Novel ZNF414 activity characterized by integrative analysis of ChIP-exo, ATAC-seq and RNA-seq data

Rodriguez-Martinez, Alejandra

2022-04

Rodriguez-Martinez , A , Vuorinen , E M , Shcherban , A , Uusi-Mäkelä , J , Rajala , N K M , Nykter , M & Kallioniemi , A 2022 , ' Novel ZNF414 activity characterized by integrative analysis of ChIP-exo, ATAC-seq and RNA-seq data ' , Biochimica et Biophysica Acta. Gene Regulatory Mechanisms , vol. 1865 , no. 3 , 194811 . https://doi.org/10.1016/j.bbagrm.2022.194811

http://hdl.handle.net/10138/343695
https://doi.org/10.1016/j.bbagrm.2022.194811

BBA - Gene Regulatory Mechanisms 1865 (2022) 194811

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Novel ZNF414 activity characterized by integrative analysis of ChIP-exo, ATAC-seq and RNA-seq data

Alejandra Rodriguez-Martinez [a,b,c,*,1], Elisa M. Vuorinen [a,b,c,1], Anastasia Shcherban [b,c,d,1], Joonas Uusi-Mäkelä [a,b,c], Nina K.M. Rajala [c], Matti Nykter [a,b,c], Anne Kallioniemi [b,c,e]

[a] Prostate Cancer Research Center, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[b] Tays Cancer Center, Tampere University Hospital, Tampere, Finland
[c] BioMediTech, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[d] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
[e] Fimlab Laboratories, Tampere, Finland

ARTICLE INFO

Keywords:
ZNF414
ChIP-exo
Transcription factor
Gene regulation
ATAC-seq

ABSTRACT

Transcription factor binding to DNA is a central mechanism regulating gene expression. Thus, thorough characterization of this process is essential for understanding cellular biology in both health and disease. We combined data from three sequencing-based methods to unravel the DNA binding function of the novel ZNF414 protein in cells representing two tumor types. ChIP-exo served to map protein binding sites, ATAC-seq allowed identification of open chromatin, and RNA-seq examined the transcriptome. We show that ZNF414 is a DNA-binding protein that both induces and represses gene expression. This transcriptional response has an impact on cellular processes related to proliferation and other malignancy-associated functions, such as cell migration and DNA repair. Approximately 20% of the differentially expressed genes harbored ZNF414 binding sites in their promoters in accessible chromatin, likely representing direct targets of ZNF414. De novo motif discovery revealed several putative ZNF414 binding sequences, one of which was validated using EMSA. In conclusion, this study illustrates a highly efficient integrative approach for the characterization of the DNA binding and transcriptional activity of transcription factors.

1. Introduction

Transcription factor (TF)-mediated regulation of gene expression is a complex process that, to a large extent, defines cellular function. Binding patterns of TFs to DNA sequences may change in disease, leading in turn to altered gene expression and subsequent abnormal cellular function. Thus, adequate approaches to characterize TF binding activity and gene-regulatory function are of great value to understand both normal and pathological cellular processes.

The Cys2-His2 zinc finger (C2H2-ZF) protein family, including 720 members, represents the largest and most diverse group of putative human TFs (out of approximately 1,600 proteins) [1,2]. The C2H2-ZF proteins may be classified based on the number and locations of C2H2-ZF domains and the presence, absence and numbers of other effector domains usually involved in protein-protein interactions, such as KRAB (Krüppel-associated box), BTB (Broad-Complex, Tramtrack, and Bric-a-brac), SCAN (SREZBP, Ctfin51, AW-1 (ZNF174), and Number 18) and SET (Su(var)3–9, Enhancer-of-zeste and Trithorax). The exact biological roles for the majority of the C2H2-ZF proteins are unknown. Recent efforts to gain a better understanding of this matter indicate that most human C2H2-ZF proteins do bind DNA, [3] and those subjected to functional characterization have shown a variety of molecular and genetic functions [2]. For example, several family members, such as ZNF281, ZNF750 and ZBP89, have been associated with features involved in tumorigenicity [4].

ZNF414, a rather recent addition to the C2H2-ZF protein family, was initially discovered in two large-scale proteomic screens aiming to map the human protein interactome [5,6]. Two protein variants are expressed from the ZNF414 gene, differing in the length of the C terminus sequence. Both forms contain three tandem classical C2H2-ZF domains in the central region (between positions 110 and 204) but lack other known effector domains [7]. The function of ZNF414 is

* Corresponding author at: Faculty of Medicine and Health Technology, Tampere University, Arvo Ylpön katu 34, 33520 Tampere, Finland.
E-mail address: alejandra.rodriguezmartinez@tuni.fi (A. Rodriguez-Martinez).
[1] A.R-M., E.M.V. and A.S. contributed equally to this work.

https://doi.org/10.1016/j.bbagrm.2022.194811
Received 16 November 2021; Received in revised form 5 March 2022; Accepted 8 March 2022
Available online 19 March 2022

largely unknown although we previously identified it as a cargo for karyopherin alpha 7 (KPNA7) [8], a nuclear import protein that promotes cancer cell growth and is involved in the maintenance of the nuclear envelope in cancer cells [8,9]. Interestingly, we also discovered a growth promoting role for ZNF414 in pancreatic cancer cells [8]. Furthermore, we demonstrated that ZNF414 is localized in the nucleus, which is compatible with its role as a TF. The C2H2-ZF domains in ZNF414 are organized in an evenly spaced array in the protein sequence [10], thus conferring a high likelihood for ZNF414 to be a sequence-specific DNA binding protein.

In this study, we evaluated the transcriptional regulatory role of ZNF414. To this end, we performed genome-wide analyses using Chromatin Immunoprecipitation with Exonuclease digestion (ChIP-exo), Assay of Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) and RNA sequencing (RNA-seq) to examine whether ZNF414 indeed binds to DNA, to identify its possible DNA binding motifs and to characterize the transcriptional response that results from its activity. ChIP-exo is a modification of the traditional chromatin immunoprecipitation method featuring an exonuclease digestion step of DNA, which allows mapping of the genomic locations of protein binding sites, such as TFs, with highly improved resolution [11]. ATAC-seq permits mapping of chromatin accessibility in a genome-wide manner [12]. The method is based on the use of Tn5 transposase, which inserts sequencing adapters into accessible regions of chromatin, and sequencing reads can then be used to infer regions of open chromatin. The combination of data from these two technologies with global transcriptome analysis by RNA-seq provides an efficient way to assess the putative role of ZNF414 as a TF.

2. Material and methods

2.1. Cell lines

The pancreatic cancer cell line Hs700T and the breast cancer cell line MCF-7 were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). Cell lines were authenticated by genotyping and were grown under recommended culture conditions. The cells were regularly tested for Mycoplasma infection.

2.2. Plasmid transfections and Western blot

A V5-tagged pcDNA6.2/EmGFP-Bsd/V5-DEST construct containing ZNF414 cds (isoform 1, UniProt identifier Q96IQ9-1) was generated using the Genome Biology Unit cloning service (Biocenter Finland, University of Helsinki, Finland). Briefly, the insert from the entry clone from the human ORFeome collaboration library was transferred into the pcDNA6.2/EmGFP-Bsd/V5-DEST destination vector using the standard LR reaction protocol. The construct was transfected into the Hs700T and MCF-7 cells using Lipofectamine 3000 reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Western blot was performed using an antibody against the V5 tag (ab9116, Abcam, Cambridge, UK) as previously described [8], except for detection, which was done with the Pierce ECL Plus Western Blotting Substrate (ThermoFisher Scientific, Waltham, MA, USA).

2.3. ChIP-exo

ChIP-exo was performed on cells transfected to overexpress V5-tagged ZNF414 due to absence of specific antibody for endogenous ZNF414. The overexpression was verified by Western blotting. Biological replicates for ChIP-exo experiments were obtained from two independent transfection experiments. Approximately 50 × 10^6 cells per sample were harvested 48 h posttransfection with pcDNA6.2/EmGFP-Bsd/V5-DEST-ZNF414 plasmid and were fixed with 1% formaldehyde for 10 min to cross-link protein-DNA complexes, followed by quenching with 125 mM glycine. Plasma membranes were lysed in Schmidt's LB1 buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Nonidet P-40, 0.25% Triton X-100) supplemented with Complete Mini protease inhibitor cocktail (Roche, Mannheim, Germany) using a Bioruptor Standard sonication device (Diagenode, Liege, Belgium) run at low power for 3 × 15 *sec* in ice water. Nuclei were pelleted and lysed with ChIP-lysis buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.1% SDS) supplemented with Complete Mini protease inhibitor cocktail (Roche) for 30 min on ice. Chromatin was sonicated to 100–400 bp fragments with a Bioruptor Standard sonication device (Diagenode) run at high power for 3 × 10 min (cycles of 30 s on, 30 s off) in ice water. Cell debris was removed by centrifugation. The IP, exonuclease digestions and library generation were performed using a commercially available ChIP-exo Kit (Active Motif, Carlsbad, CA, USA) following the manufacturer's instructions with the following alteration: DNA purification after decrosslinking was accomplished with a Monarch PCR & DNA cleanup kit (New England Biolabs, Ipswich, MA, USA). An antibody against the V5 tag (ab9116, Abcam) was used for IP. Gel extraction during library size-selection was performed with a Monarch DNA Gel extraction kit (New England Biolabs). An Active Motif's ChIP-IT qPCR control kit (human) was utilized to monitor the functionality of the protocol. Single end, 75 bp sequencing was performed on an Illumina NextSeq sequencer. *The service* was provided by the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki.

2.4. ATAC-seq

Hs700T and MCF-7 cells were transfected with the pcDNA6.2/EmGFP-Bsd/V5-DEST-ZNF414 plasmid and one sample containing 50,000 cells was harvested from both cell lines 48 h posttransfection. The ATAC-seq method was performed according to the previously described protocol [13] with the following modifications: the transposition reaction was incubated for 45 min and the final sample purification was done using a Qiagen MinElute PCR Purification Kit (Qiagen, Hilden, Germany) followed by Agencourt AMPure XP magnetic bead (Beckman Coulter, Brea, CA, USA) purification. Sequencing was performed using Illumina NextSeq High Output 2 × 75 paired end sequencing. *The service* was provided by the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki.

2.5. siRNA transfections, qPCR and proliferation assay

ZNF414 siRNA was obtained from the Dharmacon (Lafayette, CO, USA) siRNA library (siGENOME SMARTpool siRNAs). Transfections were performed on 6-well plates for RNA-seq purposes and on 24-well plates for proliferation assays. Twenty-four hours after seeding, the cells were transfected with 10 nM siRNA using Interferin reagent (Polyplus Transfection, SanMarcos, CA, USA) as per the manufacturer's instructions. An siRNA targeting the firefly luciferase (LUC) gene was used as a control in all experiments. Samples were collected 12 h and 24 h after transfection, and total RNA was extracted using a Nucleospin miRNA kit (Macherey-Nagel, Düren, Germany). qRT-PCR was performed as previously described [8]. The proliferation of MCF-7 cells was assessed using the AlamarBlue reagent (ThermoFisher Scientific) as previously described [14].

2.6. RNA-seq

RNA-seq was performed on two biological replicates per cell line, condition and timepoint (16 samples in total). Library preparation and sequencing services were provided by the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki. Total RNA was subjected to poly A capture, and the sequencing libraries were prepared using the SureSelect

Strand-specific RNA kit (Agilent, Santa Clara, CA, USA). Finally, paired-end sequencing of 75 bp long reads was done on the Illumina NextSeq.

2.7. Isolation of nuclear extracts

Nuclear protein extracts were prepared from cells harvested 48 h posttransfection with either pcDNA6.2/EmGFP-Bsd/V5-DEST-ZNF414 plasmid or the empty vector (mock transfections). All buffers, except PBS, included 1 mM DTT and Complete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, Saint Louis, MO, USA). All steps were performed on ice, with ice-cold reagents, and centrifugations were performed at 4 °C. Briefly, cells were washed with PBS and detached by scrapping. After centrifugation, the cell pellet was resuspended in 1 ml Buffer I (10 mM Hepes (pH 8), 1.5 mM $MgCl_2$, 10 mM KCl) and incubated for 15 min. After the addition of 1% Igepal-CA630, cells were vortexed for 10 s, and the nuclei were pelleted at 15,000 rpm for 3 min. Following resuspension in 175 μl Buffer II (20 mM Hepes (pH 8), 1.5 mM $MgCl_2$, 25% glycerol, 420 mM NaCl, 0.2 mM EDTA), nuclei were vortexed for 30 s and incubated under vigorous rotation for 30 min. Finally, the mixture was centrifuged at 15000 rpm for 15 min and the supernatant was collected. The protein concentration was determined using Bradford reagent (Sigma-Aldrich), and aliquots were stored at −80 °C.

2.8. Electrophoretic mobility shift assay (EMSA)

Protein-DNA complexes were detected using biotin end-labeled double stranded DNA probes. For this purpose, pairs of complementary oligonucleotides, one 5′-biotinylated and the other unmodified, were purchased from Sigma-Aldrich and dissolved in annealing buffer (10 mM Tris, pH 7.5, 50 mM NaCl and 1 mM EDTA). The oligo sequences were designed to comprise the predicted ZNF414 binding motif plus upstream and downstream sequences taken from the promoter region of a differentially expressed gene that contained the specific motif. The probe sequences for motif 10A, taken from the promoter of *LRRC6,* were Biotin-GGAGCTCAGGAAACGCGGCTTGTTTTGGTC and GACCAAAACAAGCCGCGTTTCCTGAGCTCC. For oligo 10B, the probe sequences were obtained from the promoter of *TJP2* and were as follows: Biotin-CATCCGGAAATACCGGTTCCTCCACCCGGA and TCCGGGTGGAGGAACCGGTATTTCCGGATG. Oligos were annealed by heating equal molar amounts of complementary oligonucleotides at 95 °C for 5 min followed by cooling to room temperature for 1 h. The DNA-binding activity of ZNF414 was assessed by EMSA using the LightShift Chemiluminescent EMSA Kit (ThermoFisher Scientific) as recommended by the manufacturer. Binding reactions of 20 μl contained nuclear extract (5 μg of protein), 1× binding buffer, 1 μg Poly (dI·dC), 5% Glycerol, 2.5 mM $MgCl_2$, 100 μM $ZnCl_2$ and biotinylated oligo (20 fmol). The reaction was incubated for 20 min at room temperature. In supershift experiments, 1 μg of either anti-V5 tag antibody (ab9116, Abcam) or control rabbit IgG (sc-2027, Santa Cruz, Dallas, TX, USA) was added to the reaction, followed by an additional 30 min incubation at room temperature.

2.9. Data analysis

2.9.1. ChIP-exo

Raw sequencing reads were aligned against the hg19 reference genome with the Bowtie2 v2.3.2 [15] aligning tool using its very-sensitive parameters setup. Furthermore, aligned reads with a mapping quality less than 1 along with PCR duplicates were discarded using samtools v1.3 [16]. TF binding sites were identified with the MACE v.1.2 [17] tool using the 0.1 cutoff level as the minimum coverage signal used to build the model (specified with –fold parameter). MACE, a pipeline developed specifically for the analysis of ChIP-exo data, classifies the detected peaks into three different categories: GSB, hereafter referred to as *high confidence peaks*, stands for Gale-Sharpley paired Border and has peaks that are detected in both strands; SFB stands for Single Forward Border and means that peaks are present in the forward strand; SRB stands for Single Reverse Border and has peaks that are present in the reverse strand. Peaks belonging to SFB or SRB classes are henceforth referred to as *low confidence peaks*. In addition, peaks were detected using MACS2 v2.1.1 [18] with parameters –nomodel –shift 18 –keep-dup all -f BAM.

2.9.2. ATAC-seq

Raw sequencing reads were trimmed using the TrimGalore tool v0.4.4 (Krueger F. Trim Galore!, accessible at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and aligned against the hg19 reference genome using Bowtie2 v2.3.2 [15] aligner with its very-sensitive default parameters. Furthermore, aligned reads with a mapping quality less than 20 along with PCR duplicates were filtered out using samtools v1.3 [16]. Detection of open chromatin areas was performed with the MACS2 v2.1.1 [18] peak calling tool using the paired-end alignment information setup (–BAMPE parameter), after which the peaks detected within the promoter regions of protein coding genes defined as 2 kb upstream from the Transcription Start Site (TSS) were selected for analysis.

2.9.3. RNA-seq

Raw sequencing reads were aligned with the Tophat2 v2.1.0 [19] aligning tool using parameters for the stranded sequencing library against the hg19 reference genome with subsequent filtering of the aligned reads with a mapping quality lower than 20. The gene expression of protein-coding genes was quantified with the HTSeq v0.7 tool [20] for stranded RNA-seq reads using gencode.v26lift37 annotation. Differential expression analysis using control and treatment samples in each of the two cell lines was performed using the DESeq2 v1.14.1 R package [21]. Genes were filtered based on the absolute log2-fold change (LFC $\geq$ 0.6) and base mean expression signal on a logarithmic scale (log2(baseMean) $>$ 6).

2.9.4. Gene set enrichment analysis

Differentially expressed genes identified with RNAseq with $\geq$0.6 LFC were subjected to gene set enrichment analysis using ToppGene Suite software [22,23] to reveal enriched pathways. Significance threshold was set at $p = 0.05$.

2.9.5. Transcription factor binding motif discovery

High confidence peaks (Gale-Sharpley paired Borders (GSB) detected using the MACE tool) were selected for the motif discovery analysis, which was performed using HOMER v4.11 [24] and MEME v4.10.1 [25] tools using their default parameters. The discovery analysis was repeated multiple times, where each time an increased number of MACE GSB peaks, initially preordered by the peak's score, were selected as motif binding targets. This formed approximately 30 target lists, with the largest target list comprising approximately 1000 peaks in each cell line. The discovered HOMER and MEME motifs were first filtered using a p-value cutoff of 1e-07 and 0.01, respectively, and the candidates observed only in a single motif discovery cycle across the whole range of motif discovery runs were filtered out. Furthermore, the discovered candidate TF binding motifs from the two tools were combined and compared to each other in a pairwise manner and then classified into similarity groups based on a similarity score defined by the cross-correlation coefficient. Given a pair of binding motif position weighted matrices (PWMs) the coefficient was obtained by applying cross-correlation to the arrays of weights corresponding to each nucleotide, after which the average correlation coefficient across the four nucleotides was calculated to represent a single similarity score between the motifs. When estimating the similarity score for a pair of motifs with varying lengths, the shorter motif was first extended with evenly distributed probability values between the four nucleotides. Analyses of discovered motifs were performed in RStudio using the standard built-in libraries. The motif comparison tool Tomtom [26] was used in its web

browser version using the HOCOMOCO database.

2.9.6. Integrative analysis

To identify target genes regulated by ZNF414, first a list of protein binding regions detected from ChIP-exo data by the MACE tool and overlapping or located within 300-bp distance from the areas of accessible chromatin detected from the ATAC-seq data was selected. Next, the protein-bound regions located in open chromatin areas within promoter regions of differentially expressed genes comprised the final list of targets subjected to motif scanning analysis. Motif scanning was performed using discovered motifs with the FIMO v5.0.5 tool, using its default parameters [27].

2.10. Data availability

All sequencing data has been deposited in the Gene Expression Omnibus (GEO) with the accession number GSE153779.

3. Results

3.1. ZNF414 has DNA binding activity that concentrates on the 5′ UTR and promoter regions of the genome

The putative role of human ZNF414 as a TF was studied in the Hs700T pancreatic and MCF-7 breast cancer cell lines. These models were chosen based on their phenotypic response to ZNF414 silencing, i.e., the dramatic slowing of their proliferation ([8] and Supplementary Fig. 1), which suggests the biological relevance of ZNF414 in these cells. We profiled the binding sites of ZNF414 in a genome-wide manner by performing ChIP-exo on cells overexpressing V5-tagged ZNF414 (Supplementary Fig. 2). The overall alignment rate of ChIP-exo reads for the four samples (two replicates for each cell line) was slightly over 97% (Supplementary Fig. 3A). Peak detection with the MACS2 tool resulted in 62,440 and 279,458 peaks for Hs700T and MCF-7, respectively (Supplementary Fig. 3B). More refined analysis with the MACE tool, which detects exonuclease signal borders [17], resulted in 17,039 and 20,935 *high confidence peaks* (see Methods) in Hs700T and MCF-7, respectively. In addition, 364,083 and 365,521 SFB and SRB *lower confidence peaks* (see Methods) were respectively identified for Hs700T, while the corresponding numbers for the MCF-7 cell line were 455,199 and 457,082. More than 70% of the MACS peaks were also identified with MACE (Supplementary Fig. 3B). All in all, a higher number of ChIP-exo peaks were observed in the MCF-7 cell line than in Hs700T cells with either tool (Supplementary Fig. 3B).

Next, we determined accessible chromatin regions using ATAC-seq assays to serve as a filtering tool for artifactual ChIP-exo signals. Accessible chromatin signals around the TSSs followed a typical profile with TSS enrichment scores of 6.205 for Hs700T and 5.153 for MCF-7 (Fig. 1A). Peak calling resulted in the identification of 47,375 and 52,058 peaks in ZNF414 overexpressing Hs700T and MCF-7 cells, respectively. Combining the ChIP-exo and ATAC-seq data, we selected those ChIP-exo peaks that were located in accessible chromatin areas to be used in subsequent analyses. We found that 3,433 (5%) of the ChIP-exo peaks obtained with MACS2 for the Hs700T cell line did overlap with ATAC-seq peaks, while the equivalent number for MCF-7 was 9,206 (3.3%). For both cell lines, approximately 2% of the *low confidence peaks* obtained with MACE were located in open chromatin. In contrast, approximately 5% of all the *high confidence peaks* overlapped with the ATAC-seq signal (861 and 987 peaks for Hs700T and MCF-7, respectively) (Fig. 1B). We then looked at the genomic location of the latter sets of peaks. The absolute number of peaks was highest in intergenic regions, promoters and introns (Fig. 1C). However, when assessing the enrichment of peaks in different locations, ChIP-exo peaks overlapping open chromatin signals were especially concentrated in the promoters and 5′UTR regions and, to a lesser extent, in exons. The results were similar for both cell lines (Fig. 1C).

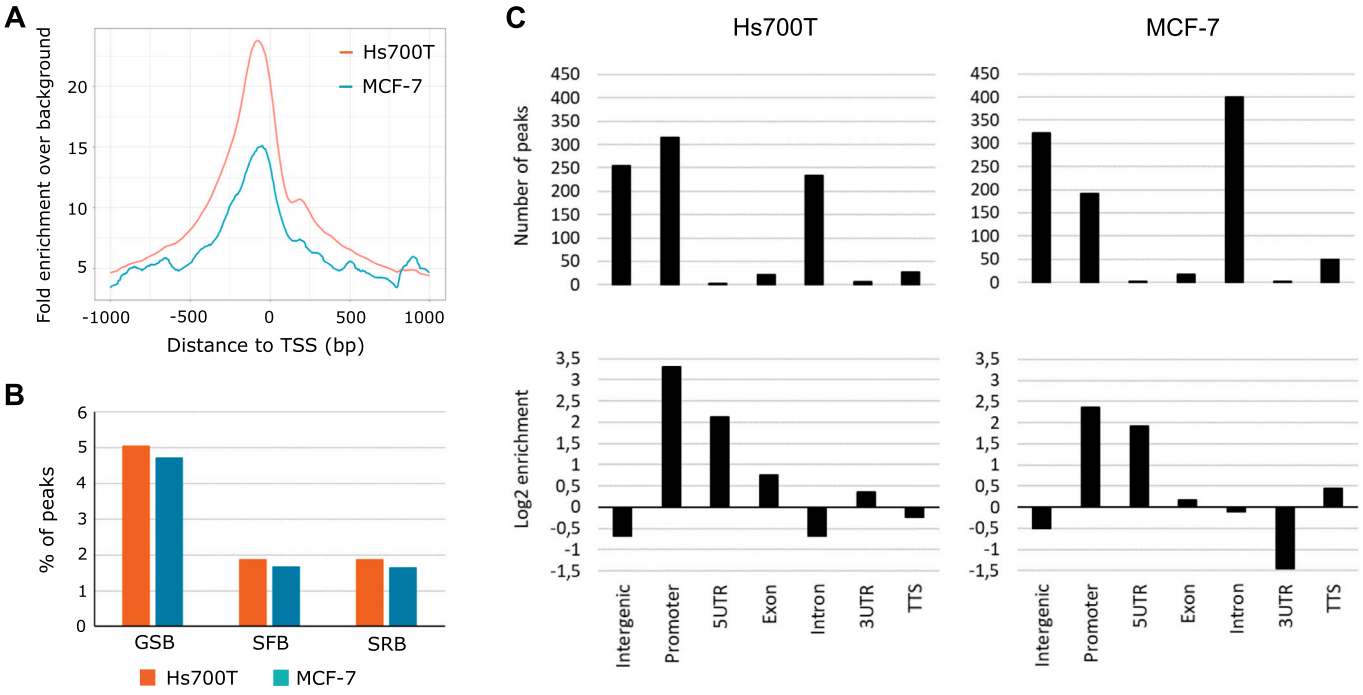


Fig. 1. Integration of ChIP-exo and ATAC-seq data. (A) ATAC-seq signal-to-noise ratio normalized to total mapped reads for Hs700T and MCF-7 cells, shown as a distribution of reads centered on TSS in a 1,000-bp window. (B) Percentage of the different types of ChIP-exo MACE peaks that overlap or are within 300 bp distance from ATAC-seq peaks. GSB, Gale-Sharpley paired borders; SFB, Single Forward Border; SRB, Single Reverse Border. (C) Genomic location and region enrichment for high confidence (GSB) ChIP-exo MACE peaks overlapping accessible chromatin.

3.2. ZNF414 levels affect the expression of genes involved in cell division and proliferation

Based on the preferential location of ZNF414 binding sites in promoters and the 5′UTR regions of coding genes (Fig. 1C), we hypothesized that ZNF414 could be a cis-acting TF. To gain insight into the putative regulatory landscape of ZNF414, the effect of ZNF414 on global gene expression patterns in the Hs700T and MCF-7 cells was studied by the RNA-seq method 12 h and 24 h after knocking down the TF. Efficient reduction of ZNF414 mRNA after siRNA transfection was confirmed by qRT-PCR (Supplementary Fig. 4). The RNA-seq results revealed 52 and 508 differentially expressed genes (DEGs) in the Hs700T cells at the 12 h and 24 h timepoints, respectively (Fig. 2, and Supplementary Table 1A and B). The corresponding numbers of DEGs in the MCF-7 cell line were 250 and 860 (Fig. 2 and Supplementary Table 2A and B). Thus, MCF-7 showed a more overt transcriptional response to ZNF414 silencing than Hs700T in terms of the number of genes regulated, which is in accordance with the ChIP-exo results demonstrating more ZNF414 binding sites in the MCF-7 cell line. Proportions of up- and downregulated DEGs varied between samples, with upregulation being a more common effect observed after ZNF414 silencing (Fig. 2A and C). There were 32 and 211 DEGs shared by the two cell lines at 12 h and 24 h, respectively, of which 27 were common to both timepoints (Fig. 3B and Supplementary Table 3).

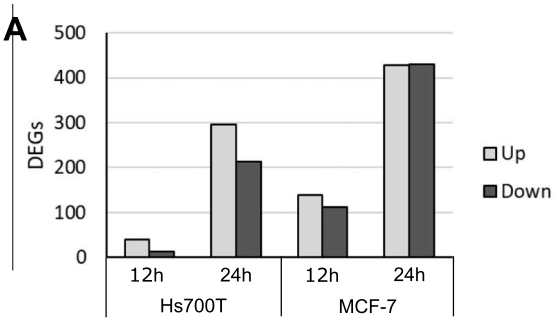
Gene ontology analyses of the DEGs in both cell lines revealed enrichment of many functional categories related to cellular proliferation, such as DNA replication and its initiation, regulation of epithelial cell proliferation and G1/S transition of the mitotic cell cycle (Tables 1 and 2). In addition, some other functional categories that have relevance in terms of the cancer-associated properties of cells were also found to be enriched, such as epithelial cell migration and differentiation in Hs700T (Table 1) and DNA recombination and DNA repair in MCF-7 (Table 2).

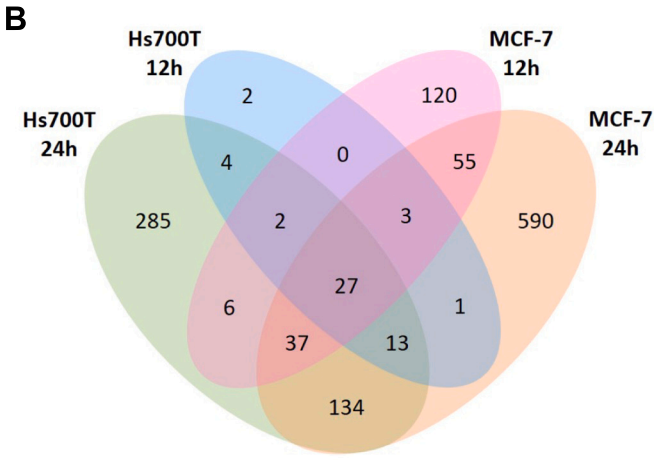
We then asked whether our approach could detect binding of ZNF414 in the promoters of DEGs. Combining ChIP-exo, ATAC-seq and RNA-seq data we found that approximately 20% of the identified DEGs harbor ZNF414 binding sites overlapping or near accessible chromatin in their promoter (Fig. 2C). Furthermore, this percentage did not vary consistently between up- and downregulated genes.

3.3. Novel DNA binding motifs for ZNF414

To search for putative DNA binding motifs of ZNF414, all ChIP-exo MACE peaks falling in the *high confidence* Gale-Sharpley paired Border (GSB) group, and thus present in both strands, were used as input for de novo motif discovery analysis. Two tools, HOMER and MEME, were used. Motif selection criteria included: a minimum of 0.5% of the target sequences must contain the motif and the same motif must occur in at least two motif discovery cycles. Additionally, *p*-value thresholds of 10^{-5} and 0.01 were applied for HOMER and MEME motifs, respectively, and consensus sequences were created based on IUPAC rules. The resulting motifs from the two tools were separately analyzed to identify conserved motifs, i.e., those with a similarity score equal to or greater than 0.95. This resulted in 62 and 34 conserved motifs from HOMER and MEME, respectively, with a total of 96 motifs (Supplementary Fig. 5). These were then further clustered into similarity groups using a similarity score threshold of 0.7 and a minimum size of 2 motifs per group. Out of the 96 conserved motifs, 57 fell into one of the 18 discovered similarity groups, which ranged in size from 2 to 8 motifs. Typically, the motifs discovered with HOMER were shorter and thereby often included in the longer motifs predicted by MEME within each similarity group (Supplementary Table 4). Depending on the cell line and motif discovery tool, different types of similarity groups were identified (Fig. 3a). For example, groups 1 and 4 contained motifs that were present in data obtained using both HOMER and MEME tools and from both cell lines. Motifs included in group 8 were present only in Hs700T data, whereas those in groups 5 and 10 were MCF-7-specific. The discovered motifs were hand curated to identify the most likely ZNF414 binding motifs based on motif characteristics, such as length, and by subjecting them to an in silico analysis with Tomtom tool to find similarities to known transcription factor binding motifs. The PWM logos of the most likely true ZNF414 binding motifs are displayed in Fig. 3b.

We used EMSA to experimentally validate the binding of ZNF414 to





C

			Hs700T			MCF-7		
			All	Up	Down	All	Up	Down
12h	Total	n	52	40	12	250	138	112
	+ChIP +ATAC	n	12	10	2	43	33	10
		%	23	25	17	17	24	9
24h	Total	n	508	295	213	860	429	431
	+ChIP +ATAC	n	99	26	43	146	67	79
		%	19	9	20	17	16	18

Fig. 2. Evaluation of the transcriptional response to ZNF414 silencing by RNA-seq. (A) The number of upregulated and downregulated DEGs in Hs700T and MCF-7 cell lines at 12 h and 24 h time points after ZNF414 silencing. (B) Venn diagram showing intersections of DEGs between the different cell lines and time points. (C) Numbers of DEGs that harbor ChIP signal (detected with MACE) near accessible chromatin (within 300 bp) in their promoters (2000 bp upstream from TSS). Percentages relative to total DEGs are also shown (blue boxes).

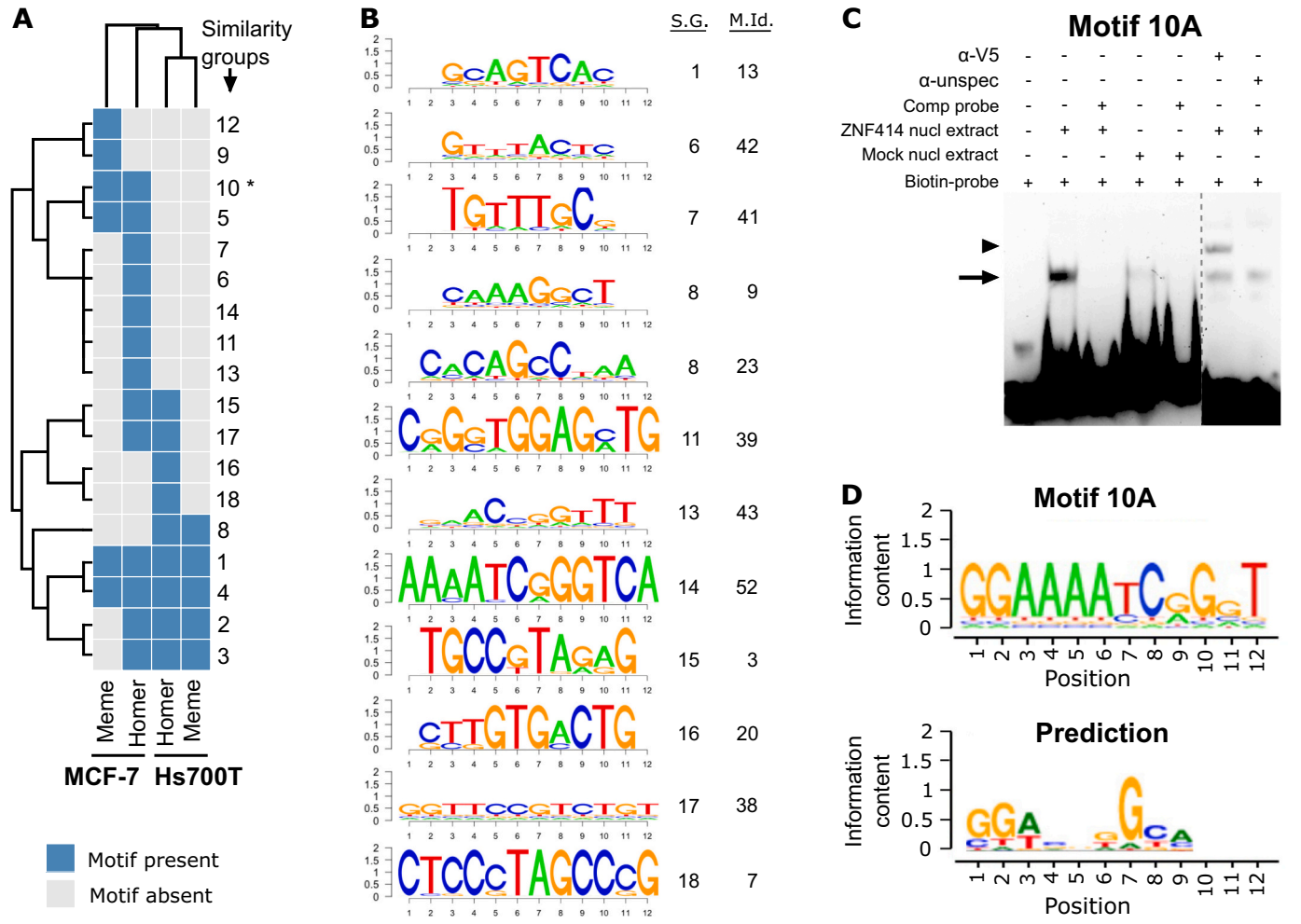


Fig. 3. Motif discovery and experimental validation. (A) Clustering of the motifs into similarity groups (right) based on the data they originate from (bottom), * similarity group that contains a validated motif. (B) Position weight matrix logos of hand-curated potentially real binding targets for ZNF414. S.G. = similarity group, M.Id. = motif ID. (C) EMSA experiments were done on nuclear protein extracts from MCF-7 cells overexpressing V5-tagged ZNF414. Mock-transfected cells were also used to detect possible binding of the oligo 10A to the endogenous ZNF414. Arrow, shift indicating any protein binding to the tested oligo; arrowhead, supershift indicating V5 antibody binding to the protein-oligo complex, which disappears when an isotype control IgG is used instead (lane 6). The dotted line represents separation between two different blots. (D) Position weight matrices of the EMSA-validated motif 10A and the in silico predicted binding motif for ZNF414 based on ZF Predictor http://zf.princeton.edu/.

an identified consensus motif from similarity group 10. The similarity group consisted of 2 motif sequences, occurring in the promoters 51 DEGs (Supplementary Table 5). To validate the motifs, we selected both motifs in group 10 (Supplementary Table 4). One motif from group 10 (Motif 10A) was experimentally validated in the MCF-7 cell line using EMSA. When the biotinylated probe was incubated with the nuclear extract from ZNF414-overexpressing cells, a band shift was observed which disappeared or was clearly reduced upon addition of excess unlabeled probe. A shifted complex which migrated to the same location in the gel was also observed, when extracts from mock-transfected cells were used. Furthermore, the probe-bound protein was identified as V5-tagged by a band supershift that became evident after addition of the V5-antibody but was absent with an isotype control antibody (Fig. 3c). We then performed a DNA-binding specificity prediction for ZNF414 using a web server for predicting a PWM for a C2H2-ZF domain protein binding site, available online at http://zf.princeton.edu [28,29]. The predicted binding motif (Fig. 3d) was similar to motif 10A.

To sum up our findings, Fig. 4 illustrates the promoter of CCNG2, one of the genes putatively regulated by ZNF414, harboring identified 10A motifs that overlap with ZNF414-binding sites and open chromatin, as evidenced by ChIP-exo and ATAC-seq peaks, respectively.

4. Discussion

In this study, we performed large-scale genomic and transcriptomic analyses in the forms of ChIP-exo, ATAC-seq and RNA-seq, and integrated all of the generated data to obtain a comprehensive view of the DNA binding and gene regulatory activity of the zinc finger protein ZNF414. This study represents the first characterization of the putative DNA binding activity of ZNF414, a recent addition to the zinc finger protein family. ZNF414 belongs to a group of over 200 proteins that do not harbor other conserved domains in addition to the zinc fingers, and are expected to bind at least to DNA [30].

To map the genomic binding sites of ZNF414, the ChIP-exo method was applied. ChIP-exo data provides higher resolution information about the binding sites than traditional ChIP-seq data. For this reason, several tools have been developed that specifically build on the characteristics of the ChIP-exo data, one being MACE. When analyzed with this tool, our data shows similar proportions of different peak types to those obtained when MACE was applied to published ChIP-exo data, such as that from Starick et al. [31]. Our results experimentally verify the binding of ZNF414 to DNA and demonstrate a clear preference for gene promoters and 5′ UTRs, indicative of a role as a conventional TF, in contrast with, for example, the prominent binding of KRAB containing

Table 1
Enrichment of gene ontology terms for genes differentially expressed in Hs700T cells 24 h after ZNF414 siRNA transfection, considering a LFC cutoff of 0.6.

Category		# genes	p-value
Biological process			
GO:0006270	DNA replication initiation	11	1.1E-06
GO:0006260	DNA replication	23	4.99E-05
GO:0050678	Regulation of epithelial cell proliferation	26	4.99E-05
GO:0010631	Epithelial cell migration	22	4.99E-05
GO:0010594	Regulation of endothelial cell migration	12	0.011
GO:0000082	G1/S transition of mitotic cell cycle	18	0.011
GO:1901342	Regulation of vasculature development	18	0.013
GO:1902531	Regulation of intracellular signal transduction	72	0.014
GO:0008285	Negative regulation of cell proliferation	34	0.024
GO:0098773	Skin epidermis development	9	0.031
GO:0031325	Positive regulation of cellular metabolic process	105	0.033
GO:0060026	Convergent extension	4	0.033
GO:0030855	Epithelial cell differentiation	30	0.033
GO:0051272	Positive regulation of cellular component movement	24	0.033
GO:0097306	Cellular response to alcohol	6	0.035
GO:0071407	cellular response to organic cyclic compound	28	0.037
GO:0022008	Neurogenesis	60	0.040
GO:0010647	Positive regulation of cell communication	63	0.040
GO:0045765	Regulation of angiogenesis	15	0.042
GO:0009967	Positive regulation of signal transduction	58	0.046
GO:1904018	Positive regulation of vasculature development	11	0.046
GO:0022616	DNA strand elongation	5	0.046
Cellular component			
GO:0043596	Nuclear replication fork	10	4.83E-06

Table 2
Gene Ontology terms enriched among the genes differentially expressed (LFC cutoff is 0.6) in MCF-7 cells 24 h after transfection with ZNF414-targeting siRNAs.

Category		# genes	p-value
Biological process			
GO:0006260	DNA replication	34	5.32E-05
GO:0006270	DNA replication initiation	11	5.32E-05
GO:0050678	Regulation of epithelial cell proliferation	29	0.007
GO:0045786	Negative regulation of cell cycle	42	0.007
GO:0000732	Strand displacement	7	0.011
GO:0010942	Positive regulation of cell death	47	0.011
GO:0010604	Positive regulation of macromolecule metabolic process	158	0.011
GO:0000082	G1/S transition of mitotic cell cycle	23	0.017
GO:0031325	Positive regulation of cellular metabolic process	155	0.023
GO:0030182	Neuron differentiation	78	0.025
GO:0006310	DNA recombination	24	0.025
GO:0048762	Mesenchymal cell differentiation	19	0.025
GO:0031572	G2 DNA damage checkpoint	7	0.030
GO:0006281	DNA repair	23	0.046
Cellular component			
GO:0043596	Nuclear replication fork	9	0.002
Molecular function			
GO:0003697	Single-stranded DNA binding	13	0.027
GO:0004668	Protein-arginine deiminase activity	3	0.027
GO:0016462	Pyrophosphatase activity	54	0.027
GO:0001077	Transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	21	0.027

ZF proteins to transposable elements [32].

To gain insight into the transcriptional regulation mediated by ZNF414, we performed RNA-seq to identify differentially expressed genes (DEGs). These analyses revealed tens to hundreds of DEGs (depending on the cell line) at the 12 h timepoint, thus providing further evidence for the function of ZNF414 as a traditional TF directly regulating gene expression. There were a number of shared DEGs between the two cell lines derived from different tissue types, suggesting that the transcriptional changes induced by ZNF414 are, at least to some extent, reproducible and not cancer-type specific. TFs have been conventionally classified as activators or repressors. However, there is abundant evidence indicating that the multifunctionality of TFs is more of a rule than an exception [33,34], and this phenomenon has also been reported for C2H2-ZF proteins [30]. Our data does not indicate a clear position of ZNF414 on either end of this scale, although based on the ratio of downregulated vs. upregulated DEGs the repression of gene expression is somewhat more common, especially in the Hs700T cells. To this regard, interrogating transcriptional regulation at an earlier timepoint might allow more decisive conclusions to be drawn.

To comprehensively understand the function of the ZNF414-regulated genes, we used gene ontology (GO) analysis to segregate the DEGs into biological process, cellular component and molecular function categories. The ontology analyses revealed enrichment of processes linked to DNA replication, cell cycle regulation, cell division and proliferation. These results are in accordance with our previous and current functional experiments in which ZNF414 knockdown led to a dramatic reduction in the proliferation of Hs700T [8] and MCF-7 cells. The enriched biological process categories were notably similar in both cell lines. Thus, ZNF414 function possibly converges in the same cellular pathways.

Integration of all generated data, namely, ChIP-exo, ATAC-seq and RNA-seq, allowed us to identify the most likely direct ZNF414 binding sites. By combining these three types of data, we showed that approximately 20% of DEGs harbored in their promoter ZNF414 binding signals overlapping or near accessible chromatin, thus most likely representing the direct target genes of ZNF414. Reports integrating all these types of data are lacking in the literature, preventing us from contrasting our results with previous studies to assess the universality of this phenomenon. DEGs without overlapping ChIP-exo and ATAC peaks in their promoter may reflect either direct or indirect regulation by ZNF414. The former could occur through binding of ZNF414 to distant sites containing enhancers or other regulatory sequences. Our results, showing abundant (although not enriched) binding of ZNF414 in intergenic regions, are not in conflict with this option. Regarding the latter, there is support in our data for this type of regulation, as illustrated by the presence of differentially expressed TFs. At any time point, there were 34 and 48 differentially expressed TFs in Hs700T and MCF-7 cell lines, respectively, of which 15 were shared by the two cell lines. TFs were defined based on the list of all known human TFs [35]. For example, E2F Transcription Factor 2 (E2F2), MYB Proto-Oncogene (MYB) and multiple zinc finger proteins were differentially expressed at the 24 h timepoint in both cell lines. E2F2 belongs to a family of TFs that play a crucial role in the control of the cell cycle and action of tumor suppressor proteins (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=1870). MYB plays an essential role in the regulation of hematopoiesis and its aberrant expression is relevant in the development of leukemias and lymphomas (https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=4602). Furthermore, the GO analysis highlights the molecular function category “transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding” as enriched in the MCF-7

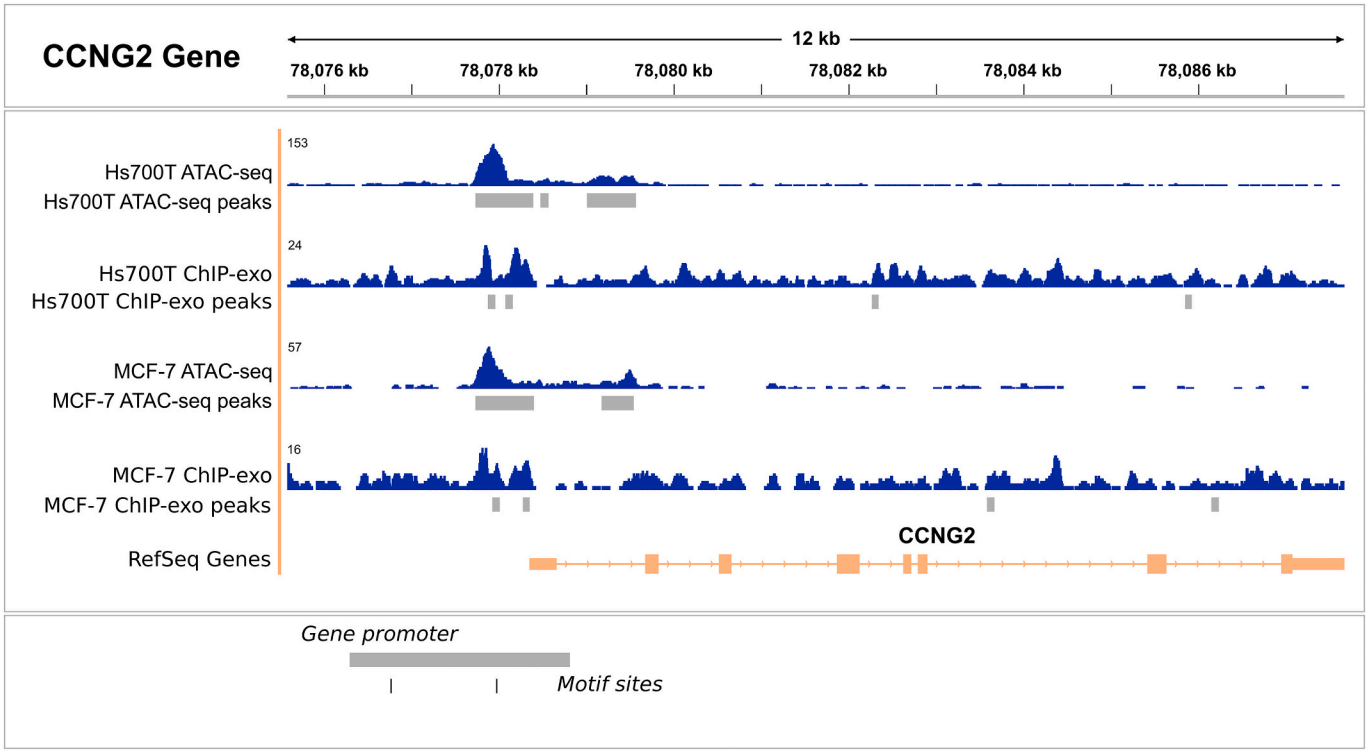


Fig. 4. Representative data from the genomic region (12 kb) harboring the CCNG2 gene and its promoter. Read densities for ATAC-seq and ChIP-exo (averaged duplicates) are shown in dark blue. Peaks called with MACS2 and MACE, respectively, are depicted in grey colour. Motif sites in the lower panel include motifs in similarity group 10.

cell line. Thus, it seems that ZNF414 induces secondary changes in gene expression by modulating the expression of other transcriptional regulators.

Previous studies have shown great diversity in the sequence preferences of ZNF proteins [3,30,36]. For this reason, we used a de novo approach for motif discovery. Despite accumulating evidence supporting the widespread contributions of sequence context, such as flanking sequences and DNA shape, in modulating sequence recognition by TFs [37], we were able to successfully experimentally validate one binding motif for ZNF414. The validated motif belonged to similarity group 10, which contains two predicted motifs. Motif 10A showed similarity to the binding motif predicted in silico for ZNF414 sequence. The motif's structure also corresponds to the structure of ZNF414: as each C2H2 domain identifies approximately four bases with a spacer sequence in between, the length of the motif is appropriate to enable the binding of at least two C2H2 domains. Based on these results, we are confident that motif 10A represents a true ZNF414 binding site. However, motif 10A is most likely not the only sequence that is bound by ZNF414 as the in vitro conditions during the EMSA assay might not be optimal for the binding of ZNF414 to all the motifs.

An example of a DEG upregulated in both cell lines and containing motif 10A in its promoter, is cyclin G2 (CCNG2). This gene participates in growth regulation and in negative regulation of cell cycle progression (https://www.uniprot.org/uniprot/Q16589). It has been reported to regulate cell proliferation by acting as a tumor suppressor, and its decreased expression is associated with malignant phenotypes in many cancer types [38].

In conclusion, by combining genome-wide experimental data with innovative, integrative computational analyses, we were able to extend our knowledge of the function of the novel transcription factor ZNF414. Furthermore, we demonstrated that combining data on transcription factor binding (ChIP-exo) and chromatin state (ATAC-seq) with gene expression (RNA-seq) is a powerful approach to identify de novo TF binding motifs.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2022.194811.

Funding

This work was supported by Jenny and Antti Wihuri Foundation, Finland (E.M.V.), Finnish Cultural Foundation's Pirkanmaa regional fund, Finland (A.R-M., E.M.V., J.U-M.), Academy of Finland, Finland project no. 312043 and 310829 (M.N.), Cancer Society of Finland, Finland (M.N., A.K.)

CRediT authorship contribution statement

Alejandra Rodriguez-Martinez: Conceptualization, Investigation, Validation, Writing – Original Draft, Writing – Review & Editing, Visualization, Funding acquisition. **Elisa M. Vuorinen**: Conceptualization, Investigation, Validation, Writing – Original Draft, Writing – Review & Editing, Funding acquisition. **Anastasia Shcherban**: Software, Formal analysis, Visualization, Writing – Review & Editing. **Joonas Uusi-Mäkelä**: Investigation, Writing – Review & Editing, Funding acquisition. **Nina K.M. Rajala**: Conceptualization. **Matti Nykter**: Resources, Conceptualization, Writing – Review & Editing, Funding acquisition. **Anne Kallioniemi**: Conceptualization, Resources, Supervision, Project Administration, Writing – Review & Editing, Funding acquisition. All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Ms. Kati Rouhento, Emilia Barannik B.Sc., and Vidal Fey Ph.D. for skillful assistance in this study.

References

[1] J.M. Vaquerizas, S.K. Kummerfeld, S.A. Teichmann, N.M. Luscombe, A census of human transcription factors: function, expression and evolution, Nat. Rev. Genet. 10 (2009) 252–263, https://doi.org/10.1038/nrg2538.

[2] L. Stubbs, Y. Sun, D. Caetano-Anolles, Function and evolution of C2H2 zinc finger arrays, Subcell. Biochem. 52 (2011) 75–94, https://doi.org/10.1007/978-90-481-9069-0_4.

[3] H.S. Najafabadi, S. Mnaimneh, F.W. Schmitges, M. Garton, K.N. Lam, A. Yang, et al., C2H2 zinc finger proteins greatly expand the human regulatory lexicon, Nat. Biotechnol. 33 (2015) 555–562, https://doi.org/10.1038/nbt.3128.

[4] M. Cassandri, A. Smirnov, F. Novelli, C. Pitolli, M. Agostini, M. Malewicz, et al., Zinc-finger proteins in health and disease, Cell Death Discov. 3 (2017) 1–12, https://doi.org/10.1038/cddiscovery.2017.71.

[5] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, et al., Towards a proteome-scale map of the human protein–protein interaction network, Nature 437 (2005) 1173–1178, https://doi.org/10.1038/nature04209.

[6] T. Rolland, M. Taşan, B. Charloteaux, S.J. Pevzner, Q. Zhong, N. Sahni, et al., A proteome-scale map of the human interactome network, Cell 159 (2014) 1212–1226, https://doi.org/10.1016/J.CELL.2014.10.050.

[7] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, et al., UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169, https://doi.org/10.1093/NAR/GKW1099.

[8] E.M. Vuorinen, N.K. Rajala, H.E. Rauhala, A.T. Nurminen, V.P. Hytönen, A. Kallioniemi, Search for KPNA7 cargo proteins in human cells reveals MVP and ZNF414 as novel regulators of cancer cell growth, Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1863 (2017) 211–219, https://doi.org/10.1016/J.BBADIS.2016.09.015.

[9] E. Laurila, E. Vuorinen, K. Savinainen, H. Rauhala, A. Kallioniemi, KPNA7, a nuclear transport receptor, promotes malignant properties of pancreatic cancer cells in vitro, Exp. Cell Res. 322 (2014) 159–167, https://doi.org/10.1016/J.YEXCR.2013.11.014.

[10] AlphaFold, Protein Structure Database, 2021. https://alphafold.ebi.ac.uk/. (Accessed 18 January 2022).

[11] H.S. Rhee, B.F. Pugh, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution, Cell 147 (2011) 1408–1419, https://doi.org/10.1016/J.CELL.2011.11.013.

[12] J.D. Buenrostro, B. Wu, H.Y. Chang, W.J. Greenleaf, ATAC-seq: a method for assaying chromatin accessibility genome-wide, Current Protocols in Molecular Biology 109 (2015), https://doi.org/10.1002/0471142727.MB2129S109, 21.29.1-9.

[13] J.D. Buenrostro, P.G. Giresi, L.C. Zaba, H.Y. Chang, W.J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position, Nat. Methods 10 (2013) 1213–1218, https://doi.org/10.1038/nmeth.2688.

[14] M. Ampuja, R. Jokimäki, K. Juuti-Uusitalo, A. Rodriguez-Martinez, E.L. Alarmo, A. Kallioniemi, BMP4 inhibits the proliferation of breast cancer cells and induces an MMP-dependent migratory phenotype in MDA-MB-231 cells in 3D environment, BMC Cancer 13 (2013) 1–13, https://doi.org/10.1186/1471-2407-13-429/FIGURES/7.

[15] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.

[16] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The sequence Alignment/Map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, https://doi.org/10.1093/BIOINFORMATICS/BTP352.

[17] L. Wang, J. Chen, C. Wang, L. Uusküla-Reimand, K. Chen, A. Medina-Rivera, et al., MACE: model based analysis of ChIP-exo, Nucleic Acids Research 42 (2014), https://doi.org/10.1093/NAR/GKU846 e156–e156.

[18] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, et al., Model-based analysis of ChIP-seq (MACS), Genome Biol. 9 (2008) 1–9, https://doi.org/10.1186/GB-2008-9-9-R137/FIGURES/3.

[19] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (2013) 1–13, https://doi.org/10.1186/GB-2013-14-4-R36/FIGURES/6.

[20] S. Anders, P.T. Pyl, W. Huber, HTSeq—a python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169, https://doi.org/10.1093/BIOINFORMATICS/BTU638.

[21] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 1–21, https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9.

[22] ToppGene Suite, 2009. https://toppgene.cchmc.org/. (Accessed 18 January 2022).

[23] J. Chen, E.E. Bardes, B.J. Aronow, A.G. Jegga, ToppGene suite for gene list enrichment analysis and candidate gene prioritization, Nucleic Acids Res. 37 (2009) W305–W311, https://doi.org/10.1093/NAR/GKP427.

[24] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Mol. Cell 38 (2010) 576–589, https://doi.org/10.1016/J.MOLCEL.2010.05.004.

[25] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, et al., MEME suite: tools for motif discovery and searching, Nucleic Acids Res. 37 (2009) W202–W208, https://doi.org/10.1093/NAR/GKP335.

[26] S. Gupta, J.A. Stamatoyannopoulos, T.L. Bailey, W.S. Noble, Quantifying similarity between motifs, Genome Biol. 8 (2007) 1–9, https://doi.org/10.1186/GB-2007-8-2-R24/TABLES/3.

[27] C.E. Grant, T.L. Bailey, W.S. Noble, FIMO: scanning for occurrences of a given motif, Bioinformatics 27 (2011) 1017–1018, https://doi.org/10.1093/BIOINFORMATICS/BTR064.

[28] A.V. Persikov, M. Singh, De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins, Nucleic Acids Res. 42 (2014) 97–108, https://doi.org/10.1093/NAR/GKT890.

[29] A.V. Persikov, R. Osada, M. Singh, Predicting DNA recognition by Cys2His2 zinc finger proteins, Bioinformatics 25 (2009) 22–29, https://doi.org/10.1093/BIOINFORMATICS/BTN580.

[30] F.W. Schmitges, E. Radovani, H.S. Najafabadi, M. Barazandeh, L.F. Campitelli, Y. Yin, et al., Multiparameter functional diversity of human C2H2 zinc finger proteins, Genome Res. 26 (2016) 1742–1752, https://doi.org/10.1101/GR.209643.116/-/DC1.

[31] S.R. Starick, J. Ibn-Salem, M. Jurk, C. Hernandez, M.I. Love, H.R. Chung, et al., ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors, Genome Res. 25 (2015) 825–835, https://doi.org/10.1101/GR.185157.114.

[32] M. Imbeault, P.Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks, Nature 543 (2017) 550–554, https://doi.org/10.1038/nature21683.

[33] S.H. Meijsing, M.A. Pufall, A.Y. So, D.L. Bates, L. Chen, K.R. Yamamoto, DNA binding site sequence directs glucocorticoid receptor structure and activity, Science 324 (2009) 407–410, https://doi.org/10.1126/SCIENCE.1164265/SUPPL_FILE/MEIJSING.SOM.PDF.

[34] K.H. Wong, K. Struhl, The Cyc8–Tup1 complex inhibits transcription primarily by masking the activation domain of the recruiting protein, Genes Dev. 25 (2011) 2525–2539, https://doi.org/10.1101/GAD.179275.111.

[35] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, et al., The human transcription factors, Cell 172 (2018) 650–665, https://doi.org/10.1016/J.CELL.2018.01.029.

[36] S.A. Wolfe, L. Nekludova, C.O. Pabo, DNA recognition by Cys2His2 zinc finger proteins, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 183–212, https://doi.org/10.1146/ANNUREV.BIOPHYS.29.1.183.

[37] S. Inukai, K.H. Kock, M.L. Bulyk, Transcription factor–DNA binding: beyond binding site motifs, Curr. Opin. Genet. Dev. 43 (2017) 110–119, https://doi.org/10.1016/J.GDE.2017.02.007.

[38] S. Hasegawa, H. Nagano, M. Konno, H. Eguchi, A. Tomokuni, Y. Tomimaru, et al., Cyclin G2: a novel independent prognostic marker in pancreatic cancer, Oncol. Lett. 10 (2015) 2986, https://doi.org/10.3892/OL.2015.3667.