Title: Mass spectrometry and CLIP-seq analysis reveal BCL11b interactions with RNA processing pathways.

Haitham Sobhy[1]*, Marco De Rovere[1]*, Amina Ait-Ammar[1,2]*, Clementine Wallet[1], Fadoua Daouad[1], Carine Van Lint[2&], Christian Schwartz[1&] and Olivier Rohr[1&]

[1] University of Strasbourg, EA 7292, DHPI, IUT Louis Pasteur, France

[2] Université Libre de Bruxelles, ULB, Gosselies, Belgium

* HS, MDR and AAA can be considered as equal contributors

& CVL, CS and OR can be considered as equal contributors

Correspondence should be addressed to Haitham Sobhy (haithamsobhy@gmail.com) and Olivier ROHR (olivier.rohr@unistra.fr)

Conflict of interest: Not known

Author contribution: HS conceived and designed bioinformatics analysis; HS, MDR and AAA conceived and performed experiments; CW and FD provided scientific and logistic supports; CVL, CS and OR conceived the experiments, supervised the work and supported with resources. HS drafted the manuscript, all authors agreed for the final version of the manuscript.

Data deposited on SRA database with ID PRJNA661202

Abstract: 168 words

Text: 4076 words (4832 words with legend)

Figures: 4

Abstract (168 words)

Although BCL11b (B-cell lymphoma/leukemia 11B, CTIP2) is a well-known transcription repressor and tumor suppressor, its functions and cellular pathways are largely unknown. Here, we show that BCL11b interacts with RNA splicing/processing and nonsense-mediated decay (NMD) proteins, including FUS, SMN1, UPF1 and Drosha. Mass spectrometry analysis (LC-MS/MS) shows that BCL11b interacts with histones, polymerases, and chromatin remodeling (CHD, SWI/SNF, and topoisomerase) proteins. BCL11b-bound RNAs were UV cross-linked and sequenced (CLIP-seq) showing that BCL11b binds to coding and noncoding RNAs (ncRNAs). Surprisingly, RNA transcripts and proteins produced by the same genes like FUS, ESWR1, CHD and Tubulin, were found bound to BCL11b. Deeper analysis of the CLIP-seq data further suggested that BCL11b binds to nonsense mediated RNA decay and retained introns transcripts. Our study is the first genome-wide study of BCL11b-protein and BCL11b-RNA interactants. Our results suggest that the functions of BCL11b are not restricted to the regulation of gene transcription. BCL11b may also control physiologic and physiopathologic pathways by direct bindings to protein complexes, coding RNA and non-coding RNA.

Importance

First genome-wide BCL11b-proteins interactomics

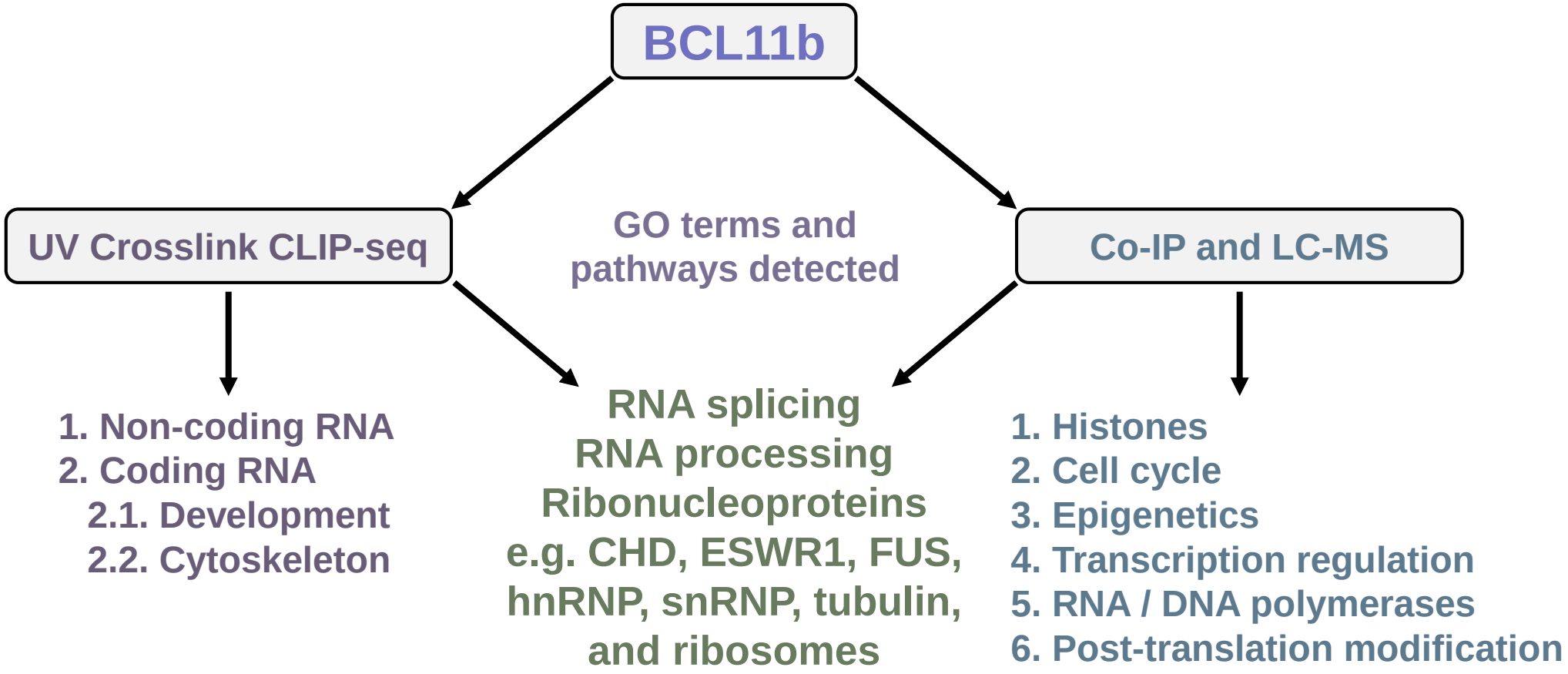
First genome-wide BCL11b-RNA interactomics

BCL11b interacts with RNA processing and RNA splicing proteins

BCL11b interacts with neurodegenerative genes and sarcoma genes

Keywords

Transcription factor BCL, zinc figure family, RNA processing, RNA splicing, gene expression regulation, T-cell development, neuron cell development,


BCL11b
UV Crosslink CLIP-seq
GO terms and pathways detected
Co-IP and LC-MS
1. Non-coding RNA
2. Coding RNA
2.1. Development
2.2. Cytoskeleton
RNA splicing
RNA processing
Ribonucleoproteins
e.g. CHD, ESWR1, FUS, hnRNP, snRNP, tubulin, and ribosomes
1. Histones
2. Cell cycle
3. Epigenetics
4. Transcription regulation
5. RNA / DNA polymerases
6. Post-translation modification

Introductions

The transcription regulator B-cell lymphoma/leukemia 11B (Bcl11b, also known as COUP-TF-interacting protein 2, CTIP2) plays crucial roles in the epigenetic regulation of gene transcription and specifically in the control of the elongation process by interacting with an inactive form of the p-TEFb complex (7SK RNA, CDK9, cyclin T1, and HEXIM1) (Cherrier et al., 2013). BCL11b, which harbors six zinc finger domains (ZnFs) of C2H2-type, shares >60% homology with the human BCL11a, as well as the mouse, chicken, and Xenopus homologues. BCL11 proteins were isolated from T-cell lines derived from patients with T-cell leukemia, suggesting the role of BCL11 proteins in blood cell development and lymphomagenesis, reviewed in (Fu et al., 2017; Lennon et al., 2016, 2017). Beside, BCL11b was deleted in γ-ray induced mouse lymphomas, suggesting a role as a tumor suppressor and modulator of radiation-induced DNA damages, reviewed in (Kominami, 2012).

Studying the function of BCL11b by knocking out the gene (loss-of-function) is challenging (Longabaugh et al., 2017). The mice that lack the two alleles of BCL11b die shortly after birth exhibiting defects in multiple tissues, including immune system, central nervous system (CNS), skin, hair cells in cochlea, teeth, and thymus among other organs (Kominami, 2012) , which suggests the importance of BCL11b during the development. As consequences of absence of the BCL11b-deficient cells, studying the functions and pathways triggered by BCL11b is challenging.

In this study, we used co-immunoprecipitation of proteins followed by mass spectrometry (LC-MS/MS, or MS for short) as well as immunoprecipitation of cross-linked RNA (CLIP-seq), to identify the proteins and RNAs bound to BCL11b respectively. Our results reveal that the partners of BCL11b are involved in cellular pathways dedicated to neuron development, the control of the cell cycle, RNA splicing and neurodegenerative diseases.

Results

Protein interactomics and pathways of BCL11b

To characterize the protein interactants and pathways triggered by BCL11b, we used HEK cells overexpressing Flag-BCL11b. After proteins pull-down using anti-flag antibodies, we performed quantitative LC-MS/MS. After statistical analysis, we identified 629 BCL11b protein partners (supplementary method file S1 and supplementary dataset S1). Among the interacting proteins, we can confirm the previously described interactions between BCL11b and the P-TEFb complex (including CDK9 and CycT1) (Cherrier et al., 2013), HDACs (Marban et al., 2007), HMGA1 (Eilebrecht et al., 2014), HP1 proteins (Rohr et al., 2003), KU proteins (Shadrina et al., 2020), and DCAF1 (Forouzanfar et al., 2019) as a validation of our experimental approach (Figure 1A).

BCL11b interacts with proteins dedicated to genetic and epigenetic regulations

Performing gene ontology (GO) analysis (figure 1B and supplementary data S2), we observed that BCL11b interacts with proteins and complexes that are involved in cellular genetics, DNA replication and regulations of mitosis (Figure 1B), including proteins that are involved in the regulation of gene transcription (CCNT1, CDK9, SMARCAs and SMARCCs), and epigenetic regulators, including HDACs, CHDs, EHMTs, KMT2A (Figure 1B). Indeed, BCL11b binds to more than 20% of the total protein components of H2AX, NuRD and SNW1 complexes, among other proteins involved in chromatin organization, and DNA recombination (Figure 1C). Additionally, we observed BCL11b is associated with proteins involved in DNA repair and cellular response to stimuli (such as BRCA2, MCMs, PARP1, and SMCs proteins). In agreement to our results, BCL11b was previously shown to interact with Ku70 and Ku80 proteins, which have an important role in the non-homologous end-joining DNA repair pathway (Shadrina et al., 2020).

BCL11b interacts with structural proteins and the nuclear pore complex

Among our novel findings, we observed BCL11b interacts with structural proteins involved in cytoskeleton, and motor activity, such as multiple isoforms of tubulin (supplementary data S1 and S2). Another significant GO term is the nucleocytoplasmic shuttling pathways, which include transport of the RNA or ribonucleoprotein complexes, such as NUP133, NUP153, NUP93 and NUP214. Noteworthy, nucleoporins could have molecular roles beyond the nucleocytoplasmic shuttling, such as regulations of gene transcription and chromatin organization (Sun et al., 2019).

BCL11b interacts with proteins involved in RNA translation and RNA splicing

The role of BCL11b during RNA processing have not been described by any previous research. We found that multiple proteins bound to BCL11b are involved in RNA processing pathways such as translation of RNA and ribosome assembly, such as 40S and 60S ribosomal subunits (figure 1C). Moreover, BCL11b interacts with RNA-binding proteins, including RNA splicing and ribonucleoproteins (RNPs), such as spliceosomes. Among these complexes, the survival motor neuron (SMN) protein complex, which is made by binding of SMN protein with other proteins so-called GEMIN and Sm proteins (Matera and Wang, 2014). It is known that SMN complex assembles in nucleus before it is translocalized to the cytoplasm. SMN complex is required for assembly of RNP complexes, which are the building block of spliceosomes. Other RNPs that are involved in post-transcriptional RNA regulation, include small nuclear RNP (snRNPs), heterogeneous nuclear RNP (hnRNP), serine/arginine (SR)-rich proteins, and small nuclear RNA (snRNAs). BCL11b interacts to 22 different hnRNPs or snRNPs, and about 17% of proteins that constitute the spliceosome (figure 1B-1D).

On the other hand, FUS is another crucial splicing factor, which is involved in regulations of RNA splicing, transport of RNA, and regulation of gene expression as response to DNA damage

(Chen et al., 2019). FUS can accumulate near the transcription start sites (TSS) and binds to transcription initiation factors to inhibit phosphorylation of the RNA polymerase 2 C-terminal domain (POLR2-CTD) (Chen et al., 2019). We performed co-IP and immunoblot to confirm that BCL11b interacts with SMN1 and FUS, which are among the hallmarks of RNA splicing processes (Figure 1E and 1F). Our IP-MS and IP-immunoblot confirm that BCL11b is able to interact with SMN1 and FUS proteins, and therefore BCL11b could have a role in regulation of RNA splicing and translation of RNA.

BCL11b interacts with the Drosha and the DGCR8 protein complexes

FUS is a member of the Drosha and the DGCR8 complexes, which are able to regulate gene expression at transcriptional and post-transcriptional levels (Pong and Gullerova, 2018). On the other hand, Ago is the key protein in RNA interference (RNAi) pathway. Ago2 can bind to Dicer or TRBP proteins forming TRBP containing complex. Drosha and TRBP complexes can regulate translation of mRNA, and degradation of mRNA of certain transcripts (Pong and Gullerova, 2018). Here, by IP-MS experiment, we found that BCL11b is associated with 15 of the 20 proteins of the Drosha complex and 10 proteins of the 11 proteins of the DGCR8 complex (Figure 1D, and 1E), in addition to 20 proteins of TRBP complex, including Ago2. To confirm these results we performed co-IP experiment, which confirms the association of BCL11b with FUS and Drosha (Figure 1F, 1G and 1H) and further suggest that BCL11b may coordinate with FUS and Drosha to regulate gene expression at transcriptional and post-transcriptional levels and to control mRNA translation and stability.

BCL11b binds to protein from the nonsense-mediated mRNA decay (NMD) pathway

NMD pathway is crucial to decay mRNA transcripts with premature stop codons (Kurosaki et al., 2019). These transcripts lead to gain- or loss-of-function of some genes, which could be deleterious during the development. UPF1 is the hallmark of the NMD pathway.

Among the significant IP-MS hits, we observed interactions between BCL11b and proteins involved in the NMD pathway. Out of 120 proteins identified in this specific GO molecular pathway, 61 were found bound to BCL11b (figure 1C). Co-IP experiments confirmed that BCL11b associates with UPF1, which is the hallmark of NMD pathway, and thereby this confirms that BCL11b interacts with the protein complex dedicated to NMD (Figure 1I). Again, these results show that BCL11b is involved in the regulation of gene expression at the post-transcriptional level (Figure 1F-1H).

BCL11b is associated with cellular RNA

Although BCL11b has six C2H2 ZnF domains, to our knowledge the RNA regulation activities has not been described for any of the two BCL11 proteins, However, BCL11b can bind the 7SK ncRNA to regulate transcription elongation (Cherrier et al., 2013; Eilebrecht et al., 2014). Our MS proteomics data show that BCL11b binds to RNA-binding proteins in complexes dedicated to RNA processing and RNA splicing. To further study the RNA-binding activity and obtain genome-wide landscape of BCL11b-bound RNAs, we cross-linked the BCL11b protein to cellular RNAs using UV radiation before immunoprecipitation and sequencing (CLIP-seq). These experiments have been performed by targeting the endogenous BCL11b in microglial cells, and an overexpressed Flag-BCL11b in HEK293 cells. We analyzed the results by two approaches, as the following:

First, number of CLIP reads mapped to a gene feature were counted and the genes that are enriched over the control samples were selected (see method in Suppl. file S1), which lead to a dataset of 740 different genes (figure 2, 3A, S1-S5 and Suppl. data S3). We found that about 96% of the BCL11b-bound RNAs are mapped to protein-coding genes (Table S1). BCL11b tends to bind to transcripts of long genes. Among the genes that have high numbers of CLIP reads, we found that about 20% are over 100kb in length, whereas the average length of these genes is 61kb, which is two-fold the genomics length (~30kb). The longest gene is Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 1 (AGAP1, NCBI GeneID:

116987), which is 637,753 bp in length. In agreement with previous studies focusing on BCL11b control of cellular gene expression (Kominami, 2012; Wakabayashi et al., 2003), we found that the GO terms of BCL11b-bound RNA include genes from neuronal development and neurogenesis of multiple parts of brain (e.g. midbrain, hindbrain and substantia nigra). BCL11b binds to RNA of proteins that have been described to regulate proliferation, regeneration and migration of neuronal and non-neuronal cells, including glial cells and Schwann cells (details are found in figures 3B and S6, Suppl. data 4). However, BCL11b was also found associated with RNA from genes involved in the development of the immune system, hemopoiesis, blood vessels, and hair follicles. Finally, we found RNA coding for actin and tubulin associated with BCL11b as well as RNA coding for proteins of the major and the minor spliceosome (Figure 1E, Table S2, Suppl. data 5).

As a second analysis approach, we divide the whole genome into 10kb regions (bins) and we count the number of reads corresponding to each bin (see method in Suppl. file S1). For additional confirmation, we performed peak calling to cluster the RNAs into bins. We then selected the regions that are enriched over the control. The results show that over 40% of CLIP reads are mapped within 0-50kb down-stream of the transcription start sites (TSS) (Figure S7, Suppl. data S6, S7). Noting that some these RNAs mapped to long genes, we concluded that about half of the CLIP reads are within the genes and at least one third could be located in the intergenic regions. Aligning the sequences, we found two conserved motifs (CUCRGCCU and UCCCAGCW) in BCL11b-bound RNA regions (figure 3C). Gene ontology analysis for the genes localized in these regions suggested their involvement into structural process, extracellular matrix organization, and response to stress (Suppl. data S6, S7). The gene products are involved in angiogenesis, p53 and inflammation pathways, cytoskeletal regulation and multiple diseases such as cancers, Huntington's disease and Alzheimer (Figure S8, Suppl. data S6, S7).

Taken together, our results demonstrate that half of the BCL11b-binding RNAs are transcribed from long genes and mapped to the genic regions further suggesting an impact of BCL11b on protein expression or function at post-transcriptional levels.

BCL11b binds to RNA transcripts and proteins product of the same gene

We found BCL11b associated with 137 RNA transcripts and proteins encoded by the same gene (Figure 3A, and Suppl. data S8). Among them, we found EWSR1, FUS, SPFQ, UPF1, tubulins, CHDs and hnRNPs transcripts and proteins. Interestingly, GO analysis reveals that these proteins are involved in RNA processing, NMD pathway and translation (Figure 3D, Suppl. data S8, S9). Interestingly, tubulin was reported to use a unique mechanism to auto-regulate its expression level within the cell to ensure a stable production of the protein in cells, and over-expression of tubulin gene does not lead to increase its protein expression level (Gasic and Mitchison, 2019). The mechanism of tubulin auto-regulation is largely unknown. However, it thought that tubulins α and β could polymerize through an unknown mediator, which then bind to the nascent tubulin peptide (Gasic and Mitchison, 2019). By unknown mechanism, the mediator can bind to ribosomes and/or ribonucleases to terminate the translation. We found that BCL11b interacts with multiple isoforms of tubulin proteins, tubulin RNAs, as well as ribonucleases, and ribosomes (Figure 3E). In addition, we have noted that BCL11b binds to multiple members of the same protein family, such as tubulins, actins, RNPs, or ribosomes. Again, these results may support the involvement of BCL11b in the regulation of protein expressions at post-transcriptional levels.

BCL11b binds differentially transcripts of the same genes

To further study the association of BCL11b with cellular RNA, we counted the number of reads per exon for CLIP reads that mapped to genes. We found that more than 75% of the CLIP reads were mapped to exons (Figure 3F). Having a deeper look to the CLIP-seq reads on

genome browsers, we found that when reads were mapped to intronic regions of genes such as EWSR1, FUS, and SRRM2, the locations of the reads corresponded to the coordinates of RNA isoforms that do not code for proteins (we refer to them here as INCP), such as retained introns (IR), nonsense mediated decay (NMD), non-stop decay or processed transcript, (Figure 2, Suppl. Figure S2-5). This result is consistent with our results demonstrating that BCL11b interacts with proteins involved in RNA splicing and NMD pathways, including FUS, SMN1, RNPs, Drosha, and UPF1. These results show that BCL11b can regulate gene expression at post-transcriptional levels (Figure 1F-1I).

To strengthen this finding, we collected the transcripts (or RNA isoforms) of 23 genes that encode for isoforms coding for proteins (ICPs), and isoforms that do not code for proteins (INCPs), such as NMD and IR. We then counted the number of reads that correspond to each isoform. We found that BCL11b binds to different isoforms of the same gene in differential manner (Figure 3G, Suppl. data S10). For example, TOP3A gene encodes for 21 RNA isoforms, similarly, ESWR1, Drosha, SIRT7 and RAD52 genes encode for 20-21 RNA isoforms. The fold increase to control of CLIP-seq for TOP3A ranges from 1.1 to 2.7 folds, whereas, the binding to RNA of RAD52 is up to 7-folds increase (log2 fold change ranges from 0.8 to 2.8 folds) (Figure 3G). The same concept applies to SMN1 (10 isoforms), COL27A1 (8 isoforms) and RAD51 (11 isoforms). We counted the CLIP-seq reads per each transcript (or isoform) and then normalized to the length of the transcripts. We found that isoforms coding for protein (ICPs) have 2 folds CLIP reads than INCPs (Figure 3H). The Spearman correlation rank (R) between the CLIP reads and transcripts length for ICPs is 0.51 and 0.44 for INCP, which suggest that BCL11b tends to bind to long isoforms of ICPs and less degree INCPs. Together these results suggest that BCL11b preferentially binds to long RNA isoforms that code for proteins.

We still noticed some CLIP reads that correspond to intronic regions within ICPs, but the CLIP reads are mapped to the coordinates of exons in other INCP isoforms. Noteworthy, NMD or IR transcripts introduce additional exons instead of introns. To explain this concept,

we highlighted the genomics introns in pink colors, see (Figure 2 and Figures S2-S5). Although these intron regions are usually removed in case of ICP transcripts (thick dark bars in e panel in Figure 2,), we found that part of these regions corresponded to exons in INCPs (light blue bars in e panel in Figure 2). Computationally, we counted number of reads mapped to exons and introns, we normalized them to the number of exons or introns as shown in Figure 3I and Suppl. data S10. If BCL11b binds to exons regions of RNA, we expect to find that most of the CLIP reads are biased to exons coordinates. However, in Figure 3I, we observed that most of the reads are mapped to introns. In case of INCPs, the number of reads were almost equally distributed throughout the exons and introns. Statistically, we calculated the Spearman correlation rank (R) between number of reads and number of exons, which show that number of exons correlate with number of CLIP reads, in case of ICPs, but not in case of INCPs (R of INCPs=0.2, and R of ICPs=0.46). This result suggests that INCPs with few numbers of exons have multiple CLIP reads. The DNA (cytosine-5-)-methyltransferase 1 (DNMT1) transcripts are good examples that support this observation. Number of reads mapped to IR transcripts are 2-3 fold higher than the number of reads mapped to INCPs transcripts. Indeed, the 5 exons IR transcripts (ENST00000591764, with 589 reads) have 3-folds more reads than the 9 exons protein coding transcripts (ENST00000588952, with 196 reads) (Figure 3J, Suppl. data S10).

Taken together, Our CLIP-seq results confirm that BCL11b binds to complexes with ncRNA and ribonucleoproteins (Figure 2 and S2-S5). The results support our proteomics findings that BCL11b interacts with RNA processing and RNA splicing proteins, such as FUS, SMN1, Drosha complex and NMD complex (Figure 1). Our results further suggest that BCL11b may be able to differentiate between the types of RNA isoforms to contribute to the selection of the one to be translated. Together, BCL11b could interact with other proteins, such as FUS or UPF1 to regulate mRNA translation of selected genes.

Involvement of BCL11b- associated factors in diseases

Described as a tumor suppressor gene, BCL11b expression level was reported to be modulated in T-Cell Lymphoma (Fu et al., 2017; Kominami, 2012). Previous reports studying Ewing Sarcoma (Orth et al., 2020; Wiles et al., 2013), spinal muscular atrophy (SMA) (Doktor et al., 2016; Maeda et al., 2014), and Parkinson's disease (PD) (Scherzer et al., 2007) showed differential expression of BCL11b. In addition, since BCL11b was reported to be involved in development, it is speculated that BCL11b could have roles in neurodegenerative diseases (Lennon et al., 2017), particularly in motor neuron amyotrophic lateral sclerosis (ALS) disease (Lennon et al., 2016), and Huntington's disease (Ahmed et al., 2015). However, the role of BCL11b in these disease was not clear, because of the lack of studies related to protein and RNA interactions of BCL11b nor its functional complexes, it was challenging to understand the exact role of BCL11b in these diseases.

As we discussed above, using IP-MS, and CLIP-seq experiments, we detected interactions between BCL11b and multiple proteins that are involved in diseases such as: EWSR1, which is known to contribute to Ewing's sarcoma (Wiles et al., 2013); SMN1, which is involved in SMA (Maeda et al., 2014); FUS, TDP-43 and TAF15 proteins that are involved in ALS (Yamazaki et al., 2012); SRRM2, which is a key factor causing Parkinson's disease (Shehadeh et al., 2010) and 7SK, MALAT1, NEAT1 and TUG1 ncRNAs, SFPQ protein, RNPs and spliceosomes proteins that are known to contribute to cancer, as well as the motor neuron SMA and ALS diseases (Bhat et al., 2020; Briese et al., 2018; Wu and Kuo, 2020). Noteworthy, FUS associates with SMN, RNPs and RNA-binding proteins to form a large complex (De Santis et al., 2017; Yamazaki et al., 2012). We found that one third of these proteins were found bound to BCL11b by LC-MS, which suggest the ability of BCL11b to form complexes with FUS or SMN1 (table S2, S3). On the other hand, DNA damage pathway may regulate RNA splicing process by selecting specific RNA isoforms to be translated (Shkreta and Chabot, 2015). FUS, Drosha and Ago2 are examples of regulators that are recruited in response to DNA damage to regulate gene expression of certain protein isoforms (Chen et al., 2019; Pong and Gullerova, 2018).

Our CLIP-seq and RNA IP results confirmed the binding of BCL11b to the RNAs coding for FUS, SFPQ and SRRM2 genes and specifically to exon and exon-intron junctions regions (Figure 2, 4B, S1-S3, S9 and S10, table S4, Suppl. data S10). Noteworthy, we could not observe significant increase of the gene expression of RNA encoded from the same regions (using RT-qPCR) after over-expression of BCL11b (Figure 4C). These results suggest that BCL1b could contribute to isoform selection by binding to RNA isoforms, without impacting the global level of the transcripts.

Discussion and Conclusion

This study describes for the first time BCL11b-associated proteome and RNAome. To our knowledge this is the first time to report a protein to bind to RNA transcript and protein encoded from the same gene, such as FUS, Figure 1F, 1G, 3 and 4B. Our results suggest that BCL11b is involved in multiple pathways and functions. In addition to its function as transcription regulator, BCL11b binds to proteins involved in cytoskeleton and development, as well as ribonucleic complexes such as ribosomes and RNPs. We confirmed that BCL11b binds to proteins involved in RNA processing, such as RNA splicing and NMD pathway, in addition to the ncRNA that are involved in RNA splicing. Our results show that BCL11b binds to DNA damage repair proteins, to chromatin remodeling and histone-modifying proteins (epigenetic modifications). One major GO term found is the assembly and disassembly of the complexes. We speculate that one of the major roles of BCL11b is to assemble various complexes to perform a function.

There are number of regulators found in close association to chromatin as checkpoint safeguard. As response to cell cycle checkpoints, stimuli, or DNA damage, the safeguard factors regulate the expression level of some specific genes. Here, it emerges the importance of a regulator that can coordinate between replication, transcription and translation (Shkreta and Chabot, 2015). First, we demonstrate that BCL11b binds to RNA processing proteins (such as FUS, Drosha, Ago2, UPF1, and RNA splicing factors, Figure 1F-1I) revealing a role in the

regulation of gene expression in response to DNA damage. Then, we highlighted the potential role of BCL11b in autoregulation of some specific gene translation such as tubulin (Figure 3E).

We found that BCL11b preferentially binds to RNA isoforms coding for proteins (Figure 3F). We also observed multiple MS hits that correspond to 40S and 60S ribosome complexes (Figure 1C). On the other hand, BCL11b was found associated with isoforms that do not code for proteins (INCPs), such as NMD and IR. Generally, translation is regulated by selecting the desired mRNA for being translated, whereas, the RNA decay pathways, such as Drosha, UPF1 (NMD pathway), and Ago (RNA interference) are able to decay or silence the undesirable RNA isoforms (Tuck et al., 2020; Wang and Aifantis, 2020). Here, we showed based on reliable methods (IP, LC-MS, CLIP-seq, and RIP-PCR) that BCL11b contributes to the selection of RNA isoforms. Selecting the isoform (so-called isoform switching) is a crucial process through which cells select a specific RNA isoform during the development or immune response. Some isoforms might be needed at specific steps during development, cell cycle or response to stimuli. Therefore, malfunction of RNA splicing or processing has a great impact on cell proliferation and cancer prognosis (Wang and Aifantis, 2020).

In conclusion, our analysis of BCL11b-protein and BCL11b-RNA interactions provide the first global overview of BCL11b partners. The results suggest that BCL11b may bind to RNA processing proteins to regulate translation and/or select the desirable RNA isoform. One can speculate that depletion of BCL11b could lead to isoform switching and translation of wrong isoform of the protein (misfolded or short) but this needs to be formally demonstrated. Although the exact role of BCL11b in diseases deserve to be studied by future research, our results offer potential genome-wide RNA and proteins datasets, which will help future efforts to understand the molecular basis of cancer and neurodegenerative diseases.

Conflict of interest: no conflict of interest is known

Human or animal experiments: no human or animal work have been performed

Funding:

This work was supported by the French agency for research on AIDS and viral hepatitis (ANRS); SIDACTION; the European Union's Horizon 2020 research and innovation programme under grant agreement No 691119-EU4HIVCURE-H2020-MSCA-RISE-2015; The Belgian Fund for Scientific Research (FRS-FNRS, Belgium), the "Fondation Roi Baudouin", the Walloon Region (Fonds de Maturation) and the university of Brussels (Action de Recherche Concertée (ARC) grant). The laboratory of CVL is part of the ULB-Cancer Research Centre (U-CRC). AAA is a fellow of the Wallonie-Bruxelles International program and of the Marie Skłodowska Curie COFUND action. HS is a fellow of ANRS. MDR is a fellow of University of Strasbourg. CVL is Directeur de Recherches of the F.R.S-FNRS, Belgium.

Figure Legends

Figure 1. BCL11b binds to multiple complexes and pathways as deciphered by MS results. A) These proteins were previously shown to interact with BCL11b and we can detect them by IP-MS method. The gene names of BCL11b-bound ncRNAs are written in blue. B) GO terms of BCL11b-interacting proteins (BIPs), X-axis show percent of BIPs to the total number of proteins identified in this category. C) Four representative GO terms with the list of proteins that are detected to bind to BCL11b. D) The complexes that constitute of BIPs, X-axis show percent of BIPs to the total number of proteins in the complex; the total number of proteins in the complex is written in dark blue outside the bars. E) Four representative complexes with the name of proteins. The BIPs are found in rectangles, BCL11b-interacting ncRNAs are in blue; whereas the non-BIPs are in hexagonal red shapes. The arrows refer to protein-protein interactions. F-I) western blot images show co-IP using antibodies directed to anti-BCL11b, anti-FUS, and anti-Drosha to detect BCL11b, FUS, SMN1, and UPF1. The antibodies (anti-) used to probe western blots are listed at left of the figure. The figures were constructed from Supplementary data files S1 and S2.

Figure 2. Representative CLIP-seq results of three protein-coding genes (EWSR1, FUS, and SRRM2), and non-protein coding gene (NEAT1 ncRNA). The olive green bars refer to the NCBI RefSeq genes. X-axis represents the coordinates of the gene and the chromosome number. The Y-axis represents the number of CLIP reads in each condition. a) Over-expressed BCL11b in HEK cells, b) endogenous BCL11b in microglial cells, whereas, c and d are their control experiments, respectively. e) The black bars represent the GENCODE transcripts of the stated genes, the sky blue transcripts refer to the protein non-coding transcripts that harbor significant CLIP reads and their coordinates correspond to intron of RefSeq gene, these intronic regions are highlighted with light rose colored boxes. The light blue horizontal boxes refer to introns without significant BCL11b CLIP reads. For additional figures and CLIP-seq reads alignments, see Figure S1-S5.

Figure 3. CLIP-seq results of BCL11b-bound RNA. A) Number of BCL11b-interacting proteins, from IP-MS experiment, that passed our cut-off (so-called MS hits); number of genes, which have CLIP reads passed our cut-off (so-called CLIP hits); 137 genes can bind to BCL11b in form of RNA or in form of protein. B) GO terms of the genes that passed CLIP-seq cut-offs. C) The predicted BCL11b-binding motif. D) GO terms of 137 shared genes (BCL11b binds to RNA and protein product). E) Schematic diagram of the hypothesis regarding tubulin autoregulation mechanism. BCL11b binds to multiple tubulin RNA, in brown box, in addition BCL11b binds to multiple ribosomes (oranges box), tubulins (blue box) and ribonucleases (red box). F) Violin shape represents the number of CLIP reads mapped to exons and introns of the significant genes. G) BCL11b binds differentially to the selected 23 genes. Number next to the names are number of transcripts. H) Numbers of CLIP reads normalized to number of exons for isoforms coding for proteins (ICPs) or isoforms do not code for proteins (INCPs). The table shows the Spearman's rank correlation coefficient (R) between numbers of CLIP reads in case of ICPs or INCPs from one side and length or number of exons from the other side. I) Number of CLIP reads mapped to exons and those mapped to introns in case of ICPs or INCPs. J) Number of CLIP reads mapped to exons of DNMT1 gene are counted and normalized to number of exons. The table shows the average number of CLIP reads and number of exons within protein-coding transcripts and non-protein-coding transcripts. The blue colors refer to ICPs, whereas red, green and orange colors refer to nonsense mediated decay, processed transcript, and retained intron transcripts, respectively. The figures 3G-3J were constructed from Supplementary data files S10.

Figure 4. A) The table shows the contribution of BCL11b and other genes in the Parkinson's disease (PD), SMA, ALS and Ewing sarcoma diseases. B) Results of RNA IP-qPCR (RIP-PCR) of three genes, the fold change of RNA bound to endogenous BCL11b to the input control. For the location of the amplicons on the gene, please refer to figure S3, and for PCR primers, Table

S4. C) Gene expression of the amplicons used in G using overexpressed BCL11b plasmid and pCDNA3 vector. For raw Ct values presented in (B) and (C), see Supplementary data S11. D) The graphical abstract shows the major proteins pathways, complexes and based on GO terms obtained after IP-MS and CLIP-seq experiments.

References

Ahmed, I., Sbodio, J.I., Harraz, M.M., Tyagi, R., Grima, J.C., Albacarys, L.K., Hubbi, M.E., Xu, R., Kim, S., Paul, B.D., *et al.* (2015). Huntington's disease: Neural dysfunction linked to inositol polyphosphate multikinase. Proceedings of the National Academy of Sciences *112*, 9751-9756.

Bhat, A.A., Younes, S.N., Raza, S.S., Zarif, L., Nisar, S., Ahmed, I., Mir, R., Kumar, S., Sharawat, S.K., Hashem, S., *et al.* (2020). Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance. Molecular Cancer *19*, 57.

Briese, M., Saal-Bauernschubert, L., Ji, C., Moradi, M., Ghanawi, H., Uhl, M., Appenzeller, S., Backofen, R., and Sendtner, M. (2018). hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons. Proc Natl Acad Sci U S A *115*, E2859-E2868.

Chen, C., Ding, X., Akram, N., Xue, S., and Luo, S.-Z. (2019). Fused in Sarcoma: Properties, Self-Assembly and Correlation with Neurodegenerative Diseases. Molecules *24*, 1622.

Cherrier, T., Le Douce, V., Eilebrecht, S., Riclet, R., Marban, C., Dequiedt, F., Goumon, Y., Paillart, J.C., Mericskay, M., Parlakian, A., *et al.* (2013). CTIP2 is a negative regulator of P-TEFb. Proc Natl Acad Sci U S A *110*, 12655-12660.

De Santis, R., Santini, L., Colantoni, A., Peruzzi, G., de Turris, V., Alfano, V., Bozzoni, I., and Rosa, A. (2017). FUS Mutant Human Motoneurons Display Altered Transcriptome and microRNA Pathways with Implications for ALS Pathogenesis. Stem Cell Reports *9*, 1450-1462.

Doktor, T.K., Hua, Y., Andersen, H.S., Brøner, S., Liu, Y.H., Wieckowska, A., Dembic, M., Bruun, G.H., Krainer, A.R., and Andresen, B.S. (2016). RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. Nucleic Acids Research *45*, 395-416.

Eilebrecht, S., Le Douce, V., Riclet, R., Targat, B., Hallay, H., Van Driessche, B., Schwartz, C., Robette, G., Van Lint, C., Rohr, O., *et al.* (2014). HMGA1 recruits CTIP2-repressed P-TEFb to the HIV-1 and cellular target promoters. Nucleic Acids Res *42*, 4962-4971.

Forouzanfar, F., Ali, S., Wallet, C., De Rovere, M., Ducloy, C., El Mekdad, H., El Maassarani, M., Ait-Ammar, A., Van Assche, J., Boutant, E., *et al.* (2019). HIV-1 Vpr mediates the depletion of the cellular repressor CTIP2 to counteract viral gene silencing. Sci Rep *9*, 13154.

Fu, W., Yi, S., Qiu, L., Sun, J., Tu, P., and Wang, Y. (2017). BCL11B-Mediated Epigenetic Repression Is a Crucial Target for Histone Deacetylase Inhibitors in Cutaneous T-Cell Lymphoma. J Invest Dermatol *137*, 1523-1532.

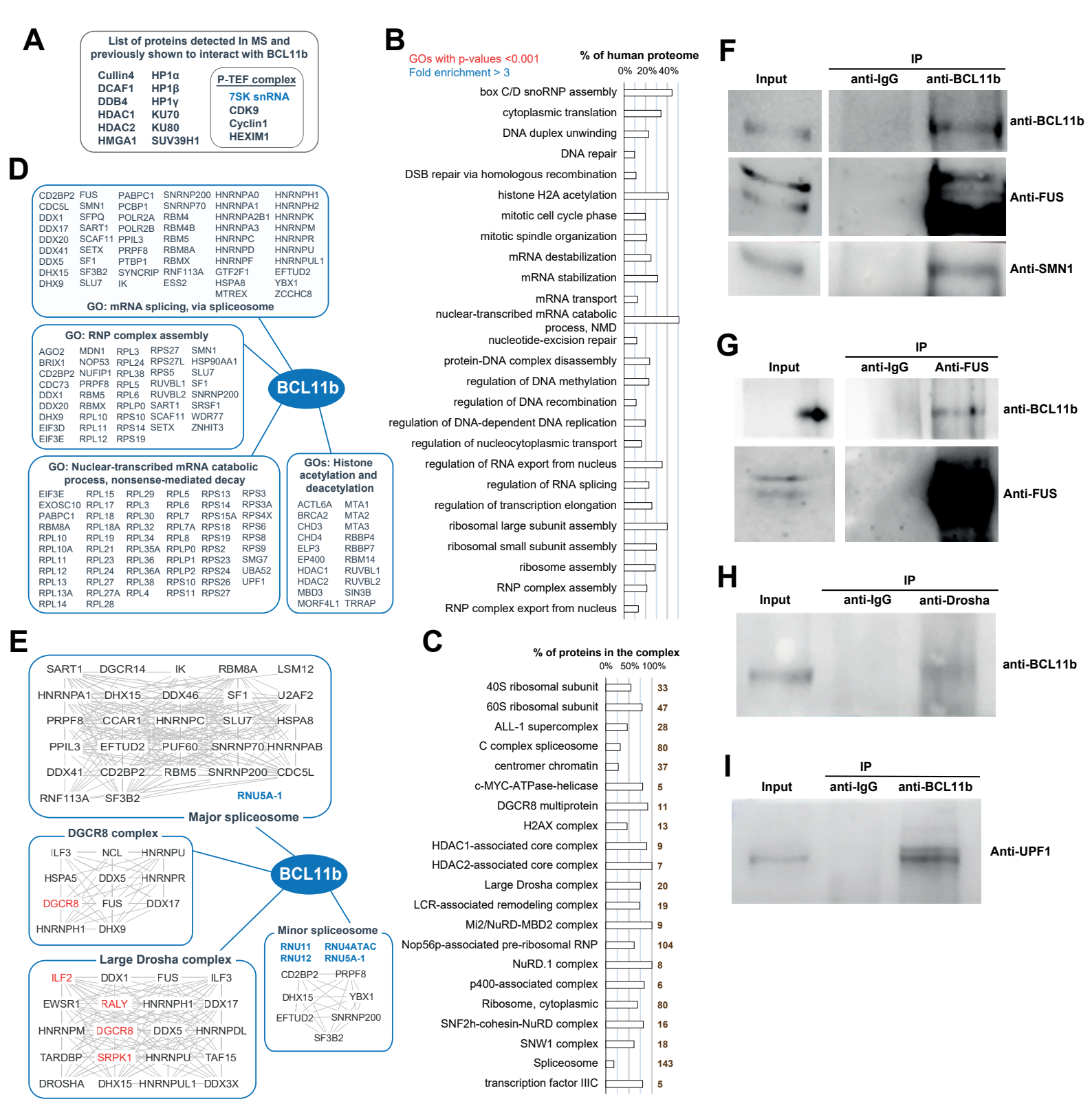
Gasic, I., and Mitchison, T.J. (2019). Autoregulation and repair in microtubule homeostasis. Current Opinion in Cell Biology *56*, 80-87.

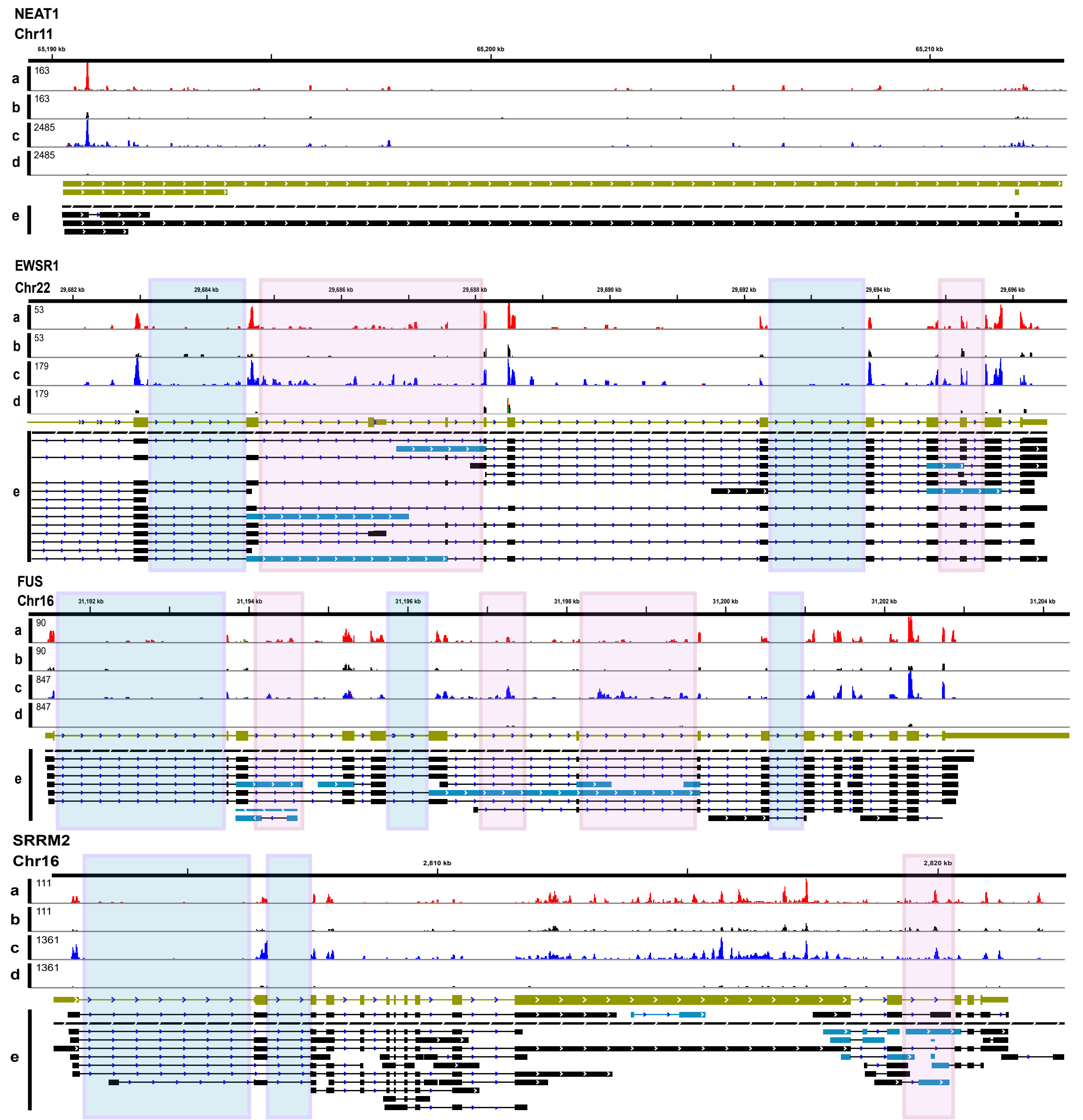
Kominami, R. (2012). Role of the transcription factor Bcl11b in development and lymphomagenesis. Proc Jpn Acad Ser B Phys Biol Sci *88*, 72-87.
Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nature Reviews Molecular Cell Biology *20*, 406-420.
Lennon, M.J., Jones, S.P., Lovelace, M.D., Guillemin, G.J., and Brew, B.J. (2016). Bcl11b: A New Piece to the Complex Puzzle of Amyotrophic Lateral Sclerosis Neuropathogenesis? Neurotoxicity Research *29*, 201-207.
Lennon, M.J., Jones, S.P., Lovelace, M.D., Guillemin, G.J., and Brew, B.J. (2017). Bcl11b—A Critical Neurodevelopmental Transcription Factor—Roles in Health and Disease. Frontiers in Cellular Neuroscience *11*.
Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., *et al.* (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proceedings of the National Academy of Sciences *114*, 5800-5807.
Maeda, M., Harris, A.W., Kingham, B.F., Lumpkin, C.J., Opdenaker, L.M., McCahan, S.M., Wang, W., and Butchbach, M.E.R. (2014). Transcriptome profiling of spinal muscular atrophy motor neurons derived from mouse embryonic stem cells. PLoS One *9*, e106818.
Marban, C., Suzanne, S., Dequiedt, F., de Walque, S., Redel, L., Van Lint, C., Aunis, D., and Rohr, O. (2007). Recruitment of chromatin-modifying enzymes by CTIP2 promotes HIV-1 transcriptional silencing. EMBO J *26*, 412-423.
Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nature Reviews Molecular Cell Biology *15*, 108-121.
Orth, M.F., Hölting, T.L.B., Dallmayer, M., Wehweck, F.S., Paul, T., Musa, J., Baldauf, M.C., Surdez, D., Delattre, O., Knott, M.M.L., *et al.* (2020). High Specificity of BCL11B and GLG1 for EWSR1-FLI1 and EWSR1-ERG Positive Ewing Sarcoma. Cancers *12*, 644.
Pong, S.K., and Gullerova, M. (2018). Noncanonical functions of microRNA pathway enzymes – Drosha, DGCR8, Dicer and Ago proteins. FEBS Letters *592*, 2973-2986.
Rohr, O., Lecestre, D., Chasserot-Golaz, S., Marban, C., Avram, D., Aunis, D., Leid, M., and Schaeffer, E. (2003). Recruitment of Tat to heterochromatin protein HP1 via interaction with CTIP2 inhibits human immunodeficiency virus type 1 replication in microglial cells. J Virol *77*, 5415-5427.
Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M., *et al.* (2007). Molecular markers of early Parkinson's disease based on gene expression in blood. Proceedings of the National Academy of Sciences *104*, 955-960.
Shadrina, O., Garanina, I., Korolev, S., Zatsepin, T., Van Assche, J., Daouad, F., Wallet, C., Rohr, O., and Gottikh, M. (2020). Analysis of RNA binding properties of human Ku protein reveals its interactions with 7SK snRNA and protein components of 7SK snRNP complex. Biochimie *171-172*, 110-123.
Shehadeh, L.A., Yu, K., Wang, L., Guevara, A., Singer, C., Vance, J., and Papapetropoulos, S. (2010). SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. PLoS One *5*, e9104.
Shkreta, L., and Chabot, B. (2015). The RNA Splicing Response to DNA Damage. Biomolecules *5*, 2935-2977.
Sun, J., Shi, Y., and Yildirim, E. (2019). The Nuclear Pore Complex in Cell Type-Specific Chromatin Structure and Gene Regulation. Trends in Genetics *35*, 579-588.
Tuck, A.C., Rankova, A., Arpat, A.B., Liechti, L.A., Hess, D., Iesmantavicius, V., Castelo-Szekely, V., Gatfield, D., and Bühler, M. (2020). Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. Molecular Cell *77*, 1222-1236.e1213.
Wakabayashi, Y., Watanabe, H., Inoue, J., Takeda, N., Sakata, J., Mishima, Y., Hitomi, J., Yamamoto, T., Utsuyama, M., Niwa, O., *et al.* (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. Nat Immunol *4*, 533-539.
Wang, E., and Aifantis, I. (2020). RNA Splicing and Cancer. Trends in Cancer.

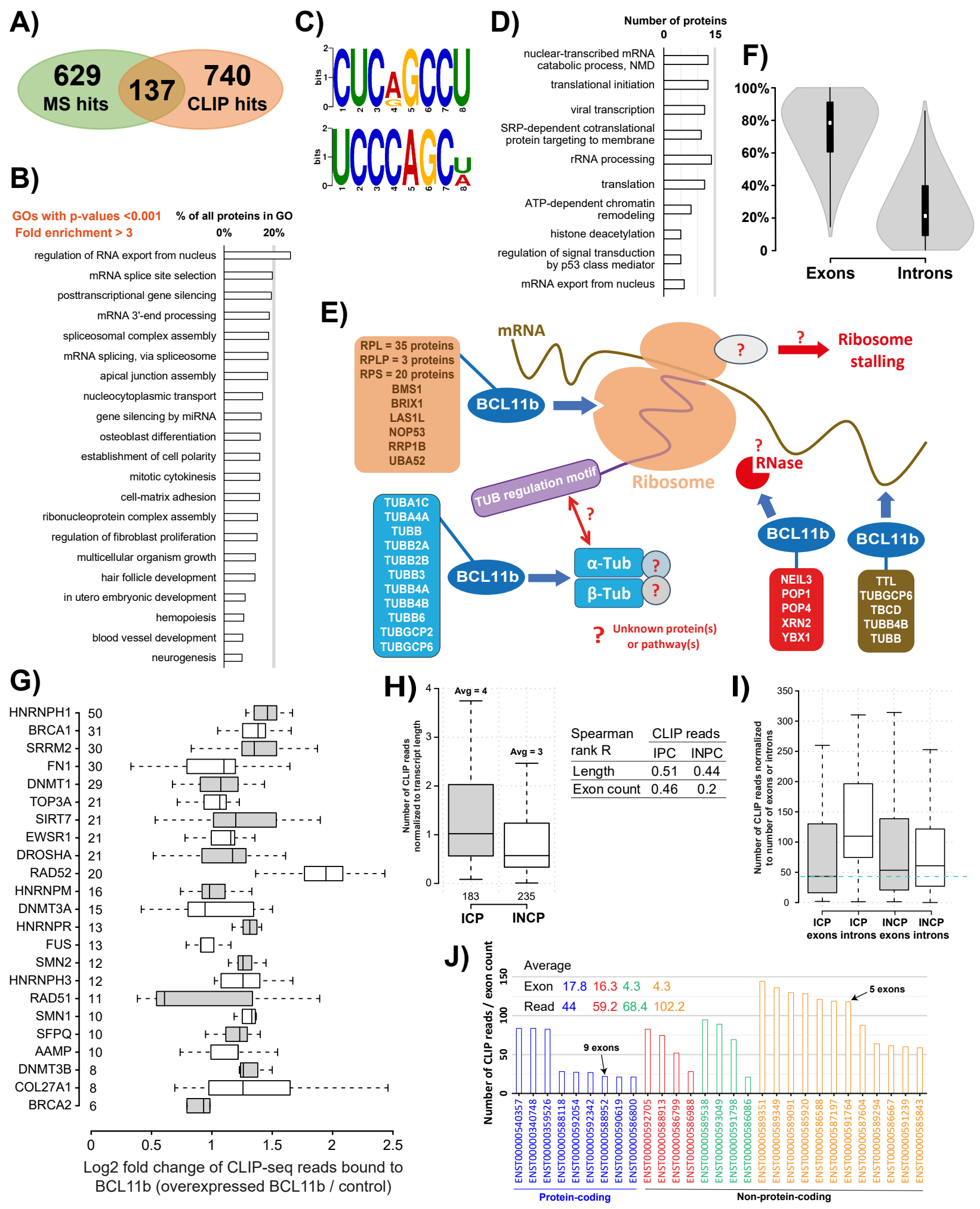
Wiles, E.T., Lui-Sargent, B., Bell, R., and Lessnick, S.L. (2013). BCL11B Is Up-Regulated by EWS/FLI and Contributes to the Transformed Phenotype in Ewing Sarcoma. PLOS ONE *8*, e59369.

Wu, Y.-Y., and Kuo, H.-C. (2020). Functional roles and networks of non-coding RNAs in the pathogenesis of neurodegenerative diseases. Journal of Biomedical Science *27*, 49.

Yamazaki, T., Chen, S., Yu, Y., Yan, B., Haertlein, Tyler C., Carrasco, Monica A., Tapia, Juan C., Zhai, B., Das, R., Lalancette-Hebert, M., *et al.* (2012). FUS-SMN Protein Interactions Link the Motor Neuron Diseases ALS and SMA. Cell Reports *2*, 799-806.


A
List of proteins detected In MS and previously shown to interact with BCL11b
Cullin4 HP1α
DCAF1 HP1β
DDB4 HP1γ
HDAC1 KU70
HDAC2 KU80
HMGA1 SUV39H1
P-TEF complex
7SK snRNA
CDK9
Cyclin1
HEXIM1
B
GOs with p-values <0.001
Fold enrichment > 3
% of human proteome
0% 20% 40%
box C/D snoRNP assembly
cytoplasmic translation
DNA duplex unwinding
DNA repair
DSB repair via homologous recombination
histone H2A acetylation
mitotic cell cycle phase
mitotic spindle organization
mRNA destabilization
mRNA stabilization
mRNA transport
nuclear-transcribed mRNA catabolic process, NMD
nucleotide-excision repair
protein-DNA complex disassembly
regulation of DNA methylation
regulation of DNA recombination
regulation of DNA-dependent DNA replication
regulation of nucleocytoplasmic transport
regulation of RNA export from nucleus
regulation of RNA splicing
regulation of transcription elongation
ribosomal large subunit assembly
ribosomal small subunit assembly
ribosome assembly
RNP complex assembly
RNP complex export from nucleus
F
Input
IP
anti-IgG
anti-BCL11b
anti-BCL11b
Anti-FUS
Anti-SMN1
G
Input
IP
anti-IgG
Anti-FUS
anti-BCL11b
Anti-FUS
H
Input
IP
anti-IgG
anti-Drosha
anti-BCL11b
I
Input
IP
anti-IgG
anti-BCL11b
Anti-UPF1
D
CD2BP2 FUS PABPC1 SNRNP200 HNRNPA0 HNRNPH1
CDC5L SMN1 PCBP1 SNRNP70 HNRNPA1 HNRNPH2
DDX1 SFPQ POLR2A RBM4 HNRNPA2B1 HNRNPK
DDX17 SART1 POLR2B RBM4B HNRNPA3 HNRNPM
DDX20 SCAF11 PPIL3 RBM5 HNRNPC HNRNPR
DDX41 SETX PRPF8 RBM8A HNRNPD HNRNPU
DDX5 SF1 PTBP1 RBMX HNRNPF HNRNPUL1
DHX15 SF3B2 SYNCRIP RNF113A GTF2F1 EFTUD2
DHX9 SLU7 IK ESS2 HSPA8 YBX1
MTREX ZCCHC8
GO: mRNA splicing, via spliceosome
GO: RNP complex assembly
AGO2 MDN1 RPL3 RPS27 SMN1
BRIX1 NOP53 RPL24 RPS27L HSP90AA1
CD2BP2 NUFIP1 RPL38 RPS5 SLU7
CDC73 PRPF8 RPL5 RUVBL1 SF1
DDX1 RBM5 RPL6 RUVBL2 SNRNP200
DDX20 RBMX RPLP0 SART1 SRSF1
DHX9 RPL10 RPS10 SCAF11 WDR77
EIF3D RPL11 RPS14 SETX ZNHIT3
EIF3E RPL12 RPS19
BCL11b
GO: Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
EIF3E RPL15 RPL29 RPL5 RPS13 RPS3
EXOSC10 RPL17 RPL3 RPL6 RPS14 RPS3A
PABPC1 RPL18 RPL30 RPL7 RPS15A RPS4X
RBM8A RPL18A RPL32 RPL7A RPS18 RPS6
RPL10 RPL19 RPL34 RPL8 RPS19 RPS8
RPL10A RPL21 RPL35A RPLP0 RPS2 RPS9
RPL11 RPL23 RPL36 RPLP1 RPS23 SMG7
RPL12 RPL24 RPL36A RPLP2 RPS24 UBA52
RPL13 RPL27 RPL38 RPS10 RPS26 UPF1
RPL13A RPL27A RPL4 RPS11 RPS27
RPL14 RPL28
GOs: Histone acetylation and deacetylation
ACTL6A MTA1
BRCA2 MTA2
CHD3 MTA3
CHD4 RBBP4
ELP3 RBBP7
EP400 RBM14
HDAC1 RUVBL1
HDAC2 RUVBL2
MBD3 SIN3B
MORF4L1 TRRAP
E
SART1 DGCR14 IK RBM8A LSM12
HNRNPA1 DHX15 DDX46 SF1 U2AF2
PRPF8 CCAR1 HNRNPC SLU7 HSPA8
PPIL3 EFTUD2 PUF60 SNRNP70 HNRNPAB
DDX41 CD2BP2 RBM5 SNRNP200 CDC5L
RNF113A SF3B2 RNU5A-1
Major spliceosome
DGCR8 complex
ILF3 NCL HNRNPU
HSPA5 DDX5 HNRNPR
DGCR8 FUS DDX17
HNRNPH1 DHX9
BCL11b
Large Drosha complex
ILF2 DDX1 FUS ILF3
EWSR1 RALY HNRNPH1 DDX17
HNRNPM DGCR8 DDX5 HNRNPDL
TARDBP SRPK1 HNRNPU TAF15
DROSHA DHX15 HNRNPUL1 DDX3X
Minor spliceosome
RNU11 RNU4ATAC
RNU12 RNU5A-1
CD2BP2 PRPF8
DHX15 YBX1
EFTUD2 SNRNP200
SF3B2
C
% of proteins in the complex
0% 50% 100%
40S ribosomal subunit 33
60S ribosomal subunit 47
ALL-1 supercomplex 28
C complex spliceosome 80
centromer chromatin 37
c-MYC-ATPase-helicase 5
DGCR8 multiprotein 11
H2AX complex 13
HDAC1-associated core complex 9
HDAC2-associated core complex 7
Large Drosha complex 20
LCR-associated remodeling complex 19
Mi2/NuRD-MBD2 complex 9
Nop56p-associated pre-ribosomal RNP 104
NuRD.1 complex 8
p400-associated complex 6
Ribosome, cytoplasmic 80
SNF2h-cohesin-NuRD complex 16
SNW1 complex 18
Spliceosome 143
transcription factor IIIC 5

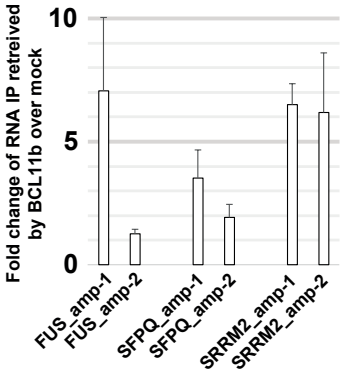

NEAT1
Chr11
65,190 kb
65,200 kb
65,210 kb
a
163
b
163
c
2485
d
2485
e
EWSR1
Chr22
29,682 kb
29,684 kb
29,686 kb
29,688 kb
29,690 kb
29,692 kb
29,694 kb
29,696 kb
a
53
b
53
c
179
d
179
e
FUS
Chr16
31,192 kb
31,194 kb
31,196 kb
31,198 kb
31,200 kb
31,202 kb
31,204 kb
a
90
b
90
c
847
d
847
e
SRRM2
Chr16
2,810 kb
2,820 kb
a
111
b
111
c
1361
d
1361
e


A)
629
MS hits
137
740
CLIP hits
B)
GOs with p-values <0.001
Fold enrichment > 3
% of all proteins in GO
0%
20%
regulation of RNA export from nucleus
mRNA splice site selection
posttranscriptional gene silencing
mRNA 3'-end processing
spliceosomal complex assembly
mRNA splicing, via spliceosome
apical junction assembly
nucleocytoplasmic transport
gene silencing by miRNA
osteoblast differentiation
establishment of cell polarity
mitotic cytokinesis
cell-matrix adhesion
ribonucleoprotein complex assembly
regulation of fibroblast proliferation
multicellular organism growth
hair follicle development
in utero embryonic development
hemopoiesis
blood vessel development
neurogenesis
C)
bits
D)
Number of proteins
0
15
nuclear-transcribed mRNA catabolic process, NMD
translational initiation
viral transcription
SRP-dependent cotranslational protein targeting to membrane
rRNA processing
translation
ATP-dependent chromatin remodeling
histone deacetylation
regulation of signal transduction by p53 class mediator
mRNA export from nucleus
F)
100%
80%
60%
40%
20%
0
Exons
Introns
E)
RPL = 35 proteins
RPLP = 3 proteins
RPS = 20 proteins
BMS1
BRIX1
LAS1L
NOP53
RRP1B
UBA52
mRNA
BCL11b
Ribosome
Ribosome stalling
RNase
TUB regulation motif
TUBA1C
TUBA4A
TUBB
TUBB2A
TUBB2B
TUBB3
TUBB4A
TUBB4B
TUBB6
TUBGCP2
TUBGCP6
α-Tub
β-Tub
Unknown protein(s) or pathway(s)
NEIL3
POP1
POP4
XRN2
YBX1
TTL
TUBGCP6
TBCD
TUBB4B
TUBB
G)
HNRNPH1 50
BRCA1 31
SRRM2 30
FN1 30
DNMT1 29
TOP3A 21
SIRT7 21
EWSR1 21
DROSHA 21
RAD52 20
HNRNPM 16
DNMT3A 15
HNRNPR 13
FUS 13
SMN2 12
HNRNPH3 12
RAD51 11
SMN1 10
SFPQ 10
AAMP 10
DNMT3B 8
COL27A1 8
BRCA2 6
0
0.5
1
1.5
2
2.5
Log2 fold change of CLIP-seq reads bound to BCL11b (overexpressed BCL11b / control)
H)
Avg = 4
Avg = 3
Number of CLIP reads normalized to transcript length
183
235
ICP
INCP
Spearman rank R
CLIP reads
IPC
INPC
Length 0.51 0.44
Exon count 0.46 0.2
I)
Number of CLIP reads normalized to number of exons or introns
ICP exons
ICP introns
INCP exons
INCP introns
J)
Average
Exon 17.8 16.3 4.3 4.3
Read 44 59.2 68.4 102.2
9 exons
5 exons
Number of CLIP reads / exon count
Protein-coding
Non-protein-coding

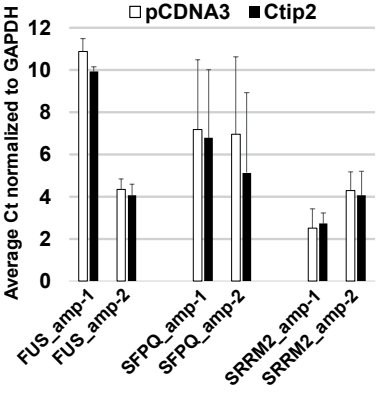
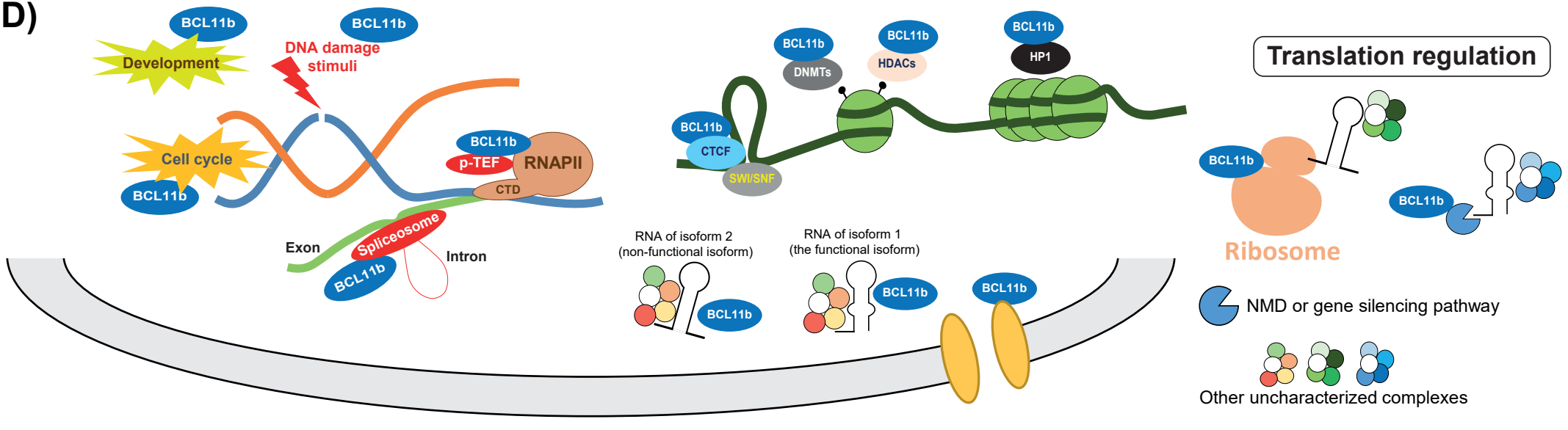
A)

	BCL11b interacts with	PD	SMA	ALS	Ewing's sarcoma
BCL11b		Down-regulated	Down-regulated		Up-regulated
SRRM2	RNA	Up-regulated	-	-	-
SMN1	Protein	-	Down-regulated	-	-
NEAT1	RNA	-	Up-regulated	-	-
EWSR1	RNA and protein	-	-	Contribute in	Contribute in
FUS	RNA and protein	-	-	Mutated	-
RNPs	RNA, protein & ncRNA	-	Defects	Defects	-
TDP43	RNA and protein	-	-	Mutated	-
TAF15	RNA and protein	-	-	Contribute in	-
VCP	RNA	-	-	Contribute in	-

B)



C)



D)