

The ALX4 dimer structure provides insight into how disease alleles impact function

Received: 11 September 2024

Accepted: 29 April 2025

Published online: 23 May 2025

Brittany Cain^{1,5}✉, Zhenyu Yuan^{2,5}, Evelyn Thoman³, Rhett A. Kovall^{1,2}✉ & Brian Gebelein^{1,4}✉

How homeodomain proteins gain sufficient DNA binding specificity to regulate diverse processes is a long-standing question. Here, we determine how the ALX4 Paired-like protein achieves DNA binding specificity for a TAAT–NNN–ATTA dimer site. We first show that ALX4 binds this motif independently of its co-factor, TWIST1, in cranial neural crest cells. Structural analysis identifies seven ALX4 residues that participate in dimer binding, many of which are conserved across the Paired-like family, but not other homeodomain proteins. Unexpectedly, the two ALX4 proteins within the dimer use distinct residues to form asymmetric protein-protein and protein-DNA interactions and mediate cooperativity. Moreover, we find that ALX4 cooperativity is required for transcriptional activation and that ALX4 disease variants cause distinct molecular defects that include loss of cooperativity. These findings provide insights into how Paired-like factors gain DNA specificity and show how disease variants can be stratified based on their molecular defects.

The homeodomain (HD) family is one of the largest families of transcription factors (TFs), consisting of ~200 proteins in humans that are classified by their highly conserved helix-turn-helix DNA binding domain¹. In vitro DNA binding assays have shown that HDs largely bind highly similar AT-rich DNA sequences^{2–4}, primarily through contacts with the N-terminal Arginine Rich Motif (ARM) and the third helix that is also known as the recognition helix⁵. However, the mechanisms by which HD TFs achieve sufficient DNA binding specificity to regulate their distinct targets in vivo are incomplete and not well understood.

Several mechanisms have been described to address this paradox. First, HD subfamilies such as the Paired, Prospero, and CUT classes encode additional DNA binding domains that bind distinct sequences. Second, studies have shown that while consensus high affinity sites are often bound by many HD factors in a relatively indiscriminate manner, low affinity sites are typically bound by fewer HD factors and thereby result in enhanced target specificity^{6–8}. Third, some HD proteins form complexes that facilitate binding to distinct and/or longer recognition sequences. For example, the HOX factors form complexes with the PBX and MEIS HD proteins to increase DNA binding specificity⁹.

Additionally, members of the Paired-like family, which share sequence conservation with the HD encoded by Paired (PAX) family members but lack the accompanying Paired DNA binding domain¹⁰, have been shown to form dimers that both increase DNA binding specificity^{11,12} and alter transcriptional output^{11,13}. Moreover, we recently used a computational pipeline to predict cooperative homodimer binding of HD TFs from high throughput systematic evolution of ligands by exponential enrichment (HT-SELEX) data and found that several, but not all, Paired-like members cooperatively bind a TAAT–NNN–ATTA (P3) dimer site¹⁴. Currently, however, it is not well understood which HD residues within the Paired-like factors confer cooperative DNA binding to these P3 sites.

The mechanisms by which Paired-like factors bind cooperatively have been previously extrapolated from a crystal structure of a modified Paired (Prd) *Drosophila* HD bound to DNA¹⁵. However, there are several caveats to using this structure as a model. First, Prd is a PAX TF, containing a Paired domain and a HD that both possess the capability to bind DNA, and Prd preferentially uses these domains to bind a distinct hybrid site that does not resemble a HD site^{16,17}. Second, the HD

¹Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, MLC 7007, Cincinnati, OH, USA. ²Department of Molecular and Cellular Biosciences, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ³Department of Biomedical Engineering, University of Cincinnati, Cincinnati, OH, USA. ⁴Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁵These authors contributed equally: Brittany Cain, Zhenyu Yuan. ✉ e-mail: Brittany.Cain@cchmc.org; kovallra@ucmail.uc.edu; Brian.Gebelein@cchmc.org

of the wildtype Prd protein, which contains a serine at position 50 (S50) within the third helix, binds cooperatively with equal affinity to a TAAT-NN-ATTA sequence (P2 site) and the P3 site, whereas the Paired-like factors encode a Q50 HD and strongly prefer binding the P3 site¹². Hence, the Prd HD only favors the P3 site when a S50Q mutation is created¹², making it unclear whether the Paired-like factors use similar interactions and mechanisms to bind cooperatively.

Here, we focus on defining the mechanisms used by the Paired-like factor, ALX4, to cooperatively bind DNA. Genetic studies in mice and humans have shown that ALX4 is critical for proper craniofacial development. *Alx4* null mutations in mice cause enlarged parietal foramina in which the parietal bones fail to fuse, resulting in a gap at the sagittal sinus of the skull as well as polydactyly and loss of cartilage in the limbs^{18,19}. Consistent with these phenotypes, ALX4 variants in humans have been associated with a variety of craniofacial defects. Patients with a homozygous nonsense variant display frontonasal abnormalities, alopecia, and sagittal suture defects²⁰, and missense variants in ALX4 are implicated in three distinct developmental disorders: Enlarged parietal foramina^{21–25}, sagittal suture craniosynostosis in which the parietal bones fuse too early in development leading to stunted brain growth, increased intracranial pressure, and scaphocephaly²⁶, and frontonasal dysplasia in which patients exhibit abnormal development of craniofacial features^{21,25,27}. The molecular mechanisms behind how many of these ALX4 missense variants cause these distinct disease phenotypes are not well understood.

A recent study found that ALX4 can also cooperatively bind DNA with the basic helix loop helix (bHLH) factor, TWIST1, to a composite site called the coordinator that contains a bHLH E-box site and a HD site spaced 6 bp apart²⁸. This motif was originally identified because its presence correlated with the divergence in acetylation patterns between human and chimp enhancers, suggesting that the coordinator motif and the TFs that bind to it drive differences between human and chimp craniofacial development²⁹. To identify the TFs that bind the coordinator site, Kim et al. generated a large amount of genomic binding data in human cranial neural crest cells (hCNCCs) to show that TWIST1 binds to the E-box sequence, drives chromatin opening, promotes HD recruitment of either ALX1, ALX4, MSX1, or PRRX1, and promotes enhancer acetylation at the coordinator motif²⁸. Taken together, these studies suggest that ALX4 uses multiple modes of DNA binding to regulate the expression of target genes required for craniofacial development.

In this study, we use a combination of structural, biochemical, and bioinformatic approaches to both determine how ALX4 homodimers cooperatively bind DNA and determine how disease variants within the HD impact ALX4's molecular functions. First, we take advantage of published ALX4 genomic binding data from differentiated hCNCCs to identify diverse ALX4 bound site types²⁸. This analysis reveals ALX4 binding to the TAAT-NNN-ATTA (P3) site is largely independent of TWIST1, whereas ALX4 binding to monomer and coordinator sites is largely TWIST1-dependent. Next, we use crystallography to solve the ALX4 HD dimer structure bound to a P3 site and find that while it resembles the Prd S50Q HD dimer¹⁵, ALX4 forms unique asymmetric interactions between the two bound ALX4 proteins. Based on this structure, we used mutagenesis studies to define residues critical for cooperative DNA binding and assess how five ALX4 missense variants associated with human birth defects impact DNA binding, cooperativity, and transcriptional output. Importantly, disease variants can be stratified based on their molecular impact which includes complete loss of DNA binding, loss of cooperative DNA binding, loss of protein stability, and subcellular mis-localization, each of which leads to decreased transcriptional output on the P3 site. Overall, these studies reveal mechanisms by which ALX4 binds cooperatively to its P3 site and begin to define the gene regulatory impact of ALX4 cooperativity and its role in ALX4-associated disease.

Results

Genomic binding of ALX4 to the P3 dimer site is independent of co-factor TWIST1

To identify the genomic DNA binding sites of ALX4, we utilized available CUT&RUN (C&R) data from human embryonic stem cells lines differentiated into hCNCCs²⁸. These cells contain an engineered TWIST1 locus that expresses a FKBP12^{F36V}-tagged TWIST1 protein that undergoes rapid degradation following treatment with a dTAG^V-1 ligand, providing temporally controlled expression of TWIST1 (Fig. 1A). With this system, Kim et al. assessed ALX4 genomic binding in the presence and absence of TWIST1²⁸. Since Kim et al. primarily analyzed this data from the perspective of TWIST1 binding (i.e., focused on co-binding to TWIST1 bound regions), we reanalyzed the ALX4 C&R data to determine if ALX4 bound to TWIST1-independent loci as well as TWIST1-dependent loci. From this analysis, we identified 3,250 ALX4 bound regions in the genome (Fig. 1B). Consistent with these regions containing active enhancers, ALX4 bound regions were highly accessible and had high signal for H3K27 acetylation (Supplementary Fig. 1). After a 1 hour treatment with the dTAG^V-1 ligand that results in rapid degradation of the FKBP12^{F36V}-tagged TWIST1 protein (Fig. 1A), ALX4 no longer bound to 2867 of these regions, indicating that many ALX4 bound regions were dependent on TWIST1 (Fig. 1B). However, 383 ALX4 genomic regions were bound similarly in the presence and absence of TWIST1.

We next performed a de novo motif search³⁰ on the ALX4 bound regions and focused on four distinct binding motifs: the Coordinator, the P3 dimer site, an independent E-box site, and an independent HD monomer site (Figs. 1C and Supplementary Fig. 2). However, the presence of a motif does not guarantee TF binding, and the presence of a composite site does not guarantee that both TFs are bound at the same time. To stringently assess for TF binding, we first scanned ALX4 peaks for potential binding sites using Homer and then filtered potential binding sites by comparing MNase digestion patterns of the motif and nearby flanking windows. If a given site is bound, the motif should be protected from MNase cleavage, and the neighboring accessible sequences will be preferentially cleaved. To ensure sites were not scored in more than one category, we first identified all digest confirmed coordinator and P3 dimer sites and removed these prior to analyzing for independent E-box and HD monomer binding (Supplementary Fig. 3A). Using this approach, we identified > 500 bound sites for each site type within the 3250 bound regions (Supplementary Fig. 3A). Figure 1C displays the average 5' MNase signal for the filtered site types, and the digestion signal of each site is shown in Supplementary Fig. 3B. Note, the length of the protected region correlates with the length of the motif as expected.

The above data provides evidence that ALX4 binds three types of sites in vivo: Coordinator sites in complex with TWIST1, **TAAT-NNN-ATTA** (P3) dimer sites, and independent HD monomer sites. Surprisingly, we also found footprint protection at 1026 E-box sites that were not affiliated with a Coordinator site. To assess if these sites were mischaracterized weak coordinator sites, we analyzed the flanking sequences surrounding the independent E-box sites and did not observe enrichment for HD sites at the appropriate spacing from the E-box site, whereas similar analysis of the called coordinator sites showed clear HD site enrichment 6 bps from the E-box sequence (Supplementary Fig. 4A, B). These findings suggest that these protected E-box sequences in the ALX4 C&R data were not misclassified coordinator motifs and are likely to represent nearby, independent bHLH binding to these sites. As a further test of this idea, we performed footprint analysis for another TF motif that is significantly enriched in the called ALX4 C&R peaks. For this analysis, we selected the enriched AP-2 motif (Supplementary Fig. 2) bound by TFAP transcription factors because this motif is GC-rich and therefore unlikely to be directly bound by ALX4 and because TFAP2A has been previously shown to

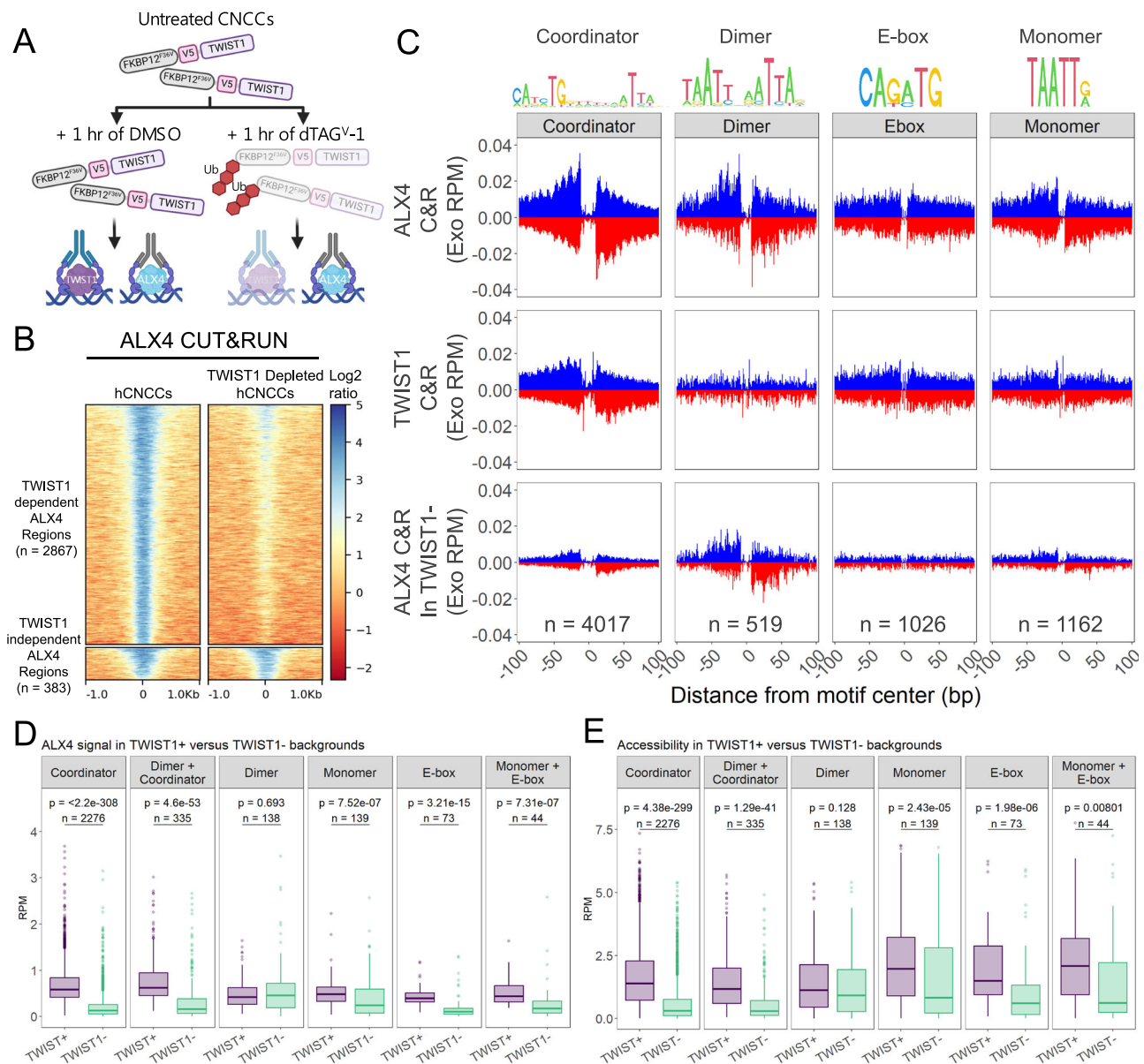


Fig. 1 | ALX4 genomic binding to the P3 dimer site is TWIST1 independent.

A Schematics of experiments performed in Kim et al.²⁸. Created in BioRender. Cain, B. (2025) <https://BioRender.com/p80e992> **(B)** ALX4 bound 3250 regions in wild-type conditions. 2867 (88%) regions were no longer bound in dTAGV-1 treated samples in which TWIST1 was depleted (TWIST1-dependent ALX4 regions). 383 (12%) regions were maintained after TWIST1 depletion (TWIST1-independent ALX4 regions). The color bar denotes the log2 ratio of the immunoprecipitated sample versus the IgG control. **C** A HOMER de novo motif analysis revealed four distinct site types that were enriched in the ALX4 C&R: the Coordinator²⁸, the P3 dimer site, an independent E-box, and a monomer HD site. MNase digestion patterns were used to discriminate between bound and unbound sites. TWIST1 dependence was assessed by footprinting the ALX4 bound sites with TWIST1 C&R and ALX4 C&R in a TWIST1 depleted background. **D** ALX4 binding signal decreases significantly after TWIST1 depletion for all peak sets except for peaks containing only the P3 dimer site. The 3250 ALX4 bound regions were categorized based on the presence of a

coordinator site, a coordinator and dimer site, a dimer site, or only a monomer site or E-box or both independent sites. The average reads per million (RPM) after E. coli spike in normalization of the 200 bp window at the peak center were compared with a two-sided Wilcoxon rank sum test. Each datapoint represents the RPM for a peak center. **E** Chromatin accessibility decreases significantly after TWIST1 depletion for all peak sets except for peaks containing only the P3 dimer site. The average RPM of a 200 bp window at the peak centers was compared with a two-sided Wilcoxon rank sum test. Each datapoint represents the RPM for a peak center. Note, accessibility signals between TWIST1+ and TWIST1- datasets were comparable at housekeeping genes: GAPDH and ACTB (Supplementary Fig. 5). For boxplots, the center line defines the median, the lower and upper box limits correspond with the 25th and 75th percentiles respectively, the lower and upper whiskers define either 1.5 times the interquartile range or the most extreme value. Data outside of the whisker limits are plotted as individual points. Source Data are provided on [Figshare](https://www.figshare.com)⁶⁹.

regulate NCC enhancers³¹. Footprint analysis revealed that 234 TFAP sites within the ALX4 C&R peaks were protected, suggesting that the binding of other TFs within accessible chromatin can be detected through footprint analysis of the ALX4 C&R peaks (Supplementary Fig. 4C).

To assess which of the footprint protected sites identified in the ALX4 C&R data were also bound by TWIST1, we similarly analyzed the

TWIST1 C&R for protected binding motifs. We found that TWIST1 was largely bound at the coordinator site, but not the three other site types (Fig. 1C). Conversely, we wanted to identify ALX4 sites that are dependent versus independent of TWIST1. To do so, we performed footprint analysis for ALX4 bound sites in the TWIST1 depleted cells. Here, we found that the overall signal of ALX4 binding to the coordinator site was largely lost (Fig. 1C). Note, the remaining weakened

ALX4 footprint signal is consistent with the short-term treatment of dTAG^{V1} being insufficient to deplete all TWIST1 at these binding sites. Interestingly, ALX4 P3 dimer site protection and signal were retained more so than other site types, suggesting that ALX4 P3 site binding is largely independent of TWIST1. In contrast, the footprint signal at monomer ALX4 binding sites was dramatically decreased.

To statistically analyze the biased retention of ALX4 P3 dimer site binding in the TWIST1 depleted background, we looked at the total binding signal and chromatin accessibility across peak centers. We first separated ALX4 peaks into six categories: (1) peaks that contained a coordinator site but not a P3 dimer site; (2) peaks that contained a coordinator and a P3 dimer site; (3) peaks that contained a P3 dimer site without a coordinator site; and (4 through 6) peaks that did not contain a coordinator site or a P3 dimer site but contained either a (4) ALX4 monomer site, (5) a E-box site, or (6) both sites. The majority of peaks contained coordinator sites, which emphasizes TWIST1's strong influence on ALX4 binding²⁸. We then compared the ALX4 C&R binding signal as well as the ATAC-seq reads between wildtype cells and TWIST1 depleted cells. Note, ATAC-seq signals were comparable between the wildtype and TWIST1 depleted cells at common house-keeping genes (Supplementary Fig. 5). All peak categories significantly decreased in ALX4 binding signal (Fig. 1D) and chromatin accessibility (Fig. 1E) after TWIST1 depletion with the exception of peaks containing only P3 dimer sites, further demonstrating that the majority of ALX4 P3 dimer sites are bound independently of TWIST1. In contrast, the loss in accessibility in peaks containing only ALX4 monomer sites provides a potential explanation as to why these sites depend on TWIST1 regulation. Taken together, these data reveal that ALX4 genomic binding to the P3 dimer site is TWIST1-independent, whereas ALX4 binding to the coordinator and monomer motifs is TWIST1-dependent.

The ALX4 HD is sufficient to mediate cooperative DNA binding to the P3 site

We previously found that ALX4, as well as several other Paired-like HDs, form cooperative homodimers on P3 sites that are identical to the TWIST1-independent ALX4 site found in the genomic analysis of hCNCCs¹⁴. However, the mechanisms used by these HDs to bind DNA cooperatively as well as the regions of the protein required to facilitate cooperativity are largely unknown. Our prior studies revealed that an ALX4 protein containing the HD as well as relatively short peptide chains N- and C-terminal to the HD (aa 169-303) was highly cooperative on a probe containing a P3 site¹⁴. To assess if the ALX4 HD alone (aa 209-274) (Fig. 2A) is sufficient to mediate cooperative DNA binding, we performed quantitative electrophoretic mobility shift assays (EMSAs) using a P3 probe with the **TAAT**-TAG-**ATTA** site found within the 20 bp sequence that appeared most frequently after four cycles of ALX4 HT-SELEX^{21,14}. To control for spacer-length dependence of cooperativity, we created a P4 sequence with a single G inserted into the center of the spacer sequence (**TAAT**-TAGG-**ATTA**). Comparative analysis of ALX4 binding to the P3 and P4 sites revealed that the ALX4 HD is sufficient to cooperatively bind the palindromic sequence in a spacer-dependent manner (Fig. 2B; all replicates are shown in Supplementary Fig. 6A). We used Tau, which measures the multiplier by which the binding of the first ALX4 protein facilitates the binding of a second ALX4 protein, to quantify the cooperativity of ALX4 bound to a P3 site and P4 site¹²; (See "Methods" for more information). A Tau greater than one demonstrates that binding of the first protein facilitates the binding of the second protein, a Tau equal to one demonstrates independent binding, and a Tau less than one demonstrates that the binding of the first protein hinders the binding of the second protein. On the P3 site, the Tau value (110 ± 14) for the ALX4 HD revealed highly cooperative binding. In sharp contrast, when a single nucleotide is added between these two palindromic sites, cooperativity is lost as ALX4 fails to preferentially bind as a dimer complex ($\text{Tau} = 1.5 \pm 1.4$, Fig. 2C). Moreover, cooperativity is not impacted by which nucleotide is inserted into the

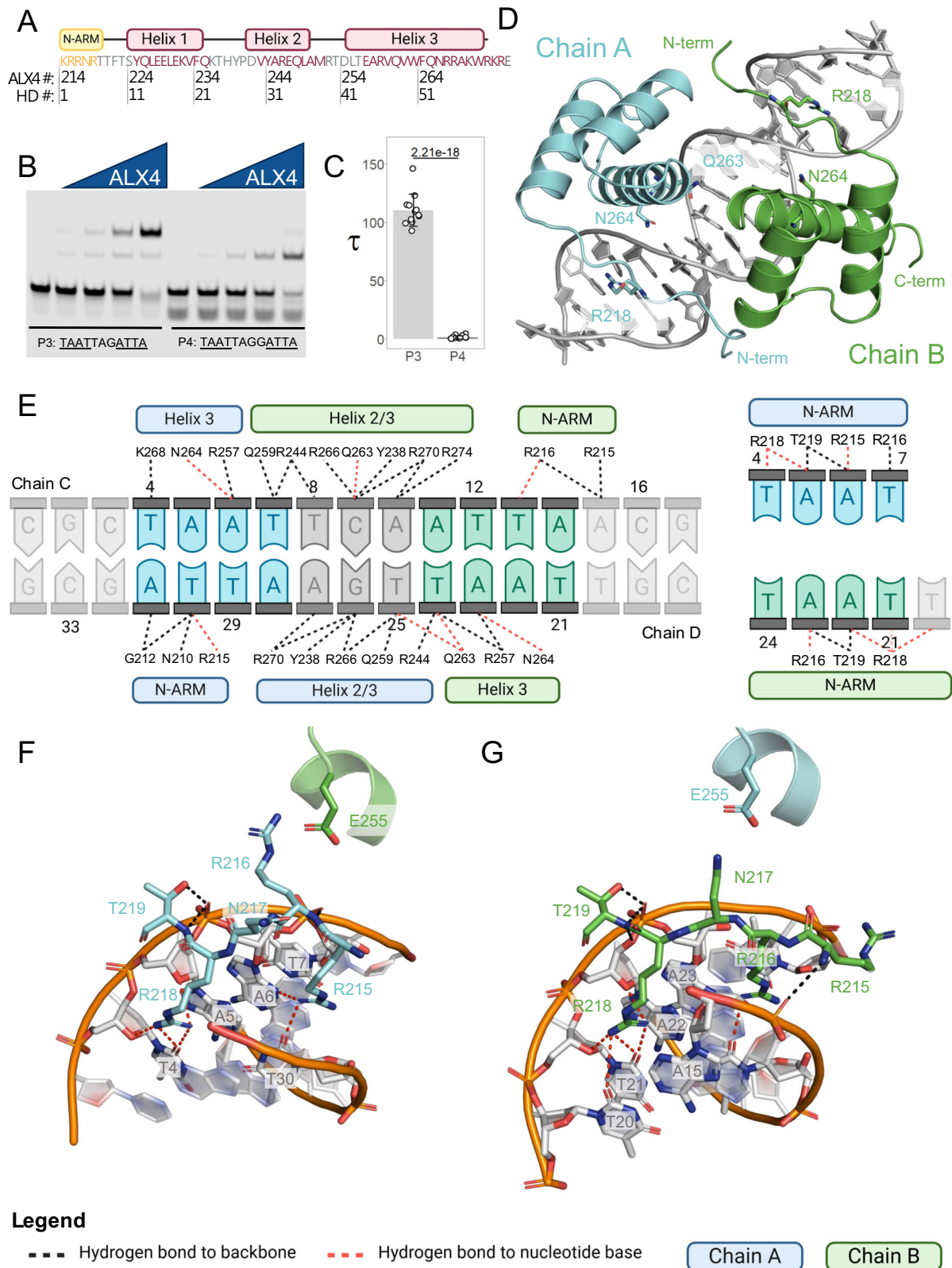
center of the spacer (Supplementary Fig. 6B), and this spacer length dependent cooperative behavior is shared across all members of the ALX family (Supplementary Fig. 6C, D). These findings demonstrate that the ALX HD is sufficient for cooperative DNA binding.

ALX4 dimeric crystal structure reveals asymmetrical DNA contacts via the N-terminal ARM

To determine the structural basis for the mechanisms that underlie ALX4 cooperativity, we solved the crystal structure of the ALX4 HD (aa 209-274) bound to a P3 site (-CGCT**TAATTCAATTA**ACG-) at 2.39 Å resolution (Fig. 2D) as well as the apo ALX4 HD at 2.39 Å resolution (Supplementary Fig. 7A; and Supplementary table 1). Both proteins in the dimer complex as well as the apo protein share the same overall structure with the exception of the N-terminal ARM, which is largely unstructured in apo ALX4 and thereby differs in both conformation and electron density resolution compared to ALX4 monomers in the dimer structure (Supplementary Fig. 7A). On the P3 DNA site, ALX4 forms a head-to-head dimer, in which DNA contacts are largely maintained, including minor groove interactions with the ALX4 N-terminal ARM and major groove interactions with helix 3 of ALX4 (Fig. 2D). Overall, the ALX4 dimer interactions are largely concentrated in two areas: (1) between the N-terminal ARM region of one molecule and the N-terminal residues of helix 2 from the other ALX4 HD molecule and (2) opposing alanine residues located in the N-terminal region of helix 3 of both ALX4 molecules (Fig. 3). Here, we label the two ALX4 proteins as Chain A (cyan) and Chain B (green), refer to residues based on their numbering in the ALX4 protein with the canonical HD numbering in parentheses, and separate our analysis of the ALX4 protein-to-DNA interactions (Fig. 2E-G) and ALX4 protein-to-protein interactions (Fig. 3).

We first used PDBePISA to map the interactions of each ALX4 protein to DNA³² (Fig. 2E). Consistent with other HD proteins, two regions of the ALX4 HD make most of the base-specific interactions with DNA: the third helix (residues 254-269) contacts the major groove, and the N-terminal ARM motif (residues 214-218) contacts the minor groove (Fig. 2D). Of these interactions, the major groove DNA contacts are largely symmetrical between the two ALX4 chains (Figs. 2E; and Supplementary Fig. 7B, C). N264 (canonical HD numbering: N51) forms base-specific hydrogen bonds with **TAAT** (dA6/dA23) (Chain A DNA contact/Chain B DNA contact), and Y238 (Y25), R244 (R31), R257 (R44), R266 (R53), and R270 (R57) form identical contacts to the DNA backbone of their respective half-site (Fig. 2E; and Supplementary Fig. 7B, C). However, the ALX4 Q263 (Q50) residue has distinct chain-specific configurations that contribute to major groove DNA binding. Prior studies have shown that the 50th residue of the HD can influence the specificity of the **TAATNN** nucleotides^{10,11,33,34}, and in the ALX4 dimer structure, Chain B: Q263 (Q50) has the potential to form base contacts with dC9, dA10, dT24, and dT25 in the major groove (Supplementary Fig. 7D). In contrast, Q263 (Q50) of Chain A does not form DNA contacts and instead forms intrachain hydrogen bonds with Q259 (Q46) (Supplementary Fig. 7E), albeit at this resolution we cannot formally exclude the possibility that Q263 (Q50) of Chain A forms water mediated contacts with DNA (Supplementary Fig. 7D').

As opposed to the largely symmetrical interactions found in the second and third helices, the ALX4 N-terminal ARM mediates several unique interactions apart from R218 (R5), which makes similar minor groove DNA contacts to each DNA half site. In the N-terminal ARM of Chain A, R215 (R2) inserts into the minor groove of DNA and forms a DNA base contact with dA6 and dT30 (Fig. 2E, F), while R216 (R3) of Chain A forms a salt bridge with E255 (E42) of Chain B (described below). Hence, in this interface, R215 (R2) forms the DNA contacts, whereas R216 (R3) forms the protein-protein interactions. In comparison, the R215 (R2) residue in Chain B is positioned outside of the minor groove, and the side



chain makes no DNA contacts (Fig. 2E, G). This residue primarily faces the protein-to-protein interface and forms contacts with the main chains of Y238 (Y25) and V241 (V28) (described below). In its stead, the R216 (R3) side chain of Chain B inserts into the minor groove and forms base-specific contacts with dT13 and dA23 as well as a backbone contact with dA15 (Fig. 2E, G). Interestingly, there was not sufficient electron density to model the residues

N210 (N -4) through G212 (G -2) and the side chains of K213 (K -1) and K214 (K1), suggesting that these regions were disordered in Chain B. It is possible that the R215 (R2) interaction to DNA stabilizes these N-terminal residues in Chain A. Thus, unlike the similar DNA-protein interactions mediated by the second and third helices, the N-terminal ARM of the two ALX4 proteins uses distinct residues to bind the minor groove of the P3 binding site.

Fig. 2 | ALX4 cooperativity is DNA mediated. **A** A map of the ALX4 HD with the ALX4 amino acid numbering (ALX4 #) and canonical homeodomain numbering (HD #) below. Created in BioRender. Cain, B. (2025) <https://BioRender.com/t12p127>. **B** The ALX4 HD is sufficient to mediate cooperativity on the P3 site but is unable to bind cooperatively to the P4 site. ALX4 protein (0, 37.5, 75, 150, and 300 nM) was combined with each respective fluorescent probe. A single replicate is shown here, and all three replicates are shown in Supplementary Fig. 6A. Schematics of the protein-DNA complexes are shown to the right of the gel. **C** Tau cooperativity factors were calculated for every lane in which protein was added. Bars depict the average Tau for each probe and each dot represents a Tau from an independent binding reaction ($n = 12$ binding reactions; biological replicates). Error bars denote standard deviation. Tau factors were compared with an unpaired, two-sided student t-test. Note, quantitation was derived from two gels from the same

experiment that were processed in parallel. **D** X-ray structure of ALX4 reveals that ALX4 binds as a dimer to the P3 site in a head-to-head orientation. ALX4 Chains A and B are colored cyan and green, respectively; DNA is colored gray. ALX4 residues R218 (canonical HD numbering: R5), Q263 (Q50), and N264 (N51) are shown in a stick representation. **E** Schematic of ALX4-DNA interactions. Hydrogen bond interactions were determined by PDBePISA and required atoms to be within 4 Å. Chain A and B proteins are labeled in blue and green respectively. Hydrogen bond interactions to the DNA backbone are shown via black dotted lines whereas hydrogen bonds to nucleotide bases are shown in red. Created in BioRender. Cain, B. (2025) <https://BioRender.com/a69f574>. **F, G** Structural comparisons of protein-DNA interactions made by the ALX4 Chains A and B. Source Data are provided on Figshare⁶⁹.

ALX4 dimer interactions are mediated via three interfaces

The two ALX4 proteins bound to DNA interact through three interfaces (Fig. 3). The first interface is between the N-terminal ARM of Chain A and the turn between helix 1 and helix 2 on Chain B. In this interface, the K213 (K -1) side chain of Chain A forms hydrogen bonds with T236 (T23) and Y238 (Y25) of Chain B; the main chain of K214 (K1) of Chain A forms a hydrogen bond with V241 (V28) of Chain B; and R216 (R3) of Chain A forms a salt bridge with E255 (E42) of Chain B (Fig. 3A-B), while R215 (R2) faces away from the interface and inserts into the DNA minor groove (Fig. 2F). The second interface features head-to-head interactions between the start of the third helices where A256 (A43) form symmetrical van der Waals interactions as well as the salt bridge between R216 (R3) of Chain A and E255 (E42) of Chain B (Fig. 3C, D). The third interface consists of the turn between helix 1 and helix 2 of Chain A and the N-terminal ARM of Chain B. Unlike in interface 1, R216 (R3) is inserted into the minor groove and forms no polar contacts with Chain A, whereas K213 (K -1) and K214 (K1) lack sufficient electron density to model their side chains. Hence, these three residues have little involvement in the protein-to-protein interface. Instead, R215 (R2) of Chain B forms hydrogen bonds with Y238 (Y25) and V241 (V28) of Chain A. In addition, N217 (N4) of Chain B forms a hydrogen bond with E245 (E32) of Chain A (Fig. 3E, F). Thus, while interfaces 1 and 3 use symmetrical regions of the ALX4 protein, the residues utilized to facilitate these interactions are vastly different.

R215 and R216 are critical for cooperative DNA binding to P3 sites

Given their chain-specific protein-protein and protein-DNA contributions, we next wanted to evaluate the roles that R215 (R2) and R216 (R3) have in cooperativity and DNA binding. We designed independent R215A (R2A) and R216A (R3A) mutations as well as a R215A;R216A (R2A;R3A) double mutant in the ALX4 DNA binding domain and evaluated their ability to cooperatively bind to the P3 site. Note, each ALX4 mutation tested in this manuscript was found to not significantly decrease protein stability as compared to wild type ALX4 in thermal stability assays unless specifically noted (Table 1). Through quantitative EMSAs, we found that R215A (R2A) and R216A (R3A) reduced cooperativity 5-fold and 10-fold, respectively (Fig. 3G; all replicates shown in Supplementary Fig. 8A). Moreover, the effects of the N-terminal ARM mutations compound as the double mutant, R215A;R216A (R2A;R3A), reduced cooperativity 20-fold (Fig. 3G).

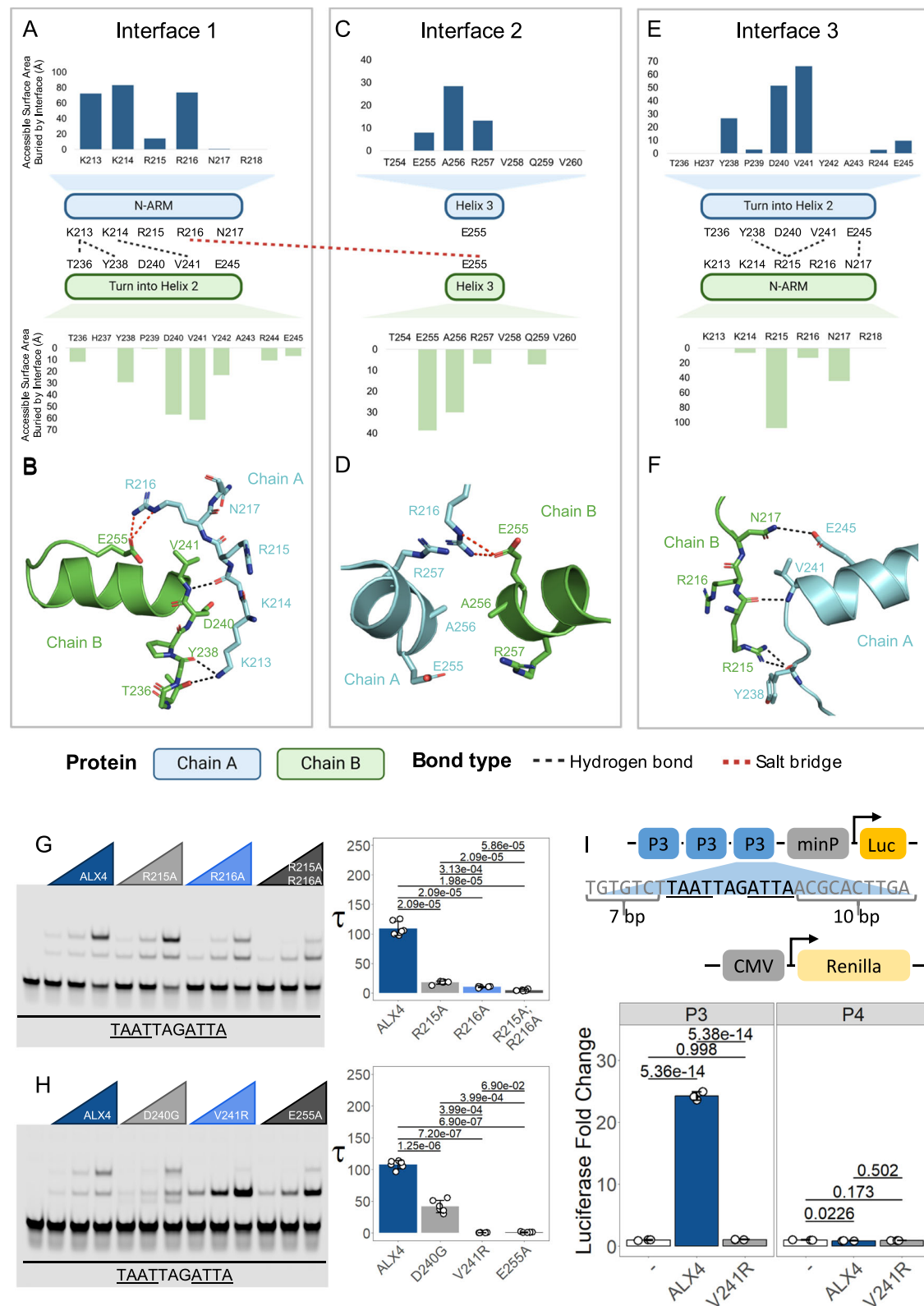
Next, we assessed the role of the ALX4 R215 (R2) and R216 (R3) residues in DNA binding affinity to a monomer binding site using isothermal titration calorimetry (ITC). Intriguingly, we found that the R215A (R2A) and R216A (R3A) mutations that significantly decreased cooperative DNA binding (Fig. 3G) did not significantly impact ALX4's ability to bind DNA (Table 2; Supplementary Fig. 12). However, the double mutant decreased DNA binding affinity -14-fold (Table 2; $p = 4.76 \times 10^{-10}$). The lack of affinity changes in the R215A (R2A) and R216A

(R3A) mutations supports the idea that either arginine residue can contribute to monomeric DNA binding by making similar minor groove contacts (Fig. 2F, G). However, the simultaneous mutation of both positions results in a loss of the ATTA minor groove DNA contacts, leading to a substantial loss in DNA binding affinity.

To test the role of protein-protein interactions in cooperative DNA binding outside of the N-terminal ARM, we next independently mutated three residues at the protein-protein interface: D240 (D27) which forms van der Waals interactions with the N-terminal arm of Chain A; V241 (V28) that forms hydrogen bonds and nonpolar interactions with either the R215 (R2) or R216 (R3) residues of the N-terminal ARM of the opposite chain; and E255 (E42) which mediates the salt bridge with R216 (R3) of Chain A. We found that a D240G (D27G) mutation caused a two-fold decrease in cooperativity ($\text{Tau} = 41.8 \pm 9.66$); a V241R (V28R) mutation, which inserts a larger charged residue predicted to sterically clash with the N-terminal ARM, disrupted cooperativity and even hindered the ability of the second protein to independently bind to the second site as reflected by the Tau value being less than 1 ($\text{Tau} = 0.351 \pm 0.330$); and E255A (E42A), which removes the side chain that forms the salt bridge, fully disrupted cooperativity ($\text{Tau} = 1.070 \pm 0.750$) (Fig. 3H; all replicates shown in Supplementary Fig. 8B). Overall, these data are consistent with a prior study showing that an I28R mutation disrupts cooperativity of the Prd S50Q protein to a P3 site¹⁵, and a recent genomic study that found that a pathogenic variant associated with cone rod dystrophy, CRX E80A (E42A), disrupts cooperative binding, leading to dysregulation of select genes involved in late stage photoreceptor development³⁵.

ALX4 mediated transcriptional activation via P3 sites is cooperativity dependent

Prior studies showed that ALX4 activates synthetic target sites containing 3 iterative - P3 sites in a spacer-dependent manner in luciferase reporter assays^{11,18}. To determine if transcriptional activation was cooperativity dependent rather than only DNA site dependent, we multimerized and cloned either 3 high affinity P3 sites (3xP3) or P4 sites (3xP4) in front of a minimal promoter and luciferase gene (schematic shown in Fig. 3I) and tested the ability of ALX4 wildtype and V241R (V28R) to activate gene expression in HEK293T cells. Importantly, the V241R (V28R) mutation did not significantly decrease monomer DNA binding affinity in ITC assays (Table 2), highlighting that this mutation selectively disrupts cooperativity. As expected, wild type ALX4 activated the 3xP3 reporter but failed to activate the non-cooperative 3xP4 reporter (Fig. 3I), demonstrating dimer site-dependent activation. Moreover, we found that the cooperativity-deficient ALX4 V241R (V28R) protein failed to activate either the 3xP3 or 3xP4 reporter (Fig. 3I). Comparative Western blot and immunostaining assays revealed similar protein levels and nuclear localization of the wild type and V241R (V28R) ALX4 proteins in cell culture (Supplementary Fig. 8D-E). Taken together, these data demonstrate that ALX4 transcriptional activation is cooperativity dependent.



The ALX4 dimer structure combines interactions found in the PRD and Engrailed structures

To better understand how the protein-protein and protein-DNA interactions in the ALX4 structure compare with other HD/DNA complexes, we analyzed the Prd S50Q/DNA dimer structure on a similar P3 site¹⁵, the Engrailed HD/DNA monomer structure⁵, and the Aristaless (al) HD/DNA monomer structure³⁶ (Supplementary Fig. 9). Overall, DNA backbone interactions as well as the base contacts between the

third helix and the major groove of DNA across these structures were very similar (Supplementary Fig. 9). However, the N-terminal ARM interactions were distinct between them due in large part to which Arg residue is inserted into the minor groove of DNA. Both the En and al monomer structures use the R3 residue to facilitate minor groove binding (Supplementary Fig. 9), whereas the Paired S50Q protein uses the R2 residue to mediate highly similar DNA interactions thereby freeing the R3 residue to mediate largely symmetrical protein-protein

Fig. 3 | ALX4 cooperativity is facilitated by three protein-protein interfaces. **A, C, E** Van der Waals interactions were calculated via the total accessible surface area of the residue that is buried by the other protein chain and are shown in the bar graphs. Polar contacts are shown via dotted lines between interacting residues; hydrogen bond contacts are shown with black dotted lines, whereas salt bridge contacts are shown via red dotted lines. **B, D, F** Protein structure views of the protein-protein interfaces. **A, B** The N-terminal ARM (N-ARM) of Chain A contacts the turn between helix 1 and 2 as well as helix 3 of Chain B. **C, D** The start of both recognition helices forms a symmetrical interface. **E, F** The N-ARM of Chain B forms contacts with the turn between helix 1 and 2 of Chain A. **G, H** ALX4 protein (0, 50, 100, and 200 nM) was tested on the P3 fluorescent probe. Tau cooperativity factors were calculated for every lane in which protein was added. Bars depict the average Tau for each protein and each dot represents a Tau from an independent binding reaction. Error bars denote standard deviation. Tau factors ($n = 6$ binding reactions;

biological replicates) were compared with a one-way ANOVA with a Holm-Bonferroni post-hoc correction. Note, quantitation was derived from two gels from the same experiment that were processed in parallel. **G** R215A (R2A), R216A (R3A), and R215A; R216A (R2A; R3A) mutations negatively affect ALX4 cooperativity. **H** D240G (D27G), V241R (V28R), and E255A (E42A) reduce ALX4 cooperativity. **I** ALX4 activates transcription in a cooperativity dependent manner. Luciferase assays were performed in transfected HEK293T cells. Bars denote average luciferase fold change of sample compared to reporter alone and dots represent an independent transfected well ($n = 3$ transfected wells; biological replicates). Error bars denote standard deviation and luciferase fold changes between proteins were compared via a one-way ANOVA with a Tukey post-hoc correction. A single replicate of each EMSA is shown and all replicates are in Supplementary Fig. 8. **A–F** Created in BioRender. Cain, B. (2025) <https://BioRender.com/z651674>. Source Data are provided on Figshare⁶⁹.

Table 1 | Melting temperatures derived from differential scanning fluorimetry for ALX4 structural mutations and ALX4 disease variants

Protein	Melting temperature	
	°C	P value
ALX4	58.4 ± 0.05	–
R215A	59.4 ± 0.12	0.013
R216A	59.5 ± 0.16	6.18e-3
R215A; R216A	59.2 ± 0.39	0.065
V241R	61.4 ± 0.08	3.38e-10
K211E	58.8 ± 0.12	0.806
R216G	59.2 ± 0.05	0.045
R218Q	58.9 ± 0.05	0.576
Q225E	49.3 ± 0	2.33e-14
R272P	56.8 ± 0.21	1.54e-5

Proteins were tested in triplicate, and melting temperatures are reported as mean ± standard deviation. Melting temperatures were compared via a one-way ANOVA with a Tukey post-hoc correction. The p-value of the variant compared to the wildtype ALX4 protein is shown. The melting curves are shown in Supplementary Fig. 13A. Source Data are provided on Figshare⁶⁹.

interactions between the two Chains (Fig. 4B, D). Intriguingly, the mechanisms used by ALX4, especially by the ALX4 N-terminal ARM, are a blend of both the Prd dimeric structure and the Engrailed and *al* monomeric structures. The protein-DNA interface at Chain A (Fig. 4A, C) resembles the interface found in the Paired dimeric structure (Fig. 4B, D, F), in which R216 (R3) interacts with E255 (E42) rather than inserting into the minor groove. In this case, R215 (R2) compensates for R216 (R3) to make the necessary protein-DNA contacts. The protein-DNA interface at Chain B (Fig. 4E) resembles the Engrailed and *al* monomeric structures (Supplementary Fig. 9), where R216 (R3) inserts into the minor groove and mediates the necessary contacts with DNA while R215 (R2) makes negligible DNA contacts. In this interface, the protein-protein interface is unique to ALX4 as R2 and N4 make no protein-protein contacts and minimal van der Waals interactions in the Prd S50Q structure (Fig. 4E, F). Overall, these data highlight how the flexible use of which arginine residue in the N-terminal ARM is used to contact DNA can allow for other N-terminal ARM residues to contribute to cooperative DNA binding through either symmetric (Paired S50Q) or asymmetric (ALX4) interactions.

ALX4 variants associated with disease differentially impact cooperativity and DNA binding

With our improved understanding of the residues implicated in DNA binding and cooperativity, we next sought to classify the functional impacts of ALX4 missense variants associated with disease. There are three conditions associated with ALX4 missense variants: 1) Enlarged parietal foramina (PFM), 2) Sagittal craniosynostosis (CS), and 3) Frontonasal dysplasia (FND). Despite the large range of phenotypes

and severities, limited functional characterization has been performed. Here, we assessed how ALX4 disease variants impact ALX4 function by measuring cooperativity with quantitative EMSAs (Supplementary Fig. 10), DNA binding with ITC assays (Table 2), protein stability with differential scanning fluorimetry assays (Table 1), and transcriptional activity with luciferase assays in transfected HEK293T cells. The expression of each ALX4 protein in HEK293T cells was confirmed via western blot (Supplementary Fig. 10G), and protein translocation into the nucleus was evaluated via immunofluorescence (Supplementary Fig. 11). In total, we characterized five ALX4 missense variants: four PFM variants (R216G, R218Q, Q225E, and R272P), one CS variant (K211E), and two FND variants (R216G and Q225E) (Fig. 5A). Note, the R216G and Q225E variants have been reported to cause more than one disease.

The ALX4 R216G (R3G) allele has been associated with both enlarged PFM and FND²⁵. Here, we found that R216G (R3G) reduces cooperativity 10-fold in EMSAs (Fig. 5B) and this variant has little impact on DNA binding affinity in ITC assays (Table 2), much like the R216A (R3A) variant tested above (Fig. 3G). The loss of cooperativity can be explained by the ALX4 crystal structure as R216 (R3) forms salt bridges with E255 (E42) on Interface 1 (Fig. 3A). Further, the preservation of DNA binding is consistent with R215's (R2) ability to insert into the minor groove of DNA to form the necessary base pair contacts and compensate for the loss of R216 (R3) (Fig. 2F). Consistent with the importance of cooperativity in ALX4-mediated transcriptional activation, we found that the R216G variant resulted in a 10-fold reduction in luciferase output on the 3xP3 reporter (Fig. 5H). Intriguingly, however, immunostaining of the ALX4 R216G variant revealed decreased nuclear localization in comparison to wild type ALX4, with the ALX4 R216G protein being observed in the cytoplasm and nucleus (Fig. 5C). NLSMapper³⁷ predicted that the N-terminal ARM is the sole nuclear localization signal in ALX4. These data are consistent with previous studies of other HD proteins, including close homologs, CART1 and ARX, that found that the basic residues in the N-terminal ARM can function as a nuclear localization signal^{38–43}. Hence, the ALX4 R216G variant results in two molecular defects: decreased nuclear localization and decreased cooperative DNA binding to the P3 dimer site, both of which likely contribute to the reduction of ALX4-mediated activation on the P3 site.

The ALX4 Q225E (Q12E) variant is associated with PFM and FND²¹. We found that this variant reduced cooperativity ~2-fold (Fig. 5D) with little change in DNA binding affinity (Table 2). This alteration in cooperativity modestly reduced transcriptional output on the P3 site (Fig. 5H). A recent protein binding microarray found that the ALX4 Q225E variant weakens DNA binding but preferentially maintains binding to "AATAAA" 6mers⁴⁴. The authors suspected that this residue may perturb the hydrophobic core due to Q225E's proximity to residues within the core (Fig. 5E). In support of this idea, we found that the Q225E variant decreased protein stability in differential scanning fluorimetry compared to the wildtype protein (Fig. 5F), while

Table 2 | Calorimetric binding data of ALX4 structural mutations and ALX4 disease variants on the 14mer oligo duplex: CGCTAATTAGCTCG

Protein	N	ΔG°	ΔH°	$-\Delta S^\circ$	K	K_d	P value
		kcal/mol	kcal/mol	kcal/mol		nM	
ALX4	1.08 ± 0.01	-10.5 ± 0.04	5.9 ± 0.08	-4.7 ± 0.09	6.47 ± 0.40	15.5 ± 0.98	–
R215A	1.00 ± 0.04	-10.3 ± 0.41	-5.3 ± 0.56	-5.0 ± 0.22	4.43 ± 3.00	30.4 ± 18.51	0.964
R216A	1.01 ± 0.02	-10.6 ± 0.18	-6.16 ± 0.08	-4.4 ± 0.22	7.02 ± 2.16	15.2 ± 4.63	1.000
R215A; R216A	1.02 ± 0.01	-9.0 ± 0.11	-4.0 ± 0.23	-5.0 ± 0.33	0.48 ± 0.08	214.0 ± 39.75	4.76e-10
V241R	0.99 ± 0.10	-11.1 ± 0.41	-6.5 ± 0.38	-4.6 ± 0.43	18.4 ± 11.5	7.45 ± 5.22	0.999
K211E	1.05 ± 0.11	-10.4 ± 0.21	-6.0 ± 0.50	-4.3 ± 0.71	5.03 ± 1.60	21.5 ± 7.68	1.000
R216G	1.01 ± 0.08	-10.0 ± 0.11	-5.73 ± 0.36	-4.2 ± 0.31	2.46 ± 0.47	41.8 ± 8.81	0.581
R218Q	–	–	–	–	–	–	–
Q225E	1.01 ± 0.05	-9.8 ± 0.15	-6 ± 0.39	-3.8 ± 0.51	1.99 ± 0.52	52.5 ± 12.29	0.192
R272P	0.99 ± 0.04	-10.0 ± 0.17	-6.3 ± 0.07	-3.7 ± 0.21	2.42 ± 0.71	43.6 ± 12.11	0.500

All proteins were tested in triplicate, and measurements are reported as mean ± standard deviation. Affinities (K_d) were compared via a one-way ANOVA with a Tukey post-hoc correction. The p value of the variant compared to the wildtype ALX4 protein is shown. The thermograms are shown in Supplementary Fig. 12. (N: stoichiometry; ΔG° : Gibb's free energy; ΔH° : enthalpy; T: temperature (K); ΔS° : entropy; K: association constant; K_d : dissociation constant). Source Data are provided on [Figshare](#)⁶⁹.

maintaining proper secondary structure content as measured by circular dichroism (Supplementary Fig. 13B). However, the question of how this destabilization leads to a biased retention of specific DNA sequences has not yet been addressed.

The ALX4 R218Q (R5Q) allele is a well-studied variant associated with PFM^{19,22,23}. We found that this variant abolishes all DNA binding in ITC assays (Fig. 5G) as well as in EMSAs (Supplementary Fig. 10B). Consistent with these results, the ALX4 R218Q variant failed to activate the 3xP3 reporter (Fig. 5H), even though it was properly localized to the nucleus (Supplementary Fig. 11). The complete loss of DNA binding is predicted by the structure as R218 (R5) forms contacts in the minor groove of DNA. Further, the R5 HD residue is highly conserved across most HD TFs¹⁰ and is considered a disease variant hotspot for HDs^{23,45}

We also tested the K211E (K–3E) variant that has been associated with increased penetrance of CS²⁶ and the R272P (R59P) variant that has been associated with PFM²⁴. We found no substantial reductions in cooperativity, DNA binding affinity, protein stability or transcriptional output for these variants (Fig. 5H; Table 1; Table 2). Interestingly, our results contradict previous findings that K211E and R272P increased and decreased transcriptional output, respectively. However, this discrepancy may be cell-type specific as we used HEK293T cells, whereas past studies used calvarial osteoblasts²⁶. It is also important to note that these variants occur on surface accessible residues on the protein. Thus, while these residues did not mediate protein-to-protein interactions in the ALX4 dimer structure, they could be implicated in protein-protein interactions with other co-factors.

ALX4 variants can be inherited in autosomal dominant or recessive patterns based on the disease condition rather than the variant with enlarged PFM having a dominant inheritance pattern and FND having a recessive pattern. Hence, the same ALX4 pathogenic variant can cause PFM in a dominant manner and FND in a recessive manner even within the same family. Since 4 of the 5 HD missense variants characterized here can cause a condition with an autosomal dominant inheritance pattern (R216G, R218Q, Q225E, and R272P), we tested the ability of these disease variants to alter wildtype ALX4 protein function using a combination of DNA binding and transcriptional reporter assays. For the DNA binding studies, we first purified a longer wildtype ALX4¹⁶⁹⁻³⁰³ protein that has a distinct binding shift in EMSAs, and we titrated in either a shorter wildtype ALX4 protein or shorter ALX4 proteins encoding each pathogenic variant to assess their ability to heterodimerize and/or compete with the longer wildtype ALX4 protein for a P3 binding site (Fig. 6A). As expected, increasing the concentrations of the shorter wildtype ALX4 protein reveals that it can both heterodimerize with and compete with the larger ALX4¹⁶⁹⁻³⁰³ homodimer protein complex for P3 binding sites (Fig. 6A-B). In contrast, the

ALX4 R218Q variant that fails to bind DNA on its own was unable to bind with the wildtype protein and failed to disrupt ALX4¹⁶⁹⁻³⁰³ binding to DNA (Fig. 6A-B), consistent with the idea that ALX4 dimer formation and cooperativity on P3 sites is DNA mediated. Each of the remaining pathogenic variants was similarly tested (see gels in Supplementary Fig. 14), and these EMSAs revealed that ALX4 R216G and Q225E variants were both able to heterodimerize and compete with the ALX4¹⁶⁹⁻³⁰³ but less so compared to wildtype ALX4. These findings are consistent with the R216G and Q225E variants being able to bind DNA but having lower cooperativity on P3 sites than wildtype ALX4. Lastly, the ALX4 K211E and R272P variants were able to heterodimerize and disrupt the ALX4¹⁶⁹⁻³⁰³ dimer similarly to the wildtype protein (Fig. 6B), consistent with their ability to cooperatively bind the P3 site much like wildtype ALX4.

Next, we assessed the impact of co-expressing the ALX4 disease variants with wildtype ALX4 in transcriptional reporter assays. For this study, we transfected HEK293T cells with a low constant amount of ALX4 wildtype construct that induces approximately a 10-fold increase in transcriptional activity on its own (black bar in Fig. 6C). We then titrated increasing amounts of constructs encoding either additional wildtype ALX4 or the five pathogenic variants. These studies revealed the following: the ALX4 K211E and Q225E constructs can synergize with ALX4 wildtype but to a lesser extent than wildtype ALX4, and ALX4 R272P can activate to a higher magnitude compared to wildtype (Fig. 6C). Note, these results were very similar to our findings when the ALX4 K211E, Q225E, and R272P constructs were transfected into HEK293T cells in the absence of wildtype ALX4 (Fig. 5H), suggesting that the co-expression of these variants with wildtype ALX4 does not dramatically alter P3-dependent transcriptional outcomes. In contrast, titrating in ALX4 R218Q or ALX4 R216G had little impact on the baseline level of activity created by the 187.5 picograms of wildtype ALX4. Although at high concentrations, ALX4 R218Q caused a slight, but significant reduction in the baseline wildtype activity (Fig. 6C). These results are consistent with ALX4 R218Q leading to more severe cases of enlarged PFM than ALX4 pathogenic nonsense variants²³. However, our EMSA data suggest that this loss in activity is not the result of direct competition for DNA binding sites (Fig. 6A-B). The inability of the ALX4 R216G variant to synergize with wildtype ALX4 to activate higher levels of expression is consistent with the weakened cooperative binding mediated by the R216G variant and the low maximum activation induced by ALX4 R216G alone (Fig. 5H). Taken together, our findings that ALX4 variants do not dramatically interfere with wildtype ALX4 DNA binding and

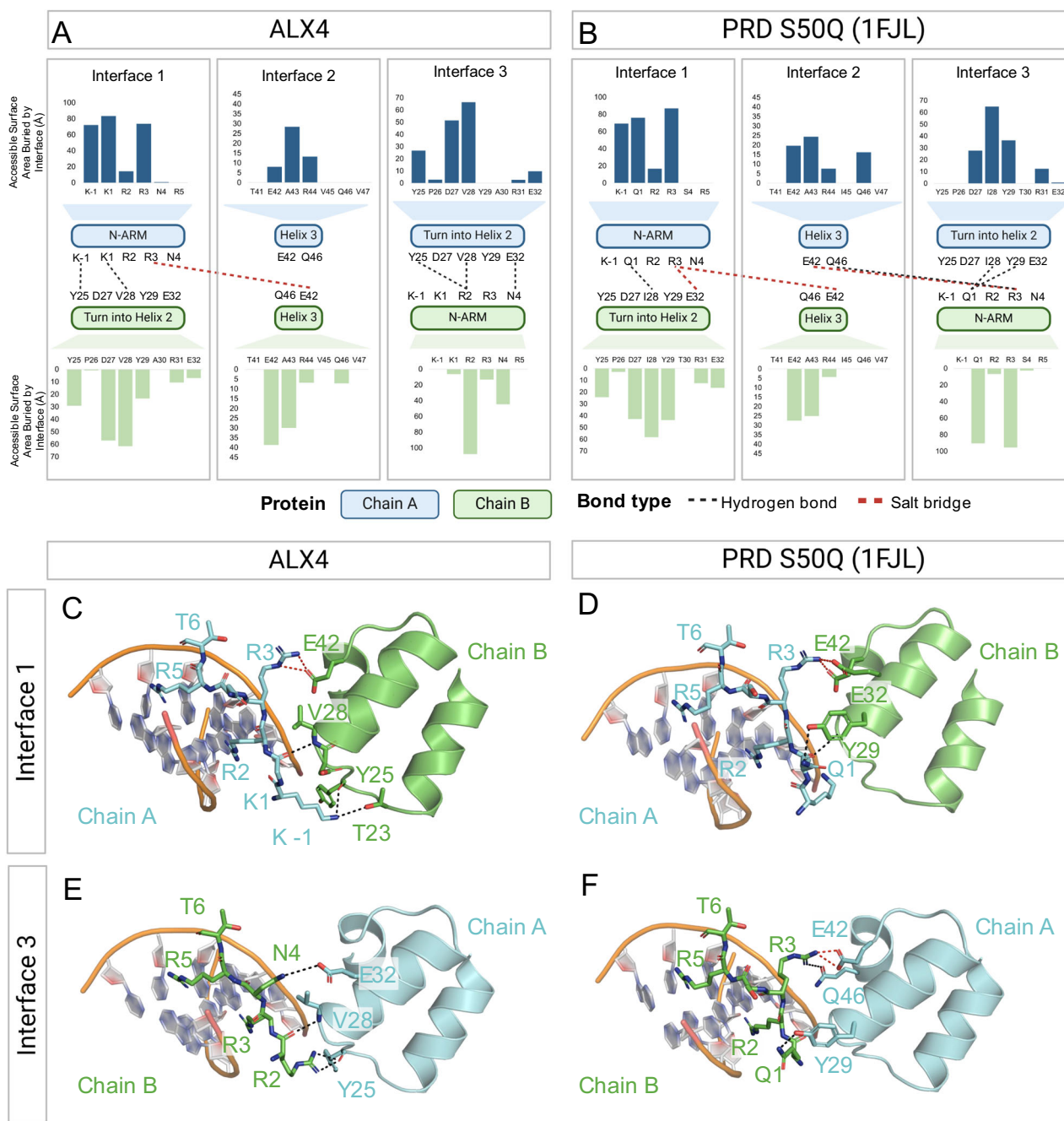


Fig. 4 | Comparison of ALX4 and PRD S50Q structures reveals asymmetric N-terminal ARM interactions in ALX4 but not PRD. A, B Protein-protein interactions of the ALX4 and PRD S50Q structure with the canonical HD numbering for comparison. Van der Waals interactions are shown in the bar graphs and are measured by the residue's accessible surface area buried by the other chain. Polar bond contacts are shown as dashed lines between contacting residues with hydrogen bonds in black and salt bridges in red. **A, C** In Chain A, K-1 forms bonds with T23 and Y25, K1 contacts V28, and R3 contacts E42. R2 inserts into the minor groove. **B, D** This interface is quite similar to the PRD S50Q structure as Q1

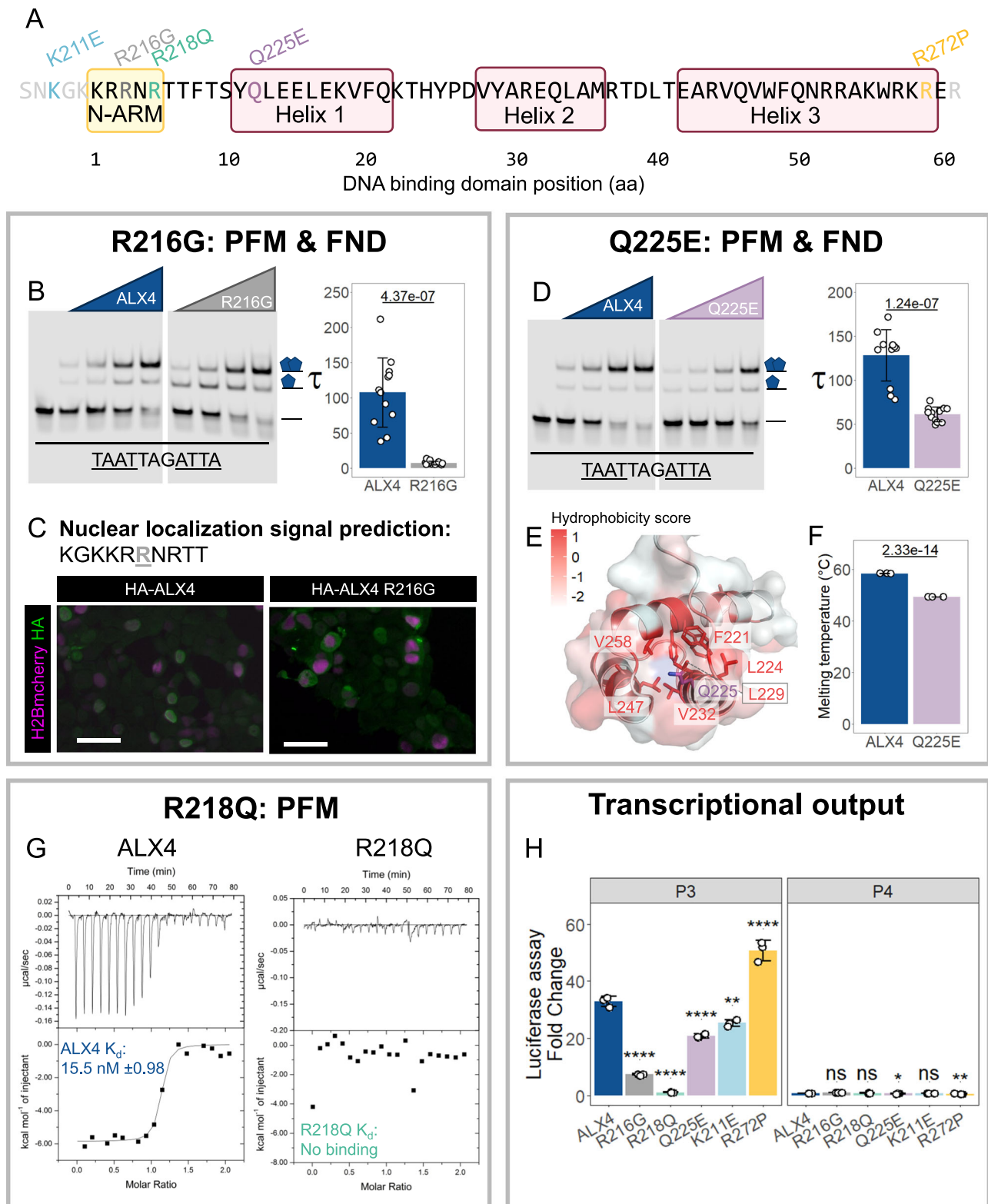
interfaces with I28 and Y29 in the turn between helix 1 and 2, and R3 interfaces with E32 and E42 in helix 2 and 3. **A, E** K-1, K1 and R3 of Chain B have limited involvement in the protein-protein interface as R3 inserts into the minor groove. Instead, R2 contacts the main chains of Y25 and V28 in the turn between helix 1 and 2, and N4 interfaces with E32 in the second helix. **B, F** This interface is not consistent with the PRD structure as Q1 and R3 are again the primary facilitators of the protein-protein interactions between the N-terminal ARM and the turn between helix 1 and 2 as well as the beginning of helix 3. Created in BioRender. Cain, B. (2025) <https://BioRender.com/s80m002>. Source Data are provided on Figshare⁶⁹.

transcriptional activation and in some cases can even form dimers with wildtype ALX4 support the idea that the pathological ALX4 variants are more likely to behave as loss-of-function or hypomorphic proteins than as dominant-negative proteins.

Discussion

In this work, we defined the mechanisms underlying how the Paired-like factor, ALX4, binds cooperatively to the TAAT-NNN-ATTA (P3)

dimer site and assessed how known disease variants impact DNA binding and transcriptional regulation. Using available genomic binding data, we first found that ALX4 binds several distinct motifs and of these only ALX4 binding to the P3 dimer site is independent of the craniofacial master regulator, TWIST1. This contrasts with ALX4's ability to bind the heterodimeric coordinator motif and independent monomer sites, both of which were dependent on TWIST1 occupancy and/or TWIST1 induced accessibility. Second, we used crystallography



to solve the ALX4 structure on a P3 site, providing insights into the protein-DNA and protein-protein interactions that mediate ALX4 cooperativity and DNA binding specificity. Third, we characterized the functional impacts of five ALX4 disease variants associated with PFM, FND, or increased penetrance of CS using a combination of quantitative DNA binding and transcriptional reporter assays. Collectively, these studies advance our understanding of the molecular mechanisms underlying HD DNA binding specificity and the molecular defects caused by disease-associated ALX4 alleles.

The diverse and flexible roles of the N-terminal ARM in mediating cooperativity and DNA binding specificity

How HDs gain sufficient DNA binding specificity to regulate distinct *in vivo* targets while using highly conserved DNA binding domains with highly similar *in vitro* DNA binding preferences has been a long-standing question in the field. Cooperativity addresses one part of this paradox as dimerization increases TF-DNA binding specificity in three ways. (i) Cooperativity typically involves binding longer DNA binding sites (i.e., the 11 bp P3 dimer site) that are less likely to occur at random

Fig. 5 | ALX4 human disease variants impact DNA binding and/or cooperativity which affect transcriptional output on a P3 site. A Five missense variants located within or near the DNA binding domain were characterized. **B, D** ALX4 and the corresponding variant protein (0, 37.5, 75, 150, and 300 nM) were combined with the P3 fluorescent probe. A single replicate is shown but all replicates are shown in Supplementary Fig. 10. Tau cooperativity factors were calculated for every lane in which protein was added. Bars depict the average Tau for each protein and each dot represents a Tau from an independent binding reaction ($n = 12$ binding reactions; biological replicates). Error bars denote standard deviation. Tau factors were compared with an unpaired, two-sided student t-test. Note, quantitation was derived from two gels from the same experiment that were processed in parallel. **C** ALX4 R216G (R3G) hindered protein's ability to localize to the nucleus in HEK293T cells. H2B-mcherry was used as a transfection control. NLS Mapper³⁷ predicted the only ALX4 nuclear localization signal to be the N-terminal ARM. Scale bar represents 50 μ M. **E** The Q225E (Q12E) variant is located near the hydrophobic

core of ALX4. Protein schematic was created in PYMOL and residues are colored based on hydrophobicity. **F** ALX4 Q225E reduced protein stability. Melting temperatures derived from differential scanning fluorimetry were compared via a one-way ANOVA with a Tukey post-hoc correction. The p value of the variant compared to the wildtype ALX4 protein is shown. The melting curves are shown in Supplementary Fig. 13A. **G** ALX4 R218Q (R5Q) failed to bind the 14mer oligo duplex, CGCTAATTAGCTCG, in ITC whereas the wildtype ALX4 protein bound the sequence with an affinity of 15.5 nM. **H** Luciferase assays were performed in transfected HEK293T cells ($n = 3$ transfected wells; biological replicates). Bars denote average luciferase fold change of the sample compared to reporter alone, while each dot represents an independent transfected well. Error bars denote standard deviation and luciferase fold changes between proteins were compared via a one-way ANOVA with a Tukey post-hoc correction where * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Source data are provided on [Figshare](#)⁶⁹.

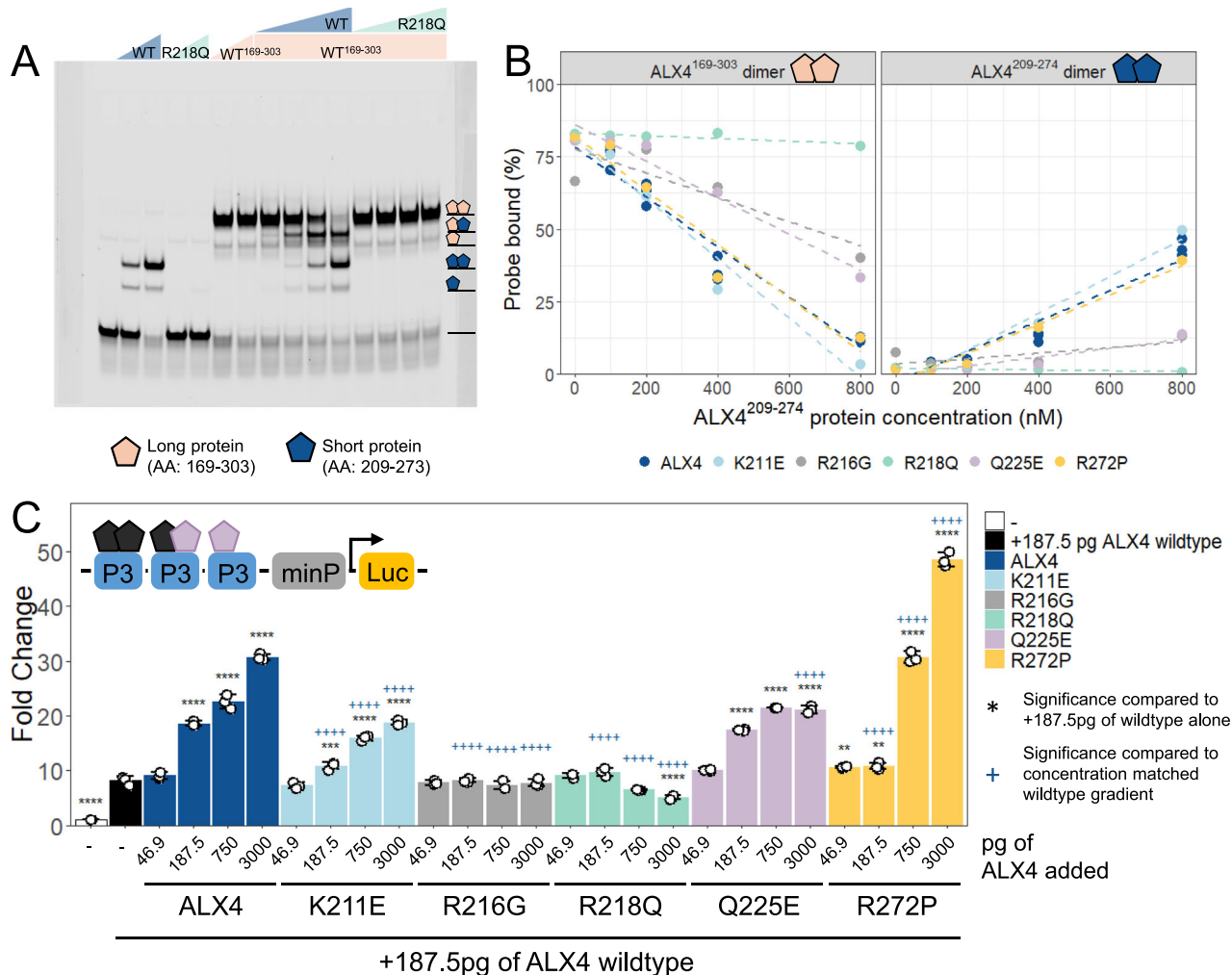


Fig. 6 | Select ALX4 pathogenic variants can bind DNA and transcriptionally synergize with wildtype ALX4. A Assessing pathogenic variants' heterodimerization capacity and ability to compete with wildtype ALX4 for a P3 DNA site via EMSA. Wildtype ALX4²⁰⁹⁻²⁷⁴ and ALX4 R218Q²⁰⁹⁻²⁷⁴, and ALX4¹⁶⁹⁻³⁰³ were tested at 200 and 400 nM individually to determine complex size (lane 2-7). Wildtype ALX4²⁰⁹⁻²⁷⁴ and ALX4 R218Q²⁰⁹⁻²⁷⁴ were then titrated (100, 200, 400, and 800 nM) with 400 nM of ALX4¹⁶⁹⁻³⁰³ (lane 8-15). Experiment was repeated twice with similar results. **B** Pathogenic variants have distinct capacities to disrupt the ALX4¹⁶⁹⁻³⁰³ dimer and form ALX4²⁰⁹⁻²⁷⁴ dimers. Percentage of probe bound by the ALX4¹⁶⁹⁻³⁰³ dimer and ALX4²⁰⁹⁻²⁷⁴ dimer across increasing concentrations of ALX4. All quantified gels are shown in Supplementary Fig. 14. **C** Luciferase assays were performed in

transfected HEK293T cells ($n = 3$ transfected wells; biological replicates) to assess the ability of ALX4 disease variants to synergize with wildtype ALX4 to induce transcriptional changes. Bars denote average luciferase fold change of the sample compared to reporter alone, while each dot represents an independent transfected well. Error bars denote standard deviation and luciferase fold changes between proteins were compared via a two-way ANOVA with a Tukey post-hoc correction where * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ for comparisons to 187.5 pg of ALX4 wildtype alone and * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ for comparisons to the concentration matched wildtype gradient. Source Data are provided on [Figshare](#)⁶⁹.

compared to a single 4 to 6 bp monomer site. (ii) A TF that cooperatively binds to a specific dimer site is likely to outcompete other TFs that non-cooperatively bind the two HD sites that compose the dimer motif. For example, we previously found that the HD, Gsx2, cooperatively bound a TAAT–7N–TAAT dimer site with higher affinity compared to its monomer site⁴⁶. This added affinity provides Gsx2 a selective advantage in binding this site over non-cooperative HDs, whereas ALX4 would have a similar advantage on the P3 dimer site. (iii) Since cooperativity typically involves protein-protein contacts, the number of residues that contribute to TF-DNA binding specificity increases. In Supplementary Fig. 9, we highlight the residues implicated in DNA side chain contacts for the monomeric Engrailed and aristaless structures. Unsurprisingly, the key residues: R3, R5, and N51 are highly conserved across all *Drosophila*¹⁰ and human HDs⁴⁴, consistent with most HDs binding the same monomer sites in *in vitro* assays^{2–4}. Here, we identified seven additional HD residue positions that indirectly contribute to ALX4 DNA specificity by facilitating P3 site cooperativity. These residues: K1, R2, N4, T23, V28, E32, and E42, are largely conserved across the Paired-like subclass⁴⁷, explaining why other Paired-like factors, including ALX1 and ALX3 (Supplementary Figs. 6C, D), have similar cooperative DNA binding behaviors^{11,14,48,49}. However, these residues have little conservation across the entire HD family, highlighting why only select HDs can bind the P3 site cooperatively¹⁴. Interestingly, AlphaFold3⁵⁰ correctly predicted that ALX4 would bind to the TAAT DNA binding sites within the P3 site, however AlphaFold3 docked the two proteins in a head-to-tail orientation rather than the correct head-to-head orientation (Supplementary Fig. 15). Because of this, AlphaFold would be unable to correctly predict the interfacing regions and residues, highlighting both the utility in solving the crystal structures of these complexes and the current limitations of using a prediction program to determine protein-protein interfaces.

An unexpected finding from the ALX4 dimer structure was the asymmetric use of the N-terminal ARM residues to mediate DNA-protein and protein-protein interactions. Given the palindromic nature of the P3 site (TAAT–3N–ATTA), one would have predicted that ALX4 would use similar mechanisms to bind each half site, much like was previously found for the Prd S50Q dimeric structure¹⁵. However, we found that the N-terminal ARM residues facilitating these interfaces are unique between the two structures. In the Prd S50Q structure, interfaces 1 and 3 mediate protein-protein interactions that are largely symmetrical and are facilitated by positions 1 and 3 in the N-terminal ARM (Fig. 4)¹⁵. In contrast, interface 3 in the ALX4 structure is driven by R2, N4, Y25, and E32, revealing the importance for these residues in P3 site cooperativity (Fig. 4E). R2, N4, and Y25 do not make protein-protein contacts in the Prd S50Q structure and were not previously thought to participate in cooperativity. Further, R2, Y25, and E32 are conserved across the Paired-like class, whereas N4 is not well conserved⁴⁷. Future studies will be required to address the importance of the fourth position in contributing to P3-mediated cooperativity.

Comparative studies across HDs reveal that there is some flexibility as to which arginine residue in the N-terminal ARM mediates minor groove DNA binding contacts. While all HDs use the highly conserved R5 residue to make DNA contacts, additional contacts are typically mediated by a second arginine residue at either position 2 or position 3. We found that the asymmetry between the two ALX4 proteins is largely driven by whether R2 or R3 inserts into the minor groove. Moreover, we showed that while both arginine residues are required for cooperativity (Fig. 3G), either arginine is sufficient for high affinity monomer DNA binding (Table 2). The preservation of DNA binding in R2A or R3A mutations suggests that either R2 or R3 can insert into the minor groove and compensate for the loss of the other for monomer binding, but by doing so, the sole arginine can no longer facilitate the protein-protein interactions required to maintain cooperativity. R2 and R3's ability to make similar DNA contacts has been

shown previously: A bacterial one-hybrid survey of 84 *Drosophila* HDs revealed that both positions have the potential to define the specificity of TAAT³³. This has been structurally confirmed as R3 makes DNA base contacts in Engrailed⁵, PITX2⁵¹, and aristaless³⁶ structures, whereas Msx-1 uses R2 to mediate similar contacts⁵². Interestingly, R2 and R3 are conserved throughout all the Paired-like class⁴⁷, but many other HDs have a lysine in one of these positions^{10,44}. While lysine has a similar charge as arginine, lysine residues insert into DNA minor grooves ~3-fold less than arginine residues⁵³ and appear less frequently in protein-protein interfaces than arginine residues⁵⁴. Thus, the structural studies on ALX4 and other HD proteins highlight how differences in key HD residues, especially within the N-terminal ARM, contribute to distinct DNA binding and cooperativity behaviors.

ALX4-induced transcriptional activation is dependent on TF cooperativity

ALX4 variants, including several missense variants within the HD, have been associated with craniofacial birth defects such as PFM, FND, and increased penetrance of CS. How such mutations disrupt the ability of ALX4 to bind DNA, regulate target gene expression, and ultimately cause disease is not well understood. Here, we used the insight gained from our ALX4 dimer structure to interrogate the ability of several HD missense alleles to bind cooperative dimer DNA binding sites as well as monomer sites. Combining quantitative DNA binding assays with a dimer-dependent transcriptional assay, we were able to identify molecular defects for three of five tested ALX4 HD missense variants. Based on their locations in the ALX4 crystal structure, it was not surprising that neither K211E nor R272P significantly impacted monomer or cooperative dimer binding. Moreover, both proteins were properly localized and activated transcription similarly as wildtype ALX4, making it unclear how these variants cause disease. In contrast, the other three disease variants have clear molecular defects that impact DNA binding. R218Q (R5Q) completely abolished monomer and dimer DNA binding, whereas R216G (R3G) and Q225E (Q12E) selectively decreased cooperative DNA binding. In addition, R216G and Q225E caused secondary deficits that are likely to contribute to their observed ~10-fold and ~2-fold reduced transcriptional output, respectively (Fig. 5H). R216G reduced protein nuclear localization (Fig. 5C) and Q225E impacted protein stability (Fig. 5E, F). Thus, these three ALX4 HD disease variants have distinct molecular defects that impact their ability to activate transcription via cooperative P3 binding sites.

Past work showed that ALX4 only transcriptionally activates reporter genes containing the P3 site, and not monomer, P2, P4, or P5 sites^{18,19}. Consistent with these data, we found that a cooperativity-deficient ALX4 protein, ALX4 V241R (V28R), could no longer transcriptionally activate the P3 site (Fig. 3I), despite maintaining its ability to bind DNA (Fig. 3H; and Table 2). The correlation between cooperativity and transcriptional output emphasizes the importance of cooperativity on ALX4 function and suggests that ALX4 requires a partner to enact a transcriptional change. A recent study found that ALX4 cooperatively binds with TWIST1 to regulate human craniofacial development²⁸, providing a second case in which ALX4 binds with a partner to induce a transcriptional change. In alignment with this idea, we found that ALX4 binding to the coordinator and HD monomer were dependent on TWIST1 (Fig. 1), whereas ALX4 binding to the P3 site was not dependent on TWIST1 binding (Fig. 1)^{18,19}. This behavior extends to close homologue, CART1 (sometimes referred to as ALX1) as it too specifically activates transcription on a P3 site¹⁸, and has high co-occupancy with TWIST1 in differentiated hCNCs²⁸. Other HDs have also been shown to alter transcriptional activity via homo-dimeric cooperativity. For example, CRX, another Paired-like HD protein, exhibited higher transcriptional changes in massively paralleled reporter assays on P3 dimer sites compared to monomer sites¹³. In addition, the Gsx2 HD protein was found to induce opposing transcriptional outcomes on dimer (gene stimulation) versus monomer

(repression) sites⁴⁶. Thus, cooperativity-dependent transcriptional activation is not unique to ALX4, but a common mechanism used by many HDs.

Methods

Genomic analyses

The ALX4 and TWIST1 genomic binding data were acquired from Kim et al.²⁸ and the accession IDs for these data are listed in Supplementary table 2. Sequencing data underwent adapter trimming with Cutadapt (v4.4)⁵⁵ and quality control with fastqc (v0.11.2)⁵⁶ via the wrapper trimalore (v0.6.6)⁵⁷. Sequences were mapped to hg19 with bowtie2 (v2.3.4.1) with the following settings: --local --very-sensitive-local --no-unal --no-mixed --no-discordant⁵⁸. Results were deduplicated with picard MarkDuplicates (v2.18.22)⁵⁹. MACS3 (v3.0.0) with the --keep-dup all --q value 0.01 options was used for peak calling of all genomic assays⁶⁰. For C&R peak calling, --call-summits --SPMR settings were also applied; for H3K27ac ChIP-seq peak calling, --broad --broad-cutoff 0.1 settings were also applied, and for ATAC-seq peak calling --call-summits --SPMR --shift -75 --extsize 150 settings were also applied. BigWigs and heatmaps were generated using deepTools (3.5.1)⁶¹. C&R and ChIP-seq bigwigs were normalized with the log2 ratio of the immunoprecipitated datasets versus the input or IgG controls. ATAC-seq bigwigs were normalized to reads per genomic content.

Homer (v4.9) de novo motif analysis identified 6, 12, and 18 bp length enriched sequences in the ALX4 C&R datasets³⁰. annotatePeaks.pl from the Homer package was used to annotate the coordinator, dimer, monomer, and E-box sites within the ALX4 bound regions. For footprint analysis, the genome coverage tool from bedtools (v2.27.0) calculated 5' read coverage of reads that were less than 120 bp in length⁶². We determined whether a predicted site was bound by comparing the MNase digestion signal of the motif (which should be protected when bound by a TF) to regions flanking the motif (which should be highly accessible). The set motif window for the coordinator, dimer, monomer, and E-box site was -10bp to 10 bp, -6bp to 6 bp, -3bp to 3 bp, and -3bp to 3 bp, respectively, where 0 is the motif center. The motif flank was ± 40 bp outside of the motif window. The following equation was used to determine if a site was bound, where M_f stands for the average 5' binding signal of the 40 bp flanking either side of the motif window and M_w stands for the average 5' binding signal of the motif window: $\frac{M_f + 0.01}{M_w + 0.01} > 1$ ⁴⁶. The 0.01 addition eliminated any errors introduced by dividing by zero. We used the GRanges package in R to remove any overlapping or palindromic repeat sites as well as any single sites that were part of a bound composite sites⁶³. Supplementary Fig. 3A shows the number of sites removed in each of the described steps.

To statistically compare the reads per million of the ALX4 C&R in wildtype conditions versus the ALX4 C&R in TWIST1 depleted conditions in Fig. 1E, ALX4 C&R in TWIST1 depleted conditions was normalized to the ALX4 C&R in wildtype conditions by multiplying the reads per million by a scaling factor derived by the *E.coli* spike-in control²⁸.

$$\text{Scaling Factor} = \frac{E.\text{coli alignment rate}_{\text{wildtype}}}{\text{Human alignment rate}_{\text{wildtype}}} / \frac{E.\text{coli alignment rate}_{\text{TWIST1depleted}}}{\text{Human alignment rate}_{\text{TWIST1depleted}}}$$

$$\text{Scaling Factor} = \frac{0.12}{96.04} / \frac{0.10}{96.79} = 1.21$$

Protein preparation

The mouse ALX4 sub-fragment was PCR amplified using Accuzyme DNA polymerase (Bioline) from mALX4 cDNA (Genscript: OMu19476) and cloned into pET14P. Note: while mouse cDNAs were used, the mouse and human ALX4 protein sequences are identical for the

regions used in this study. pET14P is a modified pET14b plasmid in which a PreScission Protease site was inserted between the 6xHis-tag and TF coding sequence¹⁴. PCR primers used for amplification include a C-terminal Strep-tag and are listed in Supplementary Table 3. ALX4 variant cDNAs were prepared via site directed mutagenesis. All DNA is available upon request. All pET14P-ALX4 constructs were transformed into C41(DE3) (Sigma-Aldrich) *E. coli*, grown in LB media, and induced with 0.1 mM IPTG. Following bacterial cell lysis, ALX4 proteins were purified using a combination of Ni-NTA and Strep-Tactin affinity, and size exclusion chromatography. The N-terminal 6xHis-tag was removed via digestion with the PreScission Protease (GE Healthcare) per manufacturer protocol. ALX4¹⁶⁹⁻³⁰³ was cloned and purified as previously described¹⁴. Briefly, the ALX4 169-303 pET14P construct was transformed into BL21 *E.coli*, grown in LB media, and induced with 0.5 mM of IPTG. Bacteria were lysed and protein was purified via Ni-NTA pulldown. Protein purity was confirmed via SDS-PAGE with Gel-Code blue staining (Thermo Scientific) (Supplementary Figs. 8C and 10F), and protein concentrations were assessed by absorbance at UV 280 nm.

Isothermal titration calorimetry

ITC experiments were performed using a Microcal VP-ITC microcalorimeter. For all experiments, DNA duplexes were placed in the syringe at -100 μ M, and all ALX4 proteins were placed in the cell at -10 μ M. Titrations consisted of an initial 1 μ l injection followed by nineteen 14 μ l injections. All experiments were performed in a buffer containing 50 mM sodium phosphate, pH 6.5 and 150 mM NaCl at 20 °C. All samples were dialyzed overnight to ensure buffer match. *c* values (*c*=K_{Mn}) were 200 < *c* < 702. All binding experiments were performed in triplicate. Final raw data were analyzed using ORIGIN and fit to a one-site binding model.

Crystallography

ALX4-DNA complexes were formed prior to crystallization by mixing purified protein in a 2:1 (protein:DNA) ratio with a final complex concentration of -10 mg/ml (-1 mM). The DNA used for crystallization was a 17-mer duplex, containing the sequence 5' - CGCTAATTCAAT-TAACG - 3' with two ALX4 binding sites (bold, underlined) spaced three base pairs apart. The ALX4-DNA complex crystallized in a solution containing 0.1 M Tris pH 7.5, 0.2 M Trimethylamine N-oxide, and 20% PEG MME 2000. Crystals diffracted to 2.39 Å resolution and belong to the space group P3₂21 with cell dimensions *a* = *b* = 70.46 *c* = 159.14 Å. The asymmetric unit of the crystal contained two copies of ALX4 and one copy of the DNA duplex. Unbound ALX4 was crystallized at 10 mg/ml in a solution containing 0.1 M MgCl₂, 0.1 M KCl, 0.1 M PIPES pH=7.0, 20% PEG Smear Medium. The crystals diffracted to 2.39 Å resolution and belong to the space group I2₁3 with cell dimensions *a* = *b* = *c* = 95.78 Å. All synchrotron diffraction data was collected at APS LS-CAT.

Structure determination, model building, and refinement

Phaser was used for molecular replacement with the PDB file 1FJL as the search model⁶⁴. COOT was used for manual model building within the observed electron density⁶⁵ and Phenix was used for refinement with TLS parameters⁶⁶. The final ALX4/DNA and ALX4 models were refined to a *R*_{work}/*R*_{free} = 22%/24% and 21%/25%, respectively, with good overall geometry and validated with MolProbity⁶⁷. Protein structure schematics were created by PYMOL (v2.5.7)⁶⁸. Accessible surface area buried by interface measurements and polar interactions were calculated by (PDBEPIA v1.52)³².

Circular dichroism

CD experiments were performed on an Aviv Circular Dichroism Spectrophotometer 215 using a 0.5 mm quartz cuvette (Hellma Analytics). The cuvette was not removed during a series of scans taken

from 300 to 190 nm in 1 nm increments. Proteins were dialyzed into a buffer containing 5 mM sodium phosphate (pH 6.5) and 150 mM NaF and diluted to the desired concentration of ~0.30 mg/ml in the same buffer. Data are plotted as mean residue ellipticity, θ , in units of deg cm²/dmol per residue.

Electrophoretic mobility shift assays

EMSA probes were prepared by annealing a 5'-IRDye 700 nM labeled oligo (IDT) to an oligo containing the binding sites of interest and filling in with Klenow (NEB). Sequences used for probe preparation are listed in Supplementary Table 4. All EMSA binding reactions were performed in the following binding buffer: 10 mM Tris, pH 7.5; 50 mM NaCl; 1 mM MgCl₂; 4% glycerol; 0.5 mM DTT; 0.5 mM EDTA; 50 µg/ml poly(dI-dC); and 200 µg/ml of BSA. 34 nM of denoted fluorescent probe and the denoted proteins and concentrations were incubated at room temperature for 20 minutes. For EMSAs in Fig. 2 and Fig. 5 where one protein was tested at a time, 4% polyacrylamide gels were run at 150 V for 2 h. For EMSAs in Fig. 6 where multiple proteins were tested at a time, 6.5% polyacrylamide gels were run at 150 V for 3 h. Gels were imaged on the Li-Cor Odyssey CLx scanner and band intensity was measured via the Li-Cor image studio software. All uncropped and unprocessed gels are provided in the Figshare⁶⁹. Tau was calculated by Eq. (1), where [P₂D] represents dimer binding, [PD] represents monomer binding, and [D] represents unbound probe¹². Equation (1) was derived by the equilibrium reactions of a single protein binding DNA as well as the equilibrium equation of the binding of a second protein to the monomeric complex.

$$\tau = \frac{4[P_2D][D]}{[PD]^2} \quad (1)$$

Luciferase assays

Full-length human ALX4 cDNA sequences of the wildtype and variants were cloned into a HA-tagged pcDNA3.1 mammalian expression vector by Genscript. 5 UAS sites, 3 P3 or P4 sites, and the minimal promoter from the *Drosophila* Ac gene were cloned between KpnI and NcoI sites of pGL3-basic (Promega). Note, while 3xP3 or 3xP4 sites were cloned 3' from the 5 upstream activation sequence (UAS) sites, no Gal4 was added in the transfections. Thus, the 5xUAS sites had no impact on luciferase output. cDNA and reporter sequences are provided in Supplementary Data 1. All DNA is available upon request.

Forty thousand HEK293T cells (ATCC: CRL-11268) were cultured in each well of a 48-well plate in DMEM and 20% fetal bovine serum for 18 h prior to transfection. Each well was transfected with 80 ng of total DNA: 5 ng of Renilla plasmid, 25 ng of indicated firefly reporter plasmid, and the denoted amount of ALX4 cDNA and pcDNA3.1 empty expression construct using the Effectene Transfection reagent (Qiagen) following manufacturer's protocol. Cells were lysed with 80 µL of passive lysis buffer (Promega) 30 h after transfection. Firefly luciferase and Renilla luciferase fluorescent values were measured with the Promega Dual Luciferase Assay kit (Promega: E1910) and the GloMax-96 Microplate Luminometer. Firefly luciferase was normalized to Renilla luciferase to control for transfection efficiency. All results are reported as expression fold change over the reporter alone sample. All samples were run in triplicate.

Western blot

Four hundred thousand HEK293T cells (ATCC: CRL-11268) were cultured in each well of a 6-well plate in DMEM and 20% fetal bovine serum for 18 h prior to transfection. Each well was transfected with 600 ng of ALX4 cDNA in pcDNA3.1 or a pcDNA3.1 empty expression construct with Effectene Transfection reagent (Qiagen) following the manufacturer's protocol. Cells were lysed by mechanically scraping cells in 750 µL of radioimmunoprecipitation assay (RIPA) buffer

(150 mM of NaCl, 50 mM of Tris HCl pH 8.0, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 1x cComplete Protease Inhibitor Cocktail (Roche: 04693116001), and 1 mM of phenylmethylsulfonyl fluoride). Lysates were incubated on ice for 5 minutes and then sonicated for 5 seconds with 25 seconds of rest time for a total of 3 sonication cycles.

Lysates were run on a 4–20% gradient acrylamide gel (BioRad) at 100 V for 90 minutes. Gels were transferred to a polyvinylidene difluoride (PVDF) membrane via a semi-dry transfer that was run for 1 h at 15 V at room temperature. Membranes were blocked in a 0.5% Casein solution for 1 h at room temperature. Membranes were probed with HA-tag (Roche; Catalog Number: 1-867-423; clone: 3F10; Dilution: 1:250) and β -actin (Li-Cor; Catalog number: 926-42212; Lot number: D01217-01; Dilution: 1:2000) overnight at 4 °C in PBS + 0.05% Tween-20 (PBST) + 0.05% Casein. The membrane was then washed 3 times with PBST and incubated with IRDye 680RD Goat anti-Rat (Li-Cor; Catalog number 926-68076; Lot number: D30124-25; Dilution: 1:5000) and IRDye 800CW Donkey anti-Mouse (Li-Cor; Catalog number: 926-32212; Lot number: C90805-13; Dilution: 1:5000) in PBST + 0.05% Casein for 1.5 h at room temperature. The membrane was washed 3 times in PBST and imaged on the Li-Cor Odyssey CLx scanner. All uncropped and unprocessed blots are provided on Figshare⁶⁹.

Cell immunofluorescence

Forty thousand HEK293T cells were cultured in each well of a 48-well plate in DMEM and 20% fetal bovine serum for 18 h prior to transfection. Each well was transfected with 40 ng of an H2B-mcherry expression vector to label transfected cells and 40 ng of the indicated ALX4 cDNA in pcDNA3.1. Thirty h after transfection, cells were washed with PBS and then fixed by incubating cells in 4% paraformaldehyde for 10 minutes at 4 °C. Cells were permeabilized in PBS + 0.5% Triton-X for 5 minutes at room temperature. Cells were blocked in PBS + 5% normal donkey serum (Jackson ImmunoResearch Laboratories Inc.: 017-000-121) for 1 h at room temperature and then probed with the HA-tag (Roche; Catalog Number: 1-867-423; clone: 3F10; Dilution: 1:4000) in PBS + 0.2% Tween-20 + 5% normal donkey serum overnight at 4 °C. Cells were washed three times in PBS + 0.2% Tween-20 and then incubated in Alexa Fluor 488 Donkey Anti-Rat (Jackson ImmunoResearch Laboratories Inc; Catalog number: 712-545-153; Lot number: 164432; Dilution: 1:1000) in PBS + 0.2% Tween-20 + 5% normal donkey serum for 1 h at room temperature. Cells were washed three times in PBS + 0.2% Tween-20. Cells were imaged on a Nikon ECLIPSE Ts2R Inverted Microscope. The LUT of the H2Bmcherry was changed from yellow to magenta in FIJI to allow for the discrimination of channels in the merged images.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The previously published genomic binding and accessibility data²⁸ analyzed in this study are available in the GEO database under accession code [GSE230319](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE230319). The X-ray crystallography data generated in this study have been deposited in the Protein Data Bank under accession codes [9D9R](https://www.rcsb.org/structure/9D9R) and [9D9V](https://www.rcsb.org/structure/9D9V). All biochemical data generated in this study are provided in the Supplementary Information and on Figshare <https://doi.org/10.6084/m9.figshare.28535987>.

Code availability

All custom code can be found at GitHub [https://github.com/cainbn97/ALX4_footprinting_analysis] or Zenodo <https://doi.org/10.5281/zenodo.14962844>⁷⁰.

References

- Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
- Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
- Berger, M. F. et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
- Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* (1979) **324**, 1720–1723 (2009).
- Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579–590 (1990).
- Crocker, J. et al. Low affinity binding site clusters confer HOX specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
- Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu Rev. Cell Dev. Biol.* **35**, 357–379 (2019).
- Zandvakili, A., Campbell, I., Gutzwiller, L. M., Weirauch, M. T. & Gebelein, B. Degenerate Pax2 and Senseless binding motifs improve detection of low-affinity sites required for enhancer specificity. *PLoS Genet* **14**, 1–25 (2018).
- Slattery, M. et al. Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell* **147**, 1270–1282 (2011).
- Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* **125**, 497–521 (2016).
- Tucker, S. C. & Wisdom, R. Site-specific heterodimerization by paired class homeodomain proteins mediates selective transcriptional responses. *J. Biol. Chem.* **274**, 32325–32332 (1999).
- Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. & Desplan, C. Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.* **7**, 2120–2134 (1993).
- Hughes, A. E. O., Myers, C. A. & Corbo, J. C. A massively parallel reporter assay reveals context-dependent activity of homeodomain binding sites in vivo. *Genome Res* **28**, 1520–1531 (2018).
- Cain, B. et al. Prediction of cooperative homeodomain DNA binding sites from high-throughput-SELEX data. *Nucleic Acids Res* **51**, 6055–6072 (2023).
- Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell* **82**, 709–719 (1995).
- Nitta, K. R. et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife* **4**, e04837 (2015).
- Jun, S. & Desplan, C. Cooperative interactions between paired domain and homeodomain. *Development* **122**, 2639–2650 (1996).
- Qu, S., Tucker, S. C., Zhao, Q. & Wisdom, R. Physical and genetic interactions between Alx4 and Cart1. *Development* **369**, 359–369 (1999).
- Qu, S. et al. Mutations in mouse *Aristaless-like4* cause strong's luxoid polydactyly. *Development* **125**, 2711–2721 (1998).
- Kayserili, H. et al. ALX4 dysfunction disrupts craniofacial and epidermal development. *Hum. Mol. Genet* **18**, 4357–4366 (2009).
- Kayserili, H., Altunoglu, U., Ozgur, H., Basaran, S. & Uygur, Z. O. Mild nasal malformations and parietal foramina caused by homozygous ALX4 mutations. *Am. J. Med Genet A* **158A**, 236–244 (2012).
- Valente, M., Valente, K. D., Sugayama, S. S. M. & Kim, C. A. Malformation of Cortical and Vascular Development in One Family with Parietal Foramina Determined by an ALX4 Homeobox Gene Mutation. *AJNR Am. J. Neuroradiol.* **25**, 1836–1839 (2004).
- Mavrogiannis, L. A. et al. Enlarged parietal foramina caused by mutations in the homeobox genes ALX4 and MSX2: From genotype to phenotype. *Eur. J. Hum. Genet.* **14**, 151–158 <https://doi.org/10.1038/sj.ejhg.5201526> (2006).
- Wuyts, W. et al. The ALX4 homeobox gene is mutated in patients with ossification defects of the skull (foramina parietalia permagna, OMIM 168500). *J. Med Genet* **37**, 916–920 (2000).
- Altunoglu, U., Satkin, B., Uygur, Z. O. & Kayserili, H. Mild nasal clefting may be predictive for ALX4 heterozygotes. *Am. J. Med Genet A* **164**, 2054–2058 (2014).
- Yagnik, G. et al. ALX4 gain-of-function mutations in nonsyndromic craniosynostosis. *Hum. Mutat.* **33**, 1626–1629 (2012).
- Wu, E., Vargevik, K. & Slavotinek, A. M. Subtypes of frontonasal dysplasia are useful in determining clinical prognosis. *Am. J. Med Genet A* **143A**, 3069–3078 (2007).
- Kim, S. et al. DNA-guided transcription factor cooperativity shapes face and limb mesenchyme. *Cell* **187**, 692–711 (2024).
- Prescott, S. L. et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Hovland, A. S. et al. Pluripotency factors are repurposed to shape the epigenomic landscape of neural crest cells. *Dev. Cell* **57**, 2257–2272.e5 (2022).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Noyes, M. B. et al. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277–1289 (2008).
- Tucker-Kellogg, L. et al. Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9Å resolution: structural basis for enhanced affinity and altered specificity. <http://biomednet.com/elecref/0969212600501047> (1997).
- Zheng, Y., Stormo, G. D. & Chen, S. Aberrant homeodomain-DNA cooperative dimerization underlies distinct developmental defects in two dominant CRX retinopathy models. *Genome Res* <https://doi.org/10.1101/gr.279340.124> (2024).
- Miyazono, K. I. et al. Cooperative DNA-binding and sequence-recognition mechanism of *aristaless* and *clawless*. *EMBO J.* **29**, 1613–1623 (2010).
- Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *PNAS* **106**, 10171–10176 (2009).
- Furukawa, K. et al. Functional domains of paired-like homeoprotein Cart1 and the relationship between dimerization and transcription activity. *Genes Cells* **7**, 1135–1147 (2002).
- Shoubridge, C., Cloosterman, D., Parkinson-Lawrence, E., Brooks, D. & Gécz, J. Molecular pathology of expanded polyalanine tract mutations in the *Aristaless*-related homeobox gene. *Genomics* **90**, 59–71 (2007).
- Shoubridge, C. et al. Mutations in the nuclear localization sequence of the *aristaless* related homeobox; sequestration of mutant ARX with IPO13 disrupts normal subcellular distribution of the transcription factor and retards. *Cell Division*. <http://www.pathogeneticsjournal.com/content/3/1/1> (2010).
- Ming Khor, J. & Ettensohn, C. A. Functional divergence of paralogous transcription factors supported the evolution of biomineralization in echinoderms. <https://doi.org/10.7554/eLife.32728.001> (2017).
- Ye, W., Lin, W., Tartakoff, A. M. & Tao, T. Karyopherins in nuclear transport of homeodomain proteins during development.

- Biochimica et Biophysica Acta - Mol. Cell Res.* **1813**, 1654–1662 <https://doi.org/10.1016/j.bbamcr.2011.01.013> (2011).
43. Lin, W. et al. The roles of multiple importins for nuclear import of murine aristaless-related homeobox protein. *J. Biol. Chem.* **284**, 20428–20439 (2009).
 44. Kock, K. H. et al. DNA binding analysis of rare variants in homeodomains reveals homeodomain specificity-determining residues. *Nat. Commun.* **15**, 3110 (2024).
 45. Chi, Y. I. et al. Homeodomain revisited: a lesson from disease-causing mutations. *Human Genetics*. **116**, 433–444 <https://doi.org/10.1007/s00439-004-1252-1> (2005).
 46. Salomone, J. et al. Conserved Gsx2/Ind homeodomain monomer versus homodimer DNA binding defines regulatory outcomes in flies and mice. *Genes Dev.* **35**, 157–174 (2021).
 47. Thai, M. H. N. et al. Constraint and conservation of paired-type homeodomains predicts the clinical outcome of missense variants of uncertain significance. *Hum. Mutat.* **41**, 1407–1424 (2020).
 48. Pérez-Villamil, B., Mirasierra, M. & Vallejo, M. The homeoprotein Alx3 contains discrete functional domains and exhibits cell-specific and selective monomeric binding and transactivation. *J. Biol. Chem.* **279**, 38062–38071 (2004).
 49. Rao, E. et al. The Leri-Weill and Turner syndrome homeobox gene SHOX encodes a cell-type specific transcriptional activator. *Hum. Mol. Genet.* **10**, 3083–3091 (2001).
 50. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
 51. Chaney, B. A., Clark-Baldwin, K., Dave, V., Ma, J. & Rance, M. Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. *Biochemistry* **44**, 7497–7511 (2005).
 52. Hovde, S., Abate-Shen, C. & Geiger, J. H. Crystal structure of the Msx-1 homeodomain/DNA complex. *Biochemistry* **40**, 12013–12021 (2001).
 53. Rohs, R. et al. The role of DNA shape in protein-DNA recognition. *Nature* **461**, 1248–1253 (2009).
 54. Tsai, C. J., Lin, S. L., Wolfson, H. J. & Nussinov, R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci.* **6**, 53–64 (1997).
 55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
 56. Andrews, S. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics* <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
 57. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. TrimGalore. *Babraham Bioinformatics* <https://doi.org/10.5281/zenodo.5127899> (2021).
 58. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
 59. Picard Toolkit. *Broad Institute, GitHub repository* <https://broadinstitute.github.io/picard/> (2019).
 60. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
 61. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
 62. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
 63. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
 64. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
 65. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
 66. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
 67. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
 68. Schrödinger, LLC. et al. The {PyMOL} Molecular Graphics System, Version-2.5.7 <https://www.pymol.org/> (2015).
 69. Cain, B. et al. The ALX4 dimer structure provides insight into how disease alleles impact function. *figshare. Dataset.* <https://doi.org/10.6084/m9.figshare.28535987> (2025).
 70. Cain, B. et al. The ALX4 dimer structure provides insight into how disease alleles impact function. *GitHub* <https://doi.org/10.5281/zenodo.14962844> (2025).

Acknowledgements

This work was supported by National Institute of Health grants to B.G. (R01-GM079428) and R.A.K. (R03-TR004875) and a Center for Pediatric Genomics grant from CCHMC to B.G.

Author contributions

This scientific study was conceived and planned by B.C., R.A.K., and B.G. Genomic analyzes were performed by B.C. All construct synthesis and protein purification were carried out by Z.Y. All crystallography and X-ray structural studies were conducted by Z.Y. and analyzed by Z.Y. and R.A.K. Protein structure visuals and schematics were generated by B.C. ITC, differential scanning fluorimetry assays, and CD assays were performed by Z.Y. EMSAs were performed by B.C. and E.T. Cell immunofluorescence assays, western blots, and reporter assays were performed by B.C. The manuscript was written by B.C. and edited by R.A.K. and B.G.

Competing interests

R.A.K. is on the scientific advisory board of Celestia Biotech AG and has received research funding from Celestia Biotech AG for projects unrelated to this manuscript. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-59728-9>.

Correspondence and requests for materials should be addressed to Brittany Cain, Rhett A. Kovall or Brian Gebelein.

Peer review information *Nature Communications* thanks Ankur Garg and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025