Research Report

Title: Novel *de novo* variant in *EBF3* is likely to impact DNA binding in a patient with a neurodevelopmental disorder and expanded phenotypes: patient report, *in silico* functional assessment, and review of published cases

Running Title: A novel *de novo* variant detected in *EBF3*

Patrick R. Blackburn[1,2]*, Sarah S. Barnett[3]*, Michael T. Zimmermann[4], Margot A. Cousin[4,5], Charu Kaiwar[6,7], Filippo Pinto e Vairo[4,5], Zhiyv Niu[3,8], Matthew J. Ferber[3,5,8], Raul A. Urrutia[9], Duygu Selcen[10], Eric W. Klee[3,4,5,8], Pavel N. Pichurin[8]

[1]*Center for Individualized Medicine, Mayo Clinic, Jacksonville, FL*

[2]*Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL*

[3]*Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN*

[4]*Department of Health Sciences Research, Mayo Clinic, Rochester, MN*

[5]*Center for Individualized Medicine, Mayo Clinic, Rochester, MN*

[6]*Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ*

[7]*Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ*

[8]*Department of Clinical Genomics, Mayo Clinic, Rochester, MN*

[9]*Laboratory of Epigenetics and Chromatin Dynamics, Epigenomics Translational Program, Center for Individualized Medicine. Mayo Clinic, Rochester, MN*

[10]*Department of Neurology, Mayo Clinic, Rochester, MN*

* Denotes equal contribution

Correspondence: Pavel N. Pichurin, M.D. Department of Clinical Genomics, Mayo Clinic, 200 First St SW. Mayo 19 West. Rochester, MN 55905. T 507 284 3215. Email: Pichurin.Pavel@mayo.edu; Eric W. Klee, Ph.D. Department of Health Sciences Research, Mayo Clinic. 200 First St SW. Harwick 3-71. Rochester, MN 55905. T 507 266 5741; 507 284 5569. Email: Klee.Eric@mayo.edu.

ABSTRACT

Pathogenic variants in *EBF3* were recently described in three back-to-back publications in association with a novel neurodevelopmental disorder characterized by intellectual disability, speech delay, ataxia, and facial dysmorphisms. In this report we describe an additional patient carrying a *de novo* missense variant in *EBF3* (c.487C>T**,** p.(Arg163Trp)) that falls within a conserved residue in the zinc knuckle motif of the DNA binding domain. Without a solved structure of the DNA binding domain, we generated a homology-based atomic model and molecular dynamics simulations for EBF3, which predicted decreased DNA affinity for p.(Arg163Trp) compared to wild-type protein and control variants. These data are in agreement with previous experimental studies of EBF1

showing the paralogous residue is essential for DNA binding. The conservation and experimental evidence existing for EBF1 and *in silico* modeling and dynamics simulations to validate comparable behavior of multiple variants in *EBF3* demonstrates strong support for the pathogenicity of p.(Arg163Trp). We show that our patient presents with phenotypes consistent with previously reported patients harboring *EBF3* variants and expands the phenotypic spectrum of this newly identified disorder with the additional feature of a bicornuate uterus.

INTRODUCTION

The early B-cell factor 3 (EBF3) is a member of the Collier/Olf/EBF (COE) family of transcription factors that have a number of crucial developmental roles (Dubois and Vincent 2001). Recently, EBF3 was implicated in a novel neurodevelopmental disorder characterized by intellectual disability with or without CNS malformations, speech delay, hypotonia, ataxia, facial dysmorphisms, and urogenital anomalies (Chao et al. 2016; Sleven et al. 2016; Harms et al. 2016). Three concurrent reports describing a total of 21 patients found several *de novo* heterozygous missense, nonsense, splice site, and small insertions and deletions involving *EBF3* (Chao et al. 2016; Sleven et al. 2016; Harms et al. 2016). Interestingly, all of the missense variants that were identified fell within the highly conserved DNA binding domain (DBD) of EBF3; 4 had *de novo* missense variants within a single codon resulting in mutation of the p.Arg163 residue (Chao et al. 2016; Sleven et al. 2016). This residue falls within the ‘zinc knuckle’ motif, is conserved across human EBF paralogs (EBF1-4), and

has a crucial role in coordinating DNA binding (Hagman et al. 1995; Fields et al. 2008). Functional analysis of EBF1 p.Arg163Ala showed loss of DNA binding (Treiber et al. 2010). Functional studies of *EBF3* variants affecting p.Arg163 have confirmed the essential role of this residue (Chao et al. 2016; Sleven et al. 2016). The EBF3 Arg163 codon falls within a CpG dinucleotide island, which could explain why it likely represents a mutation hotspot for this novel disorder (Chao et al. 2016). EBF3 binds to DNA as a homodimer, but can also heterodimerize with other EBF family members to transactivate target genes (Green and Vetter 2011a). Studies of mutant/wild-type (WT) EBF3 heterodimers show that amino acid substitutions at critical residues could exert a dominant negative effect, thereby reducing DNA binding and subsequent transcriptional activation (Sleven et al. 2016). In this report, we describe another patient with a novel heterozygous c.487C>T (p.(Arg163Trp)) variant in *EBF3* who has significant clinical overlap with other previously described patients. We compare the clinical manifestations of our patient with those of previously reported individuals and also provide structural evidence supporting pathogenicity of the novel c.487C>T (p.(Arg163Trp)) variant found in this case.

RESULTS

Clinical Presentation and Family History

The proband was the product of an uncomplicated pregnancy, delivered at term to a G3P2, 31-year-old mother and 34-year-old father. She was noted at birth to have a markedly distended abdomen and was not voiding appropriately. A blind-

end dimple was noted at the expected location of urethra and catheterization could not be performed. A percutaneous suprapubic cystoscopy was performed and vaginoscopy demonstrated a normal vagina and cervix with a bicornuate uterus. Urethral meatal stenosis was noted and the urethra was subsequently dilated. The suprapubic tube was removed but she had failed voiding trials, prompting vesicostomy and a diagnosis of atonic bladder. During her evaluation at around two years of age, she had generalized hypotonia with global developmental delay; she sat independently at 1 year, was not yet walking at 2 years, and had a 2 to 3 word vocabulary. There was no evidence of ataxia or dystonia; deep tendon reflexes were reported decreased throughout. She had bilateral esotropia, short stature with height less than the 2nd percentile, and weight and head circumference around the 25th percentile for age. On exam, she was grossly nondysmorphic, with down-turned mouth corners, mild retrognathia, a low posterior hairline, and mild pectus excavatum (Figure 1). She had normal evaluations that included brain and total spine magnetic resonance imaging (MRI), electromyography (EMG), chromosomal microarray, and comprehensive biochemical metabolic testing. There were no other family members with a similar constellation of findings; however her mother and maternal grandmother were reported to have strabismus. The proband has two older maternal half-siblings including a sister with a history of attention-deficit/hyperactivity disorder (ADHD) and a brother with ADHD and history of speech delay.

Genomic Analyses

To identify variants of interest, whole exome sequencing was performed on genomic DNA extracted from samples submitted from the proband, biological mother, and biological father; 97% of the exome-capture region was covered at a read depth of 20X or greater. A *de novo* variant (c.487C>T, p.(Arg163Trp)) was identified in *EBF3* which was considered a gene of uncertain significance at the time of analysis. This variant was reported due to *EBF3* being implicated as part of the critical region in the 10q26 microdeletion syndrome, which has features overlapping the clinical phenotype of the proband (Faria et al. 2016). *In silico* analyses predicted that this alteration was deleterious, probably damaging, and disease causing by SIFT, Polyphen-2, and MutationTaster2, respectively (Kumar et al. 2009; Adzhubei et al. 2010; Schwarz et al. 2014). This variant has not been reported in the literature or in publically available databases including ExAC and gnomAD (Lek et al. 2016). *EBF3* is highly intolerant to both missense and loss of function variation and the residue mutated in our patient is conserved across species and paralogs (Figure 2). Targeted Sanger sequencing (forward 5'-ACAACAAATGGTGCAATGCACA-3', reverse 5'-AAAATACAAGTCGGGCATAAAAGGG-3') was used to confirm the variant in the proband and absence of the alteration in parental samples.

Table 1. Variant Information

Position (hg19/GRCh37)	Type	Gene	HGVS cDNA	HGVS protein	Zygosity	Inheritance	SIFT/ Polyphen-2/ MutationTaster2	ExAC/gnomAD Allele Frequency
Chr10:131755589 G>A	Missense	EBF3	NM_001005463.2 c.487C>T	NP_001005463.1 p.(Arg163Trp)	Het	*De novo*	Deleterious/ Probably Damaging/ Disease Causing	N/R

Phenotypic Analyses of Patients with *EBF3* Variants

Comparison of our patient with other recently reported cases with variants affecting the p.Arg163 residue revealed substantial phenotypic overlap (Table 2). All patients had intellectual disability, global developmental delay, speech delay, mild facial dysmorphisms, strabismus, and some urogenital anomaly, such as micropenis, cryptorchidism, urinary retention/reflux, and/or bladder control issues (Table 2). Several structural brain abnormalities were noted in other patients on MRI, but this was not a consistent feature across all patients, including the patient described in this report. Our patient also had several urogenital anomalies including atonic bladder, distal urethral stricture, vesicoureteral reflux, bilateral hydroureter and hydronephrosis, and recurrent urinary tract infections (UTI). Recurrent UTIs have been described in several other patients, possibly suggestive of underlying urogenital malformations, but only one other patient with a c.512G>A (p.(Gly171Asp)) variant has been described who had a diagnosis of atonic bladder (Harms et al. 2016). Our patient also had a bicornuate uterus, which has not been reported in any other patients with *EBF3* intragenic variants, suggesting that this may represent an additional feature associated with this novel disorder.

Molecular Modeling and Dynamics Simulations Demonstrate Loss of DNA Binding for p.(Arg163Trp)

In this study, we employed molecular modeling and physics-based atomic simulation to investigate the effects of p.(Arg163Trp) on DNA binding. Previous work in the EBF3 human paralog EBF1 has demonstrated experimentally that p.Arg163 contacts DNA, intercalating into the minor grove, and that both p.Arg163Trp and p.Arg163Ala lead to loss of DNA binding. Additionally, p.Lys239Ala interacts with DNA, but mutation to alanine had no effect on binding (Treiber et al. 2010). Thus, we generated triplicate simulation of EBF3 dimers bound to DNA for all four sequence contexts: WT, p.Lys239Ala, p.Arg163Trp, and p.Arg163Ala. Simulations revealed consistent increases in EBF3 dynamics for p.Arg163Ala and p.Arg163Trp, while p.Lys239Ala exhibited WT-like behavior (Figure 3). The increased dynamics is due to loss of DNA contact by the zinc knuckle domain (see Supplemental Animation 1), which leads to greater deviations from the native conformation (Figure S1). To quantify the departure of the zinc knuckle from the DNA minor grove, we measured reference distances between the protein and DNA backbone. These distances are significantly and consistently longer for both p.Arg163 variants (see Figure 3 and Figure S2) while the WT and p.Lys239Ala remain stable. MD simulations capture both the short-scale random atomic fluctuations and the large-scale motions of the protein. We applied Principal Component (PC) analysis to our simulation data to identify the large-scale motions therein. We summarized these large-scale motions of EBF3, revealing that loss of contact between the zinc knuckle and DNA leads to large-scale conformations changes characterized by pivoting of EBF3 around the DNA helix axis (Figure S3). Thus, atomic simulations of EBF3 variants agree with

biochemical experiments performed for EBF1 and indicate that p.(Arg163Trp) is likely to lead to loss of DNA affinity.

DISCUSSION

EBF3 is one of four highly related transcription factors in the Collier/Olf/EBF (COE) family found in humans. EBF3 is composed of an N-terminal DNA-binding domain (DBD), an Ig-like/plexins/transcription factors (IPT) domain, a helix-loop-helix (HLH) domain, and a C-terminal domain (Figure 2) (Liberg et al. 2002). The DBD coordinates a zinc ion through a histidine and three cysteine residues that form a 14-residue zinc knuckle motif that is essential for DNA binding (Hagman et al. 1995; Liberg et al. 2002). In humans, there is 97% homology at the amino acid level between EBF1 and EBF3 within the DBD, which suggests that these related proteins have similar DNA-binding properties (Dubois and Vincent 2001). In this report, we describe a patient with a novel p.(Arg163Trp) missense mutation that falls within a highly conserved residue in the zinc knuckle motif. Molecular modeling as well as functional confirmation in patients (see Table 2 for description of functional studies performed) with other variants affecting the p.Arg163 residue, strongly suggest that the variant found in our patient is pathogenic. In addition, our patient showed extensive phenotypic overlap with recently reported patients (Chao et al. 2016; Sleven et al. 2016; Harms et al. 2016). Our patient had several urogenital abnormalities, including a bicornuate uterus, which has not been described in any other patients to date and may represent an expansion of the known phenotype.

EBF3 has been less well studied than other EBF family members, but is known to be expressed in early post-mitotic neurons during development and plays a role in neurogenesis (Garel et al. 1997). In invertebrate model organisms, the *EBF3* orthologs collier in *D. melanogaster* and unc-3 (CeO/E) in *C. elegans* are expressed in mandibular and intercalary segment primordia during head specification in flies (Crozatier et al. 1996; 1999) as well as in chemosensory and developing motor neurons during axonal outgrowth in worms (Prasad et al. 1998). Studies in *Xenopus* revealed that EBF3 ortholog (Xebf3) is activated by XNeuroD and is expressed in primary neurons where it regulates neuronal differentiation (Pozzoli et al. 2001). At the transcriptional level in Xenopus, Xebf2 and Xebf3 appear to have largely overlapping patterns of expression and may have partially redundant functions (Green and Vetter 2011a). This appears to be supported by the fact that *Ebf2* and *Ebf3* knockout mice as well as *EBF2/3* double heterozygous knockout mice have similar phenotypes with defects in neuronal development and olfactory axon growth (Wang et al. 2004). In mice, *Ebf3* is also expressed in Cajal-Retzius (C-R) cells, which are important for the development of the cerebral cortex (Chiara et al. 2012). Together, *Ebf2* and *Ebf3* appear to regulate the migration of C-R cells arising in the cortical hem during corticogenesis, and disruption of these genes leads to defects in neuronal development (Chiara et al. 2012).

Gene expression profiling comparing *Arx* mutant and E14.5 wild-type ventral telencephalic tissues in mice revealed that *Ebf3* is one of the most differentially expressed genes in *Arx* mutant ganglionic eminences where it is not typically expressed (Colasante et al. 2009). *Arx* encodes the Aristaless-related homeobox protein, which is an essential transcription factor involved in patterning, neuronal proliferation and differentiation, and axonal outgrowth (Friocourt and Parnavelas 2010). *Arx* has been shown to be sufficient to repress *Ebf3* expression, and defects in neuronal development in *Arx* mutant mice can be partially rescued through *Ebf3* silencing (Colasante et al. 2009). Interestingly, pathogenic variants in *ARX* result in a number of developmental disorders in humans including lissencephaly (LISX2; MIM# 300215), Proud syndrome (MIM# 300004), infantile spasms without brain malformations (EIEE1; MIM# 308350), and syndromic (MIM# 309510) and nonsyndromic (MIM# 300419) mental retardation (Friocourt and Parnavelas 2010). Patients with *ARX*-related disorders that have CNS malformations can also present with urogenital anomalies, suggesting that the expression patterns of *EBF3* and *ARX* are both antagonistic and tightly regulated in a number of different tissues (Friocourt and Parnavelas 2010). Prior to recent reports, EBF3 had not been implicated definitively in any disorder in humans. However, patients with 10q26 microdeletion syndrome that includes a 3.5 Mb minimally deleted region (SROII) involving *EBF3* have several overlapping clinical features with individuals carrying *EBF3* intragenic variants, including short stature, craniofacial dysmorphisms, strabismus, abnormal ears, genital anomalies, urinary tract anomalies, CNS malformations, microcephaly, and

intellectual disability (Faria et al. 2016). As in ARX-related disorders, variants in *EBF3* are associated with marked phenotypic and clinical heterogeneity, with some patients presenting with structural brain abnormalities and others having no evidence of CNS malformations.

EBF3 is Expressed in Skeletal Muscle and is Involved in Muscle Development and Muscle-Specific Transcription

Outside of the CNS, *Ebf3* shows the highest expression levels in the diaphragm, bone marrow, and skeletal muscle in developing mouse embryos (Jin et al. 2014). Adult mice express *Ebf3* at the highest levels in skeletal muscle, the uterus, the eye, and the diaphragm (Jin et al. 2014). In Drosophila, *collier* is expressed in muscle progenitors and is required for myoblast fusion (Crozatier et al. 1999). Similarly in *Xenopus, Ebf2* and *Ebf3* are expressed in muscle and are required for somite organization, migration of hypaxial muscle anlagen, and the development of jaw muscle (Green and Vetter 2011b). *MyoD* and *Myf5* appear to be direct targets of these transcription factors (Green and Vetter 2011b). MyoD can also upregulate the expression of *Ebf* genes, suggestive of a positive feedback loop between Ebf and MyoD that is necessary for differentiation of muscle cells in *Xenopus* (Green and Vetter 2011b). *Ebf3* knockout mice die of respiratory failure before postnatal day 2 due to failure of the lung to unfold (Jin et al. 2014). This lethal phenotype is caused by a hypercontractile diaphragm with impaired Ca^{2+} efflux due to downregulation of *Serca1 (Atp2a1)* in the absence of *Ebf3* (Jin et al. 2014). Ebf3 has been shown to bind to the promoter

of *Atp2a1* and synergizes with MyoD to induce expression of other muscle-specific target genes (Jin et al. 2014). *Ebf3* is also expressed in the urogenital tract of developing mice, including regions that correspond to the detrusor muscle of the bladder and the muscle layer of the pelvic urethra (Figure S4) (McMahon et al. 2008; Harding et al. 2011). Given its role in muscle development and contraction, it is possible that disruption of EBF3-mediated transcription in muscle could lead to hypotonia, atonic bladder/loss of bladder control, and other phenotypes observed in patients with pathogenic *EBF3* variants. Further studies will be essential to unraveling the role of EBF3 in the expression of disease in humans.

METHODS

Sample Collection and Whole Exome Sequencing

Whole exome sequencing was performed on genomic DNA extracted from all samples submitted. The exome was captured utilizing a custom reagent developed by Mayo Clinic and Agilent Technologies, targeting 19,456 genes and 187,715 exons using 637,923 probes to capture a 54.1Mbp total region. Sequencing was performed on an Illumina HiSeq 2500 Next Generation sequencing instrument, using HapMap Sample NA12878 as an internal control. Paired-end 101 base-pair reads were aligned to a modified human reference genome (GRCh37/hg19) using Novoalign (Novocraft Technologies, Malaysia). Sequencing quality was evaluated using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). All germline variants were

jointly called through GATK Haplotype Caller and GenotypeGVCF (McKenna et al. 2010). Each variant was annotated using the BioR Toolkit (Kocher et al. 2014) and subsequently evaluated for clinical relevance.

Table 3. Sequencing Results

10x Coverage	Mean Coverage	Yield (Gb)	>Q30 (%)	Mean Q	Filtered Variants	EBF3 Mean Exon Coverage	Variant Coverage
98.63%	140x	15.3	99.92	38.405	75117	72x	379

Molecular Modeling and Molecular Dynamics Simulations

Our modeling and analysis began from the UniProt sequence of EBF3 (also known as COE3), Q9H4W6, which is 97% identical to EBF1 across its DNA binding domain (Goujon et al. 2010). The initial configuration of our structure was generated using homology modeling from the crystal structure of EBF1 dimer bound to DNA, 3MLP (Treiber et al. 2010), using Modeller (Martí-Renom et al. 2000) and an automated modeling approach (Zhi et al. 2014). All-atom implicit environment configurations were generated using VMD (Humphrey et al. 1996). Molecular dynamics (MD) simulations were carried out using NAMD (Phillips et al. 2005) and the CHARMM36 force field (Best et al. 2012; Hart et al. 2012). Triplicate simulations for each variant were independently energy minimized for 10,000 steps, followed by heating to 300K over 600ps via a Langevin thermostat and equilibration for 1ns, with a simulation time step of 1fs and conformations recorded every 2ps. A further 5ns of simulation trajectory was generated for analysis. Prior to analysis, all trajectories were aligned to the initial WT

conformation using DNA backbone atoms. Principal Component (PC) analysis was performed in Cartesian space on protein C^{α} atoms. Analysis was carried out using custom scripts, leveraging VMD and the Bio3D R package (Grant et al. 2006). Protein structure visualization was performed in PyMol version 1.7.6. (N1) and VMD v1.9.3.

ADDITIONAL INFORMATION

Data Deposition and Access: Whole-exome sequencing data is not publicly available because patient consent could not be obtained. The variant has been submitted to ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) under accession number SCV000493129.

Ethics Statement: The proband and/or parents were consented for sample collection and subsequent analysis under a protocol approved by the institutional review board of the Mayo Clinic. Written informed consent was obtained from the proband's parents for publication and accompanying images.

Acknowledgments: We would like to thank the patient and her family for participating in this study.

Author Contributions: P.R.B., S.S.B., M.T.Z., E.W.K., and P.N.P. designed the study. P.R.B., S.S.B., M.T.Z., M.A.C., C.K., F.P.V., Z.N., M.J.F., R.A.U., D.S.,

E.W.K. and P.N.P. gathered the data. P.R.B., M.T.Z., and E.W.K. analyzed the data. P.R.B., S.S.B., M.T.Z., E.W.K, and P.N.P. wrote the paper.

Funding: We would also like to thank the Mayo Clinic Center for Individualized Medicine (CIM) for supporting this research through the CIM Investigative and Functional Genomics Program.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248–249.

Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, Mackerell AD. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J Chem Theory Comput* **8**: 3257–3273.

Chao H-T, Davids M, Burke E, Pappas JG, Rosenfeld JA, McCarty AJ, Davis T, Wolfe L, Toro C, Tifft C, et al. 2016. A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *Am J Hum Genet*.

Chiara F, Badaloni A, Croci L, Yeh ML, Cariboni A, Hoerder-Suabedissen A, Consalez GG, Eickholt B, Shimogori T, Parnavelas JG, et al. 2012. Early B-cell factors 2 and 3 (EBF2/3) regulate early migration of Cajal-Retzius cells from the cortical hem. *Developmental Biology* **365**: 277–289.

Colasante G, Sessa A, Crispi S, Calogero R, Mansouri A, Collombat P, Broccoli V. 2009. Arx acts as a regional key selector gene in the ventral telencephalon mainly through its transcriptional repression activity. *Developmental Biology* **334**: 59–71.

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A. 1996. Collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. *Curr Biol* **6**: 707–718.

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A. 1999. Head versus trunk patterning in the Drosophila embryo; collier requirement for formation of the intercalary segment. *Development* **126**: 4385–4394.

Dubois L, Vincent A. 2001. The COE--Collier/Olf1/EBF--transcription factors: structural conservation and diversity of developmental functions. *Mechanisms of Development* **108**: 3–12.

Faria ÁC, Rabbi-Bortolini E, Rebouças MRGO, de S Thiago Pereira ALA, Frasson MGT, Atique R, Lourenço NCV, Rosenberg C, Kobayashi GS, Passos-Bueno MR, et al. 2016. Craniosynostosis in 10q26 deletion patients: A consequence of brain underdevelopment or altered suture biology? *Am J Med Genet A* **170A**: 403–409.

Fields S, Ternyak K, Gao H, Ostraat R, Akerlund J, Hagman J. 2008. The “zinc knuckle” motif of Early B cell Factor is required for transcriptional activation of B cell-specific genes. *Mol Immunol* **45**: 3786–3796.

Friocourt G, Parnavelas JG. 2010. Mutations in ARX Result in Several Defects Involving GABAergic Neurons. *Front Cell Neurosci* **4**: 4.

Garel S, Marín F, Mattéi MG, Vesque C, Vincent A, Charnay P. 1997. Family of Ebf/Olf-1-related genes potentially involved in neuronal differentiation and regional specification in the central nervous system. *Dev Dyn* **210**: 191–205.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**: W695–9.

Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**: 2695–2696.

Green YS, Vetter ML. 2011a. EBF factors drive expression of multiple classes of target genes governing neuronal development. *Neural Dev* **6**: 19.

Green YS, Vetter ML. 2011b. EBF proteins participate in transcriptional regulation of Xenopus muscle development. *Developmental Biology* **358**: 240–250.

Hagman J, Gutch MJ, Lin H, Grosschedl R. 1995. EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. *EMBO J* **14**: 2907–2916.

Harding SD, Armit C, Armstrong J, Brennan J, Cheng Y, Haggarty B, Houghton D, Lloyd-MacGilp S, Pi X, Roochun Y, et al. 2011. The GUDMAP database--an online resource for genitourinary research. *Development* **138**: 2845–2853.

Harms FL, Girisha KM, Hardigan AA, Kortüm F, Shukla A, Alawi M, Dalal A, Brady L, Tarnopolsky M, Bird LM, et al. 2016. Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism. *Am J Hum Genet*.

Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, Mackerell AD. 2012. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J Chem Theory Comput* **8**: 348–362.

Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. *J Mol Graph* **14**: 33–38.

Jin S, Kim J, Willert T, Klein-Rodewald T, Garcia-Dominguez M, Mosqueira M, Fink R, Esposito I, Hofbauer LC, Charnay P, et al. 2014. Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. *Nat Commun* **5**: 3793.

Kocher J-PA, Quest DJ, Duffy P, Meiners MA, Moore RM, Rider D, Hossain A, Hart SN, Dinu V. 2014. The Biological Reference Repository (BioR): a rapid and flexible system for genomics annotation. *Bioinformatics* **30**: 1920–1922.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**: 1073–1081.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285–291.

Liberg D, Sigvardsson M, Akerblad P. 2002. The EBF/Olf/Collier Family of Transcription Factors: Regulators of Differentiation in Cells Originating from All Three Embryonal Germ Layers. *Mol Cell Biol* **22**: 8389–8397.

Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* **29**: 291–325.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303. http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110.

McMahon AP, Aronow BJ, Davidson DR, Davies JA, Gaido KW, Grimmond S, Lessard JL, Little MH, Potter SS, Wilder EL, et al. 2008. GUDMAP: the genitourinary developmental molecular anatomy project. *J Am Soc Nephrol* **19**: 667–671.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. 2005. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**: 1781–1802.

Pozzoli O, Bosetti A, Croci L, Consalez GG, Vetter ML. 2001. Xebf3 is a regulator of neuronal differentiation during primary neurogenesis in Xenopus. *Developmental Biology* **233**: 495–512.

Prasad BC, Ye B, Zackhary R, Schrader K, Seydoux G, Reed RR. 1998. unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors. *Development* **125**: 1561–1568.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **11**: 361–362. http://dx.doi.org/10.1038/nmeth.2890.

Sleven H, Welsh SJ, Yu J, Churchill MEA, Wright CF, Henderson A, Horvath R, Rankin J, Vogt J, Magee A, et al. 2016. De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. *Am J Hum Genet*.

Treiber N, Treiber T, Zocher G, Grosschedl R. 2010. Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. *Genes & Development* **24**: 2270–2275.

Wang SS, Lewcock JW, Feinstein P, Mombaerts P, Reed RR. 2004. Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. *Development* **131**: 1377–1388.

Zhi H, Ning S, Li X, Li Y, Wu W, Li X. 2014. A novel reannotation strategy for dissecting DNA methylation patterns of human long intergenic non-coding RNAs in cancers. *Nucleic Acids Res* **42**: 8258–8270.

Additional Resources:

N1. The PyMOL Molecular Graphics System, Version 1.7.6 Schrödinger, LLC.

FIGURE LEGENDS

Figure 1: Patient photographs showing front and side views. The proband was noted to have mild dysmorphic features including bilateral esotropia, retrognathia, down-turned corners of the mouth, and a low posterior hairline.

Figure 2: Schematic diagram of EBF3 (NP_001005463.1) protein structure.

Numbering corresponds to amino acids.

Figure 3: R163W induces dynamic changes throughout EBF3 that are comparable to the validated pathogenic R163A and distinct from control simulations. A) We calculated the average RMSD across replicates for each simulation condition (indicated by color). Proteins with alterations at residue 163 demonstrated increased mobility throughout both monomers (indicated by light and dark gray rectangles along the abscissa) with the greatest differences among the DNA-interacting regions (black rectangles) around R163. **B)** In order to quantify the differences between conformations, we measured the distance from residue 163 (violet sphere) to nearby phosphate atoms in the DNA backbone (orange spheres). **C)** We show a comparison using one of these reference distances (larger sphere in panel B), demonstrating that both the WT and K239A retain stable DNA interactions, while both R163A and R163W lose contact with DNA. Additional distance measures are presented in Figure S2.

Table 1: Variant Information.

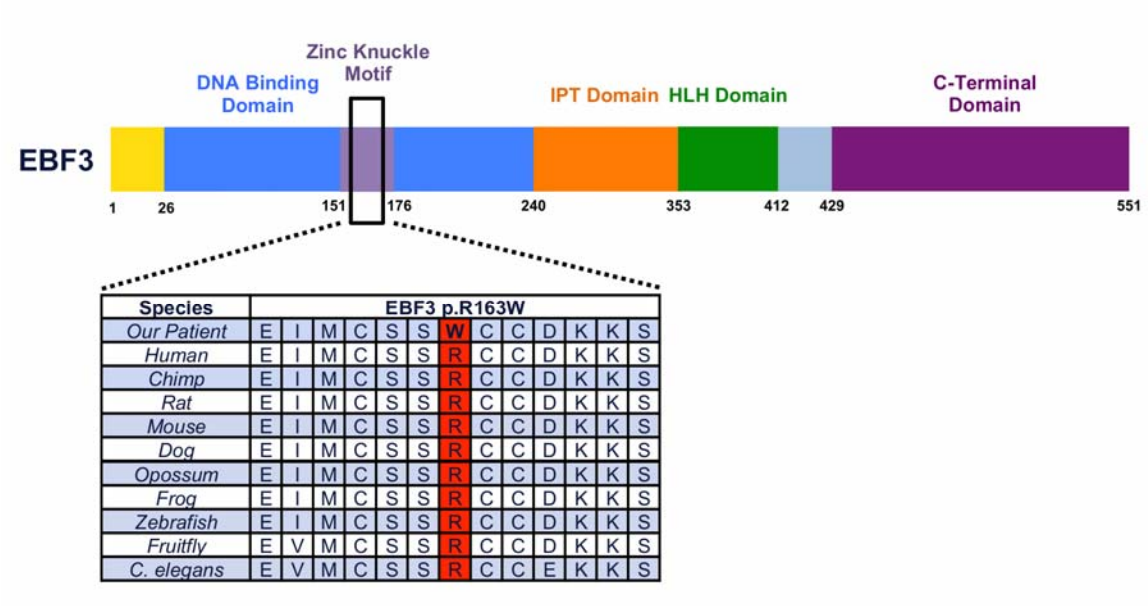
Table 2: Clinical summary of the patient described in this report and previously reported patients with variants affecting the Arg163 residue.

N/R, not reported; +, present.

Table 3: Sequencing Results.

FIGURES

Figure 1: Patient photographs showing front and side views. The proband was noted to have mild dysmorphic features including bilateral esotropia, retrognathia, down-turned corners of the mouth, and a low posterior hairline.



Species	EBF3 p.R163W												
Our Patient	E	I	M	C	S	S	**W**	C	C	D	K	K	S
Human	E	I	M	C	S	S	R	C	C	D	K	K	S
Chimp	E	I	M	C	S	S	R	C	C	D	K	K	S
Rat	E	I	M	C	S	S	R	C	C	D	K	K	S
Mouse	E	I	M	C	S	S	R	C	C	D	K	K	S
Dog	E	I	M	C	S	S	R	C	C	D	K	K	S
Opossum	E	I	M	C	S	S	R	C	C	D	K	K	S
Frog	E	I	M	C	S	S	R	C	C	D	K	K	S
Zebrafish	E	I	M	C	S	S	R	C	C	D	K	K	S
Fruitfly	E	V	M	C	S	S	R	C	C	D	K	K	S
C. elegans	E	V	M	C	S	S	R	C	C	E	K	K	S

Figure 2: Schematic diagram of EBF3 (NP_001005463.1) protein structure.

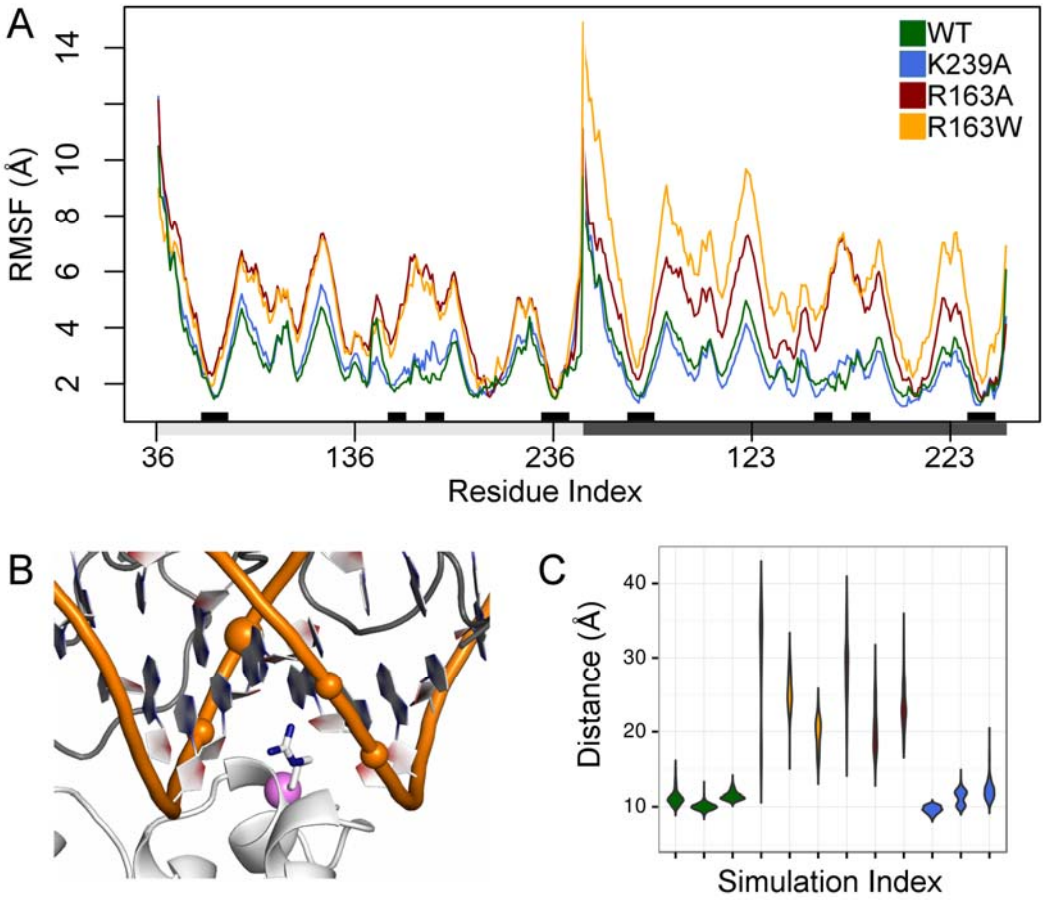


Figure 3: R163W induces dynamic changes throughout EBF3 that are comparable to the validated pathogenic R163A and distinct from control simulations. A) We calculated the average RMSD across replicates for each simulation condition (indicated by color). Proteins with alterations at residue 163 demonstrated increased mobility throughout both monomers (indicated by light and dark gray rectangles along the abscissa) with the greatest differences among the DNA-interacting regions (black rectangles) around R163. **B)** In order to quantify the differences between conformations, we measured the distance from residue 163 (violet sphere) to nearby phosphate atoms in the DNA backbone (orange spheres). **C)** We show a comparison using one of these

reference distances (larger sphere in panel B), demonstrating that both the WT and K239A retain stable DNA interactions, while both R163A and R163W lose contact with DNA. Additional distance measures are presented in Figure S2.

Supplemental Information:

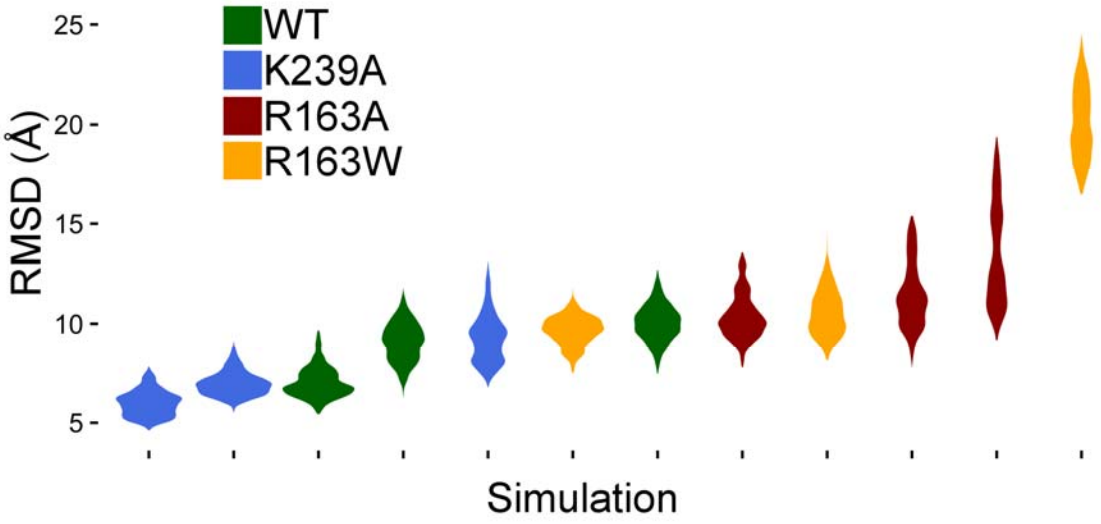


Figure S1: Greater RMSDs are observed for R163 variants, compared to WT and the control variant. For each of triplicate simulations of each variant, we plot the probability distribution (wider regions are more probable) of C^{α} RMSDs after aligning DNA atoms from each frame to the initial WT conformation.

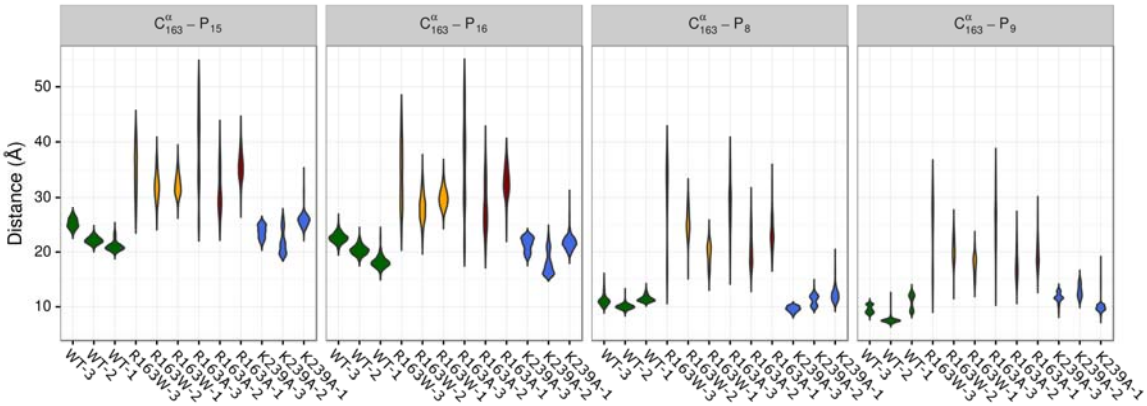


Figure S2: Distance from residue 163 to each of four DNA bases is greater for R163 variants than either control. We monitored the distance between the C^{α} of residue 163 in each EBF3 monomer to the backbone phosphate atom of four nearby bases (one per panel). Since there are two monomers in each simulation, we took the maximum of the two distances within each simulation frame (timepoint). Data are summarized by smoothed probability density plots where the probability of observing a protein-DNA distance in a simulation is proportional to width. Each variant was simulated in triplicate and probability densities are shown for each. Colors are as in previous figures and correspond to each variant – the order of variants is the same in all panels and indicated by abscissa labels. Our patient's R163W closely resembles the biochemically validated loss-of-binding variant (R163A), while the WT and biochemically verified normal-function variant (K239A) resemble one another.

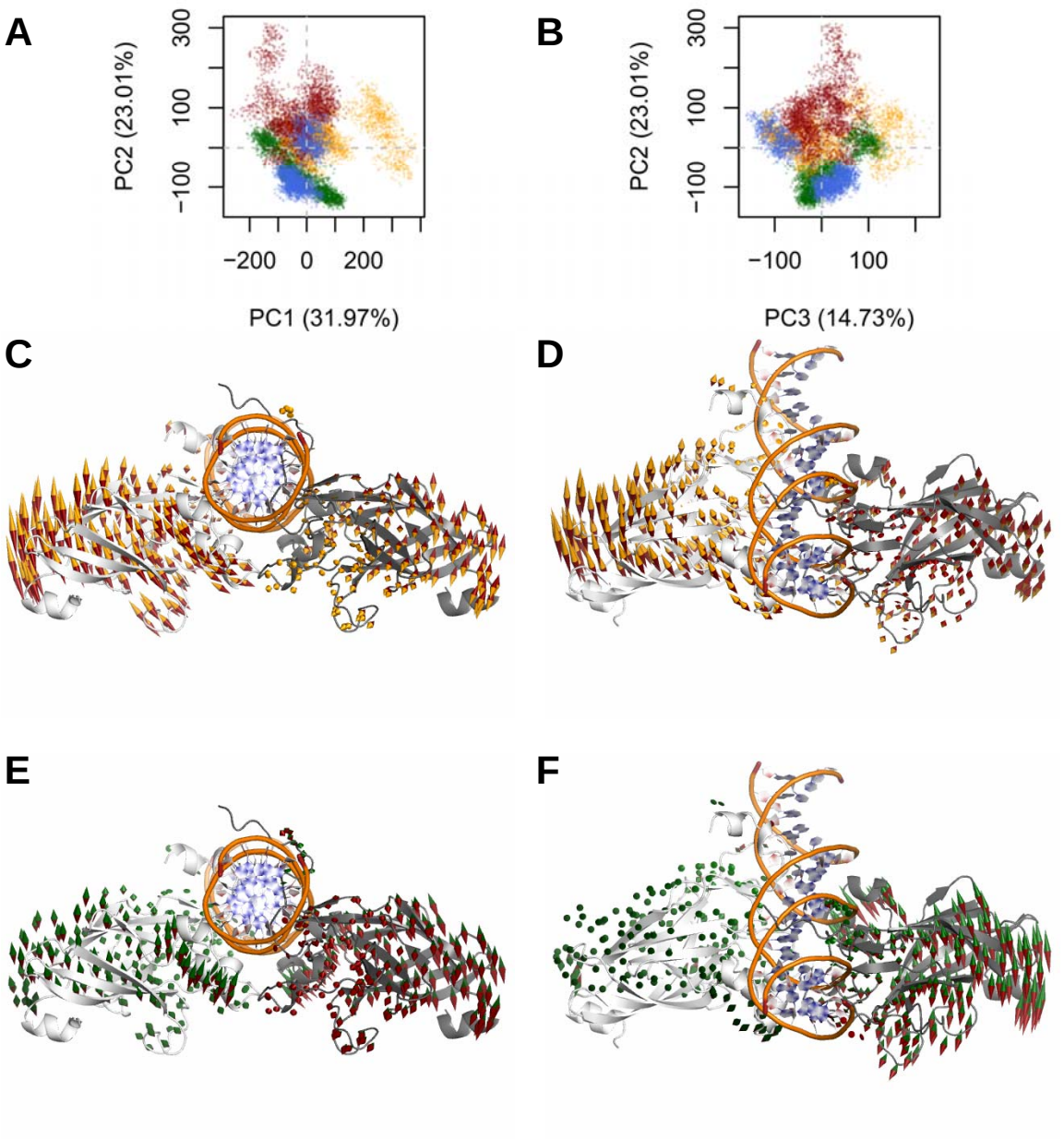
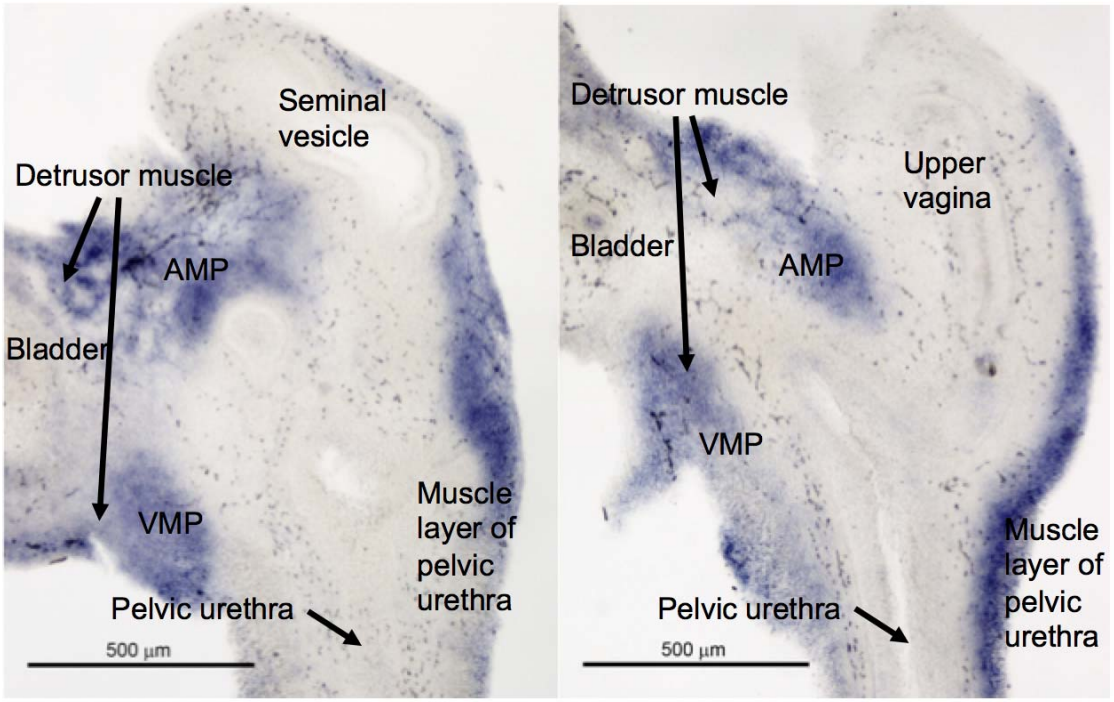


Figure S3: Visualization of the dominant motions using PCA. MD simulations capture both the short-scale random atomic fluctuations and the large-scale motions of the protein. The latter are typically of biologic interest and are typically identified using PC analysis. We applied PC analysis to our simulation data to identify the large-scale motions therein. **A)** Snapshots from each simulation are plotted as points in the first two PCs, revealing that both Arg163 variants lead to

differential sampling compared to the WT and p.Lys239Ala. Color indicates variant, as in previous figures. **B)** The analogous plot in the 2nd and 3rd PCs reveals little separation between the variants along the 3rd PC. Indicating that the first two PCs are most critical. **C,D)** Visualization of the motion indicated by the first PC shows that it is a rotation of the EBF3 dimer around the DNA helix axis. Color indicates sign: +PC1 in orange, -PC1 in red. The two EBF3 monomers are colored white and gray. **E,F)** Visualization of the motion indicated by the second PC shows that it describes a similar motion, but with the relative magnitude of each monomer's motion switched. Color indicates sign: +PC2 in red, -PC2 in green.

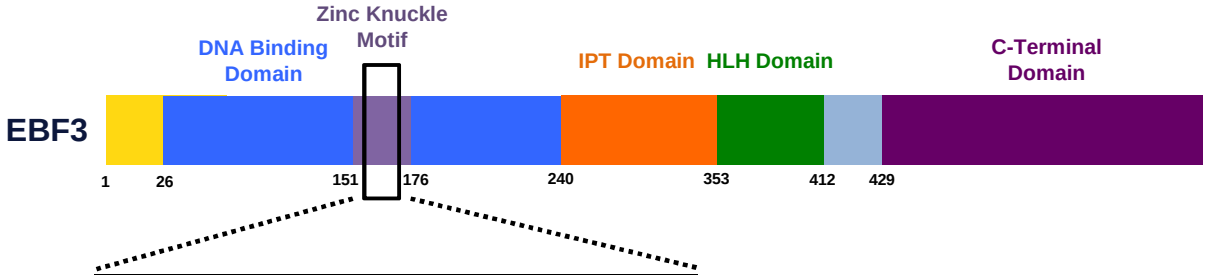


Specimen GUDMAP:14141:
Embryonic day 17.0
Male

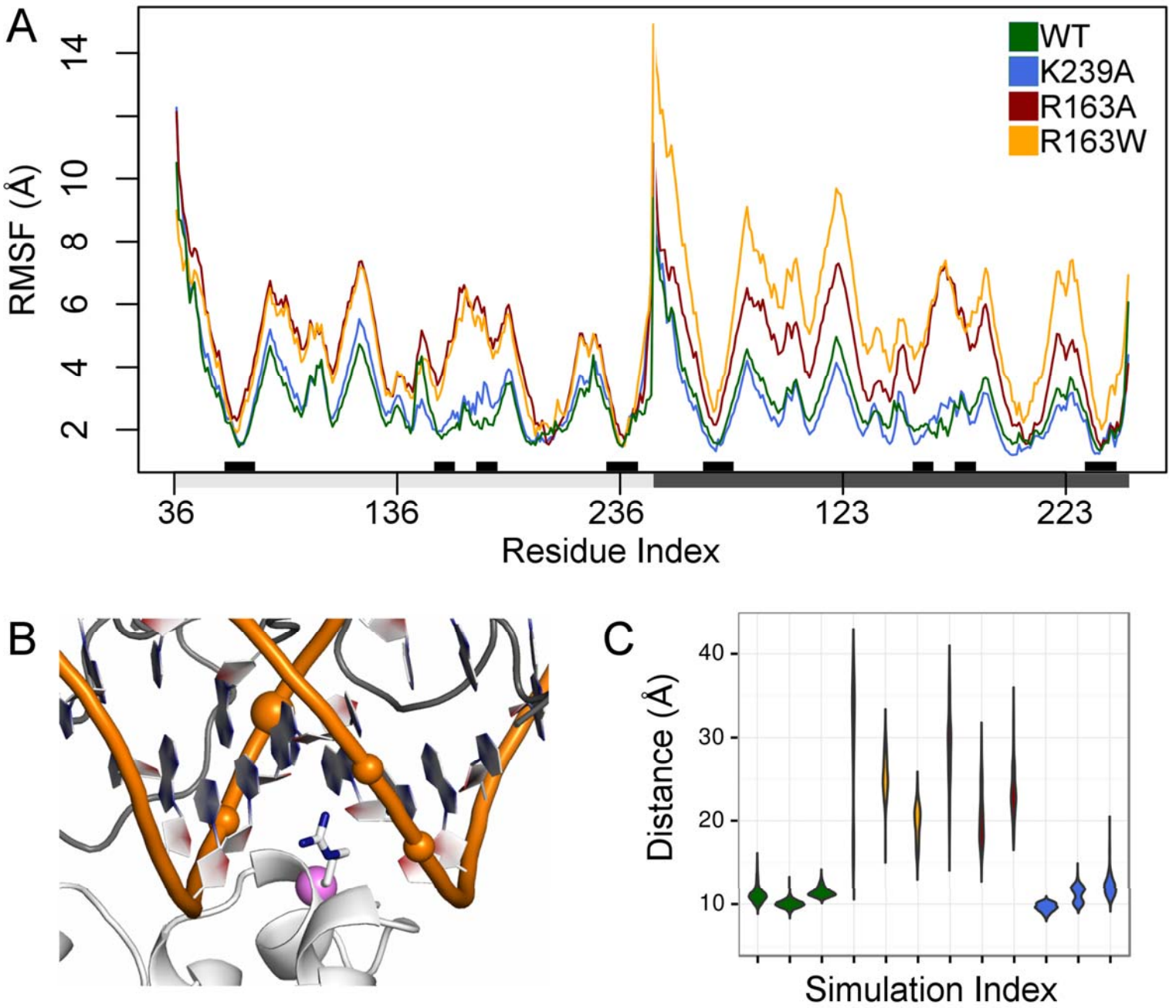
Specimen GUDMAP:14142:
Embryonic day 17.0
Female

Figure S4: *In situ* hybridization showing *Ebf3* expression in the male and female reproductive systems of Theiler stage 25 mice. Expression of *Ebf3* can be seen in the bladder, detrusor muscle, and muscle layer of the pelvic urethra. VMP; ventral mesenchymal pad; AMP; anterior mesenchymal pad**.** This figure was constructed using data from the GUDMAP database http://www.gudmap.org, accessed in January 2017, including *in situ* data (Chad Vezina, University of Wisconsin-Madison, Madison, WI, USA, cmvezina@wisc.edu).

Supplemental animation 1: Two representative simulations from WT (blue) and R163W (orange) are shown in cartoon representation with residue 163 in sticks representation; bound DNA is colored dark blue and gold, respectively. The viewing angle is chosen to highlight the zinc knuckle interaction with DNA for the closer monomer. The trajectories have been superimposed using DNA backbone atoms. The trajectories begin from the energy minimized conformations and show the time-dependent dynamics of the system as it is heated (atom motion accelerates through the animation) and equilibrated. While residue 163 intercalates into the minor grove of DNA in both energy minimized structures, R163W dislocates from this position about halfway through the animation. While R163 has dynamic interactions with DNA throughout the simulation, it remains within the minor grove.



Species	EBF3 p.R163W												
Our Patient	E	I	M	C	S	S	**W**	C	C	D	K	K	S
Human	E	I	M	C	S	S	R	C	C	D	K	K	S
Chimp	E	I	M	C	S	S	R	C	C	D	K	K	S
Rat	E	I	M	C	S	S	R	C	C	D	K	K	S
Mouse	E	I	M	C	S	S	R	C	C	D	K	K	S
Dog	E	I	M	C	S	S	R	C	C	D	K	K	S
Opossum	E	I	M	C	S	S	R	C	C	D	K	K	S
Frog	E	I	M	C	S	S	R	C	C	D	K	K	S
Zebrafish	E	I	M	C	S	S	R	C	C	D	K	K	S
Fruitfly	E	V	M	C	S	S	R	C	C	D	K	K	S
C. elegans	E	V	M	C	S	S	R	C	C	E	K	K	S



Position (hg19/GRCh37)	Type	Gene	HGVS cDNA	HGVS protein	Zygosity	Inheritance	SIFT/ Polyphen-2/ MutationTaster2	ExAC/ gnomAD Allele Frequency
Chr10:131755589G>A	Missense	EBF3	NM_001005463.2 c.487C>T	NP_001005463.1 p.(Arg163Trp)	Het	*De novo*	Deleterious/ Probably Damaging/ Disease Causing	N/R

Variant in EBF3	Chr10: 131755589G>A (hg19); NM_001005463.2; c.487C>T; p.(Arg163Trp)	Chr10: 131755588C>T (hg19); NM_001005463.2; c.488G>A; p.(Arg163Gln)	Chr10: 131755588C>T (hg19); NM_001005463.2; c.488G>A; p.(Arg163Gln)	Chr10: 131755588C>A (hg19); NM_001005463.2; c.488G>T; p.(Arg163Leu)	Chr10: 131755588C>G (hg19); NM_001005463.2; c.488G>C; p.(Arg163Pro)
Reference	This report	Chao et al. 2016	Chao et al. 2016	Chao et al. 2016	Sleven et al. 2016
Inheritance	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*
Sex	Female	Male	Female	Female	Male
Ethnicity	Irish, Mexican, German, and French descent	Pacific Islander of Chinese and Japanese descent	African American	English, Irish, German, and Polish descent	English descent
Age at most recent assessment	23 months	7 years	5 years	3 years	13 years
Birth	Birth was at term by vaginal delivery and was significant for meconium aspiration and distended bladder requiring catheterization	Birth was at 38 weeks gestation by repeat Caesarean section and was significant for a loose nuchal cord wrapped around the neck and a fractured clavicle	Patient's prenatal history is significant for decreased fetal movements, and birth was at 40 weeks via induced vaginal delivery for oligohydramnios	Birth was at 39 weeks of gestation by Caesarean section due to breech position	Birth was at 38 weeks by elective Caesarian section due to breech position
Birth weight in grams (SD)	3490 grams	3400 grams	3350 grams	2700 grams	3200 grams
Birth length in cm (SD)	49.5 cm	52 cm	50.8 cm	N/R	N/R
OFC at birth in cm (SD)	N/R	36.2 cm	33.5 cm	N/R	N/R
Weight in kg (SD)	10.6 kg (30th percentile) at 23 months	19.5 kg (25th percentile)	17.5 kg (41st percentile)	10th percentile	N/R
Height in cm (SD)	71 cm (< 2nd percentile) at 23 months	109.7 cm (25th percentile)	114.8 cm (91st percentile)	10th percentile	132.5cm (< 1st percentile) at 13 years
OFC in cm (SD)	46.7 cm (30th percentile) at 23 months	52.5 cm (50th-75th percentile)	51 cm (85th percentile)	20th percentile	56.2 cm (50th–75th percentile) at 13 years
Intellectual disability	+	+	+	+	+
Global developmental delay	+	+	+	+	+
Speech delay	+	+	+	+	+
Hypotonia	+	+	+	+	+
Ataxia	Not present at time of evaluation	Gait ataxia	Gait ataxia, dysmetria	Gait ataxia	Gait and truncal ataxia
Seizures	N/R	N/R	N/R	N/R	N/R
Brain MRI findings	Brain MRI revealed no abnormalities	Brain MRI (performed at 7 years of age) revealed small inferior posterior cerebellar lobes, hypoplasia of the posterior vermis, and mild prominence of the ventricles and sulci	Brain MRI (performed at 18 months of age) revealed hypoplasia of the anterior and posterior vermis, overfolding of the superior helices, with normal cerebellar hemispheres	Brain MRI (performed at 1 year of age) showed normal cerebellar vermis and hemispheres	Brain imaging studies (performed at age 1 and 5) revealed cerebellar cleft or absent vermis, cerebellar atrophy, and atrophy of pontine tegmentum
Facial dysmorphisms	Low posterior hairline, down turned mouth corners, and mild retrognathia	Myopathic facies and short anteverted nostrils	Triangular-shaped facies	Triangular-shaped facies	Dolichocephaly, prominent forehead and occiput, and deep-set eyes
Strabismus	Bilateral esotropia	+	+	+	Esotropia
Ears	Ears neutrally set and well formed	Overfolding of the superior helices	Overfolding of the superior helices	N/R	N/R
Urogenital anomalies	Bilateral hydroureter and hydronephrosis after birth due to extreme urinary retention/atonic bladder, vesicoureteral reflux (vesicostomy was performed), distal urethral stricture, recurrent urinary tract infections, bicornate uterus	Micropenis and cryptorchidism	Mild reduction in volume of the labia majora, lacks bladder control	Urinary retention associated with incomplete bladder emptying and grade 1 urinary reflux	Left cryptorchidism
Additional clinical findings	Patient has generalized low muscle tone with a normal EMG study with no evidence for myopathy, defect in neuromuscular transmission, or peripheral neuropathy, deep tendon reflexes were decreased (-2/-3 in upper and lower), normal spinal MRI, mild pectus excavatum, uses 2-3 words purposefully	Clinical features include facial weakness, expressive speech disorder, dysarthria, dysphagia, gastroesophageal reflux disease, and hockey-stick palmar creases	Patient's clinical features include facial weakness, abnormal palmar creases, fifth-finger clinodactyly, expressive speech disorder, apraxia, dysarthria, dysphagia, and perseverative social behaviors. Patient also has marked insensitivity to pain	Patient's clinical features include facial weakness, expressive speech disorder, dysphagia, motor stereotypies, small feet, and torticollis. Patient speaks only one word, did not walk until late, and has a pincer grasp. Patient also has marked insensitivity to pain	Patient's clinical features include pectus excavatum, tapering fingers, pes planus, shortened great toes, dysarthria, and high-pitched voice. The patient stood with support at 2 years of age, walked independently at 5 years and 8 months, has limited speech (50 words), and an IQ of 71.
Functional evidence	Not done	Activation of COE-binding sequence reporter-gene was assessed in vitro as the ratio of NanoLuc to firefly luciferase. A 92-fold induction was observed with WT EBF3. EBF3 p.Arg163Gln showed a very poor induction of transcription and was indistinguishable from the negative control.	Activation of COE-binding sequence reporter-gene was assessed in vitro as the ratio of NanoLuc to firefly luciferase. A 92-fold induction was observed with WT EBF3. EBF3 p.Arg163Gln showed a very poor induction of transcription and was indistinguishable from the negative control.	Activation of COE-binding sequence reporter-gene was assessed in vitro as the ratio of NanoLuc to firefly luciferase. A 92-fold induction was observed with WT EBF3. EBF3 p.Arg163Leu had only a 45-fold induction, suggesting that the variant is a hypomorphic alteration.	Used μM2.21 cell system to measure the relative transactivation abilities of mutant and wild-type EBF3 proteins across graded levels of expression (assay measures the percentage of surface mIgM expression as a direct readout of EBF3 function). The p.Arg163Pro variant was inactive across all expression levels, suggesting that mutation of this residue ablates EBF3 function.

10x Coverage	Mean Coverage	Yield (Gb)	>Q30 (%)	Mean Q	Filtered Variants	EBF3 Mean Exon Coverage	Variant Coverage
98.63%	140x	15.3	99.92	38.405	75117	72x	379

Novel de novo variant in EBF3 is likely to impact DNA binding in a patient with a neurodevelopmental disorder and expanded phenotypes: patient report, in silico functional assessment, and review of published cases

Patrick R. Blackburn, Sarah S. Barnett, Michael T. Zimmermann, et al.

Cold Spring Harb Mol Case Stud published online March 10, 2017
Access the most recent version at doi:10.1101/mcs.a001743