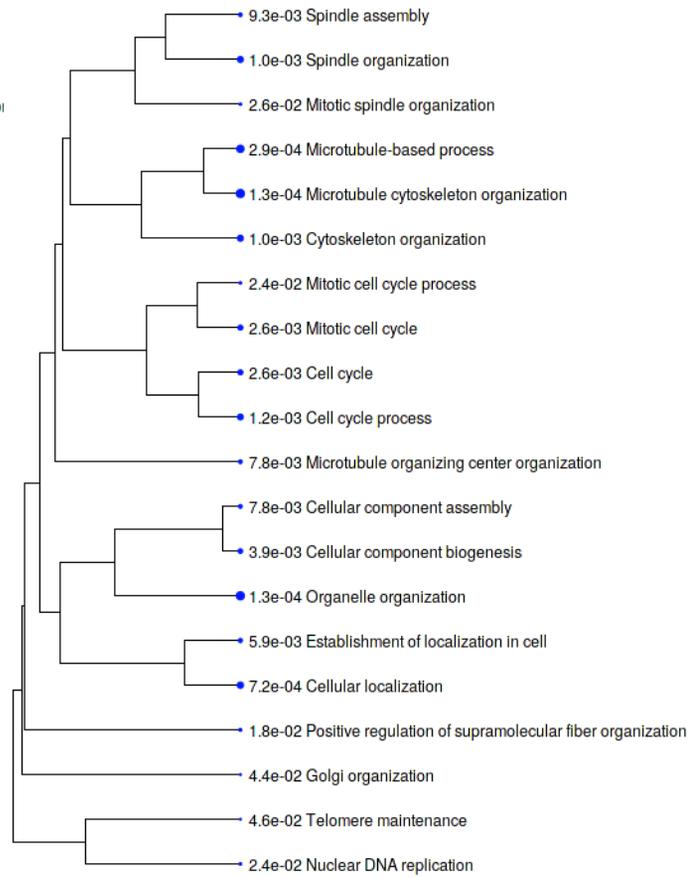
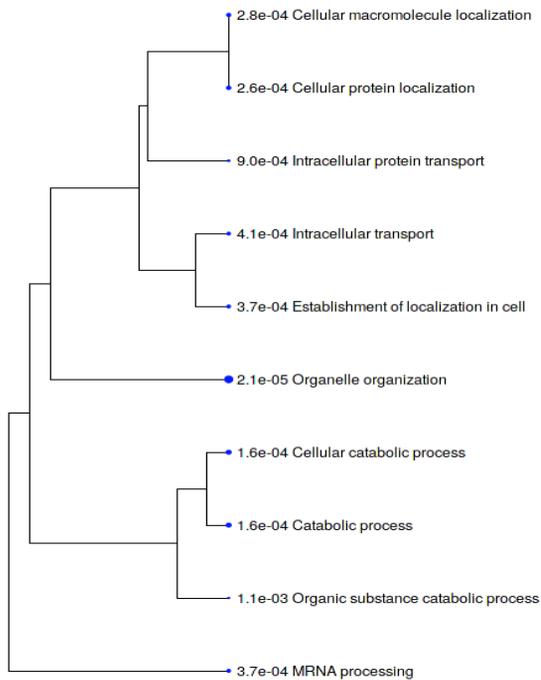
Supplementary Figure 5: (a) Western blotting of whole cell extracts probed with the depicted antibodies. (b) Venn diagram showing the overlap between differential AS events upon hnRNP L and hnRNP LL depletion as compared to scramble siRNA treated cells. The overlap between different type of AS events are shown. A3SS-Alternate 3' Splice Site, A5SS-Alternate 5' Splice Site, MXE-Mutually Exclusive Exons, RI-Retained Intron, SE-Skipped Exon.



Supplementary Figure 6: (a) Microscopy images and (b) western blot showing the expression of mCherry-HA-hnRNP L/LL in sihnRNPL expressing cells. The lower expression of ectopically expressed hnRNP L as compared to LL is because the exogenous hnRNP L is not resistant to sihnRNPL. The scale bar is 1 mm. LE- Lower Exposure, HE- Higher Exposure. VC – Vector Control.

| Protein | Cancer Type | Protein Change | Number of Samples |
|---------|---------------------------------------|----------------|-------------------|
| SETD2 | Uterine Endometrioid Carcinoma | P2192H | 1555 |
| hnRNP L | Oligoastrocytoma | Y257C | 27 |
| hnRNP L | Head and Neck Squamous Cell Carcinoma | I214V | 150 |

Supplementary Figure 7: Mutations in residues of SETD2 and hnRNP L found in cancer according to <https://www.cbioportal.org/> that are important for SETD2-hnRNP L interaction.

a**b****c**

Supplementary Figure 8: GO-term analysis using ShinyGo of differential AS events (a) upon hnRNP L depletion, (b) upon hnRNP LL depletion, and (c) unique to hnRNP L depletion as compared to hnRNP LL depletion.

| Oligo | Sequence (5'-3') |
|-------------|--------------------------|
| hnRNPL_F | TTCTGCTTATATGGCAATGTGG |
| hnRNPL_R | GACTGACCAGGCATGATGG |
| GAPDH_F | TTCGACAGTCAGCCGCATCTTCTT |
| GAPDH_R | CAGGCGCCCAATACGACCAAATC |
| hnRNPLL_F | AAAAGGATCACTCGGCCAGG |
| hnRNPLL_R | TTGTCTCTGGCGACCCTTTC |
| TJP1_e21_F | TGAAGGTATCAGCGGAGGGA |
| TJP1_e21_R | TCACGCAGTTACGAGCAAGT |
| TJP1_e20_F | GAGCTGCCTCAGTACTTGGT |
| TJP1_e20_R | TCCAGTCCCTTACCTTTTCGC |
| BPTF_e18_F | GCCGACAGTGATTGCAACTT |
| BPTF_e18_R | AACTGAGCCACTTATGGGGG |
| BPTF_e18a_F | AGGCAATCCAGTCACTATGGC |
| BPTF_e18a_R | CCTTGTGTTGCTGGCACATTT |

Supplementary Figure 9: Sequence of oligos used to perform RT-PCR.