Human Molecular Genetics, 2015, Vol. 24, No. 17 4848–4861

doi: 10.1093/hmg/ddv208
Advance Access Publication Date: 8 June 2015
Original Article

ORIGINAL ARTICLE

ZC4H2, an XLID gene, is required for the generation of a specific subset of CNS interneurons

Melanie May[1,†], Kyu-Seok Hwang[2,†], Judith Miles[3], Charlie Williams[4], Tejasvi Niranjan[5], Stephen G. Kahler[6], Pietro Chiurazzi[7], Katharina Steindl[8], Peter J. Van Der Spek[9], Sigrid Swagemakers[9], Jennifer Mueller[4], Shannon Stefl[10], Emil Alexov[10], Jeong-Im Ryu[2], Jung-Hwa Choi[2], Hyun-Taek Kim[2], Patrick Tarpey[11], Giovanni Neri[7], Lynda Holloway[1], Cindy Skinner[1], Roger E. Stevenson[1], Richard I. Dorsky[12], Tao Wang[5], Charles E. Schwartz[1,‡,*] and Cheol-Hee Kim[2,‡,*]

[1]Greenwood Genetic Center, Greenwood, SC 29646, USA, [2]Department of Biology, Chungnam National University, Daejeon 305-764, Korea, [3]Department of Child Health, University of Missouri School of Medicine, Columbia, MO 65212, USA, [4]Division of Genetics and Metabolism, University of Florida College of Medicine, Gainesville, FL 33612, USA, [5]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21287, USA, [6]Arkansas Children's Hospital, Little Rock, AR 72202, USA, [7]Institute of Medical Genetics, Catholic University, Rome 00-168, Italy, [8]Institute of Medical Genetics, University of Zurich, Schwerzenbach 8006, Switzerland, [9]Department of Bioinformatics, Erasmus University Medical Center, Rotterdam 3015, Netherlands, [10]Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA, [11]Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK and [12]Department of Neurobiology and Anatomy, University of Utah, Salt Lake City, UT 84112, USA

*To whom correspondence should be addressed. Tel: +1 864 9418140; Fax: +1 864 3881703; Email: ceschwartz@ggc.org (C.E.S.); Email: zebrakim@cnu.ac.kr (C.H.K.)

Abstract

Miles–Carpenter syndrome (MCS) was described in 1991 as an XLID syndrome with fingertip arches and contractures and mapped to proximal Xq. Patients had microcephaly, short stature, mild spasticity, thoracic scoliosis, hyperextendable MCP joints, rocker-bottom feet, hyperextended elbows and knees. A mutation, p.L66H, in *ZC4H2*, was identified in a XLID re-sequencing project. Additional screening of linked families and next generation sequencing of XLID families identified three *ZC4H2* mutations: p.R18K, p.R213W and p.V75in15aa. The families shared some relevant clinical features. *In silico* modeling of the mutant proteins indicated all alterations would destabilize the protein. Knockout mutations in *zc4h2* were created in zebrafish and homozygous mutant larvae exhibited abnormal swimming, increased twitching, defective eye movement and

[†]M.M. and K.S.H. contributed equally to this work.
[‡]C.E.S. and C.H.K. made equal senior author contributions to the paper.
Received: February 19, 2015. **Revised:** May 28, 2015. **Accepted:** June 1, 2015

pectoral fin contractures. Because several of the behavioral defects were consistent with hyperactivity, we examined the underlying neuronal defects and found that sensory neurons and motoneurons appeared normal. However, we observed a striking reduction in GABAergic interneurons. Analysis of cell-type-specific markers showed a specific loss of V2 interneurons in the brain and spinal cord, likely arising from mis-specification of neural progenitors. Injected human wt *ZC4H2* rescued the mutant phenotype. Mutant zebrafish injected with human p.L66H or p.R213W mRNA failed to be rescued, while the p.R18K mRNA was able to rescue the interneuron defect. Our findings clearly support *ZC4H2* as a novel XLID gene with a required function in interneuron development. Loss of function of *ZC4H2* thus likely results in altered connectivity of many brain and spinal circuits.

Introduction

Miles–Carpenter syndrome (MCS) is an X-linked intellectual disability (XLID) syndrome first described in 1991 (1). The syndrome was characterized by males having short stature, microcephaly, exotropia, long hands, digit contractions, rocker-bottom feet and spasticity along with severe intellectual disability. A distinguishing feature was the presence of an excess of arch fingerprints or a low total ridge count. Additionally, it was one of the few XLID syndromes in which the carrier females have some somatic features, as well as mild cognitive impairment.

X-linked intellectual disability conditions account for about two out of 1000 males with intellectual disability (ID) and are quite heterogeneous. An estimated 150–200 genes are presumed to be responsible for XLID of which only about 100 have been identified. Indeed 30 XLID syndromes and 50 mapped families with non-syndromal XLID are without a known causative gene mutation (http://www.ggc.org/research/molecular-studies/xlid.html). In order to address this gap, a large collaborative resequencing project of 718 X genes in 208 XLID probands, (IGOLD, International Genetics of Learning Disability), was undertaken using Sanger sequencing to identify novel XLID genes (2). Besides identifying nine XLID genes, it generated over 500 non-recurrent missense mutations. As a partial follow-up of this project, 15 probands were sequenced for the same 718 genes using next generation sequencing. Novel, non-recurrent missense mutations were then pursued further. One such missense mutation, p.L66H, was identified in *ZC4H2*, a zinc finger gene for which very little information was available. Therefore, additional testing of families was undertaken. Screening of families linked to Xq11 and next generation X-exome sequencing of small XLID families identified two additional *ZC4H2* mutations, a missense mutation, p.R213W, and an in-frame insertion of 15 amino acids, p.V75in15aa. Independently, a third *ZC4H2* missense mutation, p.R18K, was identified in a family with XLID using whole exome sequencing (WES). Biophysical *in silico* modeling indicated all three missense mutations were destabilizing. Functional studies, using two KO zebrafish mutants created using the TALEN methodology, allowed us to determine that ZC4H2 is important for the specification and generation of a subset of central nervous system (CNS) interneurons. In addition, rescue experiments indicated that the L66H and R213W mutations result in loss of protein function. Further studies of *ZC4H2* will provide a better appreciation for the important role this gene plays in brain development and function.

Results

Mutations in *ZC4H2*

Family K8070

Next generation resequencing of 718 genes located on the X-chromosome identified a mutation, c.197T > A; p.L66H, in a male from K8070 (Fig. 1). This family was previously published as having MCS, an XLID condition with severe ID, macrocephaly, contractures and exotropia (Table 1) (1). The syndrome was mapped to a broad region of the X, spanning from Xp11.4 to Xq22.2. The location of *ZC4H2* in Xq11.2 was consistent with this localization and segregation of the c.197T > A mutation was found in K8070 (Fig. 1A). The c.197T > A mutation was not listed in dbSNP nor in the 1000 Genome database. Additionally, a separate analysis of 1302 normal X chromosomes failed to detect the change, further substantiating c.197T > A was not a rare polymorphism. Bioinformatic analysis revealed the mutation was likely pathogenic (Table 2). The L66 residue was also observed to be highly conserved down to *Candida tropicalis*.

Family K8615

Once the L66H mutation in *ZC4H2* was found in K8070, 30 families within the Greenwood Genetic Center (GGC) cohort of XLID families with linkage to Xq11 were screened for additional *ZC4H2* mutations. The proband in family K8615 was found to have a p.R213W (c.637C > T) missense mutation. The mutation segregated in the family (Fig. 1B). The family has three affected males in three different sibships. The presentation was different from that of K8070 (Table 1). Contractures were only present at the knees and only one male had a club foot. Spasticity was noted in one male, but both males had motor developmental delay and ID. The carrier females did not exhibit any somatic features, but one had mild ID. The R213 residue is highly conserved down to *Drosophila* and bioinformatic analysis indicated it was very likely to be pathogenic (Table 2).

Family K9333

As part of a focused, NINDS funded, next generation sequencing of the X chromosome-exome of families with XLID, another *ZC4H2* mutation, c.225 + 5G > A, was found in a family with three affected males in three generations (Fig. 1C). No localization data were available due to the structure of the family and the lack of a sample on one affected male (II-5, Fig. 1C). The alteration segregated in the family and was found to cause a splicing defect resulting in an in-frame insertion of 15 novel amino acids, p.V75in15aa (Fig. 2). Additionally, the wt transcript was not observed by RT–PCR. The affected males presented with a phenotype different from MCS and K8615 (Table 1). They had hypotonia, which progressed to spasticity, 3+ reflexes, microcephaly and severe developmental delay. No contractures were noted. Nor did the affected males have seizures. The carrier females were normal.

Family K9611

After whole-genome sequencing of the proband (II-4, Fig. 1D) and his mother (I-2), X-linked variants were identified comparing the genome sequences of patient II-4 and his mother I-2. Fifty-two variants present in the heterozygous mother were transmitted to II-4, 16 of which are not present in the dbSNP or in the 1000 genomes database. Eight of these 16 variants map to the linkage region and cosegregation in the family was determined by Sanger resequencing of individual amplicons. Four of the eight variants caused a missense mutation, but only one of these, R18K in *ZC4H2*, had a high Polyphen-2 score (0.97) suggesting a deleterious effect on protein function (Table 2).

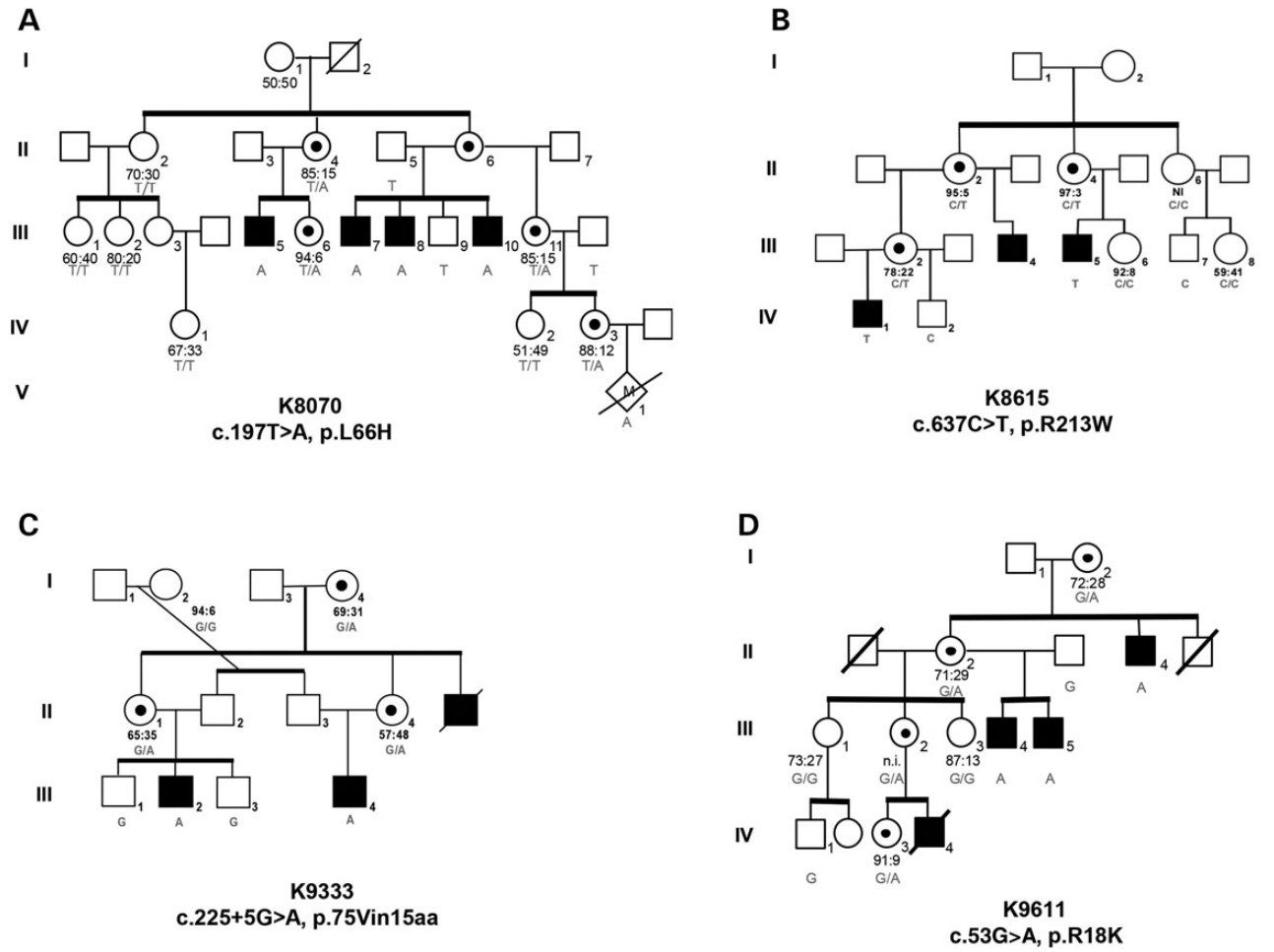

Figure 1. Pedigrees of the four families with mutations in *ZC4H2*. The genotype for each family mutation is given for individuals studied. X-inactivation data for the women is given below each female. (■) affected male; (⊙) carrier female as determined either by pedigree structure or *ZC4H2* analysis.

Tissue expression of *ZC4H2*

ZC4H2 expression analysis of human tissue was conducted using RT–PCR and a multiple human fetal tissue cDNA panel (Clontech). The expression of two different *ZC4H2* transcripts, ENST00000374839 (224 amino acids) and ENST00000337990 (201 amino acids), was examined because, although, the R18H mutation does exist in the primary transcript of 224 amino acids, it was not present in the shorter form of 201 amino acids (Supplementary Material, Fig. S1A). Thus, the expression level across multiple tissues was explored for both transcripts. As seen in Supplementary Material, Figure S1B, both transcripts are expressed in the brain with the shorter one perhaps at a higher level. The long form is apparently not present in the liver. Since both transcripts were present in fetal brain, we examined the level of expression in different regions of human brain using a Rapid Scan Human Brain Panel (Origene). The findings were quite revealing (Supplementary material, Fig. S1C). The short form, missing exon 1 and the R18 residue, was highly expressed in all regions of the brain and the spinal cord. The long form was weakly expressed in all regions with perhaps higher expression in the hypothalamus, pons and medulla. It was not observed in the spinal cord sample. These observations might explain the mildness of the phenotype observed in the family with the R18H mutation relative to the other families, such as the lack of hypotonia, seizures and hyperreflexia in the affected males (Table 1).

In silico modeling of *ZC4H2* mutations

Although the bioinformatic analyses indicated all four *ZC4H2* mutations were likely pathogenic (Table 2), we undertook *in silico* analyses to predict the effect the three missense mutations had on the stability of the ZC4H2 protein. The experimental 3D structures of wild-type and mutant ZC4H2 proteins are not available and were modeled *in silico* as outlined in the Materials and Methods section.

R18K mutation

In comparing the predicted structures for the WT and mutant R18K, we were able to predict that the mutation will likely be destabilizing. Using the criteria described in the Materials and Methods section, we were able to reduce the five models in three force fields for a total of 15 structure options down to three viable options: model2 for Amber and Charmm and model3 for Amber force fields. A visual inspection of these models revealed that R18 forms two hydrogen bonds with the glutamic acid at position 106 (E106) in the model2 Amber WT structure (Supplementary Material, Fig. S2A), and three hydrogen bonds with E106 in model2 Charmm (Supplementary Material, Fig. S2B). No hydrogen bonds were formed between these residues in model3 Amber (Supplementary Material, Fig. S2C). Instead R18 is completely exposed to the water phase, which would allow it to form favorable hydrogen bonds with water.

Table 1. Clinical findings in four families with *ZC4H2* mutations

Kindred *ZC4H2* mutation	K8070[a] p.L66H		K8615[a,b] p.R213W		K9333 p.V75in15aa		K9611[a,b,c] p.R18K		Totals	
Affected individuals	4M	6F	3M	4F	3M	3F	4M	4F	14M	17F
Growth										
Short stature (<3)	4/4	0/6	0/3	0/4	3/3	—	3/3	1/4	10/13	1/14
Microcephaly (<3)	3/4	1/6	0/3	0/4	3/3	—	3/3	0/4	9/13	1/14
Adaptation										
Respiratory distress	0/4	0/6	0/3	0/4	0/2	—	0/3	—	0/12	0/10
Poor feeding	2/4	0/6	0/3	0/4	1/2	—	0/3	1/1	3/12	1/11
Motor developmental delay	4/4	2/6	3/3	0/2	3/3	—	3/3	1/1	13/13	3/9
Facies										
Ptosis	2/4	2/6	0/3	0/4	0/2	—	3/3	0/1	5/12	3/11
Exotropia	4/4	5/6	1/2	0/2	—	—	0/3	0/4	5/9	5/11
Long philtrum	2/4	4/6	0/3	—	0/1	—	3/3	0/1	5/11	4/7
High-arched palate	4/4	2/6	0/2	0/2	0/1	—	3/3	—	7/10	2/8
Broad alveolar ridges	3/4	2/6	0/2	—	—	—	0/3	0/1	3/9	2/7
Carp-shaped mouth	2/4	0/6	0/3	0/2	0/2	—	0/3	0/1	2/12	0/9
Skeletal										
Narrow shoulders/thorax	4/4	2/6	0/3	0/4	1/2	—	3/3	—	8/10	2/10
Kyphosis, lordosis, scoliosis	4/4	2/5	0/3	0/4	3/3	—	3/3	—	10/13	2/9
Hip dislocation or contracture	0/4	1/6	0/3	0/4	0/1	—	0/3	0/1	0/11	1/11
Knee or elbow contractures	3/4	6/6	2/3	0/4	0/2	—	3/3	—	9/13	6/10
Camptodactyly	4/4	3/6	0/3	0/4	0/2	—	0/3	0/1	4/12	3/11
Ulnar deviation of fingers	4/4	0/5	0/3	0/4	0/2	—	0/3	0/1	4/12	1/10
Club foot/rocker-bottom feet/flat feet	4/4	0/6	1/3	0/4	0/2	—	3/3	0/1	8/12	0/11
Short neck	0/4	0/6	0/3	0/4	0/2	—	3/3	—	3/12	0/10
Neuromuscular										
Distal muscle weakness	4/4	1/6	3/3	0/4	0/3	—	3/3	—	10/13	1/10
Spasticity/hyperreflexia	3/4	0/6	3/3	0/4	3/3	—	0/3	0/1	9/13	0/11
Seizures	1/4	0/6	0/3	1/4	2/3	1/1	0/3	0/1	3/13	2/12
Drooling	4/4	0/6	2/3	0/4	3/3	—	0/3	—	9/13	0/10
Hypotonia	4/4	0/6	0/3	0/4	2/2	—	0/3	—	6/12	0/10
Other										
Intellectual disability	4/4	4/5	3/3	2/4	3/3	—	3/3	1/1	13/13	7/11
MRI abnormality	0/1	—	0/1	—	1/1	—	1/1	—	0/4	—

[a]K8070 showed an excess of fingerprint arches in all males and females and cataracts in 1/4 males and 1/5 females.
[b]K8615 showed athetoid/dystonic movements of the hands in 2/2 males.
[c]K9611 showed palmar hyperkeratosis in 3/3 males, Sidney lines in 3/3 males and 1/1 female, and small testes and penis in 3/3 males.

The folding free energy (ΔΔG) of model2 Amber and Charmm were both predicted to be destabilizing; however, the ΔΔG of model3 Amber was predicted to be stabilizing. The overall averaged ΔΔG is destabilizing but with a large deviation. The results of the energy calculations and the potential loss of favorable hydrogen bonds when R18 is mutated to K18 lead to our prediction that this mutation will be destabilizing and therefore potentially disease-causing. This prediction is consistent with third party predictors and online webservers FoldX, Eris, PoPMuSiC, I-Mutant 2.0 and PolyPhen-2, all of which predicted destabilization of the gene due to the mutation (Supplementary Material, Table S1).

L66H mutation

For the L66H mutation, our criteria allowed for only one model to be included in the analysis: model5 Charmmforce field. The model showed L66 being buried within the protein and protected from the water with no charged residues surrounding it (Supplementary Material, Fig. S3). This is an ambiguous alteration to predict, since it involves a titratable residue, the His residue, whose standard pKa is 6.5. No attempt was made to calculate the apparent pKa of H66 in the models because none of the existing predictors allows for protonation state adjustment. In our modeling, the H66 was kept fully ionized. The prediction of the average ΔΔG is destabilizing but no deviation can be given since only one model was used. All of the third-party servers agreed in their prediction of the mutation to be destabilizing (Supplementary Material, Table S1). This is what is expected since a replacement of the hydrophobic Leu with a charged His at position 66 in the hydrophobic core should be destabilizing.

R213W mutation

We were able to include three models for the R213W mutation including: model2 Charmm, and model4 Amber and Charmm (see Supplementary Material). Model4 Amber showed R213 forming two hydrogen bonds with the backbone oxygen at position 221 (Supplementary Material, Fig. S4), while all other models had R213 exposed to the water. Our energy results showed the R213W mutation would result in destabilization but the standard deviation is significant. All of the other servers also predicted this mutation to be destabilizing (Supplementary material, Table S1).

Table 2. Bioinformatic analysis of missense mutations in *ZC4H2*

	R18K	L66H	R213W
Consurf results	Exposed residue /Highly conserved	Buried residue /Highly conserved	Exposed residue /Highly conserved
ipTree	Destabilizing	Destabilizing	Destabilizing
MuStab	Decreased stability	Decreased stability	Increased stability
Mutation tasting	Disease causing	Disease causing	Disease causing
pMut	Neutral	Pathological	Pathological
Polyphen	Damaging	Damaging	Damaging
SIFT	Tolerated	Not tolerated	Not tolerated
Conserved	Conserved down to gallus	Conserved down to tropicalis	Conserved down to Fugu
Domain	zf-C4H2 coiled-coil domain	zf-C4H2 coiled-coil domain	No

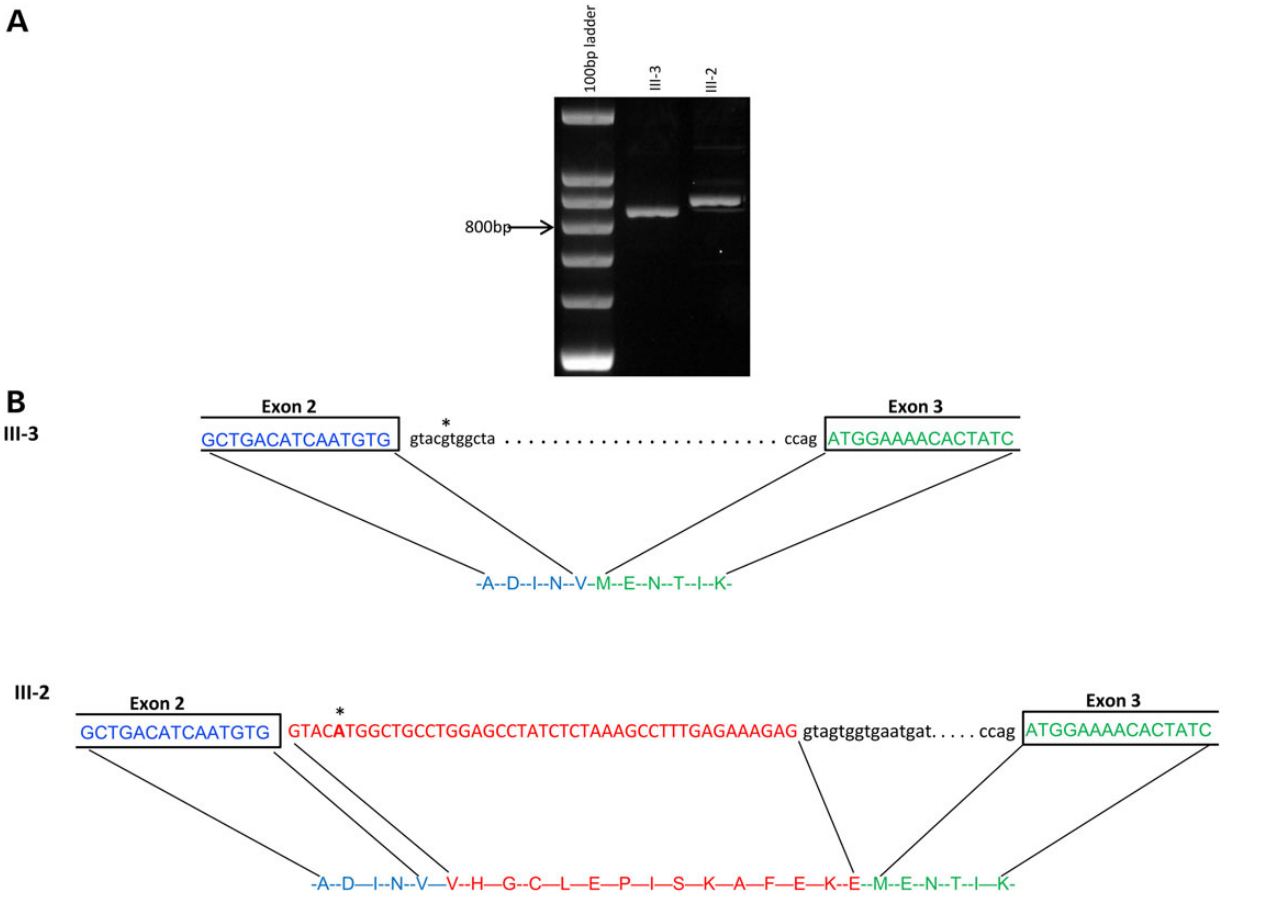

Figure 2. Partial analysis of *ZC4H2* cDNA in family K9333. (**A**) RT–PCR of *ZC4H2* mRNA prepared from individuals III-2 and III-3 in family K9333 using 4% agarose. (**B**) Sequence of *ZC4H2* from individuals III-2 and III-3 flanking the exon 2/exon 3 boundary. The schematic shows how the insertion of 45 bp in III-2 results from the c.225 + 5G > A mutation. Bases in the transcripts are in capital letters. Red letters indicate intron 2 sequence. The mutation site is indicated with an asterisk.

Immunofluorescence studies of wild-type and R213W in Cos-7 cells

As a preliminary analysis of the effect of the missense mutations on the *ZC4H2* protein, immunofluorescence studies were conducted in Cos-7 cells using GFP-tagged wt-type and mutant constructs. We noticed that in Cos-7 cells transfected with *ZC4H2*-wt, most of the protein resided in the nucleus (Supplementary Material, Fig. S5A). For two of the missense mutations, *ZC4H2*-R18K and *ZC4H2*-L66H, their location pattern was not much different from wild-type. However, for Cos-7 cells transfected with *ZC4H2*-R13W, the majority of the protein still resided outside of the nucleus (Supplementary Material, Fig. S5A). Because of this observation, the ZC4H2 protein was analyzed to see if it contained a nuclear localization signal (NLS). Using an NLS predictor (http://www.moseslab.csb.utoronto.ca/NLStradamus), a NLS domain was predicted to be contained within amino acids 207–224 (KAKSRSRNPKKPKRKQDE). Residue R213 (underlined) is located in this stretch of amino acids. Thus, the immunofluorescence data indicate the R213W mutation perturbs the NLS domain which might influence whatever function the *ZC4H2* protein has within the nucleus, thereby contributing to the phenotype of males in family K8615 relative to males with other *ZC4H2* mutations.

Spatiotemporal expression pattern of zebrafish *zc4h2*

To determine the potential function of ZC4H2 and the pathogenicity of the missense mutations in human patients, we chose to use a zebrafish model, which allows rapid genetic, behavioral and embryological experimental analysis of a vertebrate nervous system. Using whole-mount *in situ* hybridization, the spatiotemporal expression of the *zc4h2* ortholog during zebrafish embryonic development was explored. *zc4h2* transcripts were mainly detected throughout the developing CNS, but were excluded from the most medial ventricular zone suggesting expression in differentiating progenitors and mature neurons and/or glia (Fig. 3). Based on our observation of human ZC4H2 nuclear localization of in Cos-7 cells, the cellular localization of the zebrafish *zc4h2* gene product was examined. As shown in Supplementary Material, Figure S5B, ecotopically expressed Zc4h2 was located mainly in the nucleus of zebrafish cells.

Generation of zebrafish *zc4h2* mutants

Two independent zebrafish *zc4h2* null mutant alleles were created using TALEN methodology (Supplementary Material, Fig. S6). Homozygous mutant fish were viable beyond embryogenesis, and all mutant larvae exhibited abnormal flexion of the pectoral fins and abnormally positioned eyes when compared with wild-type siblings (Fig. 4A–D). These findings are consistent with the contractures and exotropia observed in the human patients with *ZC4H2* mutations. Additionally, at 5 days post-fertilization (dpf), all mutant larvae exhibited active movements of the pectoral fins, continuous swimming movements and balance problems (Supplementary material, Videos S1 and S2), compared with wild-type siblings that had saccadic eye movements and were mostly stationary (Supplementary material, Videos S3 and S4). The mutants also exhibited an open mouth as well as continuous jaw movements (Supplementary material, Video S5).

Based on the expression of *zc4h2* in the CNS during early development and the apparent hyperactivity exhibited by the fish, an anatomical analysis was undertaken using markers for specific classes of neurons. To address a potential motoneuron defect reported after morpholino knockdown of the gene (3), we stained wild-type and mutant fish at 24 h with an antisense probe for *isl1* mRNA and a znp-1 (anti-synaptotagmin 2) antibody, both known markers for zebrafish motoneurons (4,5). Surprisingly, we observed no differences between wild-type and *zc4h2* mutant embryos for either of these markers (Supplementary material, Fig. S7). However, in mutant embryos, expression of both *vsx2*, which labels V2a interneurons (6), and *vsx1*, which labels V2a/b interneuron precursors (7), was significantly decreased throughout the CNS (Figs 4E,F and 5A–D). To assess whether loss of V2 neuron markers was correlated with altered neural progenitor specification, we examined the expression of *dbx2* and *nkx6.1*, markers for adjacent p1 and p2 progenitor domains, respectively (8). In mutant embryos, *dbx2* expression was clearly expanded, with a corresponding loss of *nkx6.1* expression in the midbrain tegmentum, hindbrain and spinal cord (Fig. 5E,F and I–L). Interneuron defects

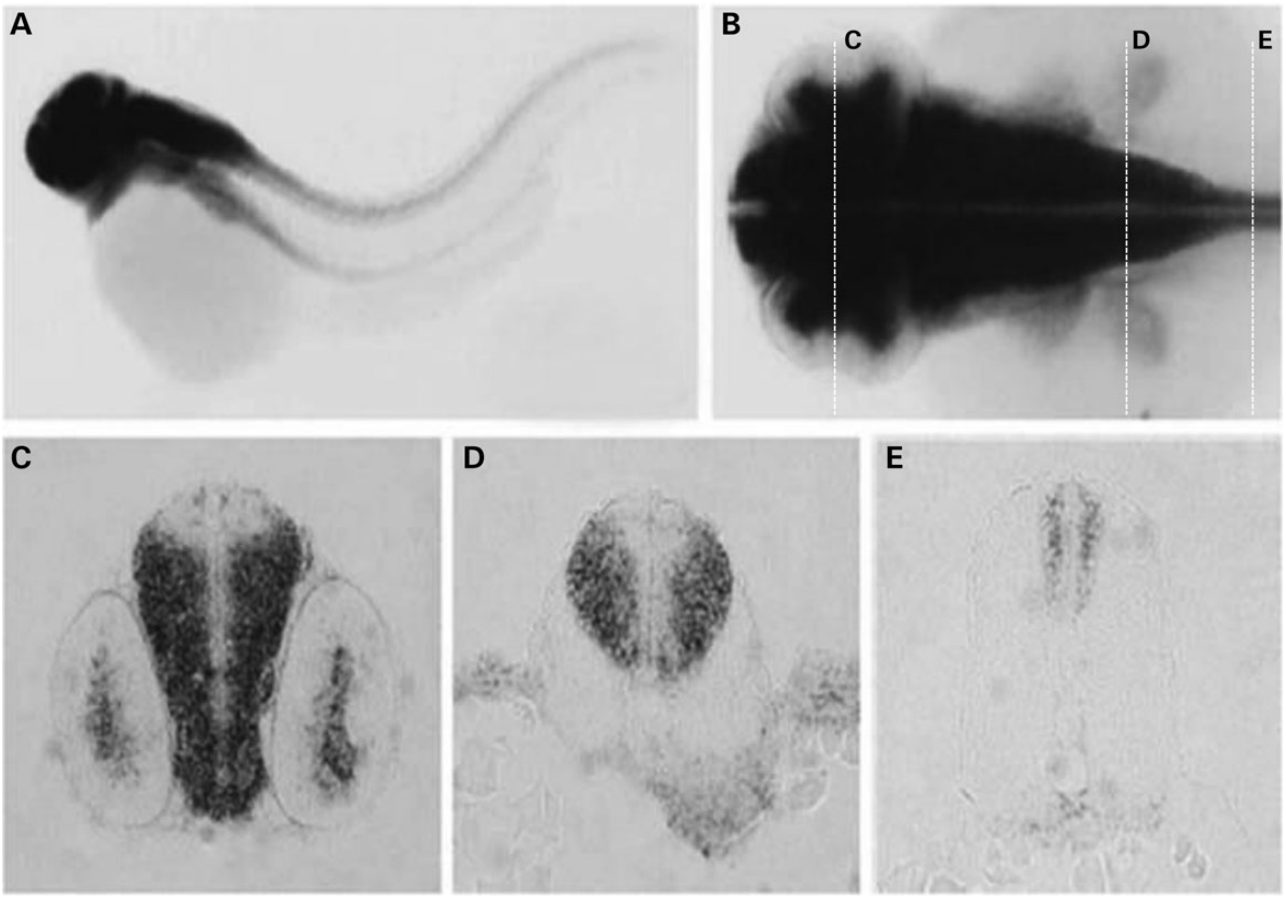

Figure 3. Whole-mount *in situ* hybridization of the zebrafish *ZC4H2* ortholog, *zc4h2*. (**A** and **B**) Whole mount *in situ* hybridization at 48 hpf. (**C–E**) Sections cut at the positions indicated in (B). *zc4h2* shows CNS-specific expression outside the ventricular zone where stem cells are found.

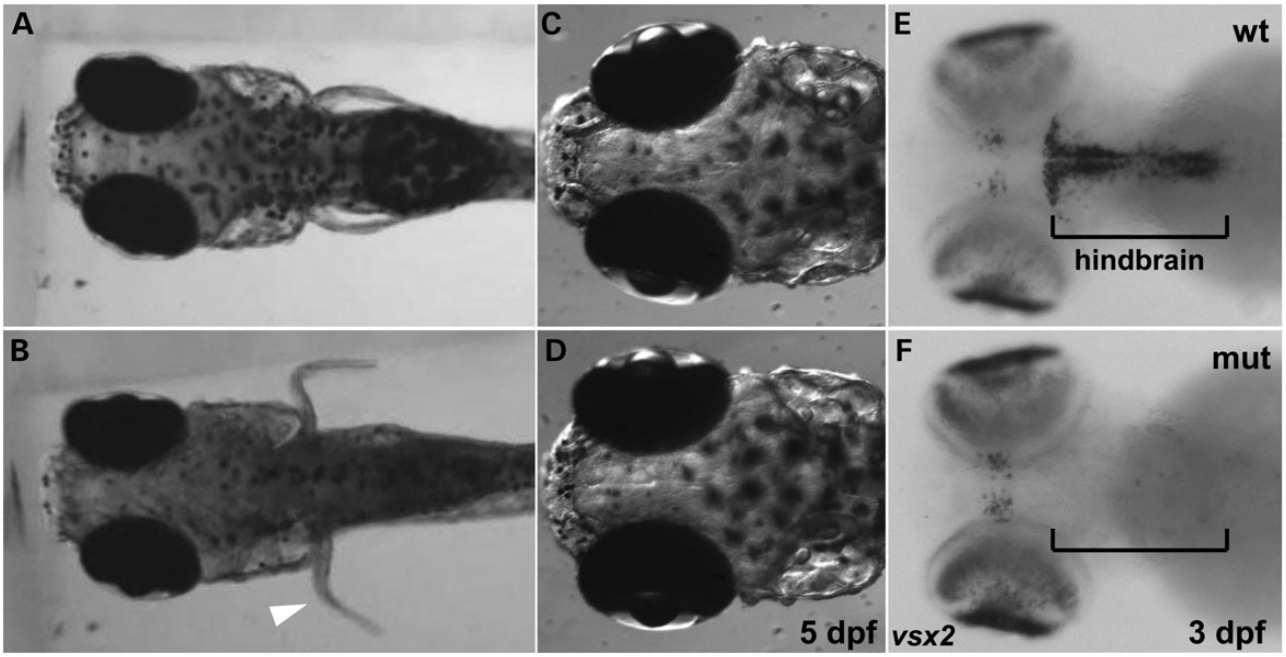

Figure 4. Frames from movies of wild-type sibling (wt) and *zc4h2* KO mutant (mut) embryos. (**A–D**) At, 5 dpf, dorsal views showing normal positioning of fins and eyes in a wt larva (A and C) and abnormally flexed fins and outward positioning of eyes in a *zc4h2* homozygous mutant larva (B and D). Larvae were positioned in 3% methylcellulose. (**E** and **F**) Dorsal views showing absence of *vsx2* expression, which labels hindbrain interneurons, in 3 dpf mutant larvae.

appeared to be restricted to populations arising from these progenitor domains, as expression of *dlx2a*, which marks interneuron precursors in the forebrain, was unaffected (Fig. 5G–H).

Consistent with the defects observed in V2 markers, we found a significant reduction in the number of GABAergic neurons in the intermediate spinal cord of mutant embryos as evidenced by decreased staining for *gad1* (Fig. 5M and N). Decreased expression of *gata3*, which encodes a transcription factor important for the generation of V2b inhibitory neurons (9), was also observed in mutant embryos (Fig. 5O and P). In addition, ectopic expression of *glyt2a*, a marker for glycinergic neurons, was observed in this same region in mutants (Fig. 5Q and R). In contrast, no difference between mutants and wild-type siblings was observed when *vglut2.1*, a marker for glutamatergic excitatory neurons, was examined (Fig. 5S and T). By 33 hpf, a reduction in GABAergic interneurons was observed in the midbrain tegmentum region of *zc4h2* mutants, which is thought to control pectoral fin and jaw movements (Supplementary Material, Fig. S8A and B) (10). Further analysis showed similar reductions in the expression of both *gata2* and *gata3* in the spinal cord, hindbrain neurons and the midbrain tegmentum at 24 hpf (Supplementary Material, Fig. S8C–F). Additionally, the expression of *scn1Lab*, a voltage-gated sodium channel expressed in inhibitory interneurons (11), was slightly decreased (Supplementary material, Fig S8G and H). Together, these data suggest that zc4h2 functions to regulate the specification of p2 CNS progenitors, which normally generate a population of primarily GABAergic interneurons (Fig. 5U).

Human *ZC4H2* mutations affect the ability to rescue zebrafish mutants

To explore the pathogenic nature of the three missense mutations in *ZC4H2*, wild-type and mutant human *ZC4H2* mRNA constructs were injected into *zc4h2* mutant zebrafish 1-cell embryos and expression of *gad1* was examined at 24 h. Wild-type human *ZC4H2* mRNA was able to restore *gad1* expression (Fig. 6A and C, G and I), as well as the behavioral phenotype (Supplementary material, Video S6). However, both L66H and R213W mRNAs were only able to partially rescue *gad1* expression (Fig. 6D and E, J and K), consistent with both mutations being pathogenic as predicted both bioinformatically and by *in silico* structural modeling. Interestingly, the R18K mRNA was able to restore *gad1* expression (Fig. 6F and L), consistent with family K9611 having a somewhat milder phenotype (Table 1).

Discussion

XLIDs comprise many entities, some syndromic and many non-syndromic without any distinguishing clinical finding other than ID. With respect to the more than 110 syndromal XLID conditions, many present with microcephaly or spastic paraplegia. However, few present with scoliosis or congenital contractures. Therefore, identification of the genetic cause for a syndrome with all four of these clinical findings would likely prove to be quite informative as to their interconnection.

We have identified four mutations in *ZC4H2*, three missense and one in-frame insertion, in the original family with MCS and three other families with a constellation of microcephaly, short stature, contractures, spasticity and ID.

ZC4H2 is located in Xq11.2 and finding mutations in four families in our study along with mutations in another five families (3) is consistent with the clustering of both XLID genes and families in the pericentric region of the X chromosome. Of the 170 XLID entities (syndromic and non-syndromic) with a gene finding, 70 (41%) of them are localized to the pericentric region and of the known 114 XLID genes, 50 (44%) are located in this same region (http://www.ggc.org/research/molecular-studies/xlid.html). The reason for this clustering is not evident at this time.

ZC4H2 is a zinc-finger protein belonging to the family of proteins with a C-terminal zinc-finger domain characterized by four cysteine residues and two histidine residues. Additionally, ZC4H2 has a coiled-coil domain. However, little was known about the

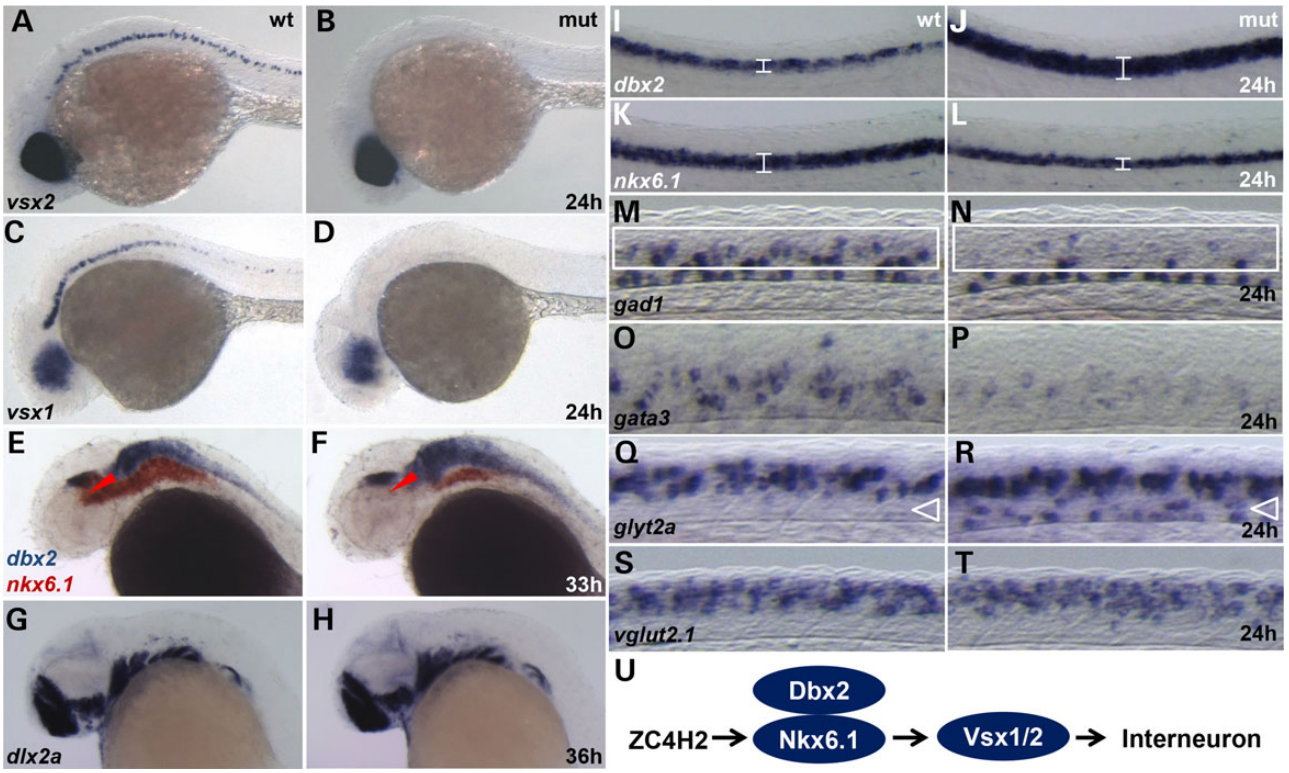

Figure 5. Whole-mount *in situ* hybridization for neural markers in wild-type sibling (wt) and *zc4h2* KO mutant (mut) zebrafish. (**A–D**) At 24 h, expression of *vsx2* and *vsx1*, markers for V2a interneurons and V2a/b precursors, respectively, was significantly reduced in the brain and spinal cord of mutants. (**E** and **F**) At 33 hpf, expression of *dbx2* (blue) was increased and expression of *nkx6.1* (red) was decreased in the midbrain tegmentum (*). and hindbrain of mutants. (**G** and **H**) At 35 hpf, expression of *dlx2a*, a marker of forebrain GABAergic precursors, was unaffected in mutants. (**I–L**) At 24 hpf, expression of *dbx2* was increased and expression of *nkx6.1* was decreased (brackets) in the spinal cord of mutants. (**M** and **N**) Expression of *gad1*, a marker for GABAergic neurons, was decreased in V2 interneuron territories of the mutant spinal cord (White rectangles). (**O** and **P**) *gata3* expression, a marker for V2b inhibitory interneurons, was dramatically reduced in the mutant spinal cord. (**Q** and **R**) *glyt2a*, a marker for glycinergic neurons, was ectopically expressed in the mutant spinal cord. (**S** and **T**) *vglut2.1* expression, a marker for glutamatergic excitatory neurons, was unaffected in *zc4h2* homozygous mutant zebrafish. (**U**) Model for Zc4h2 function in zebrafish interneuron specification.

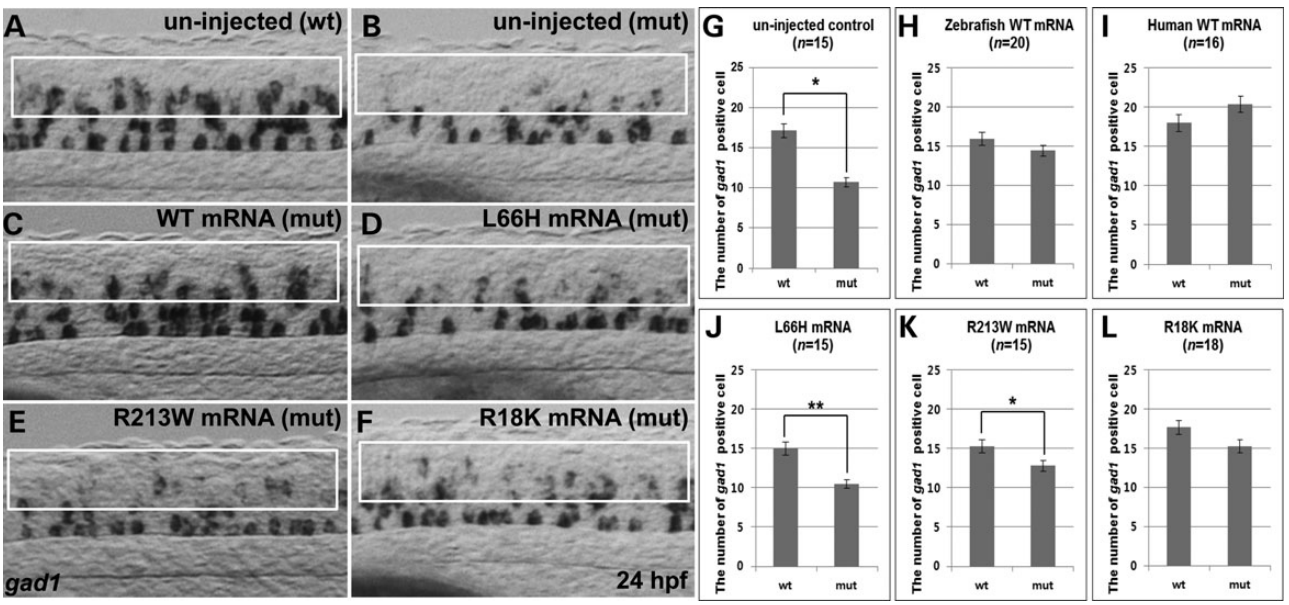

Figure 6. Rescue of *gad1* expressing interneurons in zebrafish *zc4h2* KO mutant by injection of human *ZC4H2* mRNAs. (**A** and **B**) *gad1* transcripts in uninjected control wild-type sibling (wt) and *zc4h2* homozygous mutant (mut) embryos at 24 hpf. (**C–F**). Rescue experiments with mRNA containing the indicated human *ZC4H2* WT and mutations. White rectangles indicate V2 interneuron territories. (**G–L**) Quantitative analysis of *gad1* transcripts in *zc4h2* mutant (mut) embryos with and without injected mRNA. $^{**}P < 0.001$, $^{*}P < 0.05$.

function of the ZC4H2 protein other than it was proposed as a candidate gene for XLID based on a computational approach (12). Our *in silico* modeling of the effects of these three missense mutations of the wt properties of ZC4H2 protein predicted that the mutations would destabilize the structure of ZC4H2 and thus would indirectly affect its function.

The phenotypes of the families reported by Hirata *et al*. (3) and here (Table 1) have quite similar clinical findings. Family 1 in Hirata *et al*. (3), reported as Wieacker–Wolff syndrome (13), was not originally considered as having arthrogryposis multiplex congenita. Rather, both Wieacker–Wolff syndrome and MCS appear to have had some lower limb musculoskeletal findings at birth, but clearly not something suggestive of early onset arthrogryposis. It is possible that both syndromes, as well as our three other families represent a condition that is progressive in nature.

In order to unravel the function(s) of the ZC4H2 protein, two zebrafish mutant alleles were generated using TALEN methodology. Homozygous mutant larvae for both alleles showed motor hyperactivity, abnormal swimming and continuous movement of their jaws. The fish also exhibited contractures of the pectoral fins and abnormal eye positioning suggestive of extropia. Together, these phenotypes are highly reminiscent of the human patients carrying *ZC4H2* mutations, indicating that our zebrafish model can be used to determine the underlying cellular and molecular mechanisms of motor dysfunction.

Contrary to a previous study carried out using morpholino knockdown of *zc4h2* (3), we observed no defects in motoneurons or tail coiling of mutant larvae (Supplementary Material, Video S7 and Figs S7 and S9). It is therefore possible that some of the zebrafish phenotypes observed in this earlier work may have been due to off-target effects of morpholinos (14). Instead, we observed a clear loss of multiple markers of the V2a and V2b interneuron lineages in the hindbrain and spinal cord, as well as loss of GABAergic neurons in the midbrain tegmentum. In addition, we observed accompanying ectopic glycinergic neurons in the ventral spinal cord, consistent with a defect in neuronal fate specification. Analysis of neural progenitor markers strongly suggested that the loss of V2 interneurons results from mis-specification of adjacent spinal cord progenitor domains. Intriguingly, the neuronal subtypes that are specifically affected in mutant larvae provide a potential physiological basis for the behavioral phenotypes observed in both zebrafish and humans. While we cannot rule out an additional role for ZC4H2 in excitatory synaptic function as previously reported (3), these findings were based on overexpression studies of GFP-tagged protein and may not reflect an endogenous role. Furthermore our data indicate that wild-type protein tagged with small epitopes is primarily localized to the nucleus when expressed in both Cos-7 cells and zebrafish embryos.

The presence of two isoforms for *ZC4H2* with different levels of expression in tissues and sections of the brain likely contributes to the broad clinical spectrum of involvement of the central and peripheral nervous systems observed in our four families and those reported by Hirata *et al*. (3). Additionally, it is quite possible *ZC4H2* has more than one function as it contains both a zinc-finger domain and a coiled-coil domain. These possible explanations for the clinical variability are evident in that males in family K8615 whose mutation, R213W, affects nuclear localization of the proteins, do not exhibit the narrow shoulders, kyphosis nor degree of distal contractures observed in the other families and males in family K9333 whose mutation, R18K, would not be present in the short form of *ZC4H2* do not present with spasticity, drooling and hypotonia.

In summary, our work provides a novel mechanistic model for the etiology of MCS and other XLID entities with mutations in *ZC4H2*. The expression of *ZC4H2* in the developing CNS, as well as the neuroanatomical defects in zebrafish mutants, suggest a required function in neural tube progenitor specification that ultimately affects interneuron fate and connectivity throughout the brain and spinal cord. Because loss of GABAergic neurons is the most prominent phenotype arising from this mis-specification, *ZC4H2* loss of function may therefore result in hyperexcitability of multiple brain and spinal circuits.

Materials and Methods

Clinical reports

Fourteen males were affected in the four kindreds (Table 1). The 14 prominent findings, i.e. those occurring in over 50% of males, included intellectual disability, motor and speech delay, short stature, microcephaly, distal muscle weakness, foot abnormalities (rocker-bottom, pes planus, club foot), knee and/or elbow contractures, narrow shoulders and upper torso, ptosis and/or exotropia, long philtrum, highly arched palate, spinal curvature (kyphosis, scoliosis, lordosis), hypotonia and/or drooling and spasticity. An excessive number of fingerprint arches was present in all affected males and females in K8070 but was not noted in the other kindreds.

Seventeen carrier females were identified in the four kindreds. Clinical information on the females is most complete in K8070 and K8615, somewhat complete in K9611 and not available in K9333 (Table 1). Growth was typically normal but with head circumference and stature in the lower centiles: only one carrier female had short stature (<3rd centile) and only one had microcephaly (<3rd centile). Although IQ measurements were not available, most carrier females (seven out of 11) were considered to have low or borderline cognitive function. None had hypotonia, drooling or spasticity. Six females had elbow or knee contractures, two had seizures, three had camptodactyly and two had narrow shoulders/thorax. Craniofacial features were not distinctive although several carrier females in K8070 had ptosis, long philtrum, high-arched palate and broad alveolar ridges and all except 1 (five out of six) had exotropia. Regarding X-inactivation (XI), it varied within families (Fig. 1). Thus, it was difficult to observe a correlation between skewing of XI and the lack of phenotype in the carrier females or the severity of the phenotype in males in the family.

K8070

Clinical details of the four affected males and six affected females were reported in 1991 (1). The prominent findings are given in Table 1 and the pedigree shown in Figure 1A. The gene was linked to DXYS1 at Xq 21.31 (lod score 2.788).

K8615

Three males in three generations are affected in this kindred (Fig. 1B and Table 1). The carrier II-4 has intelligence in the mildly disabled range and has seizures. II-2 is said to function in the low normal range. Neither have spasticity. III-3 is considered to be normal.

II-1 had a birth weight of 4 kg. Club feet were present. Development was globally delayed with sitting at 2 years, crawling at 2 years and only two indistinct words by 2 years. At 2 4/12 years, his height was 85 cm (10th centile), weight 12.6 kg (20th centile) and OFC 50.5 cm (75th–90th centile). The face appeared normal. There was no nystagmus but some tongue protrusion, drooling and dystonic movements of the oropharynx. Fifth finger clinodactyly and horizontal palmer creases were present. The lower

limbs showed weak muscles, hypertonia, hyperreflexia, knee contractures and several scars from club foot surgery. MRI at age 1 year was normal. He now lives in a community living facility and has a reported IQ of 60.

Few details are available on III-4 and III-5 except they are intellectually disabled and have no speech, use wheelchairs for ambulation and have lower limb weakness and spasticity.

K9333

Three males in two generations are affected in this kindred (Fig. 1C and Table 1). Ptosis and a long philtrum have not been recognized. Nor have narrow shoulders, contractures, camptodactyly or ulnar deviation of the fingers. All have had scoliosis.

II-4 (female) is reported to have a speech impediment, seizures and abnormal gait. The father of III-4 had congenital cataracts.

II-5 was globally delayed from birth. He fed poorly. He never spoke, nor walked. He kept his arms and knees bent, but he did not have contractures. He was considered to have cerebral palsy. His maximum weight was 40 lbs. He had significant scoliosis by the age of 10, but was not regarded as suitably robust for any surgery. A prolonged seizure occurred on at least one occasion. He died suddenly at age 22 years.

III-2, the nephew of II-5, had poor feeding in infancy and delayed development. It was suspected early that he had the same condition as his uncle. He has had seizures, stenosis of the foramen magnum and upper cervical spine (not requiring surgery) and scoliosis.

III-4, born 17 months after III-2, had a birth length of 48 cm and weight of 3.2 kg. Bilateral cataracts, like his father's, were present. Development was globally delayed, and he was soon recognized as having the familial condition. Motor delay was apparent by 6 months. At 23 months, he was not sitting alone and had poor muscle tone and head control. Examination then showed length 79.5 cm (<3rd centile), weight 11 kg (20th centile) and OFC 44.5 cm (<3rd centile), roving eye movements, increased startle reflex, hypotonia, everted lower lip, hyperreflexia and Babinski reflexes. MRI was unremarkable. He has continued to have profound delays, but is aware of his surroundings. He does not have seizures.

K9611

The Italian family was ascertained through individual III-3 (Fig. 1D) who sought advice for her reproductive options, considering the presence of psychomotor delay and intellectual disability in her two stepbrothers (III-4 and III-5) and her maternal uncle II-4. Individual IV-4 also reportedly had poor growth and psychomotor delay, still did not speak, and had not started primary school at age of 7 years because of complicated pneumonia. Clinical findings are reported in Table 1 and are briefly summarized here.

Patient II-4 is 59 years old. He is a gentle person living in a sheltered residence. He speaks a few words and walks with great difficulty. He has short stature (159 cm) and small head circumference (52.5 cm), small hands with short fingers and no fingerprint arches. Patient III-4 is 36 years old. He has short stature (157 cm) and small head circumference (52 cm), small hands with short fingers and one arch on the right index finger. Patient III-5 is 33 years old. He also has short stature (159 cm) and small head circumference (52 cm), short palm with normal middle finger and three fingerprint arches. All patients have palmar hyperkeratosis and complete or partial Sidney line that is also observed in examined female carriers. Also patients III-4 and III-5 had persistence of deciduous teeth (premolars), while III-4 had two connatal teeth in the lower jaw. All affected males (II-4, III-4 and III-5) had small testes and short penis.

While non-carrier females, like proband III-3 and her sister III-1, have normal intelligence, heterozygous females have variable degree of intellectual disability and have never excelled at school. IV-3 is the only one who completed high school but with individual support and personalized programs and, according to her own mother, is the most intellectually impaired of the carriers. The X-inactivation ratio using the androgen receptor CAG repeat showed that IV-3 had markedly skewed X-inactivation with a 91:9 ratio. Given the close proximity of the ZC4H2 and AR genes in Xq11.2-q12 (2.6 Mb apart), we deduce from haplotype analysis that the ZC4H2 mutation is on the active X chromosome in informative carrier females and is expressed in 91% of blood cells of IV-3, compared to 72 and 71% in I-2 and II-2, respectively.

ZC4H2 mutation detection

K8070

Exon capture and deep sequencing of 718 genes on the X-chromosome was done as detailed in (15).

K8615

Screening of XLID families linked to the Xq12 region of the X chromosome was conducted using M13 tagged primers for the exons of *ZC4H2*. The PCR products were amplified and then purified with ExoSap from USB. Using M13 primers and the ABI 3730 sequencer the PCR products were sequenced with BIGDye Terminator v3.1 Cycle sequencing Kit from Life Technologies. The sequence data were analyzed using SEQMAN from DNASTAR. The R213W alteration was found in an affected male and then tested and confirmed in the entire family.

K9333

Genomic DNA samples from 82 XLID male samples [30 unrelated and 52 related (26 sib-pairs)] were used to generate Illumina sequencing library using TruSeq DNA sample Preparation kits. The X chromosome exome was enriched using an Agilent SureSelect X chromosome exome kit. Individual libraries were uniquely barcoded and sequenced on the Illumina HiSeq2000 platform using either 75 bp or 100 bp pair-end read modules. INDEL realignment and base recalibration were conducted using GATK. Unified Genotyper (GATK) was used for variant calling [parameters: –ploidy 1, all else at default]. The final variant output was pre-processed by removing variants with zero coverage from one strand and by removing variants in close proximity (<10 bp) to another variant in the same sample. Such variants are most likely to be erroneous.

Potential XLID variants were identified using a sib-pair comparison filter. Given the rarity of XLID, this filter assumes that unique XLID mutations are rare and shared only between affected family members. Variant output from sequencing was filtered by retaining only those variants shared between related samples and by removing variants shared between unrelated samples in our cohort. Additionally, variants shared with 162 male samples (unaffected controls) from the 1000 Genomes project were removed. This filter effectively reduced the total number of variants from 1774 ± 240 variants per samples to 31 ± 5 variants per sample, easing identification of potential disease-causing mutations.

K9611

DNA was extracted from all individuals of generation II and III as well as from I:2, IV:1 and IV:3 (Fig. 1D). Linkage was established

between markers DXS993 (Xp11.4) and DXS8055 (Xq23) identifying a minimal interval of 70 Mb. Considering the large interval and the corresponding elevated number of candidate genes, whole-genome diploid sequencing was performed on individuals I:2 and II:4 by Complete Genomics Inc. (Mountain View, CA, USA) using a sequencing-by-ligation method as described by Drmanac *et al.* (16). Briefly, the human genome sequencing procedures include DNA library construction, DNA Nano-Balls (DNBs) generation, DNB array self-assembling, cPAL-based sequencing, imaging, image data analyses including basecalling, DNB mapping and sequence assembly. Data were provided as lists of sequence variants (SNPs and short indels) relative to the reference genome. Human Genome Build 36 was used as a reference to identify sequence variations in each sample. Analysis was done using Complete Genomics analysis tools (cga tools) and TIBCO Spotfire.

Confirmation of *ZC4H2* variants

The variants in *ZC4H2* were confirmed using Sanger Sequencing in a second sample of genomic DNA. Once confirmed, other members of the four families were also sequenced.

Polymorphism analysis

Screening of normal control individuals (male and female) for the *ZC4H2* c.197T > A variant consisted of utilizing allele specific amplification (ASO) using primers:

ZC4H2 ASO K8070F 5′ATGGCCCATGTGGAGGAtCa 3′; *ZC4H2* ASO K8070R 5′CACTACCACAGCCTGCACTGA 3′. A control amplicon was amplified using primers for *ZNF711*: *ZNF711* ex 4AF TTCCTGGTTGTTTGATTTT; *ZNF711* ex 4AR ACAAGGTCAGCCACG AAAAC. The *ZC4H2* product was 177 bp and the *ZNF711* product was 498 bp. PCR products were used.

ZC4H2 tissue expression analysis

Human Fetal MTC panel (Clontech) was utilized for analysis of *ZC4H2* expression in fetal tissues. The MTC cDNA preparations utilized for RT–PCR analysis were conducted as outlined by the manufacture in a final reaction volume of 20 µl. Primers for *ZC4H2* long were: *ZC4H2* long cDNA F CCCTTGGCTGGTGTA TTTGT; *ZC4H2* long cDNA R TAGGAGACTTCGTGGGGTTG and primers for *ZC4H2* short were: *ZC4H2* short cDNAF ATGGAAGAT CAAGGCTCGTT; *ZC4H2* short cDNAR TTATTCATCCTGCTTCC GTTTC. Primers for the control gene, *GAPDH* were RT-F 5′-ATG GGTCAGAAGGATTCCTATGTG-3′ and RT-R 5′-TGTTGAAGGTCTC AAACATGATCTGG-3′. PCR conditions for *ZC4H2* were: initial denature 95°C at 5 min, followed by 30 cycles of denature at 95°C, 30 s, annealing at 57°C, 30 s and extension at 72°C, 40 s. Final extension was at 72° for 5 min. For *GAPDH*, the PCR conditions were the same except the annealing was done at 55°C, the extension was for 30 s and only 25 cycles were used. Ten microliters of the PCR reaction were run on a 3% agarose gel in 1XTBE.

Cos7 transfection

RNA

RNA was prepared from a control cell line and from patient lymphoblast cell lines using GenElute Mammalian Total RNA Miniprep (Sigma catalogue number RTN-70).

Plasmid construction

cDNA was prepared with Superscript First Strand Synthesis Kit for RT–PCR (Invitrogen catalogue number 11904-018) from 2 µg of RNA prepared from lymphoblast cell lines. A PCR reaction with the oligos 5′GGAGAGGGGGAAGCTTGTAA3′and 5′TAGGAG ACTTCGTGGGGTTG3′ and PFU Turbo (Stratagene catalogue number 600250) was employed to generate the insert. The fragment was run on a 1% TAE agarose gel and purified with a Gel Extraction Kit (Qiagen catalogue number 287040.) The purified product was cloned into pcDNA3.1D/V5-His-Topo vector using the pcDNA3.1 Directional Topo Expression Kit (Invitrogen catalogue number K4900-01). All plasmids generated have a V5 tag on the c-terminus. The LacZ V5 plasmid was included as a positive control with the kit. Plasmids were sequenced to confirm the insert with the vector specific primers PC (5′ GGG AGA CCC AAG CTG GCT AGT 3′) and BGH (5′ TAG AAG GCA CAG TCG AGG 3′) and insert specific primers 5′CACCATGGCAGATGAGCAAGAAATC3′ and 5′TTCATCCTGCTTCCGTTTCGGCTT3′. Site-directed mutagenesis was done to create the following alterations: L66H and R213W using the QuikChange II Site-Directed mutagenesis kit (Stratagene catalogue number 200524). All constructs were sequenced to verify the alteration.

Cell culture

Cos-7 cells were obtained from American Tissue Culture Collection. The cells were cultured in DMEM (Sigma catalogue number D5796) supplemented with 10% FBS (Atlanta Biologicals catalogue number S12450H), 1 × Penicillin/Streptomycin (Sigma catalogue number P0781), 2 mM glutamine (Sigma catalogue number G7513) in a 5% CO_2 humidified 37C incubator.

Transfection

Cos-7 cells were cultured on poly-L-lysine (Sigma catalogue number P4707) coated 24-well tissue culture dishes in growth media 18–24 h priors to transfection. A transfection complex containing DMEM (Sigma catalogue number D5796), 1 µg of plasmid and 2 µl of Lipofectamine 2000 (Invitrogen catalogue number 11668-027) was prepared and added to each well. After 24 h, the transfection complexes were removed, the cells washed one time with PBS and growth media was added back to the cells. Twenty-four hours post-transfection, the cells were moved to a poly-L-lysine coated glass slip and cultured in growth media for 24 h.

Immunofluorescence

The cells were washed twice with PBS, fixed with 4% paraformaldehyde/PBS, permeabilized with 0.1% Trtiton X-100/PBS and blocked with block (2% horse sera, 0.4% BSA in PBS). The cells were incubated with anti-V5 mouse monoclonal antibody (Invitrogen catalogue number R960.25), washed with block and incubated with Alexa Fluor 594 anti-Mouse IgG and Alexa Fluor 488 Phalloidin (Invitrogen catalogue numbers A21200 and A12379), washed with block, incubated with DAPI and mounted with Gold Prolong antifade medium (Invitrogen catalogue number P36930).

ZC4H2 modeling

Structures

X-ray structures are unavailable for Zinc Finger ZC4H2; therefore, it was necessary to utilize the online webserver I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (17) to create models for the wild-type structure of ZC4H2. Using the five WT models and the SCAP program, which is part of Jackal package (http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Jackal), we created separate structures for the three independent point mutations investigated in this work (R18K, L66H and R213W). Then the WT and the mutant structures were energetically minimized with two force fields: Amber (18) and Charmm (19). This resulted in total of 40 structures (five general models,

each model including a WT, R18K, L66H and R213W version, and each WT or mutated structure being minimized using the two aforementioned fields).

The sequence similarity between the ZC4H2 sequence and any available file in the Protein Data bank (PDB) (20) is very low and it is anticipated that *in silico* generated models will be of low quality. Since the goal is to assess the effect of single point mutations on ZC4H2 protein stability, one can argue that global structural feature will be less important than the local structural arrangement. Because of that, we applied local structural considerations to eliminate insufficient models. The considerations are as follows: (i) all of the models were submitted to a third party server, Verify3D (http://nihserver.mbi.ucla.edu/Verify_3D/), which provides an analysis of the quality of a model. Scores are marked on a scale of –0.5 to 1.0. Models below a particular manually selected threshold value of 0.20 were eliminated; all models were then visually inspected to verify acceptable structural features including; (ii) the formation of favorable hydrogen bonds between the target and near-by residues; (iii) the hydrophilic R18 and R213 being exposed to the water phase and hydrophobic L66 being at least partially buried or protected from the water phase; (iv) to assure that like charges are either at distances greater that 6Å or are involved in hydrogen bonding; and finally (v) to verify that the surrounding residue environment close to the target sites does not include unacceptable features such as buried charged residues that are not involved in hydrogen bonds and over packing. Only structural models satisfying these conditions were used for the energy analysis (all models can be downloaded from http://compbio.clemson.edu/ZC4H2).

Modeling the effect of mutations on ZC4H2 protein stability

The folding free-energy changes were calculated with the Molecular Mechanics Generalized Born (MMGB) (21):

$$\begin{aligned}\Delta G(\text{folding}) &= G(\text{folded}) - G(\text{unfolded}) \\ &= G(\text{folded}) - G_0(\text{unfolded}) - G_x(\text{unfolded}). \quad (1)\end{aligned}$$

Where G(folded) is the total potential energy of the folded state and the G(unfolded) is the total potential energy of the unfolded state. The energy, G(unfolded), of unfolded state, is split into two terms, G_0(unfolded) and G_x(unfolded), as discussed in our previous works (22,23). G_x(unfolded) is the energy of unfolded state of 'x' residue segments at the center of mutation site, while G_0(unfolded) is the energy of unfolded state of the rest of protein. Under such an assumption, the G_0(unfolded) is identical for WT and mutants and cancels out in Eq. (2) and therefore does not need to be calculated. The folding energy change due to a mutation is calculated with the following equation:

$$\begin{aligned}\Delta\Delta G(\text{folding\_mutation}) &= \Delta G(\text{folding\_WT}) - \Delta G(\text{folding\_mutant}) \\ &= G(\text{folded\_WT}) - G_x(\text{unfolded\_WT}) \\ &\quad - G(\text{folded\_mutant}) + G_x(\text{unfolded\_mutant}). \quad (2)\end{aligned}$$

Where $\Delta\Delta G$(folding_mutation) represents the folding energy change due to a mutation; ΔG(folding_WT) and ΔG(folding_mutant) are the folding energies of the WT protein and the mutant, respectively. In addition, many third party structure- and sequence-based servers were employed to verify our prediction results including Eris, FoldX, I-Mutant 2.0, PolyPhen 2, and PoPMuSiC (http://troll.med.unc.edu/eris/, http://foldx.crg.es/, http://folding.uib.es/i-mutant/i-mutant2.0.html, http://genetics.bwh.harvard.edu/pph2/, http://babylone.ulb.ac.be/popmusic/, respectively).

Maintenance of zebrafish lines

Zebrafish (*Danio rerio*) were raised and maintained under standard conditions as previously described (24). All experiments involving zebrafish were approved by the Institutional Animal Care and Use Committees of Chungnam National University (CNU-00393).

Generation of zebrafish mutant lines

The TALEN vectors targeting the second exon of *zc4h2* (Supplementary material, Fig. S5) were designed and constructed by ToolGen (Korea, www.toolgen.com). TALEN vectors were linearized by *Pvu*II and purified by ethanol precipitation. Subsequently, mRNAs encoding left and right TALENs were synthesized using the mMESSAGE mMACHINE T7 Transcription kit (Ambion catalogue number AM1344) and purified by the phenol/chloroform precipitation. One-cell fertilized embryos were mincroinjected with TALEN mRNAs. Site-specific TALEN function was examined by isolation of genomic DNA, target-specific PCR and T7 endonuclease I assay. Genomic DNA was extracted from single larvae or fin-clipped tissues of adults and lysed with genomic DNA isolation buffer (10 mM Tris, 50 mM EDTA, 200 mM NaCl, 0.5% SDS and 0.5 mg/ml Proteinase K) for 12 h at 55°C. Genomic DNA was purified using the phenol/chloroform method. PCR was performed using the following primer pairs and PCR conditions: *zc4h2* exon2 forward primer, 5′-GACCGAGTTTGAAGTCTGG-3′; *zc4h2* exon2 reverse primer, 5′-CACGTTAATATCCGCATGAATG-3′. Consequently, 130 bp PCR products were purified by gel extraction and analyzed by T7 endonuclease I assay (NEB catalogue number M0302).

Analysis of zebrafish mutants

Plasmid construction

The zebrafish *zc4h2* gene was isolated from a 24 hpf zebrafish cDNA library and first cloned into the pGEM-T Easy Vector System (Promega catalogue number A1360) and then subcloned into the *Eco*RI site in the pCS2+ expression vector. Primers used for this PCR were: *zc4h2* forward primer 5′-ATGGCGGCGAGCAAGAGATAA-3′, *zc4h2* reverse primer 5′-TGTTACTCGTCAGGTTTTCTCTTG-3′. To construct a Flag-tagged *zc4h2* expression vector, PCR was performed using the same primers and the product was subcloned into the *Nco*I site in the pCS2 + Flag vector.

Microinjection

For mutant rescue experiments, synthetic capped mRNAs for zebrafish *zc4h2* and human wild-type, L66H, R213W and R18K *ZC4H2* forms were transcribed via the mMESSAGE mMACHINE SP6 Transcription kit (Ambion catalogue number AM1340) using the linearized plasmid DNA as a template. To confirm subcellular localization in zebrafish cells, plasmid DNA for the *zc4h2-Flag* construct was injected into zebrafish embryos. mRNAs and plasmid DNA were dissolved in 0.2 M KCl with 0.2% Phenol Red as a tracking dye, and then microinjected into one to two-cell stage embryos using a PV820 Pneumatic PicoPump (WPI).

Whole-mount in situ hybridization and immunohistochemistry

Antisense digoxigenin-labeled RNA probes for *zc4h2*, *gad1*, *glyt2a*, *vglut2.1*, *gata2*, *gata3*, *isl1*, *scn1Lab*, *vsx2*, *vsx1*, *dbx2*, *nkx6.1* and

dlx2a were synthesized using a DIG and Fluorescein RNA Labeling Mix (Roche catalogue number 11277073910, 11685619910) according to the manufacturer's protocols. Whole-mount *in situ* hybridization was carried out as published (25). Whole-mount immunohistochemistry for znp1 was performed as described (24).

Paraffin sectioning

For paraffin embedding, embryos were dehydrated with ethanol for 5 min and washed for 10 min in isopropanol. They were then cleared three times for 10 min in Xylene. Embryos were transferred to Xylene:paraffin (1:1) solution at 60°C for 20 min and then embedded three times for 20 min with paraffin in stainless steel molds. Embryos were held in position in the pre-warmed plastic embedding cassette. After trimming, sections (7 μm) were cut by a microtome and deparaffinated with Xylene. The preparations were mounted with Canada Balsam and examined using a LEICA DM5000B microscope.

Behavior analysis

For observation of free swimming, 5 dpf wild-type sibling and *zc4h2* mutant larvae were placed on slide glass in embryo media. Larvae were monitored using a stereomicroscope (LEICA, MZ16), camera (LEICA DC300FX) and microscope imaging software (LEICA, IM50). For recording of free swimming, monitor display was recorded by using Camtasia Studio software (TechSmith, Vesion 7.0.0). For observation of eye, pectoral fin and jaw movement, 5 dpf larvae were transferred onto a glass slide coated with 3% methylcellulose.

Statistical analysis

To assess the significance of differences between wild-type sibling and mutant groups, all data were analyzed using one-way ANOVA with Dunnett's post-test. The significance level was set at $^{**}P < 0.001$, $^{*}P < 0.05$ versus control group and data were represented as the means ± SEM (Standard Error of the Means).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We would like to thank the families, especially those enrolled in the Greenwood Genetic Center X-linked ID project, who have been patient with us while we searched for the genetic cause for 20 years. Special thanks to the Raymond C. Phillips Unit, State of Florida and Heather Stalker, M.Sc. GGC. Dedicated to the memory of Ethan Francis Schwartz (1996–1998).

Conflict of Interest statement. None declared.

Funding

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government Ministry of Science, ICT and Future Planning (MSIP) (2014R1A2A1A11053562 to C.H.K.), National Institute of Neurological Disorders and Stroke (NINDS) (R01NS073854 to C.E.S.), and the South Carolina Department of Disabilities and Special Needs (SCDDSN) (2015-45 to C.E.S.). Funding to pay the Open Access publication charges for this article was provided by the Center for Translational Molecular Medicine (CTMM), Erasmus MC, University Medical Center Rotterdam, The Netherlands.

References

1. Miles, J.H. and Carpenter, N.J. (1991) Unique X-linked mental retardation syndrome with fingertip arches and contractures linked to Xq21.31. *Am. J. Med. Genet.*, **38**, 215–223.
2. Tarpey, P.S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., O'Meara, S., Latimer, C., Dicks, E., Menzies, A. *et al.* (2009) A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat. Genet.*, **41**, 535–543.
3. Hirata, H., Nanda, I., van Riesen, A., McMichael, G., Hu, H., Hambrock, M., Papon, M.A., Fischer, U., Marouillat, S., Ding, C. *et al.* (2013) ZC4H2 mutations are associated with arthrogryposis multiplex congenita and intellectual disability through impairment of central and peripheral synaptic plasticity. *Am. J. Hum. Genet.*, **92**, 681–695.
4. Hutchinson, S.A. and Eisen, J.S. (2006) Islet1 and Islet2 have equivalent abilities to promote motoneuron formation and to specify motoneuron subtype identity. *Development*, **133**, 2137–2147.
5. Fox, M.A. and Sanes, J.R. (2007) Synaptotagmin I and II are present in distinct subsets of central synapses. *J. Comp. Neurol.*, **503**, 280–296.
6. Kimura, Y., Okamura, Y. and Higashijima, S. (2006) alx, a zebrafish homolog of Chx10, marks ipsilateral descending excitatory interneurons that participate in the regulation of spinal locomotor circuits. *J. Neurosci.*, **26**, 5684–5697.
7. Kimura, Y., Satou, C. and Higashijima, S. (2008) V2a and V2b neurons are generated by the final divisions of pair-producing progenitors in the zebrafish spinal cord. *Development*, **135**, 3001–3005.
8. Stone, D. and Rosenthal, A. (2000) Achieving neuronal patterning by repression. *Nat. Neurosci.*, **3**, 967–969.
9. Lundfald, L., Restrepo, C.E., Butt, S.J., Peng, C.Y., Droho, S., Endo, T., Zeilhofer, H.U., Sharma, K. and Kiehn, O. (2007) Phenotype of V2-derived interneurons and their relationship to the axon guidance molecule EphA4 in the developing mouse spinal cord. *Eur. J. Neurosci.*, **26**, 2989–3002.
10. Butler, A.B. and Hodos, W. (2005) *Comparative Vertebrate Neuroanatomy: Evolution and Adaptation*, 2nd edn. John Wiley & Sons, Inc., Hoboken, NJ.
11. Baraban, S.C., Dinday, M.T. and Hortopan, G.A. (2013) Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. *Nat. Commun.*, **4**, 2410.
12. Lombard, Z., Park, C., Makova, K.D. and Ramsay, M. (2011) A computational approach to candidate gene prioritization for X-linked mental retardation using annotation-based binary filtering and motif-based linear discriminatory analysis. *Biology Direct*, **6**, 30.
13. Wieacker, P., Wolff, G., Wienker, T.F. and Sauer, M. (1985) A new X-linked syndrome with muscle atrophy, congenital contractures, and oculomotor apraxia. *Am. J. Med. Genet.*, **20**, 597–606.
14. Kok, F.O., Shin, M., Ni, C.W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I. *et al.* (2015) Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell*, **32**, 97–108.
15. Takano, K., Liu, D., Tarpey, P., Gallant, E., Lam, A., Witham, S., Alexov, E., Chaubey, A., Stevenson, R.E., Schwartz, C.E. *et al.* (2012) An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity. *Hum. Mol. Genet.*, **21**, 4497–4507.

16. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G. *et al.* (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, **327**, 78–81.
17. Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, **5**, 725–738.
18. Case, D.A., Cheatham, T.E. III, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
19. Brooks, B.R., Brooks, C.L. III, Mackerell, A.D. Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
20. Kouranov, A., Xie, L., de la Cruz, J., Chen, L., Westbrook, J., Bourne, P.E. and Berman, H.M. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.*, **34**, D302–D305.
21. Zhang, Z., Wang, L., Gao, Y., Zhang, J., Zhenirovskyy, M. and Alexov, E. (2012) Predicting folding free energy changes upon single point mutations. *Bioinformatics*, **28**, 664–671.
22. Witham, S., Takano, K., Schwartz, C. and Alexov, E. (2011) A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins*, **79**, 2444–2454.
23. Zhang, Z., Teng, S., Wang, L., Schwartz, C.E. and Alexov, E. (2010) Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Hum. Mutat.*, **31**, 1043–1049.
24. Westerfield, M. (2000) *The Zebrafish Book. A Guide for the Laboratory use of Zebrafish (Danio Rerio)*. 4th edn. University of Oregon Press, Eugene.
25. Thisse, C. and Thisse, B. (2008) High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.*, **3**, 59–69.