MRG proteins are shared by multiple protein complexes with distinct functions

Maëva Devoucoux[1], Céline Roques[1], Charles Joly-Beauparlant[2], Arnaud Droit[2], Samer M.I. Hussein[1] and Jacques Côté[1*]

1. St. Patrick Research Group in Basic Oncology, Laval University Cancer Research Center, Oncology division of CHU de Québec-Université Laval Research Center, Quebec City, QC G1R 3S3, Canada.

2. Axe Neurosciences, Centre de Recherche du CHU de Québec-Université Laval, Pavillon CHUL, 2705 boulevard Laurier, Quebec City, QC G1V 4G2, Canada ; Faculty of Medicine, Université Laval, 1050 ave de la Médecine, Quebec City, QC, Canada.

*Corresponding author: jacques.cote@crhdq.ulaval.ca

Running title: Various functions of MRG proteins interactome

Highlights

- MRG15 and MRGX are stably associated with several different protein complexes important for genome expression and stability.
- Several MRG-containing complexes are chromatin modifiers
- Specific point mutations in the MRG domain differentially affect associated complexes.
- A major human complex homologous to the yeast TINTIN complex is identified.
- The protein EP400NL competes with EP400 to functionally separate TINTIN from the NuA4/TIP60 complex .
- TINTIN contains a bromodomain and a chromodomain to regulate transcription

Abbreviation

AP-MS, affinity purification combined to mass spectrometry; ASH1L, Absent, Small, or Homeotic discs 1-Like; BRD, bromodomain; CHD, chromodomain; CTD, C-terminal domain; EP400NL, EP400-N-terminal like; FPLC, Fast Protein Liquid Chromatography; K562, human erythroleukemic cells; KAT, lysine acetyltransferase; KDAC, lysine deacetylase; MRG15, MORF-related gene on chromosome 15; NuA4, Nucleosome Acetyltransferase of H4; RNAPII, RNA polymerase II; SETD2, SET Domain containing 2; TINTIN, Trimer Independent of NuA4 involved in Transcription Interactions with Nucleosomes; TIP60, Tat Interactive Protein 60kDa; U2OS, human osteosarcoma epithelial cells.

ABSTRACT

MRG15/MORF4L1 is a highly conserved protein in eukaryotes that contains a chromodomain recognizing H3K36me3 in chromatin. Intriguingly, it has been reported in the literature to interact with several different factors involved in chromatin modifications, gene regulation, alternative mRNA splicing and DNA repair by homologous recombination. In order to get a complete and reliable picture of associations in physiological conditions, we used genome editing and tandem affinity purification to analyze the stable native interactome of human MRG15, its paralog MRGX/MORF4L2 that lacks the chromodomain, and MRGBP (MRG-binding protein) in isogenic K562 cells. We found stable interchangeable association of MRG15 and MRGX with the NuA4/TIP60 histone acetyltransferase/chromatin remodeler, Sin3B histone deacetylase/demethylase, ASH1L histone methyltransferase and PALB2/BRCA2 DNA repair protein complexes. These associations were further confirmed and analyzed by CRISPR-tagging of endogenous proteins and comparison of expressed isoforms. Importantly, based on structural information, point mutations could be introduced that can specifically disrupt MRG15 association with some complexes but not others. Most interestingly, we also identified a new abundant native complex formed by MRG15/X-MRGBP-BRD8-EP400NL that is functionally similar to the yeast TINTIN (Trimer Independent of NuA4 for Transcription Interactions with Nucleosomes) complex. Our results show that EP400NL, being homologous to the N-terminal region of NuA4/TIP60 subunit EP400, creates TINTIN by competing for BRD8 association. Functional genomics indicate that human TINTIN plays a role in transcription of specific genes. This is most likely linked to the H4ac-binding bromodomain of BRD8 along the H3K36me3-binding chromodomain of MRG15 on the coding region of transcribed genes. Taken together,

our data provide a complete detailed picture of human MRG proteins-associated protein complexes which is essential to understand and correlate their diverse biological functions in chromatin-based nuclear processes.

INTRODUCTION

The eukaryote genome is organized into a structure called chromatin, with the nucleosomes being its basic units, playing central roles in the regulation of all DNA-based processes for genome expression and maintenance (1). One major element governing chromatin structure and functions is the diverse post-translational modifications (PTMs) deposited on the external tails of the nucleosomal histones (2). PTMs on specific residues of specific histones include acetylation, methylation, phosphorylation and ubiquitination, which are recognized by dedicated reader domains in factors to regulate biological processes (3, 4). Disruption of the enzymes that "write" or "erase" these histone marks, or their "readers", occurs in several human diseases, including cancer, leading to intense research efforts during the past several years to understand these regulatory interactions during cellular life (5).

One mark that has been studied extensively is the methylation of lysine 36 on histone H3 (H3K36me), which can occur as mono- (me1), di- (me2), or tri-(me3) methylation, levels being linked to distinct functions. While all methylation levels are generated by Set2 in *Saccharomyces cerevisiae* (6), mammalian methyltransferases like ASH1L and NSD1/2/3 only do H3K36me1/me2. Mammalian SETD2 on the other hand is responsible for H3K36me3 (7). Set2 associates with the elongating RNA polymerase II (RNAPII) phosphorylated on Ser2 of its C-terminal repeats (CTD). This leads to enrichment of H3K36me3 on the coding region of transcribed genes which recruits the RPD3S histone deacetylase complex to stabilize chromatin in the wake of RNAPII in order to avoid cryptic initiation and spurious transcription (8, 9). Mammalian SETD2 does not seem to affect global chromatin acetylation levels (7), even if H3K36me3 plays a central role in transcriptional activation and is distributed throughout the

coding regions of active genes (10, 11). More precisely, H3K36me3 has been observed to be specifically enriched on exons, compared to introns, independent of nucleosome occupancy, in both yeast and mammals (12-14). These data were used to define a function of this histone mark in alternative mRNA splicing (15). It has been proposed that H3K36me3 is used as a platform recognized by the factor MORF-related gene on chromosome 15 (MRG15, a.k.a. MORF4L1) through its N-terminal chromodomain (CHD), an interaction that recruits the splicing machinery (16, 17). The yeast homolog of MRG15, Eaf3, is also known to affect splicing through its binding with the splicing factor Prp45/SKIP, thereby leading to the recruitment of spliceosome on intron-containing genes (18).

Mammalian MRG15 has a paralog expressed for the X-chromosome, MRGX (a.k.a. MORF4L2), which lacks the chromodomain. MRG15/X seem to regulate cell proliferation and senescence is specific contexts, while only MRG15 is required for mouse development (19, 20). They have been found associated with both lysine acetyltransferase (KAT) and deacetylase (KDAC) complexes, namely NuA4/TIP60 and Sin3B, also conserved in yeast (NuA4 and RPD3S) and containing the homologous protein Eaf3 (21-25). While MRGX does not possess a N-terminal CHD, it shares with MRG15 a highly conserved C-terminal MRG domain, crucial for their non-histone protein-protein interactions (20). Interestingly, it has been suggested that MRG15 can form an homodimer which is dissociated by its acetylation, regulating MRG15 association with other factors (26). MRG15 is also involved in DNA repair through its interaction with DNA repair factor PALB2, thereby favoring repair of double strand breaks by homologous recombination (27-29). It has been recently described that the MRG15-PALB2 complex is associated with undamaged chromatin on active genes to protect them from

genotoxic/replicative stress (30). Finally, MRG15 was very recently implicated in diurnal rhythm of epigenomic remodeling through chromatin acetylation as well as lipid metabolism through gene activation with LRH-1 (31)

Although current knowledge on MRG15 points towards quite distinct important roles in the cell and during development, its described functions and associated cofactors are often studied in very artificial non-physiological conditions. Here, we aimed at defining the complete detailed interactomes of MRG15/X in the most native conditions possible. Tandem-affinity purification from isogenic cell lines identified several stable complexes containing MRG15 or MRGX. Endogenous protein tagging by CRISPR allowed validation and characterization of a new tetrameric protein complex with striking resemblance to the yeast TINTIN (Trimer Independent of NuA4 for Transcription Interactions with Nucleosomes) complex that we have previously identified (25). The presence of a bromodomain and a chromodomain in this complex and its effect on transcription supports the claim that we have identified a human TINTIN complex that associates with the body of active genes.

EXPERIMENTAL PROCEDURES

Recombinant Proteins and Pull-down Assay

Recombinant 6xHis-BRD8, GST-MRG15 CHD long and short, GST-EP400(aa194-446), GST-EP400NL(aa51-297) and GST-ASH1L(aa2040-2634) were expressed in E. coli BL21 cells grown in

LB media (from pET15b, pGEX4T3 or pGEX6P2 vectors). After induction with IPTG (30μM) overnight at 16C, bacteria were harvested and GST or 6xHistagged proteins were purified as previously described with minor modifications (32).

Protein and chromatin pull-down experiments were performed as previously described (33). The GST pulldown assays were performed using 300 to 600 ng of GST-fused protein and an equivalent amount of His-tagged protein or 5μg of native chromatin from HeLa cells, which was precleared using glutathione-Sepharose beads. Equivalent protein levels were estimated through Coomassie blue-stained SDS-PAGE by comparison with known amounts of bovine serum albumin (BSA) standards. After pre-clearing, the His-tagged proteins were incubated with GST-immobilized protein in binding buffer (25 mM HEPES pH 7.9, 10% glycerol, 100 μg/ml BSA, 1 mM PMSF, 0.5 mM dithiothreitol (DTT), 0.1% Tween 20, and protease inhibitors, 250mM NaCl was used for chromatin pull-down) at 4°C for 1h to 4h followed by washing the beads three times. To visualize the proteins, the beads were loaded on SDS-PAGE gels, followed by Western blotting with anti-GST (Sigma G1160), anti-His (Clontech 631212), anti-H3K36me2 (Upstate 07-369), anti-H3K36me3(Abcam ab9050), anti-H3K79me3 (Abcam ab2621), and anti-H4 (Abcam ab7311). A GST-only protein and beads were used as a control.

Experimental Model and Subject Details

K562 and U2OS cells were obtained from the ATCC and maintained at 37°C under 5% CO2. K562 were cultured in RPMI medium supplemented with 10% newborn calf serum and GlutaMAX. U2OS cells were cultured in DMEM medium supplemented with 10% fetal bovine serum.

Establishment of Isogenic Cell Lines using the AAVS1 Safe Harbor or CRISPR/Cas9-mediated Genome Editing

K562 cells stably expressing 3xFLAG-2xStrep-tagged MRG15 spliced variants (long CHD or short CHD), MRGBP, MRGX, BRD8 spliced variants (one or two BRDs), EPC1 or EP400NL spliced variants (isoforms Q6ZTU2-5 and Q6ZTU2-6) were established as described before (34). BRD8, ASH1L, MRGX, and EP400NL were also endogenously 3xFlag-2xStrep-tagged using CRISPR/Cas9 with ouabain selection as previously described (35). The following gRNAs were used to generate these cell lines:

Gene	Part	Sequence for	Sequence rev
BRD8	N-terminal	CCTCAGGGTTGGAGACTTCG	CGAAGTCTCCAACCCTGAGG
BRD8 Q9H0E9-1	C-terminal	AAAGGCTCAAGTAGTCTGGA	TCCAGACTACTTGAGCCTTT
MRGX	C-terminal	TGTGAGCGTCTACAGACAGC	GCTGTCTGTAGACGCTCACA
EP400NL	C-terminal	TTTGAGGACTCGATGTGCTG	CAGCACATCGAGTCCTCAAA
ASH1L	C-terminal	CTCAAAGAATGAGAACCTCAA	TTGAGGTTCTCATTCTTTGAG

Affinity Purification of Complexes

Native complexes were purified essentially as previously described (34). Briefly, Nuclear extracts were prepared from 2.5.10^9 cells and pre-cleared with CL6B sepharose beads. FLAG immunoprecipitations with anti-Flag agarose affinity resin (Sigma M2) were performed followed by two elutions with 3xFLAG peptide in elution buffer (20 mM HEPES-KOH [pH 7.9],

10% glycerol, 150 mM KCl, 0.1% Tween 20, 1mMDTT, 1 mM PMSF, 2 mg/mL Leupeptin, 5 mg/mL Aprotinin, 2 mg/mL Pepstatin, 10 mM Na-Butyrate, 10 mM β-glycerophosphate) with 200 μg/mL 3xFLAG peptide (Sigma). Then, STREP immunoprecipitations with Strep-tactin sepharose beads (Cedarlane) were performed followed by two elutions with elution buffer supplemented with 4mM biotin. Typically, 20μL of the first elution Strep was loaded on Nu-PAGE 4%–12% Bis-Tris gels (Invitrogen) and analyzed via silver staining. Fractions were then analyzed by mass spectrometry and western blotting. For purification after inducing DNA damage, cells were treated with 50ng/mL of neocarzinostatin (NCS) for 3h before beung collected to prepare the nuclear extract.

Mass Spectrometry Analysis

Mass spectrometry analyses were performed by the Proteomics Platform of the CHU de Québec-Université Laval Research Center (Québec, Qc, Canada). Samples were analyzed by nano-LC/MSMS either with a 5600 triple TOF or an Orbitrap Fusion.

5600 triple TOF: Peptides were analyzed using an Ekspert NanoLC425 coupled to a 5600+ triple TOF mass spectrometer (Sciex, Framingham, MA, USA). Peptides were trapped at 4 µl/min in loading solvent (0.1% formic acid) on a 5mm x 300 µm C18 pepmap cartridge pre-column (Thermo Fisher Scientific / Dionex Softron GmbH, Germering, Germany) for 10 minutes. Then, the pre-column was switched online with a self-packed picofrit column (New Objective) packed with reprosil 3u, 120A C18, 15 cm x 0.075 mm internal diameter, (Dr. Maisch). Peptides were eluted with a linear gradient from 5-35% solvent B (A: 0.1% formic acid, B: acetonitrile, 0.1% formic acid) in 35 minutes, at 300 nL/min for a total run time of 60min. Mass spectra were

acquired using a data-dependent acquisition mode using Analyst software version 1.7. Each full scan mass spectrum (400 to 1250 m/z) was followed by collision-induced dissociation of the twenty most intense ions. Dynamic exclusion was set for 12 sec and tolerance of 100 ppm.

Orbitrap fusion: Peptides were analyzed using a Dionex UltiMate 3000 nanoRSLC chromatography system (Thermo Fisher Scientific) connected to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Peptides were trapped at 20 µl/min in loading solvent (2% acetonitrile, 0.05% TFA) on a 5mm x 300 µm C18 pepmap cartridge pre-column (Thermo Fisher Scientific / Dionex Softron GmbH, Germering, Germany) for 5 minutes. Then, the pre-column was switched online with a Pepmap Acclaim column (ThermoFisher) 50 cm x 75µm internal diameter separation column and the peptides were eluted with a linear gradient from 5-40% solvent B (A: 0,1% formic acid, B: 80% acetonitrile, 0.1% formic acid) in 60 minutes, at 300 nL/min for a total runtime of 90 min. Mass spectra were acquired using a data-dependent acquisition mode using Thermo XCalibur software version 4.3.73.11. Full scan mass spectra (350 to 1800m/z) were acquired in the orbitrap using an AGC target of 4e5, a maximum injection time of 50 ms, and a resolution of 120 000. Internal calibration using lock mass on the m/z 445.12003 siloxane ion was used. Each MS scan was followed by MSMS fragmentation of the most intense ions for a total cycle time of 3 seconds (top speed mode). The selected ions were isolated using the quadrupole analyzer in a window of 1.6 m/z and fragmented by Higher energy Collision-induced Dissociation (HCD) with 35% of collision energy. The resulting fragments were detected by the linear ion trap at a rapid scan rate with an AGC target of 1e4 and a maximum injection time of 50ms. Dynamic exclusion of previously fragmented peptides was set for 20 sec and a tolerance of 10 ppm.

Database searching: MGF peak list files were created using Protein Pilot version 4.5 software (Sciex) for the data obtained with the 5600+ triple TOF and with Proteome Discoverer 2.3 software (Thermo) for the orbitrap data. The MGF sample files were then analyzed using Mascot (Matrix Science, London, UK; version 2.5.1). Mascot was set up to search a contaminant database and Uniprot Homo sapiens database assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da (orbitrap) or 0.1Da (5600+) and a parent ion tolerance of 10.0 ppm (orbitrap) and 0.1Da (5600+). Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Deamidation of asparagine and glutamine and oxidation of methionine were specified in Mascot as variable modifications. 2 missed cleavages were allowed.

Criteria for protein identification: Scaffold (version Scaffold_4.8.7, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 8.0% probability to achieve an FDR less than 1.0% by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability to achieve an FDR less than 1.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (36). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Data were further analyzed using the CRAPome online tool (www.crapome.org).

All MS files generated as part of this study were deposited at MassIVE (http://massive.ucsd.edu). The MassIVE ID is MSV000087245 and the MassIVE link for download is

http://massive.ucsd.edu/ProteoSAFe/status.jsp?task=9534fb8b77634860938b1139a53d9df8.

The password for download prior to final acceptance is MRG15.

Antibodies and siRNAs

The following antibodies were used for Western blotting at the indicated dilution: anti-FLAG - HRP conjugate (Sigma M2, 1:10000); anti-Brd8 (Bethyl A300-219A, 1:10000); anti-DMAP1 (Thermoscientific PA1-886, 1:1000); anti-P400 (Abcam ab5201, 1:1000); anti-MRG15 (Active Motif 39361, 1:1000); anti-MRGBP (Abnova H00055257-B01, 1:2000); anti-MRGX (Abnova PAB6152, 1:1000); anti-KAT5/Tip60 (Abcam ab137518, 1:1000); anti-ASH1L (Bethyl A301-749, 1:1000); anti-KDM5A (Cell signaling 3876, 1:1000); anti-Pf1 (Bethyl A301-647, 1:1000); anti-PALB2 (Bethyl A301-246, 1:3000); anti-AKAP8 (Abcam ab72196, 1:2500); anti-GAPDH (Thermofisher 39-8600, 1:10000); anti-KAP1 (MAB3662, 1/4000).

The following siRNA were used against the indicated proteins: KAT5/TIP60 smartpool (Dharmacon, CCACAGAUCACCAUCAAUG, GAACAAGAGUUAUUCCAG, CACAGGAACUCACCACAUU, GGACAGCUCUGAUGGAAUA), BRD8 from Sigma (SASI_Hs01_00131635; target sequences start at nucleotide 2223); MRG15 smartpool (Dharmacon, GAAGAGCCUUGCUUUAUUA, UAAAGUACCUGGCAAAGAA, GUACCCAGCUACUCUAUAA, GAUGACUGGGACUUAAUUA); MRGX smartpool (Dharmacon, GCACUCAGCUGCUCUACAA, GAGGAGGCGUUUAAGAAUA, GCAGGGAAAUGUUGAUAAU, GCUCCAAUGUCCCAGGUUU), MRGBP smartpool (Dhaarmacon, GAAGAACUCCUCAGACUUG, GAACUUCGUCCUUCCAGAA, GAACCGACACUUCCACAUG, ACAAAGUCCUGACCGCAAA), and

siLuciferase as control is used (DOYYA, AACUGACGCGGAAUACUUCGA). Knockdown efficiency was validated by RT-qPCR with the following primers:

Targets	primers for	primers rev
BRD8	ATGGTGGGGAGATACAGCAA	AGTATGTGGATCCCCCACAG
KAT5/TIP60	CATCCTCCAGGCAATGAGAT	ACTTGGCCAAAAGACACAGG
MRG15	CACCCATGTCCCAGGTGTAT	GCAAGGCTCTTCTCATCCAG
MRGX	GTGTGCTTGCTTGGAGATCA	TCAGGGAAGGTTCTGCAATC
MRGBP	GGAGGAGACAGTGGTGTGG	CATGTGGAAGTGTCGGTTCA

In Vitro Histone Methyltransferase Assay

Purified endogenous ASH1L (or truncated recombinant GST-ASH1L(aa2040-2634)) was first preincubated with recombinant MRG15 purified from baculovirus (or the same amount of BSA) and 0.5µg of native Short OligoNucleosomes (SON) purified from HeLa cells for 4h at 30C in a 20µl reaction containing 100mM Tris-HCl pH8.0, 25% glycerol, 0.5mM EDTA, 5mM DTT, 5mM PMSF, 10mM $MgCl_2$ for 30min at 4C. Then, 2 µL of ^{3}H S- adenosyl methionine (0.55uCi/ul) was added and the mixture was incubated for 3h at 30C. the reaction was then spotted onto p81 filter paper, washed in carbonate buffer and scintillation counting was used to determine the incorporation.

RNA-sequencing Analysis

2 days post-Knockdown (KD) of MRG proteins, BRD8 or KAT5 in U2OS cells, RNA was extracted with the Monarch Total RNA Miniprep Kit (NEB #T2010), following the manufacturer's

indications. Then, at least 1.5μg of RNA were depleted of rRNAs using the NEB rRNA-depletion kit (HMR) followed by Illumina library preparation. Finally, cDNAs were sequenced using the NovaSeq6000 S2 for 50M of PE100 reads. Two replicates of each KD were sequenced. Raw sequences and processed data of short reads RNA-sequencing from U2OS cells were deposited in the GEO database under accession number **GSE181533**.

Gene expression analysis

Raw fast5 files were base called using Guppy v.4.2.2 with the guppy_basecaller command and the dna_r9.4.1_450bps_fast.cfg configuration file and default settings. Barcodes were detected and reads were separated with the guppy_barcoder using the default settings. The resulting FASTQ reads were aligned to the hg19 human reference with Minimap2 with the following parameters: -aLx splice --cs=long. Raw read counts were obtained with the featureCounts tool from the Subread package v 2.0.0, using the exon counting mode (37, 38). EdgeR R-package (v3.12.1) was then used to normalize the data, calculated RNA abundance at the gene and transcript levels (as counts per million reads (CPM), and perform statistical analysis (39). Briefly, a common biological coefficient of variation (BCV) and dispersion (variance) were estimated based on a negative binomial distribution model. This estimated dispersion value was incorporated into the final EdgeR analysis for differential gene expression, and the generalized linear model (GLM) likelihood ratio test was used for statistics, as described in EdgeR user guide. Analysis of variations in gene expression was performed by comparing KDs BRD8, MRGBP, MRG15, MRGX, and KAT5/TIP60 to siControl (siLuciferase). From all genes showing differential expression, a cutoff of 2-fold difference was applied to the Log2(fold change) values between the selected sample and control.

mRNA splice variants analysis

Reads were trimmed using fastp v0.20.0 (40). Quality check was performed on raw and trimmed data to ensure the quality of the reads using FastQC v0.11.8 and MultiQC v1.8 (41, 42). The quantification was performed with Kallisto v0.46.2 (43). Differential expression analysis was performed in R v4.0.0 using the DESeq2 v1.28.1 (44, 45). Analysis of variations in splice isoform expression was performed by comparing differential expression at the gene and transcript levels. From all genes showing no differential expression with the selected control (padj ≥ 0.1), those with significant differences in transcript expression (padj ≤ 0.1) were selected. To highlight the genes with high changes in transcript expression compared to gene expression, a cutoff of 2-fold difference was applied to the Log2(fold change) values between the selected sample and control.

Reverse Transcription-qPCR

siRNAs for KD of BRD8 or MRG15 or MRGBP or MRGX or KAT5/TIP60 were transfected in U2OS cells. siLuciferase (siLuc) was used as control. 48hrs later total RNA was extracted with the RNeasy Plus Mini kit (QIAGEN). 500 ng of RNA was reverse transcribed by oligo-dT and random priming into cDNA with a qScript cDNA SuperMix kit (QuantaBio-VWR), according to the manufacturer's instructions. Quantification of the amount of cDNA was done with SYBR Green I (Roche) on a LightCycler 480 (Roche) for real-time PCR. The *RPLP0* gene was used as a housekeeping internal control for normalization. The error bars represent the range based on independent experiments. The oligonucleotide sequences used for expression analysis by RT-qPCR are listed below:

Targets	primers for	primers rev
RFX8	GTCATGCTGATGTCATTGCCT	CCAAGAAGTCATGAAGGAGAACA
HHIP	GGGCGCCTGGAGAATAAGATA	GGCTTTGAGAATGTGGAGAGC
SYTL5	TATGCAGGGAGTGTCGAGTTG	GTGCCGAGAACATTGACTTGC
CD180	TGAAGGGTCTGAACTTGCTCA	TAGAGGTGTGTAAGGGAGGGG
AKT2	CCGCTACTACGCCATGAAGAT	AACGGGTGCCTGGTGTTC
PRRG4	GCATTTTGGCAGGAATATTCAGC	CCAGCAGCAATTAATCCAGTCAG
PCDHB9	AGGATCTGGGACTAGCAGAGG	CAGCTTCTCTCGGTCCAGTTT
RPLP0	CGACCTGGAAGTCCAACTAC	ATCTGCTGCATCTGCTTG

Experimental Design and Statistical Rationale

A mock K562 cell line expressing an empty 3xFlag-2xStrep tag from the *AAVS1* locus is used as control. Each Knockdown in U2OS cells was performed in biological duplicates. Control with a siLuciferase was generated concomitantly to experimental samples. RT-qPCRs to validate Knockdown and RNA-sequencing were performed in biological duplicates. Error bars represent the range from two biological replicates. Statistical analyses were performed via two-way ANOVA using Prism version 7 (GraphPad software inc California, USA) followed by Tukey's test. P-values <0.05 were considered significant.

RESULTS

Impact of the CHD on the binding to of MRG proteins to chromatin.

Although MRG15 is highly conserved from yeast to human, it is interesting to observe that two isoforms are produced in mammals through alternative mRNA splicing and that a paralogous protein, MRGX, is produced by another gene. Interestingly, the major difference between these 3 proteins resides in the chromodomain (CHD). MRG15 isoforms contain either a "long CHD" (Q9UBU8-1 in Uniprot) with an extra 39 amino acids (aa) insert or a "short CHD" (Q9UBU8-2 in Uniprot) which is considered canonical. In contrast, MRGX lacks the CHD region at its N-terminus (**Fig. 1A**). Yeast Eaf3 in comparison also has a "long CHD" but the extra insert is located on the opposite side of the hydrophobic cage that recognizes H3K36me3 compared to MRG15 (**Fig. S1A**). The proteins are characterized by a ~175aa MRG domain at their C-terminus, important for their non-histone protein-protein interactions (20). Eaf3 CHD has been shown to bind H3K36me2/me3 which is required for suppression of spurious intragenic transcription by the RPD3S histone deacetylase complex (24, 46, 47). While Eaf3 CHD has been argued to also bind H3K4me3 *in vitro*, MRG15 short CHD does not show any interaction with this mark (48). We then wondered if the insert in MRG15 long CHD isoform could change its binding properties. To test that, we performed GST pull-down experiments with purified native chromatin and recombinant proteins containing the short or long CHD of MRG15. The short CHD showed great specificity for H3K36me3, compared to H3K36me2, while the long CHD seem to have weaker affinity (**Fig. 1B**).

To further examine the potential impact of the CHD size and or its absence on chromatin binding *in vivo*, we generated isogenic K562 cell lines with a single copy of the epitope-tagged

(3xFlag-2xStrep) MRG15 isoforms or MRGX cDNA inserted at the *AAVS1* safe-harbor locus, as described before (**Fig. S1B**)(34). We then analyzed their genome-wide locations using ChIP-seq experiments with an anti-Flag antibody. MRG15-S, MRG15-L and MRGX seem the share the vast majority (>70%) of their genomic targets, independently of the CHD status (**Fig. S1C**). On the other hand, MRG15-L had a relatively much smaller number of bound regions that could be mapped, possibly reflecting a looser interaction with chromatin. It remains to be seen if the small proportion of genomic regions uniquely bound by MRGX or MRG15 are functionally significant.

MRG proteins are associated with several distinct protein complexes but are interchangeable.

As mentioned above, several associated factors have been identified for mammalian canonical MRG15 (short CHD/MRG15-S), often using methods like transient or stable over-expression and simple immunoprecipitations. Thus, we wanted to get a full picture of the associated complexes in the most native conditions using our validated approaches. At the same time, we investigated if MRG15-L, MRG15-S and MRGX have different interactomes. Even though the *AAVS1* system usually allows near physiological level of expression (within 2-2.5-fold (34)), it does not recapitulate native regulation by endogenous promoters and 5'/3'UTRs. We were able to tag the endogenous *MRGX* gene using an improved CRISPR/Cas9 genome editing method (35). We used a gRNA targeting the codon stop of MRGX to introduce the 3xFlag-2xStrep tag (**Fig. S1D**). Comparing endogenous versus *AAVS1*-mediated expression levels indicates

significantly lower endogenous expression, but similar to MRG15 expression from AAVS1 (**Fig. S1B**).

We then performed tandem affinity purifications with native elution at both steps, as previously described (34, 49). The final elutions with biotin were then analyzed by silver staining, mass spectrometry and western blotting (**Fig. 1C-E**, **Table S1**). In agreement with the highly conserved MRG domain between MRG15 and MRGX, we observed very similar interactomes for all MRG proteins. We could detect the full set of subunits for the NuA4/TIP60 acetyltransferase complex as well as the Sin3B deacetylase complex (homologous to yeast RPD3S) that includes the KDM5A H3K4 demethylase. It is interesting to point out that MRG15 and MRGX proteins are mutually exclusive in their associations since we cannot MRGX in MRG15 preps and vice versa (**Fig. 1D** and **Table S1**). MRG15-S and MRGX have also been previously linked to PALB2/BRCA2, playing a crucial function in DNA repair by homologous recombination (27, 29, 30) and we confirmed that (**Fig. 1E**, **Table S1**). This interaction seems stable and not modulated by a cellular DNA damage response since we obtained very similar results purifying MRG15-S from cells treated with DNA double-strand break-inducing agent neocarzinostatin (NCS), for PALB2/BRCA2 or any other interacting partners (Table S1, right panel), in agreement with the proposed role of MRG15/PALB2/BRCA2 in protecting undamaged chromatin on active genes (30). MRFAP1 is also an interesting interactor of both MRG15 spliced isoforms and MRGX as it was known to form a complex with MRG15-S through regulation of NEDDylation (50). Finally, a particularly interesting shared interaction is with the H3K36 methyltransferase ASH1L (**Fig. 1E**, **Table S1**).

Endogenous ASH1L is associated with MRG proteins.

ASH1L is a member of the trithorax group proteins involved in activation of Hox genes in *Drosophila*. Its methyltransferase activity is crucial for H3K36me1/me2 maintenance, thereby leading to MLL recruitment on its specific promoter targets (51, 52). The catalytic activity of ASH1L seems also important for DNA repair via nucleotide excision (53). Moreover, previous studies have been shown that MRG15 and MRGX enhance the methyltransferase activity of a truncated version of ASH1L (54, 55). Indeed, this binding can affect the auto-inhibitory loop of ASH1L (56, 57). Importantly, this stimulatory effect is therefore independent of the H3K36me-binding CHD of MRG15. To confirm this association and biochemical studies in the physiological context, we were able to tag and purify endogenous ASH1L using the CRISPR/Cas9 system in K562 cells (**Fig. 2A**, **S1E**). We then analyzed by western blots and mass spectrometry the ASH1L interactome. We confirmed association of MRG15 and MRGX as well as Nurf55, an already known factor link to ASH1L (**Fig. 2B**, **Table S2**). Interestingly, the A Kinase Anchoring protein 8 (AKAP8) and multiple RNA-processing factors were also found associated with ASH1L, assuming a potential function in alternative splicing, as it was already described for AKAP8 (58).

To confirm the role of MRG15 in stimulating ASH1L methyltransferase activity, we performed an *in vitro* histone methyltransferase assay (HMT) to measure its activity. We used a recombinant truncated version of ASH1L (aa 2040-2684), preincubated with recombinant MRG15 or BSA. As expected, MRG15 greatly enhanced the methyltransferase activity of ASH1L (**Fig. 2C**). In parallel, the methyltransferase activity of the purified endogenous ASH1L was not affected by the preincubation with recombinant MRG15. This strongly suggests that the vast

majority of endogenous ASH1L is stably bound to an MRG protein *in vivo,* being fully active and non-responsive to addition of exogenous MRG15.

Genetic mutations to modulate MRG15 association with distinct complexes

Previous work has been done to characterize the structure of the MRG domain as an interface for protein-protein interactions (20, 59). Based on this information, we investigated if we could manipulate the association of MRG proteins with distinct complexes in native conditions. We generated K562 cell lines expressing tagged MRG15 carrying the W172A and Y235A substitutions in its MRG domain (**Fig. 3A**). After purification and analysis by silver-stained gel, western blot and mass spectrometry, we can conclude that these mutations specifically disrupt MRG15 association with the Sin3B deacetylase complex, BRCA2/PALB2, MRFAP1 and ASH1L (**Fig. 3C-D**, **Table S3**). Strikingly, MRG15 association with the NuA4/TIP60 complex is not affected, indicating that distinct molecular features of the MRG domain allow the multi-specificity of protein interactions and therefore associations to complexes. These mutants could help address the functional implication of MRG proteins association with these partners versus its presence in NuA4/TIP60. To see if it would be possible to address the opposite question, we again used published structural information (20) and introduce point mutations in the MRG-binding protein MRGBP, a subunit of NuA4/TIP60 directly associated with MRG15/X in the complex. We generated K562 cell lines expressing tagged wild-type and mutant MRGBP carrying the W78A F105A substitutions (**Fig. 3A**). Analysis of the purified fractions showed that the double mutant disrupted the association of MRG15 and MRGX with the native NuA4/TIP60

complex without affecting other components (**Fig. 3D-E**). Together, these MRG mutants have the potential to help us understand the specific function of MRG protein in distinct complexes. For example, as mentioned before, MRG15 is known to have a function in the repair of DNA double-strand breaks by homologous recombination, certainly in part through its association with PALB2/BRCA2 (27, 29, 30, 60). Furthermore, the NuA4/TIP60 complex plays also plays a role in DNA repair, promoting homologous recombination (61-63). Thus, taking advantage of its mutants, it will be possible to characterize the function of MRG15 in DNA repair, distinguishing its role from within the NuA4/TIP60 complex versus its association with PALB2/BRAC2.

The MRG interactomes identify a new tetrameric protein complex

When we analyzed the interactomes of MRG15-S/L and MRGX, we noticed by gel, western and mass spectrometry the over-representation of specific subunits of NuA4/TIP60. This was the case for MRGBP and the bromodomain-containing protein BRD8. That drew a parallel with a similar observation we made with yeast NuA4 components a few years ago (25). At that time, we noticed that MRG15 and MRGBP homologs, Eaf3 and Eaf7, were preferentially enriched when used as bait, along the protein Eaf5 which has no clear homolog in higher eukaryotes. We went on to characterize a yeast trimeric complex of Eaf5-Eaf7-Eaf3 that exists independently of NuA4, binds the coding region of active genes and interacts with the elongating RNAPII (25). TINTIN (Trimer Independent of NuA4 for Transcription Interactions with Nucleosomes) is involved in transcription elongation presumably by facilitating disruption and

recycling of nucleosomes from the front to the back of the elongating polymerase, also helping to suppress spurious transcription (64).

Since our purification of MRGBP, the human homolog of Eaf7, also yielded over overrepresentation of MRG15/X and BRD8 compared to the rest of NuA4/TIP60 subunits (**Fig. 3D, 4A**), we wondered if MRG15 in higher eukaryotes could also form an independent functional complex like yeast TINTIN, in which BRD8 would be the functional homolog of yeast Eaf5. To test this hypothesis, we generated K562 cell lines using the *AAVS1* system to purify BRD8 in comparison to MRGBP (**Fig. S2A**). Interestingly, BRD8 possesses two isoforms produced by alternative mRNA splicing, a shorter version with one H4ac-binding bromodomain (BRD) (Q9H0E9-2 to -4) and a longer version with two BRDs (Q9H0E9-1). Using the CRISPR/Cas9 system, we were able to tag both endogenous isoforms. We designed gRNAs to target the C-terminal of the two-BRD isoform or the N-terminal of the *BRD8* gene, tagging both isoforms (**Fig. S3B**). Although clonal selection shows the introduction of our tag in the C-terminus of the long isoform, we were not able to detect expression of this specific variant compared to the N-terminal tagging (**Fig. S3C**). Strangely, even expressing the 2BRD isoform from the *AAVS1* locus yielded very low signal, arguing that this isoform is unstable in K562 cells.

We tandem affinity purified MRGBP-3xFlag-2xStrep and BRD8-3xFlag-2xStrep expressed from *AAVS1* or endogenously tagged and analyzed the fractions by silver staining, mass spectrometry and western blots with the annotated antibodies (**Fig. 4A-B, Table S4**). Surprisingly, in those purifications, but also in our previous MRG15/X purifications, we detected an uncharacterized factor, called EP400 N-terminal like or EP400NL (**Fig. 1D and 4B**, **Tables S1, S3 and S4**). EP400NL was first described as a pseudogene and is homologous to a short region

at the 5' of the EP400 gene (**Fig. S2D**). It can produce multiple spliced variants but only two of them (Q6ZTU2-5 and Q6ZTU2-6) with a specific C-terminal sequence were found associated with MRG proteins and BRD8 by mass spectrometry results (**Table S1, S3 and S4**). Thus, EP400NL proteins share a similar N-terminal sequence with the very large EP400 protein, an essential ATP-dependent remodeler and scaffolding subunit of the NuA4/TIP60 complex (**Fig. S3D**)(34, 65).

EP400NL creates a human TINTIN complex by competing with EP400 for association to trimeric BRD8-MRGBP-MRG15/X

To prove the existence of this functional module independent of the NuA4/TIP60 complex, we purified BRD8 and MRGBP as above, but in parallel to the NuA4/TIP60-specific subunit EPC1 (61). We then loaded the tandem affinity-purified fractions on a Superose 6 size exclusion column to visualise distinct assemblies. Fractions from the gel filtration were analyzed by Western blot, silver-stained gel and mass spectrometry (**Fig. 4C-D**, **Table S5**). While the purified EPC1 fraction eluted early as expected with the large size of the NuA4/TIP60 (~2MDa), BRD8 eluted as two populations with the strongest signal associated with a smaller protein complex (~700–800 KDa) (**Fig. 4C-D**). This smaller population also co-eluted with the strongest signal of MRGBP and MRG15/X proteins. Importantly, in contrast to these proteins, EP400NL was only detected in the smaller size fraction, with no signal eluting with NuA4/TIP60. These results demonstrate the existence of a human TINTIN complex formed by EP400NL-BRD8-MRGBP-MRG15/X. Gel filtration of the purified MRGBP fraction gave similar results but with an important difference. In this case, a third much smaller population was detected below 150kDa

and containing likely only MRGBP and MRG15/X (**Fig. 4C-D**, **Table S5**). Altogether, these data support the idea that the human TINTIN exists independently of the NuA4/TIP60 complex with potential functions in transcription as it was observed in yeast.

Since EP400NL was previously described as a pseudogene, we generated K562 cell lines using both *AAVS1* and CRISPR/Cas9 to, on one hand, tag the specific isoform Q6ZTU2-5 and, on the other hand, the endogenous gene at the C-terminus to tag the two specific isoforms found by mass spectrometry (**Fig. S2E-F**). We then used those cell lines to demonstrate physiological expression of the endogenous *EP400NL* gene and characterize the encoded protein interactome. Analysis of the tandem affinity-purified fractions confirmed expression of EP400NL by the endogenous gene and showed strong, apparently stoichiometric, association with BRD8, MRGBP and MRG15/X (**Fig. 5A-B, Table S6**)(note that MRGBP does not stain efficiently with silver, while showing strong signal by mass spectrometry). Importantly, did not purifiy with any other known components of NuA4/TIP60. Interestingly, an unusually high number of peptides from mRNA processing factors were found, suggesting a potential function of human TINTIN in mRNA processing or at least a close proximity during transcription elongation (**Table S6**).

If EP400NL competes with EP400 to keep human TINTIN separated from the NuA4/TIP60 complex, it is important to identify its binding target. Direct physical interactions between BRD8 and MRGBP as well as MRGBP and MRG15 have been reported (20, 66). We have shown that Eaf5 is responsible for anchoring TINTIN to the yeast NuA4 complex, where it binds the first 85aa of Eaf1, the scaffolding subunit homologous to EP400 (25, 67, 68). Thus, if BRD8 is

really the human functional homolog of yeast Eaf5, it should be responsible to bridge MRGBP-MRG15/X to EP400 in NuA4/TIP60 or EP400NL in TINTIN. To answer this question, we performed a pull-down assay with full-length recombinant BRD8 and GST fusions with homologous N-terminal regions of EP400 (aa194-446) and EP400NL (aa51-297) (see **Fig. S2D**). BRD8 was very efficiently pulled down by both fusions but not GST control, thereby confirming our hypothesis (**Fig. 5C**). Thus, BRD8 acts like the bridge that anchors the MRG proteins to the NuA4/TIP60 complex and EP400NL can block this interaction by associating with BRD8 through its homology to EP400, freeing the TINTIN complex in the cell (**Fig. 5D**).

The human TINTIN complex regulates both transcription and transcript isoforms of specific genes

The yeast TINTIN was shown to interact with the elongating RNAPII and to be involved in recycling disrupted nucleosomes in its wake to suppress spurious transcription (25). To test if human TINTIN also has a role in transcription, we performed siRNA-mediated knockdowns (KDs) of TINTIN components in comparison to KAT5/Tip60 knockdown, in an attempt to distinguish functions associated with NuA4/TIP60 vs independent TINTIN (**Fig. S3A-B**). RNAs were extracted 48h post-KDs and mRNA-sequencing was performed from samples depleted either for BRD8, MRGBP, MRG15 or KAT5/Tip60. 431 common genes were found downregulated in all KDs compared to siControl while and 529 common genes were upregulated (**Fig. 6A**). These likely reflect the function of the NuA4/TIP60 complex. Due to the mutual exclusion of MRG15 and MRGX, we also compared KD of MRGX with the other KDs and

found over 500 common genes downregulated or upregulated (**Fig. S3C**). Thus, some specific genes seem differently affected by the MRGX-containing NuA4/TIP60 complex versus the MRG15-containing one. Furthermore, when we excluded specific targets affected by the KD of KAT5/Tip60, we observed 161 downregulated genes and 124 upregulated genes commonly affected by KDs of TINTIN members BRD8, MRGBP and MRG15 (**Fig. 6A**). These numbers change to 253 downregulated and 142 upregulated when MRG15 KD is replaced by MRGX KD (**Fig. S3C**). Gene ontology term (GO-term) enrichment analyses were performed on those gene sets in order to identify specific biological processes and molecular function pathways in which TINTIN could be implicated (**Fig. 6B-C**, **S3D**). We also used RT-qPCR to validate some targets specifically regulated by the TINTIN complex but not the NuA4/TIP60 complex, such as *RFX8*, *HHIP*, *SYTL5*, *CD180*, which are downregulated by the KDs, and *AKT2* which is upregulated. *PRRG4* and *PCDHB9* are used as control where all KDs were similar to siControl (**Fig. 6D**).

MRG15, its yeast homolog Eaf3 and MRGBP having already been linked to mRNA splicing processes (16-18, 69), we determined if the ratio of alternative transcripts per gene were changed by KDs of TINTIN components. We first compared differential transcript expressions in KDs of BRD8, MRGBP, MRG15/X and KAT5/TIP60 to siControl. Then, we again excluded the transcripts that were significantly changed in KD KAT5/TIP60 (padj ≤ 0.1). The components of the TINTIN complex seem to commonly affect the regulation of only a few alternative transcripts (12 downregulated and 12 upregulated for TINTIN-MRG15, vs 25 down and 18 up for TINTIN-MRGX), even if each KD by itself shows a defect in the expression of much higher numbers of transcripts (**Fig. 6E**). Overall, these data demonstrate a significant function of the human TINTIN complex, independently of the NuA4/TIP60, in transcription regulation by

modulating the expression of specific genes (**Fig. 6F**). We also showed that BRD8 and MRG proteins play also likely important roles in the regulation of some genes from within the NuA4/TIP60 acetyltransferase/remodeling complex or through colocalization on the same target of NuA4/TIP60 at the promoter and TINTIN on the coding region, as we have described in yeast (**Fig. 6F**)(25).

Discussion

In this work, we characterized the full interactomes of human MRG domain-containing proteins MRG15 and MRGX, factors that have been implicated in several distinct nuclear processes through association with diverse proteins. These experiments were done as to reflect the most native interactions, avoiding over-expression and when possible, using endogenous proteins as bait. We observed that MRG15 isoforms with long or short CHD and MRGX share the same interactome, such as stable association with the NuA4/TIP60 histone acetyltransferase complex and the Sin3B histone deacetylase/demethylase complex. Interestingly, MRG15 and MRGX are mutually exclusive in their associated complexes. Since MRGX does not possess the H3K36me3-binding chromodomain, it will be important to characterize the impact of this difference on the function of their partners. *MRG15* knock-out is embryonic lethal in mice, while *MRGX* knock-out does not lead to any obvious phenotype in development or cell proliferation, which may in part due to some unidirectional compensation mechanism (19, 70, 71).

Interestingly, our data show that MRGX and MRG15 are stably associated with the endogenous H3K36me1/me2 methyltransferase ASH1L, in an apparent stoichiometric ratio since its activity cannot be stimulated *in vitro* by addition of exogenous MRG15 to relieve an auto-inhibitory loop (**Fig. 2**) (54-57). It will be interesting to study the functional impact of the MRG15 mutations that disrupt its association with endogenous ASH1L *in vivo*, through genome-wide location analysis of ASH1L and H3K36me2. The same mutants also disrupt association with the Sin3B complex. It was shown in yeast that Eaf3 plays a crucial role in the function of the homologous RPD3S complex to deacetylate nucleosomes in the wake of the elongating polymerase to suppress spurious transcription (24). A related function was proposed for the mammalian Sin3B complex (22, 72). Thus, it will be interesting to do transcriptomics to analyze if spurious transcription, i.e. cryptic initiation, anti-sense transcription, is detected with the MRG15 mutants. Since Sin3B also contains the KDM5 H3K4me3 demethylase that blocks the spreading of this mark from the TSS region to the body of the genes, it will also be important to verify the impact of the MRG15 mutant on the localization of this histone mark (72).

Most importantly, this study allowed the identification and characterization of a new functional MRG-containing complex, human TINTIN. We discovered a new partner of MRG15, EP400NL, previously dismissed as a probable pseudogene. EP400NL is key to create a BRD8-MRGBP-MRG15/X-containing TINTIN complex physically independent of NuA4/TIP60. EP400NL is homologous to the N-terminal region of NuA4/TIP60 scaffold subunit EP400, the interface on which the BRD8-MRGBP-MRG15/X trimer is normally anchored to the complex through BRD8. EP400NL competes with this binding and by doing so free the trimer from NuA4/TIP60, creating TINTIN. Interestingly, the *EP400* gene is also known to produce a short splice variant encoding a

short protein corresponding to EP400NL, which could play a similar role with TINTIN. Importantly, these proteins lack the HSA and SANT domains of EP400 that are required to assemble the NuA4/TIP60 complex (67, 68).

Since the yeast TINTIN complex is found on the coding region of active genes and plays a role during transcription elongation, we investigated if human TINTIN also affects gene expression. Indeed, several genes are deregulated in the absence of TINTIN components, independently of NuA4/TIP60 (**Fig. 6**). Further work will be necessary to better understand the function of TINTIN in mammals. EP400NL is likely an important target to use but, as mentioned above, EP400 short isoforms produced by alternative splicing may play a redundant role. Mapping the precise region of interaction between BRD8 and EP400NL/EP400 may be useful to design better tools to study TINTIN specific function.

Although MRG15 and MRGBP have been proposed to regulate mRNA splicing (16, 17, 69), we could not find a significant number of alternate transcripts specifically affected by the human TINTIN complex. Our mass spectrometry data also did not detect previously identified partners of MRG15 (PTB) and MRGBP (VEZF) for such function. It is possible that these interactions mostly occur on chromatin, making them difficult to detect in extracts. Independent KDs of MRG15, MRGX and MRGBP do show a larger number of transcripts being misregulated (**Fig. 6E**). This may be linked instead to the function of the small but abundant MRGBP-MRG15/X dimer that our gel filtration experiment identified (**Fig. 4**).

In conclusion, this study presented a highly confident native stable interactome of MRG proteins and uncovered a new mammalian complex implicated in gene regulation. It highlights

the multi-specificity of the MRG domain for protein-protein interactions and associations with distinct protein complexes implicated in several chromatin-based nuclear processes. The presence of H4ac-binding bromodomain and H3K35me3-binding chromodomain in TINTIN also underlines again the cross-talk between transcription and chromatin modifications.

Acknowledgments

We thank Catherine Lachance, Anahita Lashgari, Alexandre Prudente and Suk Min Jang for work linked to this study and Tatiana Kutateladze for the recombinant ASH1L construct. This work was supported by a grant from the Canadian Institutes of Health Research (FDN-143314) to J.C. J.C. holds the Canada Research Chair in Chromatin Biology and Molecular Epigenetics.

Author contributions

MD and CR performed experiments. CB and SH performed bioinformatic analyses. AD and JC supervised and secured funding. MD and JC designed experiments, analyzed the data, and prepared the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A˚ resolution. Nature 389:251-260.
2. Strahl BD, Allis CD. 2000. The language of covalent histone modifications. Nature 403.
3. Torres IO, Fujimori DG. 2015. Functional coupling between writers, erasers and readers of histone and DNA methylation. Current opinion in structural biology 35:68-75.
4. Rothbart SB, Strahl BD. 2014. Interpreting the language of histone and DNA modifications. Biochimica et biophysica acta 1839:627-43.
5. Dawson MA, Kouzarides T, Huntly BJ. 2012. Targeting epigenetic readers in cancer. The New England journal of medicine 367:647-57.
6. Venkatesh S, Workman JL. 2013. Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. Wiley interdisciplinary reviews Developmental biology 2:685-700.
7. Edmunds JW, Mahadevan LC, Clayton AL. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. The Embo Journal 27:406–420.
8. Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Molecular and cellular biology 25:3305-16.
9. Lee JS, Shilatifard A. 2007. A site to remember: H3K36 methylation a mark for histone deacetylation. Mutation research 618:130-4.
10. Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. The Journal of biological chemistry 280:17732-6.
11. Vakoc CR, Sachdeva MM, Wang H, Blobel GA. 2006. Profile of histone lysine methylation across transcribed mammalian chromatin. Molecular and cellular biology 26:9185-95.
12. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. Nature genetics 41:376-81.
13. Wilhelm BT, Marguerat S, Aligianni S, Codlin S, Watt S, Bahler J. 2011. Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. Genome biology 12:R82.
14. Spies N, Nielsen CB, Padgett RA, Burge CB. 2009. Biased chromatin signatures around polyadenylation sites and exons. Molecular cell 36:245-54.
15. de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I, Andrau JC, Ferrier P, Carmo-Fonseca M. 2011. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nature structural & molecular biology 18:977-83.

16. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. Science 327:996-1000.
17. Iwamori N, Tominaga K, Sato T, Riehle K, Iwamori T, Ohkawa Y, Coarfa C, Ono E, Matzuk MM. 2016. MRG15 is required for pre-mRNA splicing and spermatogenesis. PNAS:E5408–E5415.
18. Leung CS, Douglass SM, Morselli M, Obusan MB, Pavlyukov MS, Pellegrini M, Johnson TL. 2019. H3K36 Methylation and the Chromodomain Protein Eaf3 Are Required for Proper Cotranscriptional Spliceosome Assembly. Cell reports 27:3760-3769 e4.
19. Chen M, Tominaga K, Pereira-Smith OM. 2010. Emerging role of the MORF/MRG gene family in various biological processes, including aging. Annals of the New York Academy of Sciences 1197:134-41.
20. Xie T, Zmyslowski AM, Zhang Y, Radhakrishnan I. 2015. Structural Basis for Multi-specificity of MRG Domains. Structure 23:1049-57.
21. Doyon Y, Cote J. 2004. The highly conserved and multifunctional NuA4 HAT complex. Current opinion in genetics & development 14:147-54.
22. Jelinic P, Pellegrino J, David G. 2011. A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. Molecular and cellular biology 31:54-62.
23. Hayakawa T, Ohtani Y, Hayakawa N, Shinmyozu K, Saito M, Ishikawa F, Nakayama J. 2007. RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. Genes Cells 12:811-26.
24. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123:581-92.
25. Rossetto D, Cramet M, Wang AY, Steunou AL, Lacoste N, Schulze JM, Cote V, Monnet-Saksouk J, Piquet S, Nourani A, Kobor MS, Cote J. 2014. Eaf5/7/3 form a functionally independent NuA4 submodule linked to RNA polymerase II-coupled nucleosome recycling. The Embo Journal 33:1397-415.
26. Chen Y, Li J, Dunn S, Xiong S, Chen W, Zhao Y, Chen BB, Mallampalli RK, Zou C. 2014. Histone deacetylase 2 (HDAC2) protein-dependent deacetylation of mortality factor 4-like 1 (MORF4L1) protein enhances its homodimerization. The Journal of biological chemistry 289:7092-7098.
27. Sy SM, Huen MS, Chen J. 2009. MRG15 is a novel PALB2-interacting factor involved in homologous recombination. The Journal of biological chemistry 284:21127-31.
28. Ducy M, Sesma-Sanz L, Guitton-Sert L, Lashgari A, Gao Y, Brahiti N, Rodrigue A, Margaillan G, Caron MC, Cote J, Simard J, Masson JY. 2019. The Tumor Suppressor PALB2: Inside Out. Trends in biochemical sciences 44:226-240.
29. Hayakawa T, Zhang F, Hayakawa N, Ohtani Y, Shinmyozu K, Nakayama J, Andreassen PR. 2010. MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. Journal of cell science 123:1124-30.
30. Bleuyard JY, Fournier M, Nakato R, Couturier AM, Katou Y, Ralf C, Hester SS, Dominguez D, Rhodes D, Humphrey TC, Shirahige K, Esashi F. 2017. MRG15-mediated tethering of PALB2 to unperturbed chromatin protects active genes from genotoxic stress.

Proceedings of the National Academy of Sciences of the United States of America 114:7671-7676.

31. Wei Y, Tian C, Zhao Y, Liu X, Liu F, Li S, Chen Y, Qiu Y, Feng Z, Chen L, Zhou T, Ren X, Feng C, Liu Y, Yu W, Ying H, Ding Q. 2020. MRG15 orchestrates rhythmic epigenomic remodelling and controls hepatic lipid metabolism. Nat Metab 2:447-460.
32. Lalonde ME, Avvakumov N, Glass KC, Joncas FH, Saksouk N, Holliday M, Paquet E, Yan K, Tong Q, Klein BJ, Tan S, Yang XJ, Kutateladze TG, Cote J. 2013. Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. Genes & development 27:2009-24.
33. Saksouk N, Avvakumov N, Champagne KS, Hung T, Doyon Y, Cayrou C, Paquet E, Ullah M, Landry AJ, Cote V, Yang XJ, Gozani O, Kutateladze TG, Cote J. 2009. HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. Molecular cell 33:257-65.
34. Dalvai M, Loehr J, Jacquet K, Huard CC, Roques C, Herst P, Cote J, Doyon Y. 2015. A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. Cell reports 13:621-633.
35. Agudelo D, Duringer A, Bozoyan L, Huard CC, Carter S, Loehr J, Synodinou D, Drouin M, Salsman J, Dellaire G, Laganiere J, Doyon Y. 2017. Marker-free coselection for CRISPR-driven genome editing in human cells. Nature methods 14:615-620.
36. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75:4646-58.
37. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.
38. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923-30.
39. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-40.
40. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884-i890.
41. de Sena Brandine G, Smith A, D. 2021. Falco: high-speed FastQC emulation for quality control of sequencing data. f1000research 8:1874.
42. Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047-8.
43. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology 34:525-7.
44. Kirby KN, Gerlanc D. 2013. BootES: an R package for bootstrap confidence intervals on effect sizes. Behavior research methods 45:905-27.
45. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15:550.
46. Ruan C, Lee CH, Cui H, Li S, Li B. 2015. Nucleosome contact triggers conformational changes of Rpd3S driving high-affinity H3K36me nucleosome engagement. Cell reports 10:204-15.

47. Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, Natarajan K, Workman JL. 2012. Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature 489:452-5.
48. Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, Huang Q, Liu Q, Hao Q, Ding J. 2006. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic acids research 34:6621-8.
49. Doyon Y, Cote J. 2016. Preparation and Analysis of Native Chromatin-Modifying Complexes. Methods Enzymol 573:303-18.
50. Larance M, Kirkwood KJ, Xirodimas DP, Lundberg E, Uhlen M, Lamond AI. 2012. Characterization of MRFAP1 turnover and interactions downstream of the NEDD8 pathway. Molecular & cellular proteomics : MCP 11:M111 014407.
51. Zhu L, Li Q, Wong SH, Huang M, Klein BJ, Shen J, Ikenouye L, Onishi M, Schneidawind D, Buechele C, Hansen L, Duque-Afonso J, Zhu F, Martin GM, Gozani O, Majeti R, Kutateladze TG, Cleary ML. 2016. ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia. Cancer discovery 6:770-83.
52. An S, Yeo KJ, Jeon YH, Song JJ. 2011. Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. The Journal of biological chemistry 286:8369-8374.
53. Balbo Pogliano C, Gatti M, Ruthemann P, Garajova Z, Penengo L, Naegeli H. 2017. ASH1L histone methyltransferase regulates the handoff between damage recognition factors in global-genome nucleotide excision repair. Nature communications 8:1333.
54. Huang C, Yang F, Zhang Z, Zhang J, Cai G, Li L, Zheng Y, Chen S, Xi R, Zhu B. 2017. Mrg15 stimulates Ash1 H3K36 methyltransferase activity and facilitates Ash1 Trithorax group protein function in Drosophila. Nature communications 8:1649.
55. Schmahling S, Meiler A, Lee Y, Mohammed A, Finkl K, Tauscher K, Israel L, Wirth M, Philippou-Massier J, Blum H, Habermann B, Imhof A, Song JJ, Muller J. 2018. Regulation and function of H3K36 di-methylation by the trithorax-group protein complex AMC. Development 145.
56. Lee Y, Yoon E, Cho S, Schmahling S, Muller J, Song JJ. 2019. Structural Basis of MRG15-Mediated Activation of the ASH1L Histone Methyltransferase by Releasing an Autoinhibitory Loop. Structure 27:846-852 e3.
57. Hou P, Huang C, Liu CP, Yang N, Yu T, Yin Y, Zhu B, Xu RM. 2019. Structural Insights into Stimulation of Ash1L's H3K36 Methyltransferase Activity through Mrg15 Binding. Structure 27:837-845 e3.
58. Hu X, Harvey SE, Zheng R, Lyu J, Grzeskowiak CL, Powell E, Piwnica-Worms H, Scott KL, Cheng C. 2020. The RNA-binding protein AKAP8 suppresses tumor metastasis by antagonizing EMT-associated alternative splicing. Nature communications 11:486.
59. Xie T, Graveline R, Kumar GS, Zhang Y, Krishnan A, David G, Radhakrishnan I. 2012. Structural basis for molecular interactions involving MRG domains: implications in chromatin biology. Structure 20:151-60.
60. Wu J, Chen Y, Lu LY, Wu Y, Paulsen MT, Ljungman M, Ferguson DO, Yu X. 2011. Chfr and RNF8 synergistically regulate ATM activation. Nature structural & molecular biology 18:761-8.

61. Jacquet K, Fradet-Turcotte A, Avvakumov N, Lambert JP, Roques C, Pandita RK, Paquet E, Herst P, Gingras AC, Pandita TK, Legube G, Doyon Y, Durocher D, Cote J. 2016. The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation. Molecular cell 62:409-421.
62. Bassi C, Li YT, Khu K, Mateo F, Baniasadi PS, Elia A, Mason J, Stambolic V, Pujana MA, Mak TW, Gorrini C. 2016. The acetyltransferase Tip60 contributes to mammary tumorigenesis by modulating DNA repair. Cell death and differentiation 23:1198-208.
63. Li ML, Jiang Q, Bhanu NV, Wu J, Li W, Garcia BA, Greenberg RA. 2019. Phosphorylation of TIP60 Suppresses 53BP1 Localization at DNA Damage Sites. Molecular and cellular biology 39.
64. Bhat W, Ahmad S, Cote J. 2015. TINTIN, at the interface of chromatin, transcription elongation, and mRNA processing. RNA biology 12:486-9.
65. Pradhan SK, Su T, Yen L, Jacquet K, Huang C, Cote J, Kurdistani SK, Carey MF. 2016. EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. Mol Cell 61:27-38.
66. Yamaguchi K, Sakai M, Shimokawa T, Yamada Y, Nakamura Y, Furukawa Y. 2010. C20orf20 (MRG-binding protein) as a potential therapeutic target for colorectal cancer. British journal of cancer 102:325-31.
67. Setiaputra D, Ahmad S, Dalwadi U, Steunou AL, Lu S, Ross JD, Dong MQ, Cote J, Yip CK. 2018. Molecular Architecture of the Essential Yeast Histone Acetyltransferase Complex NuA4 Redefines Its Multimodularity. Mol Cell Biol 38.
68. Wang X, Ahmad S, Zhang Z, Cote J, Cai G. 2018. Architecture of the Saccharomyces cerevisiae NuA4/TIP60 complex. Nat Commun 9:1147.
69. Gowher H, Brick K, Camerini-Otero DR, Felsenfeld G. 2012. Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. Proc Natl Acad Sci U S A 109(7):2370-5.
70. Tominaga K, Matzuk MM, Pereira-Smith OM. 2005. MrgX is not essential for cell growth and development in the mouse. Molecular and cellular biology 25:4873-80.
71. Tominaga K, Kirtane B, Jackson JG, Ikeno Y, Ikeda T, Hawks C, Smith JR, Matzuk MM, Pereira-Smith OM. 2005. MRG15 regulates embryonic development and cell proliferation. Mol Cell Biol 25:2924-37.
72. Xie L, Pelz C, Wang W, Bashar A, Varlamova O, Shadle S, Impey S. 2011. KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. EMBO J 30:1473-84.

FIGURE LEGENDS

Figure 1. The interactome of MRG proteins highlights the association with several distinct complexes.

(A) Schematic representation of MRG15 spliced variants, MRGX, and the yeast Eaf3 functional domains.

(B) Short chromodomain MRG15 binds preferentially H3K36me3 *in vitro*. Recombinant GST-tagged MRG15 isoforms or GST or beads were incubated with LON followed by washes and western blotting. Anti-H4 and anti-H3K79me3 are used as control.

(C) 4-12% SDS-page of 3xFlag-2xStrep elutions of indicated purification from K562 cells were analyzed by silver staining. MRGX and MRG15 spliced variants bind the acetyltransferase NuA4/TIP60 complex. Subunits were identified by mass spectrometry (Table S1).

(D-E) Purified complexes from C) were analyzed by western blot with the indicated antibodies to confirm the equivalent presence of known subunits of the NuA4/TIP60 complex (D) and the other known interactors of MRG proteins (E) (* non-specific band).

Figure 2. Endogenous ASH1L is stably associated with MRG proteins.

(A) Purification of endogenous ASH1L from K562 cells modified by CRISPR/Cas9 system to add 3xFlag 2xStrep. The elution was analyzed on 4-12% SDS-PAGE followed by silver staining.

(B) Purified ASH1L from A) was analyzed by western blot with the indicated antibodies to confirm the interaction with ASH1L and its interactors such as MRG proteins and AKAP8.

(C) *In vitro* histone methyltransferase assay performed with recombinant ASH1L (aa2040-2634)-pGEX6P2 (recAHS1L) and strep elution from endogenous ASH1L (endoASH1L). The graph shows the scintillation counts of the liquid assays with 3H S- adenosyl methionine (SAM) with native short oligonucleosomes (SON). An empty vector pGEX6P2 and a mock cell line were used as control and subtracted from final counts. Error bars represent the range of two technical replicates.

Figure 3. Structure-based mutations can differentially affect specific MRG15-containing complexes.

(A) Schematic representation of W172A Y235A MRG15 mutant leading only MRGBP interaction, and W78A F105A MRGBP mutant preventing MRG15 interaction.

(B) 4-12% SDS-page of 3xFlag-2xStrep elutions of indicated purification from K562 cells were analyzed by silver staining. Strep elutions were analyzed by mass spectrometry in total spectrum count (Table 4).

(C) Purified complexes from (B) were analyzed by western blot with the indicated antibodies (* non-specific band).

Figure 4. Identification of the TINTIN complex in human cells, composed of BRD8-MRGBP-MRG15/X and new subunit EP400NL.

(A) Purification of native NuA4/TIP60 complex from K562 cells. Strep elutions of indicated purifications were analyzed on 4-12% SDS-PAGE followed by silver staining. MRGBP and spliced variants of BRD8 (1BRD, one bromodomain, 2BRD, two bromodomains, and CRISPR for N-term tagged BRD8) are linked to NuA4/TIP60 complex.

(B) Western blots of purified complexes in (A) confirmed the interaction with NuA4/TIP60 complex and the new subunit EP400NL.

(C) Western blots of gel filtration (FPLC) from fractions 17 to 33 representing annotated complex from flag elutions of EPC1, BRD8, and MRGBP AAVS1 K562 cell lines.

(D) 4-12% SDS-PAGE followed by silver staining of annotated fractions showing NuA4/TIP60 complex (17), TINTIN complex (23), and dimer MRGBP/MRG15 or MRGX (33) from flag elutions of BRD8, MRGBP, and EPC1.

Figure 5. Endogenous EP400NL isolates TINTIN through an interaction with BRD8.

(A) 3xFlag-2xStrep elutions of Q6ZTU2-5 (AAVS1) or endogenous (CRISPR) purified EP400NL from K562 cells analyzed by silver staining.

(B) Purified complexes from (A) were analyzed by western blot with the indicated antibodies to confirm the interaction with BRD8, MRGBP, and MRG15/X representing the human TINTIN complex.

(C) BRD8 interacts with the N-term region of EP400NL *in vitro*. Recombinant His-tagged BRD8 was incubated with GST fusion proteins on beads followed by washes and western blotting.

(D) Schematic representation of BRD8-MRGBP-MRG15/X that can bind EP400 to form NuA4/TIP60 complex or the new subunit EP400NL to form TINTIN complex.

Figure 6. Human TINTIN is implicated in the expression of specific genes.

(A) 161 genes are commonly downregulated (left) and 124 are upregulated in KDs of BRD8, MRGBP, and MRG15 but not in KD of TIP60, in U2OS cells. We kept the genes with high changes in expression, using a cutoff of 2-fold difference was applied to the Log2(fold change) values between the selected KDs and control (siLuc).

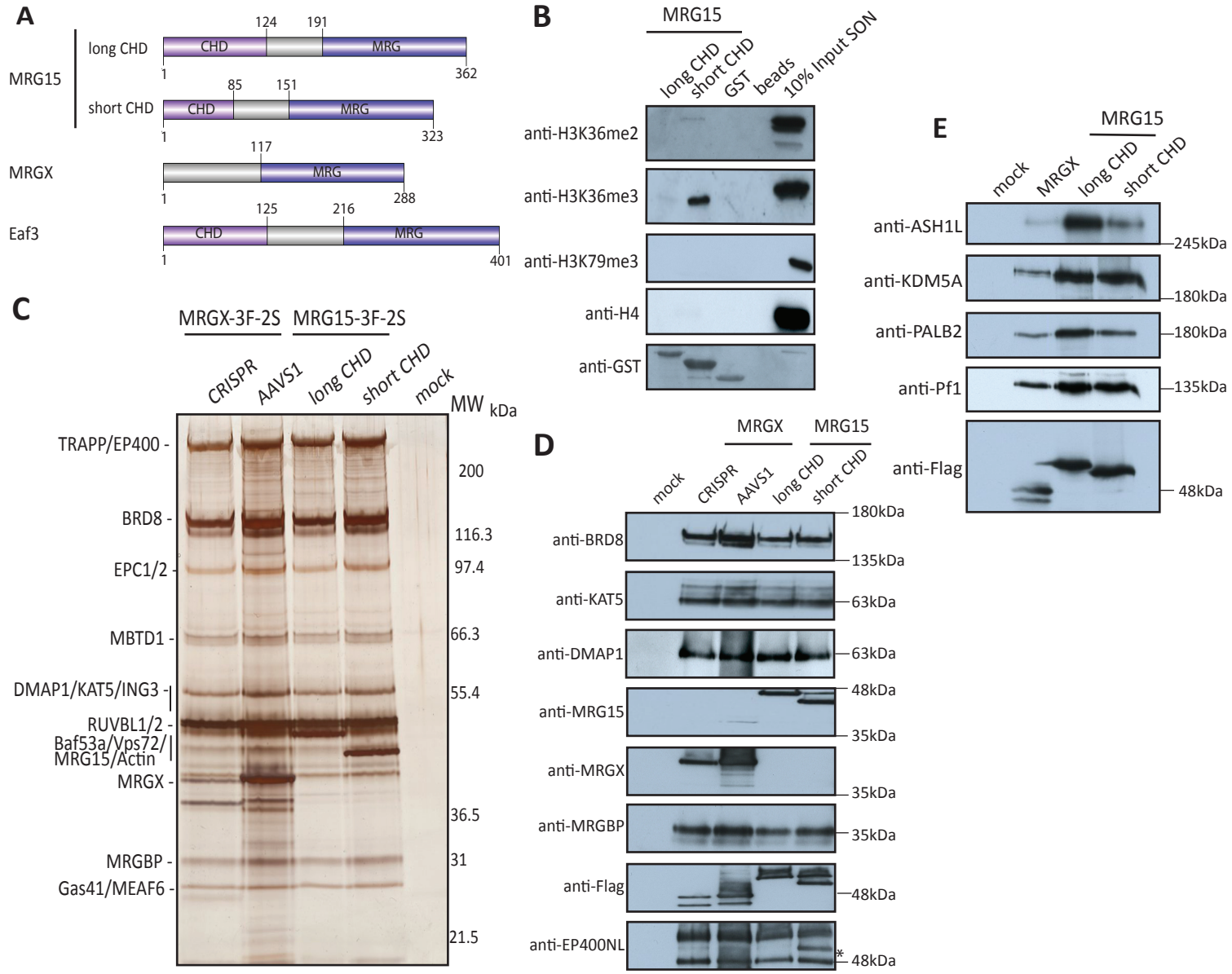
(B) and (C) Gene ontology analysis for downregulated (D) and upregulated genes (C) showing a significant enrichment (pValue <0.01) using DAVID 6.8.

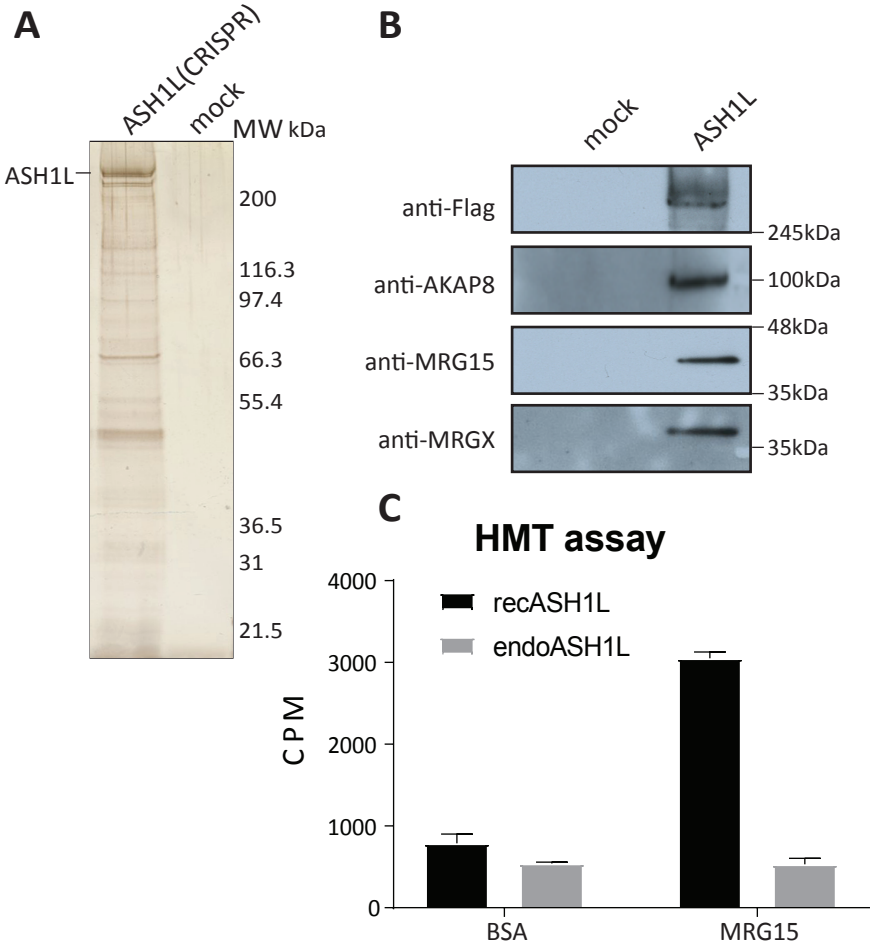
(D) RT-qPCR to validate target genes of TINTIN. *RFX8*, *HHIP*, *SYTL5*, and *CD180* are found downregulated, and *AKT2* is upregulated in KD TINTIN subunits. *PRRG4* and *PCDHB9* are used as control. Values represent means relative to control (siLuc). Error bars represent the range from two biological replicates.

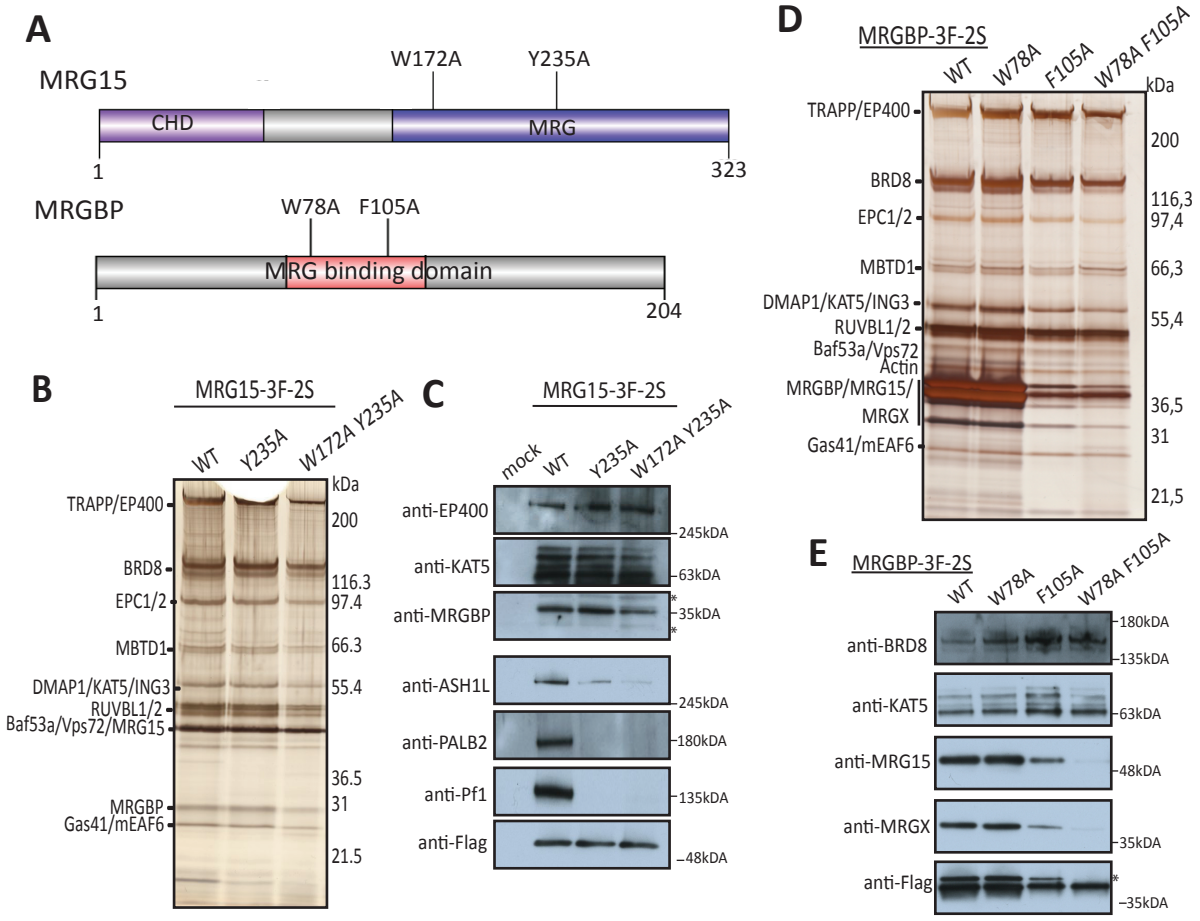
(E) Venn diagrams of downregulated (left) and upregulated (right) transcripts, specifically regulated by KDBRD8, MRGBP, MRG15, or MRGX after exclusion of transcripts regulated by KDKAT5/TIP60, using a cutoff of 2-fold difference. siLuciferase was used as control. Two replicates of each KD were performed in U2OS cells.

(F) BRD8 brings MRG proteins into the acetyltransferase NuA4/TIP60 complex due to its direct interaction with the EP400-N-terminal fraction. These subunits are important for gene expressions of specific targets. Furthermore, those subunits form a new submodule, called TINTIN, when BRD8 interacts with a new factor, EP400NL. TINTIN is also important for transcription regulation of specific genes.

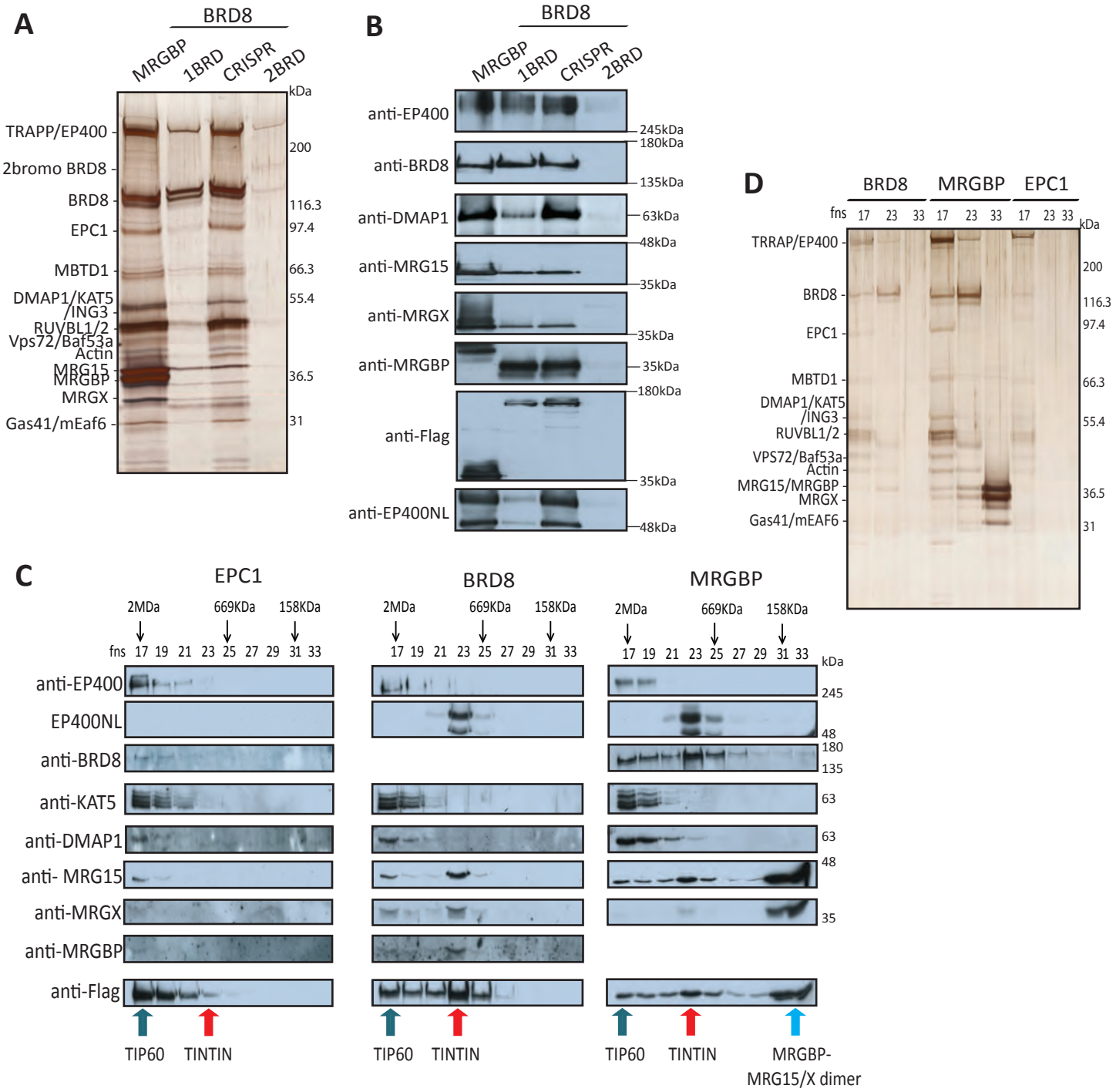
Devoucoux Figure 1

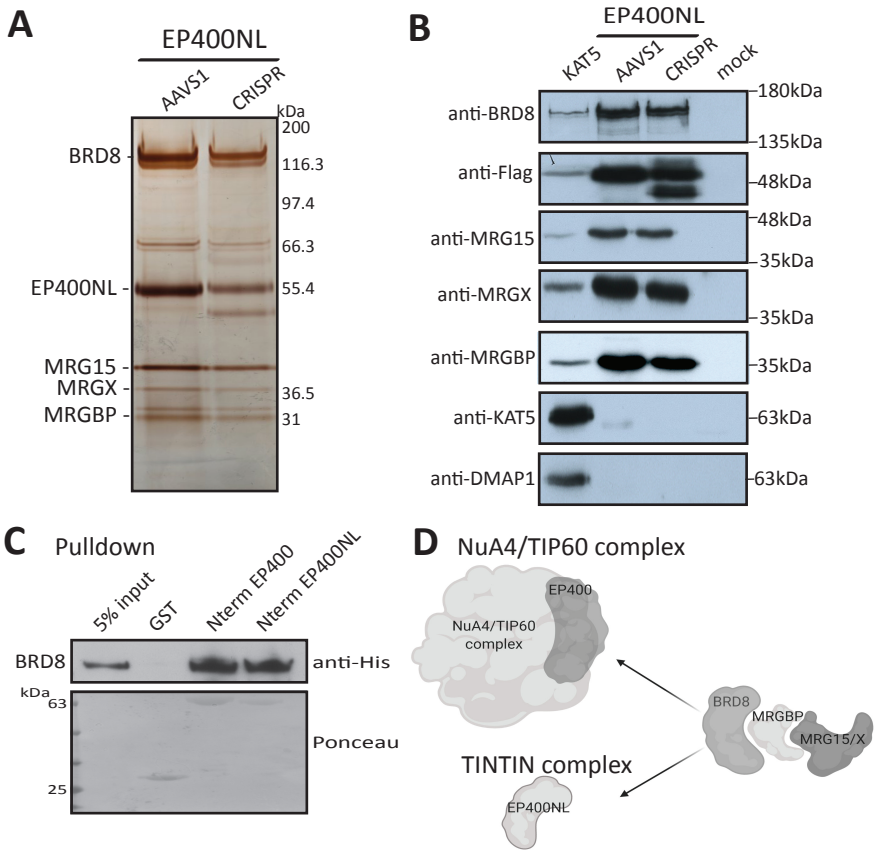



A
ASH1L(CRISPR)
mock
MW kDa
ASH1L
200
116.3
97.4
66.3
55.4
36.5
31
21.5
B
mock
ASH1L
anti-Flag
245kDa
anti-AKAP8
100kDa
48kDa
anti-MRG15
35kDa
anti-MRGX
35kDa
C
HMT assay
recASH1L
endoASH1L
C P M
4000
3000
2000
1000
0
BSA
MRG15


A
MRG15
W172A
Y235A
CHD
MRG
1
323
MRGBP
W78A
F105A
MRG binding domain
1
204
B
MRG15-3F-2S
WT
Y235A
W172A Y235A
kDa
TRAPP/EP400
BRD8
EPC1/2
MBTD1
DMAP1/KAT5/ING3
RUVBL1/2
Baf53a/Vps72/MRG15
MRGBP
Gas41/mEAF6
200
116.3
97.4
66.3
55.4
36.5
31
21.5
C
MRG15-3F-2S
mock
WT
Y235A
W172A Y235A
anti-EP400
anti-KAT5
anti-MRGBP
anti-ASH1L
anti-PALB2
anti-Pf1
anti-Flag
245kDA
63kDA
35kDA
245kDA
180kDA
135kDA
48kDA
D
MRGBP-3F-2S
WT
W78A
F105A
W78A F105A
kDa
TRAPP/EP400
BRD8
EPC1/2
MBTD1
DMAP1/KAT5/ING3
RUVBL1/2
Baf53a/Vps72
Actin
MRGBP/MRG15/
MRGX
Gas41/mEAF6
200
116,3
97,4
66,3
55,4
36,5
31
21,5
E
MRGBP-3F-2S
WT
W78A
F105A
W78A F105A
anti-BRD8
anti-KAT5
anti-MRG15
anti-MRGX
anti-Flag
180kDA
135kDA
63kDA
48kDA
35kDA
35kDA

Devoucoux Figure 4




A
EP400NL
AAVS1
CRISPR
kDa
200
116.3
97.4
66.3
55.4
36.5
31
BRD8
EP400NL
MRG15
MRGX
MRGBP
B
EP400NL
KAT5
AAVS1
CRISPR
mock
anti-BRD8
anti-Flag
anti-MRG15
anti-MRGX
anti-MRGBP
anti-KAT5
anti-DMAP1
180kDa
135kDa
48kDa
48kDa
35kDa
35kDa
35kDa
63kDa
63kDa
C
Pulldown
5% input
GST
Nterm EP400
Nterm EP400NL
BRD8
anti-His
kDa
63
25
Ponceau
D
NuA4/TIP60 complex
EP400
NuA4/TIP60 complex
BRD8
MRGBP
MRG15/X
TINTIN complex
EP400NL

Devoucoux Figure 6

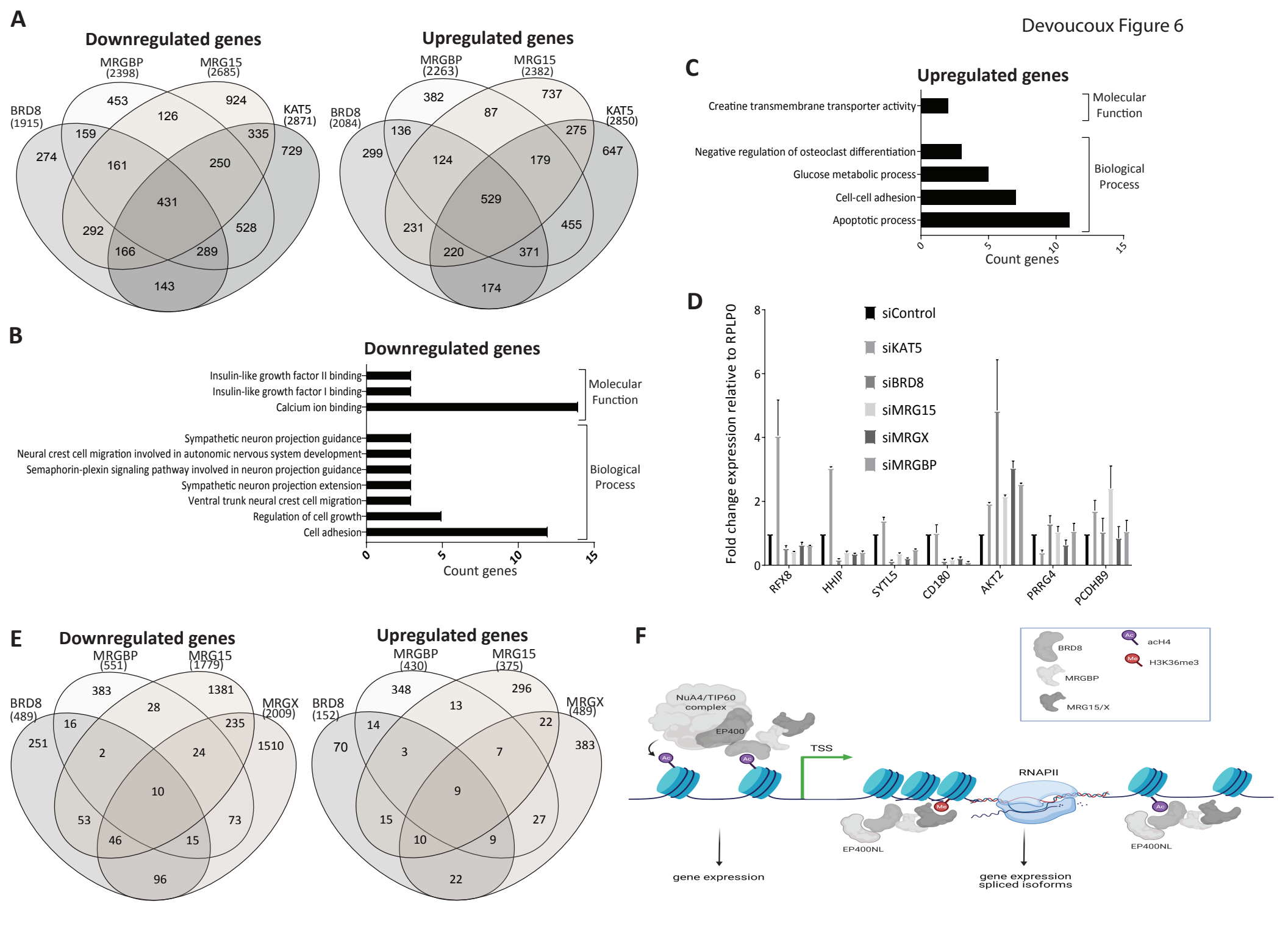


Table S1. Mass spectrometry analysis of proteins that copurified with MRGX, MRG15 long and short CHD from K562 nuclear extracts and after DNA damage (NCS) (Total spectral counts), Related to Fig. 1.

				MRG15		MRG15+NCS
NuA4/Tip60	**MW**	**Mock**	**MRGX**	**long CHD**	**Short CHD**	**short CHD**
TRRAP	436 kDa	0	174	184	392	547
EP400	336 kDa	0	175	143	301	348
BRD8	97 kDa	0	85	80	138	162
EPC1	91 kDa	0	22	16	36	53
EPC2	91 kDa	0	21	28	59	76
MBTD1	71 kDa	0	11	21	35	38
KAT5/TIP60	53 kDa	0	11	9	32	45
DMAP1	53 kDa	0	37	27	45	51
RUVBL2	51 kDa	0	61	58	93	106
RUVBL1	50 kDa	0	53	49	103	103
Baf53a	47 kDa	0	37	28	47	60
ING3	45 kDa	0	9	11	31	39
ACTB	42 kDa	0	30	30	38	47
Vps72	42 kDa	0	8	10	23	21
MRG15	40 kDa	0	17	54	152	152
MRGX	32 kDa	0	138	6	0	0
GAS41	27 kDa	0	20	17	21	22
MRGBP	22 kDa	0	21	9	34	42
MEAF6	22 kDa	0	5	7	10	8
H2B	14 kDa	0	2	2	8	8

Sin3B	**MW**	**Mock**	**MRGX**	**long CHD**	**Short CHD**	**short CHD**
KDM5A	196 kDa	0	3	9	29	36
Sin3B	129 kDa	0	1	7	26	39
Pf1/PHF12	110 kDa	0	8	9	27	41
HDAC1	52 kDa	0	0	0	3	7
RBBP7	47 kDa	0	1	2	0	0
EMSY	131 kDa	0	14	23	48	46

Others	**MW**	**Mock**	**MRGX**	**long CHD**	**Short CHD**	**short CHD**
EP400NL	45 kDa	0	32	21	67	73
BRCA2	384 kDa	0	8	16	60	83
PALB2	131 kDa	0	5	3	22	22
ASH1L	332 kDa	0	0	2	27	33
MRFAP1	15 kDa	0	12	3	6	2

Table S2. Mass spectrometry analysis of endogenous ASH1L purified from K562 cells (total spectral counts). Related to Figure 2.

Identified Proteins	MW	ASH1L
ASH1L	332 kDa	222
MRG15	40 kDa	17
MRGX	32 kDa	10
RBBP4/ NuRF55	48 kDa	12
AKAP8	76 kDa	13
AKAP8L	72 kDa	7
KMT2A	432 kDa	4
HDAC1	55 kDa	7

Table S3. Mass spectrometry analysis of MRG15 WT versus mutant interactomes purified from K562 cells through MRG15-3xFlag 2xStrep integrated at AAVS1 (total spectral counts). Related to Figure 3.

			MRG15		
	Identified proteins	**MW**	**WT**	**Y235A**	**W172A Y235A**
NuA4/TIP60 complex	TRRAP	436 kDa	94	25	14
	EP400	343 kDa	183	159	52
	BRD8	103 kDa	73	47	23
	EPC2	91 kDa	38	20	11
	EPC1	91 kDa	24	16	6
	MBTD1	71 kDa	9	5	1
	KAT5	62 kDa	25	10	4
	KAT5	62 kDa	25	10	4
	DMAP1	53 kDa	34	10	9
	RUVBL2	51 kDa	126	56	31
	RUVBL1	50 kDa	62	43	17
	ACTL6A	47 kDa	35	21	15
	ING3	45 kDa	13	1	2
	ACTB	42 kDa	33	20	45
	VPS72	41 kDa	16	4	4
	MRG15	37 kDa	143	117	82
	YEATS4	27 kDa	12	4	3
	MRGBP	22 kDa	20	13	7
	MEAF6	22 kDa	2	0	0
TINTIN	EP400NL	52 kDa	51	30	12
Sin3B complex	EMSY	141 kDa	37	0	0
	PHF12/Pf1	110 kDa	13	0	0
	SIN3B	133 kDa	5	0	0
	KDM5A	192 kDa	8	0	0
	RBBP4	48 kDa	2	0	0
Other factors	BRCA2	384 kDa	15	0	0
	ASH1L	333 kDa	7	0	0
	PALB2	131 kDa	9	0	0
	MRFAP1	15 kDa	5	0	0

Table S4. Mass spectrometry analysis of NuA4/TIP60 complex and EP400NL that interact with MRGBP and BRD8 transcripts (with 2BRDs or all transcripts (N-term CRISPR) or 1BRD), purified from K562 cells (total spectral counts). Related to Figure 4.

		BRD8			
Identified Proteins	**MW**	**Endogenous**	**1 BRD**	**2BRD**	**MRGBP**
TRRAP	436 kDa	330	57	33	292
EP400	340 kDa	275	74	42	296
RUVBL2	51 kDa	127	39	26	119
BRD8	135 kDa	117	68	23	114
RUVBL1	50 kDa	84	26	18	93
EPC2	91 kDa	68	15	5	53
DMAP1	53 kDa	59	18	7	44
ACTL6A	47 kDa	47	17	8	43
ATCB	42 kDa	46	13	11	40
MRG15	40 kDa	44	35	7	200
MRGX	32 kDa	42	34	9	118
MBTD1	71 kDa	30	7	3	21
EPC1	91 kDa	33	11	2	38
YEATS4	27 kDa	27	13	3	21
MRGBP	22 kDa	23	17	5	89
ING3	45 kDa	23	2	0	24
KAT5	53 kDa	20	8	0	22
VPS72	42 kDa	14	6	3	18
MEAF6	22 kDa	8	3	0	8
H2B	14 kDa		2		2
EP400NL	45 kDa	64	11	0	63

Table S5. Mass spectrometry analysis of FPLC fraction number 23 of BRD8 and MRGBP representing TINTIN complex and number 33 of MRGBP for MRGBP/MRG15 or MRGX dimer (total spectral counts). Related to Figure 4.

		BRD8	MRGBP	
Identified Proteins	**MW**	**Fraction 23**	**Fraction 23**	**Fraction 33**
BRD8	103 kDa	76	110	8
MRGBP	22 kDa	16	30	24
MRG15	41 kDa	9	39	15
MRGX	32 kDa	12	35	20
EP400NL	38 kDa	23	56	3

Table S6. Mass spectrometry analysis of EP400NL interactome purified from K562 cells through EP400NL-3xFlag 2xStrep integrated at *AAVS1* or from CRISPR/Cas9-mediated tagging of the endogenous gene (total spectral counts). Related to Figure 5.

		EP400NL	
TINTIN	**MW**	**CRISPR**	**Q6ZTU2-5 AAVS1**
BRD8	103 kDa	192	313
EP400NL	52 kDa	95	133
MRG15	37 kDa	117	164
MRGX	32 kDa	71	107
MRGBP	22 kDa	41	63

		EP400NL	
mRNA processing factors	**MW**	**CRISPR**	**Q6ZTU2-5 AAVS1**
SRRM2	300 kDa	11	4
SF3B1	146 kDa	5	2
SF3B3	136 kDa	13	13
GEMIN4	119 kDa	2	0
U2SURP	118 kDa	6	6
SF3B2	100 kDa	7	8
RBM5	92 kDa	5	2
DDX20	92 kDa	2	1
HNRNPM	78 kDa	14	15
PRPF31	55 kDa	12	6
HNRNPK	51 kDa	7	4
HNRNPF	46 kDa	3	5
RBM17	45 kDa	2	4
HNRNPC	32 kDa	2	0
SNRPD3	14 kDa	2	1

Supplemental Experimental procedures

ChIP-sequencing Experiments

For Flag ChIP, 1mg of cross-linked chromatin from K562 cells was incubated with 10µg of anti-Flag antibody (Sigma, M2) pre-bound on 300 µl of Dynabeads Prot-G (Invitrogen) overnight at 4°C. The beads were washed extensively and eluted in 0.1% SDS, 0.1M NaHCO3. Crosslink was reversed with 0.2M NaCl and incubation overnight at 65C. Samples were treated with RNase and Proteinase K for 2h and recovered by phenol-chloroform and ethanol precipitation. Libraries for sequencing were prepared with TruSeq LT adaptors (Illumina). Samples were sequenced by 50 bp single reads on HiSeq 4000 platform (Illumina).

Processing, Alignment and Peak Calling of ChIP-seq Data

FastQ format reads were aligned to the hg19 human reference using the Bowtie alignment algorithm [63]. Bowtie2 version 2.1.6 was used with the pre-set sensitive parameter to align ChIP sequencing reads. MACS version 2.0.10 (model-based analysis of ChIPseq) peak finding algorithm was used to identify regions of ChIP-Seq enrichment over the background [64]. The pipeline, commands, and parameters that were used are: Trimming of sequence (filter out 39 adaptor, and remove last 2 bases and 3 extra bases if it matches with adaptor sequence). Mapping sequences to human genome (hg19) using Bowtie: (i) command: bowtie2 -p 8 –sensitive -x genome/genome -U sequence.reads.fastq – S sample.sam. Peak calling algorithm MACS: (i) command: macs2 callpeak –t ChIPseq.data.bam -c input.sample.file.bam --broad -f BAM -g hs -n [directory] –outdir MACS2_files --nomodel --shiftsize 100 –B. Unique mapped read values were normalized to library size. Peaks were annotated as per human genome EnsDb.Hsapiens.v75 – hg19. Raw sequences and processed data of ChIP-sequencing from K562 cells were deposited in the GEO database under accession number **GSE181533**.

Cell Cycle Analysis

Fluorescence-activated cell sorting (FACS) analysis was used for cell cycle profiling. The cells were harvested by trypsinization, fixed with 70% ethanol, treated with propidium iodide for FACS analysis on a BD Accuri C6 Plus.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. MRG15 isoforms and MRGX chromatin binding *in vitro* and *in vivo*, related to Figs 1 and 2.

(A) Alignment of MRG15 isoforms that have a "long" or "short" chromodomain, the paralog MRGX without chromodomain, and the Eaf3 homolog from *Saccharomyces cerevisiae*.

(B) Whole-cell extract of isogenic K562 cells stably expressing the indicated proteins with 3xFlag 2xStrep from the AAVS1 safe harbor locus or endogenously tagged (CRISPR MRGX).

(C) ChIP-seq analysis of MRG15-S/L and MRGX in k562 cells using the AAVS1 isogneic cell lines and anti-Flag. Venn diagram of mapped significant peaks for each and overlaps. Short MRG15 isoform and MRGX bind mostly the same loci. The long MRG15 isoform shows a very small number of bound locations. Numbers represent the number of significant peaks mapped.

(D) Genomic editing with the CRISPR/Cas9 system for tagging of MRGX. Schematic of the MRGX locus, Cas9 target site, and donor construct used to insert the cassette containing 3xFlag-2xStrep in the last exon. Annotated are the positions of the stop codon the target site, the PAM motif, and homology arms left and right (HAL and R) (left). Anti-Flag analysis of MRGX whole-cell extracts from AAVS1 and tagging-CRISPR K562 cell lines (right).

(E) Methyltransferase activity of ASH1L is enhanced by MRG15 *in vitro*, related to Fig.2.

Schematic of the ASH1L locus, Cas9 target site, and donor construct used to insert the cassette containing 3xFlag-2xStrep in the last exon. Annotated are the positions of the stop codon the target site, the PAM motif, and homology arms left and right (HAL and R).

Figure S2. CRISPR/Cas9 BRD8 and EP400NL, in K562 cells, related to Figs 4-5

(A) Whole-cell extracts (WCE) were analyzed by immunoblotting with anti-Flag. Anti-KAP1 was used as loading control.

(B) Genomic editing with the CRISPR/Cas9 system for tagging double bromodomain (2BRD) BRD8 in C-term and all BRD8 transcripts in N-term. Schematic of the BRD8 locus, Cas9 target site, and donor construct used to insert the cassette containing 3xFlag-2xStrep in the last exon for 2BRD or first exon for all transcripts. Annotated are the positions of the stop codon the target site, the PAM motif, and homology arms left and right (HAL and R).

(C) Anti-Flag analysis of BRD8 whole-cell extracts from AAVS1 and tagging-CRISPR K562 cell lines.

(D) Alignment to see the similarity of human N-term EP400 and new subunit human EP400NL.

(E) Genomic editing with the CRISPR/Cas9 system for tagging EP400NL in C-term. Schematic of the EP400NL locus, Cas9 target site, and donor construct used to insert the cassette containing 3xFlag-2xStrep in the last exon. Annotated are the positions of the stop codon the target site, the PAM motif, and homology arms left and right (HAL and R).

(F) EP400NL expression from CRISPR or AAVS1 K562 cells. Anti-GAPDH is used as loading control.

Figure S3. KDs TIP60 versus TINTIN subunits in U2OS, related to Fig 6.

(A) Validation of the different siRNAs used for RNA sequencing analyses by RT-qPCR. RPLP0 was used as control. Different sets of primers were annotated. Knowdowns were normalized with the siControl (siLuciferase) (mean ± s.e.m, n=2).

(B) Cell cycle profile analysis of U2OS cells 48hrs after transfection of the different siRNAs. Error bars represented range of two independent experiments.

Statistical analyses were performed by two-way ANOVA test followed by Tukey's test, *, $p < 0.1$, **, $p < 0.01$, ****, $p < 0.0001$.

(C) 253 genes are downregulated (left) and 142 genes are upregulated (right) by KDs of TINTIN subunits (BRD8, MRGBP, and MRGX) but not NuA4/TIP60 complex. To highlight the genes with high changes in

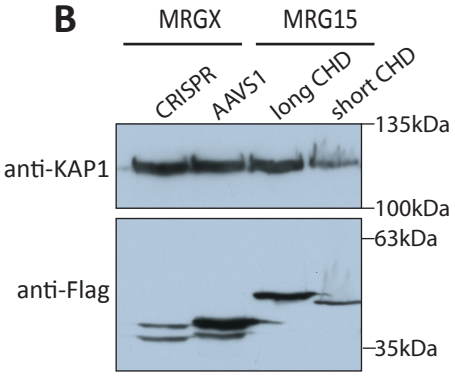
expression, a cutoff of 2-fold difference was applied to the Log2(fold change) values between the selected KDs and control (siLuc).

(D) Gene ontology analysis for downregulated (left) and upregulated genes (right) showing a significant enrichment (pValue <0.01) using DAVID 6.8.

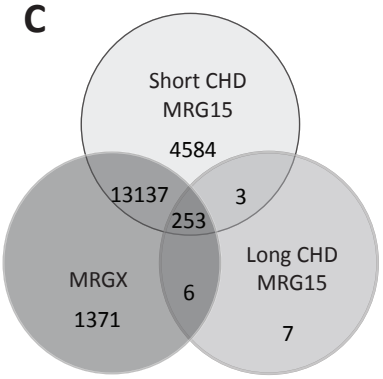
A

```
                                   chromodomain
MRG15  short MAPKQDPKPKFQEGERVLCFHGPLLYEAKCVKVAIKDKQ--------------------- 39
       long  MAPKQDPKPKFQEGERVLCFHGPLLYEAKCVKVAIKDKQ--------------------- 39
MRGX         ------------------------------------------------------------
Eaf3         ---MVDLEQEFALGGRCLAFHGPLMYEAKILKIWDPSSKMYTSIPNDKPGGSSQATKEIK 57

MRG15  short -----------------VKYFIHYSGWNK------------------------------- 51
       long  -----------------VKYFIHYSGWNKKSAVRPRRSEKSLKTHEDIVALFPVPEGAPS 82
MRGX         ------------------------------------------------------------
Eaf3         PQKLGEDESIPEEIINGKCFFIHYQGW--------------------------------- 84
                                cage

MRG15  short --------NWDEWVPESRVLKYVDTNLQKQRELQKANQEQ-------YAE----GKMRGA 92
       long  VHHPLLTSSWDEWVPESRVLKYVDTNLQKQRELQKANQEQ-------YAE----GKMRGA 131
MRGX         -----MSSRKQGSQPR-------GQQSAEEENFKKPTRSN-------MQR----SKMRGA 37
Eaf3         ------KSSWDEWVGYDRIRAYNEENIAMKKRLANEAKEAKKSLLEQQKKKKLSTSLGGP 138
                         :              :   :..: :  :.          .     .: *

MRG15  short APGKKTSGLQQ-----------KNVE--------------------VKTKKNKQKTPGNG 121
       long  APGKKTSGLQQ-----------KNVE--------------------VKTKKNKQKTPGNG 160
MRGX         SSGKKTAGPQQ-----------KNLEPALPGRWGGRSAENPPSGSVRKTRKNKQKTPGNG 86
Eaf3         SNGGKRKGDSRSNASISKSTSQSFLTSSVSGRKSGRSSANS-------------LHPGSS 185
             : * *  * .:           . :                                **..

MRG15  short DGGSTSETPQPPRKKRARVDPTVENEETFMNRVEVKVKIPEELKPWLVDDWDLITRQKQL 181
       long  DGGSTSETPQPPRKKRARVDPTVENEETFMNRVEVKVKIPEELKPWLVDDWDLITRQKQL 220
MRGX         DGGSTSEAPQPPRKKRARADPTVESEEAFKNRMEVKVKIPEELKPWLVEDWDLVTRQKQL 146
Eaf3         LRSSSDQNGNDDRRRSSSLSPNMLHHIAGYPTPKISLQIPIKLKSVLVDDWEYVTKDKKI 245
               .*:.:  :  *:: :  .*.:  . :     ::.::** :**  **:**: :*::*::

MRG15  short FYLPAKKNVDSILEDYANYKKSRGNTDNKEYAVNEVVAGIKEYFNVMLGTQLLYKFERPQ 241
       long  FYLPAKKNVDSILEDYANYKKSRGNTDNKEYAVNEVVAGIKEYFNVMLGTQLLYKFERPQ 280
MRGX         FQLPAKKNVDAILEEYANCKKSQGNVDNKEYAVNEVVAGIKEYFNVMLGTQLLYKFERPQ 206
Eaf3         CRLPADVTVEMVLNKYEHEVSQELESPGSQSQLSEYCAGLKLYFDKCLGNMLLYRLERLQ 305
               ***. .*: :*:.* :  ... :  ..:  :.*  **:* **:  **. ***::** *

MRG15  short YAEILADHPDA-PMSQVYGAPHLLRLFVRIGAMLAYTPLDEKSLALLLNYLHDFLKY    297
       long  YAEILADHPDA----PMSQVYGAPHLLRLFVRIGAMLAYTPLDEKSLALLLNYLHDFLKY 336
MRGX         YAEILLAHPDA----PMSQVYGAPHLLRLFVRIGAMLAYTPLDEKSLALLLGYLHDFLKY 262
Eaf3         YDELLKKSSKDQKPLVPIRIYGAIHLLRLISVLPELISSTTMDLQSCQLLIKQTEDFLVW 365
             * *:*    .        ::*** *****:  :  ::: * :* :*  **:   .*** :

MRG15  short LAKNSATLFSASDYEV--------APPEYHRKAV-- 323
       long  LAKNSATLFSASDYEV--------APPEYHRKAV-- 362
MRGX         LAKNSASLFTASDYKV--------ASAEYHRKAL-- 288
Eaf3         LLMHVDEYFNDKDPNRSDDALYVNTSSQYEGVALGM 401
             *  :    *. .* :         :  :*.  *:
```

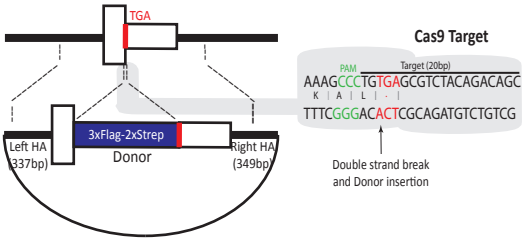
B

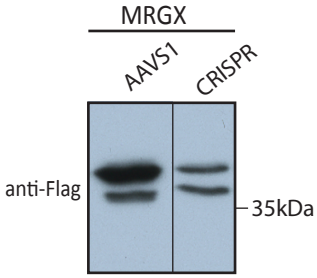


C

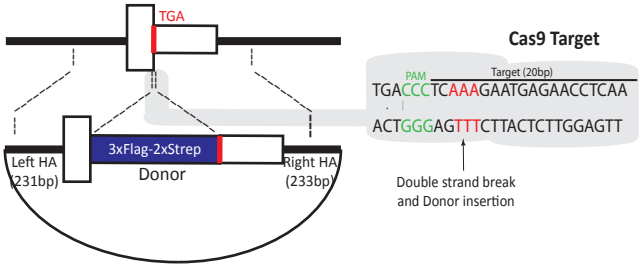


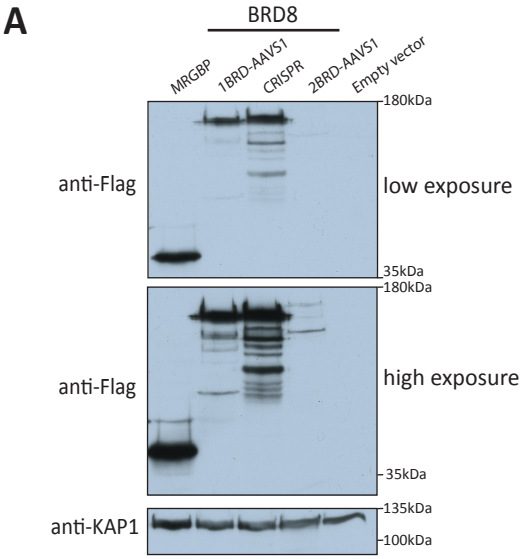
D

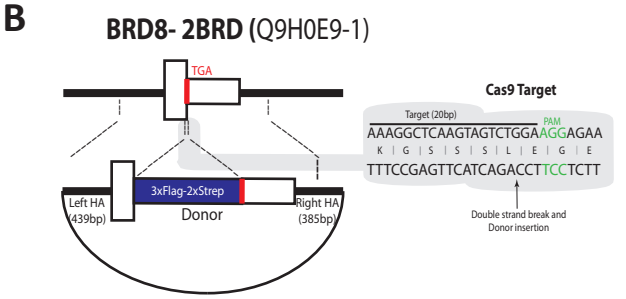


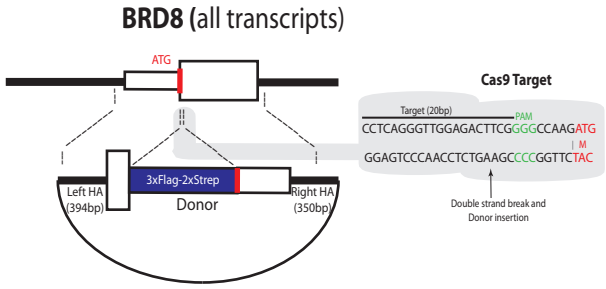


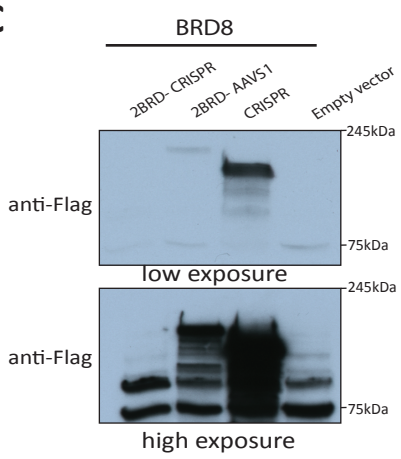
E



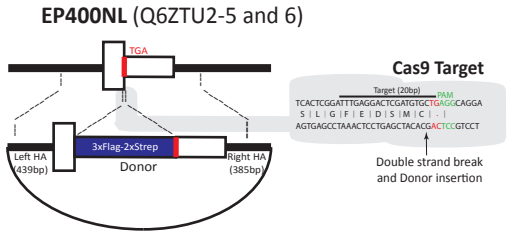

A
BRD8
MRGBP
1BRD-AAVS1
CRISPR
2BRD-AAVS1
Empty vector
180kDa
anti-Flag
low exposure
35kDa
180kDa
anti-Flag
high exposure
35kDa
135kDa
anti-KAP1
100kDa

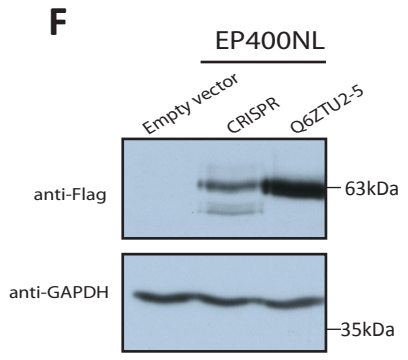

B
BRD8- 2BRD (Q9H0E9-1)
TGA
Cas9 Target
Target (20bp)
PAM
AAAGGCTCAAGTAGTCTGGAAAGGAGAAA
K | G | S | S | S | L | E | G | E
TTTCCGAGTTCATCAGACCTTCCTCTT
3xFlag-2xStrep
Left HA (439bp)
Donor
Right HA (385bp)
Double strand break and Donor insertion

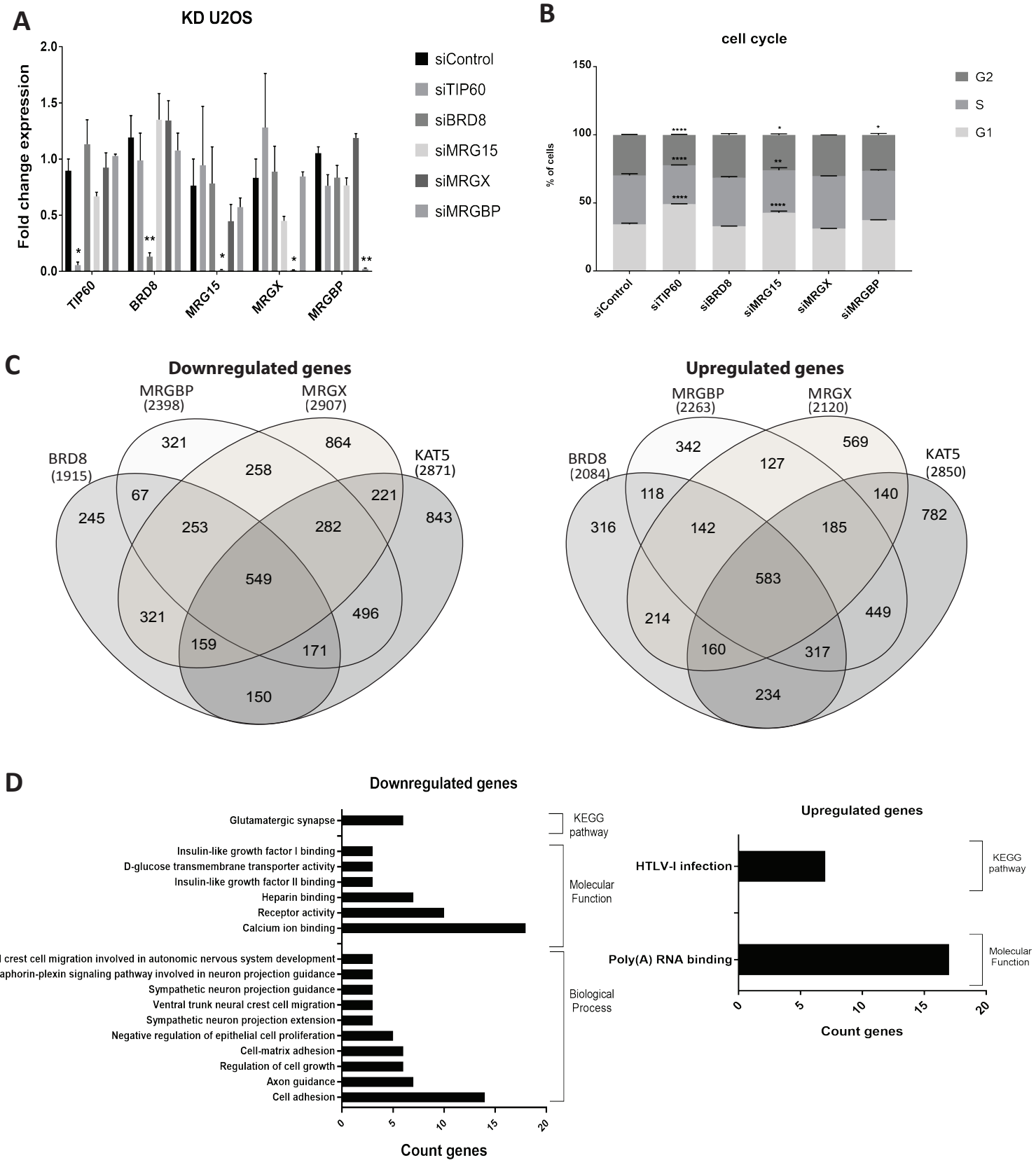

BRD8 (all transcripts)
ATG
Cas9 Target
Target (20bp)
PAM
CCTCAGGGTTGGAGACTTCGGGGCCAAGATG
M
GGAGTCCCAACCTCTGAAGCCCCGGTTCTAC
3xFlag-2xStrep
Left HA (394bp)
Donor
Right HA (350bp)
Double strand break and Donor insertion


C
BRD8
2BRD- CRISPR
2BRD- AAVS1
CRISPR
Empty vector
245kDa
anti-Flag
75kDa
low exposure
245kDa
anti-Flag
75kDa
high exposure

D
EP400 MHHGTGPQNVQHQLQRSRACPGSEGEEQPAHPNPPPSPAAPFAPSASPSAPQSPSYQIQQ 60
EP400NL-5 MQHVSSSQSSQRHVQWPGACPGAGE-EQPACSQPS----------LPLTLPSPSHQLQQ 48
EP400NL-6 MQHVSSSQSSQRHVQWPGACPGAGE-EQPACSQPS----------LPLTLPSPSHQLQQ 48
EP400 LM-NRSPATGQNVNITLQSVGPVVGGNQQITLAPLPLPSPTSPGFQFSAQPRRFEHGSPS 119
EP400NL-5 LM---------- 50
EP400NL-6 LMVRGGPAGGQNMNVDLQGVGPGLQGSPQVTLAPLPLPSPTSPGFQFSAQPRRFEHGSPS 108
EP400 YIQVTSPLSQQVQTQSPTQPSPGPGQALQNVRAGAPGPGLGLCSSSPTGGFVDASVLVRQ 179
EP400NL-5 ----------
EP400NL-6 YIQVTSPLSQQVQTQSPTQPSPGPGQALQNVRAGAPGPGLGLCSSSPTGDFVDASVLVRQ 168
EP400 ISLSPSSGGHFVFQDGSGLTQIAQGAQVQLQHPGTPITVRERRPSQPHTQSGGTIHHLGP 239
EP400NL-5 ----------DGSGLTQIAQGAQVQLQHPGTPITVRERRPSQPHTQSGGTIHHLGP 96
EP400NL-6 ISLSPSSGGHFVFQDGSGLTQIAQGAQVQLQHPGTPITVRERRPSQPHTQSGGTIHHLGP 228
EP400 QSPAAAGGAGLQPLASPSHITTANLPPQISSIIQGQLVQQQQVLQGPPLPRPLGFERTPG 299
EP400NL-5 QSPAAAGGAGLQPLASPSHITTANLPPQISSIIQGQLVQQQQVLQGPPLPRPLGFERTPG 156
EP400NL-6 QSPAAAGGAGLQPLASPSHITTANLPPQISSIIQGQLVQQQQVLQGPPLPRPLGFERTPG 288
EP400 VLLPGAGGAAGFGMTSPPPPTSPSRTAVPPGLSSLPLTSVGNTGMKKVPKKLEEIPPASP 359
EP400NL-5 VLLPGAGGAAGFGMTSPPPPTSPSRTAVPPGLSSLPLTSVGNTGMKKVPKKLEEIPPASP 216
EP400NL-6 VLLPGAGGAAGFGMTSPPPPTSPSRTAVPPGLSSLPLTSVGNTGMKKVPKKLEEIPPASP 348
EP400 EMAQMRKQCLDYHYQEMQALKEVFKEYLIELFFLQHFQGNMMDFLAFKKKHYAPLQAYLR 419
EP400NL-5 EMAQMRKQCLDYHHQEMQALKEVFKEYLIELFFLQHFQGNMMDFLAFKERLYGPLQAYLR 276
EP400NL-6 EMAQMRKQCLDYHHQEMQALKEVFKEYLIELFFLQHFQGNMMDFLAFKERLYGPLQAYLR 408
EP400 QNDLDIEEEEEEEEEEEEKSEVINDEVKVVTGKDGQTGTPVAIATQLPPKVSAAFSSQQQ 479
EP400NL-5 QNDLDIEEEE------EEHFEVINDEVKVVARKHGQPGTPVAIATQLPPRTSAAFPAQQQ 330
EP400NL-6 QNDLDIEEEE------EEHFEVINDEVKVVARKHGQPGTPVAIATQLPPRTSAAFPAQQQ 462
EP400 PFQQALAGSLVAGAGSTVETDLFKRQQAMPSTGMAEQSKRPRLEVGHQGVVFQHPGADAG 539
EP400NL-5 PLQVLSDGSTVQLPRL----------SSLGFEDSMC---------- 356
EP400NL-6 PLQVLSDGSTVQLPRL----------SSLGFEDSMC---------- 488
EP400 VPLQQLMPTAQGGMPPTPQAAQLAGQRQSQQQYDPSTGPPVQNAASLHTPLPQLPGRLPP 599
EP400NL-5 ----------
EP400NL-6 ----------


E
EP400NL (Q6ZTU2-5 and 6)
TGA
Cas9 Target
Target (20bp)
PAM
TCACTCGGATTTGAGGACTCGATGTGCTGAGGCAGGA
S | L | G | F | E | D | S | M | C | G | A
AGTGAGCCTAAACTCCTGAGCTACACGACTCCGTCCT
3xFlag-2xStrep
Left HA (439bp)
Donor
Right HA (385bp)
Double strand break and Donor insertion


F
EP400NL
Empty vector
CRISPR
Q6ZTU2-5
anti-Flag
63kDa
anti-GAPDH
35kDa


A
KD U2OS
Fold change expression
2.0
1.5
1.0
0.5
0.0
TIP60
BRD8
MRG15
MRGX
MRGBP
siControl
siTIP60
siBRD8
siMRG15
siMRGX
siMRGBP
B
cell cycle
% of cells
150
100
50
0
G2
S
G1
siControl
siTIP60
siBRD8
siMRG15
siMRGX
siMRGBP
C
Downregulated genes
MRGBP (2398)
MRGX (2907)
BRD8 (1915)
KAT5 (2871)
321
864
258
67
221
245
253
282
843
549
321
496
159
171
150
Upregulated genes
MRGBP (2263)
MRGX (2120)
BRD8 (2084)
KAT5 (2850)
342
569
127
118
140
316
142
185
782
583
214
449
160
317
234
D
Downregulated genes
Glutamatergic synapse
KEGG pathway
Insulin-like growth factor I binding
D-glucose transmembrane transporter activity
Insulin-like growth factor II binding
Heparin binding
Receptor activity
Calcium ion binding
Molecular Function
Neural crest cell migration involved in autonomic nervous system development
Semaphorin-plexin signaling pathway involved in neuron projection guidance
Sympathetic neuron projection guidance
Ventral trunk neural crest cell migration
Sympathetic neuron projection extension
Negative regulation of epithelial cell proliferation
Cell-matrix adhesion
Regulation of cell growth
Axon guidance
Cell adhesion
Biological Process
0
5
10
15
20
Count genes
Upregulated genes
HTLV-I infection
KEGG pathway
Poly(A) RNA binding
Molecular Function
0
5
10
15
20
Count genes