700–715 Nucleic Acids Research, 2019, Vol. 47, No. 2
doi: 10.1093/nar/gky1148
Published online 22 November 2018

MIR sequences recruit zinc finger protein ZNF768 to expressed genes

Michaela Rohrmoser[1], Michael Kluge[2], Yousra Yahia[3], Anita Gruber-Eber[1], Muhammad Ahmad Maqbool[3], Ignasi Forné[4], Stefan Krebs[5], Helmut Blum[5], Ann Katrin Greifenberg[6], Matthias Geyer[6], Nicolas Descostes[7,8], Axel Imhof[4], Jean-Christophe Andrau[3], Caroline C. Friedel[2] and Dirk Eick[1,*]

[1]Department of Molecular Epigenetics, Helmholtz Center Munich and Center for Integrated Protein Science Munich (CIPSM), Marchioninistrasse 25, 81377 Munich, Germany, [2]Institute for Informatics, Ludwig-Maximilians-Universität München, Amalienstrasse 17, 80333 Munich, Germany, [3]Institut de Génétique Moléculaire de Montpellier (IGMM), Univ Montpellier, CNRS-UMR5535, Montpellier, France, [4]Biomedical Center Munich, ZFP, Großhadener Strasse 9, 82152 Planegg-Martinsried, Germany, [5]Laboratory for Functional Genome Analysis (LAFUGA) at the Gene Center, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany, [6]Institute of Structural Biology, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany, [7]Department of Biochemistry and Molecular Pharmacology, New York University Langone School of Medicine, New York, NY 10016, USA and [8]Howard Hughes Medical Institute, New York University Langone School of Medicine, New York, NY 10016, USA

Received July 06, 2018; Revised October 25, 2018; Editorial Decision October 29, 2018; Accepted October 29, 2018

ABSTRACT

Mammalian-wide interspersed repeats (MIRs) are retrotransposed elements of mammalian genomes. Here, we report the specific binding of zinc finger protein ZNF768 to the sequence motif GCTGTGTG (N_{20}) CCTCTCTG in the core region of MIRs. ZNF768 binding is preferentially associated with euchromatin and promoter regions of genes. Binding was observed for genes expressed in a cell type-specific manner in human B cell line Raji and osteosarcoma U2OS cells. Mass spectrometric analysis revealed binding of ZNF768 to Elongator components Elp1, Elp2 and Elp3 and other nuclear factors. The N-terminus of ZNF768 contains a heptad repeat array structurally related to the C-terminal domain (CTD) of RNA polymerase II. This array evolved in placental animals but not marsupials and monotreme species, displays species-specific length variations, and possibly fulfills CTD related functions in gene regulation. We propose that the evolution of MIRs and ZNF768 has extended the repertoire of gene regulatory mechanisms in mammals and that ZNF768 binding is associated with cell type-specific gene expression.

INTRODUCTION

Approximately half of mammalian genomes is of repetitive nature and composed of long (LINE) and short interspersed sequences (SINE) (1,2). Mammalian-wide interspersed repeats (MIRs) are an ancient family of retrotransposed SINEs that spread genome-wide before and during mammalian radiation (3,4). MIRs are ∼240 bp long and consist of tRNA-derived sequences, a 70 bp MIR-specific core region, and sequences similar to the 3′ ends of LINEs. MIRs are enriched at gene loci in euchromatin, harbor putative transcription-factor binding sites, provide insulator and enhancer function (5–8), encode microRNAs, are transcribed by RNA polymerase III (9,10), are associated with tissue-specific gene expression (5,11), and sometimes provide splicing signals and contribute to exonization (12). MIRs constitute 5–16% of the genome in marsupials and monotremes and 0.5-3% in placentalia (13). Like other transposable elements, MIRs have shaped gene regulatory networks in vertebrates (14–17), but our understanding how MIRs regulate gene activity is still elusive.

Similarly to MIRs, the family of zinc finger proteins (ZNFs) strongly expanded in mammals (18,19). Widespread binding of ZNFs to regulatory regions indicates that mammalian genomes contain an extensive ZNF regulatory network that targets a diverse range of genes and pathways (20,21). Zinc finger protein 768 (ZNF768) evolved in mammals and is defined by a domain of ten zinc fingers with >96% (Figure 1) identity in placentals and marsupials, but is less conserved in monotremes (Supplementary Figure S1). Placentalia additionally evolved an array of 10–20 heptad repeats in the amino-terminus of ZNF768, which is absent in marsupials and monotremes. This array has a striking similarity to the carboxy-terminal domain (CTD)

*To whom correspondence should be addressed. Tel: +49 89 3187 1512; Email: eick@helmholtz-muenchen.de

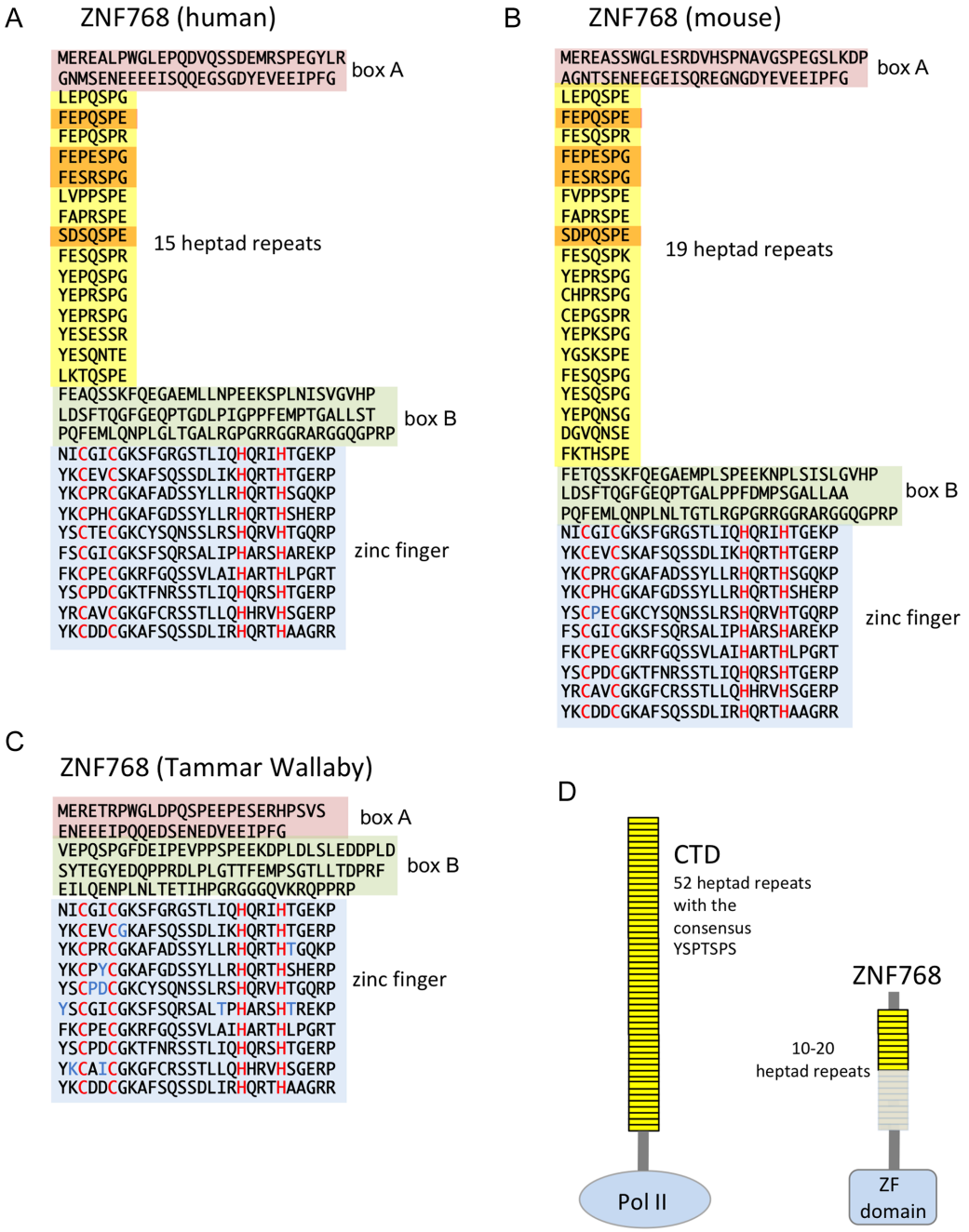

Figure 1. Domain structure of ZNF768 in placentalia and marsupials and comparison with the CTD of RNA polymerase II. (**A**) Human ZNF768 is composed of domains box A (red box) and box B (green box) at the N-terminus interrupted by an array of 15 heptad repeats (yellow box) and a domain of 10 zinc fingers at the C-terminus (blue box). (**B**) Mouse ZNF768 evolved an array of 19 heptad repeats. (**C**) ZNF768 of the marsupial Tammar Wallaby contains conserved A, B, and zinc finger domains, while the array of heptad repeats is absent. (**D**) Number of heptad repeats in RNA polymerase II in vertebrates and ZNF768 in placentalia (see also Supplementary Figure S1).

of the large subunit (Rpb1) of RNA polymerase II (Pol II), which is composed of 52 heptad repeats with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$.

The CTD functions as a platform for recruitment and dissociation of cellular factors to the transcription machinery and is mainly regulated during the transcription cycle by phosphorylation of heptad repeats by various kinases (22–26). It is required for initiation, elongation, and termination of transcription, but also for capping, splicing, and 3′processing of the nascent transcript. Interestingly, CTD can function as transcriptional activator after fusion to a GAL4 DNA binding domain (27). Furthermore, transition of Pol II through the transcription cycle is also observed if CTD is fused to other subunits of Pol II (28). Recent reports further provide evidence that CTD of Pol II can aggregate reversibly alone, or with low complexity domains of other transcription factors, like FUS, and that the ability for

phase separation in liquid droplets is an important feature for the regulation of transcriptional activity (29–32).

Due to the striking similarity of the heptad repeat array in ZNF768 with the array of heptad repeats in CTD of Pol II we investigated if ZNF768 can act as a transcription factor and fulfill gene regulatory functions in cells.

MATERIALS AND METHODS

Tissue culture and recombinant gene expression

U2OS osteosarcoma cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco) and Raji B-cells in RPMI 1640 medium (Gibco) supplemented with 10% fetal calf serum (FCS, Bio&Sell), 2 mM L-glutamine (Gibco), 100 U/ml penicillin (Gibco), and 100 μg/ml streptomycin (Gibco) at 37°C at 8% or 5% CO_2, respectively. Stably transfected U2OS cell lines were generated with the expression vector pRTS-1 (33) using Polyfect (QIAGEN) followed by hygromycin B (200 μg/ml) selection. Conditional gene expression was induced with 1 μg/ml doxycycline. Recombinant ZNF768 and mutants are tagged C-terminally by a hemagglutinin (HA) tag and synthesized with an optimized codon usage (Gene Art, Regensburg). Details for cloning in pRTS has been described elsewhere (34). All plasmids were confirmed by DNA sequencing prior to expression.

Monoclonal antibody

The generation of monoclonal antibodies has been described previously (34). The ZNF768-specific peptide RSPESDSQSPEFESQSPRYEPQSPGYEPRSPG (synthesized by PSL GmbH, Heidelberg) was coupled to ovalbumin for immunization. The rat monoclonal antibody 7D6 used in this study (IgG2c) specifically recognizes human ZNF768.

Immunoprecipitation (IP) and SDS-PAGE

Cells were washed twice with cold phosphate-buffered saline (PBS) and lysis was performed in 100 μl lysis buffer per 2×10^6 cells (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1% NP-40 (Roche), 1x PhosSTOP (Roche), 1× protease inhibitor cocktail (Roche)) at 4°C for 30 min, followed by sonication on ice using a BRANSON Sonifier 250 (15 s on, 15 s off, 50% duty) and centrifugation at 16 400 rpm (FA-45-24-11 rotor) for 10 min at 4°C. Immunoprecipitation was performed using Dynabeads® Protein A und G (1:1) (Invitrogen). Lysates were incubated with antibody-coupled beads (2.5 μg of antibodies for 4 h at 4°C, followed by three washes with 1 ml lysis buffer) overnight. Beads were washed three times with 1 ml lysis buffer and boiled in laemmli buffer (2% SDS, 10% glycerol, 60 mM Tris–HCl, pH 6.8, 10 mM EDTA, 1 mM PMSF, 100 mM DTT, 0.01% bromophenol blue) for SDS-PAGE. Whole cell lysates or IP samples were resolved by SDS-PAGE (10% or 15%) and transferred onto nitrocellulose transfer membranes (GE Healthcare). The membrane was blocked with 5% milk/TBS-T for 1 h. Incubation with primary antibodies was performed over night at 4°C, followed by incubation with HRP-conjugated secondary antibodies for 1 h and chemiluminescence detection with ECL (GE Healthcare).

Immunofluorescence microscopy

U2OS cells were seeded on a coverslip and grown for 24 h. Cells were washed with PBS and fixed with 2% paraformaldehyde (PFA) at RT for 5 min. After permeabilization with 0.15% TritonX-100, samples were blocked with 1% BSA and incubated with 7D6 or HA-specific mAbs over night at 4°C. Samples were washed with PBS for 5 min at RT, 0.15% Triton X-100 for 10 min at RT, blocked with 1% BSA for 7 min and incubated with Cy5-conjugated donkey anti-rat immunoglobulin (Dianova) in the dark for 45 min. Cells were washed again, stained with 4′,6-diamidino-2-phenylindole (DAPI) (Sigma) and mounted on slides using fluorescent mounting medium (Dako). Confocal microscopy was performed on a Leica LSCM SP2 fluorescence microscope using the objective HCX PL APO 63× 1.4. Images were processed using ImageJ 1.37 V and Fuji software and the plug-in RGB profiler. Scale bars were calculated as follows:

$B \times 5\ \mu m / P$ (B = picture length in μm, P = (512 pixel × voxel size) in μm)

siRNA transfection

siRNA transfection was performed according to the manufacturer's protocol using HS_ZNF768_1 FlexiTube siRNA (Qiagen) and the HiPerFect Transfection Reagent (Qiagen) with the exception that transfection was repeated after 24 h. Negative (non-silencing) siRNA (Qiagen) was used as control.

Cell proliferation assay

Cell proliferation was determined using the Real-time xCELLigence System (Roche). U2OS cells were seeded at a density of 3.000 cells per 100 μl in equilibrated 96-well microtiter xCELLigence assay plates (E-plates). Conditional gene expression was induced with 1 μg/ml doxycycline at the indicated time points. Alternatively, siRNA transfection was performed according to the manufacturer's protocol.

Purification of ZNF768

Expression plasmids of human ZNF768 (UniProt accession number Q9H5H4) were cloned from a synthetic gene that was codon optimized for expression in *Escherichia coli* cells (Gene Art, Regensburg). A full length ZNF768 (1–540) construct and a construct consisting of the N-terminal heptad-repeats only (1–197) were cloned by PCR with restrictions sites NcoI/EcoRI and ligated into a pGEX-4T1 vector modified with a TEV protease cleavage site. All plasmids were confirmed by DNA sequencing prior to expression.

Plasmids were transformed into *E. coli* BL21(DE3) cells and induced at an OD_{600} of 0.6 to 1.0 with 0.3 mM IPTG for 4 h growth. Cells were harvested in lysis buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM DTE) and lysed by ultrasound. Fusion proteins were isolated with GSH Sepharose FastFlow (GE Healthcare) affinity chromatography methods. Cleavage of the GST-tag was achieved by adding TEV protease in a 1/50 ratio and was

performed for 20 h at 4°C. Protein solution was concentrated and loaded on a preparative HiLoad 16/60 Superdex 200 prep grade gel filtration column (GE Healthcare) for full length ZNF768 or on a HiLoad 16/60 Superdex 75 column for the truncated version of ZNF768 (1–197), respectively, and equilibrated in gel filtration buffer (50 mM HEPES pH 7.5, 150 mM NaCl and 1 mM TCEP). Fractions of the peak containing ZNF768 proteins determined by SDS PAGE analysis were pooled and concentrated to 5 mg/ml. The protein was aliquoted, snap frozen in liquid nitrogen and stored at −80°C.

Gel shift assay

Gel shift assay was performed according to the manufacturer's protocol using the DIG Gel Shift Kit, second generation (Sigma-Aldrich). Briefly, oligonucleotides were annealed to equimolar amounts of their complementary strands (M1: 5′- CAGTGCTGTGTGAC CTTGGGCAAGTCACTTAACCTCTCTGCAGT-3′, M2: 5′- CAGTGCTGTGTGCAGTCAGTCAGT CAGTCAGTCCTCTCTGCAGT-3′ and M3: 5′- CAGTCAGTTGTGACCTTGGGCAAGTCACTT AACCTCCAGTCAGT-3′) by heating to 95°C for 5 min and cooling slowly to room temperature. Double-stranded oligonucleotide probes were labelled at the 3′ end using DIG-11-dUTP and terminal transferase. Binding reactions were performed in 20 μl volumes containing binding buffer [20 mM HEPES, pH 7.6, 1 mM EDTA, 10 mM $(NH4)_2SO_4$, 5 mM DTT, 0.2% (w/v) Tween 20, 30 mM KCl], 50 ng/μl Poly [d(I-C)] and 5 ng/μl Poly L-lysine at room temperature for 15 min. 0.6 ng of DIG-labelled DNA and extract of 1.5–15 μg purified ZNF768-WT or ZNF768 1–197 was used. For competition experiments, unlabeled competitor DNA was added in excess. Protein-DNA-complexes were separated by a native 6% (w/v) polyacrylamide 0.5× TBE gel, transferred onto a positively charged Nylon membrane (GE Healthcare), fixed by Stratagene cross-linker and detected by chemiluminescent substrate CSPD (Roche).

Chromatin immunoprecipitation for ChIP-seq

Cells were crosslinked using a formaldehyde containing solution (10 mM NaCl, 0.1 mM EDTA pH 8.0, 0.05 mM EGTA pH 8.0, 5 mM HEPES pH 7.8 and 1% formaldehyde) for 10 min at 20°C, the reaction was quenched by the addition of glycin to a final concentration of 250 μM for 5 min. Crosslinked cells were collected and washed twice with PBS before snap freezing in liquid nitrogen and storage at −80°C until subsequent use.

Prior to sonication, the crosslinked cells were resuspended in lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100, 1× protease inhibitor cocktail) at 4°C for 20 min. Nuclei were collected by centrifugation and washed in a second buffer (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 1× protease inhibitor cocktail) for 10 min at 4°C then collected by centrifugation and resuspended in the shearing buffer (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 100 mM NaCl, 0.1% Na-Deoxycholate, 0.5% *N*-lauroylsarcosine, 1× protease inhibitor cocktail). Sonication was carried out in a Bioruptor Pico ultrasounds water bath (Diagenode B01060001) for 30 cycles of 30 s ON and 30 s OFF pulses in 4°C water. Sonicated extracts were centrifuged at high speed in the presence of 0.1% of Triton X-100 and snap frozen in liquid nitrogen and then stored at −80°C until subsequent use.

Prior to ChIP, the ZNF768 mAb was coupled to protein-G coated magnetic beads (Dynabeads, life technologies) by incubation in 0.5% BSA PBS overnight at 4°C. Pre-coated beads were then washed and incubated with the sonicated chromatin extracts. ChIP was carried out overnight at 4°C on a rotating wheel. The equivalent of 10×10^7 cells sonicated extract was used for each ChIP experiment for both cell lines. After incubation, the beads were washed 7× with wash buffer (50 mM HEPES pH 7.6, 500 mM LiCl, 1 mM EDTA pH 8.0, 1% NP-40, 0.7% Na-Deoxycholate, 1× protease inhibitor cocktail) followed by one wash with TE-NaCl buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0, 50 mM NaCl). Immunoprecipitated chromatin was eluted by two sequential incubations with 100 μl elution buffer (50 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 1% SDS) at 65°C for 15 min. The two eluates were pooled and incubated at 65°C for 12 h to reverse-crosslink of chromatin, followed by treatment with RNase A (0.2 μg/ml) at 37°C for 2 h and proteinase K (0.2 μg/ml) at 55°C for 2 h. The DNA was isolated by phenol:chloroform:isoamylalcohol (25:24:1 pH 8.0) extraction followed by Qiaquick PCR Purification (Qiagen, Germany) and quantified with Qubit DS DNA HS Assay (ThermoFisher Scientific, USA).

At least 1 ng of ChIP DNA was used to prepare sequencing library with Illumina ChIP Sample Library Prep Kit (Illumina, USA) with a few optimizations to the protocol. The ChIP DNA was size selected using Ampure beads (Life technologies) to enrich for fragments <400 bp prior to end-repair, 3′end adenylation and adapter ligation. Library fragments were then directly amplified by 10 cycles of PCR. Barcoded libraries from different samples were pooled together and sequenced on Illumina HiSeq2000 platform in paired-end sequencing runs.

RNA-seq libraries

For preparation of total RNA cells (0.9 Mio/ml) were harvested and resuspended in TRIzol reagent (Life Technologies) and snap-frozen in liquid nitrogen. After thawing RNA was extracted from 0.4 ml of TriZol lysate using the direct-zol RNA Miniprep (Zymo Research, Irvine CA, USA) as described in the manufacturer's protocol. RNA was assessed for purity by UV–vis spectrometry (Nanodrop) and for integrity by Bioanalyzer (Agilent Bioanalyzer 2100, Agilent, Santa Clara USA). RNA was of high purity (abs. 260/280 > 1.9, abs. 269/239 > 2.1) and integrity (Bioanalyzer RIN > 9) and thus used for further processing. For production of RNA-seq libraries total RNA was DNAse treated (dsDNAse, Fermentas) and 100 ng of this RNA was processed with a strand-specific protocol (RNA-seq complete kit, NuGEN, San Carlos, USA). In brief the RNA was reverse transcribed to cDNA with a reduced set of hexamer primers, avoiding excessive representation of rRNA in the cDNA. Second strand cDNA synthesis

was done in presence of dUTP. After ultrasonic fragmentation of the cDNA and end repair, Illumina-compatible adapter were ligated. Adapters contained uracil in one strand, allowing complete digestion of the second-strand derived DNA. After strand selection the libraries were amplified, assessed for correct insert size on the Agilent Bioanalyser and diluted to 10 nM. Barcoded libraries were mixed in equimolar amounts and sequenced on an Illumina HiSeq1500 in single-read mode with a read length of 100 bp.

Deep sequencing

ChIP-seq and RNA-seq analysis was performed as previously described (35). Four biological replicates of U2OS and Raji cells were used for RNA-seq library construction.

ChIP-seq data processing

Raw sequencing reads were aligned to the human genome (hg38) using BWA (36). Sequence reads with an alignment score <30 for paired-end reads and <20 for single-end reads were discarded as well as all reads that aligned equally well to different positions in the genome. Peak calling was performed using GEM (37) in GPS mode and with a *q*-value cutoff of 0.01. Overlapping peaks (peak centers: ±100 bp) were merged both within samples and across all four samples to obtain the final list of unique peaks. Motif discovery was performed using MEME-ChIP (38) and sequence logos of the binding motif and surrounding regions were created using weblogo (39). Annotation of peaks relative to gene features was performed using the ChipSeeker package in R (40). Gene annotations were taken from GENCODE version 25 (41). Repeat annotations by RepeatMasker and phyloP100 conservation scores (42) for hg38 were downloaded from the UCSC genome browser. Visualization of ZNF768 binding on the genome and corresponding peaks was performed using Gviz (43). For the analysis of binding frequencies of ZNF768 in promoter, UTR, exon and intron regions, the same number of 200 bp regions were randomly selected using BEDtools (44).

RNA-seq data processing

Quality check of sequencing reads was performed using FastQC (available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Sequencing reads were mapped against the human genome (hg38) and human rRNA sequences using ContextMap version 2.7.9 (45) (using BWA as short read aligner and default parameters). Number of read fragments per gene were determined from mapped RNA-seq reads in a strand-specific manner using featureCounts (46) and GENCODE version 25 gene annotations. RPKM values were calculated using edgeR and averaged between replicates (47). Differential gene expression analysis was performed using limma (48). Functional enrichment analysis for UniProt keywords and Gene Ontology terms was performed with the DAVID webserver (49). Significantly enriched terms were determined using a cutoff of 0.05 on the *P*-value adjusted for multiple testing using the method by Benjamini and Hochberg (50). Analysis workflows were implemented and run using the workflow management system Watchdog (51).

Purification of ZNF768 for mass spectrometric analysis

For purification of ZNF768, Raji or U2OS cells (3×10^8) were collected and IP was performed in 3 biological replicates as described in the respective paragraph of immunoprecipitation. Simultaneously, αZNF768 antibody (7D6) and αPes1 (8E9) antibody, respectively, was coupled to Sepharose A and G beads for 4 h at 4°C. αZNF768 antibody (7D6) was used to identify the interactome of ZNF768 whereas αPes1 (8E9) antibody served as a subclass control (52,53).

On beads digestion

After the last washing step with lysis buffer, beads were washed three times by adding 100 μl of 50 mM NH_4HCO_3. For trypsin digest, beads were transferred to a clean tube and incubated with 100 μl of 10 ng/μl trypsin-solution in 1M urea and 50 mM NH_4HCO_3 for 30 min at 25°C. Samples were centrifuged at 800 rpm and supernatant was transferred into a fresh tube. Beads were washed twice with 50 μl of 50 mM NH_4HCO_3. The supernatants were pooled into the corresponding tube and incubated overnight at 25°C after addition of 1mM DTT. Iodoacetamide (IAA) 10 μl (5mg/ml) was added and incubated for 30 min in the dark at 25°C. To quench the IAA, 1 μl of 1M DTT was added and samples were incubated for 10 min at 25°C, followed by addition of 2.5 μl of trifluoroacetic acid (TFA) and desalting using 2× C18 Stagetips (54). Stagetips were washed three times with 20 μl of 100% ACN (1000 rpm, 1 min) and three times by adding 20 μl of 0.1% TFA (1800 rpm, 1 min). Subsequently, samples were added (800 rpm, 30 min) and washed 3 times with 20 μl of 0.1% TFA, followed by elution into a clean tube by washing three times with 20 μl of 80% ACN/25% TFA solution. Finally, samples were evaporated to dryness, resuspended in 20 μl formic acid solution and stored at −20°C until LC–MS analysis.

Protein quantification by liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS)

Purified peptides (5μl) were automatically injected into in an Ultimate 3000 RSLC HPLC system (Dionex Thermo), separated on an analytical column C18 micro column (75 μm i.d. × 15 cm, packed in-house with Reprosil Pur C18 AQ 2.4 μm, Doctor Maisch) using a 50-min gradient from 5 to 60% acetonitrile in 0.1% formic acid. The effluent from the HPLC was subsequently electrosprayed into a LTQ Orbitrap XL mass spectrometer (Thermo). The MS instrument was operated in a data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 300 to 1800) were acquired in the Orbitrap with a resolution of $R = 60\,000$ at m/z 400 (after accumulation to a 'target value' of 500,000 in the linear ion trap). The six most intense peptide ions with charge state between 2 and 4 were sequentially isolated to a target value of 10,000 and fragmented in the linear ion trap by collision induced dissociation (CID). For all measurements with the Orbitrap mass analyzer, three lock-mass ions from ambient air ($m/z = 371.10123$, 445.12002, 519.13882) were used for internal. Usual MS conditions were: spray voltage, 1.5 kV; no sheath and

auxiliary gas flow; heated capillary temperature, 200°C; normalized collision energy 35% for CID in LTQ. The threshold for ion selection was 10 000 counts for MS2. The used activation was 0.25 and activation time 30 ms. MaxQuant 1.5.2.8 was used to identify proteins and quantify by iBAQ with the following parameters: Database, Uniprot_Hsapiens_3AUP000005640_170526; MS tol, 10ppm; MS/MS tol, 0.5 Da; Peptide FDR, 0.1; protein FDR, 0.01 Min. peptide length, 5; variable modifications, oxidation (M); fixed modifications, carbamidomethyl (C); peptides for protein quantitation, razor and unique; min. peptides, 1; min. ratio count, 2. Identified proteins were considered as interaction partners if their MaxQuant iBAQ values displayed a greater value than $\log_2$ 5-fold enrichment (FC) and *P*-value 0.05 (*t*-test adjusted for multiple comparisons) when compared to the control.

RESULTS

ZNF768 domain structure and conservation

An array of ten zinc fingers at the C-terminus of ZNF768 shows high conservation in placentalia and marsupials (>96%) but is less conserved in monotremes (blue boxes in Figure 1 and Supplementary Figure S1). At the N-terminus, two sequence blocks (box A and box B, red and green boxes in Figure 1) are conserved in placentalia and marsupials, but replaced by unrelated sequences in monotremes. In addition, ZNF768 of placentalia has evolved an array of heptad repeats that is positioned between box A and box B. The number of repeats varies between 20 repeats in mouse lemur and 10 repeats in pika and malayan pangolin. Mouse ZNF768 contains 19 repeats, chimpanzee 16 and human 15 repeats (Supplementary Figure S1). Similar to heptad repeats in CTD of Pol II, the heptad repeats in ZNF768 show no length variation (except a single extended repeat in pika). However, the composition of amino acids in the heptad repeats shows higher variation in ZNF768, both, within and between species (Supplementary Figures S1 and S2B). Serine-5 and proline-6 residues show the highest conservation between ZNF768 and Pol II, followed by the residues corresponding to tyrosine-1 and proline-3 in the CTD, the position of threonine-4 and serine-7 show only little or almost no conservation. The position corresponding to serine-2 in the CTD is particularly remarkable in ZNF768. It is replaced in almost all repeats by an acidic amino acid (mostly glutamic acid). The phosphorylation of serine-2 residues in CTD by P-TEFb is a hallmark in RNA elongation control (55) and a replacement of serine-2 may mimic its phoyphorylation. It is thus tempting to speculate that the array of heptad repeats in ZNF768 potentially can mimic a CTD phosphorylated at serine-2 residues.

ZNF768 is associated with euchromatin and required for growth and cell viability

To study the cellular function of ZNF768 we first raised a monoclonal antibody (7D6) towards human ZNF768 using a peptide containing heptad repeats 8–12 as epitope (Figure 1A, materials and methods). This antibody preferentially stains euchromatic regions in the nucleus of fixed U2OS cells (Figure 2A and Supplementary Figure S3A). Expression of the zinc finger-containing C-terminal domain of ZNF768 (Figure 2E) caused a similar staining pattern, while expression of the N-terminus containing the array of heptad repeats resulted in a more diffuse staining of the nucleus (Supplementary Figure S3B), suggesting that the zinc finger domain is responsible for the association of ZNF768 with euchromatin. Antibody 7D6 immunoprecipitated endogenous ZNF768 protein quantitatively from extracts of osteosarcoma cell line U2OS and B-lymphoid cell line Raji (Figure 2B), proving its high specificity and suitability to study the binding of ZNF768 to DNA in chromatin immunoprecipitation (ChIP) experiments. Knockdown experiments of ZNF768 confirmed the specificity of mAb 7D6 (Figure 2C) and showed further that ZNF768 is required for viability and proliferation of U2OS cells (Figure 2D). In line with its essential function, expression of mutants with deletions of either the N- or C-terminal domain of ZNF768 have a dominant-negative phenotype and inhibit cell proliferation (Figure 2E–G). Finally, ZNF768 is a phosphoprotein and can be phosphorylated at almost all heptad repeat serine-5 residues (www.cellsignal.com, Supplementary Figure S2A). Treatment of cellular extracts of U2OS cells with alkaline phosphatase causes a shift of the hyperphosphorylated form of ZNF768 (Supplementary Figure S2C) and reveals that a large fraction of ZNF768 is hyperphosphorylated in U2OS cells.

Identification of the ZNF768 binding motif in cellular DNA

To investigate if ZNF768 can bind to specific DNA sequences, we performed ChIP experiments with mAb 7D6 using extracts of U2OS and Raji cells. DNA libraries of two biological replicates were prepared for each cell line and analyzed by next generation sequencing. Peak calling identified a total of 21 012 unique peaks and 13.1% of these peaks were consistently identified in all four samples and an additional 28.8% at least in both replicates for the same cell type (Supplementary Figure S4). Generally, ZNF768 binding sites distributed over all chromosomes (Supplementary Figure S5). Motif discovery identified several potential binding motifs for ZNF768 (Supplementary Figure S6A). The top two identified motifs were found in 46% and 37% of peaks, respectively, and for both motifs the other motif was often found as a secondary motif at a distance of ∼20 bp. We thus hypothesized that the ZNF768 binding motif consists of anchor regions connected by a linker region of ∼20 bp. In fact, 58.1% of identified peaks contained this consensus motif with at most three mismatches in the anchor regions and a linker region of 20 ± 3 bp (Figure 3A, Supplementary Figures S4 and S6B). For peaks identified in all 4 samples, this number was as high as 98.3% and the vast majority of peaks with motif hits (83.5%) had a linker length of 20 bp (Supplementary Figure S6B). Gelshift experiments with recombinant ZNF768 protein confirmed the motif, GCTGTGTG (N_{20}) CCTCTCTG, and revealed that the nucleotide sequence of the spacer between the two anchor regions is likely not critical for binding (Supplementary Figure S7).

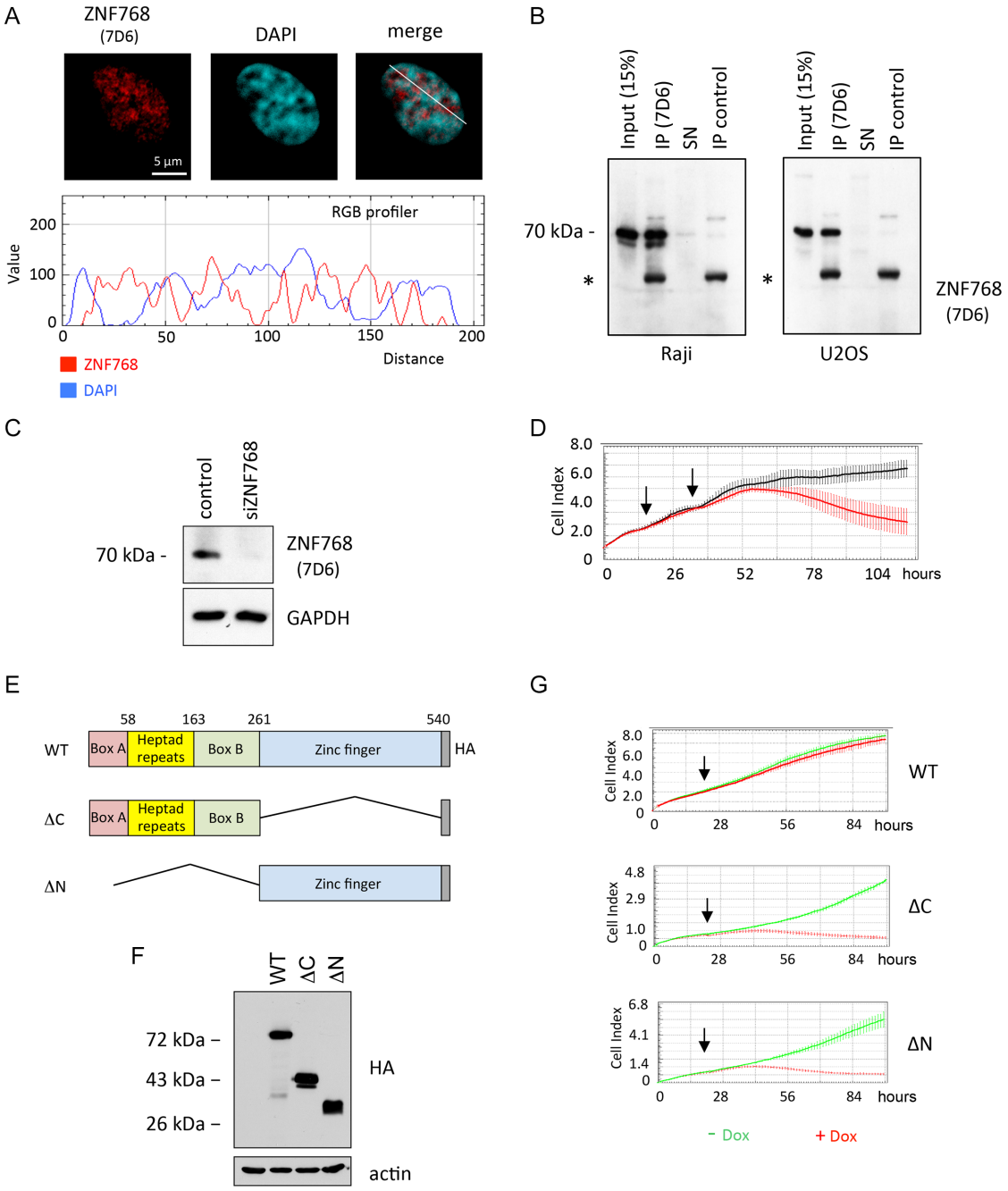

Figure 2. ZNF768 is associated with euchromatin and essential for cell viability and proliferation. (**A**) Confocal image of U2OS cells stained with DAPI and with the ZNF768-specific mAb 7D6; merge of images on the right hand site. White line marks the area of the RGB profiler, profiles of DAPI and ZNF768 at the bottom. (**B**) mAb 7D6 immunoprecipitates a 70 kD protein from extracts of Raji and U2OS cells. SN: supernatant, *: Ig heavy chain. (**C**) siRNA mediated knockdown of ZNF768 in U2OS cells. (**D**) Growth kinetics of U2OS cells after knockdown of ZNF768 measured by xCelligence (Roche). Arrows indicate consecutive addition of siRNA. (**E**) Expression constructs of HA-tagged ZNF768 wild-type and deletion mutants and (**F**) expression control in U2OS cells. (**G**) Growth kinetics of U2OS cells after expression of ZNF768-WT and ZNF768 mutants measured by xCelligence (Roche). Arrows indicate addition of doxycycline.

ZNF768 binds to MIR sequences

A systematic comparison of ZNF768 binding sites to repeats in the human genome showed an enrichment of binding sites within all four types of MIRs (Figure 3B). 12,488/21,012 peaks overlapped with MIRs. Furthermore, almost all peaks (92%) with the binding motif were contained in a MIR sequence and the consensus sequence for all MIR types actually contains the ZNF768 binding motif (Supplementary Figure S8A). Despite this fact, only a small fraction (12.2%) of MIRs in the human genome contains the ZNF768 binding motif, which is not surprising giving a per-base identity <80% for human MIR sequences. Although only 15.8% of MIRs containing the binding mo-

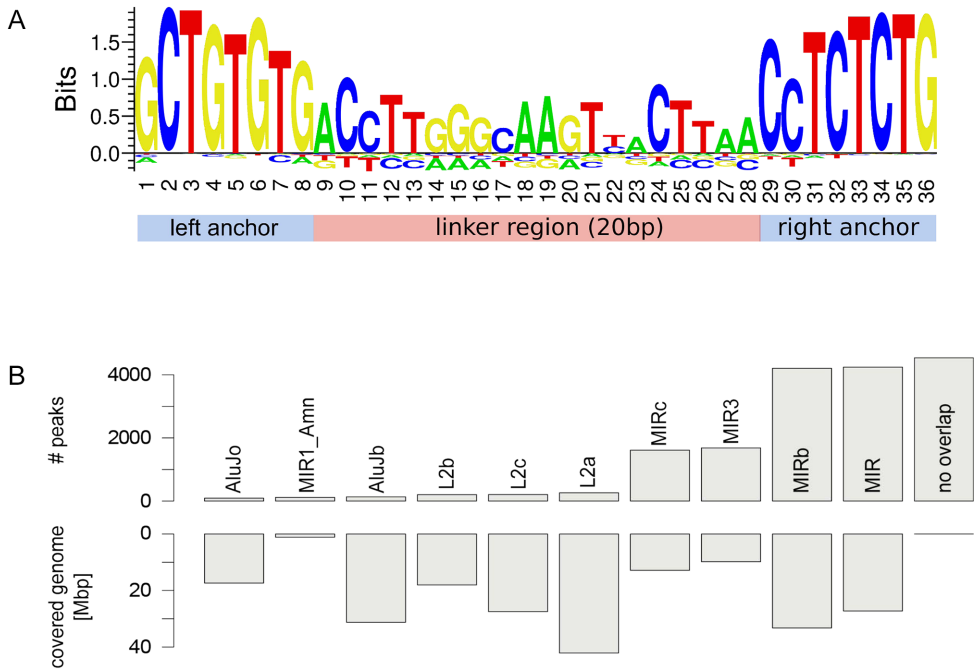

Figure 3. DNA binding motif and genomic binding of ZNF768 to MIRs. (**A**) Consensus ZNF768 binding motif identified by ChIP-seq experiments in Raji and U2OS cells and determined from peaks containing the motif (±3 mismatches in the anchors and linker length of 20 ± 3 bp). (**B**) Number of ZNF768 binding sites overlapping with particular type of repetitive sequences (top; in case of multiple overlaps the largest overlap is used) compared to the genomic length covered by the corresponding type of repetitive sequence (bottom). The right-most bar shows the number of ZNF768 binding sites with no overlap to repetitive sequences.

tif were found to be bound by ZNF768, this fraction increased to 54.2% when considering MIRs with a more stringent and strict version of the motif (linker length: 19/20bp, 1 mismatch in anchors). Thus, most MIRs diverged so far from the consensus that the binding motif was lost and no general conservation of the binding motif in human MIRs was observed (Supplementary Figure S8B). This divergence also allowed reliably aligning reads to MIR sequences despite their repetitive origin. Only reads that could be aligned uniquely to the genome were used for peak calling. Although 8524 detected peaks (40.6%) were not within MIRs, 13.5% of these peaks contained the binding motif. The remaining peaks may contain a weaker version of the binding motif, recruit ZNF768 to chromatin by other mechanisms (e.g. looping), or represent spurious binding.

Interestingly, MIRs with ZNF768 binding show a clear conservation of the two anchor motifs in the human genome. Sequences of the linker in the binding motif and outside of the binding motif were not particularly conserved, similar to MIRs without binding of ZNF768 (Supplementary Figure S8B). We further investigated whether ZNF768 binding sites in MIRs were also conserved across species by analyzing phyloP100 conservation scores determined from a multiple alignment of 99 vertebrate genomes against the human genome. Positive PhyloP scores indicate slower than expected evolution. The analysis of phyloP100 scores within and around the binding motif (±25 bp) in MIR sequences bound by ZNF768 showed increased conservation for most positions within the anchor regions (Figure 4A). Sequences outside of the binding motif or within the linker region, however, were mostly not conserved. Unbound MIR sequences showed no particular conservation (Figure 4B) indicating that ZNF768 binding represents a conserved function of a subset of MIR sequences in mammals.

ZNF768 binding is associated with transcribed genes

We next asked if ZNF768 binding sites were enriched in regulatory elements of genes. We found a strong enrichment of ZNF768 binding in promoters and a slight enrichment of binding in exons and introns, while the binding frequency in intergenic sequences was reduced (Figure 5A). Interestingly, the 1061 ZNF768 binding sites outside of MIRs that contained the binding motif showed an even higher enrichment at promoters. To investigate whether genes with ZNF768 binding tended to be more highly expressed, we analyzed RNA-seq data of four replicates of total RNA of Raji and U2OS cells (Supplementary Table S1). In both cell lines, protein-coding genes with ZNF768 binding in the promoter or 5′UTR were more highly expressed on average than the remaining protein-coding genes (Figure 5B). In contrast, binding in intronic regions showed only a small but significant effect (Figure 5B). This provides evidence that ZNF768 regulates transcription by binding in or near promoter regions of active genes.

ZNF768 binds to genes with cell type-specific expression

Raji and U2OS cells revealed common and cell type-specific binding sites of ZNF768. Common sites were for instance associated with genes for RNA polymerase II subunit E

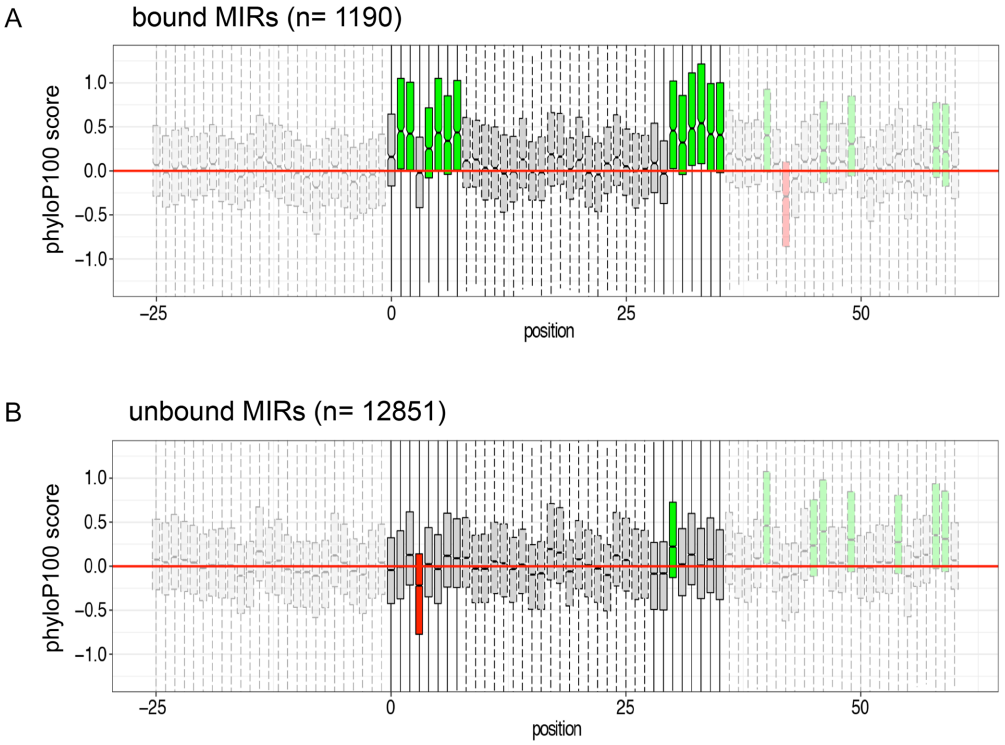

Figure 4. Conservation of the ZNF768 motif (+25 bp on either side) in MIRs either (**A**) bound or (**B**) not bound by ZNF768. Only MIRs were considered that align without gaps to the MIR consensus sequence in the region of the ZNF768 binding motif +25 bp on either side. Distribution of PhyloP100 scores are indicated by boxplots for each position (green = median PhyloP100 score > 0.2) and regions within the motif region are indicated in more intensive colors.

(POLR2E) (Figure 6A) or Solute Carrier Family 1 Member 5 (SLC1A5) and Mitochondrial Ribosomal Protein S5 (MRPS5) (Supplementary Figure S9A,B). These genes are expressed in Raji and U2OS cells and show similar peaks in both cell lines. Thus, many of the 2747 identified common binding sites in Raji and U2OS may be associated with commonly expressed genes. We also identified a large number of peaks that were present either in Raji or U2OS cells. In U2OS cells, strong peaks for ZNF768 were associated with the promoter region of the GAS2L1 gene (Figure 6B), the ID1 and SNPH genes (Supplementary Figure S9C,D) and the gene body of the ANXA2 and ALDH7A1 genes (Supplementary Figure S9E, F), but were absent or only faintly detectable in Raji cells. These genes are expressed in U2OS but not in Raji cells. Inversely, Raji cells showed a strong ZNF768 binding site in the promoter region of the B-Lymphocyte Surface Antigene (CD19) gene (Figure 6C), which is a B cell-specific non-receptor tyrosine kinase required for B cell receptor signaling. Strong Raji-specific peaks were further detected for the genes CD86, ATP2A3, RHOH, PLCG2, LYN, and ARHGDIB (Supplementary Figure S9G–L). These genes are expressed in Raji but not U2OS cells.

A global analysis of differential gene expression between U2OS and Raji cells showed significant differences in fold-changes for genes with peaks specific to either cell line (Figure 6D and Supplementary Table S1). In particular, genes with U2OS-specific peaks were on average 30-fold higher expressed in U2OS cells. For genes with Raji-specific peaks, the fold-changes in gene expression were lower. This may be due to the higher number of peaks identified in Raji cells, indicating a higher sensitivity but lower specificity compared to peaks in U2OS cells. Thus, a significant fraction of seemingly Raji-specific peaks may simply have been missed in U2OS. We conclude that binding of ZNF768 occurs preferentially at expressed genes and at least in part in a cell type-specific manner. The underlying mechanisms regulating the cell type-specific binding of ZNF768 in Raji and U2OS cells are currently unclear, but may involve, e.g. DNA methylation or other epigenetic marks.

ZNF768-regulated genes in U2OS cells

To study the gene regulatory potential of ZNF768, we induced expression of the dominant-negative mutant ZNF768-ΔN (Figure 2E) in U2OS cells and analyzed changes in the transcriptome after 12 h. A >2-fold change in RNA levels was detected for 500 downregulated and 155 upregulated genes (Supplementary Table S2). Functional enrichment analysis of repressed genes revealed several significantly enriched gene sets including two gene sets containing DNA binding proteins (105 genes) and zinc finger proteins (103 genes) (Figure 7A and Supplementary Table S3). Repressed genes in both gene sets show a large overlap (63 genes) with repressed transcription-associated genes (Figure 7B). We conclude that ZNF768 can act as transcriptional regulator and is required particularly for the expression of other transcription factors.

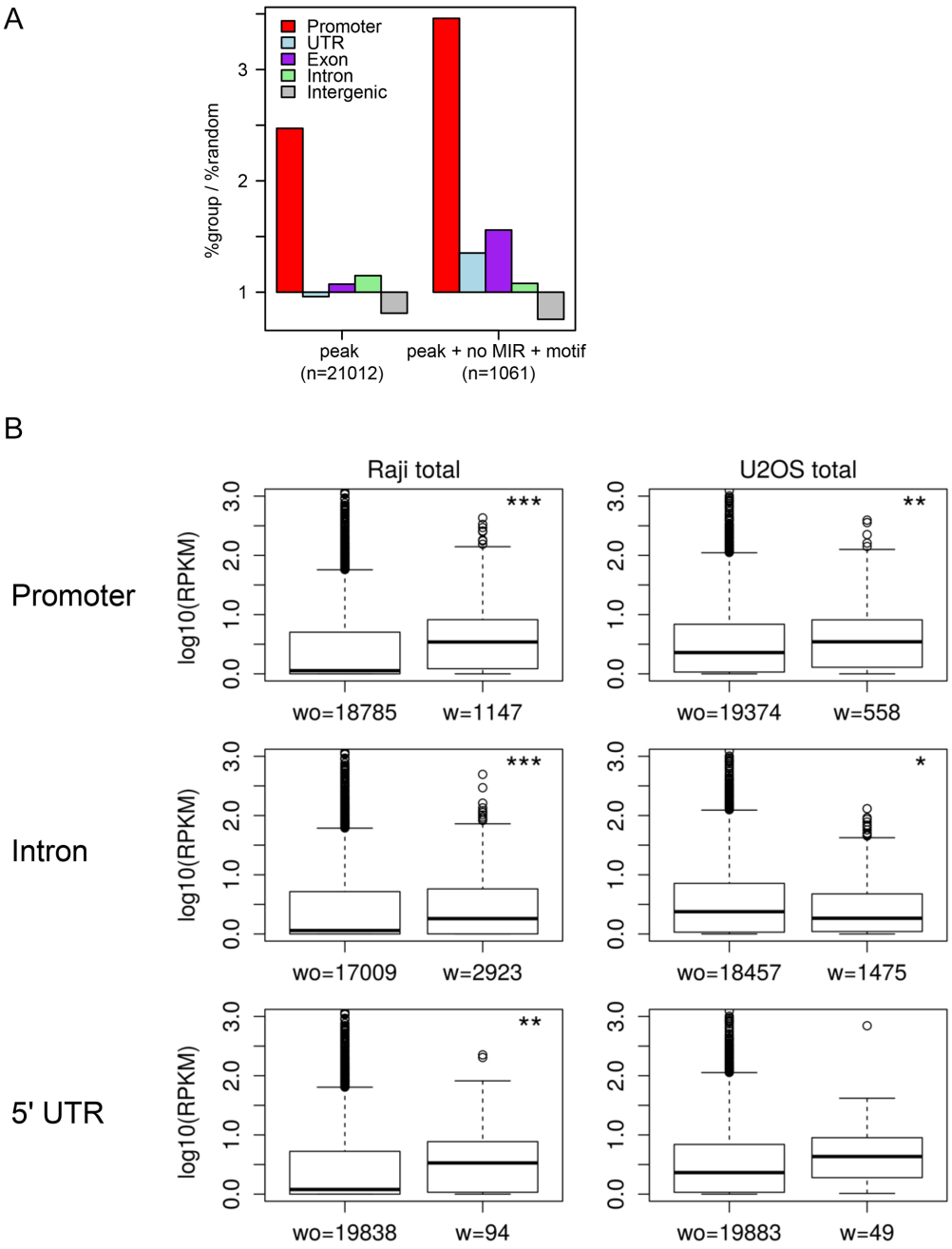

Figure 5. Genomic distribution of the ZNF768 binding motif in Raji and U2OS cells. (**A**) Frequency of ZNF768 binding sites in promoters (–1 kb to transcription start site) and other genomic regions compared to randomly selected binding sites with the same peak length distribution. This shows an enrichment of genomic binding of ZNF768 in promoters, in particular for motif-containing peaks outside of MIRs. (**B**) Boxplots illustrating the distribution of expression levels in total RNA (quantified as RPKM = reads per kilobase per Million mapped reads) in Raji or U2OS cells for genes without (wo) and with (w) peaks in the respective cells. A pseudocount of 1 was added to all RPKM values before plotting. *P*-values for a Wilcoxon rank sum test comparing RPKM levels between the two groups are indicated as: $*P < 10^{-3}$, $**P < 10^{-5}$, $***P < 10^{-10}$

Mass spectrometric analysis of ZNF768 associated factors

We used the mAb 7D6 for a combined immunoprecipitation (IP) and mass spectrometric (MS) assay to identify ZNF768 associated factors. The ZNF768 interactome of Raji and U2OS cells showed a large overlap and twenty of the best thirty interactors were found in both cell lines (Figure 8 A,B, Supplementary Figure S10). Among the common factors we identified three subunits of the Elongator complex (Elp1, Elp2 and Elp3), SR rich splicing factor (SUGP2), centromere protein E (CENPE), several E3 ligases (USP13, Trim33, and HERC2), proteins with centrosomal functions (CEP170-1, Cep170-2 and NIN), and other factors. The binding of Elongator subunit Elp3 to ZNF768 was confirmed in IP experiments with an Elp3-specific antibody (Figure 8C). mAb 7D6 could immunoprecipitate a significant fraction of Elp3 protein of cellular extracts of Raji cells.

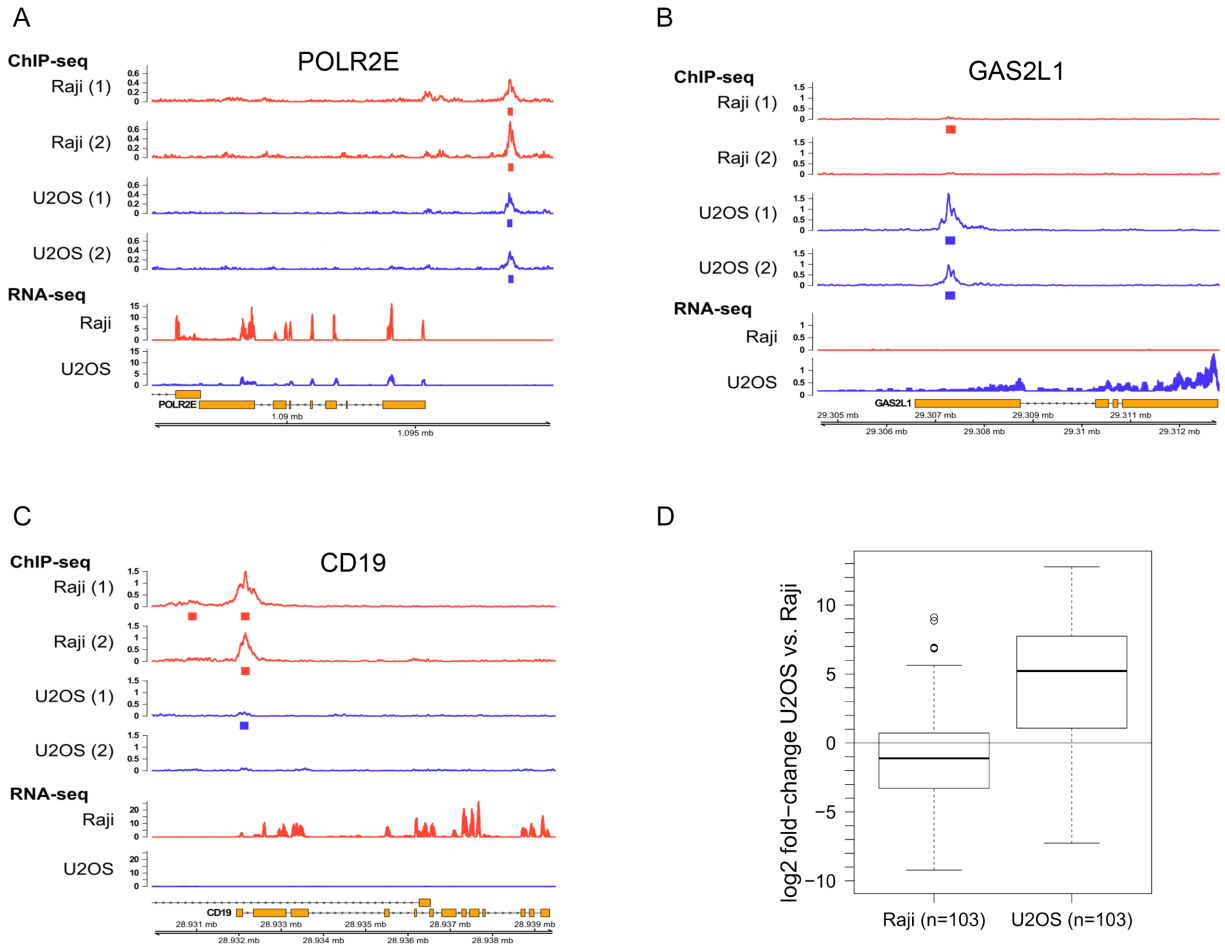

Figure 6. Common and cell type-specific peaks of ZNF768 in Raji and U2OS cells. ChIP-seq (replicates shown separately) and RNA-seq (mean of four replicates) read coverage (in counts per million) for example genes. Identified peaks are shown as rectangles below the corresponding ChIP-seq sample. Genomic coordinates and gene annotation (boxes = exons, lines = introns, strand indicated by arrowheads) are shown in the bottom row. (**A**) ZNF768 binding in the promoter upstream region of the RNA Polymerase II Subunit E (POLR2E) gene, which is expressed in both Raji and U2OS cells. (**B**) Binding of ZNF768 in the promoter region of the Growth Arrest Specific 2-Like (GAS2L1) gene, which is expressed in U2OS but not Raji cells. (**C**) Binding of ZNF768 to the promoter region of the B-Lymphocyte Surface Antigene (CD19) gene, which is expressed in Raji but not U2OS cells. (**D**) Genes in Raji and U2OS cells with cell-specific peaks differ in gene expression. Boxplots illustrate the distribution of fold-changes in gene expression between both cell lines (determined with limma) for genes with cell-specific peaks (= peaks identified in gene body or 1 kb upstream in both replicates of the corresponding cell line but not for the other cell line; to account for the differences in sensitivity between Raji and U2OS ChIP-seq, for Raji only the 103 genes with the top-scoring Raji-specific peaks were evaluated, i.e. the same number of genes as with U2OS-specific peaks). Significance of the difference in median values was determined using the Wilcoxon rank sum test ($***P \leq 10^{-10}$).

The results suggests that ZNF768 can recruit Elongator and other factors to expressed genes in Raji and U2OS cells.

DISCUSSION

ZNF768 binds to MIR sequences

ZNF768 proteins in mammals contain an array of ten zinc fingers that allow the specific binding to DNA. ChIP-seq experiments revealed approximately ten to twenty thousand ZNF768 binding sites in the genome of Raji and U2OS cells. The majority of these sites is contained within MIR sequences and shares a common binding motif that is part of the MIR consensus sequence. The motif of the binding site is 36 bp long and consists of two anchor sequences of 8 bp separated by a linker of 20 bp, which probably does not contribute to the binding specificity of ZNF768 as revealed by gel shift experiments. ZNF768 binds preferentially at or near promoters, suggesting that binding of ZNF768 is associated with gene expression. In agreement with this assumption we observed ZNF768 binding preferentially in euchromatic regions of the nucleus. Likewise, MIR sequences have been reported to be associated with transcriptional active euchromatin but not heterochromatin (5,6). Strikingly, the number of MIR sequences in mammals varies considerably from about 20% of the total genome in monotremes to 1% or 3% of the genome in mice and humans, respectively. Furthermore, the ZNF768 binding motif, although part of the MIR consensus sequence, is not conserved in all MIRs, but only in those displaying a peak in ZNF768 ChIP-seq experiments. Notably, we also detected ∼1000 peaks containing the ZNF768 binding motif outside of MIRs. This category of peaks showed the highest association with promoters.

Given the length of the detected DNA binding motif and the position of the two anchor sequences at its flanks it is

A

description	adj. p–value	# genes	% of downregulated genes
DNA–binding	3.9e–09	105	21%
Zinc–finger	7.2e–09	103	21%
Coiled coil	1.2e–08	144	29%
Phosphoprotein	1.5e–08	325	65%
Zinc	5.4e–08	117	23%
Polymorphism	1.2e–07	343	69%
Transcription regulation	5.2e–07	111	22%
Nucleus	5.8e–07	216	43%
Transcription	8.0e–07	113	23%
Metal–binding	4.4e–06	148	30%
Cell division	4.5e–06	37	7.4%
Cell cycle	4.6e–06	51	10%
Mitosis	7.3e–06	29	5.8%
Chromosome	1.1e–05	33	6.6%
ATP–binding	1.5e–05	76	15%
Nucleotide–binding	1.5e–04	87	17%
Helicase	2.2e–03	16	3.2%
Cytoskeleton	4.8e–03	58	12%
DNA repair	5.6e–03	24	4.8%
DNA damage	5.9e–03	27	5.4%
Ubl conjugation	9.2e–03	81	16%
Microtubule	9.8e–03	21	4.2%
TPR repeat	1.7e–02	15	3%

B

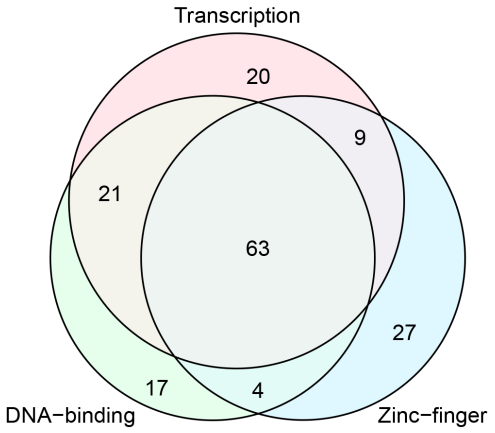

Figure 7. Functional enrichment analysis for UniProt keywords for genes downregulated upon expression of the dominant-negative mutant ZNF768-ΔN in U2OS cells. (**A**) Significantly enriched UniProt keywords (identified with DAVID at an adjusted P-value < 0.05) for downregulated genes (>2-fold down-regulated, adjusted P-value <0.01, for full details see also Supplementary Table S3). (**B**) Venn diagram of downregulated genes annotated with the keywords Transcription, DNA-binding, and Zinc finger. The data indicates that inhibition of ZNF768 downregulates other transcription factors with zinc finger domains.

likely that the proximal and distal, but not the central zinc fingers, of the array of 10 zinc fingers contribute to DNA binding. The potential function of the central zinc fingers is currently unknown. We have currently no evidence for other conserved motifs upstream or downstream of the ZNF768 binding motif. Notably, we also observed ZNF768 peaks at gene loci that do not contain the DNA binding motif. It is currently unclear if binding to these loci requires the zinc finger domain and/or other parts of the protein.

ZNF768 is an essential gene for cell proliferation

Knockdown experiments as well as the expression of dominant-negative mutants revealed the functional requirement of ZNF768 for cell viability and proliferation. Expression of a mutated form of ZNF768 containing only the C-terminal or N-terminal domain, respectively, led to a decline of the cell index in cell proliferation assays. A decline of this index was also seen after siRNA-mediated knockdown of ZNF768 expression. This indicates that the functional loss of ZNF768 cannot be compensated by other cellular factors. Our results suggest that ZNF768, despite being an evolutionary young gene, gained essential function(s) for the expression of growth related genes. A detailed genetic analysis combined with mass spectrometry experiments will be required in the future to analyze the function of ZNF768 in the context of growth control in more detail.

Cell type-specific binding of ZNF768 to gene loci

ChIP-seq analysis of ZNF768 revealed common but also a large number of differential binding sites in Raji and U2OS cells. Furthermore, many putative binding sites containing the binding motif were not occupied by ZNF768 in either Raji or U2OS cells. This observation suggests that binding of ZNF768 to DNA is regulated and that not all binding sites are equally accessible in Raji and U2OS. The mechanism(s) regulating the different accessibility is currently unknown but may include DNA methylation, histone composition at binding motifs or specific histone marks. Additionally, other cellular factors may block or permit binding of ZNF768 to the binding motif. In this context it will be important to determine at which stage of cell differentiation the access of ZNF768 to its binding motif is regulated.

The observed differential binding of ZNF768 in Raji and U2OS cells further prompted us to ask whether binding of ZNF768 can mark differentially expressed genes in both cell lines. In fact, we found a general correlation between ZNF768 binding and the activity of adjacent genes. In particular, we found a correlation between ZNF768 binding and gene expression for those genes that are active only in Raji or U2OS cells. From these data we conclude that binding of ZNF768 can mark commonly as well as cell type-specifically expressed genes in Raji and U2OS cells.

A

Raji

gene name	enrichment factor (log2)	p-value (-log10)
NEURL4	11.50	5.68
ZNF768	11.18	3.27
MSANTD4	10.95	2.10
CXorf38	10.68	5.11
NIPSNAP2	10.50	6.40
MRPS28	10.23	1.67
ELP3	10.05	6.41
CEP170-1	10.00	2.90
ARL1	9.97	5.35
USP13	9.68	2.36
CCDC92	9.50	2.23
CEP170-2	9.50	7.20
MRPS23	8.85	1.47
MCCC1	8.55	3.21
VARS	8.49	5.70
ELP2	8.44	3.13
PRRC1	8.40	1.45
CENPE	8.36	5.04
MRPS31	8.35	1.34
MRPS34	8.34	1.49
TWNK	8.22	4.86
TRIM33	8.16	4.68
MRPS35	7.72	1.39
SUGP2	7.64	1.57
ELP1	6.76	3.66
HERC2	6.26	2.17
MCCC2	6.19	3.06
BMS1	5.82	5.52
KIF2A	5.73	4.20
NIN	4.72	1.57

B

U2OS

gene name	enrichment factor (log2)	p-value (-log10)
USP13	17.24	13.37
ZNF768	17.12	3.87
GAGE2D	16.91	5.71
ELP3	15.92	3.68
ZNF837	13.97	5.98
AEBP1	13.83	2.62
ELP2	13.67	2.37
SRSF5	12.80	6.61
CXorf38	11.78	9.08
MRPS34	11.20	5.91
MCCC2	11.18	2.68
CEP170-2	10.72	5.71
CEP170-1	10.35	3.50
CCDC92	10.32	5.21
HLA-DRB1	10.03	7.30
MRPS14	9.69	1.31
SUGP2	9.44	2.32
ELP1	9.35	3.68
TRIM33	9.06	4.54
MRPS23	8.73	1.37
CCDC47	8.27	1.58
NIN	7.99	3.49
GRPEL1	7.76	1.63
CENPE	7.74	3.25
HERC2	7.44	3.14
MCCC1	7.11	1.32
MRPS26	6.01	1.33
KIF2A	5.53	3.67
MSANTD4	5.39	3.71
AHNAK	5.15	4.03

C

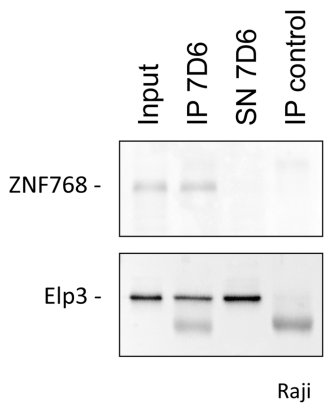

Figure 8. ZNF768 interactome. ZNF768 was immunoprecipitated from cellular extracts of (**A**) Raji and (**B**) U2OS cells. Thirty interaction factors with the highest enrichment are shown. Common factors in both cell lines are depicted in red. The list of all interaction factors is shown in Supplementary Table S4. (**C**) ZNF768 mAb 7D6 specifically co-immunoprecipitates Elp3 from cellular extracts of Raji cells.

ZNF768 functions as transcription factor

Finally, we asked if binding of ZNF768 is required for expression of specific genes. To demonstrate this we studied the transcriptome of U2OS cells 12 h after overexpression of a ZNF768 mutant lacking the N-terminal domain. We found several hundred genes that were significantly repressed after expression of this dominant-negative mutant. We also found a few induced genes, which may be upregulated indirectly. The gene ontology analysis of repressed genes revealed several gene classes related to transcriptional regulation suggesting that ZNF768 is hierarchically located upstream of a network of transcription factor genes and may function as a regulatory master gene for this network.

The notion that ZNF68 may act as a transcription factor was further supported by mass spectrometric analysis of the ZNF768 interactome in Raji and U2OS cells. In both cell

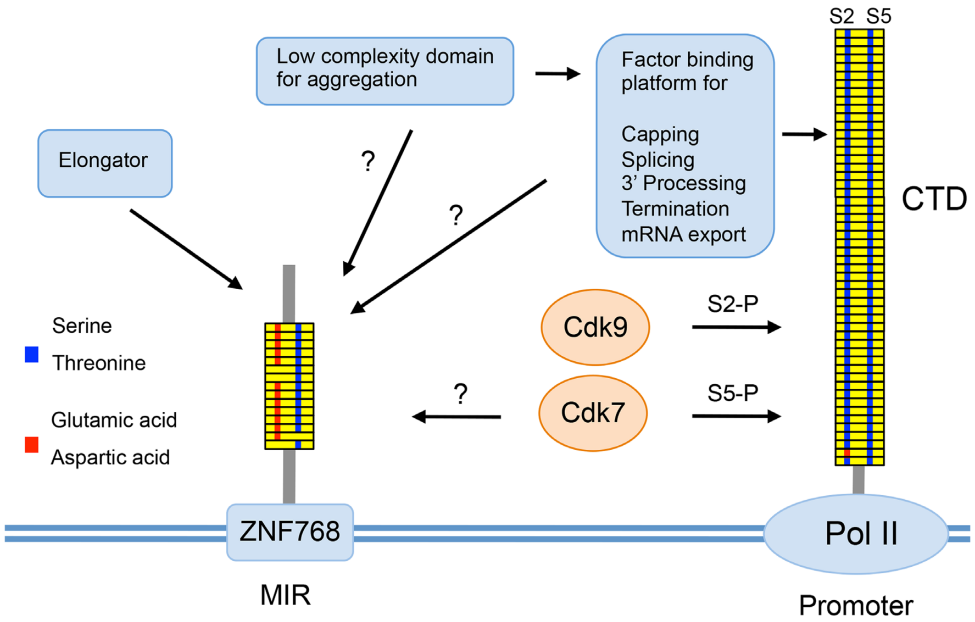

Figure 9. Known functions and regulation of Pol II CTD via its heptad repeat array and the implication with regards to possible functions and regulation of the heptad array in ZNF768 of placentalia (represented by human ZNF768).

lines ZNF768 interacts with subunits Elp1, Elp2 and Elp3 of the Elongator complex. The complex is conserved from yeast to mammals, consists of six subunits, Elp1-6, and has been proposed to function in the control of RNA elongation (56). The Elongator was found associated to the hyperphoshorylated form of Pol II, but the mode of interaction and the involved Elongator subunits are still elusive. Our data suggest that recruitment of Elongator to active genes may also occur by ZNF768. ZNF768 binds first a subcomplex of Elongator consisting of Elp1-3 that subsequently may assemble with subunits Elp4-6. In the future it will be interesting to study if heptad repeats of ZNF768 are involved in the recruitment of Elongator, as suggested for the CTD of Pol II, and if ZNF768 of marsupials lacks the ability of Elongator recruitment.

Originally, the array of heptad repeats in ZNF768 attracted our attention to study the function of ZNF768 as transcriptional activator due to its similarity to the array of heptad repeats in CTD of Pol II. This raises a couple of intriguing questions. First, can this array fulfill similar or related functions as the array of heptad repeats in CTD? If so, can the acidic amino acids that are present at many positions in heptad repeats of ZNF768 mimic a hyperphosphorylated form of Pol ll? Such a mimicry is most likely for position 2 of heptad repeats in ZNF768, which contains glutamic acid in almost all repeats across all species. It is tempting to speculate that binding of ZNF768 can recruit cellular factors to genomic loci that otherwise are recruited only if serine-2 of CTD is phosphorylated, e.g. by Cdk9, or other kinases (see model in Figure 9). In contrast, serine-5 residues are conserved between ZNF768 and the CTD and may depend on phosphorylation in ZNF768, similar as in the CTD, to allow interaction with other factors. Future work will address these and other questions and illuminate if and how the new regulatory network of ZNF768 and MIR sequences has contributed to speciation of placentalia.

DATA AVAILABILITY

GEO submissions: ChIP-Seq (GSE111879), RNA-seq Raji cells (GSE111880), RNA-seq U2OS cells (GSE111881). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (57) partner repository with the dataset identifier PXD010831.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Elisabeth Kremmer for help with the generation of ZNF768 mAb.

FUNDING

D.E. and A.I. were supported by the Deutsche Forschungsgemeinschaft (DFG), SFB1064, Chromatin Dynamics and DFG excellence cluster CIPSM. In D.E. and J.C.A. labs, the work was supported by a German-French BMBF-ANR grant 'EpiGlyco'. CNRS; 'Agence Nationale de la Recherche' (ANR); 'amorcage jeunes équipes' Fondation pour la Recherche Medicale FRM [AJE20130728183 to J.C.A.]; Deutsche Forschungsgemeinschaft (DFG) [FR2938/7-1 and CRC 1123 (Z2) to C.C.F. and M.K.]; Deutsche Forschungsgemeinschaft (DFG) [GE 976/9-2 to M.G.] and is a member of the DFG excellence cluster ImmunoSensation. Funding for open access charge: Helmholtz Center Munich.

Conflict of interest statement. None declared.

REFERENCES

1. Kramerov,D.A. and Vassetzky,N.S. (2011) Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb.)*, **107**, 487–495.
2. Redi,C.A. and Capanna,E. (2012) Genome size evolution: sizing mammalian genomes. *Cytogenet Genome Res.*, **137**, 97–112.
3. Jurka,J., Zietkiewicz,E. and Labuda,D. (1995) Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. *Nucleic Acids Res.*, **23**, 170–175.
4. Smit,A.F. and Riggs,A.D. (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.*, **23**, 98–102.
5. Jjingo,D., Conley,A.B., Wang,J., Marino-Ramirez,L., Lunyak,V.V. and Jordan,I.K. (2014) Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mob. DNA*, **5**, 14.
6. Jjingo,D., Huda,A., Gundapuneni,M., Marino-Ramirez,L. and Jordan,I.K. (2011) Effect of the transposable element environment of human genes on gene length and expression. *Genome Biol. Evol.*, **3**, 259–271.
7. Smith,A.M., Sanchez,M.J., Follows,G.A., Kinston,S., Donaldson,I.J., Green,A.R. and Gottgens,B. (2008) A novel mode of enhancer evolution: the Tal1 stem cell enhancer recruited a MIR element to specifically boost its activity. *Genome Res.*, **18**, 1422–1432.
8. Wang,J., Vicente-Garcia,C., Seruggia,D., Molto,E., Fernandez-Minan,A., Neto,A., Lee,E., Gomez-Skarmeta,J.L., Montoliu,L., Lunyak,V.V. *et al.* (2015) MIR retrotransposon sequences provide insulators to the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E4428–E4437.
9. Varshney,D., Vavrova-Anderson,J., Oler,A.J., Cowling,V.H., Cairns,B.R. and White,R.J. (2015) SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat. Commun.*, **6**, 6569.
10. Yeganeh,M., Praz,V., Cousin,P. and Hernandez,N. (2017) Transcriptional interference by RNA polymerase III affects expression of the Polr3e gene. *Genes Dev.*, **31**, 413–421.
11. Carnevali,D., Conti,A., Pellegrini,M. and Dieci,G. (2017) Whole-genome expression analysis of mammalian-wide interspersed repeat elements in human cell lines. *DNA Res.*, **24**, 59–69.
12. Krull,M., Petrusma,M., Makalowski,W., Brosius,J. and Schmitz,J. (2007) Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res.*, **17**, 1139–1145.
13. Smit,A.F., Hubley,R. and Green,P. (2015). *Repeatmasker Open 4.0*. http://www.repeatmasker.org.
14. Cournac,A., Koszul,R. and Mozziconacci,J. (2016) The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res.*, **44**, 245–255.
15. Feschotte,C. (2008) Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.*, **9**, 397–405.
16. Medstrand,P., van de Lagemaat,L.N. and Mager,D.L. (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.*, **12**, 1483–1495.
17. Testori,A., Caizzi,L., Cutrupi,S., Friard,O., De Bortoli,M., Cora,D. and Caselle,M. (2012) The role of Transposable Elements in shaping the combinatorial interaction of Transcription Factors. *BMC Genomics*, **13**, 400.
18. Matthews,J.M. and Sunde,M. (2002) Zinc fingers–folds for many occasions. *IUBMB Life*, **54**, 351–355.
19. Yang,P., Wang,Y. and Macfarlan,T.S. (2017) The role of KRAB-ZFPs in transposable element repression and Mammalian evolution. *Trends Genet.*, **33**, 871–881.
20. Imbeault,M., Helleboid,P.Y. and Trono,D. (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, **543**, 550–554.
21. Najafabadi,H.S., Mnaimneh,S., Schmitges,F.W., Garton,M., Lam,K.N., Yang,A., Albu,M., Weirauch,M.T., Radovani,E., Kim,P.M. *et al.* (2015) C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.*, **33**, 555–562.
22. Bentley,D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.*, **15**, 163–175.
23. Eick,D. and Geyer,M. (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.*, **113**, 8456–8490.
24. Harlen,K.M. and Churchman,L.S. (2017) The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.*, **18**, 263–273.
25. Jeronimo,C., Collin,P. and Robert,F. (2016) The RNA polymerase II CTD: the increasing complexity of a low-complexity protein domain. *J. Mol. Biol.*, **428**, 2607–2622.
26. Zaborowska,J., Egloff,S. and Murphy,S. (2016) The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.*, **23**, 771–777.
27. Seipel,K., Georgiev,O., Gerber,H.P. and Schaffner,W. (1994) Basal components of the transcription apparatus (RNA polymerase II, TATA-binding protein) contain activation domains: is the repetitive C-terminal domain (CTD) of RNA polymerase II a "portable enhancer domain"? *Mol. Reprod. Dev.*, **39**, 215–225.
28. Suh,H., Hazelbaker,D.Z., Soares,L.M. and Buratowski,S. (2013) The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. *Mol. Cell*, **51**, 850–858.
29. Burke,K.A., Janke,A.M., Rhine,C.L. and Fawzi,N.L. (2015) Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell*, **60**, 231–241.
30. Hnisz,D., Shrinivas,K., Young,R.A., Chakraborty,A.K. and Sharp,P.A. (2017) A phase separation model for transcriptional control. *Cell*, **169**, 13–23.
31. Kwon,I., Kato,M., Xiang,S., Wu,L., Theodoropoulos,P., Mirzaei,H., Han,T., Xie,S., Corden,J.L. and McKnight,S.L. (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell*, **155**, 1049–1060.
32. Schwartz,J.C., Ebmeier,C.C., Podell,E.R., Heimiller,J., Taatjes,D.J. and Cech,T.R. (2012) FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.*, **26**, 2690–2695.
33. Bornkamm,G.W., Berens,C., Kuklik-Roos,C., Bechet,J.M., Laux,G., Bachl,J., Korndoerfer,M., Schlee,M., Holzel,M., Malamoussi,A. *et al.* (2005) Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res.*, **33**, e137.
34. Rohrmoser,M., Holzel,M., Grimm,T., Malamoussi,A., Harasim,T., Orban,M., Pfisterer,I., Gruber-Eber,A., Kremmer,E. and Eick,D. (2007) Interdependence of Pes1, Bop1, and WDR12 controls nucleolar localization and assembly of the PeBoW complex required for maturation of the 60S ribosomal subunit. *Mol. Cell Biol.*, **27**, 3682–3694.
35. Shah,N., Maqbool,M.A., Yahia,Y., El Aabidine,A.Z., Esnault,C., Forne,I., Decker,T.M., Martin,D., Schuller,R., Krebs,S. *et al.* (2018) Tyrosine-1 of RNA polymerase II CTD controls global termination of gene transcription in mammals. *Mol. Cell*, **69**, 48–61.
36. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.*, **25**, 1754–1760.
37. Guo,Y., Mahony,S. and Gifford,D.K. (2012) High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.*, **8**, e1002638.
38. Machanick,P. and Bailey,T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697.
39. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
40. Yu,G., Wang,L.G. and He,Q.Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
41. Harrow,J., Frankish,A., Gonzalez,J.M., Tapanari,E., Diekhans,M., Kokocinski,F., Aken,B.L., Barrell,D., Zadissa,A., Searle,S. *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.*, **22**, 1760–1774.
42. Pollard,K.S., Hubisz,M.J., Rosenbloom,K.R. and Siepel,A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
43. Hahne,F. and Ivanek,R. (2016) Visualizing genomic data using gviz and bioconductor. *Methods Mol. Biol.*, **1418**, 335–351.
44. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
45. Bonfert,T., Kirner,E., Csaba,G., Zimmer,R. and Friedel,C.C. (2015) ContextMap 2: fast and accurate context-based RNA-seq mapping. *BMC Bioinformatics*, **16**, 122.

46. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
47. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
48. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
49. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
50. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Nr.*, **57**, 289–300.
51. Kluge,M. and Friedel,C.C. (2018) Watchdog - a workflow management system for the distributed analysis of large-scale experimental data. *BMC Bioinformatics.*, **19**, 97.
52. Holzel,M., Rohrmoser,M., Schlee,M., Grimm,T., Harasim,T., Malamoussi,A., Gruber-Eber,A., Kremmer,E., Hiddemann,W., Bornkamm,G.W. *et al.* (2005) Mammalian WDR12 is a novel member of the Pes1-Bop1 complex and is required for ribosome biogenesis and cell proliferation. *J. Cell Biol.*, **170**, 367–378.
53. Kellner,M., Rohrmoser,M., Forne,I., Voss,K., Burger,K., Muhl,B., Gruber-Eber,A., Kremmer,E., Imhof,A. and Eick,D. (2015) DEAD-box helicase DDX27 regulates 3′ end formation of ribosomal 47S RNA and stably associates with the PeBoW-complex. *Exp. Cell Res.*, **334**, 146–159.
54. Ishihama,Y., Rappsilber,J. and Mann,M. (2006) Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J. Proteome Res.*, **5**, 988–994.
55. Guo,J. and Price,D.H. (2013) RNA polymerase II transcription elongation control. *Chem. Rev.*, **113**, 8583–8603.
56. Otero,G., Fellows,J., Li,Y., de Bizemont,T., Dirac,A.M., Gustafsson,C.M., Erdjument-Bromage,H., Tempst,P. and Svejstrup,J.Q. (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell*, **3**, 109–118.
57. Vizcaino,J.A., Csordas,A., del-Toro,N., Dianes,J.A., Griss,J., Lavidas,I., Mayer,G., Perez-Riverol,Y., Reisinger,F., Ternent,T. *et al.* (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.*, **44**, 447–456.