8270–8282 *Nucleic Acids Research, 2023, Vol. 51, No. 15* *Published online 6 July 2023*
https://doi.org/10.1093/nar/gkad583

Structural basis for specific DNA sequence motif recognition by the TFAP2 transcription factors

Ke Liu [*,†], **Yuqing Xiao**[†], **Linyao Gan, Weifang Li, Jin Zhang and Jinrong Min** [*]

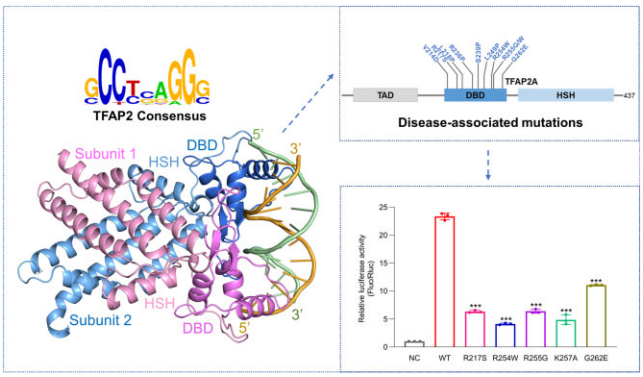
Hubei Key Laboratory of Genetic Regulation and Integrative Biology, School of Life Sciences, Central China Normal University, Wuhan 430079, PR China

Received April 24, 2023; Revised June 01, 2023; Editorial Decision June 20, 2023; Accepted June 27, 2023

ABSTRACT

The TFAP2 family regulates gene expression during differentiation, development, and organogenesis, and includes five homologs in humans. They all possess a highly conserved DNA binding domain (DBD) followed by a helix-span-helix (HSH) domain. The DBD-HSH tandem domain specifically binds to a GCC(N3)GGC consensus sequence, but the precise recognition mechanisms remain unclear. Here, we found that TFAP2 preferred binding to the GCC(N3)GGC sequence, and the pseudo-palindromic GCC and GGC motifs and the length of the central spacer between the two motifs determined their binding specificity. Structural studies revealed that the two flat amphipathic α-helical HSH domains of TFAP2A stacked with each other to form a dimer via hydrophobic interactions, while the stabilized loops from both DBD domains inserted into two neighboring major grooves of the DNA duplex to form base-specific interactions. This specific DNA binding mechanism controlled the length of the central spacer and determined the DNA sequence specificity of TFAP2. Mutations of the TFAP2 proteins are implicated in various diseases. We illustrated that reduction or disruption of the DNA binding ability of the TFAP2 proteins is the primary cause of TFAP2 mutation-associated diseases. Thus, our findings also offer valuable insights into the pathogenesis of disease-associated mutations in TFAP2 proteins.

GRAPHICAL ABSTRACT



INTRODUCTION

The TFAP2 transcription factor, initially identified as an enhancer binder of SV40, plays a crucial role in regulating gene expression and is conserved from prokaryote to eukaryotes (1). In humans, five TFAP2 homologs have been identified, known as TFAP2A-E or AP-2α, AP-2β, AP-2γ, AP-2δ and AP-2ε (2–4). These homologs are believed to have evolved from a single chordate ancestral gene, suggesting that they may share conserved functions (2). The TFAP2 proteins act as either transcriptional repressors or activators during differentiation, development, and organogenesis (5–12). In addition to their critical roles in normal biological functions, mutations and abnormal expression of the TFAP2 proteins have been linked to various diseases (13,14). Thus, the diverse functions of the TFAP2 transcription factors make them a promising area of research for both developmental biology and disease pathology.

The TFAP2 proteins contain a less conserved transactivation domain (TAD) at the N-terminus, and a highly conserved basic DNA binding domain (DBD) followed by a helix-span-helix (HSH) domain at the C-terminus (Supplementary Figure S1A). The HSH domain contributes to the dimerization of the TFAP2 proteins to form either homodimers or heterodimers, which is necessary for the

[*]To whom correspondence should be addressed. Email: minjinrong@ccnu.edu.cn
Correspondence may also be addressed to Ke Liu. Email: keliu2015@mail.ccnu.edu.cn
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

sequence-specific binding of their DBD domains (8,15,16). The DBD domain and the HSH domain share over 95% and 75% sequence similarity among the TFAP2 proteins, respectively (Figure 1A), which explains why both the TFAP2 homodimers and heterodimers bind to a similar GCC(N3)GGC or GCC(N3)GGC-like (N3 = any three nucleotides) consensus sequence (2,8,16–18). This consensus sequence of TFAP2 is found in both promoter and enhancer regions of different TFAP2 target genes (3,8,19–24). For instance, TFAP2A activates the expression of transglutaminase TGM2 that contains a CCC(N3)GGC consensus sequence in its promoter (24). TFAP2 protein also regulates the expression of Insulin-like growth factor binding protein-5 (IGFBP5) by binding to the GCC(N3)GGC motif in its regulatory region (23). TFAP2C binds to the GCC(N3)GGC consensus sequence in the enhancer of human epidermal growth factor receptor-2 (HER2), which reduces the expression of HER2 in breast cancer (19).

TFAP2A is considered the ancestral paralog of the TFAP2 proteins, which is ubiquitously expressed in both vertebrates and invertebrates (2,25,26). TFAP2A plays a critical role in transcriptional regulation through various mechanisms. For example, TFAP2A co-regulates the gene expression of melanocyte differentiation and melanin synthesis with MITF, thereby contributing to normal melanocyte function and melanoma progression (27,28). Additionally, TFAP2A controls the expression of the epidermal growth factor receptor (EGFR), and is involved in maintaining a balance between growth and differentiation in the epidermis (29). Furthermore, by activating BMP7A, TFAP2A inhibits both Fgf and Notch signaling in SAG neural development, contributing to the specification and maturation of neurons (10). As a pioneer factor, TFAP2A can co-occupy active chromatin with other transcription factors such as GATA2 and GATA3, as well as the histone acetyltransferase p300 and the NuRD complex. This facilitates the modulation of chromatin accessibility and gene expression (20,30–32). Notably, TFAP2A knockout in mice results in craniofacial malformations and embryonic lethality, underscoring its importance in development (12).

Given the highly conserved sequence of the DBD-HSH tandem domain between the five TFAP2 homologs, TFAP2A and its TFAP2 paralogs exhibit functional redundancy. For instance, TFAP2B plays redundant roles with TFAP2A in regulating nephric duct morphogenesis (33), retinogenesis (9), sympathetic neuron development (34), craniofacial development (21) and nephron differentiation (35). TFAP2C also exhibits similar expression patterns and redundant functions with TFAP2A during mouse embryogenesis (36), lipid droplet biogenesis (22), neural crest development (37) and epidermis differentiation (38). Moreover, TFAP2E and TFAP2A exhibit redundant functions in melanogenesis (39). In addition to homodimers, TFAP2A/C heterodimers and TFAP2A/B heterodimers respectively mediate neural plate border induction and promote neural crest specification during neural crest development (8). It is believed that these homodimers or heterodimers modulate different functions by recruiting diverse cofactors to target sites. Therefore, a deeper understanding of the molecular mechanisms underlying DNA recognition by the TFAP2 proteins is crucial for understanding how they carry out their diverse transcriptional activities. However, the precise mechanisms by which TFAP2 proteins recognize the consensus binding motif remain unclear.

To advance our comprehension of the diverse transcriptional activity of TFAP2 proteins, in this study, we measured the DNA binding specificities of TFAP2A and TFAP2B, and determined the crystal structures of the DBD-HSH tandem domains of TFAP2A and TFAP2B in their apo form and the TFAP2A-DNA complexes to demonstrate the DNA recognition mechanism of the TFAP2A DBD-HSH tandem domain. We further evaluated the DNA binding affinity and transcriptional activation of wild-type TFAP2A and its disease-associated mutants by ITC and luciferase reporter assays. Thus, our findings not only provide insights into the molecular basis of the DNA recognition by TFAP2A, but also offer valuable insights into the pathogenesis of disease-associated mutations of TFAP2 proteins.

MATERIALS AND METHODS

Cloning, expression and purification

Human TFAP2A (aa 203–420 and aa 279–411) and TFAP2B (aa 219–457) fragments were cloned into the pET28-MKH8-SUMO vector to generate N-terminal His_6 and Sumo-tagged fusion proteins with a tobacco etch virus (TEV) cleavage site. Mutants of TFAP2A and TFAP2B were constructed using site-directed mutagenesis and confirmed by sequencing. All of the recombinant plasmids were overexpressed using *E. coli* BL21 (DE3) under the induction with 0.5 mM IPTG at 14°C overnight. The cells were harvested and then resuspended in a lysis buffer with 500 mM NaCl, 20 mM Tris (pH 7.5) and 5% glycerol, followed by sonication at 4°C. After centrifugation at 16 000 g, the supernatant was collected and further purified using the Ni-NTA resin (Qiagen). The recombinant protein was eluted and treated with TEV protease to remove the His_6 and Sumo-tag followed by affinity chromatography, anion-exchange chromatography, and gel-filtration column chromatography (GE Healthcare). The final purified wild-type and mutant proteins of TFAP2A and TFAP2B were stored in a buffer containing 20 mM Tris (pH 7.5) and 150 mM NaCl. For ITC binding assay, TFAP2A and TFAP2B proteins were stored in a buffer containing 20 mM Tris (pH 7.5) and 200 mM NaCl.

Isothermal titration calorimetry (ITC) assay

The DNA oligonucleotides in this study were obtained from General Biosystems Co. Ltd. (Anhui), and then annealed to DNA duplexes in a buffer containing 20 mM Tris (pH 7.5) and 200 mM NaCl. In the ITC binding assays, the final concentrations of proteins and DNAs were in the range from 7 to 30 μM and 100–350 mM, respectively. The ITC assays were carried out using MicroCal PEAQ-ITC (Malven) at 25°C. Each titration consisted of 19 injections, in which the first injection was set at 0.4 μl and the following injections were set at 2 μl. The dissociation constant (K_d) was determined by fitting the integrated titration data using 'One Set of Sites' fitting model by a nonlinear least-squares method implemented in MicroCal ITC200 analysis software Origin

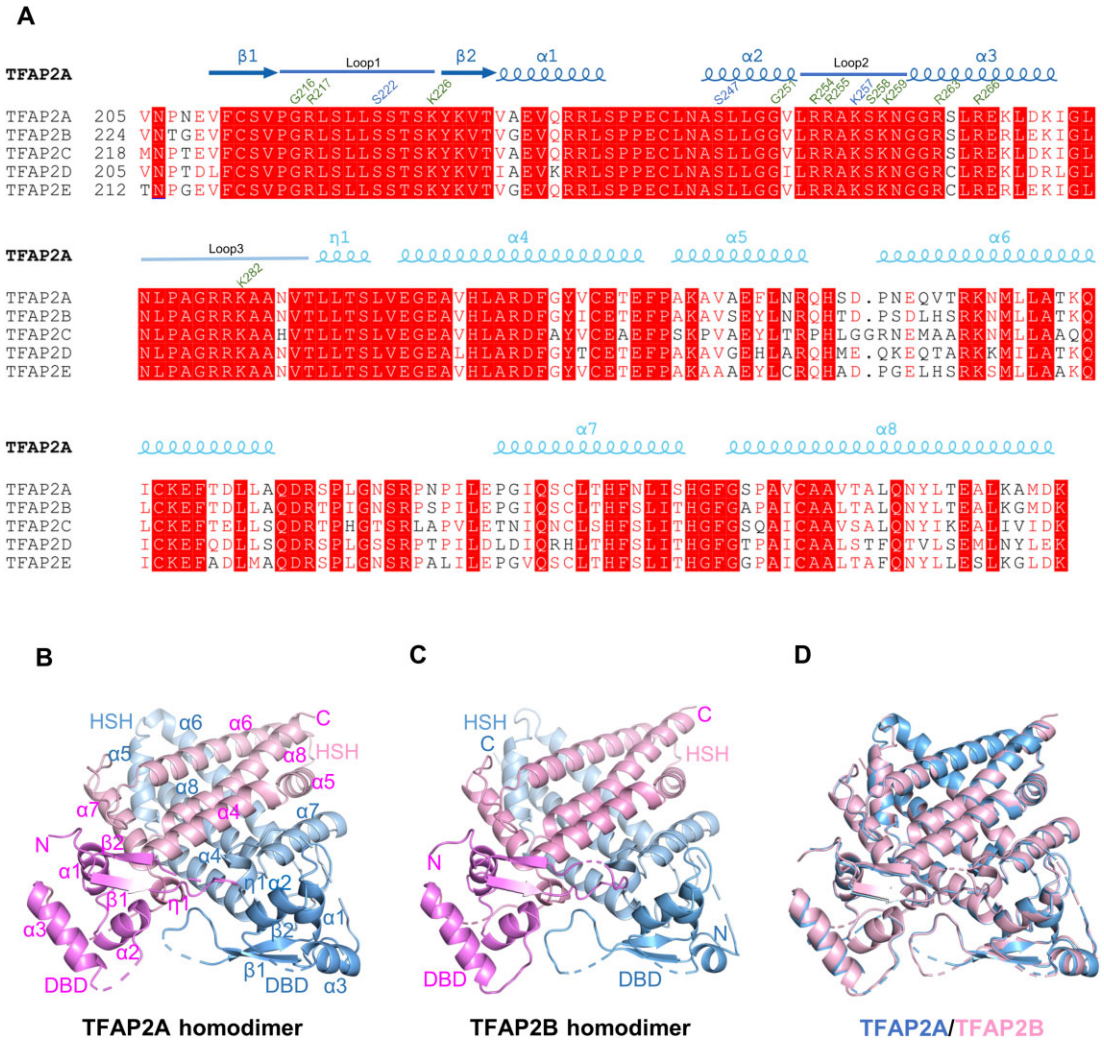


Figure 1. Overall structure of the TFAP2 DBD-HSH tandem domain. (**A**) Sequence alignment of DBD-HSH tandem domains of human TFAP2 proteins. TFAP2A (UniProt: P05549), TFAP2B (UniProt: Q92481), TFAP2C (UniProt: Q92754), TFAP2D (UniProt: Q7Z6R9) and TFAP2E (UniProt: Q6VUC0). The secondary structure elements of human TFAP2A are indicated above with colors corresponding to Supplementary Figure S1A. The DBD domain is colored deep blue, while the HSH domain and the loop linking both domains are colored blue. Identical residues are marked by red background, the base and backbone interacting residues of TFAP2A are numbered in blue and green above the sequence, respectively. (**B**) Overall structure of the TFAP2A DBD-HSH tandem domain. The two TFAP2A subunits are colored pink and blue, respectively, wherein the N-terminal DBD domains from both subunits are highlighted in dark pink and deep blue, respectively. (**C**) Overall structure of the TFAP2B DBD-HSH tandem domain. The two TFAP2B subunits are shown in the same way as TFAP2A in (B). (**D**) Superposition of the structures of the DBD-HSH tandem domains of TFAP2A (blue) and TFAP2B (pink) in a cartoon representation.

7.0 (Malven). The ITC assays for samples in Table 1 were conducted with three replicates ($n = 3$), and the errors represent the standard errors from the three replicates. For the ITC binding assays of the mutant samples, the standard errors are the fitting errors of the ITC titration curves if the binding is detectable.

Differential scanning fluorometry (DSF) assay

In this study, protein stability was assessed with 0.4 mg/ml of protein in a buffer of 20 mM Tris (pH 7.5) and 150 mM NaCl by different scanning fluorometry (DSF) (40). The fluorescent dye SYPRO Orange (Sigma) was used at a final concentration of 8 × . A total volume 10 μl of protein sample was used to measure the melting curves, and the temperature was gradually increased from 25°C to 90°C at a

Table 1. Binding affinities of the DBD-HSH tandem domains of TFAP2A (aa 203-420) and TFAP2B (aa 219-457) to different DNA sequences

DNA Sequences	TFAP2A K_d (nM)	TFAP2B K_d (nM)
5′-GT**GCC**CGA**GGC**AG-3′[a]	19 ± 1	18 ± 2
5′-GT**GCC**TGA**GGC**AG-3′[a]	17 ± 1	19 ± 2
5′-GT**GCC**GA**GGC**AG-3′	WB	WB
5′-GT**GCC**GAT**GGC**AG-3′	341 ± 20	340 ± 17
5′-T**GCC**CCCGG**GGC**A-3′	WB	WB
5′-GT**GmCC**CGA**GGC**AG-3′	347 ± 3	299 ± 22
5′-GT**GCmC**CGA**GGC**AG-3′	107 ± 3	126 ± 12

[a]DNA sequence used for crystallization in this study. WB: weak binding.

rate of 2°C/min using the real-time PCR instrument (Light Cycler 480, Roche, Switzerland). The excitation wavelength was 465 nm and the emission wavelength was 580 nm. Melting curves were analyzed using GraphPad Prism 9 and the Tm values for each sample were calculated based on three independent experiments.

Size-exclusion chromatography (SEC) analysis

We carried out the size-exclusion chromatography (SEC) analysis to analyze the oligomeric state of TFAP2A. The DBD-HSH tandem domain (aa 203–420) and HSH domain (aa 279–411) proteins of TFAP2A were analyzed by size-exclusion chromatography using Superdex 75 10/300 GL (GE Healthcare) with a buffer containing 20 mM Tris (pH 7.5), 150 mM NaCl and 1 mM DTT.

Protein crystallization

Crystallization was carried out using the sitting drop vapor diffusion method at 18°C by mixing 0.6 μl samples and 0.6 μl reservoir solution. Crystals of the TFAP2A (15 mg/ml) in apo form were obtained under conditions with 0.2 M sodium tartrate dibasic dihydrate and 20% PEG 3350 (w/v). Crystals of TFAP2B (15 mg/ml) in apo form were grown in a reservoir solution containing 0.2 M Sodium citrate tribasic dihydrate and 20% PEG 3350 (w/v). To obtain crystals of the TFAP2A-DNA complex, the TFAP2A protein (8 mg/ml) was mixed with different DNA oligonucleotides at a molar ratio of 1:1.2, and both complex crystals were grown in a reservoir solution containing 0.1 M citric acid (pH 3.5) and 34% PEG 200 (w/v).

Data collection and structure determination

Prior to diffraction data collection, the crystals were protected in a cryoprotectant consisting of their respective crystallization condition supplemented with 20–25% (v/v) glycol or ethylene glycol, and the protected crystals were then flash-frozen in liquid nitrogen. Diffraction data of the TFAP2A and TFAP2B in apo form were collected at SSRF 02U1 beamline at 100 K, respectively, while the diffraction data of the TFAP2A-CGA DNA complex were collected at SSRF 17B beamline at 100 K, TFAP2A-TGA DNA complex were collected at SSRF 19U beamline at 100 K. All the data were then processed with the HKL 2000 suite (41), XDS (42) and CCP4 (43). The structure of the TFAP2A in apo form was solved by the molecular replacement with the program PHASER (44) using the TFAP2A structural model from the AlphaFold2 Protein Structure Database (45,46) as the search model. The TFAP2B and TFAP2A-DNA structures were solved by the molecular replacement with the program PHASER using the TFAP2A apo form as the search model. Model building was performed with Coot (47), and structure refinement was performed with REFMAC (Version 5.8.0257) (48). The crystal data collection and refinement statistics were summarized in Supplementary Table S1.

Cell culture and western blot

The HEK293T cells were maintained in DMEM (Dulbecco's modified Eagle's medium) high-glucose media (Cytiva) containing 10% fetal bovine serum (Hyclone) and 0.1 × Penicillin-Streptomycin (Hyclone) at 37°C in 5% CO_2 atmosphere. HEK293T cells (1×10^6 cells/well) cultured in 6-well plates were transfected with 1 μg pCDNA3.1-TFAP2A (encoding Flag- and His_{10}-tagged full-length TFAP2A wild-type or its mutants) with Lipofectamine 2000 (Invitrogen). 24 h after transfection, the cell was washed with 1× PBS, followed by lysis and SDS-PAGE analysis. After transferring to a PVDF membrane, the membrane was blocked by 5% milk in PBST for 1 h at room temperature, and then incubated with 1 μl primary antibody (anti-Flag, Cat no.: 66008-4-Ig, proteintech; anti-GAPDH, Cat no.: 60004-1-Ig, proteintech) overnight at 4°C. After washing the membrane with TBST three times, HRP-conjugated antibody (Cat no.: SA00001-1, proteintech) was used for the secondary antibody. After washing by TBST, membrane was incubated with 1 ml ECL reagent (Meilunbio) and exposed to Multi color fluorescent gel imaging system (FluorChem R).

Luciferase activity analysis

The promoter sequence from −537 to −1 of the *IGFBP5* was predicted to contain one TFAP2A binding motif (23). We then subcloned this *IGFBP5*-promoter sequence into pGL4.20 vector. Meanwhile, according to our ITC binding assay, we designed the *IGFBP5*-promoter region containing different TFAP2A binding motif mutants. HEK293T cultured in 24-well plates (2.5×10^5 cells/well) were co-transfected with pCDNA3.1-TFAP2A or its mutants (pCDNA3.1 as control vector) (400 ng), pGL4.20-*IGFBP5*-promoter or its mutants (pGL4.20 as control vector) (100 ng), and pRL-TK (5 ng, *Renilla* luciferase as an internal control) using Lipofectamine 2000 (Invitrogen). 24 h after transfection, cells were washed using 1 × PBS and lysed for 15 min using passive lysis buffer (Promega). Luciferase activity was measured with the Dual-Luciferase reporter assay system (Promega) according to the manufacturer's instructions using GloMax 20/20 Luminometer (Promega). Firefly luciferase activity was normalized based on the Renilla activity for each sample. Error bars represent SD from three replicates ($n = 3$), which were calculated using one-way ANOVA by GraphPad Prism 9.

RESULTS AND DISCUSSION

Overall structures of the DBD-HSH tandem domains of TFAP2A and TFAP2B

The TFAP2 transcription factors have been extensively studied in different species. However, the molecular mechanisms underlying their binding to the GCC(N3)GGC consensus DNA sequence remain unclear. To this end, we first determined the crystal structures of the DBD-HSH tandem domains of TFAP2A (aa 203–420) and TFAP2B (aa 219–457) (Supplementary Table S1). Superposition of our apo form structures of TFAP2A and TFAP2B with that predicted by alphafold2 showed that they overlaid well with a root mean square deviation (RMSD) of 0.70 Å (over 168 Cα atoms of TFAP2A) or 0.67 Å (over 179 Cα atoms of TFAP2B), respectively, confirming the protein structure prediction power of Alphafold (45). As expected, both

TFAP2A and TFAP2B formed homodimers via their HSH domains, and the dimerization surface of the HSH domain is rich in hydrophobic residues (Figure 1B and C), which was further confirmed by our gel filtration chromatography results (Supplementary Figure S1B). Superposition of the TFAP2A and TFAP2B structures showed that these two structures superimposed well with a RMSD of 0.26 Å over their 160 Cα atoms (Figure 1D), suggesting that the TFAP2 proteins share a highly conserved structure, consistent with their high sequence similarity. In the following, we will use the structure of the TFAP2A DBD-HSH tandem domain to describe their structural features.

The DBD-HSH structure of TFAP2A adopted an α-helical architecture. The DBD domain (aa 202–273) consisted of two short antiparallel β-strands (β1 and β2) followed by three α-helices (α1–α3). The HSH domain (aa 293–422) was of a flat amphipathic α-helical fold made up of α-helices α4-α8. The two domains were linked by a long loop containing a 3_{10}-helix (Figure 1A and B). The two flat α-helical HSH domains stacked with each other to form a dimer via hydrophobic interactions in a reverse orientation, with their α-helices nearly perpendicular to each other (Figure 1B). Mutating the hydrophobic residues of the dimerization surface, such as V307, F379, V391 and L398, to the polar aspartic acid resulted in insoluble proteins of these TFAP2A mutants, implying the importance of the hydrophobic interaction in the dimerization of the TFAP2A and its stability. The two DBD domains of the dimer looked like a pair of levers of pliers joined at the dimerized HSH domains and formed a wide-open positively charged binding groove that could potentially recognize double-stranded DNA. However, we found that three loops, namely, Loop1 (linking β1 and β2), Loop2 (connecting α2 and α3), and Loop3 (linking α3 and η1), were disordered in the TFAP2A apo form structure (Figure 1B), and the disordered loops were also found in the TFAP2B apo form structure (Figure 1C and D).

DBD-HSH tandem domain of TFAP2 specifically binds to a consensus DNA sequence of GCC(N3)GGC

The TFAP2 proteins have been shown to specifically recognize a consensus DNA sequence of GCC(N3)GGC *in vivo* and *in vitro* (21–23,49). To confirm the DNA binding sequence specificity of the TFAP2 proteins quantitatively, we performed isothermal titration calorimetry (ITC) binding assays using two 13-mer DNA oligonucleotides with the sequences of GTGCCCGAGGCAG and GTGCCTGAGGCAG, which were designed based on the previously reported ChIP-seq data (Figure 2A) (18). Our ITC results showed that the DBD-HSH tandem domains of TFAP2A and TFAP2B displayed similar binding affinities to GTGCCCGAGGCAG DNA with a K_d of ∼19 and 18 nM, respectively (Figure 2B, Table 1). TFAP2A and TFAP2B also exhibited comparable binding affinity to the GTGCCTGAGGCAG DNA, meaning that the sequence identity of the central spacer did not affect the binding affinities (Table 1, Supplementary Figure S2).

In addition to the 3bp central spacer between the pseudo-palindromic GCC and GGC motifs, several reports have suggested that DNA sequences with the other number of nucleotides as the central spacer could also bind to TFAP2 (1,49,50). To explore the effect of the number of nucleotides in the central spacer on TFAP2 binding, we tested DNA sequences with different spacer lengths using ITC. Our results showed that TFAP2A bound to the 4bp-spacer DNA of GTGCCCGATGGCAG with a K_d of ∼341 nM, which is ∼18-fold weaker than that of the 3bp-spacer containing DNA, and displayed barely detectable binding to DNA with a 2 bp- or 5 bp-spacer between the GCC and GGC motifs (Table 1, Supplementary Figure S2). Thus, our findings suggest that the length of the central spacer, but not its sequence identity, determined the DNA binding ability of TFAP2A to the pseudo-palindromic GCC(N3)GGC sequence.

Moreover, our ITC assays revealed that the DBD-HSH tandem domain of TFAP2B exhibited a similar DNA binding selectivity to that of TFAP2A (Table 1, Supplementary Figure S2), which supports the previous report that TFAP2 proteins share similar DNA binding specificity (2). Overall, our quantitative analyses provide further evidence of the specific recognition of the GCC(N3)GGC motif by TFAP2 proteins and shed light on the importance of the central spacer length in DNA binding.

Complex structures of TFAP2 with the GCC(N3)GGC DNA

To understand why the DBD-HSH tandem domain of TFAP2 selectively bound to the GCC(N3)GGC motif, crystal structures of TFAP2A bound to two different DNA sequences were determined, one with a central CGA and the other with a central TGA trinucleotide in the GTGCC(N3)GGCAG sequence (Figure 2C–E, Supplementary Figure S3A–C and Supplementary Table S1). The DNA binding in both complex structures was almost identical (Figure 2C–E, Supplementary Figure S3A–C). Moreover, the TFAP2A structures in the complexes were almost identical to its structure in the apo form, except that the Loop1, Loop2 and Loop3 became ordered when bound to the DNA duplex (Figure 2C, Supplementary Figure S3A).

In the complex structures, the TFAP2A dimer used its two positively charged DBD domains to bind a single DNA duplex, while the HSH domain was not directly involved in DNA binding (Figure 2C–E, Supplementary Figure S3A–C). This finding is consistent with previous reports that the HSH domain lacks the DNA binding ability (Supplementary Figure S1C) (15). Without the DBD domain, the HSH domain could still form a dimer (Supplementary Figure S1B), meaning that the HSH domain was only responsible for dimerization, bringing the two DBD domains together to hold the DNA duplex firmly. To explore whether the DNA binding mode is novel, we searched the DALI and FATCAT servers. We found that the closest structure to TFAP2A is the transcriptional regulator phenolic acid decarboxylase regulator (PadR) of *Bacillus subtilis* (*Z*-score 5.3) (51,52) and Max bHLH domain (*P*-value 1.46 e-4) (53,54). However, their dimerization and DNA recognition mode are different from the TFAP2 family, indicating that the DNA binding by the DBD-HSH tandem domain of TFAP2A is novel.

In the TFAP2A-DNA complex structures, the two DBD domains bound to the DNA duplex in a 2-fold symmetry-

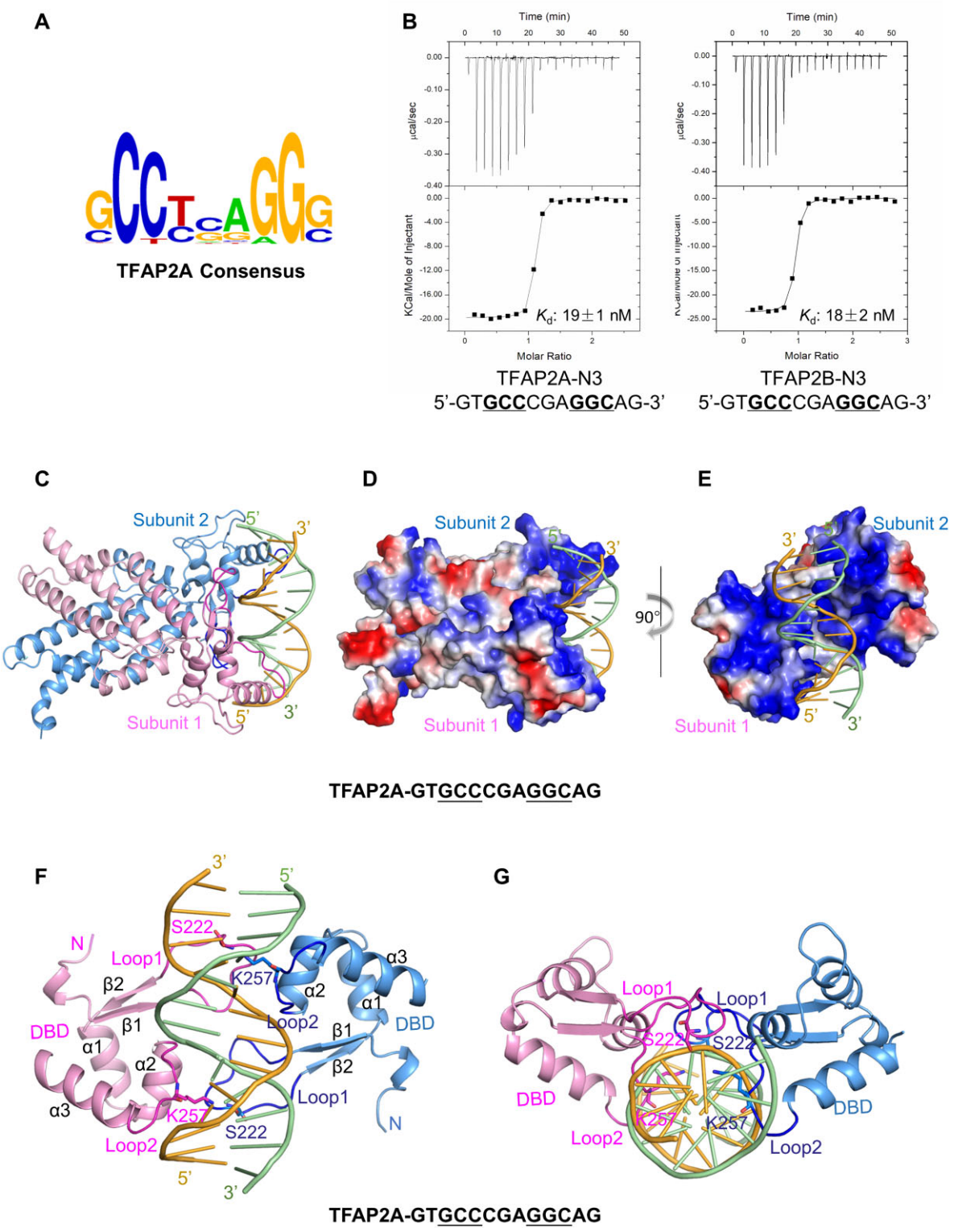
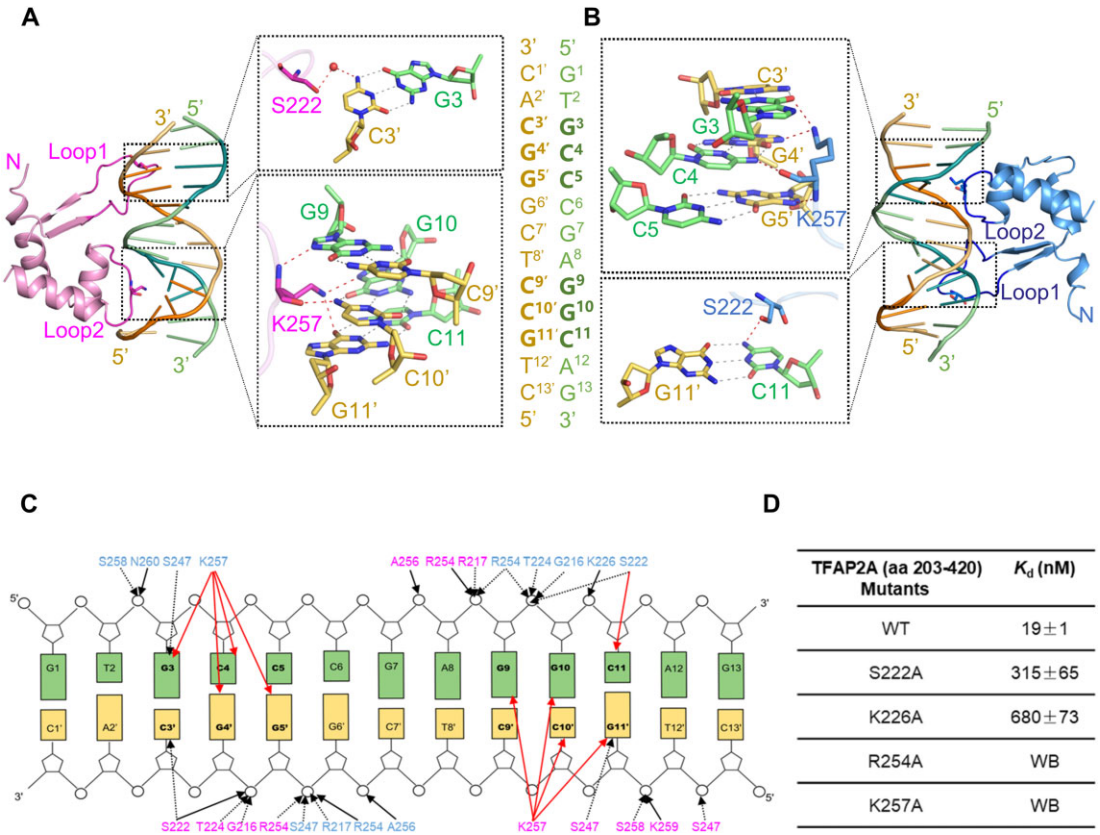


Figure 2. Crystal structure of the TFAP2A DBD-HSH tandem domain bound to DNA of the GCC(N3)GGC consensus sequence. (**A**) Predominant DNA binding motif of the TFAP2A DBD-HSH tandem domain. The predominant motif is derived from https://jaspar.genereg.net/ with the ID MA0003.2. (**B**) ITC curves of the DBD-HSH tandem domains of TFAP2A and TFAP2B bound to a GCC(N3)GGC sequence containing DNA. Only one strand of the DNA duplex is shown. (**C**) Overall structure of the TFAP2A DBD-HSH tandem domain in complex with the GTGCCCGAGGCAG DNA in a cartoon representation. DNA is shown in cartoon mode with one strand colored in yellow and the other one in green. (**D**) Electrostatic surface potential of the TFAP2A DBD-HSH tandem domain in complex with the GTGCCCGAGGCAG DNA viewed in the same orientation as (C). (**E**) Electrostatic surface potential view of (D) after 90° clockwise rotation. (**F**) Structure of TFAP2A bound to the DNA duplex through the Loop1 and Loop2 of both DBD domains. The DBD domain and DNA duplex are shown in the same way as (C). (**G**) Structure of TFAP2A bound to the DNA duplex through DBD domains viewed from the top of (F).



TFAP2A (aa 203-420) Mutants	K_d (nM)
WT	19±1
S222A	315±65
K226A	680±73
R254A	WB
K257A	WB

Figure 3. Structural basis for preferential recognition of the GCC(N3)GGC DNA by the TFAP2A DBD-HSH tandem domain. (**A** and **B**) Detailed base interactions between the TFAP2A DBD domain and the GTGCCCGAGGCAG DNA. The protein residues and DNA bases are shown in stick models, and detailed base interactions are also shown in the black zoom-in boxes. Hydrogen bonds formed between protein and DNA are marked as red dashed lines, while hydrogen bonds between DNA bases are marked as gray dashed lines. (**C**) Schematic diagram of the TFAP2A DBD-HSH tandem domain bound to the GTGCCCGAGGCAG DNA. The conserved GCC and GGC motifs are highlighted in black fonts. The direct base and phosphate group interactions are marked as red and black solid arrows, respectively, while the water-mediated hydrogen bonds are indicated by black dashed arrows. (**D**) Binding affinities of the TFAP2A mutants to the GTGCCCGAGGCAG DNA.

related fashion. Positively charged residues from both DBD domains, including R217, K226, R254, R255, K259, R263, R266 and K282, formed a positively charged U-shaped surface to bind the DNA backbone via electrostatic interactions (Figure 2D and E, Supplementary Figure S3B-D). Furthermore, Loop1 and Loop2 of one DBD domain inserted into two neighboring major grooves of the DNA duplex from one side, and Loop1 and Loop2 of the other DBD domain inserted into the neighboring major grooves of the DNA duplex from almost the opposite side (Figure 2F and G). The side chain of K257 in Loop2 from subunit 1 inserted into the major groove and formed two hydrogen bonds with the bases of G10 and G11′, and its main chain amine and carbonyl groups formed hydrogen bonds with the bases of G9 and C10′, respectively (Figure 3A and C). This binding mode is also conserved in subunit 2, wherein K257 bound to G3C4C5/C3′G4′G5′ trinucleotide base pairs in a similar manner (Figure 3B and C). In addition, the residue S222 on Loop1 in both subunits recognized C11/G11′ and G3/C3′ base pairs via direct hydrogen bond or water-mediated interactions (Figure 3A–C), and S247 on α2 in both subunits formed water-mediated interactions with C11/G11′ and G3/C3′ base pairs (Figure 3C). Except for the two pseudo-palindromic GCC and GGC motifs, no other base-mediated interactions were observed between TFAP2A and DNA. Hence, K257, S222 and S247 were the major structural determinants in binding the pseudo-palindromic GCC and GGC motifs in the consensus GCC(N3)GGC sequence.

In addition to the base recognitions, the side chains of R254 from both subunits pointed to the DNA minor groove and formed salt bridges with the phosphate backbone of G9 and G5′, respectively (Figure 3C). Residues G216, S222, T224 and R254 from subunit 1, and R217, S247, R254, A256, S258 and N260 from subunit 2 formed extensive hydrogen bonding interactions with the backbones of the G3C4C5/C3′G4′G5′ motif as well as its franking nucleotides, and the interactions were mostly conserved for the G9G10C11/C9′C10′G11′ half-site as well (Figure 3C).

Structural basis for specific binding of the GCC(N3)GGC consensus sequence by TFAP2

In order to investigate the sequence tolerance of the pseudo-palindromic GCC and GGC motifs in the GCC(N3)GGC consensus sequence, we designed a series of DNA sequence mutants based on the GTGCCCGAGGCAG sequence.

Table 2. Binding affinities of the DBD-HSH tandem domains of TFAP2A (aa 203–420) and TFAP2B (aa 219–457) to different DNA variants

DNA sequences	TFAP2A K_d (nM)	TFAP2B K_d (nM)
5′-GT**GCC**CGA**GGC**AG-3′	19 ± 1	18 ± 2
5′-GT**c**C**C**CGA**GGC**AG-3′	78 ± 16	55 ± 7
5′-GT**a**C**C**CGA**GGC**AG-3′	373 ± 46	313 ± 69
5′-GT**t**C**C**CGA**GGC**AG-3′	565 ± 87	248 ± 49
5′-GT**c**C**C**CGA**GG**gAG-3′	336 ± 53	355 ± 57
5′-GT**a**C**C**CGA**GGt**AG-3′	WB	WB
5′-GT**t**C**C**CGA**GGa**AG-3′	WB	WB
5′-GT**Ga**CCGA**GGC**AG-3′	WB	WB
5′-GT**Gt**CCGA**GGC**AG-3′	WB	WB
5′-GT**Gg**CCGA**GGC**AG-3′	WB	WB
5′-GT**GCa**CGA**GGC**AG-3′	WB	WB
5′-GT**GCg**CGA**GGC**AG-3′	WB	WB
5′-GT**GCt**CGA**GGC**AG-3′	694 ± 71	461 ± 71
5′-GT**GCt**CGA**aGC**AG-3′	WB	WB
5′-GT**atg**CGA**GGC**AG-3′	WB	WB

Only one strand of the DNA duplex is shown in the table, and the mutated nucleotides are shown in lower case. WB: weak binding.

Our ITC results showed that substituting G in the GCC motif with C, A or T reduced its binding affinities by ∼4, ∼19 or ∼30-fold, respectively (Table 2, Supplementary Figure S4). When we simultaneously mutated G in the GCC motif and C in the pseudo-palindromic GGC motif, the binding was further reduced (Table 2, Supplementary Figure S4). Our structures explained why nucleotide substitution of the GTGCCCGAGGCAG sequence weakened its binding to TFAP2A. When G3 in the G3C4C5 motif was mutated to C, T or A, some weak interactions were observed, however, the hydrogen bonding interactions between the side chain of K257 and the G3 were disrupted (Supplementary Figure S5A–C).

When we mutated the second nucleotide in the GCC motif, we found that any substitution almost completely disrupted its binding to TFAP2A (Table 2, Supplementary Figure S4), which is consistent with a previous report that the second C/G base pair of the GCC/GGC could not tolerate any substitution (49). Structural analysis revealed that when the second nucleotide in the G3C4C5 trinucleotide was replaced by A, G or T, there was steric clash between K257 and the new base, aside from the disrupted hydrogen bonding interactions between the K257 and C4/G4′ base pair (Supplementary Figure S5D–F).

When we mutated the third nucleotide in the GCC motif, the mutations severely diminished its binding to TFAP2A, except for the GCt mutant, which still displayed a modest binding affinity of ∼694 nM. However, when its pseudo-palindromic position was also mutated, the DNA duplex almost lost binding to TFAP2A (Table 2, Supplementary Figure S4). Our structural modeling showed that when mutating C5 in the G3C4C5 motif to A or G, T5′ or C5′ that is complementary with A5 or G5 either clashed or lost the hydrogen bond with the amino group of K257 (Supplementary Figure S5H and I), while when mutating the third C5 to T5, the amino group of K257 and the N7 group of A5′ that is complementary with T5 could still form a hydrogen bond to stabilize the complex (Supplementary Figure S5J). Our ITC binding studies also showed that TFAP2B displayed a similar binding specificity as TFAP2A (Table 2, Supplementary Figure S6), further confirming that TFAP2B has the same preference and binding mode for GCC(N3)GGC motif DNA as TFAP2A.

Effect of cytosine methylation of the GCC(N3)GGC consensus sequence on its binding to TFAP2

TFAP2C is a member of the TFAP2 protein family and has been reported to regulate the expression of estrogen receptor-α (ERα) and human epidermal growth factor receptor-2 (HER2) in breast cancers by binding to the GCC(N3)GGC consensus sequence of their promoters. However, the binding is disrupted when the promoters are methylated (19,55). To investigate whether DNA methylation of the consensus sequence directly inhibited its binding to the TFAP2 proteins, we synthesized three methylated DNA sequences (GTGmCCCGAGGCAG, GTGCmCCGAGGCAG and GTGCmCGGAGGCAG) and measured their binding ability to TFAP2A and TFAP2B. Our ITC results showed that the GmCC or GCmC methylated DNA bound to both TFAP2 proteins, but with a reduced binding affinity (Table 1, Supplementary Figure S2, S4 and S6), which is consistent with a previous report that TFAP2A could bind to the cytosine-methylated DNA *in vitro* (56). Structural analysis revealed that methylation of C4 caused a mild steric clash with the carbonyl group of K257 (Supplementary Figure S5G), while methylation of C5 did not significantly affect its binding with TFAP2A, which was reflected by its slightly reduced binding affinity to TFAP2A (Supplementary Figure S5K).

Validation of key residues of TFAP2 in sequence-specific binding to the GCC(N3)GGC motif

To illustrate the key residues of TFAP2A in DNA binding, we made several TFAP2A point mutants and examined their DNA binding ability by ITC assays. Our ITC results showed that mutating S222 or K226 of TFAP2A to alanine reduced their binding affinity by ∼16- or 35-fold when compared to the wild-type TFAP2A, while the R254A and K257A mutants exhibited negligible or very weak binding (Figure 3D, Supplementary Figure S7). Since these DNA interacting residues are highly conserved among the TFAP2 homologs (Figure 1A), we speculated that all TFAP2 proteins might share the same DNA binding mode.

Our structures also explained why the TFAP2A dimer preferred a 3bp central spacer between the pseudo-palindromic GCC and GGC motifs. We observed that the conformations of Loop1 and Loop2 in both subunits were stabilized by surrounding residues via hydrogen bonding interactions in the complexes. Only the 3bp central spacer allowed optimal insertion of the two loops into the major grooves of the DNA duplex to form base-specific interactions. Therefore, the fixed distance between Loop1 and Loop2 served as a ruler to control the length of the central spacer (Supplementary Figure S3E).

Table 3. Binding affinities of the TFAP2A (aa 203–420) and TFAP2B (aa 219–457) disease-associated mutants to GCC(N3)GGC DNA

	Mutants	K_d (nM)
TFAP2A	WT	19 ± 1
	R217S	ND
	R254W	WB
	R255G	WB
	G262E	129 ± 22
TFAP2B	WT	18 ± 2
	R236C	ND
	A275D	WB
	K276R	382 ± 36
	L284S	ND
	R285Q	178 ± 43
	R300C	88 ± 20
	V336I	44 ± 9

The DNA sequence of one strand of the DNA duplex is GTGCCCGAGGCAG. ND: no detectable binding; WB: weak binding.

Validations of the disease-associated mutations of TFAP2A and TFAP2B in DNA binding

Mutations of the TFAP2 proteins are implicated in various diseases, including cancers (57–59). For example, mutations in TFAP2A have been found in individuals with Non-syndromic cleft lip with palate (NSCLP) or Branchio-oculo-facial syndrome (BOFS). These mutations include V214D, R217S, L218P, R236P, S239P, L249P, R254W, R255G/W and G262E (60–62). Mutations in TFAP2B, including P62R, G217A, R236C/S, A275D, K276R, L284S, R285Q, R300C, T306M, V336I and R382X, can cause patent ductus arteriosus (PDA), Char syndrome, and Syndromic craniosynostosis (63–65). To explore how these mutations in TFAP2A and TFAP2B cause diseases, we made these disease-associated mutants of TFAP2A and TFAP2B to measure their DNA binding ability. We successfully obtained soluble and stable proteins for most mutants, except the V214D, L218P, R236P, S239P and L249P mutants of TFAP2A. Our ITC results showed that the TFAP2A R217S, R254W and R255G mutants exhibited significantly reduced or no detectable DNA binding ability, while the G262E mutant had a ∼7-fold reduced binding affinity compared to the wild-type TFAP2A (Table 3, Supplementary Figure S8). Among the TFAP2B mutants, the R236C, A275D and L284S mutants lost DNA binding ability, while the K276R, R285Q, R300C and V336I mutants had a DNA binding affinity reduced by ∼2- to 21-fold compared to the wild-type TFAP2B (Table 3, Supplementary Figure S8).

We then mapped the disease-associated mutations in TFAP2A and TFAP2B, and found that, except for the R300C, T306M and V336I mutations that locate on the Loop3 or dimerization surface of TFAP2B, the vast majority of these mutations are located in the DBD domains of TFAP2A and TFAP2B (Figure 4A, Supplementary Figure S9A). Our structural analysis revealed that mutations such as R254W, R255G on Loop2, and R217S on Loop1 of TFAP2A, disrupted their interactions with the backbone or destabilized the conformation of the DBD domain (Supplementary Figure S3D and E). Similarly, A275D and K276R mutations in TFAP2B led to electrostatic repulsion or steric clash with the DNA duplex, while the R236 mutation on Loop2 of TFAP2B caused protein instability when mutated to cysteine (Figure 1A). Additionally, although L284 did not directly interact with DNA, mutating L284 to S on the α3 of TFAP2B resulted in protein instability. The poor protein stability of TFAP2A and TFAP2B caused by the R217S, R236C and L284S mutations was further supported by our differential scanning fluorimetry (DSF) assay, which exhibited lower melting temperatures when compared with the wild-type TFAP2A or TFAP2B protein (Supplementary Figure S9B and C). Based on our binding studies and structural analysis, we concluded that the identified disease-associated mutations in TFAP2A and TFAP2B mainly lead to a decrease or disruption of their DNA binding ability, either directly or indirectly, resulting in various diseases.

Transcriptional activity of TFAP2A requires a fully functional DNA binding domain

To assess whether disease-associated mutations of TFAP2A affect its ability to activate transcription, we conducted luciferase reporter analysis to evaluate the transcriptional activity of the TFAP2A disease-associated mutants using a pGL4.20-*IGFBP5*-promoter plasmid with one TFAP2A binding site upstream of the Firefly luciferase gene in HEK293T cells (Figure 4B). Before luciferase reporter analysis, we measured the protein expression levels of TFAP2A WT and its variants, and observed that all of the TFAP2A variants exhibited comparable expression levels to that of the TFAP2A WT (Figure 4C). Subsequently, we performed luciferase reporter analysis and found that all TFAP2A disease variants exhibited a ∼2- or 5-fold decrease in luciferase activity compared to the wild-type TFAP2A (Figure 4D), which is consistent with our ITC binding results (Table 3).

To investigate the importance of the TFAP2A binding motif in transcriptional activation, we mutated the TFAP2A binding site in the pGL4.20-*IGFBP5*-promoter, and conducted luciferase reporter analysis with TFAP2A. We found that replacing the 3bp central spacer of the GCC(N3)GGC motif with a 4bp-spacer led to a ∼2-fold decrease in transcriptional activation, while using a 2bp-spacer containing pGL4.20-*IGFBP5*-promoter reduced transcriptional activation by ∼5-fold compared to the wild-type pGL4.20-*IGFBP5*-promoter (Figure 4E). Similarly, mutations at the second nucleotide in the GCC motif, or the third nucleotide of the GCC motif and its symmetry-related position, or substitution of the GCC motif with atg, also resulted in a ∼6-fold decrease in transcriptional activation compared to the wild-type pGL4.20-*IGFBP5*-promoter (Figure 4E). These findings are in line with our ITC DNA binding data using the mutated DNA binding sequences (Table 2, Supplementary Figure S4). Therefore, our luciferase reporter analysis demonstrated that the decrease or disruption of the DNA binding ability of TFAP2A is the primary cause of TFAP2A mutation-mediated diseases.

CONCLUSIONS

The TFAP2 family plays essential roles in regulating a variety of biological processes, and its dysfunction is associated with several human diseases and oncogenesis, such as

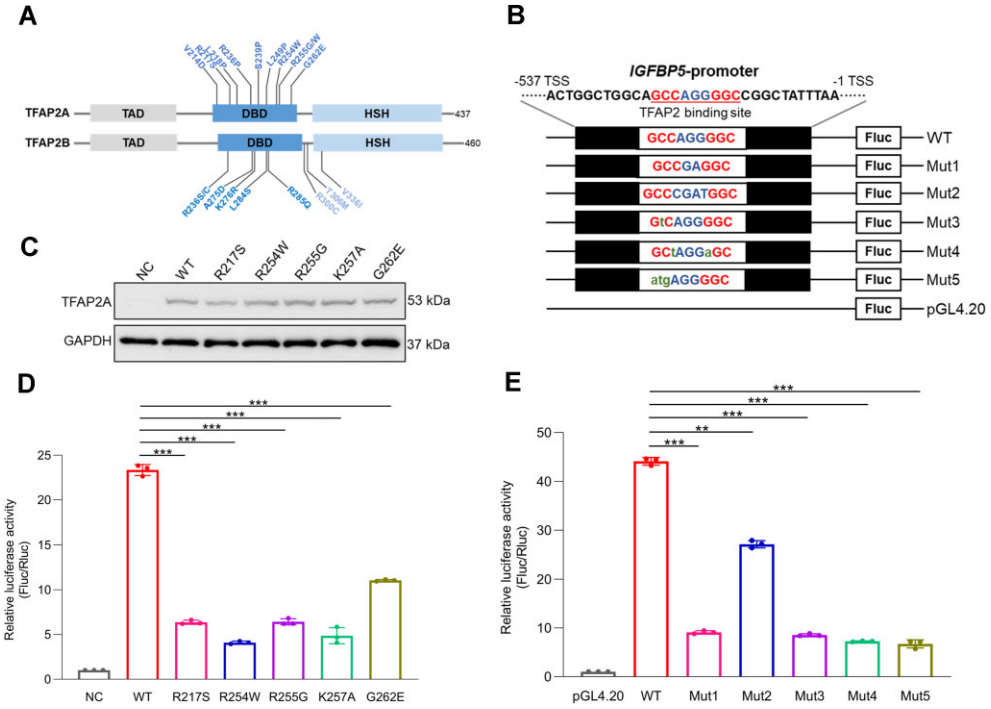


Figure 4. Transcriptional activity analysis of TFAP2A disease-associated mutants. (**A**) Disease-associated TFAP2A and TFAP2B mutations. (**B**) Schematic representation of the *IGFBP5*-promoter reporter assay. WT: normal *IGFBP5*-promoter region; Mut1-5: differently mutated *IGFBP5*-promoters; pGL4.20: negative control. The mutated nucleotides are shown in lower case. (**C**) Western blot analysis of TFAP2A wild-type and mutants in HEK293T 24 h after transfection. (**D**) Dual luciferase assays of wild-type TFAP2A and its mutants binding to WT *IGFBP5*-promoter in HEK293T. NC: pcDNA3.1 negative control. (**E**) Dual luciferase assays of TFAP2A binding to WT and mutated *IGFBP5*-promoters in HEK293T. pGL4.20: negative control. Error bars represent SD from three replicates ($n = 3$) (**$P < 0.01$ and ***$P < 0.001$).

BOFS, PDA, metastasis, lung carcinoma, carcinogenesis, and breast cancer (59,66–68). The TFAP2 family proteins are not only biomarkers for disease diagnosis but also important prognostic and therapeutic targets (14,69–71). Additionally, the 3D structure and DNA binding ability of the human TFAP2 is also conserved in *Drosophila* and *C. elegans* (72,73). Despite the increasing interest in the critical functions of the TFAP2 family, how TFAP2 homodimer and heterodimer specifically bind to a particular promoter to regulate the expression of target genes, and how mutations of the TFAP2 family lead to diseases remain unresolved.

In this study, we found that TFAP2 preferred binding to the GCC(N3)GGC sequence, and the pseudo-palindromic GCC and GGC motifs and the length of the central spacer between the two motifs determined their binding specificity. Furthermore, we first determined the structures of TFAP2A and TFAP2B homodimers in their apo form, and the structures of the TFAP2A homodimer bound to its consensus DNA sequence. Structural analysis revealed that two amphipathic α-helical HSH domains of TFAP2 stacked with each other to form a dimer via tight hydrophobic interactions, while the stabilized loops from both DBD domains inserted into two neighboring major grooves of the DNA duplex to form base-specific interactions. Although we did not obtain the TFAP2 heterodimer structure, sequence and structural comparisons of the HSH-DBD domains of TFAP2A and TFAP2B showed that both TFAP2 proteins share a high sequence and structure similarity. These characteristics, together with our mutagenesis and binding data, suggest that the TFAP2A-TFAP2B heterodimer should dimerize and recognize the consensus DNA sequence in the same mode as that of the TFAP2A homodimer.

It has been well known that transcription factors (TFs) can bind to their target site DNA sequences as either homodimers or heterodimers with other TFs, such as the PadR proteins, bZIP (basic-region leucine zipper) and bHLH (basic helix-loop-helix) (51,74,75), however, their dimerization and DNA recognition mode differ from that of the TFAP2 family. For example, PadR forms a Dolmen shape dimerization mode and binds to the major and minor grooves of dsDNA simultaneously via a winged helix-turn-helix (wHTH) domain (51). bZIP TFs usually form homodimers or heterodimers through a Leucine zipper, and they recognize the dsDNA mainly using an α-helix. Furthermore, the bZIP TF homodimers or heterodimers usually display different DNA binding sequence preferences depending on their dimerization partners (75). Thus, the structures of the TFAP2 family determined in this study exhibit a novel fold and DNA binding mode.

DATA AVAILABILITY

The crystal structures of the DBD-HSH tandem domain of TFAP2A in apo form and in complex with different DNAs, and the DBD-HSH tandem domain of TFAP2B in apo

form have been deposited in the Protein Data Bank with accession codes 8J0L, 8J0K, 8J0R and 8J0Q, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff members at BL19U1 and BL10U2 of the Shanghai Synchrotron Radiation Facility (SSRF) of China for support in diffraction data collection.

FUNDING

Central China Normal University (CCNU) [KJ02502022-0435] and National Natural Science Foundation of China [31770834]. Funding for open access charge: National Natural Science Foundation of China [31770834]; Central China Normal University (CCNU) [KJ02502022-0435].
Conflict of interest statement. None declared.

REFERENCES

1. Mitchell,P.J., Wang,C. and Tjian,R. (1987) Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell*, **50**, 847–861.
2. Eckert,D., Buhl,S., Weber,S., Jager,R. and Schorle,H. (2005) The AP-2 family of transcription factors. *Genome Biol.*, **6**, 246.
3. Zhao,F., Satoda,M., Licht,J.D., Hayashizaki,Y. and Gelb,B.D. (2001) Cloning and characterization of a novel mouse AP-2 transcription factor, AP-2delta, with unique DNA binding and transactivation properties. *J. Biol. Chem.*, **276**, 40755–40760.
4. Moser,M., Ruschoff,J. and Buettner,R. (1997) Comparative analysis of AP-2 alpha and AP-2 beta gene expression during murine embryogenesis. *Dev. Dyn.*, **208**, 115–124.
5. Eloranta,J.J. and Hurst,H.C. (2002) Transcription factor AP-2 interacts with the SUMO-conjugating enzyme UBC9 and is sumolated in vivo. *J. Biol. Chem.*, **277**, 30798–30804.
6. Williams,M.J., Goergen,P., Rajendran,J., Zheleznyakova,G., Hagglund,M.G., Perland,E., Bagchi,S., Kalogeropoulou,A., Khan,Z., Fredriksson,R. *et al.* (2014) Obesity-linked homologues TfAP-2 and Twz establish meal frequency in *Drosophila melanogaster*. *PLos Genet.*, **10**, e1004499.
7. Magnusdottir,E., Dietmann,S., Murakami,K., Gunesdogan,U., Tang,F., Bao,S., Diamanti,E., Lao,K., Gottgens,B. and Azim Surani,M. (2013) A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.*, **15**, 905–915.
8. Rothstein,M. and Simoes-Costa,M. (2020) Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Res.*, **30**, 35–48.
9. Bassett,E.A., Korol,A., Deschamps,P.A., Buettner,R., Wallace,V.A., Williams,T. and West-Mays,J.A. (2012) Overlapping expression patterns and redundant roles for AP-2 transcription factors in the developing mammalian retina. *Dev. Dyn.*, **241**, 814–829.
10. Kantarci,H., Edlund,R.K., Groves,A.K. and Riley,B.B. (2015) Tfap2a promotes specification and maturation of neurons in the inner ear through modulation of bmp, fgf and notch signaling. *PLos Genet.*, **11**, e1005037.
11. Bamforth,S.D., Braganca,J., Eloranta,J.J., Murdoch,J.N., Marques,F.I., Kranc,K.R., Farza,H., Henderson,D.J., Hurst,H.C. and Bhattacharya,S. (2001) Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. *Nat. Genet.*, **29**, 469–474.
12. Schorle,H., Meier,P., Buchert,M., Jaenisch,R. and Mitchell,P.J. (1996) Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature*, **381**, 235–238.
13. Raap,M., Gierendt,L., Kreipe,H.H. and Christgen,M. (2021) Transcription factor AP-2beta in development, differentiation and tumorigenesis. *Int. J. Cancer*, **149**, 1221–1227.
14. Kolat,D., Kaluzinska,Z., Bednarek,A.K. and Pluciennik,E. (2022) Prognostic significance of AP-2alpha/gamma targets as cancer therapeutics. *Sci. Rep.*, **12**, 5497.
15. Williams,T. and Tjian,R. (1991) Characterization of a dimerization motif in AP-2 and its function in heterologous DNA-binding proteins. *Science*, **251**, 1067–1071.
16. Williams,T. and Tjian,R. (1991) Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. *Genes Dev.*, **5**, 670–682.
17. Hoffman,T.L., Javier,A.L., Campeau,S.A., Knight,R.D. and Schilling,T.F. (2007) Tfap2 transcription factors in zebrafish neural crest development and ectodermal evolution. *J. Exp. Zool. B Mol. Dev. Evol.*, **308**, 679–691.
18. Castro-Mondragon,J.A., Riudavets-Puig,R., Rauluseviciute,I., Lemma,R.B., Turchi,L., Blanc-Mathieu,R., Lucas,J., Boddie,P., Khan,A., Manosalva Perez,N. *et al.* (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **50**, D165–D173.
19. Liu,Q., Kulak,M.V., Borcherding,N., Maina,P.K., Zhang,W., Weigel,R.J. and Qi,H.H. (2018) A novel HER2 gene body enhancer contributes to HER2 expression. *Oncogene*, **37**, 687–694.
20. Rada-Iglesias,A., Bajpai,R., Prescott,S., Brugmann,S.A., Swigut,T. and Wysocka,J. (2012) Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell*, **11**, 633–648.
21. Van Otterloo,E., Milanda,I., Pike,H., Thompson,J.A., Li,H., Jones,K.L. and Williams,T. (2022) AP-2alpha and AP-2beta cooperatively function in the craniofacial surface ectoderm to regulate chromatin and gene expression dynamics during facial development. *Elife*, **11**, e70511.
22. Scott,C.C., Vossio,S., Rougemont,J. and Gruenberg,J. (2018) TFAP2 transcription factors are regulators of lipid droplet biogenesis. *Elife*, **7**, e36330.
23. Duan,C. and Clemmons,D.R. (1995) Transcription factor AP-2 regulates human insulin-like growth factor binding protein-5 gene expression. *J. Biol. Chem.*, **270**, 24844–24851.
24. Beck,A.C., Cho,E., White,J.R., Paemka,L., Li,T., Gu,V.W., Thompson,D.T., Koch,K.E., Franke,C., Gosse,M. *et al.* (2021) AP-2alpha regulates S-phase and is a marker for sensitivity to PI3K inhibitor Buparlisib in colon cancer. *Mol. Cancer Res.*, **19**, 1156–1167.
25. Meulemans,D. and Bronner-Fraser,M. (2002) Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. *Development*, **129**, 4953–4962.
26. Yu,J.K., Meulemans,D., McKeown,S.J. and Bronner-Fraser,M. (2008) Insights from the amphioxus genome on the origin of vertebrate neural crest. *Genome Res.*, **18**, 1127–1132.
27. Seberg,H.E., Van Otterloo,E., Loftus,S.K., Liu,H., Bonde,G., Sompallae,R., Gildea,D.E., Santana,J.F., Manak,J.R., Pavan,W.J. *et al.* (2017) TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF. *PLoS Genet.*, **13**, e1006636.
28. Kenny,C., Dilshat,R., Seberg,H.E., Van Otterloo,E., Bonde,G., Helverson,A., Franke,C.M., Steingrimsson,E. and Cornell,R.A. (2022) TFAP2 paralogs facilitate chromatin access for MITF at pigmentation and cell proliferation genes. *PLoS Genet.*, **18**, e1010207.
29. Wang,X., Bolotin,D., Chu,D.H., Polak,L., Williams,T. and Fuchs,E. (2006) AP-2alpha: a regulator of EGF receptor signaling and proliferation in skin epidermis. *J. Cell Biol.*, **172**, 409–421.
30. White,J.R., Thompson,D.T., Koch,K.E., Kiriazov,B.S., Beck,A.C., van der Heide,D.M., Grimm,B.G., Kulak,M.V. and Weigel,R.J. (2021) AP-2alpha-mediated activation of E2F and EZH2 drives melanoma metastasis. *Cancer Res.*, **81**, 4455–4470.
31. Pastor,W.A., Liu,W., Chen,D., Ho,J., Kim,R., Hunt,T.J., Lukianchikov,A., Liu,X., Polo,J.M., Jacobsen,S.E. *et al.* (2018) TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.*, **20**, 553–564.
32. Krendl,C., Shaposhnikov,D., Rishko,V., Ori,C., Ziegenhain,C., Sass,S., Simon,L., Muller,N.S., Straub,T., Brooks,K.E. *et al.* (2017) GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E9579–E9588.

33. Sanchez-Ferras,O., Pacis,A., Sotiropoulou,M., Zhang,Y., Wang,Y.C., Bourgey,M., Bourque,G., Ragoussis,J. and Bouchard,M. (2021) A coordinated progression of progenitor cell states initiates urinary tract development. *Nat. Commun.*, **12**, 2627.
34. Schmidt,M., Huber,L., Majdazari,A., Schutz,G., Williams,T. and Rohrer,H. (2011) The transcription factors AP-2beta and AP-2alpha are required for survival of sympathetic progenitors and differentiated sympathetic neurons. *Dev. Biol.*, **355**, 89–100.
35. Chambers,B.E., Gerlach,G.F., Clark,E.G., Chen,K.H., Levesque,A.E., Leshchiner,I., Goessling,W. and Wingert,R.A. (2019) Tfap2a is a novel gatekeeper of nephron differentiation during kidney development. *Development*, **146**, dev172387.
36. Winger,Q., Huang,J., Auman,H.J., Lewandoski,M. and Williams,T. (2006) Analysis of transcription factor AP-2 expression and function during mouse preimplantation development. *Biol. Reprod.*, **75**, 324–333.
37. Li,W. and Cornell,R.A. (2007) Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. *Dev. Biol.*, **304**, 338–354.
38. Wang,X., Pasolli,H.A., Williams,T. and Fuchs,E. (2008) AP-2 factors act in concert with Notch to orchestrate terminal differentiation in skin epidermis. *J. Cell Biol.*, **183**, 37–48.
39. Van Otterloo,E., Li,W., Bonde,G., Day,K.M., Hsu,M.Y. and Cornell,R.A. (2010) Differentiation of zebrafish melanophores depends on transcription factors AP2 alpha and AP2 epsilon. *PLoS Genet.*, **6**, e1001122.
40. Liu,K., Zhang,J., Xiao,Y., Yang,A., Song,X., Li,Y., Chen,Y., Hughes,T.R. and Min,J. (2023) Structural insights into DNA recognition by the BEN domain of the transcription factor BANP. *J. Biol. Chem.*, **299**, 104734.
41. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
42. Kabsch,W. (2010) Xds. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
43. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta. Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
44. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
45. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
46. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A. *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
47. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
48. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta. Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
49. Mohibullah,N., Donner,A., Ippolito,J.A. and Williams,T. (1999) SELEX and missing phosphate contact analyses reveal flexibility within the AP-2(alpha) protein: DNA binding complex. *Nucleic Acids Res.*, **27**, 2760–2769.
50. Jolma,A., Yan,J., Whitington,T., Toivonen,J., Nitta,K.R., Rastas,P., Morgunova,E., Enge,M., Taipale,M., Wei,G. *et al.* (2013) DNA-binding specificities of human transcription factors. *Cell*, **152**, 327–339.
51. Park,S.C., Kwak,Y.M., Song,W.S., Hong,M. and Yoon,S.I. (2017) Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR. *Nucleic Acids Res.*, **45**, 13080–13093.
52. Holm,L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res.*, **50**, W210–W215.
53. Wang,D., Hashimoto,H., Zhang,X., Barwick,B.G., Lonial,S., Boise,L.H., Vertino,P.M. and Cheng,X. (2017) MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.*, **45**, 2396–2407.
54. Li,Z., Jaroszewski,L., Iyer,M., Sedova,M. and Godzik,A. (2020) FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res.*, **48**, W60–W64.
55. Woodfield,G.W., Hitchler,M.J., Chen,Y., Domann,F.E. and Weigel,R.J. (2009) Interaction of TFAP2C with the estrogen receptor-alpha promoter is controlled by chromatin structure. *Clin. Cancer Res.*, **15**, 3672–3679.
56. Hu,S., Wan,J., Su,Y., Song,Q., Zeng,Y., Nguyen,H.N., Shin,J., Cox,E., Rho,H.S., Woodard,C. *et al.* (2013) DNA methylation presents distinct binding sites for human transcription factors. *Elife*, **2**, e00726.
57. Thewes,V., Orso,F., Jager,R., Eckert,D., Schafer,S., Kirfel,G., Garbe,S., Taverna,D. and Schorle,H. (2010) Interference with activator protein-2 transcription factors leads to induction of apoptosis and an increase in chemo- and radiation-sensitivity in breast cancer cells. *BMC Cancer*, **10**, 192.
58. Kolat,D., Kaluzinska,Z., Bednarek,A.K. and Pluciennik,E. (2019) The biological characteristics of transcription factors AP-2alpha and AP-2gamma and their importance in various types of cancers. *Biosci. Rep.*, **39**, BSR20181928.
59. Cheng,C., Ai,Z. and Zhao,L. (2020) Comprehensive analysis of the expression and prognosis for TFAP2 in human lung carcinoma. *Genes Genomics*, **42**, 779–789.
60. Li,H., Sheridan,R. and Williams,T. (2013) Analysis of TFAP2A mutations in Branchio-Oculo-Facial Syndrome indicates functional complexity within the AP-2alpha DNA-binding domain. *Hum. Mol. Genet.*, **22**, 3195–3206.
61. Meshcheryakova,T.I., Zinchenko,R.A., Vasilyeva,T.A., Marakhonov,A.V., Zhylina,S.S., Petrova,N.V., Kozhanova,T.V., Belenikin,M.S., Petrin,A.N. and Mutovin,G.R. (2015) A clinical and molecular analysis of branchio-oculo-facial syndrome patients in Russia revealed new mutations in TFAP2A. *Ann. Hum. Genet.*, **79**, 148–152.
62. Thomeer,H.G., Crins,T.T., Kamsteeg,E.J., Buijsman,W., Cruysberg,J.R., Knoers,N.V. and Cremers,C.W. (2010) Clinical presentation and the presence of hearing impairment in branchio-oculo-facial syndrome: a new mutation in the TFAP2A gene. *Ann. Otol. Rhinol. Laryngol.*, **119**, 806–814.
63. Edward,H.L., D'Gama,A.M., Wojcik,M.H., Brownstein,C.A., Kenna,M.A., Grant,P.E., Majzoub,J.A. and Agrawal,P.B. (2019) A novel missense mutation in TFAP2B associated with Char syndrome and central diabetes insipidus. *Am. J. Med. Genet. A*, **179**, 1299–1303.
64. Lingaiah,K., Sosalagere,D.M., Mysore,S.R., Krishnamurthy,B., Narayanappa,D. and Nallur,R.B. (2011) Mutations of TFAP2B in congenital heart disease patients in Mysore, South India. *Indian J. Med. Res.*, **134**, 621–626.
65. Mani,A., Radhakrishnan,J., Farhi,A., Carew,K.S., Warnes,C.A., Nelson-Williams,C., Day,R.W., Pober,B., State,M.W. and Lifton,R.P. (2005) Syndromic patent ductus arteriosus: evidence for haploinsufficient TFAP2B mutations and identification of a linked sleep disorder. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 2975–2979.
66. Campbell,N.R., Rao,A., Hunter,M.V., Sznurkowska,M.K., Briker,L., Zhang,M., Baron,M., Heilmann,S., Deforet,M., Kenny,C. *et al.* (2021) Cooperation between melanoma cell states promotes metastasis through heterotypic cluster formation. *Dev. Cell*, **56**, 2808–2825.
67. Cyr,A.R., Kulak,M.V., Park,J.M., Bogachek,M.V., Spanheimer,P.M., Woodfield,G.W., White-Baer,L.S., O'Malley,Y.Q., Sugg,S.L., Olivier,A.K. *et al.* (2015) TFAP2C governs the luminal epithelial phenotype in mammary development and carcinogenesis. *Oncogene*, **34**, 436–444.
68. Park,J.M., Wu,T., Cyr,A.R., Woodfield,G.W., De Andrade,J.P., Spanheimer,P.M., Li,T., Sugg,S.L., Lal,G., Domann,F.E. *et al.* (2015) The role of Tcfap2c in tumorigenesis and cancer growth in an activated Neu model of mammary carcinogenesis. *Oncogene*, **34**, 6105–6114.
69. Al-Sabri,M.H., Nikpour,M., Clemensson,L.E., Attwood,M.M., Williams,M.J., Rask-Anderson,M., Mwinyi,J. and Schioth,H.B. (2022) The regulatory role of AP-2beta in monoaminergic neurotransmitter systems: insights on its signalling pathway, linked disorders and theragnostic potential. *Cell Biosci*, **12**, 151.

70. Wu,V.T., Kiriazov,B., Koch,K.E., Gu,V.W., Beck,A.C., Borcherding,N., Li,T., Addo,P., Wehrspan,Z.J., Zhang,W. *et al.* (2020) A TFAP2C gene signature is predictive of outcome in HER2-positive breast cancer. *Mol. Cancer Res.*, **18**, 46–56.
71. Anttila,M.A., Kellokoski,J.K., Moisio,K.I., Mitchell,P.J., Saarikoski,S., Syrjanen,K. and Kosma,V.M. (2000) Expression of transcription factor AP-2alpha predicts survival in epithelial ovarian cancer. *Br. J. Cancer*, **82**, 1974–1983.
72. Bauer,R., McGuffin,M.E., Mattox,W. and Tainsky,M.A. (1998) Cloning and characterization of the Drosophila homologue of the AP-2 transcription factor. *Oncogene*, **17**, 1911–1922.
73. Turek,M., Besseling,J., Spies,J.P., Konig,S. and Bringmann,H. (2016) Sleep-active neuron specification and sleep induction require FLP-11 neuropeptides to systemically induce sleep. *Elife*, **5**, e12499.
74. Vinson,C., Myakishev,M., Acharya,A., Mir,A.A., Moll,J.R. and Bonovich,M. (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol. Cell. Biol.*, **22**, 6321–6335.
75. Rodriguez-Martinez,J.A., Reinke,A.W., Bhimsaria,D., Keating,A.E. and Ansari,A.Z. (2017) Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *Elife*, **6**, e19272.