Disease-associated missense variants in ZBTB18 disrupt DNA binding and impair the development of neurons within the embryonic cerebral cortex

Isabel A. Hemming^{1,2,3,†}, Olivier Clément^{1,2,3,5,†}, Ivan E. Gladwyn-Ng^{1,2}, Hayley D. Cullen^{1,2,3,4}, Han Leng Ng⁵, Heng B. See^{1,2}, Linh Ngo^{1,2}, Daniela Ulgiati⁵, Kevin D.G. Pflieger^{1,2,6}, Mark Agostino^{4,7,8,*}, Julian I-T. Heng^{4,6,7,*}

¹The Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, Australia

²Centre for Medical Research, The University of Western Australia, Crawley, Australia

³The University of Western Australia, Medical School, Faculty of Health and Medical Sciences, Crawley, Australia

⁴Curtin Health Innovation Research Institute, Curtin University, Bentley, Australia

⁵The University of Western Australia, School of Biomedical Sciences, Faculty of Health and Medical Sciences, Crawley, Australia

⁶ARC Centre for Personalised Therapeutics Technologies, Australia

⁷School of Pharmacy and Biomedical Sciences, Curtin University, Bentley, Australia

⁸Curtin Institute for Computation, Curtin University, Bentley, Australia

* To whom correspondence should be addressed. Tel:(61)-413-509-384, Email:

Julian.Heng@curtin.edu.au

Correspondence may also be addressed to Mark Agostino. Tel:(61)-8-9266-9719, Email:

Mark.Agostino@curtin.edu.au

† The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

ABSTRACT

The activities of DNA-binding transcription factors, such as the multi-zinc finger protein ZBTB18 (also known as RP58, or ZNF238), are essential to coordinate mammalian neurodevelopment, including the birth and radial migration of newborn neurons within the fetal brain. In humans, the majority of disease-associated missense mutations in ZBTB18 lie within the DNA-binding zinc finger domain and are associated with brain developmental disorder, yet the molecular mechanisms explaining their role in disease remain unclear. To address this, we developed *in silico* models of ZBTB18 bound to DNA, and discovered that half of the missense variants map to residues (Asn461, Arg464, Glu486) predicted to be essential to sequence-specific DNA contact, while others map to residues (Leu434, Tyr447, Arg495) with limited contributions to DNA binding. We studied pathogenic variants to residues with close (N461S) and limited (R495G) DNA contact and found that each bound DNA promiscuously, displayed altered transcriptional regulatory activity *in vitro*, and influenced the radial migration of newborn neurons *in vivo* in different ways. Taken together, our results suggest that altered transcriptional regulation could represent an important pathological mechanism for ZBTB18 missense variants in brain developmental disease.

INTRODUCTION

The activities of DNA-binding transcription factors are crucial to the formation of the mammalian cerebral cortex, during which, coordinated waves of gene expression orchestrate a step-wise process of neurogenesis, cell migration, and circuit formation within the developing organ (Gupta, Tsai, & Wynshaw-Boris, 2002; Nord, Pattabiraman, Visel, & Rubenstein, 2015). ZBTB18 encodes a transcriptional repressor of the broad complex tramtrack bric-a-brac (BTB) zinc-finger family, comprising an N-terminal BTB domain for protein-protein interaction, as well as four Cys2-His2-like (C2H2) zinc fingers at its C-terminus for DNA binding (Mitchelmore et al., 2002). In mice, *Zbtb18* is essential to neurodevelopment (Okado et al., 2009; Xiang et al., 2011), including the radial positioning of newborn postmitotic neurons within the developing cerebral cortex (Heng et al., 2013; Ohtaka-Maruyama et al., 2013). Notably, loss of *Zbtb18* results in derepression of downstream target genes, including the cytoskeleton gene *Rnd2*, which, in turn, leads to defective neuronal migration within the embryonic cerebral cortex (Heng et al., 2013; Ohtaka-Maruyama et al., 2013). We reported that ZBTB18 directly binds enhancer motifs within the *Rnd2* gene identified as *E1* and *E2* (Heng et al., 2013), in order to modulate its levels for the control of radial migration by immature neurons. The transcriptional repressor functions for ZBTB18 are essential to its role in the regulation of multiple target genes essential to embryo development (Ohtaka-Maruyama et al., 2013; Xiang et al., 2011).

In humans, genetic mutations in *ZBTB18* are associated with structural brain abnormalities, neuronal migration disorder and intellectual disability (Cohen et al., 2017; de Munnik et al., 2014; Deciphering Developmental Disorders, 2017; Depienne et al., 2017; Hemming et al., 2016; van der Schoot et al., 2018). More than half of all pathogenic variants identified in ZBTB18 are predicted to be truncating, suggesting that haploinsufficiency/loss-of-function represents a general pathological mechanism for disease (Depienne et al., 2017). However, of the remaining variants that are not predicted to be truncating, it is noteworthy that the majority of these map to the C-terminal zinc finger DNA-binding domain of the polypeptide (Supplementary Figure S1 and Supplementary Table S1). This suggests that missense variants could influence the functions for ZBTB18 in DNA binding and transcriptional regulation which, in turn, leads to brain developmental disease.

Here, we have used homology modelling, molecular dynamics (MD) simulations, and molecular mechanics-generalised Born/surface area (MM-GB/SA) calculations to investigate the structural and functional properties of wildtype ZBTB18 and two disease-associated variants – NP_991331.1:p.Arg495Gly (R495G) (Rauch et al., 2012) and NP_991331.1:p.Asn461Ser (N461S) (Farwell et al., 2015; rs797044885) – bound to cognate DNA motifs. Through this approach, we propose key residues within ZBTB18 mediating its contact with DNA, and demonstrate that disease-associated missense variants to contact and non-base-contact residues within ZBTB18 influence its sequence-specific DNA binding, as well as its transcriptional activity *in vitro* and neuronal functions *in vivo*.

MATERIAL AND METHODS

Homology modelling of ZBTB18 zinc finger domain in complex with DNA probe sequences

The structures of the wild-type ZBTB18 zinc finger region from residues 373-501 in complex with DNA fragments from the E1 and E2 probe sequences were prepared by homology modelling. These motifs were derived from a previously characterised 3' regulatory enhancer sequence for *Rnd2*, a downstream target gene which mediates neuronal migration during brain development (Heng et al., 2013). Separate templates for each of the four C2H2 subdomains comprising the ZBTB18 zinc finger domain were identified using BLAST searches of each subdomain against the Protein Data Bank (PDB) using Prime Structure Prediction (Supplementary Table S2). The relevant C2H2 subdomains of the templates were overlaid to the relevant C2H2 subdomains of the designed zinc finger protein Aart (PDB 2I13) (Segal, Crotty, Bhakta, Barbas, & Horton, 2006), which afforded the highest sequence identity to the overall ZBTB18 zinc finger domain. The Composite/Chimera feature of Prime was used to build protein-DNA complexes for the wild-type ZBTB18 zinc finger domain, initially retaining the DNA from PDB 2I13. This DNA was trimmed to a 15 base pair fragment, representing the length of DNA most closely bound by the zinc finger domain.

C2H2 zinc finger domain structures in complex with DNA similar in sequence to the E1 sequence were then identified; these structures guided the placement of the desired DNA sequences with respect to the protein. The PDB was searched for C2H2 zinc finger domain structures (PFAM PF00096) with DNA bound. Pairwise alignment of the E1 sequence to each DNA sequence from the identified structures was performed using the *needleall* tool of EMBOSS (Rice, Longden, & Bleasby, 2000), this identified the DNA contained in PDB 5KE6 as affording the best alignment (Hashimoto et al., 2017). PDB 5KE6 was then aligned to the initially built DNA-protein complex (i.e., the complex of ZBTB18 with the DNA from PDB 2I13). A sequence alignment between the DNA in PDB 5KE6 and the DNA in PDB 2I13 was generated based on the structural alignment between the two DNA fragments. Combining this structure-based sequence alignment with the sequence alignment of the DNA probe sequences with the template DNA allowed an alignment between the DNA probe sequences and the DNA from PDB 2I13 to be generated (Supplementary Table S3). The DNA from PDB 2I13 was then mutated in UCSF Chimera to the relevant sequences based on this alignment.

Following model building and inclusion of the appropriate DNA sequences, the complex was energy minimised using Prime Minimization in a step-wise manner; non-template residues were first minimised, followed by all protein and ion components of the complex (i.e., the entire structure, excluding DNA), and finally, the entire complex. Erroneous *cis*-amides (i.e., *cis*-amides not derived from the template or immediately prior to a proline residue) and incorrect stereochemistry generated during the model building process was identified and corrected using Maestro. The structures of the zinc finger domain with each of the E1 and E2 probe sequences were produced in this manner; complexes with mutant proteins were prepared by mutating the relevant residue(s) in these representative structures using the mutation facilities of Maestro and used without further minimisation.

Molecular dynamics simulations

Parameterization of the complexes was performed using AmberTools (Salomon-Ferrer, Case, & Walker, 2013). Simulations were performed using GROMACS 5.0.2 (Abraham et al., 2015). ZBTB18 was parameterised with the AMBER ff14SB force field (Maier et al., 2015), while bound DNA was parameterised with the parmbsc0 parameter set (Perez et al., 2007) with ϵ/ζ OL1 (Zgarbova et al., 2013) and χ OL4 (Krepl et al., 2012) modifications. The Zinc AMBER force field (ZAFF) (Peters et al., 2010) was used to parameterise the zinc centres in ZBTB18. The resulting topology was then ported to GROMACS format using *acpype* (Sousa da Silva & Vranken, 2012) and the remaining system setup completed in GROMACS.

Complexes were solvated in TIP3P water (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) in a dodecahedral box with a minimum of 10 Å distance from the protein to the box edge. The system was charge neutralised with the addition of sodium ions, with further sodium and chloride ions added to a concentration of 0.1 M. Equilibrations of the system in the NVT and NPT ensembles were adapted from previously described procedures (Perez et al., 2007; Shields, Laughton, & Orozco, 1997, 1998); briefly, short simulations (0.1 ns) with gradually reducing position restraints on the complex heavy atoms were performed (initially an NVT simulation with 10000 kJ/(mol nm²) applied, followed by NPT simulations with 2000 kJ/(mol nm²) applied, gradually decreasing by 400 kJ/(mol nm²)). Following this, the production MD simulation was performed in the NPT ensemble for 50 ns, with a target temperature of 300K and a target pressure of 1 atm. The modified Berendsen thermostat (V-rescale) was employed for temperature coupling (Bussi, Donadio, & Parrinello, 2007). The Parrinello-Rahman barostat was employed for pressure coupling (Parrinello & Rahman, 1981). The smooth particle mesh Ewald method was employed for long-range electrostatics (Essmann et al., 1995). All bonds were constrained using the LINCS algorithm (Hess, Bekker, Berendsen, & Fraaije, 1997). Coordinates were saved every 10ps.

Five simulations of each complex were performed, starting from new random velocities in each replicate. For binding energy calculations, the final 10 ns of three of these simulations were used, determined by examining the root-mean-squared deviation (RMSD) for heavy atoms over this time range; simulations where the RMSD appeared stable and maintaining under 6.0 Å were selected for binding energy calculations.

Binding free energy calculations and decomposition

Binding energies were calculated using the (MM-GB/SA) approach, facilitated by the MMPBSA.py tool of AmberTools (Miller et al., 2012). The single-trajectory protocol was utilised (Gohlke, Kiel, & Case, 2003), with the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} = \Delta H + \Delta G_{\text{solvation}} - T\Delta S = \Delta G_{\text{vdw}} + \Delta G_{\text{ele}} + \Delta G_{\text{GB}} + \Delta G_{\text{SA}} - T\Delta S$$

where ΔG_{vdw} is the molecular mechanics van der Waals interaction energy, ΔG_{ele} is the molecular mechanics electrostatic interaction energy, ΔG_{GB} is the change in polar desolvation energy upon complex formation and ΔG_{SA} is the change in nonpolar desolvation energy upon complex formation. The entropic term ($-T\Delta S$) was not calculated, due to high computational cost and poor accuracy. The modified GB model developed by Onufriev et al. (*igb* = 5) was used to calculate the polar desolvation energy (Onufriev, Bashford, & Case, 2004); this GB model is suggested as a suitable “general purpose” GB model and has been previously applied to protein-DNA complexes (Ruscio & Onufriev, 2006). The non-polar component of the desolvation energy was calculated via solvent accessible surface areas calculated with the LCPO method (Weiser, Shenkin, & Still, 1999). MM-GB/SA energies were also decomposed per-residue (*idecomp* = 1). The GROMACS tool *g_cluster* was used to identify the most frequently observed structures from the regions of the simulations submitted for MM-GB/SA analysis. The cutoff for the *g_cluster* calculation was set to 0.25°nm.

Mouse handling and Animal Ethics License.

All animal procedures were conducted according to standard operating procedures approved by the Animal Ethics Committees (Perkins Institute: AE021, University of Western Australia: RA/3/100/1361), as well as guidelines provided by the National Health and Medical Research Council of Australia. Unsexed C57BL/6J wild-type mouse embryos were employed for *in utero* electroporation studies.

DNA constructs and cloning

A *Zbtb18* shRNA vector used in this study has been previously reported (Heng et al., 2013). The cDNA sequence encoding isoform 1 of human ZBTB18 (NM_205768.2) and ZBTB18 (R495G) were commercially synthesised (Integrated DNA Technologies). Both cDNAs were individually sub-cloned into pCIG-FLAG vector in which the GFP cassette was excised by restriction digestion and subsequent cloning (pCIG-F[NG]) (Heng et al., 2013). The N461S variant was generated via site-directed mutagenesis using a GeneArt® Site-Directed Mutagenesis PLUS kit (A14604; Life Technologies, Carlsbad, California, United States) on the pCIG-F[NG] ZBTB18(WT) construct with the following primers: sense, 5'-TCCAGTACTCGCACAGCCTGAGCCGCCATGC-3'; antisense, 5'-GCATGGCGGCTCAGGCTGTGCGAGTACTGGA-3'. All constructs were verified by Sanger sequencing (Australian Genome Research Facility, Perth, Australia).

Cell Culture and Transfection

Human (HeLa, HEK293T) and mouse (P19 embryocarcinoma) cell lines were cultured in complete growth media (DMEM (10313-021; Gibco by Life Technologies, Carlsbad, California, United States), 2 mM L-glutamine (25030-081; Gibco by Life Technologies, Carlsbad, California, United States), 10% fetal bovine serum (SFBSF; Bovogen Biological, Keilor East, Victoria, Australia)), in an incubator at 37°C supplied with humidified 5% CO₂. HeLa cells were seeded in 500 µl of growth media in a 24-well plate, at a density of 50 000 cells per well for localisation experiments. P19 cells were seeded in 500 µl of growth media in a 24-well plate, at a density of 25 000 cells per well for luciferase assays.

HEK293T cells were seeded in 2ml of growth media in a 6-well plate, at a density of 500 000 cells per well for Electrophoretic Mobility Shift Assay (EMSA) experiments and initial western blotting, and in a 48-well plate at a density of 25 000 cells per well for experiments with MG132 exposure. 24 hours after seeding, cells were transiently transfected with Lipofectamine 2000 (11668-10019; Invitrogen by Life Technologies, Carlsbad, California, United States) as per manufacturer's instructions. The media was changed 5 hours post-transfection (media without Lipofectamine). For immunostaining experiments, HeLa cells were cultured on uncoated 13 mm round coverslips (CS13100; Grate by Trajan, Ringwood, Victoria, Australia) and fixed onto the coverslips at either 24 or 48 hours post-transfection. P19 cells were collected 48 hours post-transfection for luciferase assays. Nuclear extract from HEK293T cells was collected 48 hours post-transfection for EMSAs. For MG132 experiments HEK293T cells were treated with 10 μ M MG132 (C2211, Sigma-Aldrich, St. Louis, Missouri, United States) at 40 hours post-transfection, and then washed with 1x PBS buffer and lysed using KALB buffer at either 0, 4, or 8, hours post MG132 treatment.

Luciferase assays

Firefly luciferase reporter constructs harbouring a synthetic decamerised binding site (BS10), as well as a construct comprising mouse *Rnd2* 3'-enhancer sequence (Heng et al., 2008), were used. A *Renilla* luciferase construct was co-transfected to normalise for transfection efficacy (Promega, Madison, Wisconsin, United States). For competition assays between WT, N461S, and R495G, 250 ng of each ZBTB18 construct was added, together with 50 ng of firefly reporter construct, and 25 ng of *Renilla* construct, per well. DNA was added to 100 μ l Opti-MEM I (31985-070; Gibco by Life Technologies, Carlsbad, California, United States) per sample. The four constructs were pCIG-F[NG]-empty, pCIG-F[NG]-ZBTB18(WT), pCIG-F[NG]-ZBTB18(N461S), and pCIG-F[NG]-ZBTB18(R495G). 48 hours post-transfection, P19 cells were washed with 500 μ l 1x PBS and lysed with 150 μ l 1x Passive lysis buffer as per manufacturer's instructions (E1960; Dual Luciferase Assay Kit, Promega, Madison, Wisconsin, United States). The lysate was then centrifuged at 15000 x g for 15 min at 4°C. 20 μ l of supernatant for each 24-well was added to two 96-wells, and reagent added to each well manually. Firefly and *Renilla* signals were read in each well of the 96-well plate in order of rows (5 second recordings per well), using a VictorLight plate reader (Perkin Elmer, Waltham, Massachusetts, United States) according to manufacturer's instructions.

Western Blotting

HEK293T cells were collected 48 hours after transfection and Western blotting. Briefly, cells were washed with 1x PBS buffer and were lysed using KALB buffer (1% Triton X detergent, 150mM NaCl, 0.02% sodium azide, 1mM EDTA, 50mM Tris-HCl pH 7.4, and protease inhibitor (4693159001; Roche, Basel, Switzerland)), and protein concentration measured using Bradford reagent (500205; Bio-Rad, Hercules, California, United States). Protein samples were analysed on a 10% SDS gel, transferred to

nitrocellulose membrane, and incubated overnight with primary antibodies: mouse anti- β -actin (1:5,000 (A5441; Sigma-Aldrich, St. Louis, Missouri, United States)), mouse anti-FLAG (1:5,000 (F1804; Sigma-Aldrich, St. Louis, Missouri, United States)), mouse anti-ZBTB18 (1:5,000 (H00010472 -M04; Abnova, Jhongli, Taiwan)), rabbit anti-FLAG (1:5,000 (2368; Cell Signaling, Danvers, Massachusetts, United States)) and rabbit anti-ZBTB18 (1:5,000 (12714-1-AP; Proteintech, Chicago, Illinois, United States)). Membranes were incubated with anti-rabbit IRDye 680LT or anti-mouse IRDye 800 secondary antibodies (1:10,000 (LI COR Biosciences, Lincoln, Nebraska, United States) for 2 hours before analysis using the Odyssey imaging system (LI-COR Biosciences, Lincoln, Nebraska, United States).

Electrophoretic Mobility Shift Assays (EMSAs)

EMSAs were performed as previously described (Cruickshank et al., 2015), using nuclear extracts from HEK293T transfected cells collected using the NE-PER™ Nuclear and Cytoplasmic Extraction Kit according to manufacturer's instructions (78833; Life Technologies Carlsbad, California, United States). In short, 2 μ g of crude nuclear protein extract was incubated with 75 fmol of double-stranded oligonucleotides in EMSA reaction buffer (4% Ficoll, 20mM HEPES [pH 7.9], 1mM EDTA [pH8.0], 1.5mM DTT, 0.5 μ g Poly dI:dC) for 30 minutes, before running on a 6% non-denaturing polyacrylamide gel (6% Bis-acrylamide, 2.5% glycerol, 0.75x TBE buffer) for 90 minutes at 150 V. Commercially sourced, HPLC-purified 5'-biotinylated oligonucleotides (Sigma-Aldrich, St. Louis, Missouri, United States) comprising sense and anti-sense strands of DNA stretches encompassing the binding sites of E1 (sense 5'-CTCTGCTGTTACTCCTAAATAACAGATGTCTGTCTGCATA-3'; antisense 5'-TATGCAGACAGACATCTGTTATTTAGGAGTAACAGCAGAG-3'); E1^{mut} (sense 5'-CTCTGCTGTTACTCCTAAATAACAGTGATCTGTCTGCATA-3'; antisense 5'-TATGCAGACAGATCACTGTTATTTAGGAGTAACAGCAGAG-3'); E2 (sense 5'-TTTGATCCACCAAAGGGAGGGGCAGATGGGAGTAGGGAAG-3'; antisense 5'-CTTCCCTACTCCCATCTGCCCCTCCCTTTGGTGGATCAAA-3'); E2^{mut} (sense 5'-TTTGATCCACCAAAGGGAGGGGCAGTGAGGAGTAGGGAAG-3'; antisense 5'-CTTCCCTACTCCTCACTGCCCCTCCCTTTGGTGGATCAAA-3'). Strand pairs were annealed by resuspension in 1X TE buffer and heated to 95°C for 10 minutes, then slowly cooled to room temperature. Oligonucleotides were run on a 6% polyacrylamide gel and double-stranded oligonucleotides were extracted from the gel in elution buffer (0.1% sodium dodecyl sulfate (SDS), 0.5 M ammonium acetate, 10 mM magnesium acetate) at room temperature for 16 hours. Oligonucleotides were precipitated with 100% ethanol at -80°C for two hours, then pelleted by centrifugation at 16 000g for 30 min at 4°C and washed with 800 μ L of 70% ethanol and air dried thoroughly. Oligonucleotides were resuspended in nuclease-free water and diluted to 25 pmol/ μ L stock just before use. Biotin detection was undertaken with a Chemiluminescent Nucleic Acid Detection Module Kit according to the manufacturer's instructions (89880; Thermo Fisher, Waltham, Massachusetts, United States).

***In utero* electroporation and tissue collection for immunostaining**

In utero electroporation was performed on E14.5 embryos as described previously (Haas et al., 2016). Plasmids were delivered at a final concentration of each species at 1 µg/µl, pre-mixed with Fast Green to visualise site of injection. Three days after electroporation, the pregnant dam was euthanatised by cervical dislocation, embryos (E17.5) were decapitated, and their brains were dissected then fixed overnight at 4°C in 4% paraformaldehyde in PBS. Then, the brains were equilibrated in 30% sucrose in PBS for 3 days at 4°C, before embedding in OCT mounting medium, frozen on dry ice and finally stored at -80°C. Coronal sections 16 µm thick were prepared from each brain using a cryostat (Leica, Biosystems), and recovered on Superfrost-PLUS glass slides (Menzel-Glaser by Trajan, Ringwood, Victoria, Australia). Slides were dried for at least 60 minutes at ambient temperature conditions before storage at -80°C until immunostaining was performed, as follows. Briefly, after a blocking step with 10% of normal goat serum in PBS, brain sections were incubated overnight at 4°C with the primary antibodies: chicken anti-GFP (1:700 (Ab13970; Abcam, Cambridge, United Kingdom)), mouse anti-FLAG (1:500 (F-1804; Sigma-Aldrich, St. Louis, Missouri, United States)). Then sections were incubated with fluorescent secondary antibodies (488-Goat anti-Chicken (1:1,000 (A11039; Invitrogen by Life Technologies, Carlsbad, California, United States))); 568-Goat antimouse (1:1,000 (A11031; Invitrogen by Life Technologies, Carlsbad, California, United States))) at room temperature for 2 hr before sections were counterstained with 4'6-Diamidino-2-Phenylindole (DAPI, 1:10000 (D1306; Invitrogen by Life Technologies, Carlsbad, California, United States)). A coverslip was mounted on each slide using Fluorescent Mounting Medium (DAKO, Berlin, Germany).

Microscopy, imaging, quantification studies

Images of brain sections were captured on an epifluorescence microscope (Olympus, Shinjuku, Tokyo, Japan) equipped with a CCD camera (SPOT, Sterling Heights, Michigan, United States) for neuronal migration, and on a confocal microscope (Nikon, Minato, Tokyo, Japan) for the analysis of neuronal shape. Subdivisions of the embryonic cortex (VZ/SVZ, IZ and CP) were identified based on cell density as visualised with DAPI (4'6-Diamidino-2-Phenylindole) staining, as described previously (Haas et al., 2016). Cell counting was performed blind to the condition on representative fields of sections of electroporated brains using ImageJ software.

Statistics

Data from all experiments are expressed as the mean ± standard error of the mean (SEM). A Kruskal Wallis analysis was employed to test multiple conditions, while a Mann-Whitney U test was applied for statistical analyses between two treatments. All statistics were performed through Prism software (GraphPad).

RESULTS

Structural prediction and binding energy calculations of ZBTB18 zinc finger domain for wildtype and missense variants, in complex with DNA probe sequences

We used homology modelling, molecular dynamics (MD) simulations, and molecular mechanics-generalised Born/surface area (MM-GB/SA) calculations to investigate the structural and functional properties of the C-terminal zinc finger domain of ZBTB18 (residues 373-501) bound to two DNA regulatory motifs, E1 and E2, within the *Rnd2* gene (Heng et al., 2008; Heng et al., 2013). To explore sequence-specific binding properties, we also modelled ZBTB18 binding to mutated versions of each motif, named E1^{mut} and E2^{mut}, which we also previously reported as refractory to ZBTB18-mediated transcriptional repression (Heng et al., 2013). The predicted structure of the ZBTB18 zinc finger domain comprises four C2H2-type subdomains, featuring a moderately sized insertion between the first two C2H2 subdomains relative to the template structures (Figure 1, Supplementary Video S1 and Supplementary Figures S2A-C). MD simulations indicated stability of the majority of protein-DNA complexes, which were typically reproducible when simulations were commenced from new random velocities (Supplementary Figure S3A,D,G,J).

Based on our MM-GB/SA calculations, we found that wildtype ZBTB18 formed energetically more favourable interactions with both E1 and E2, compared to the mutated sequences E1^{mut} and E2^{mut} (Table 1). Per-residue binding energy decomposition identified residues – including Tyr458, Asn461 and His493 – as critical to wildtype ZBTB18 to bind E1 and E2 (Figure 2, Table 2 and Supplementary Table S4). For example, Asn461 contributes as much as 7.2% and 7.6% of the total binding energy of ZBTB18 to E1 and E2 respectively. Importantly, binding energy contributions made by these residues were markedly lower for ZBTB18 modelled with E1^{mut} and E2^{mut} mutants, with the exception of His493, which displayed an increased contribution when complexed with E2^{mut}, altogether suggesting roles for each residue in sequence-specific binding (Figure 2B). Based on our models of ZBTB18 with E1 and E2, we also observed that half of the disease-associated missense variants map to residues (Asn461, Arg464, Glu486) essential to sequence-specific DNA contact, while others map to residues exhibiting comparably reduced (Tyr447, Arg495) or negligible (Leu434) contributions to DNA-binding (Table 2).

Next, we focussed our attention on disease-associated ZBTB18 variants which featured residues with close (N461S) or limited (R495G) DNA contact. Separate models were generated for N461S (Figure 3A-B, and Supplementary Figure S2D-F) and R495G (Figure 3C-D, and Supplementary Figure S2G-I) binding E1, E1^{mut}, E2 and E2^{mut}. Binding energy calculations suggested that N461S exhibited more selective binding for E1 over E1^{mut}, and preferential binding for E2^{mut} over E2 (Table 1). On the other hand, R495G exhibited similar binding energies for each of the E1, E1^{mut} and E2 motifs, suggesting non-selective binding for all three sequences. Calculations with E2^{mut} indicated that R495G bound with substantially lower energy compared to E1, E1^{mut} and E2 motifs.

We performed binding energy decomposition to understand the interactions taking place in the complexes in finer detail. In the case of a N461S substitution, we observed a reduced contribution to the binding energy of this position between E1 and E2, concomitant with a structural rearrangement of Tyr458 and His493, which augmented their binding with E1^{mut} and E2^{mut}, as well as enhanced the

contribution to the binding energy of His493 with E2 (Figure 3B and Table 2). In contrast, Arg495 weakly contributes to binding of E1 and E2, and substitution to glycine eliminates any contribution to binding energy made by this position (Figure 3D and Table 2). However, with this substitution, we also observed reduced contributions made by Tyr458 and Asn461 when binding to the E1 and E2 motifs, as well as an increase in contribution to E2 binding by His493, altogether suggesting indirect effects of this missense variant on DNA binding. When modelled with E1^{mut}, Tyr458 also shows a five-fold elevation in binding energy contribution in the R495G variant (-3.5 kcal/mol for this residue within R495G complexed with E1^{mut}, versus -0.7 kcal/mol within WT ZBTB18 complexed with E1^{mut}, see Table 2). Therefore, our models suggest that missense variants to residues relevant to close (N461S) or limited (R495G) DNA contact differentially affect the capacity for ZBTB18 to bind DNA, and could lead to spurious binding.

Missense variants in ZBTB18 disrupt sequence-specific DNA-binding and transcriptional repression

To substantiate our molecular predictions, we performed electrophoretic mobility shift assays (EMSAs) with DNA probes comprising each motif. We found that WT ZBTB18 formed distinct molecular complexes with a DNA probe comprising E1 (Figure 4A, arrows), with an additional complex formed by R495G (Figure 4A, arrowhead). In agreement with our modelling results, these complexes were either detected at very low intensity or not detected at all in assays with WT ZBTB18 and E1^{mut} (Figure 4B). In contrast to WT ZBTB18, however, each missense variant formed distinct complexes with the E1^{mut} motif (Figure 4B, arrow and closed arrowhead), with R495G forming a prominent high-molecular weight complex (Figure 4B, open arrowhead). In the case of the E2 motif, ZBTB18 and its missense variants formed molecular complexes (Figure 4C, arrows), with additional complexes formed by R495G (Figure 4C, open and closed arrowheads). However, such protein-DNA complexes were not identified in parallel experiments using a probe comprising the E2^{mut} motif (Figure 4D). Together, these modelling and *in vitro* data indicate that ZBTB18 binds its DNA motifs E1 and E2 in a sequence-specific manner, while its missense variants bind E1, E2, as well as E1^{mut}, but not E2^{mut}.

To understand the impact of each missense variant ZBTB18 on transcriptional regulation *in vitro*, we performed luciferase reporter assays using a previously validated synthetic promoter construct (termed BS10) (Aoki et al., 1998), as well as a construct harbouring the *Rnd2* 3' enhancer (comprising E1 and E2 motifs) (Heng et al., 2013). We found that wildtype ZBTB18 suppressed luciferase reporter activity, while N461S failed to suppress reporter activity mediated by the BS10 synthetic enhancer, but stimulated luciferase signal mediated by the *Rnd2* 3' enhancer (Figure 4E-F). Unexpectedly, we found that R495G significantly enhanced luciferase signal in both reporter constructs. The capacity for N461S and R495G to stimulate reporter activity was extinguished when equal quantities of the WT construct were co-transfected (Supplementary Figure S4). Next, we explored transcriptional reporter activity of ZBTB18 and each missense ZBTB18 variant using reporter constructs bearing either E1^{mut}, E2^{mut}, or both E1^{mut} and E2^{mut} motifs within the *Rnd2* 3' enhancer sequence. For wildtype ZBTB18, transcriptional repression was only abolished in experiments with a

reporter bearing both E1^{mut} and E2^{mut} motifs, confirming its sequence-specific binding to native E1 and E2 motifs (Figure 4G). In contrast, statistically significant transcriptional activation by N461S was observed in constructs bearing E1+E2 or E1^{mut}+E2^{mut} motifs, but not an E1^{mut}+E2 motif or an E1+E2^{mut} motif. When compared to WT ZBTB18, the N461S variant has reduced capacity for repression in the context of E1+E2, E1^{mut}+E2 and E1+E2^{mut} motif pairs, and activation in context of the E1^{mut}+E2^{mut} motif pair (Figure 4H and summarised in Supplementary Figure S5B). These data suggest that N461S variant has reduced repressor activity compared to WT, and may influence transcriptional activation in some contexts. Conversely, R495G robustly stimulated reporter activity in all four reporter constructs, suggesting this variant has strong capacity for activation in all contexts (Figure 4I and Supplementary Figure S5C).

Next, we performed transfection experiments utilising FLAG-tagged ZBTB18 constructs to explore the impact of each missense variant on subcellular localisation and expression. Twenty four hours after transfection we observed reduced nuclear localisation of N461S compared to wildtype ZBTB18, concomitant with a corresponding increase in pancellular signal distribution (Supplementary Figure S6A-C). The proportion of cells with a punctate distribution of R495G within the nucleus was also significantly increased compared to wildtype ZBTB18 (Supplementary Figure S6D). These effects were transient and they were not observed 48 hours post-transfection (Supplementary Figure S6E-G). We also routinely observed significantly fewer FLAG-N461S immunopositive cells compared with wildtype and R495G-transfected cells suggesting that N461S might be unstable (Supplementary Figure S7A-B). Furthermore, low steady-state levels of FLAG-N461S protein from lysates of transiently transfected cells were observed (Supplementary Figure S7C-D). To further investigate the hypothesis that the N461S variant might be unstable, we examined FLAG-tagged ZBTB18 expression in transfected cells treated with the proteasome inhibitor MG132 for up to 8 hours followed by Western blotting analysis (Supplementary Figure S7E-F). In this experiment, we reproducibly observed a very low immunoblotted FLAG signal in the lysate of N461S transfected cells compared to WT and R495G conditions at 0 hours of MG132 treatment (Supplementary Figure S7E). However, after 4 hours of MG132 treatment the FLAG signal was significantly increased for N461S, compared to no MG132 treatment. Following 8 hours MG132 exposure, N461S, WT and R495G variant levels were significantly increased (Supplementary Figure S7F). These results indicate that the N461S variant exhibits an exaggerated protein instability compared to WT, and is prone to degradation in cells.

Disease-associated missense variants to ZBTB18 disrupt radial migration within the embryonic cerebral cortex.

The transcriptional regulatory activities of ZBTB18 are essential to mammalian neurodevelopment, and suppression of *Zbtb18* impairs radial migration of cortical projection neurons from their birthplace within the germinal ventricular zone (VZ), including their multipolar-to-bipolar transition within the intermediate zone (IZ) as they reach the cortical plate (CP) of the embryonic cerebral cortex (Heng et al., 2013; Ohtaka-Maruyama et al., 2013). Hence, we assessed the capacity for ZBTB18 and its

missense variants to influence the radial migration of E14.5-born cortical neurons within the E17.5 cortex through a series of *in utero* electroporation experiments. Treatment with *Zbtb18* shRNA disrupted the capacity for cells to migrate into the CP; however, this impairment is corrected by co-delivery of wildtype ZBTB18 (Figure 5A-G), consistent with previous findings (Heng et al., 2013; Ohtaka-Maruyama et al., 2013). Strikingly, we found that co-delivery of N461S significantly augmented the radial migration of *Zbtb18* shRNA-treated cells into the CP and markedly potentiated their intracortical positioning within the upper CP (Figure 5F-G). We could not detect N461S immunofluorescence in rescued cells (Supplementary Figure S8), consistent with the notion that this protein variant is unstable (see Supplementary Figure S7), and beyond the limit of immunodetection in electroporated cortical cells. In contrast, co-delivery of R495G exacerbated their migration defect (Figure 5F-G). We also performed electroporation experiments on E14.5 cortices collected two days later (ie at E16.5) to observe that the proportion of cells within the CP was not significantly different across all conditions (Supplementary Figure S9A-B). Therefore, our data suggests that the differential effects on radial migration by N461S and R495G variants are relevant to the transition from multipolar to bipolar migration as cells migrate within the IZ and CP.

Next, we analysed the cell shapes of IZ and CP neurons across each treatment group and found that knockdown of *Zbtb18* led to a significant reduction in the proportion of uni/bipolar-shaped cells in the CP and a concomitant increase in the proportion of multipolar shaped cells, as well as significantly shortened, leading neurites (Figure 5H-O). Co-delivery of WT ZBTB18 resulted in significant restorative effects on cell shape and leading neurite length (Figure 5M-O). In contrast, co-delivery of N461S did not affect IZ cell shape but instead led to an increase in multipolar-shaped cells within the CP with shorter leading neurites (Figure 5O), suggestive of dysregulated locomotion which results in their enhanced localisation within the upper CP. On the other hand, co-delivery of R495G led to an increase in multipolar-shaped cells in the IZ, which we interpret as a disruption of multipolar-to-bipolar transition within the cortical subcompartment. Cell shape profiles and leading neurite lengths of the few cells arriving within the CP were significantly different to wildtype ZBTB18 rescue, consistent with a detrimental impact of R495G expression in CP cells. Taken together, missense variants of ZBTB18 impair radial migration in a cell autonomous fashion, and disrupt neurite outgrowth in immature cortical neurons.

DISCUSSION

In this study, we have applied molecular modelling approaches to investigate the binding of ZBTB18 to two native DNA motifs (E1 and E2) within the 3' enhancer region for the migration-promoting gene, *Rnd2*; as well as their mutagenised counterparts (E1^{mut} and E2^{mut}). We demonstrate that MM-GB/SA analysis is effective to model ZBTB18 wildtype and variant complexes with different DNA sequences. This is remarkable considering the difficulties associated with the development of accurate DNA force fields for simulations, and the assumptions required for use of the MM-GB/SA approach (Genheden & Ryde, 2015). A strength to our approach has been to incorporate information

from multiple templates, as well as using a sequence homology-based approach to identifying relevant DNA structural templates, during the initial homology modelling. As a result, we have developed stable DNA-protein complexes and calculated binding energies for ZBTB18 and its variants, with observations substantiated by biochemical observations in this study.

A recent study by van der Shoot and colleagues identified via exome sequencing ZBTB18 variants at Arg464 in patients with neurological abnormalities, and used homology modelling against a single template structure and a model DNA fragment to illustrate its potential for direct involvement in DNA binding (van der Schoot et al., 2018). While our study also identified Arg464 as a major contributor to DNA binding by ZBTB18, our rigorous modelling approach utilising multiple templates to develop a model bound to cognate DNA motifs can provide fine detail into the energetic contributions to binding made by each residue, including Asn461 and His493. Crucially, we applied this modelling approach also to enable us to investigate the impact of ZBTB18 missense variants on DNA binding. Indeed, we found that substitution of Asn461 to serine alters the DNA-binding specificity of ZBTB18, and disrupts the capacity for this transcription factor to mediate transcriptional repression and radial migration. Furthermore, our models of ZBTB18 with mutated DNA motifs indicated that residues with very limited DNA contact, such as Arg495, might be relevant to binding of spurious DNA motifs. Indeed, we find that the missense variant R495G dramatically altered the DNA-binding specificity, transcriptional repressor activity and neurobiological functions of ZBTB18. It has recently been established that non-base-contacting residues are crucial to the evolution of C2H2 zinc finger DNA binding (Najafabadi et al., 2017) and our results suggest that molecular modelling approaches are capable of informing the structural and functional consequences of substitution variants and non-base-contacting residues.

Our biochemical characterisation of N461S and R495G variants demonstrates that each variant displays distinct effects on its subcellular localisation, as well as its capacity to mediate transcriptional repression in a luciferase reporter assay. The altered capacity for N461S to traffic into the nucleus could arise as a consequence of disruptions to its association with nuclear import proteins, thereby influencing its availability within the nucleus to mediate transcription. Also, we reported low steady-state levels of N461S observed in transfected cells and electroporated neurons, demonstrating that the N461S variant is unstable both *in vitro* and *in vivo*. This suggests that the N461S variant has a pathophysiological role with very low expression levels in cells. In the case of the R495G variant, we find that the presence of this variant influences the sub-nuclear localisation of ZBTB18, leading to its accumulation in nuclear puncta, suggestive of protein aggregation. Over time, we find that the distribution patterns of N461S and R495G are not significantly different to wildtype, hence we conclude that each missense variant likely causes a transient delay in the appropriate localisation of ZBTB18 within cells, leading to asynchronisation of its nuclear functions, ultimately disrupting its capacity to orchestrate neurodevelopment.

In addition to transient disruptions in subcellular localisation, we find that the transcriptional repressor functions for each missense variant are disrupted in different ways. For example, we find that the N461S variant has lost the capacity for suppressing transcription via the BS10 and *Rnd2* 3'

enhancer constructs in our luciferase assays, and can potentiate transcriptional activation in certain contexts. We speculate that the impaired capacity for ZBTB18-mediated repression, together with low steady-state levels, explains the pathological impact of the N461S variant, leading to an imbalance in the regulation of downstream target genes. In contrast, we find that the R495G variant appears to have acquired strong transcriptional activation potential *in vitro*, with attendant consequences on gene regulation in cells. Given that ZBTB18 mediates its transcriptional regulatory functions in part through its recruitment of co-repressors such as Dnmt3a (Fuks et al., 2001), we further speculate that N461S and R495G are altered in their capacity to bind protein partners for transcriptional regulation, and may recruit protein complexes that potentiate gene expression. Our capacity to visualise such changes in protein-DNA complexes in our EMSA experiments was confounded by the presence of endogenous ZBTB18 binding to the E1 motif in the lysates of mock-transfected cells. Also, our experiments with E2^{mut} motifs did not yield specific complexes for ZBTB18 and its missense variants. Nevertheless, we did observe unique protein-DNA complexes formed by N461S on the E1^{mut} motif, as well as by R495G on the E1, E1^{mut} and E2 motifs. Thus, we surmise that luciferase reporter activity mediated by N461S can be explained by its preferential binding to E1, E1^{mut} and E2 motifs, while R495G recruits additional factors to potentiate reporter activity upon binding to these motifs (Supplementary Figure S5). As such, the “loss-of-repression” variant N461S and the “gain-of-transactivation” variant R495G are postulated to influence gene expression during neurodevelopment, likely through dysregulation of downstream genes which influence the radial positioning and differentiation of neurons.

Through a series of functional assays, we found that the presence of N461S and R495G affects the development and positioning of newborn neurons within the embryonic cerebral cortex in different ways. Notably, the impact on transcriptional regulation during radial migration is likely to be two-fold. Firstly, the presence of each missense variant could lead to spurious binding of DNA regulatory sequences, resulting in changes in gene expression levels beyond ZBTB18 targets. Secondly, the presence of each missense variant has a unique effect on radially migrating neurons within the embryonic cerebral cortex. Thus, our results support a scenario in which the “loss-of-repression” variant, N461S, has a different impact on radial migration, in contrast to the “gain-of-transactivation” variant, R495G (see Supplementary Figure S10 for an illustrative summary). In such a model, a loss-of-repression owing to the presence of the N461S variant influences gene expression regulation within radially migrating neurons, leading to their exuberant migration and preferential insertion within the Cortical Plate as multipolar-shaped cells, suggestive of disrupted radialglial-guided locomotion. In contrast, the R495G variant potentiates gene expression so as to severely impair their radial migration, culminating in a failure in multipolar-to-bipolar transition of cortical cells, and their accumulation within the IZ. Indeed, we previously reported that too much or too little *Rnd2* disrupts radial migration, such that its overexpression in cortical neurons impairs the capacity for cells to adopt appropriate morphologies to undergo movement (Heng et al., 2008; Pacary et al., 2011). Consistent with this interpretation, our analysis of IZ and CP neurons in the rescue assays revealed that the presence of the R495G variant augmented the multipolar phenotype of *Zbtb18* shRNA-treated neurons, thereby suggesting an additive effect on cell shape which could underlie its negative impact

on radial migration. In contrast, treatment with N461S did not have such an effect in *Zbtb18* shRNA-treated cells in the IZ or the CP. In addition, *Zbtb18* shRNA-treated CP neurons co-electroporated with R495G extended longer leading neurites while N461S treatment did not have such an effect, further suggesting different cellular mechanisms that underlie the aberrant migration of treated neurons in each respective condition. Based on our results, we surmise that the presence of each substitution variant has a significant effect on the development of neurons within the mammalian brain. Future studies will address how each substitution variant disrupts the long-term positioning, dendritic branching and synaptic connectivity of cortical neurons within the postnatal cortex.

It has been reported that more than half of all pathogenic variants identified in ZBTB18 are predicted to be truncating, suggesting that haploinsufficiency/loss-of-function represents a general pathomechanism (Depienne et al., 2017). Yet, the majority of disease-associated missense variants to ZBTB18 lie within the C-terminal zinc finger DNA-binding domain. Therefore, our investigation of the N461S and R495G variants has been crucial to demonstrate that altered transcriptional regulation also represents an important pathological mechanism for ZBTB18 variants in brain developmental disease. Altogether, our *in silico* models and functional studies demonstrate that missense variants to base-contact and non-base-contact residues influence DNA binding, transcriptional repression and neurodevelopmental signalling.

The differences in functions for each pathogenic missense variant draws parallels with the differences in phenotypic presentation of the two human subjects diagnosed with *ZBTB18* variants (Cohen et al., 2017; Rauch et al., 2012). Particularly, while intellectual disability was diagnosed in both subjects, the subject harbouring an N461S variant was also diagnosed with microcephaly and hypoplasia of the corpus callosum (Cohen et al., 2017). In contrast, the subject diagnosed with the R495G variant was diagnosed with macrocephaly. We speculate that the presence of missense variants to the zinc finger domain of ZBTB18 influences its capacity to bind DNA in a sequence-specific fashion, affect its transcriptional regulatory functions, and negatively impact on neural development to cause a spectrum of brain disorders in children. Functional studies of common missense variants that lie within the DNA-binding domain of ZBTB18 will also be important to clarify their potential impact on transcriptional regulation in human health and disease.

DATA AVAILABILITY

The datasets analysed during the current study are available from the corresponding author on reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

I.A.H. and O.C. performed biochemistry and molecular biology studies with J.I.-T.H., H.L.N., H.B.S., D.U. and K.D.G.P.. H.D.C., I.E.G.-N. performed *in utero* electroporation experiments with L.N. and O.C.. M.A. performed structural modelling calculations in consultation with J.I.-T.H., I.A.H. and O.C.. J.I.-T.H., I.A.H., O.C. and M.A. wrote the manuscript and prepared figures, with all authors providing comment.

FUNDING

Curtin Research Fellowship [CRF130006 to M.A.]; Raine Priming Grant from the Raine Medical Research Foundation [M.A.]; National Health and Medical Research Council of Australia [1011505 to J.I.-T.H.]; Telethon-Perth Children's Hospital Research Fund [J.I.-T.H.]; NHMRC RD Wright Biomedical Research Fellowship [1085842 to K.D.G.P.]; Work was supported by computational resources provided by the Australian Government through the Pawsey Supercomputing Centre under the National Computational Merit Allocation Scheme [pa6/pawsey0196].

CONFLICT OF INTEREST

None declared.

REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19-25.
- Aoki, K., Meng, G., Suzuki, K., Takashi, T., Kameoka, Y., Nakahara, K., . . . Kasai, M. (1998). RP58 associates with condensed chromatin and mediates a sequence-specific transcriptional repression. *J Biol Chem*, 273(41), 26698-26704.
- Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *Journal of Chemical Physics*, 126(1). doi:Artn 014101 10.1063/1.2408420
- Cohen, J. S., Srivastava, S., Farwell Hagman, K. D., Shinde, D. N., Huether, R., Darcy, D., . . . Fatemi, A. (2017). Further evidence that de novo missense and truncating variants in ZBTB18 cause intellectual disability with variable features. *Clin Genet*, 91(5), 697-707. doi:10.1111/cge.12861
- Cruickshank, M. N., Dods, J., Taylor, R. L., Karimi, M., Fenwick, E. J., Quail, E. A., . . . Ulgiati, D. (2015). Analysis of tandem E-box motifs within human Complement receptor 2 (CR2/CD21)

promoter reveals cell specific roles for RP58, E2A, USF and localized chromatin accessibility. *Int J Biochem Cell Biol*, 64, 107-119. doi:10.1016/j.biocel.2015.03.016

de Munnik, S. A., Garcia-Minaur, S., Hoischen, A., van Bon, B. W., Boycott, K. M., Schoots, J., . . . Brunner, H. G. (2014). A de novo non-sense mutation in ZBTB18 in a patient with features of the 1q43q44 microdeletion syndrome. *Eur J Hum Genet*, 22(6), 844-846. doi:10.1038/ejhg.2013.249

Deciphering Developmental Disorders, S. (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature*, 542(7642), 433-438. doi:10.1038/nature21062

Depienne, C., Nava, C., Keren, B., Heide, S., Rastetter, A., Passemard, S., . . . Mignot, C. (2017). Genetic and phenotypic dissection of 1q43q44 microdeletion syndrome and neurodevelopmental phenotypes associated with mutations in ZBTB18 and HNRNPU. *Hum Genet*, 136(4), 463-479. doi:10.1007/s00439-017-1772-0

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh ewald potential. *J. Chem. Phys.*, 103, 8577-8592.

Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin Davis, B., . . . Tang, S. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med*, 17(7), 578-586. doi:10.1038/gim.2014.154

Fuks, F., Burgers, W. A., Godin, N., Kasai, M., & Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J*, 20(10), 2536-2544. doi:10.1093/emboj/20.10.2536

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*, 10(5), 449-461. doi:10.1517/17460441.2015.1032936

Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes. *Journal of Molecular Biology*, 330(4), 891-913. doi:10.1016/S0022-2836(03)00610-7

Gupta, A., Tsai, L. H., & Wynshaw-Boris, A. (2002). Life is a journey: a genetic look at neocortical development. *Nat Rev Genet*, 3(5), 342-355.

Haas, M. A., Ngo, L., Li, S. S., Schleich, S., Qu, Z., Vanyai, H. K., . . . Heng, J. I. (2016). De Novo Mutations in DENR Disrupt Neuronal Development and Link Congenital Neurological Disorders to Faulty mRNA Translation Re-initiation. *Cell Rep*, 15(10), 2251-2265. doi:10.1016/j.celrep.2016.04.090

Hashimoto, H., Wang, D., Horton, J. R., Zhang, X., Corces, V. G., & Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol Cell*, 66(5), 711-720 e713. doi:10.1016/j.molcel.2017.05.004

Hemming, I. A., Forrest, A. R., Shipman, P., Woodward, K. J., Walsh, P., Ravine, D. G., & Heng, J. I. (2016). Reinforcing the association between distal 1q CNVs and structural brain disorder: A case of a complex 1q43-q44 CNV and a review of the literature. *Am J Med Genet B Neuropsychiatr Genet*. doi:10.1002/ajmg.b.32427

635 Heng, J. I., Nguyen, L., Castro, D. S., Zimmer, C., Wildner, H., Armant, O., . . . Guillemot, F. (2008).
636 Neurogenin 2 controls cortical neuron migration through regulation of Rnd2. *Nature*,
637 455(7209), 114-118. doi:nature07198 [pii] 10.1038/nature07198

638 Heng, J. I., Qu, Z., Ohtaka-Maruyama, C., Okado, H., Kasai, M., Castro, D., . . . Tan, S. S. (2013).
639 The zinc finger transcription factor RP58 negatively regulates Rnd2 for the control of neuronal
640 migration during cortical development. *Cereb Cortex*. doi:bht277 [pii] 10.1093/cercor/bht277

641 Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint
642 solver for molecular simulations. *Journal of Computational Chemistry*, 18(12), 1463-1472.
643 doi:10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.3.Co;2-L

644 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison
645 of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics*, 79(2),
646 926-935. doi:10.1063/1.445869

647 Krepl, M., Zgarbova, M., Stadlbauer, P., Otyepka, M., Banas, P., Koca, J., . . . Sponer, J. (2012).
648 Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER
649 force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J Chem Theory Comput*, 8(7),
650 2506-2520. doi:10.1021/ct300275s

651 Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015).
652 ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.
653 *J Chem Theory Comput*, 11(8), 3696-3713. doi:10.1021/acs.jctc.5b00255

654 Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012).
655 MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of*
656 *Chemical Theory and Computation*, 8(9), 3314-3321. doi:10.1021/ct300418h

657 Mitchelmore, C., Kjaerulff, K. M., Pedersen, H. C., Nielsen, J. V., Rasmussen, T. E., Fisker, M. F., . . .
658 Jensen, N. A. (2002). Characterization of two novel nuclear BTB/POZ domain zinc finger
659 isoforms. Association with differentiation of hippocampal neurons, cerebellar granule cells,
660 and macroglia. *J Biol Chem*, 277(9), 7598-7609. doi:10.1074/jbc.M110023200

661 Najafabadi, H. S., Garton, M., Weirauch, M. T., Mnaimneh, S., Yang, A., Kim, P. M., & Hughes, T. R.
662 (2017). Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc
663 finger DNA binding. *Genome Biol*, 18(1), 167. doi:10.1186/s13059-017-1287-y

664 Nord, A. S., Pattabiraman, K., Visel, A., & Rubenstein, J. L. R. (2015). Genomic perspectives of
665 transcriptional regulation in forebrain development. *Neuron*, 85(1), 27-47.
666 doi:10.1016/j.neuron.2014.11.011

667 Ohtaka-Maruyama, C., Hirai, S., Miwa, A., Heng, J. I., Shitara, H., Ishii, R., . . . Okado, H. (2013).
668 RP58 regulates the multipolar-bipolar transition of newborn neurons in the developing
669 cerebral cortex. *Cell Rep*, 3(2), 458-471. doi:S2211-1247(13)00019-3 [pii]
670 10.1016/j.celrep.2013.01.012

671 Okado, H., Ohtaka-Maruyama, C., Sugitani, Y., Fukuda, Y., Ishida, R., Hirai, S., . . . Kasai, M. (2009).
672 The transcriptional repressor RP58 is crucial for cell-division patterning and neuronal survival

in the developing cortex. *Dev Biol*, 331(2), 140-151. doi:S0012-1606(09)00277-2 [pii]
10.1016/j.ydbio.2009.04.030

Onufriev, A., Bashford, D., & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins-Structure Function and Bioinformatics*, 55(2), 383-394. doi:10.1002/prot.20033

Pacary, E., Heng, J., Azzarelli, R., Riou, P., Castro, D., Lebel-Potter, M., . . . Guillemot, F. (2011). Proneural transcription factors regulate different steps of cortical neuron migration through Rnd-mediated inhibition of RhoA signaling. *Neuron*, 69(6), 1069-1084. doi:S0896-6273(11)00116-4 [pii] 10.1016/j.neuron.2011.02.018

Parrinello, M., & Rahman, A. (1981). Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Journal of Applied Physics*, 52(12), 7182-7190. doi:Doi 10.1063/1.328693

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, 92(11), 3817-3829. doi:10.1529/biophysj.106.097782

Peters, M. B., Yang, Y., Wang, B., Fusti-Molnar, L., Weaver, M. N., & Merz, K. M., Jr. (2010). Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput*, 6(9), 2935-2947. doi:10.1021/ct1002626

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endeke, S., Schwarzmayr, T., . . . Strom, T. M. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*, 380(9854), 1674-1682. doi:10.1016/S0140-6736(12)61480-9

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, 16(6), 276-277.

Ruscio, J. Z., & Onufriev, A. (2006). A computational study of nucleosomal DNA flexibility. *Biophysical Journal*, 91(11), 4121-4132. doi:10.1529/biophysj.106.082099

Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 3(2), 198-210. doi:10.1002/wcms.1121

Segal, D. J., Crotty, J. W., Bhakta, M. S., Barbas, C. F., 3rd, & Horton, N. C. (2006). Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J Mol Biol*, 363(2), 405-421. doi:10.1016/j.jmb.2006.08.016

Shields, G. C., Laughton, C. A., & Orozco, M. (1997). Molecular dynamics simulations of the d(T·A·T) triple helix. *J. Am. Chem. Soc.*, 119, 7463-7469.

Shields, G. C., Laughton, C. A., & Orozco, M. (1998). Molecular dynamics simulation of a PNA.DNA.PNA triple helix in aqueous solution. *J. Am. Chem. Soc.*, 120, 5895-5904.

Sousa da Silva, A. W., & Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes*, 5, 367. doi:10.1186/1756-0500-5-367

van der Schoot, V., de Munnik, S., Venselaar, H., Elting, M., Mancini, G. M. S., Ravenswaaij-Arts, C. M. A., . . . Stevens, S. J. C. (2018). Toward clinical and molecular understanding of

- pathogenic variants in the ZBTB18 gene. *Mol Genet Genomic Med*, 6(3), 393-400.
doi:10.1002/mgg3.387
- Weiser, J., Shenkin, P. S., & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry*, 20(2), 217-230. doi:10.1002/(Sici)1096-987x(19990130)20:2<217::Aid-Jcc4>3.0.Co;2-A
- Xiang, C., Baubet, V., Pal, S., Holderbaum, L., Tatard, V., Jiang, P., . . . Dahmane, N. (2011). RP58/ZNF238 directly modulates proneurogenic gene levels and is required for neuronal differentiation and brain expansion. *Cell Death Differ.* doi:cdd2011144 [pii] 10.1038/cdd.2011.144
- Zgarbova, M., Luque, F. J., Sponer, J., Cheatham, T. E., 3rd, Otyepka, M., & Jurecka, P. (2013). Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput*, 9(5), 2339-2354. doi:10.1021/ct400154j

FIGURES LEGENDS

Figure 1. Homology model for ZBTB18 in complex with a 15-residue duplex derived from E1 sequence (TAACAGATGTCTGTC), shown in “front” (C2H2 domains 1 and 4 facing viewer) and “rear” (C2H2 domains 2 and 3 facing viewer) views. Zinc Fingers 1-4 (coloured as N- to C-terminal rainbow, blue to red), as well as zinc atoms (grey spheres) and DNA residues (dA – white; dC – pink; dG – pale green; dT – tan) are shown. Residues of ZBTB18 impacted by disease-associated missense variants (Leu434, Tyr447, Asn461, Asn464, Glu486 and Arg495) are shown in the lateral panels, with hydrogen bonds in dashed green lines. Arg464, Gln486 and Asn461 are rotated slightly around the x-axis relative to the full complex views. See Supplementary Figure S1B and Table S1 for degree of amino acid sequence conservation and binding energy decomposition for these six residues, respectively.

Figure 2. Structural analysis of ZBTB18-DNA complexes. (A) Model of ZBTB18 zinc fingers 1-4 bound to the E1 sequence, coloured rainbow (blue to red) from N- to C-terminal, with zinc atoms as grey spheres, and DNA residues coloured as follows: dA – white; dC – pink; dG – pale green; dT – tan. (B) Major interactions of Tyr458, Asn461, and His493 of ZBTB18 with E1, E1^{mut}, E2 and E2^{mut} sequences. Hydrogen bonds (dashed green lines), coordination bonds (dashed purple lines), and residues involved in hydrophobic contacts in at least one panel (transparent spheres) highlighted. Views rotated slightly relative to panel (A) to improve visibility of contacts.

Figure 3. Structural analysis of DNA complexes with ZBTB18 missense variants. (A) Model of ZBTB18 N461S missense variant zinc fingers 1-4 bound to the E1 sequence. (B) Major interactions of Tyr458, Ser461, and His493 of the N461S variant with E1, E1^{mut}, E2 and E2^{mut} sequences. Hydrogen bonds (dashed green lines), coordination bonds (dashed purple lines), and residues involved in hydrophobic contacts in at least one panel (transparent spheres) highlighted. Views rotated slightly relative to panel (A) to improve visibility of contacts. In all panels, the protein is coloured rainbow (blue to red) from N- to C-terminal, with zinc atoms as grey spheres, and DNA residues coloured as follows: dA – white; dC – pink; dG – pale green; dT – tan. (C) Model of ZBTB18 R495G missense variant zinc fingers 1-4 bound to the E1 sequence. (D) Major interactions of Tyr458, Asn461, and His493 of N461S variant with E1, E1^{mut}, E2 and E2^{mut} sequences. Hydrogen bonds (dashed green lines), coordination bonds (dashed purple lines), and residues involved in hydrophobic contacts in at least one panel (transparent spheres) highlighted. Views rotated slightly relative to panel (C) to improve visibility of contacts.

Figure 4. Functional analysis of ZBTB18 and its disease-associated variants during DNA-binding and transcriptional regulation, *in vitro*. (A-D) EMSAs conducted with biotinylated DNA probes in the

presence of WT, N461S and R495G. Control represents experiments with lysate harvested from mock-transfected cells. Square brackets indicate binding complexes of very high molecular weight detected in EMSAs with all four DNA probes, suggesting these bands to comprise non-specific factors in all cell lysates. (A) WT, N461S and R495G produced complexes with E1 (arrows), with R495G forming higher molecular weight complexes (arrowhead). (B) N461S and R495G formed complexes (closed arrowhead and arrow) with an E1^{mut} probe, with R495G forming an additional, prominent high molecular weight complexes (open arrowhead). A very weak signal representing a complex between WT and E1^{mut} was detected (arrow). (C) WT, N461S and R495G formed complexes with the E2 probe (arrows), with WT and R495G forming additional molecular weight complexes which were common (closed arrowhead) and unique (open arrowhead) to these two species. (D) No specific binding complexes were observed when assessed with E2^{mut} probe. (E) Luciferase assays with a BS10 reporter show that WT ZBTB18 suppresses reporter activity while N461S does not. R495G potentiates reporter activity. (F) Luciferase reporter assays with an *Rnd2* 3' enhancer show that N461S and R495G potentiate reporter activity. (G) Luciferase reporter assays utilising *Rnd2* 3' enhancer constructs harbouring native E1 and E2 motifs (black segments); or mutated E1^{mut} and E2^{mut} motifs (grey segments) confirm sequence-specificity of WT ZBTB18 to repress activity from this reporter through E1 and E2, and not via E1^{mut} and E2^{mut}. (H) With N461S, statistically significant transcriptional activation by N461S was observed in constructs bearing E1+E2 (p<0.05) or E1^{mut}+E2^{mut} (p<0.01) motifs, but not an E1^{mut}+E2 motif or an E1+E2^{mut} motif. When compared to WT ZBTB18, the N461S variant has reduced capacity for repression in the context of E1+E2 (p<0.1), E1^{mut}+E2 (p<0.05) and E1+E2^{mut} (p<0.0001) motif pairs; as well as apparent activation of the E1^{mut}+E2^{mut} (p<0.01) motif pair. (I) R495G augments luciferase reporter activity across all constructs. Values represent fold-change of signal relative to control condition ± SEM. For all graphs Kruskal–Wallis performed, followed by a two-way Mann–Whitney U test; *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 compared to control; #P<0.05, ##P<0.01, ###P<0.001, ####P<0.0001 compared to WT ZBTB18.

Figure 5. The effects of ZBTB18 and its disease-associated variants on radial migration of E14.5-born cortical cells within the embryonic E17.5 mouse cerebral cortex. (A–E) *In utero* electroporation of E14.5 mouse cortex with either; control shRNA and an empty vector (A), *Zbtb18* shRNA and an empty vector (B), *Zbtb18* shRNA and WT (C), *Zbtb18* shRNA and N461S (D), or *Zbtb18* shRNA and R495G (E). Brains were collected three days later (E17.5), white bars indicate the cortical plate (CP), intermediate zone (IZ), and ventricular zone/subventricular zone (VS/SVZ). (F) Quantification of distribution of GFP-labelled cells within the CP, IZ, and VS/SVZ. (G) Quantification of GFP-labelled cells within the upper, middle, and lower CP (uCP, mCP, and ICP, respectively). (H–L) Representative images of neurons within either the cortical plate (CP) or the intermediate zone (IZ) of *in utero* electroporation brains (E14.5, collected E17.5) for either; control shRNA and an empty vector (H), *Zbtb18* shRNA and an empty vector (I), *Zbtb18* shRNA and WT (J), *Zbtb18* shRNA vector and N461S

(K), or *Zbtb18* shRNA and R495G (L). White arrowheads indicate round cells, unfilled arrowheads indicate multipolar cells, and indented arrowheads indicate bipolar cells. (M-N) Proportions of cells exhibiting different morphologies (round, uni/bipolar, and multipolar) in each brain region. (O) Leading processes of GFP+ cells within the CP were analyzed. FLAG-tagged WT and R495G was immunodetectable in cortical cells, but not in N461S treatment (Supplementary Figure S8). For all graphs Kruskal – Wallis performed, followed by a two-way Mann – Whitney U test (Mean $\pm$ SEM $n \geq 6$ (≥ 161 processes per brain for (o)); * $P < 0.05$, ** $P < 0.01$ compared to control; # $P < 0.05$, ### $P < 0.01$ compared to *Zbtb18* shRNA. Scale bar in (e) represents 200 μm . Scale bar in (l) represents 25 μm .

TABLES

Table 1. Binding energies for ZBTB18-DNA complexes.

ZBTB18 variant	ΔG_{bind} (kcal/mol)			
	DNA motifs			
	E1	E1 ^{mut}	E2	E2 ^{mut}
WT	-119.8 $\pm$ 0.3	-117.1 $\pm$ 0.3	-125.2 $\pm$ 0.3	-94.0 $\pm$ 0.3
N461S	-129.6 $\pm$ 0.3	-119.2 $\pm$ 0.4	-108.1 $\pm$ 0.3	-124.5 $\pm$ 0.4
R495G	-126.4 $\pm$ 0.3	-126.0 $\pm$ 0.3	-126.9 $\pm$ 0.4	-102.0 $\pm$ 0.3

Table 2. Binding energy decompositions for selected residues in ZBTB18-DNA complexes. Three residues (Tyr458, Asn461 and His493) with significant contributions to DNA-binding are indicated in grey. Binding energy calculations for additional residues (Leu434, Tyr447, Arg464, Glu486, Arg495) impacted by pathogenic missense variants are also listed. A comprehensive list describing binding energy decomposition values for all residues making a major contribution to the binding energy in at least one complex is provided as Supplementary Table S4.

Protein	ΔG_{bind} (kcal/mol)											
	WT				N461S				R495G			
DNA	E1	E1 ^{mut}	E2	E2 ^{mut}	E1	E1 ^{mut}	E2	E2 ^{mut}	E1	E1 ^{mut}	E2	E2 ^{mut}
Leu434	-0.1	+0.1	-0.0	-0.1	-0.1	-0.1	-0.1	+0.0	-0.0	-0.1	+0.1	-0.1
Tyr447	-2.1	-0.5	+0.0	-0.1	-1.6	-2.9	+0.1	-1.7	-0.9	-3.4	+0.1	+0.0
Tyr458	-3.2	-0.7	-3.4	-0.7	-2.2	-2.5	-3.1	-4.3	-1.9	-3.5	-1.1	-4.2
Asn/Ser461	-8.6	-2.3	-9.6	-1.4	-2.5	-1.5	-2.3	-3.3	-7.7	-3.9	-3.6	-5.5
Arg464	-4.0	-3.1	-4.9	-3.8	-3.9	-2.7	-4.0	-4.3	-4.0	-5.5	-4.5	-3.1
Glu486	-3.1	-2.7	-4.3	-2.5	-2.8	-2.9	-2.9	-3.4	-3.6	-2.4	-3.5	-2.8
His493	-5.9	-0.3	-0.7	-2.9	-5.3	-3.9	-5.8	-4.2	-2.3	-1.3	-4.9	-4.2

819

Arg/Gly495	-0.7	-2.5	-0.3	-2.7	-1.2	-1.4	-0.6	-1.0	+0.0	+0.0	+0.0	-0.0

Figure 1

Figure 2

A

B

Figure 3

A

C

B

D

Figure 4

Figure 5

Supplementary Figure 2

Supplementary Figure 3

Supplementary Figure 4

Supplementary Figure 5

Supplementary Figure 6

Supplementary Figure 7

Supplementary Figure 8

Supplementary Figure 9

Supplementary Figure 10

SUPPLEMENTARY DATA

Supplementary data are available at Human Mutation online.

Legends to Supplementary Figures

Supplementary Figure S1. (A) Planar representation of the ZBTB18 polypeptide featuring an N-terminal BTB domain (dark blue) and four zinc fingers (in yellow) at its C-terminus, with pathogenic missense mutations indicated (Cohen et al., 2017; Depienne et al., 2017; Rauch et al., 2012). (B) CLUSTAL Omega (ver1.2.4) alignment of the C-terminal region (aa371-531) of ZBTB18 comprising four zinc finger regions (ZF1 to ZF4). Amino acids highlighted with yellow and red represent conservative and non-conservative changes, respectively. Disease-associated missense variants are identified with an arrow. Affected residues are highly conserved among vertebrates.

Supplementary Figure S2. Representative structures of ZBTB18-DNA complexes from molecular dynamics simulations. (A) Wildtype-E1mut complex. (B) Wildtype-E2 complex. (C) Wildtype-E2mut complex. (D) N461S-E1mut complex. (E) N461S-E2 complex. (F) N461S-E2mut complex. (G) R495G-E1mut complex. (H) R495G-E2 complex. (I) R495G-E2mut complex.

Supplementary Figure S3. RMSDs over time for ZBTB18-DNA complexes in the three selected simulations for each complex. The grey region indicates the regions of the simulations used for MMGB/SA calculations. (A) Wildtype-E1 complex. (B) N461S-E1 complex. (C) R495G-E1 complex. (D) Wildtype-E1 mutant complex. (E) N461S-E1 mutant complex. (F) R495G-E1 mutant complex. (G) Wildtype-E2 complex. (H) N461S-E2 complex. (I) R495G-E2 complex. (J) Wildtype-E2 mutant complex. (K) N461S-E2 mutant complex. (L) R495-E2 mutant complex.

Supplementary Figure S4. Luciferase reporter assays performed with the Rnd2 3' enhancer in the presence of 125 ng of ZBTB18 WT construct alone, or in combination with 125ng of N461S (A) or R495G (B). Quantities of expression plasmid were balanced to 250ng per well of transfected cells with an empty construct. Signals represent fold change relative to baseline control (reporter transfected with empty mammalian constructs encoding only the FLAG epitope tag). All transfection conditions were balanced for equal quantities of plasmid DNA. Bars represent average $\pm$ SEM. A Kruskal-Wallis analysis was employed to test multiple conditions, while a Mann-Whitney U test was applied for statistical analyses between 2 treatments. **p < 0.01 compared to control.

Supplementary Figure S5. An illustrative summary of the transcriptional activity and DNA binding of ZBTB18, and its missense variants, in the context of the Rnd2 3' enhancer reporter. The strength of reporter signal is represented by the number of arrows, as well as the size of the arrows. (A) WT ZBTB18 binds E1 and E2 motifs to repress reporter activity, but repressor activity is attenuated when these motifs are mutated to E1mut or E2mut. (B) In contrast, the N461S variant binds E1, E1mut and E2 to mediate reporter activity; however, the intensity of augmented reporter signal is dependent on the contextual presentation of each of these DNA motifs, since N461S has reduced capacity for repression in the context of the E1mut+E2 and E1mut+E2mut motif pairs. (C) On the other hand, the R495G variant binds E1, E1mut and E2 and significantly augments reporter activity, possibly through the formation of a protein complex with additional factors, as evidenced in our EMSA assays.

Supplementary Figure S6. Localisation of FLAG-tagged wildtype ZBTB18, as well as the N461S and R495G variants in transiently transfected HeLa cells. (A) Representative images of the two main localisation patterns (Pancellular [a'] and Exclusively Nuclear), and two unique nuclear localisation patterns (Uniform [a''] and Punctate [a''']). Outline of visibly transfected cells (dashed green lines) are marked by GFP co-expression (not shown). (B) The nuclear localisation of N461S was significantly different from WT 24 hours post-transfection. (C) The pancellular localisation of N461S was significantly different from WT 24 hours post-transfection. (D) For the proportion of cells exhibiting exclusively nuclear localisation, punctate distribution of R495G was significantly increased compared to WT at 24 hours. (E-G) The nuclear localisation and nuclear immunostaining patterns for WT, N461S and R495G were not significantly different at 48 hours post-transfection. Bars represent mean $\pm$ SEM n=3. Scale bar for a''' represents 12.5 μ m. ***p<0.001 compared to WT.

Supplementary Figure S7. Low steady-state levels of N461S in transiently transfected cells. (A-B) There was a significant decrease in the proportion of FLAG+GFP+/GFP+ cells transfected with the N461S construct, yet the proportion of GFP-transfected cells across conditions was not significantly different. (C) Western blot to immunodetect WT, N461S and R495G as FLAG-epitope tagged proteins. (D) There was a significant decrease in the steady-state levels of FLAG-N461S signal compared to WT (n=3 biological replicates per condition). (E) Western blot to immunodetect WT, N461S and R495G as FLAG-epitope tagged proteins 0, 4, and 8 hrs post treatment with 10 μ M MG132. (F) There was a significant increase in the levels of FLAG-N461S signal during the time course; with the trend not significantly different to either WT or R495G, but significantly different to control (n=3 biological replicates per condition). **p<0.05, ***p<0.001, ****p<0.0001 compared to control, while ##p<0.05, ###p<0.001 compared to WT.

Supplementary Figure S8. Representative coronal sections of E17.5 embryonic cerebral cortices describing the Cortical Plate (CP) distribution and immunoreactivity of Zbtb18 shRNA-treated neurons co-electroporated with expression constructs encoding FLAG-tagged WT, N461S and R495G. The lower- (ICP), medial- (mCP) and upper CP (uCP) subdivisions are defined. Co-delivery of N461S augments the uCP accumulation of Zbtb18 shRNA-treated cells, compared to cells rescued with WT construct. In contrast, co-delivery of R495G failed to rescue the radial positioning of shRNA-treated neurons (see Figure 5G for quantification). Robust immunoreactivity was observed for WT and R495G electroporated cells, but N461S immunoreactivity was not detected on any section for all brains tested (n=6-8 per condition). Asterisk represents spurious immunoreactivity adjacent to a blood vessel.

Supplementary Figure S9. The effects of ZBTB18 and its disease-associated variants on radial migration E14.5-born cortical cells within the embryonic E16.5 mouse cerebral cortex. (A) In utero electroporation of E14.5 mouse cortex with either; control shRNA together with an empty expression vector; Zbtb18 shRNA and an empty vector; Zbtb18 shRNA and WT; Zbtb18 shRNA and N461S; or Zbtb18 shRNA and R495G. Brains were collected two days later (E16.5), white bars delineate the cortical plate (CP), intermediate zone (IZ), and ventricular zone/subventricular zone (VS/SVZ). (B) Quantification of distribution of GFP-labelled cells within the CP, IZ, and VS/SVZ. *P<0.05, **P<0.01 compared to control; #P<0.05, ##P<0.01 compared to Zbtb18 shRNA. Scale bar in (e) represents 200 μ m.

Supplementary Figure S10. Summary illustration of the effects of ZBTB18 pathogenic variants on radial migration in our rescue assays. (A) Bipolar-shaped cortical neurons leave the VZ/SVZ and adopt a multipolar morphology to initiate somal translocation within the IZ before transitioning to a bipolar shaped neuron to migrate into the CP by radialglial-guided locomotion. Zbtb18 is critical to the development of embryonic cortical project neurons at multiple steps. For example, Zbtb18 expression is critical for mediating neuroprogenitor cell cycle exit within the VZ/SVZ (Okado et al., 2009), as well as for the generation of appropriate numbers of neurons, rather than glia (Xiang et al., 2011). As immature neurons transit within the SVZ through to the IZ and CP, Zbtb18 regulates downstream target genes including Neurog2 and Rnd2 to drive multipolar-to-bipolar transition (Ohtaka-Maruyama et al., 2013) and radial migration into the CP. (B) Knockdown of Zbtb18 disrupts the radial migration of newborn cortical neurons, culminating in their multipolar-to-bipolar transition within the IZ and lower CP, and this migration defect is corrected by co-delivery of ZBTB18 (Figure 5A-C,F,G; and Ohtaka- Maruyama et al., 2013; Heng et al., 2013). (C) In contrast, co delivery of N461S leads to exuberant radial migration into the CP, with the accumulation of multipolar-shaped cells which is suggestive of disrupted radialglial-guided locomotion (Figure 5D,F,G). (D) Co-delivery of R495G leads to an exacerbation of the migration defect of shRNA treated cells, culminating in a failure in multipolar-to-bipolar transition of cells and their accumulation in the IZ (Figure 5E-G).

Supplementary Video Legend

Supplementary Video S1. Three-dimensional rotating model of the WT ZBTB18-E1 complex, with zinc finger domains highlighted.

Supplementary Tables

Supplementary Table S1. Summary of disease-associated missense variants of *ZBTB18*, their predicted impact on the polypeptide product, and their associated neuropathological traits.

Variant	Effect of missense variant	PolyPhen2 and SIFT scores	Intellectual disability? (Y/N)	neuropathological traits ^(reference)
L434P	Non-polar hydrophobic to non-polar hydrophobic (conserved)	PolyPhen2: 0.999 (probably damaging) SIFT:0.00 (damaging)	Y	Epilepsy and flapping of hands ⁽¹¹⁾
Y447H	Polar neutral hydrophilic to charged basic hydrophilic (non-conserved)	PolyPhen2: 0.969 (probably damaging) SIFT:0.00 (damaging)	Y	Seizures, Microcephaly (OFC 52 cm [-2.0]) ⁽¹³⁾
N461S	polar neutral hydrophilic to polar neutral hydrophilic (conserved)	PolyPhen2: 0.996 (probably damaging) SIFT:0.48 (tolerated)	Y	Global CC hypoplasia, inferior cerebellar vermis hypoplasia ⁽¹⁵⁾
R464H	Charged basic hydrophilic to Charged basic hydrophilic (conserved)	PolyPhen2: 0.999 (probably damaging) SIFT:0.00 (damaging)	Y	Hyperactive, tick disorder, obsessive–compulsive behaviour, short and dysgenic CC, likely microcephaly (OFC 50 cm [-2.25]) ⁽¹¹⁾
R464C	Charged basic hydrophilic to polar neutral hydrophilic (non-conserved)	PolyPhen2: 1.00 (probably damaging) SIFT:0.00 (damaging)	Y	Limited speech and likely macrocephaly (OFC 58cm [90 th]) ⁽⁸⁾
Q486E	Polar neutral hydrophilic to Charged acidic hydrophilic (non-conserved)	PolyPhen2: 0.954 (probably damaging) SIFT:0.00 (damaging)	Y	Unknown ⁽⁸⁾
R495G	Charged basic hydrophilic to non-polar neutral hydrophobic (non-conserved)	PolyPhen2: 0.996 (probably damaging) SIFT:0.00 (damaging)	Y	Unknown ⁽¹⁴⁾

PolyPhen2 run using <http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi>

SIFT run using http://sift.jcvi.org/www/SIFT_enst_submit.html

Supplementary Table S2. Templates for each of the C2H2 subdomains of the ZBTB18 zinc finger domain.

Region	Template (PDB ID) (Reference)	Sequence alignment/region modelled ^a
C2H2 1	3UK3	FMCPLCNKVFSPHILQIHLSTH RECSYCGKFFRSNYLNIHLRTH
C2H2 2	1MEY (Kim and Berg, 1996)	PTCSLCGKTFSCMYTLKRHERTH YKCPECGKSFSQSSNLQKHQRTH
C2H2 3	1MEY (Kim and Berg, 1996)	YTCTQCGKSFQYSHNLSRHAVVH YKCPECGKSFSQSSDLQKHQRTH
C2H2 4	5K5H (Hashimoto et al., 2017)	HACKWCERRFTQSGDLYRHIRKFH YECYICHAFTQSGTMKMHILOKH

^aIn this column, the first row indicates the ZBTB18 sequence; the second row indicates the template sequence. Regions not shown in this table were modelled *de novo* by Prime during model building.

Supplementary Table S3. Templates for the E1 and E2 probe sequences.

DNA probe	Template (PDB ID) (Reference)	Sequence alignments ^a
E1	5KE6 (Hashimoto et al., 2017)	TAACAG <u>ATG</u> TCTGTC GAGGTGTGGC AGGGAAAAGCCCGGG
E2	5KL2 (Hashimoto et al., 2016)	GGGCAG <u>ATG</u> GGAGTA GCGTGGGAGT AGGGAAAAGCCCGGG

^aIn each row of this column, the first line indicates the DNA probe sequences to be modelled, the second line indicates the DNA sequence contained in the PDB template and the third line indicates the DNA sequence from PDB 2I13.

Supplementary Table S4. Binding energy decomposition for all residues making a major contribution to the binding energy in at least one complex.^a

Residue	WT				N461S				R495G			
	E1	E1 ^{mut}	E2	E2 ^{mut}	E1	E1 ^{mut}	E2	E2 ^{mut}	E1	E1 ^{mut}	E2	E2 ^{mut}
Lys386	-2.9	-4.4	-1.2	-2.3	-2.7	-3.4	-1.4	-1.9	-2.2	-4.0	-1.3	-4.5
Ser390	-1.2	-0.9	-3.3	-2.1	-2.1	-0.4	-1.8	-0.4	-0.6	-1.8	-3.6	-1.9
His392	-2.1	-0.8	-1.7	-5.0	-3.8	-1.5	-3.1	-3.4	-2.0	-1.1	-1.8	-4.2
His397	-3.4	-4.2	-1.5	-1.4	-3.7	-4.6	-0.3	-2.2	-4.6	-3.6	-0.3	-2.0
Thr400	-3.2	-2.6	-1.3	-0.6	-4.4	-3.8	-0.4	-1.9	-4.8	-3.6	-0.7	-0.3
Arg403	-1.5	-3.3	-2.2	-1.9	-4.9	-5.2	-1.8	-3.5	-5.0	-2.1	-0.9	-0.4
Arg409	-0.5	-0.2	-0.3	-0.9	-0.3	-0.7	-2.7	-3.8	-0.2	-1.8	-0.3	-1.7
Lys426	-2.5	-3.8	-4.0	-3.1	-4.2	-3.8	-2.7	-3.2	-4.0	-3.3	-4.4	-3.2
Phe428	-0.2	-3.4	-2.4	-1.4	-0.3	-0.4	-1.9	-4.6	-3.0	-3.4	-2.9	-0.4
Ser429	-0.2	-4.9	-2.8	-1.2	+0.1	+0.1	-1.5	-5.9	-2.1	-3.3	-3.5	-1.0
Cys430	-0.5	+0.0	-0.9	-1.2	-0.4	-0.2	-0.8	-0.4	-0.8	-0.3	-0.7	-2.6
Tyr432	-4.1	-4.7	-5.9	-5.4	-5.3	-4.3	-5.7	-3.6	-4.5	-4.8	-5.5	-5.4
Thr433	-0.5	-4.1	-2.2	-0.9	-0.6	-1.0	-2.4	-5.4	-2.6	-3.2	-4.1	-1.3
Leu434 ^b	-0.1	+0.1	-0.0	-0.1	-0.1	-0.1	-0.1	+0.0	-0.0	-0.1	+0.1	-0.1
Arg436	-2.5	-4.1	-6.9	-5.0	-3.4	-4.8	-3.7	-2.4	-5.8	-3.5	-5.1	-3.1
His437	-2.0	-2.6	-4.9	-2.4	-2.1	-3.4	-3.7	-4.3	-4.6	-2.8	-4.1	-3.7
Arg439	-1.4	-2.2	-1.0	-0.9	-1.6	-1.5	-1.3	-3.0	-1.5	-4.0	-1.3	-1.2
Thr440	-2.8	-1.4	-2.8	-1.8	-2.3	-1.5	-3.0	-1.5	-3.0	-1.4	-3.6	-2.5
Lys445	-3.5	-0.7	-2.4	-1.2	-4.0	-0.9	-2.3	-0.6	-1.9	-0.6	-2.1	-2.5
Tyr447	-2.1	-0.5	+0.0	-0.1	-1.6	-2.9	+0.1	-1.7	-0.9	-3.4	+0.1	+0.0
Lys454	-4.5	-3.9	-4.4	-3.9	-4.1	-3.0	-4.0	-2.2	-4.4	-4.2	-4.3	-4.9
Phe456	-4.4	-1.0	-3.4	-1.4	-3.7	-0.6	-2.7	-1.1	-2.8	-1.4	-2.5	-4.0
Gln457	-5.3	-1.1	-4.5	-1.0	-5.4	+0.3	-4.0	-0.5	-3.8	-0.4	-3.5	-5.1
Tyr458	-3.2	-0.7	-3.4	-0.7	-2.2	-2.5	-3.1	-4.3	-1.9	-3.5	-1.1	-4.2
Ser459	-3.3	-3.0	-0.3	-1.2	-2.1	-3.5	-0.6	-0.4	-0.8	-3.9	-1.2	-0.8
His460	-3.3	-4.5	-3.3	-2.7	-3.5	-4.4	-4.2	-4.6	-3.0	-4.8	-3.6	-3.8
Asn/Ser461	-8.6	-2.3	-9.6	-1.4	-2.5	-1.5	-2.3	-3.3	-7.7	-3.9	-3.6	-5.5
Ser463	-0.0	-4.4	+0.2	+0.1	-1.0	-4.7	-0.2	-1.0	+0.2	-4.5	-0.4	-0.1
Arg464	-4.0	-3.1	-4.9	-3.8	-3.9	-2.7	-4.0	-4.3	-4.0	-5.5	-4.5	-3.1
His465	-2.6	-4.6	-4.0	-4.4	-1.9	-2.2	-2.8	-2.7	-4.3	-5.6	-3.9	-2.8
Arg471	-0.7	-5.7	-4.0	-4.7	-6.3	-4.8	-1.4	-4.6	-2.9	-6.8	-7.8	-3.2
Arg482	-6.5	-2.3	-3.8	-4.0	-6.5	-6.1	-7.4	-4.6	-4.0	-3.2	-3.9	-4.6
Gln486	-3.1	-2.7	-4.3	-2.5	-2.8	-2.9	-2.9	-3.4	-3.6	-2.4	-3.5	-2.8
Asp489	+2.4	+0.4	+2.8	+2.0	+3.8	+2.9	+1.4	+1.5	+3.3	-0.0	+1.9	+2.2
Tyr491	-1.6	-4.5	-0.2	-3.9	-1.9	-3.8	-0.8	-2.2	-0.2	-3.2	-1.8	-0.1
Arg492	-4.6	-2.7	-4.6	-3.7	-3.2	-4.0	-5.7	-4.5	-4.7	-2.2	-4.3	-4.3
His493	-5.9	-0.3	-0.7	-2.9	-5.3	-3.9	-5.8	-4.2	-2.3	-1.3	-4.9	-4.2
Arg/Gly495	-0.7	-2.5	-0.3	-2.7	-1.2	-1.4	-0.6	-1.0	+0.0	+0.0	+0.0	-0.0
dA/dG 3	+2.3	+1.2	+0.4	+0.2	-0.0	+0.6	-2.8	-1.2	+3.1	+1.2	-1.8	-1.4
dC 4	-3.8	-5.2	-1.2	-3.8	-2.2	-3.2	-4.3	-4.2	-2.2	-4.1	-5.1	-4.1
dA 5	-2.2	-0.3	-1.7	-0.8	-2.1	-2.3	-1.3	-3.5	-1.2	-2.5	-1.7	-1.1
dG 6	-7.5	-1.8	-5.5	-2.8	-5.4	-3.4	-4.8	-3.3	-5.8	-1.5	-3.7	-6.0
dG/dA 9	-2.9	-1.1	-4.1	-0.2	-6.1	-1.3	-4.7	-2.9	-3.5	-2.5	-4.9	+0.9
dG 4'	-3.3	-5.4	-1.9	-4.5	-4.0	-6.0	-0.5	-1.8	-2.5	-4.0	-1.8	-1.9
dG/dC 11'	+0.9	-4.1	-0.2	+0.7	-0.2	-0.2	+0.2	+0.4	-0.3	+0.1	-0.1	+0.5
dG/dT 15'	-0.9	-0.4	-1.4	-2.2	-0.9	-0.5	-1.4	-1.2	-5.1	-1.3	-3.5	-1.9

^aResidues individually accounting for greater than 2.5% of the total binding energy shown in this table. Only total energies (in kcal/mol) are shown. ^bThis residue is also included here as it is the site of a known disease-associated variant (see Table S1) but has very minor contribution to the binding energy in all complexes generated.

Supplementary References

- Hashimoto, H., Wang, D., Steves, A. N., Jin, P., Blumenthal, R. M., Zhang, X., & Cheng, X. (2016). Distinctive Klf4 mutants determine preference for DNA methylation status. *Nucleic Acids Res*, 44(21), 10177-10185. doi:10.1093/nar/gkw774
- Kim, C. A., & Berg, J. M. (1996). A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat Struct Biol*, 3(11), 940-945.