

Functional characterization of RebL1 highlights the evolutionary conservation of oncogenic activities of the RBBP4/7 orthologue in *Tetrahymena thermophila*

Syed Nabeel-Shah¹, Jyoti Garg^{1,2,†}, Alejandro Saettone^{1,†}, Kanwal Ashraf², Hyunmin Lee^{3,4}, Suzanne Wahab¹, Nujhat Ahmed^{4,5}, Jacob Fine², Joanna Derynck¹, Shuye Pu⁴, Marcelo Ponce⁶, Edyta Marcon⁴, Zhaolei Zhang^{3,4,5}, Jack F. Greenblatt^{4,5}, Ronald E. Pearlman², Jean-Philippe Lambert^{7,*} and Jeffrey Fillingham^{1,*}

¹Department of Chemistry and Biology, Ryerson University, 350 Victoria St., Toronto M5B 2K3, Canada, ²Department of Biology, York University, 4700 Keele St., Toronto M3J 1P3, Canada, ³Department of Computer Science, University of Toronto, Toronto M5S 1A8, Canada, ⁴Donnelly Centre, University of Toronto, Toronto M5S 3E1, Canada, ⁵Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada, ⁶SciNet HPC Consortium, University of Toronto, 661 University Avenue, Suite 1140, Toronto M5G 1M1, Canada and ⁷Department of Molecular Medicine, Cancer Research Center, Big Data Research Center, Université Laval, Quebec City, Canada; CHU de Québec Research Center, CHUL, 2705 Laurier Boulevard, Quebec City G1V 4G2, Canada

Received March 29, 2021; Revised April 22, 2021; Editorial Decision April 29, 2021; Accepted May 04, 2021

ABSTRACT

Retinoblastoma-binding proteins 4 and 7 (RBBP4 and RBBP7) are two highly homologous human histone chaperones. They function in epigenetic regulation as subunits of multiple chromatin-related complexes and have been implicated in numerous cancers. Due to their overlapping functions, our understanding of RBBP4 and 7, particularly outside of Opisthokonts, has remained limited. Here, we report that in the ciliate protozoan *Tetrahymena thermophila* a single orthologue of human RBBP4 and 7 proteins, RebL1, physically interacts with histone H4 and functions in multiple epigenetic regulatory pathways. Functional proteomics identified conserved functional links for *Tetrahymena* RebL1 protein as well as human RBBP4 and 7. We found that putative subunits of multiple chromatin-related complexes including CAF1, Hat1, Rpd3, and MuvB, co-purified with RebL1 during *Tetrahymena* growth and conjugation. Iterative proteomics analyses revealed that the cell cycle regulatory MuvB-complex in *Tetrahymena* is composed of at least five subunits including

evolutionarily conserved Lin54, Lin9 and RebL1 proteins. Genome-wide analyses indicated that RebL1 and Lin54 (Anqa1) bind within genic and intergenic regions. Moreover, Anqa1 targets primarily promoter regions suggesting a role for *Tetrahymena* MuvB in transcription regulation. RebL1 depletion inhibited cellular growth and reduced the expression levels of Anqa1 and Lin9. Consistent with observations in glioblastoma tumors, RebL1 depletion suppressed DNA repair protein Rad51 in *Tetrahymena*, thus underscoring the evolutionarily conserved functions of RBBP4/7 proteins. Our results suggest the essentiality of RebL1 functions in multiple epigenetic regulatory complexes in which it impacts transcription regulation and cellular viability.

INTRODUCTION

Eukaryotic chromatin provides a means to compact the genome as well as contributing to regulating DNA-mediated processes, such as replication, recombination, repair, and transcription (1). Histone chaperones and chromatin remodelers, as well as histone posttranslational modifications (PTMs), play pivotal roles in chromatin-

*To whom correspondence should be addressed. Tel: +1 416 979 5000 (Ext 2123); Email: jeffrey.fillingham@ryerson.ca
Correspondence may also be addressed to Jean-Philippe Lambert. Tel: +1 418 525 4444 (Ext 49399);
Email: Jean-Philippe.Lambert@crchudequebec.ulaval.ca

[†]The authors wish it to be known that, in their opinion, second and third authors should be regarded as Joint Second Authors.
Present addresses:

Syed Nabeel-Shah, Donnelly Centre, University of Toronto, Toronto M5S 3E1, Canada.
Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada.

related processes (2). For example, multiple histone chaperones, such as nuclear autoantigenic sperm protein (NASP) and anti-silencing factor 1 (Asf1), have been reported to function in the transport pathway of newly synthesized H3/H4 (3,4). Furthermore, chromatin assembly factor-1 (CAF1) and histone-regulator-A (HIRA) function in DNA replication-dependent (RD) and replication-independent (RI) chromatin assembly processes to deposit either H3–H4 or H3.3–H4, respectively (5,6). ATP-dependent chromatin remodelers, on the other hand, mobilize DNA around nucleosomes and function in transcription regulation (7). Notably, the ‘Nucleosome Remodeling and Deacetylase’ (NuRD) complex is widely present at gene promoter regions and enhancer elements and plays an important role in regulating transcription and genome integrity, as well as cell cycle progression (8,9). Finally, histone PTMs are important for chromatin assembly, as well as regulation of gene expression (10). Newly synthesized H4 histones are di-acetylated at lysine (K) 5 and K12 residues, which has been shown to be important for chromatin assembly (11). These deposition-related H4K5/12ac marks are installed by the HAT1-complex, which is composed of at least two subunits, including catalytic Hat1 and histone-binding WD40 repeat protein Hat2 (RBBP4 in humans) (12,13). In *Saccharomyces cerevisiae* an additional nuclear subunit called Hif1 (human NASP) also interacts with the HAT1-complex at roughly stoichiometric levels (14). Whereas the deposition-related H4 PTMs are conserved across species (15), the corresponding enzymatic complex has not been well described in evolutionarily distant protozoan lineages.

Retinoblastoma-binding proteins 4 and 7 (RBBP4 and RBBP7) (Cac3 and Hat2 respectively in budding yeast), also known as Retinoblastoma Associated protein 48 (RbAp48) and 46 (RbAp46), respectively, are two highly homologous histone chaperones (~90% sequence identity) with similar structures (16) that were initially identified as retinoblastoma-associated proteins (17,18). Both proteins contain WD40-repeats forming a seven-bladed β -propeller fold, which is thought to function as a protein interaction scaffold in multiprotein complexes (16,19). While human and mouse encode both homologues, only a single homologue of RBBP4 and RBBP7 has been identified in some organisms, including *Caenorhabditis elegans* (Lin53) and *Drosophila melanogaster* (p55) (20–22). RBBP4 and RBBP7 histone chaperones are components of several multi-protein complexes that are involved in diverse chromatin-related functions, including chromatin assembly and remodeling, histone post-translational modifications and regulation of gene expression (23,24). For example, RBBP4 and RBBP7 have been identified as subunits of the NuRD complex (25), switch independent 3A (Sin3A) complex (26), and polycomb repressive complex 2 (PRC2) (26–28). Furthermore, RBBP4 (Cac3) has been found to be a subunit of the CAF1 complex (29,30), and RBBP7 (Hat2) has been reported to be an essential component of the HAT1-complex (31). Within multi-subunit protein complexes, RBBP4 and RBBP7 are thought to function as chromatin adaptors via their direct interaction with histones H3/H4 (16,32).

RBBP4 also functions as a component of transcriptional regulator ‘multi-vulva class B’ (MuvB)-complex (33–35). The mammalian core MuvB complex, composed of five

subunits including Lin9, Lin37, Lin52, Lin54 and RBBP4, is a master regulator of cell-cycle-dependent gene expression (33). The composition of the core MuvB remains unaltered during different cell cycle stages, whereas its interaction partners change, which reverses its function from a transcriptional repressor to an activator (33,36,37). The core MuvB interacts with E2F4/5, DP1/2 and p130/p107 proteins to form the repressor ‘dimerization partner (DP), RB-like, E2F and MuvB’ (DREAM) complex, which represses the G1/S and G2/M genes during G0 and early G1 (35,36). The DREAM-specific proteins dissociate during late G1, and the MuvB core then forms an interaction with the B-MYB transcription factor during S phase (35,38). The interaction of B-MYB with MuvB is necessary for FOXM1 recruitment during G2 (38). These B-MYB-MuvB (MMB) and FOXM1-MuvB complexes stimulate the expression of genes that peak during G2 and M phases (34). Among the MuvB core subunits, Lin54 is a DNA-binding protein and has two tandem cysteine-rich (CXC) domains that share sequence similarity with Tesmin, a testis-specific metallothionein-like protein (39,40). Lin54 directly binds to DNA by recognizing the ‘cell cycle genes homology region’ (CHR) element (41–43). The CHR is considered to be the central promoter element required for the cell cycle-dependent regulation of G2/M phase-expressed genes (43). Several studies have shown that the DREAM-complex localizes to its target promoters via E2F/DP and LIN-54 DNA sequence motifs (44–46). In addition to Lin54, DREAM also likely comes in contact with nucleosomes by utilizing RBBP4 via its interaction with histones (33). Furthermore, *Drosophila* p55 (RBBP4) has been shown to be important for DREAM-mediated gene repression (47).

It is increasingly recognized that epigenetic aberrations contribute to tumor initiation and development (48). As such, it is not surprising that RBBP4 and RBBP7 have been extensively implicated in many cancers and are now being pursued as valuable drug targets (reviewed by 24). For example, RBBP7 is upregulated in many cancers, including non-small cell lung cancer, renal cell carcinoma, and breast cancer (49–51). Similarly, RBBP4 has been found to be overexpressed in hepatocellular carcinoma and acute myeloid leukemia (52,53). In human glioblastoma (GBM) tumor cells, knockdown of RBBP4 causes sensitization of tumor cells to temozolomide (TMZ) chemotherapy and suppresses the expression of several DNA repair genes, including RAD51, a key enzyme of the homologous recombination repair pathway (54).

Despite their critical role in multiple epigenetic pathways, the function of RBBP4 and RBBP7 and their associated protein complexes has not been well documented outside of the Opisthokonta. *Tetrahymena thermophila*, a unicellular ciliate protozoan, is an excellent experimental system to study evolutionarily conserved chromatin-related processes, including gene expression and cell cycle regulation (55–59). *Tetrahymena* has two structurally and functionally distinct nuclei, a germ-line diploid micronucleus (MIC) and a somatic polyploid macronucleus (MAC), that are physically separated within the same cell (60). The MAC essentially regulates all gene expression during vegetative growth, whereas the MIC ensures stable genetic inheritance (61). Both the MAC and MIC arise from the same zygotic nu-

cleus during sexual reproduction (conjugation) (61). During conjugation, two developing nuclei undergo substantial chromatin alterations, including DNA rearrangements and removal of ‘internally eliminated sequences’ (IES), resulting in functionally and structurally different MAC and MIC with distinct epigenetic states (62–64). Many of the epigenetic regulatory pathways in *Tetrahymena* that ensure the establishment of distinct chromatin states within the two nuclei are widely conserved across eukaryotes (65).

Here, we utilized functional proteomics and genomics approaches to characterize the single orthologue of human RBBP4 and RBBP7, RebL1, in *Tetrahymena*. Iterative proteomic analyses uncovered the composition of multiple *Tetrahymena* epigenetic regulatory complexes, such as the CAF1-, Sin3A/Rpd3- and MuvB-complexes, which co-purified with RebL1 during vegetative growth and conjugation. Furthermore, ChIP-seq analysis of Lin54 (Anqa1), combined with previously reported RNA-seq data (66), implicated *Tetrahymena* MuvB in transcription regulation. Consistently, the knockdown (KD) of *REBL1* significantly decreased cellular viability and altered the expression profile of some functionally critical genes, including *RAD51*. Overall, our results revealed a general role of RebL1 in multiple transcriptional regulatory pathways important for cell cycle and chromatin regulation in *Tetrahymena*.

MATERIALS AND METHODS

Cell strains

Tetrahymena strains CU428 [Mpr/Mpr (VII, mp-s)] and B2086 [Mpr+/Mpr+ (II, mp-s)] of inbreeding line B were acquired from the *Tetrahymena* Stock Center, Cornell University, Ithaca, NY (<http://tetrahymena.vet.cornell.edu/>). Cells cultured in 1× SPP were maintained axenically at 30°C as previously described (67). HEK293 cells (Flp-In 293 T-REx cell lines) were obtained from Life Technologies (R780-07) (Invitrogen) and cells were cultured in Dulbecco’s modified Eagle’s medium with 10% FBS and antibiotics as described (68).

Macronuclear gene replacement

Epitope tagging vectors for histone *HHF1*, *HHF2*, *REBL1*, *ANQA1* and *LIN9^{Tt}* were constructed as previously described (67). Two separate ~1 kb fragments up and downstream of the predicted stop codons were amplified using wild-type *T. thermophila* genomic DNA as template and primers as indicated in Supplemental Table S13. The resulting PCR products were digested with KpnI and XhoI (upstream product) or NotI and SacI (downstream product). The digested products were cloned into the tagging vector pBKS-FZZ kindly provided by Dr. Kathleen Collins (University of California, Berkeley, CA). Plasmids were amplified in *Escherichia coli* and the final plasmid was digested with KpnI and SacI prior to transformation. One micrometer gold particles (60 mg/mL; Bio-Rad) were coated with at least 5 µg of the digested plasmid DNA. The DNA coated gold particles were introduced into the *T. thermophila* MAC using biolistic transformation with a PDS-1000/He Biolistic particle delivery system (Bio-Rad) (69). The transformants were selected using paromomycin (60 µg/mL).

MAC homozygosity was achieved through phenotypic assortment by growing the cells in increasing concentrations of paromomycin to a final concentration of 1 mg/mL.

Epitope tagging in HEK293 cells

We cloned gateway-compatible entry clones for RBBP4 and RBBP7 into the pDEST pcDNA5/FRT/TO-eGFP vector according to the manufacturer’s instructions. The cloned vectors were co-transfected into Flp-In T-REx 293 cells together with the pOG44 Flp recombinase expression plasmid. Cells were selected for FRT site-specific recombination into the genome with hygromycin (Life Technologies, 10687010) at 200 µg/mL. Expression of the gene of interest was induced by addition of doxycycline (1 µg/mL) to the culture medium 24 hours before harvesting.

Experimental design for mass spectrometry

At least two biological replicates of each bait were processed independently along with negative controls in each batch of sample. *Tetrahymena* cells without tagged bait (i.e. empty cells) and HEK293 cells expressing GFP alone were used as controls in our analysis. Extensive washes between each sample (see details for each instrument type) were performed to minimize carry-over. The order of sample acquisition on the mass spectrometer was reversed for the second replicate to avoid systematic bias. On the LTQ mass spectrometer, a freshly made column was used for each sample as previously described (67).

Affinity purification and Mass Spectrometry sample preparation

For *Tetrahymena* samples, affinity purification was carried out essentially as described (67,70). Briefly, *T. thermophila* was grown in ~500 mL of 1× SPP to mid-log phase to a final concentration of 3×10^5 cells/mL for vegetative samples. For conjugation, *Tetrahymena* cells of different mating types were grown to mid-log phase and starved in 10 mM Tris–HCl pH 7.4 for 16 to 24 hours. Starved cells were mixed in equal numbers to initiate conjugation and pelleted at 5-hours post mixing. Vegetative and conjugating cells were pelleted and frozen at –80°C. The frozen pellets were thawed on ice and suspended in lysis buffer (10 mM Tris–HCl pH 7.5, 1 mM MgCl₂, 300 mM NaCl and 0.2% NP40 plus yeast protease inhibitors (Sigma)). For nuclease treatment, 500 units of Benzonase (Sigma E8263) was added and extracts were rotated on a Nutator for 30 min at 4°C. Whole cell extracts (WCE) were clarified by centrifugation at 16 000 × g for 30 min. The resulting soluble material was incubated with 50 µL of packed M2-agarose (Sigma) at 4°C for at least 2 hours. The M2-agarose was washed once with 10 mL IPP300 (10 mM Tris–HCl pH 8.0, 300 mM NaCl, 0.1% NP40), twice with 5 mL of IP100 (10 mM Tris–HCl pH 8.0, 100 mM NaCl, 0.1% NP40), and twice with 5 mL of IP100 without detergent (10 mM Tris–HCl pH 8.0, 100 mM NaCl). 500 µL of 0.5 M NH₄OH was added to elute the proteins by rotating for 20 min at room temperature. Preparation of protein eluates for mass spectrometry acquisition is detailed in supplementary methods.

AP-MS procedure in HEK293 cells

The AP-MS procedure for HEK293 cells was performed essentially as previously described (68,71,72). In brief, $\sim 20 \times 10^6$ cells were grown independently in two batches representing two biological replicates. Cells were harvested after 24-hour induction of protein expression using doxycycline. HEK293 cell pellets were lysed in high-salt NP-40 lysis buffer (10 mM Tris-HCl pH 8.0, 420 mM NaCl, 0.1% NP-40, plus protease/phosphatase inhibitors) with three freeze-thaw cycles. The lysate was sonicated as described (68) and treated with Benzonase for 30 min at 4°C with end-to-end rotation. The WCE was centrifuged to pellet any cellular debris. We immunoprecipitated GFP-tagged RBBP4 and RBBP7 proteins with anti-GFP antibody (G10362, Life Technologies) overnight followed by a 2-hour incubation with Protein G Dynabeads (Invitrogen). The beads were washed three times with buffer (10 mM Tris-HCl, pH 7.9, 420 mM NaCl, 0.1% NP-40) and twice with buffer without detergent (10 mM Tris-HCl, pH 7.9, 420 mM NaCl). The immunoprecipitated proteins were eluted with NH_4OH and lyophilized. Proteins for MS analysis were prepared by in-solution trypsin digestion. The protein pellet was suspended in 44 μL of 50 mM NH_4HCO_3 , the sample was reduced with 100 mM TCEP-HCl, alkylated with 500 mM iodoacetamide, and digested with 1 μg of trypsin overnight at 37°C. Samples were desalted using ZipTip Pipette tips (EMD Millipore) via standard procedures. The desalted samples were analyzed with an LTQ-Orbitrap Velos mass spectrometer (ThermoFisher Scientific). Raw MS data were searched with Maxquant (v.1.6.6.0) (73) as described previously (72), yielding spectral counts and MS intensities for each identified protein for individual human proteins for each experiment. The resulting data were filtered using SAINTexpress to obtain confidence values utilizing the two biological replicates. The AP-MS data generated from HEK293 cells expressing GFP alone were used as negative control for SAINT analysis.

MS data visualization and archiving

Cytoscape (V3.4.0; (74)) was used to generate protein-protein interaction networks. For better illustration, individual nodes were manually arranged in physical complexes. Dot plots and heatmaps were generated using ProHitsviz (75). The annotation of the co-purifying partners was carried out using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). SMART domain analysis (<http://smart.embl-heidelberg.de/>) of the predicted proteins was carried out. All MS files used in this study were deposited at MassIVE (<http://massive.ucsd.edu>). Additional details (including MassIVE accession numbers and FTP download links) can be found in Supplemental Table S14.

ChIP-Seq analysis

The ChIP-seq experiments and related analyses were performed in vegetatively growing *Tetrahymena*, as described previously (76–78), see supplemental methods for complete experimental details. Illumina adaptor sequences were removed from the 3' ends of 51-nt reads, and the remaining

reads were mapped to the *Tetrahymena* genome (2014 annotation) using Bowtie 2 with default settings (79). After removal of duplicate reads, peaks were called jointly on immunoprecipitated and input samples with MACS2 (version 2.1.2) where the peak's q -value must be lower than 0.05 (80). The ChIP inputs were used as the control. The metagene analysis was performed using ChIP-Seq reads normalized over the inputs and by BPM (bins per million). The plots were generated using deeptools (81). The identified summits (apex of the peak) with 50 bp extension in both directions were analyzed using MEME-ChIP software for *de novo* motif discovery (<http://meme-suite.org/>) (82). The occurrence of Anqa1 motif along the peaks/genome was identified using FIMO with default settings (83). The closest TSS sites to the motifs were found using HOMER software's 'annotatePeaks' program (84).

Gene expression data

For gene expression analysis, we used microarray data (accession number GSE11300) (<http://tfgd.ihb.ac.cn/>) (85) and the expression values were represented in the heatmap format. To assess the similarities in gene expression profiles hierarchical clustering was performed. RNA-seq data corresponding to the *DPL2* knockout was acquired from GSE104524 and analyzed for differential expression as described previously (86).

RNAi vector construction and RT-qPCR experiments

RNAi vectors were constructed essentially as described (87). RebL1 target sequence used in hairpin RNA constructs was amplified from CU428 genomic DNA using PrimeSTAR Max DNA polymerase with primers as listed in Supplemental Table S13. To create the hairpin cassette, amplified target fragments were cloned into the BamHI-BamHI and PstI-PstI sites, respectively, of pAkRNAi-NEO5, which was kindly provided by Dr Takahiko Akematsu, with the NEBuilder HF DNA Assembly kit. In this system, the NEO5 cassette of the backbone plasmid, which confers paromomycin resistance, has been replaced by a puromycin resistance marker (PAC) under the MTT2 copper-inducible promoter (87). The resulting plasmid was linearized with SacI and KpnI before biolistic transformation. The PAC cassette was activated by adding 630 μM CuSO_4 to the cells, with the addition of 200 $\mu\text{g}/\text{mL}$ puromycin dihydrochloride (Cayman Chemical, Ann Arbor, MI, USA, Cat. CAYM13884-500) to select transformants. RNAi was induced in cells carrying the hairpin construct by the addition of 0.5 $\mu\text{g}/\text{mL}$ CdCl_2 during vegetative growth. The expression of the target gene was examined using RT-qPCR analysis. To this end, we isolated total RNA from the RNAi-treated or WT *Tetrahymena* cells using TRIzol (Life Technologies) as per the supplier's instructions. The isolated RNA was treated with Deoxyribonuclease I (RNase-free, Thermo). cDNA was prepared using iScriptTM Reverse Transcription Supermix for RT-qPCR. qPCR was performed in technical triplicates using the cDNA prepared from three individual KD cell lines. The data were normalized to the expression levels of *HTA3*. Primers are indicated in Supplemental Table S13.

Viability assays

Viability assays were carried out essentially as described (88). We initiated our analysis by isolating single cells of wild-type or *REBL1*-RNAi treated lines in drops of media with and without 0.5 $\mu\text{g}/\text{mL}$ CdCl₂. After 2 days of growth, the approximate number of cells in each drop was determined and drops containing > 1000 cells were considered as viable. The experiment was conducted in five independent biological replicates with 50 cells isolated in each replicate. Finally, the number of viable drops from each replicate for each condition was represented as a bean plot. We used Student's *t*-test to assess statistically significant difference in the averages of viable drops of examined conditions.

RESULTS

RebL1 physically interacts with histone H4

We sought to identify the *Tetrahymena* RBBP4 and RBBP7 histone chaperones by first identifying the interaction partners for the core histone H4, one of their predicted partners. The *Tetrahymena* genome contains two core histone H4s encoded by *HHF1* and *HHF2* (89). We generated *Tetrahymena* cell lines stably expressing *HHF1* and *HHF2* with a C-terminal FZZ epitope tag from their endogenous MAC loci. The FZZ epitope tag comprises 3 \times FLAG and two protein A moieties separated by a TEV cleavage site, which can be utilized in affinity purification experiments and indirect immunofluorescence (IF), as well as genome-wide studies. The successful expression of the endogenously tagged proteins was confirmed by western blotting analysis in whole cell extracts (WCEs) prepared either from the *HHF1*- and *HHF2*-FZZ expressing cells or untagged wild-type *Tetrahymena* (Supplemental Figure S1A). Similar to the core histone H3 (90), our IF analysis indicated that during vegetative growth Hhf1- and Hhf2-FZZ localize to both MAC and MIC (Supplemental Figure S1B). This observation indicates that the FZZ-tagged histones remain functionally competent, as previously reported (58,91). Since *HHF1* and *HHF2* encode identical proteins, we used only the *HHF1*-FZZ cells (designated as H4) for our downstream analyses.

Using our established pipeline (67,70,77), we performed affinity purifications (AP) coupled to tandem mass spectrometry (AP-MS) in biological replicates to identify the interaction partners of H4-FZZ. The MS data were scored using the SAINTexpress algorithm, which utilizes semiquantitative spectral counts for assigning a confidence value to individual protein-protein interactions (PPIs) (92). SAINTexpress analysis of the H4-FZZ AP-MS data filtered against several control purifications identified 18 high-confidence co-purifying proteins (≤ 0.01 false discovery rate [FDR]) (Figure 1A; Supplemental Table S1). The H4-FZZ interaction partners included histone chaperones Asf1 and Nrp1, two heat shock Hsp70s, an Importin β 6, MCM-complex proteins including Mcm-2 and Mcm-6, FACT-complex subunit Spt16 (Cet1), Poly [ADP-ribose] polymerase 6 (Parp6) and Hat1, all of which have been shown to function in histone regulation and/or co-purify with H3 (58,67,93,94).

We identified THERM_00688660 as one of the high-confidence (FDR ≤ 0.01) H4 co-purifying proteins which shared sequence similarity with human RBBP4 and

RBBP7. The co-purification with H4 of an RBBP4/7-like protein (RebL1, retinoblastoma binding protein 4/7-like 1) is consistent with previous reports describing human and budding yeast orthologues to function as histone chaperones (95). We subsequently investigated THERM_00688660 as a putative RBBP4/7 *Tetrahymena* orthologue in greater detail.

RBBP4/7 is conserved in *Tetrahymena*

WD40-repeat family proteins are encoded by multiple genes in most organisms (19,96). We examined the *Tetrahymena* genome for the existence of additional RBBP4/7-like proteins with predicted WD40-repeats. Using the human RBBP4 and RBBP7 proteins to search against the *Tetrahymena* genome, we identified two proteins, including THERM_000709589 and the H4 co-purifying THERM_00688660 (RebL1). Both identified proteins have conserved domain architecture with an N-terminal H4-binding domain followed by six WD40 repeats (Figure 1B; Supplemental Figure S1C). Multiple sequence alignment analysis, however, indicated that THERM_000709589, which we named as Wdc1 (WD40 domain containing 1), shares only weak sequence identity ($\sim 18\%$) with its human and yeast counterparts; in comparison, RebL1 had an overall $\sim 46\%$ sequence identity (Figure 1B). It is worth noting that *Tetrahymena* also encodes another WD40 repeat protein, THERM_00467910, which appears to be an orthologue of *S. cerevisiae* Rrb1 (known as GRWD1 in humans) (Supplemental Figure S1C), as assessed by reciprocal BLAST searches. Rrb1 is known to function in ribosome biogenesis (97), making THERM_00467910 an unlikely candidate to be an RBBP4/7 orthologue.

To further categorize the identified WD40 proteins in *Tetrahymena*, we carried out a phylogenetic analysis and observed that RebL1 and Wdc1 are related to the RBBP4/7-family of histone chaperones (Supplemental Figure S1D). Wdc1 and *Arabidopsis thaliana* Msi4 clustered together, distinctly within the RBBP4/7 group on the phylogenetic tree, suggesting that these proteins might be functionally divergent (Supplemental Figure S1D). In contrast, THERM_00467910 was found within a well differentiated clade representing the Rrb1 protein family (Supplemental Figure S1D). We conclude that RebL1 and Wdc1 are related to the RBBP4/7-family of histone chaperones in *Tetrahymena*, whereas THERM_00467910 belongs to a phylogenetically distinct WD40 sub-family.

RebL1 interacts with diverse chromatin-associated protein complexes

To test the hypothesis that THERM_00688660 is a *bona fide* RBBP4/7 orthologue, we generated endogenously tagged *Tetrahymena* cell lines of two different mating types stably expressing *REBL1* with a C-terminal FZZ epitope tag from its native MAC locus (Figure 1C). Next, we performed AP-MS analysis in vegetative *Tetrahymena* cells using RebL1-FZZ as a bait. Application of SAINTexpress to the raw AP-MS data identified at least 14 high-confidence (FDR ≤ 0.01) interaction partners of RebL1

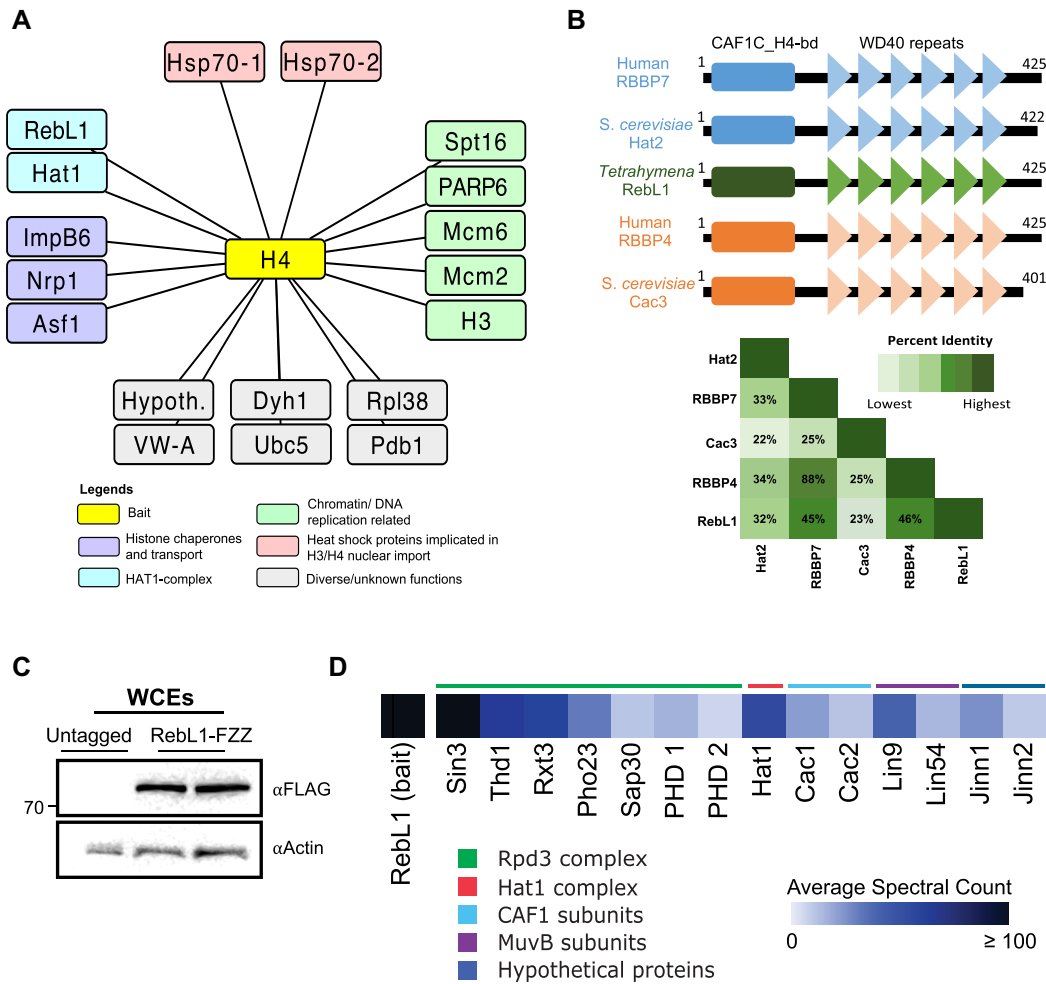


Figure 1. Identification and proteomic characterization of RebL1 in *Tetrahymena*. (A) Network representation of high confidence ($FDR \leq 0.01$) H4 co-purifying proteins. Note: Hypothetical (Hypoth.) is used to indicate proteins that lack annotation on the *Tetrahymena* Genome Database (TGD; www.ciliate.org). See Supplemental Table S1 for complete AP-MS results. (B) Comparative domain analysis of *Tetrahymena* RebL1 protein against human and budding yeast orthologues. Overall sequence identity among the orthologues is shown as half-square rectangle (bottom). (C) Western blotting analysis using whole cell lysates prepared from two different mating type *Tetrahymena* cells expressing RebL1-FZZ. The blot was probed with the indicated antibodies. (D) Heatmap of high confidence ($FDR \leq 0.01$) RebL1 co-purifying proteins from vegetatively growing *Tetrahymena*. Definition of the color code is indicated. See Supplemental Table S2 for complete AP-MS results.

during vegetative growth (Figure 1D; Supplemental Table S2). These interaction partners include the putative orthologues of *S. cerevisiae* Hat1, as well as THERM_00309890 and THERM_00219420, which shared sequence similarity with the human p150 and p60 (*S. cerevisiae* Cac1 and Cac2) subunits of the CAF1 complex, respectively (Figure 1D). These observations highlight the potentially conserved heterotrimeric and heterodimeric subunit composition of the CAF1- and HAT1-complexes, respectively. Next, we identified seven RebL1 co-purifying proteins, including Sin3, Thd1, Rxt3, Pho23, Sap30 and two PHD-containing proteins that, based on similarity with the *S. cerevisiae* orthologues, define a putative *Tetrahymena* Rpd3 histone deacetylase (HDAC) complex (Figure 1D). The high-confidence RebL1 interactions also included putative orthologues of the Lin54 (THERM_00766460) and Lin9 (THERM_00591560) subunits of the mammalian MuvB-complex (Figure 1D). An additional two unanno-

tated or hypothetical proteins, THERM_00227070 and THERM_00147500 (named as Jinn1 and Jinn2, respectively), co-purified with RebL1 and lacked any recognizable domains and/or sequence similarity to any known protein.

As RebL1 expression is highest after meiosis, a state that is marked by a series of rapid post-zygotic nuclear divisions (Supplemental figures S1E and S2A), we investigated whether its interactome was remodeled during sexual development of *Tetrahymena*. To address this, we performed AP-MS in conjugating cells harvested 5-hour post-mixing. After scoring the individual PPIs using SAINTexpress, we detected a total of 30 reproducible, high-confidence interaction partners of RebL1 during conjugation (Supplemental Table S3). All the RebL1 interaction partners identified during vegetative growth were also detected as high-confidence hits in the conjugating *Tetrahymena* (Figure 2A). Functional enrichment analysis of the 30 RebL1 interaction partners identified gene ontology (GO) terms enriched for

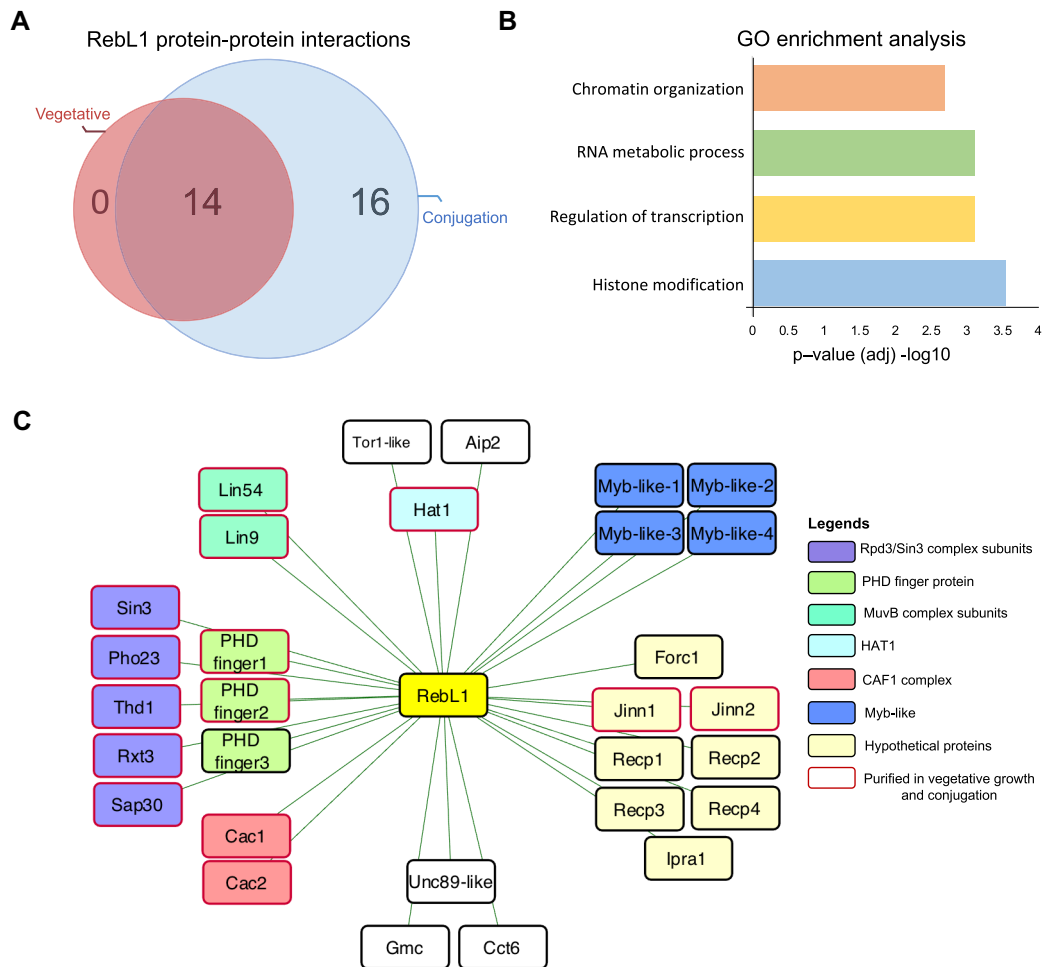


Figure 2. RebL1 has a conserved protein interaction network in *Tetrahymena*. (A) Venn diagram illustrating the overlap of high confidence (FDR ≤ 0.01) RebL1 co-purifying interaction partners during growth and conjugation (5h post-mixing the cells) in *Tetrahymena*. See Supplemental Tables S2-S3 for complete AP-MS results from vegetative and conjugating cells, respectively. (B) GO enrichment analysis of RebL1-FZZ co-purifying proteins. (C) Network view of RebL1-FZZ PPIs during conjugation (FDR ≤ 0.01). Bait node is shown in yellow and the co-purifying partners are grouped according to their putative complexes.

chromatin organization, transcription regulation and histone modifications (Figure 2B).

Among the RebL1 interaction partners identified exclusively during conjugation were four putative MYB-like transcription factors (TFs) (Figure 2C). Additionally, we identified THERM.00841280, which appears to be ciliate-specific and lacks sequence similarity to any known metazoan protein. The expression analysis of RebL1 co-purifying proteins using publicly available micro-array data revealed that THERM.00841280 is exclusively expressed during conjugation with a peak at early stages correlating with the onset of MIC meiosis (Supplemental Figure S2B) (85). We named THERM.00841280 as ‘Friend of RebL1 in conjugation’ (Forc1). Among the remaining conjugation-specific RebL1 co-purifying partners were a Tor1-like, Asf1-interacting protein 2 (Aip2) and several unannotated proteins, including Recp1-4 (RebL1 co-purifying protein 1-4) and Ipra1 (see below), without any recognizable domains (Figure 2C). We have previously shown that Aip1 and Aip2 are two highly similar ciliate-specific proteins that interact with Asf1 and likely function in chromatin assembly

pathways in *Tetrahymena* (67). We also detected Aip1 as a conjugation-specific interaction partner of RebL1, albeit at a slightly relaxed threshold (FDR ≤ 0.03) (Supplemental Table S3). The co-purification of Aip1/2 proteins that function with ImpB6 and Asf1 (67) reinforces the idea that RebL1 functions in multiple chromatin-related pathways in *Tetrahymena*, consistent with our GO enrichment analysis of the RebL1 interacting partners (Figure 2B).

RebL1 protein interaction profile is analogous to human RBBP4 and RBBP7

To examine the functional conservation of distantly related orthologues, we directly compared the protein interaction profiles of RebL1 and human RBBP4 and RBBP7. We generated inducible HEK293 cell lines expressing EGFP-tagged RBBP4 and RBBP7 and used them for anti-GFP AP-MS experiments. Using SAINTexpress, we found that RBBP4 and RBBP7 physically interact with multiple chromatin-related, as well as transcriptional regulatory, protein complexes (Supplemental Table S4,5), including the

CAF1-, Hat1-, MuvB- and Sin3A-complexes (FDR \leq 0.01; Figure 3A), as reported previously (98). However, 37.5% (18 proteins out of 48) of the high-confidence protein-protein interactions for RBBP4 and 29% (9 proteins out of 31) for RBBP7 had not been previously reported in the BioGRID database (99). While there are many shared interaction partners between the RBBP4 and RBBP7 homologues, non-overlapping distinct PPIs were also observed (Figure 3B). For example, consistent with previous studies (35), MuvB subunits co-purified with only RBBP4 and not with RBBP7 (Figure 3A). Furthermore, CAF1-complex subunits Chaf1A and Chaf1B co-purified with RBBP4 whereas Hat1 was identified exclusively in RBBP7 interaction partners. Importantly, we observed that many of the transcription- and chromatin-related complexes that we found for human RBBP4 and RBBP7 had their putative orthologous complexes in *Tetrahymena*. Examples of these conserved interaction partners include the Hat1-, CAF1-, and MuvB complexes (Figure 3C; Supplemental Figure S2C; Supplemental Table S3). While PRC2 subunits, including Suz12, Ezh2 and EED, were co-purified with RBBP4 and RBBP7 (Figure 3A), we did not identify any *Tetrahymena* orthologous subunits in our RebL1 AP-MS data. We conclude that *Tetrahymena* RebL1 physically interacts with multiple epigenetic regulatory complexes, and these functional links are conserved in humans.

RebL1, Lin9^{Tt} and Lin54^{Tt} are the conserved MuvB subunits

The identification of putative MuvB subunits in *Tetrahymena* piqued our interest due to the described role of mammalian MuvB as the master regulator of cell-cycle-dependent gene expression (33). To identify all the putative MuvB components, we used the human MuvB subunits as queries to search the *Tetrahymena* genome. Our analysis indicated that the *Tetrahymena* genome does not appear to encode any orthologues of Lin37 and Lin52. Lin9 appears to be present as a single gene copy in *Tetrahymena*, similar to humans. We identified at least 14 genes that appeared to encode the Lin54-like proteins in *Tetrahymena*. All the identified Lin54-like proteins contained two tandem TESMIN/TSO1-like CXC domains (Pfam: PF03638), consistent with the architecture of human Lin54 (39,40) (Supplemental Figure S3A). We next examined the expression profiles of the *Tetrahymena* Lin54-like genes during growth and conjugation using publicly available microarray data (85). Most of the Lin54-like genes were weakly expressed during the growth and late conjugation stages (6-hour and onwards after mixing the cells) (Supplemental Figure S3B). At least 9 out of the 14 genes, including the RebL1 interaction partner THERM.00766460, exhibited an expression peak at early conjugation stages (2–6 hours after mixing the cells). For example, the expression of THERM.00120820, THERM.00577120 and THERM.00600840 peaked at 2 hours, whereas THERM.00766460 had its highest expression at 4 hours after mixing the cells (Supplemental Figure S3B). These observations suggest that the expression of Lin54-like proteins is tightly regulated during early conjugation, a period when the MIC becomes transcriptionally active and undergoes meiosis (Supplemental Figure S1E).

We named Lin54-like proteins as ‘Anqa-1 to 14’ (after the ‘ANQA’ mythological creature thought to appear once in ages), considering that several of them exhibited an expression peak only once at a distinct conjugation stage.

We generated endogenously tagged Lin9^{Tt}-FZZ and Anqa1-FZZ (THERM.00766460) cell lines and employed them for AP-MS (Supplemental Figure S3C). These analyses showed reciprocal purification of RebL1 and Anqa1 when Lin9^{Tt}-FZZ was employed as the bait protein (Figure 4A; Supplemental Table S6). Similarly, Lin9^{Tt} and RebL1 were identified as Anqa1-FZZ high-confidence co-purifying partners (Supplemental Table S7). These results are consistent with the idea that RebL1, Anqa1 and Lin9^{Tt} represent the conserved subunits of a putative MuvB complex in *Tetrahymena*. Two unannotated proteins, which we named as Jinn1 and Jinn2 (see above), co-purified with both Lin9^{Tt}-FZZ and Anqa1-FZZ as high-confidence hits (Figure 4A). Based on their expression profiles (Supplemental Figure S2A) and co-purification with all three conserved subunits (RebL1, Lin9^{Tt} and Anqa1) (Figure 4B), we suggest that Jinn1/2 represent *Tetrahymena*-specific putative MuvB components.

In addition to functionally diverse proteins such as kinases and helicases, THERM.00317260 and THERM.00842380 also co-purified with Anqa1-FZZ during vegetative growth (FDR \leq 0.01). THERM.00317260, which interacted with RebL1 during conjugation, is a *Tetrahymena*-specific protein and lacks any recognizable domains. We named it Ipra1 (interaction partner of RebL1 and Anqa1). THERM.00842380 is the putative MYB-like TF (MybL1) that was also detected as a high-confidence RebL1 interaction partner during both growth and conjugation (FDR \leq 0.01) (Figure 2C). This observation is consistent with the reported interaction between mammalian MuvB and B-MYB TF to form an activator MMB complex (33).

We next identified Anqa1 PPIs in conjugating *Tetrahymena* cells (5 hours post mixing the cells). Anqa1 co-purified with 20 high-confidence partners, of which 8 were common to both vegetative growth and conjugation (Figure 4C; Supplemental Table S8). Among shared interactions were Lin9, RebL1, Jinn1, Jinn2, MybL1 and Ipra1, suggesting that MuvB and its derivative complexes remain functional during *Tetrahymena* development. Conjugation-specific interaction partners include several transcription regulatory proteins including RNA polymerase subunit Rpc19 (THERM.00011760), transcription factors TFIID (THERM.00160910) and TFIID subunit 5 (THERM.00227160), transcriptional regulatory protein Spt6 (THERM.01093550) and small nuclear ribonucleoprotein Sm (THERM.01289080). These results suggest that MuvB functions in transcription regulation during *Tetrahymena* development.

Genome-wide analysis reveals distinct RebL1 and Anqa1 DNA-binding profiles

Among the human MuvB subunits, RBBP4 and Lin54 are thought to bind DNA, with Lin54 exhibiting sequence specificity by recognising the CHR element (TTTGAA)

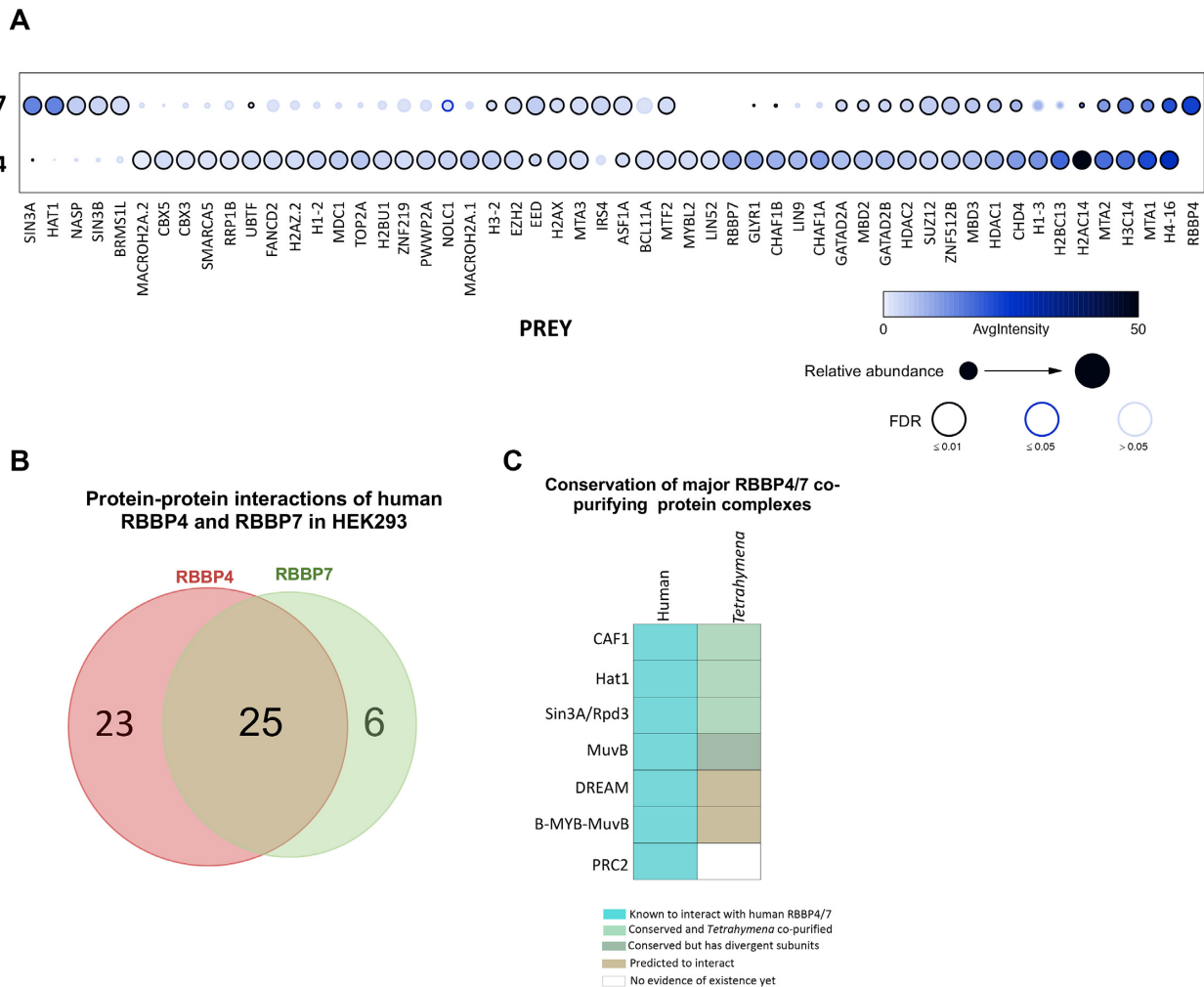


Figure 3. Proteomic analysis of human RBBP4 and RBBP7. (A) Dot plot representation of the interaction partners identified with RBBP4 and RBBP7 in HEK293 cells. Inner circle color shows the average spectral count, the circle size indicates the relative prey abundance, and the circle outer edge is the SAINT FDR. (B) Venn diagram of the overlap of PPIs for human RBBP4 and RBBP7 proteins (FDR \leq 0.01) from HEK293 cells. See Supplemental Tables S4 and S5 for complete AP-MS results for RBBP4 and RBBP7, respectively. (C) Schematic illustration of evolutionary conservation for the RebL1 co-purifying putative chromatin-related complexes that are also known to interact with human RBBP4 and RBBP7 proteins. The legend for the colors is provided.

(33). To investigate their genome-wide localization in *Tetrahymena*, we utilized our endogenously tagged RebL1 and Anqa1-FZZ cell lines and performed chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) experiments in vegetatively growing cells (Supplemental Figure S4A). ChIP-seq analysis revealed that RebL1 predominantly localizes to genic regions with 64% of the peaks being found within exons (Supplemental Figure S4B, left; Supplemental Table S9; 5531 total peaks; q value $<$ 0.05). Consistent with the role of RebL1 in multiple chromatin-related processes, metagene analysis indicated that RebL1 binds throughout the gene body of its targets without any specific sequence preferences (Supplemental Figure S4B, right).

Unlike RebL1, Anqa1 metagene analysis indicated a clear enrichment near the transcription start sites (TSS) and transcript end sites (TES) of the targeted genes (Figure 5A). Peak distribution analysis revealed that roughly half of the

Anqa1 ChIP peaks are found within 1 kilobase of the annotated TSS (Figure 5A; Supplemental Table S10). As an enrichment near the promoter regions suggests a role in transcription regulation, we utilized publicly available RNA-seq data that has been used to rank *Tetrahymena* genes based on their expression patterns during vegetative growth (66) and correlated it with the Anqa1 ChIP data. We found that the majority of the Anqa1-FZZ bound genes (290 of 422 genes) were moderately to very weakly expressed and only 12% of the targets were classified as highly expressed during *Tetrahymena* growth (Figure 5B). These results suggest that Anqa1 may function to regulate the expression of its target genes.

We carried out gene ontology (GO) enrichment analysis related to molecular functions and/or biological processes. While RebL1-binding genes were not enriched in any GO terms, in keeping with its role in multiple protein complexes, Anqa1 targets were however enriched for terms asso-

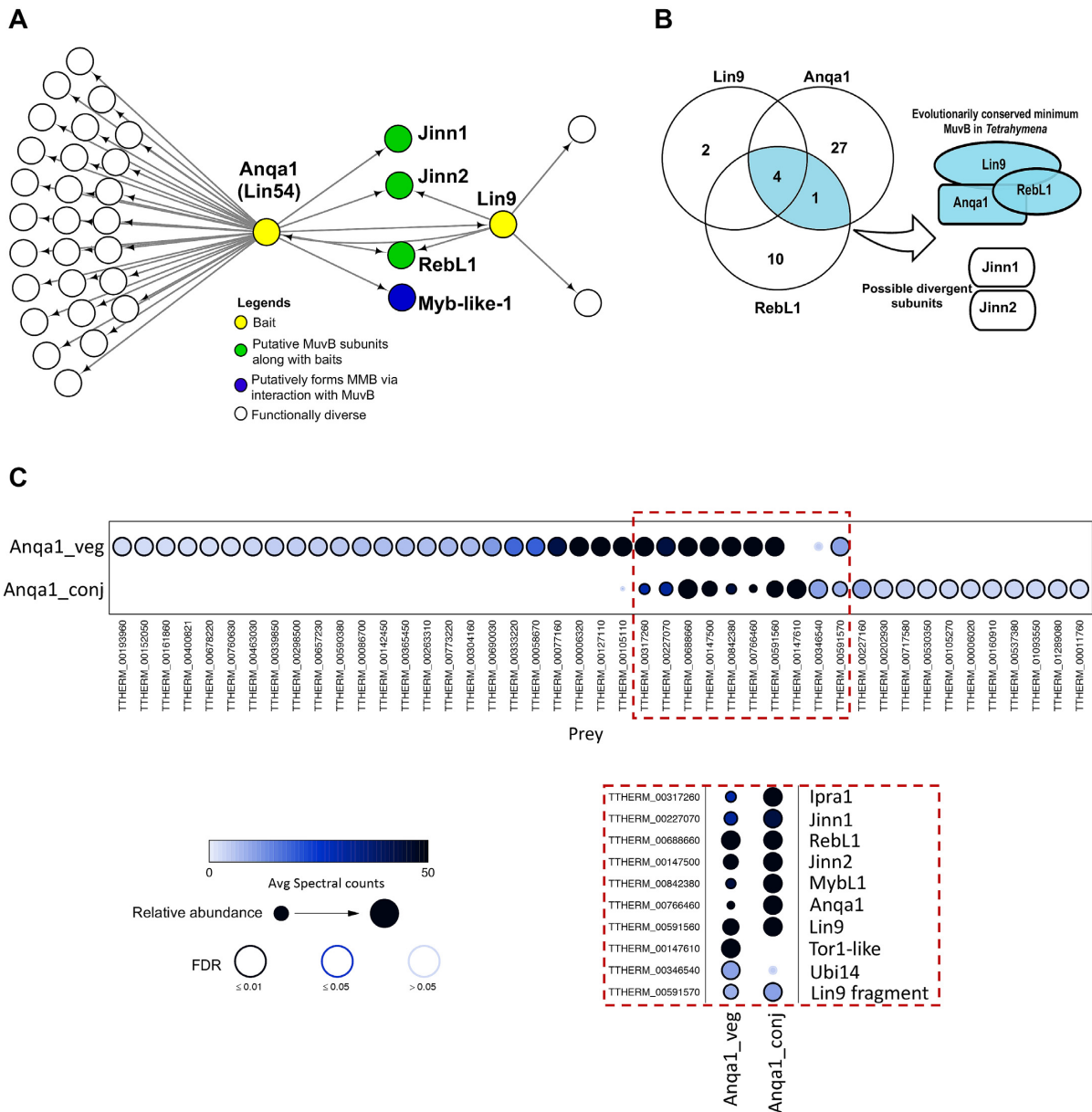


Figure 4. Proteomic analysis of putative MuvB subunits in *Tetrahymena*. (A) Network depiction of Lin9 and Anqa1 (Lin54) protein-protein interactions (FDR ≤ 0.01). Bait nodes are shown in yellow. See Supplemental Tables S6 and S7 for complete AP-MS results for Lin9 and Anqa1, respectively. (B) Venn diagram indicating the overlapping PPIs of RebL1, Lin9 and Anqa1 during vegetative growth. (C) Dot plot representation of the interaction partners identified for Anqa1 during vegetative growth and conjugating *Tetrahymena*. Inner circle color shows the average spectral count, the circle size indicates the relative prey abundance, and the circle outer edge is the SAINT FDR. Red box shows enlarged view of overlapping interaction partners. See Supplemental Table S8 for complete AP-MS results for Anqa1 during conjugation.

ciated with gene expression regulation, chromatin assembly and organization, translation regulation, ribosome biogenesis and cellular metabolic processes, suggesting a specific function in regulation of cellular processes (Figure 5C).

A hybrid E2F/DP and Anqa1 DNA-binding motif

The ChIP-seq analysis of Lin54 in metazoans has revealed the existence of a hybrid DREAM-recruiting motif representing the E2F-binding sequence upstream of the CHR element (44–46). To identify a putative DNA-binding motif, we performed a *de novo* motif discovery analysis using the

Anqa1 ChIP peaks. The identified Anqa1 motif appeared to share similarity with the characteristic DREAM-recruiting E2F/DP-CHR hybrid motif (Figure 5D). By scanning across the *Tetrahymena* genome, we identified ~ 3000 loci, with roughly half (1345 loci) being found within ± 1 kb of TSS, that contain this putative DREAM-recruiting motif (Supplemental Table S11). We cannot rule out the possibility of additional DREAM-recruiting loci given that the *Tetrahymena* genome encodes multiple Lin54- and E2F-like factors which could potentially recognize sequences with variations in the identified motif. The *Tetrahymena* genome

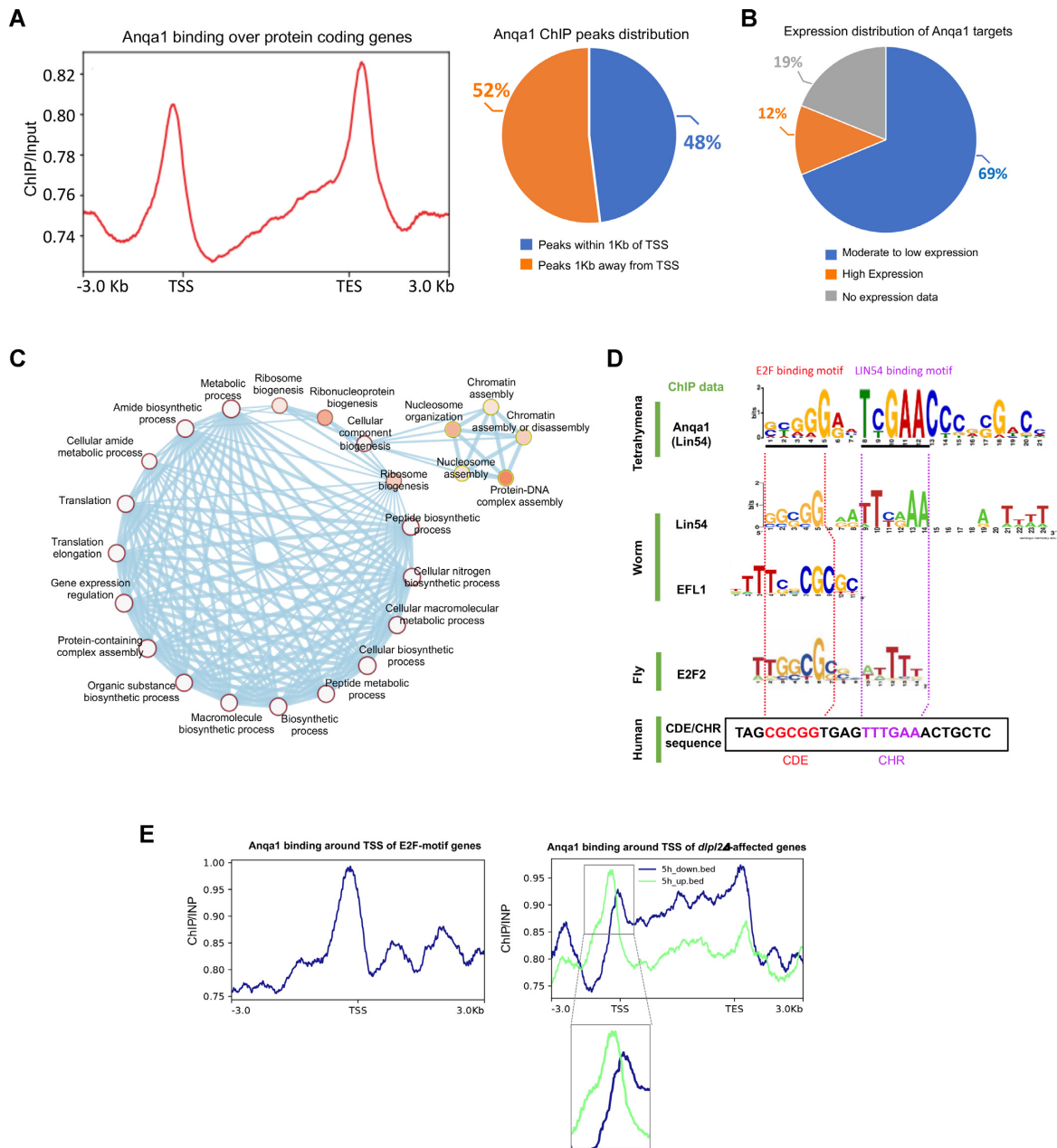


Figure 5. Genome-wide analysis of Anqa1 DNA-binding profile in *Tetrahymena*. (A) Left: Metagenesis of Anqa1 ChIP-seq data indicates enrichment near the TSS and TES of target genes. Right: Anqa1 ChIP peak distribution with respect to annotated TSS \pm 1 kb. (B) Pie chart showing the distribution of Anqa1 bound genes based on their expression levels. (C) Network depiction of gene ontology (GO) enrichment analysis among Anqa1 bound targets. Each node depicts a GO term whereas each edge represents the degree of gene overlap that is found between two gene sets corresponding to the two GO terms. GO term significance 0.01 (white) to 0 (orange). (D) An overrepresented motif in Anqa1 bound *Tetrahymena* genes identified by MEME (E value: $8.0e-223$). Below are shown previously identified motifs for the *C. elegans* EFL-1, an extended motif enriched among E2F2, LIN-9 and LIN-54 co-regulated genes in *Drosophila* and the CDE/CHR motif from the human Cdc2 promoter. Regions bound by human E2F4 and LIN54 at the Cdc2 promoter as well as homologous motif sequences in other organisms are highlighted by the dotted lines. (E) Metagenesis analyses indicating the binding profile of Anqa1 around TSS of those genes that 1) contain the E2FL1 motif in their promoters (left), 2) those genes whose expression is affected during conjugation upon DPL2 depletion (right).

encodes at least seven E2F family members (E2FL1-4 and DPL1-3) and two p107/p130 (Retinoblastoma-like1/2)-like proteins (86). Previous *in silico* analysis has identified 714 *Tetrahymena* genes to contain some version of human E2F DNA-binding motif (TTTSSCGC where S is either a G or a C) in their promoters (100). Although we did not detect any DREAM-specific proteins in our AP-MS (FDR \leq 0.01), we found that Anq1 exhibited enriched binding in the promoter regions of putative E2F motif containing genes (Figure 5E, left), consistent with the identification of an E2F/DP-CHR-like hybrid motif.

E2FL1 and Dpl2 heterodimer has been described to function in gene expression regulation during conjugation (86). To further explore the connection between Anq1 and E2Fs, we utilized previously generated RNA-seq data in *dpl2* Δ cells during conjugation (86), and examined the Anq1 binding profile around the TSS of affected genes. Our analyses indicated that Anq1 binds relatively evenly across the gene bodies of both up- and down-regulated genes at 3-, 4-, 6- and 7-hour time points after mixing the *dpl2* Δ cells during conjugation (Supplemental Figure S4C). Remarkably, however, we observed a strong binding enrichment for Anq1 specifically within the promoter regions of genes that are upregulated in *dpl2* Δ cells 5-hours post mixing cells (Figure 5E, right), a period in conjugation immediately prior to the observed meiotic defect in *dpl2* Δ cells (86). In contrast, the downregulated genes had wide-spread binding within their bodies without any enrichment near the TSS (Figure 5E, right). These *dpl2* Δ -dependent upregulated genes were enriched in pathways such as DNA replication and DNA repair (Supplemental Figure S4C). We suggest that the putative MuvB subunit Anq1 binds in the promoter region, possibly in conjunction with DREAM-specific E2F/DP proteins, to regulate the expression of some of its target genes.

RebL1 depletion results in growth defects

Previous studies have correlated the overexpression of RBBP4 and RBBP7 with tumor cell proliferation in certain cancer types (49,54). To analyze the RBBP4 and RBBP7 expression levels and correlation with cancer prognosis, we utilized archived patient RNA-seq data from 'The Cancer Genome Atlas' (TCGA), focusing on multiple tumor types for which the corresponding normal tissue data were available. The expression levels of both RBBP4 and RBBP7 were significantly higher in several cancers, including stomach adenocarcinoma, breast cancer, glioblastoma multiforme, cholangio carcinoma and liver hepatocellular carcinoma, compared with the normal tissues (P -value cutoff \leq 0.01) (Supplemental Figure S5A; Supplemental Table S12 for TCGA abbreviations). Furthermore, these high expression levels of RBBP4 and RBBP7 significantly correlated with unfavorable clinical outcomes ($P \leq$ 0.05) in many tumor types (Supplemental Figure S5B). For example, Kaplan–Meier curves indicated that hepatocellular carcinoma (LIHC) high expression groups for both RBBP4 and RBBP7 had significantly worse overall survival ($P <$ 0.01) compared with low expression groups (Figure 6A). The observed overall poor survival of LIHC high-expression group was further enhanced when the RBBP4

and RBBP7 expression was combined in our analyses, suggesting an additive effect (Figure 6B).

We examined the DepMap database, which encompasses data from genome-wide RNAi and CRISPR loss-of-function screens across hundreds of cancer cell lines, and observed that RBBP4 behaves as a common essential gene (Supplemental Figure S5C) (101,102). RBBP7, on the other hand, exhibited high selectivity in terms of cancer cell line dependency, reinforcing the concept of orthologue-specific functions for RBBP4 and RBBP7. Considering the sequence/structural similarity of RebL1 with its human orthologues and formation of similar functional links (Supplemental Figure S5D), we explored the role of RBBP4 and seven in cellular proliferation and/or gene expression using *Tetrahymena* as a model system. To this end, a cadmium inducible *RebL1* RNA interference (RNAi) construct was introduced into vegetative *Tetrahymena* cells. Upon CdCl₂ treatment, the RebL1 expression levels were significantly reduced in *RebL1*-RNAi cells in comparison with the WT and/or uninduced *Tetrahymena* (Figure 6C), as assessed by reverse-transcription quantitative real-time PCR (RT-qPCR) analysis. We used growth assays to examine the functional consequences of RebL1 depletion in growing *Tetrahymena* cells. We found that *RebL1*-RNAi cells gave ~19% viable clones, significantly fewer compared to 90% viable cells ($n = 5$) in wildtype *Tetrahymena* when treated with CdCl₂ (0.5 μ g/mL; see methods) (Figure 6D), consistent with the role of its human orthologues in cell proliferation.

RebL1 depletion causes gene expression changes

The *REBL1*-RNAi dependent reduction in cellular viability could be related to indirect changes in expression levels of some functionally important genes. In human glioblastoma cells (GBMs), RBBP4 depletion results in sensitization of tumor cells to temozolomide (TMZ) in conjunction with the downregulation of DNA repair protein Rad51 (54). We examined Rad51^{Tt} expression levels in *REBL1*-RNAi cells. Rad51^{Tt} was significantly suppressed in *Tetrahymena* depleted of RebL1 in comparison to the WT and/or uninduced cells (Figure 6E).

Previous studies have identified an auto-regulatory feedback loop for the DREAM complex (103). To test the possibility of an auto-regulatory feedback loop for MuvB, we examined the expression levels of Anq1 and Lin9^{Tt} using the RNAs derived from either the *REBL1*-RNAi or WT cells. We found that *REBL1* knockdown suppressed the expression of both Anq1 and Lin9^{Tt} (Figure 6E), consistent with the disruption of a putative *Tetrahymena* MuvB. Consistent with these data, we observed that expression levels of Lin54 and Lin9 positively correlate with RBBP4 in multiple tumor types (Supplemental Figure S6). Furthermore, the high expression of MuvB subunits correlate with significantly worse overall survival in many different cancer types ($P \leq$ 0.05), particularly in LIHC patients where increased levels of 3 out of the 5 MuvB subunits were associated with unfavourable outcomes (Figure 6F). These results are consistent with the highly conserved physical interactions of MuvB subunits with each other.

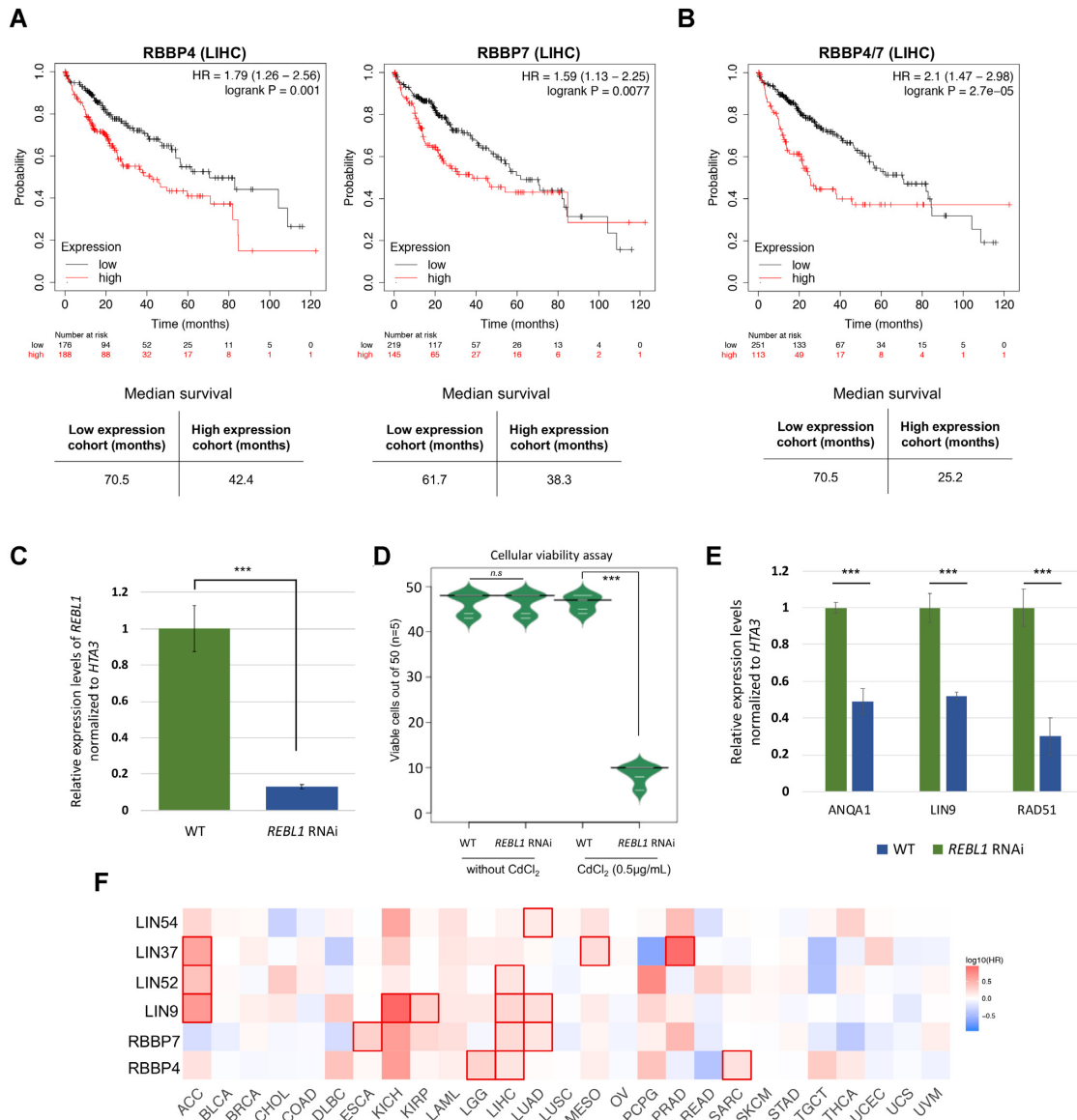


Figure 6. Conserved oncogenic properties of RBBP4/7. (A) Kaplan–Meier curves of overall survival in patients with primary hepatocellular carcinoma using TCGA-LIHC. Patients were grouped according to median expression of either RBBP4 (left) or RBBP7 (right) using log-rank tests. Bottom panels show the median survival in months for high and low expression groups. (B) Kaplan–Meier curves of overall survival in LIHC patients using combined expression of RBBP4 and RBBP7 to examine the survival rates in the TCGA-LIHC cohort. Log-rank test was used in this analysis and the bottom panel depicts the median survival in months for high and low expression groups. (C) Expression levels of *Reb1* in *REB1* RNAi *Tetrahymena* cells treated with CdCl₂ (0.5 µg/mL) or untreated, as assessed by qRT-PCR. The qRT-PCR data were normalized to *HTA3* expression. *** indicates *t*-test $P \leq 0.001$. The individual error bars indicate the standard deviation for each sample. (D) Bean plot representation of cellular viability of single cells wild-type or *REB1*-RNAi treated lines isolated in drops of media with and without CdCl₂ (0.5 µg/mL). Cell growth was observed after at least 48 hours to a maximum of 72 hours. The number of viable drops were counted from each replicate ($n = 5$; 50 cells in each replicate) and plotted. *** indicates *t*-test $P \leq 0.001$ and n.s.: non-significant. (E) Analysis of expression levels of target genes in *REB1*-RNAi cells treated with CdCl₂ (0.5 µg/mL) or untreated. The qRT-PCR data were normalized to *HTA3* expression. Statistical significance was assessed using Student's *t*-test (** $P \leq 0.01$, * $P \leq 0.05$, n.s.: non-significant). Error bars represent standard deviation. (F) Heat map showing the hazard ratios in logarithmic scale (\log_{10}) for the indicated genes in multiple cancer types. The red blocks denote higher whereas blue blocks indicate lower clinical risk, with an increase in the gene expression. The blocks with darkened frames denote statistical significance in prognostic analyses ($P \leq 0.05$; Mantel–Cox test). The TCGA cancer abbreviations are provided in Supplemental Table S12.

DISCUSSION

Effective transcriptional regulation requires the ordered assembly of multiple protein complexes on chromatin in response to specific cellular cues. In this study, we report that *Tetrahymena* contains at least three WD40 repeat family proteins with putative scaffolding roles. Specifically, we determined THERM_00688660 (RebL1) to be the major structurally and functionally conserved orthologue of human RBBP4 and 7 in *Tetrahymena*. We demonstrated that RebL1 is a component of multiple chromatin-related complexes and is important for gene expression regulation and cell growth.

Previous studies have indicated that the structure/function of various MuvB-derived complexes can differ across different species (33). For example, structurally and functionally distinct MuvB complexes in *Drosophila* salivary glands and testis have been identified (104–106). It has remained an open question, however, whether tissue-specific or development-specific MuvB complexes also exist in other species. In *Tetrahymena*, the presence of multiple Lin54 paralogues and their distinct expression pattern suggest the existence of multiple MuvB-like complexes. Although further studies will explore this interesting possibility, a previous report has shown the interaction between E2FL1/Dpl2 and Anq14 (86) supporting the existence of multiple structurally/functionally distinct MuvB-containing DREAM-like complexes in *Tetrahymena*. Several observations reported in our study indirectly support the formation of a DREAM-like complex in *Tetrahymena*: (i) the identification of a putative E2F/DP-CHR-like DNA motif; (ii) the enrichment of Anq1 ChIP peaks in the promoter regions of putative E2F target genes in *Tetrahymena* and (iii) the targeting of *dpl2*Δ-dependent upregulated genes by Anq1. We propose that *Tetrahymena* MuvB functions in collaboration with E2F/DPs to form a DREAM-like complex that negatively regulates the expression of target genes. Although E2F DNA-binding motifs are largely conserved during evolution (107), CHR elements have not yet been studied in *Tetrahymena*. Further studies are required to thoroughly examine the role of CHR and other *cis* elements in *Tetrahymena* cell cycle regulation.

The role of RBBP4 and RBBP7 in cellular proliferation has been widely reported, although the mechanistic details have remained unclear. Our results suggest that one of the mechanisms through which RBBP4 and RBBP7 might function in cellular growth is via regulating the cell-cycle regulatory complex MuvB. Consistent with this, we found that the expression profiles of MuvB subunits are interconnected in tumor cells. In previous reports, autoregulatory feedback loops have been reported involving B-Myb and the DREAM complex (103). It will be interesting to examine whether B-Myb also functions in the MuvB autoregulation pathway. Considering that MuvB subunits were identified as biomarkers of poor prognosis for multiple cancers, our mechanistic findings have direct implications for better understanding tumor cell proliferation.

RBBP4 and RBBP7 function in multiple epigenetic complexes which might have a role in cell proliferation independent of MuvB. For example, knockdown of RBBP4 in

GBM tumor cells results in suppression of the DNA-repair protein Rad51 due to disruption of the RBBP4/p300/CBP chromatin-modifying complex (54). Loss of RBBP4 in GBMs was found to be associated with suppression of tumor growth (54). Consistently, we observed that RebL1 depletion resulted in suppression of the Rad51^{Tt} orthologue in *Tetrahymena*. Disruption of *Tetrahymena* *RAD51^{Tt}* has been reported to lead to severe abnormalities during vegetative growth, as well as in conjugation (108,109). It remains to be seen whether Rad51^{Tt} also has a role in the reduction of cellular viability observed in *REBL1*-RNAi treated *Tetrahymena* cells. In summary, our research has extended the current understanding of transcription regulation in ciliates and, more broadly, on the multifaceted role(s) of RBBP4 and RBBP7 proteins in eukaryotes.

DATA AVAILABILITY

The mass spectrometry raw data reported in this paper are deposited at MassIVE (<https://massive.ucsd.edu/>). Additional files include the complete SAINTexpress outputs for each dataset as well as a “README” file that describes the details of mass spectrometry files deposition to the MassIVE repository (Supplemental Table S14 for accession numbers). ChIP-seq data generated can be found online at Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). NGS and reads files produced in this study were deposited at GEO database with unique identifier GSE156091.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Anita Samardzic for her technical assistance with *Tetrahymena* growth. We thank Guoqing Zhong for her technical help with HEK293 cell lines and related MS experiments. The authors also wish to thank Dr. Anne-Claude Gingras at the Network Biology Collaborative Centre, Lunenfeld-Tannenbaum Research Institute, Mt. Sinai Hospital, Toronto, ON, Canada for access to mass spectrometers for a portion of the *Tetrahymena* MS analyses. Tanja Durbin and team members at the Donnelly Sequencing Centre are gratefully acknowledged for their assistance with next-generation sequencing.

Authors' contribution: S.N.-S. and J.F. jointly conceived the project. S.N.-S. generated cell lines, performed western blots and affinity purifications, conducted TCGA analysis, ChIP data analysis and evolutionary studies, AP-MS of HEK293 cells, and wrote the manuscript. J.G. generated cell lines, participated in affinity purifications, performed IF microscopy and RNAi experiments with contribution from J. Fine. A.S. performed ChIP-seq and participated in affinity purifications with contribution from S.W. and J.D. K.A. generated cell lines, performed affinity purifications, and participated in TCGA analysis. H.L. and M.P. carried out ChIP data analysis. N.A. performed qPCR experiments and participated in figure preparation. S.P. and E.M. contributed to MS analysis related to

HEK293 cell lines. Z.Z. supervised H.L. J.F.G. edited the manuscript, participated in analyses, and provided reagents. R.P. edited the manuscript, supervised trainees, and provided reagents. J.-P.L. performed all *Tetrahymena* MS analyses, contributed to data analysis, partly supervised the project and edited the manuscript. J.F. was responsible for study design, manuscript editing, data analyses, project supervision and provided reagents. All authors have approved the final manuscript.

FUNDING

Fillingham and Lambert laboratories were supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants [RGPIN-2020-07036, RGPIN-2017-06124]; Fillingham laboratory was also supported by a grant from the Ryerson University Health Research Fund and the Ryerson University Faculty of Science Dean's Research Fund; J.-P.L. holds a Junior 1 salary award from the Fonds de Recherche du Québec-Santé (FRQ-S) and was also supported through a John R. Evans Leaders Fund from the Canada Foundation for Innovation [37454, 41426]; Pearlman laboratory was supported by Canadian Institutes of Health Research (CIHR) [MOP13347]; NSERC Discovery Grant [539509]; Greenblatt laboratory was supported by the CIHR Foundation Grant [FDN-154338]; H.L. was supported by Canada Graduate Scholarship from NSERC. Funding for open access charge: NSERC Discovery Grants of Lambert and Fillingham.

Conflict of interest statement. None declared.

REFERENCES

- Venkatesh, S. and Workman, J.L. (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.*, **16**, 178–189.
- Mendiratta, S., Gatto, A. and Almouzni, G. (2018) Histone supply: multitiered regulation ensures chromatin dynamics throughout the cell cycle. *J. Cell Biol.*, **218**, 39–54.
- Grover, P., Asa, J.S. and Campos, E.I. (2018) H3–H4 histone chaperone pathways. *Annu. Rev. Genet.*, **52**, 109–130.
- Campos, E.I., Fillingham, J., Li, G., Zheng, H., Voigt, P., Kuo, W.-H.W., Seepany, H., Gao, Z., Day, L. a, Greenblatt, J.F. *et al.* (2010) The program for processing newly synthesized histones H3.1 and H4. *Nat. Struct. Mol. Biol.*, **17**, 1343–1351.
- Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X. *et al.* (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, **140**, 678–691.
- Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N.A., Adams, P.D., Jansen, L.E.T. *et al.* (2011) Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol. Cell*, **44**, 928–941.
- Clapier, C.R., Iwasa, J., Cairns, B.R. and Peterson, C.L. (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.*, **18**, 407–422.
- Ahringer, J. (2000) NuRD and SIN3: Histone deacetylase complexes in development. *Trends Genet.*, **16**, 351–356.
- Hoffmann, A. and Spengler, D. (2019) Chromatin remodeling complex NuRD in neurodevelopment and neurodevelopmental disorders. *Front. Genet.*, **10**, 682.
- Lawrence, M., Daujat, S. and Schneider, R. (2016) Lateral thinking: how histone modifications regulate gene expression. *Trends Genet.*, **32**, 42–56.
- Zhang, H., Han, J., Kang, B., Burgess, R. and Zhang, Z. (2012) Human histone acetyltransferase 1 protein preferentially acetylates H4 histone molecules in H3.1-H4 over H3.3-H4. *J. Biol. Chem.*, **287**, 6573–6581.
- Barman, H.K., Takami, Y., Nishijima, H., Shibahara, K., Sanematsu, F. and Nakayama, T. (2008) Histone acetyltransferase-1 regulates integrity of cytosolic histone H3-H4 containing complex. *Biochem. Biophys. Res. Commun.*, **373**, 624–630.
- Ruiz-García, A.B., Sendra, R., Galiana, M., Pamblanco, M., Pérez-Ortín, J.E. and Tordera, V. (1998) HAT1 and HAT2 proteins are components of a yeast nuclear histone acetyltransferase enzyme specific for free histone H4. *J. Biol. Chem.*, **273**, 12599–12605.
- Poveda, A., Pamblanco, M., Tafrov, S., Tordera, V., Sternglanz, R. and Sendra, R. (2004) Hif1 is a component of yeast histone acetyltransferase B, a complex mainly localized in the nucleus. *J. Biol. Chem.*, **279**, 16033–16043.
- Sobel, R.E., Cook, R.G., Perry, C.A., Annunziato, A.T. and Allis, C.D. (1995) Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 1237–1241.
- Murzina, N. V., Pei, X.-Y., Zhang, W., Sparkes, M., Vicente-García, J., Prapat, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F. *et al.* (2008) Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure*, **16**, 1077–1085.
- Qian, Y.W., Wang, Y.C.J., Hollingsworth, R.E., Jones, D., Ling, N. and Lee, E.Y.H.P. (1993) A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. *Nature*, **364**, 648–652.
- Huang, S., Lee, W.H. and Lee, E.Y.H.P. (1991) A cellular protein that competes with SV40 T antigen for binding to the retinoblastoma gene product. *Nature*, **350**, 160–162.
- Xu, C. and Min, J. (2011) Structure and function of WD40 domain proteins. *Protein Cell*, **2**, 202–214.
- Song, J.-J., Garlick, J.D. and Kingston, R.E. (2008) Structural basis of histone H4 recognition by p55. *Genes Dev.*, **22**, 1313–1318.
- Martínez-Balbás, M.A., Tsukiyama, T., Gdula, D. and Wu, C. (1998) Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 132–137.
- Lu, X. and Horvitz, H.R. (1998) Lin-35 and lin-53, two genes that antagonize a C. elegans Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. *Cell*, **95**, 981–991.
- Reed Moody, R., Lo, M.C., Meagher, J.L., Lin, C.C., Stevers, N.O., Tinsley, S.L., Jung, I., Matvekas, A., Stuckey, J.A. and Sun, D. (2018) Probing the interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. *J. Biol. Chem.*, **293**, 2125–2136.
- Abbey, M., Trush, V., Gibson, E. and Vedadi, M. (2018) Targeting human retinoblastoma binding protein 4 (RBBP4) and 7 (RBBP7). bioRxiv doi: <https://doi.org/10.1101/303537>, 18 April 2018, preprint: not peer reviewed.
- Millard, C.J., Varma, N., Saleh, A., Morris, K., Watson, P.J., Bottrill, A.R., Fairall, L., Smith, C.J. and Schwabe, J.W.R. (2016) The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *Elife*, **5**, e13941.
- Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. and Reinberg, D. (1997) Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell*, **89**, 357–364.
- Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R. and Nogales, E. (2012) Molecular architecture of human polycomb repressive complex 2. *Elife*, **1**, e00005.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. and Reinberg, D. (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein. *Genes Dev.*, **16**, 2893–2905.
- Parthun, M.R., Widom, J. and Gottschling, D.E. (1996) The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell*, **87**, 85–94.
- Verreault, A., Kaufman, P.D., Kobayashi, R. and Stillman, B. (1996) Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell*, **87**, 95–104.
- Verreault, A., Kaufman, P.D., Kobayashi, R. and Stillman, B. (1998) Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr. Biol.*, **8**, 96–108.

32. Suganuma, T., Pattenden, S.G. and Workman, J.L. (2008) Diverse functions of WD40 repeat proteins in histone recognition. *Genes Dev.*, **22**, 1265–1268.
33. Fischer, M. and Müller, G.A. (2017) Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Crit. Rev. Biochem. Mol. Biol.*, **52**, 638–662.
34. Pilkinton, M., Sandoval, R. and Colamonici, O.R. (2007) Mammalian Mip/LIN-9 interacts with either the p107, p130/E2F4 repressor complex or B-Myb in a cell cycle-phase-dependent context distinct from the Drosophila dREAM complex. *Oncogene*, **26**, 7535–7543.
35. Litovchick, L., Sadasivam, S., Florens, L., Zhu, X., Swanson, S.K., Velmurugan, S., Chen, R., Washburn, M.P., Liu, X.S. and DeCaprio, J.A. (2007) Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol. Cell*, **26**, 539–551.
36. Sadasivam, S. and DeCaprio, J.A. (2013) The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer*, **13**, 585–595.
37. Jiang, J., Benson, E., Bausek, N., Doggett, K. and White-Cooper, H. (2007) Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the Drosophila male germline and physically interacts with always early. *Development*, **134**, 1549–1559.
38. Sadasivam, S., Duan, S. and DeCaprio, J.A. (2012) The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev.*, **26**, 474–489.
39. Andersen, S.U., Algreen-Petersen, R.G., Hoedl, M., Jurkiewicz, A., Cvitanich, C., Braunschweig, U., Schausser, L., Oh, S.-A., Twell, D. and Jensen, E.Ø. (2007) The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc in vitro and TSO1 is required for both male and female fertility in *Arabidopsis thaliana*. *J. Exp. Bot.*, **58**, 3657–3670.
40. Sugihara, T., Wadhwa, R., Kaul, S.C. and Mitsui, Y. (1999) A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation. *Genomics*, **57**, 130–136.
41. Marceau, A.H., Felthousen, J.G., Goetsch, P.D., Iness, A.N., Lee, H.W., Tripathi, S.M., Strome, S., Litovchick, L. and Rubin, S.M. (2016) Structural basis for LIN54 recognition of CHR elements in cell cycle-regulated promoters. *Nat. Commun.*, **7**, 12301.
42. Schmit, F., Cremer, S. and Gaubatz, S. (2009) LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdc2 promoter in a sequence-specific manner. *FEBS J.*, **276**, 5703–5716.
43. Müller, G.A. and Engeland, K. (2010) The central role of CDE/CHR promoter elements in the regulation of cell cycle-dependent gene transcription: review article. *FEBS J.*, **277**, 877–893.
44. Tabuchi, T.M., Deplancke, B., Osato, N., Zhu, L.J., Barrasa, M.I., Harrison, M.M., Horvitz, H.R., Walhout, A.J.M. and Hagstrom, K.A. (2011) Chromosome-biased binding and gene regulation by the *Caenorhabditis elegans* DRM complex. *PLoS Genet.*, **7**, e1002074.
45. Goetsch, P.D., Garrigues, J.M. and Strome, S. (2017) Loss of the *Caenorhabditis elegans* pocket protein LIN-35 reveals MuvB's innate function as the repressor of DREAM target genes. *PLoS Genet.*, **13**, e1007088.
46. Georlette, D., Ahn, S., MacAlpine, D.M., Cheung, E., Lewis, P.W., Beall, E.L., Bell, S.P., Speed, T., Manak, J.R. and Botchan, M.R. (2007) Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb-MuvB/dREAM complex in proliferating cells. *Genes Dev.*, **21**, 2880–2896.
47. Taylor-Harding, B., Binné, U.K., Korenjak, M., Brehm, A. and Dyson, N.J. (2004) p55, the Drosophila ortholog of RbAp46/RbAp48, is required for the repression of dE2F2/RBF-regulated genes. *Mol. Cell Biol.*, **24**, 9124–9136.
48. Flavahan, W.A., Gaskell, E. and Bernstein, B.E. (2017) Epigenetic plasticity and the hallmarks of cancer. *Science (80-.)*, **357**, eaal2380.
49. Kim, D.S., Choi, Y.P., Kang, S., Gao, M.Q., Kim, B., Park, H.R., Choi, Y.D., Lim, J.B., Na, H.J., Kim, H.K. *et al.* (2010) Panel of candidate biomarkers for renal cell carcinoma. *J. Proteome Res.*, **9**, 3710–3719.
50. Thakur, A., Rahman, K.M.W., Wu, J., Bollig, A., Biliran, H., Lin, X., Nassar, H., Grignon, D.J., Sarkar, F.H. and Liao, J.D. (2007) Aberrant expression of X-linked genes RbAp46, Rsk4, and Cldn2 in breast cancer. *Mol. Cancer Res.*, **5**, 171–181.
51. Wang, C.L., Wang, C.I., Liao, P.C., Chen, C. De, Liang, Y., Chuang, W.Y., Tsai, Y.H., Chen, H.C., Chang, Y.S., Yu, J.S. *et al.* (2009) Discovery of retinoblastoma-associated binding protein 46 as a novel prognostic marker for distant metastasis in non-small cell lung cancer by combined analysis of cancer cell secretome and pleural effusion proteome. *J. Proteome Res.*, **8**, 4428–4440.
52. Song, H., Xia, S.L., Liao, C., Li, Y.L., Wang, Y.F., Li, T.P. and Zhao, M.J. (2004) Genes encoding Pir51, Beclin 1, RbAp48 and Adolase b are up or down-regulated in human primary hepatocellular carcinoma. *World J. Gastroenterol.*, **10**, 509–513.
53. Sakhinia, E., Faranghpour, M., Yin, J.A.L., Brady, G., Hoyland, J.A. and Byers, R.J. (2005) Routine expression profiling of microarray gene signatures in acute leukaemia by real-time PCR of human bone marrow. *Br. J. Haematol.*, **130**, 233–248.
54. Kitange, G.J., Mladek, A.C., Schroeder, M.A., Pokorny, J.C., Carlson, B.L., Zhang, Y., Nair, A.A., Lee, J.H., Yan, H., Decker, P.A. *et al.* (2016) Retinoblastoma binding protein 4 modulates temozolomide sensitivity in glioblastoma by regulating DNA repair proteins. *Cell Rep.*, **14**, 2587–2598.
55. Saettone, A., Nabeel-Shah, S., Garg, J., Lambert, J.-P., Pearlman, R.E. and Fillingham, J. (2019) Functional proteomics of nuclear proteins in *Tetrahymena thermophila*: a review. *Genes (Basel)*, **10**, 333.
56. Ma, Y., Yan, G., Han, X., Zhang, J., Xiong, J. and Miao, W. (2020) Sexual cell cycle initiation is regulated by CDK19 and CYC9 in *Tetrahymena thermophila*. *J. Cell Sci.*, **133**, jcs235721.
57. Wu, M., Allis, C.D. and Gorovsky, M.A. (1988) Cell-cycle regulation as a mechanism for targeting proteins to specific DNA sequences in *Tetrahymena thermophila*. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2205–2209.
58. Ashraf, K., Nabeel-Shah, S., Garg, J., Saettone, A., Derynck, J., Gingras, A.-C., Lambert, J.-P., Pearlman, R.E. and Fillingham, J. (2019) Proteomic analysis of histones H2A/H2B and variant Hv1 in *Tetrahymena thermophila* reveals an ancient network of chaperones. *Mol. Biol. Evol.*, **36**, 1037–1055.
59. Wahab, S., Saettone, A., Nabeel-Shah, S., Dannah, N. and Fillingham, J. (2020) Exploring the histone acetylation cycle in the protozoan model *Tetrahymena thermophila*. *Front. Cell Dev. Biol.*, **8**, 509.
60. Orias, E., Cervantes, M.D. and Hamilton, E.P. (2011) *Tetrahymena thermophila*, a unicellular eukaryote with separate germline and somatic genomes. *Res. Microbiol.*, **162**, 578–586.
61. Martindale, D.W., Allis, C.D. and Bruns, P.J. (1982) Conjugation in *Tetrahymena thermophila*. A temporal analysis of cytological stages. *Exp. Cell Res.*, **140**, 227–236.
62. Yao, M.-C.C., Choi, J., Yokoyama, S., Austerberry, C.F. and Yao, C.-H.H. (1984) DNA elimination in *Tetrahymena*: a developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell*, **36**, 433–440.
63. Yao, M.C., Yao, C.H. and Monks, B. (1990) The controlling sequence for site-specific chromosome breakage in *Tetrahymena*. *Cell*, **63**, 763–772.
64. Mochizuki, K. and Gorovsky, M.A. (2004) Small RNAs in genome rearrangement in *Tetrahymena*. *Curr. Opin. Genet. Dev.*, **14**, 181–187.
65. Ruehle, M.D., Orias, E. and Pearson, C.G. (2016) *Tetrahymena* as a unicellular model eukaryote: genetic and genomic tools. *Genetics*, **203**, 649–665.
66. Xiong, J., Lu, X., Zhou, Z., Chang, Y., Yuan, D., Tian, M., Zhou, Z., Wang, L., Fu, C., Orias, E. *et al.* (2012) Transcriptome analysis of the model protozoan, *Tetrahymena thermophila*, using Deep RNA sequencing. *PLoS One*, **7**, e30630.
67. Garg, J., Lambert, J.P., Karsou, A., Marquez, S., Nabeel-Shah, S., Bertucci, V., Retnasothie, D. V., Radovani, E., Pawson, T., Gingras, A.C. *et al.* (2013) Conserved Asf1-importin β physical interaction in growth and sexual development in the ciliate *Tetrahymena thermophila*. *J. Proteomics*, **94**, 311–326.
68. Marcon, E., Ni, Z., Pu, S., Turinsky, A.L., Trimble, S.S., Olsen, J.B., Silverman-Gavrila, R., Silverman-Gavrila, L., Phanse, S., Guo, H. *et al.* (2014) Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation. *Cell Rep.*, **8**, 297–310.
69. Bruns, P.J. and Cassidy-Hanley, D. (2000) Biolistic transformation of macro- and micronuclei. *Methods Cell Biol.*, **62**, 501–512.
70. Nabeel-Shah, S., Garg, J., Kougnassoukou Tchare, P.E., Pearlman, R.E., Lambert, J.P. and Fillingham, J. (2021) Functional

- proteomics protocol for the identification of interaction partners in *Tetrahymena thermophila*. *STAR Protoc.*, **2**, 100362.
71. Schmitges, F.W., Radovani, E., Najafabadi, H.S., Barazandeh, M., Campitelli, L.F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T. *et al.* (2016) Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res.*, **26**, 1742–1752.
 72. Nabeel-Shah, S., Lee, H., Ahmed, N., Marcon, E., Farhangmehr, S., Pu, S., Burke, G.L., Ashraf, K., Wei, H., Zhong, G. *et al.* (2020) SARS-CoV-2 nucleocapsid protein attenuates stress granule formation and alters gene expression via direct interaction with host mRNAs. bioRxiv doi: <https://doi.org/10.1101/2020.10.23.342113>, 23 October 2020, preprint: not peer reviewed.
 73. Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.*, **11**, 2301–2319.
 74. Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
 75. Knight, J.D.R., Choi, H., Gupta, G.D., Pelletier, L., Raught, B., Nesvizhskii, A.I. and Gingras, A.C. (2017) ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nat. Methods*, **14**, 645–646.
 76. Saettone, A., Garg, J., Lambert, J.-P., Nabeel-Shah, S., Ponce, M., Burtch, A., Thuppu Mudalige, C., Gingras, A.-C., Pearlman, R.E. and Fillingham, J. (2018) The bromodomain-containing protein Ibd1 links multiple chromatin-related protein complexes to highly expressed genes in *Tetrahymena thermophila*. *Epigenetics Chromatin*, **11**, 10.
 77. Garg, J., Saettone, A., Nabeel-Shah, S., Cadorn, M., Ponce, M., Marquez, S., Pu, S., Greenblatt, J., Lambert, J.-P., Pearlman, R.E. *et al.* (2019) The Med31 conserved component of the divergent mediator complex in *Tetrahymena thermophila* participates in developmental regulation. *Curr. Biol.*, **29**, 2371–2379.
 78. Saettone, A., Ponce, M., Nabeel-Shah, S. and Fillingham, J. (2019) RACS: rapid analysis of ChIP-Seq data for contig based genomes. *BMC Bioinformatics*, **20**, 533.
 79. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
 80. Feng, J., Liu, T., Qin, B., Zhang, Y. and Liu, X.S. (2012) Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.*, **7**, 1728–1740.
 81. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
 82. Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME Suite. *Nucleic Acids Res.*, **43**, W39–W49.
 83. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
 84. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
 85. Miao, W., Xiong, J., Bowen, J., Wang, W., Liu, Y., Braguinets, O., Grigull, J., Pearlman, R.E., Orias, E. and Gorovsky, M.A. (2009) Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. *PLoS One*, **4**, e4429.
 86. Zhang, J., Yan, G., Tian, M., Ma, Y., Xiong, J. and Miao, W. (2018) A DP-like transcription factor protein interacts with E2f1 to regulate meiosis in *Tetrahymena thermophila*. *Cell Cycle*, **17**, 634–642.
 87. Akematsu, T., Sánchez-Fernández, R., Kosta, F., Holzer, E. and Loidl, J. (2020) The transmembrane protein Semil positions gamete nuclei for reciprocal fertilization in *Tetrahymena*. *iScience*, **23**, 100749.
 88. Howard-Till, R.A. and Yao, M.-C. (2006) Induction of gene silencing by hairpin RNA expression in *Tetrahymena thermophila* reveals a second small RNA pathway. *Mol. Cell Biol.*, **26**, 8731–8742.
 89. Horowitz, S., Bowen, J.K., Bannon, G.A. and Gorovsky, M.A. (1987) Unusual features of transcribed and translated regions of the histone H4 gene family of *Tetrahymena thermophila*. *Nucleic Acids Res.*, **15**, 141–160.
 90. Cui, B., Liu, Y. and Gorovsky, M.A. (2006) Deposition and function of histone H3 variants in *Tetrahymena thermophila*. *Mol. Cell Biol.*, **26**, 7719–7730.
 91. Nabeel-Shah, S., Ashraf, K., Saettone, A., Garg, J., Derynck, J., Lambert, J.P., Pearlman, R.E. and Fillingham, J. (2020) Nucleus-specific linker histones Hho1 and Mh1 form distinct protein interactions during growth, starvation and development in *Tetrahymena thermophila*. *Sci. Rep.*, **10**, 168.
 92. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.-C. and Choi, H. (2014) SAINTexpress: improvements and additional features in significance analysis of INTeractome software. *J. Proteomics*, **100**, 37–43.
 93. Latreille, D., Bluy, L., Benkirane, M. and Kiernan, R.E. (2014) Identification of histone 3 variant 2 interacting factors. *Nucleic Acids Res.*, **42**, 3542–3550.
 94. Nabeel-Shah, S., Ashraf, K., Pearlman, R.E. and Fillingham, J. (2014) Molecular evolution of NASP and conserved histone H3/H4 transport pathway. *BMC Evol. Biol.*, **14**, 139.
 95. Haigney, A., Ricketts, M.D. and Marmorstein, R. (2015) Dissecting the molecular roles of histone chaperones in histone acetylation by type B histone acetyltransferases (HAT-B). *J. Biol. Chem.*, **290**, 30648–30657.
 96. Hennig, L., Bouveret, R. and Grussem, W. (2005) MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. *Trends Cell Biol.*, **15**, 295–302.
 97. Iouk, T.L., Aitchison, J.D., Maguire, S. and Wozniak, R.W. (2001) Rbl1p, a yeast nuclear WD-repeat protein involved in the regulation of ribosome biosynthesis. *Mol. Cell Biol.*, **21**, 1260–1271.
 98. Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatzer, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M. and Beisel, C. (2016) A high-density map for navigating the human polycomb complexome. *Cell Rep.*, **17**, 583–595.
 99. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'donnell, L., Leung, G., Mcadam, R. *et al.* (2018) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, 529–541.
 100. Zhang, J., Tian, M., Yan, G.X., Shodhan, A. and Miao, W. (2017) E2f1 is a meiosis-specific transcription factor in the protist *Tetrahymena thermophila*. *Cell Cycle*, **16**, 123–135.
 101. Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N. V., Montgomery, P.G., Cowley, G.S., Pantel, S. *et al.* (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.*, **49**, 1779–1784.
 102. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M. *et al.* (2017) Defining a cancer dependency map. *Cell*, **170**, 564–576.
 103. Iness, A.N., Felthousen, J., Ananthapadmanabhan, V., Sesay, F., Saini, S., Guiley, K.Z., Rubin, S.M., Dozmorov, M. and Litovchick, L. (2019) The cell cycle regulatory DREAM complex is disrupted by high expression of oncogenic B-Myb. *Oncogene*, **38**, 1080–1092.
 104. Blanchard, D.P., Georgette, D., Antoszewski, L. and Botchan, M.R. (2014) Chromatin reader L(3)mbt requires the Myb-MuvB/DREAM transcriptional regulatory complex for chromosomal recruitment. *Proc. Natl. Acad. Sci. USA*, **111**, E4234–E4243.
 105. Beall, E.L., Lewis, P.W., Bell, M., Rocha, M., Jones, D.L. and Botchan, M.R. (2007) Discovery of tMAC: a *Drosophila* testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev.*, **21**, 904–919.
 106. Lu, C. and Fuller, M.T. (2015) Recruitment of mediator complex by cell type and stage-specific factors required for tissue-specific TAF dependent gene activation in an adult stem cell lineage. *PLoS Genet.*, **11**, e1005701.
 107. Rabinovich, A., Jin, V.X., Rabinovich, R., Xu, X. and Farnham, P.J. (2008) E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res.*, **18**, 1763–1777.
 108. Howard-Till, R.A., Lukaszewicz, A. and Loidl, J. (2011) The recombinases Rad51 and Dmcl1 play distinct roles in DNA break repair and recombination partner choice in the meiosis of *Tetrahymena*. *PLoS Genet.*, **7**, e1001359.
 109. Smith, J.J., Cole, E.S. and Romero, D.P. (2004) Transcriptional control of RAD51 expression in the ciliate *Tetrahymena thermophila*. *Nucleic Acids Res.*, **32**, 4313–4321.