Cut-like homeobox 1 and nuclear factor I/B mediate *ENGRAILED2* autism spectrum disorder-associated haplotype function

Jiyeon Choi[1], Myka R. Ababon[1], Paul G. Matteson[1] and James H. Millonig[1,2,]*

[1]Center for Advanced Biotechnology and Medicine and [2]Department of Neuroscience and Cell Biology,
UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

Both common and rare variants contribute to autism spectrum disorder (ASD) risk, but few variants have been established as functional. Previously we demonstrated that an intronic haplotype (*rs1861972–rs1861973* A–C) in the homeobox transcription factor *ENGRAILED2* (*EN2*) is significantly associated with ASD. Positive association has also been reported in six additional data sets, suggesting *EN2* is an ASD susceptibility gene. Additional support for this possibility requires identification of functional variants that affect *EN2* regulation or activity. In this study, we demonstrate that the A–C haplotype is a transcriptional activator. Luciferase (luc) assays in mouse neuronal cultures determined that the A–C haplotype increases expression levels (50%, *P* < 0.01, 24 h; 250%, *P* < 0.0001, 72 h). Mutational analysis indicates that the A–C haplotype activator function requires both associated A and C alleles. A minimal 202-bp element is sufficient for function and also specifically binds a protein complex. Mass spectrometry identified these proteins as the transcription factors, Cut-like homeobox 1 (Cux1) and nuclear factor I/B (Nfib). Subsequent antibody supershifts and chromatin immunoprecipitations demonstrated that human CUX1 and NFIB bind the A–C haplotype. Co-transfection and knock-down experiments determined that both CUX1 and NFIB are required for the A–C haplotype activator function. These data demonstrate that the ASD-associated A–C haplotype is a transcriptional activator, and both CUX1 and NFIB mediate this activity. These results provide biochemical evidence that the ASD-associated A–C haplotype is functional, further supporting *EN2* as an ASD susceptibility gene.

INTRODUCTION

Autism spectrum disorder (ASD) is a common human neurodevelopmental disorder with an incidence of ∼1 in 110. It includes a range of phenotypes. Autism is the most severe form, whereas individuals with Asperger's syndrome and pervasive developmental disorder-not otherwise specified have less severe phenotypes. Core symptoms of ASD include deficits in social interaction, impairments in verbal and non-verbal communication as well as stereotypic and repetitive behaviors and interests. Twin and family studies indicate that ASD has strong genetic basis. However, genetic risk likely involves both common and rare variants in multiple genes. Numerous genes have been associated with ASD but few studies have determined whether these associated alleles are functional (1–7).

Our research has focused on the homeobox transcription factor, *ENGRAILED2* (*EN2*). Animal studies have determined that *En2* is expressed in the midbrain and hindbrain throughout development and regulates multiple developmental processes relevant to ASD (8–14). Numerous *in vitro* and *in vivo* analyses have demonstrated that *En2* regulates brain connectivity which is implicated in ASD (15). Both the *En2* knock-out and an *En2* over-expression transgenic mice result in the improper mapping of cerebellar mossy fibers (13,16). In addition, ∼5% of En2 protein is secreted and forms a rostral–caudal extracellular gradient in the tectum (17,18). Inhibition of this extracellular form results in abnormal targeting of retinal axons to the tectum (19). *En2* knock-out mice also display a disruption of excitatory/inhibitory (E/I) circuit balance, and converging evidence suggests that a defect in E/I balance may contribute to ASD etiology (20,21). Finally, *En2* is expressed in the

*To whom correspondence should be addressed at: Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854-5638, USA. Tel: +1 7322353391; Fax: +1 7322355562; Email: millonig@cabm.rutgers.edu

developing locus coeruleus and raphe nuclei of the ventral mid-hindbrain and is required for norepinephrine and serotonin neurotransmitter system development (14). Abnormal norepinephrine and serotonin levels have also been associated with ASD (22–25).

Our previous association analysis determined that *EN2* is significantly associated with ASD (26,27). The common alleles (underlined) of two intronic single-nucleotide polymorphisms (SNPs), *rs1861972* (A/G) and *rs1861973* (C/T), are over-transmitted to individuals with ASD [Associated alleles are underlined]. The minor haplotype (*rs1861972– rs1861973* G–T) is over-represented in unaffected siblings. Significant association for each individual SNP as well as *rs1861972–rs1861973* A–C haplotype was first observed in 167 Autism Genetic Resource Exchange pedigrees (27) and then independently replicated in two additional data sets (three data sets, 518 families; $P = 0.00000035$) (26). Six other groups have also reported *EN2* association with ASD (28–33). These data suggest that the A–C haplotype is segregating with a DNA variant that increases risk for ASD.

To identify common risk alleles segregating with the A–C haplotype, the following criteria were applied. We expected candidates to display high r^2 with both *rs1861972* and *rs1861973* and exhibit significant association with ASD. Risk alleles should also be functional, affecting the activity or expression of *EN2*. Re-sequencing, linkage disequilibrium (LD) mapping and association analysis determined that the *rs1861972–rs1861973* A–C haplotype was the best candidate to test for function (34). Bioinformatics determined that the *rs1861972-* and *rs1861973*-associated alleles are situated in transcription factor consensus sequences.

We now extend these findings using a series of molecular genetics and biochemical approaches. Luciferase (luc) assays are performed at two different stages of neuronal development. We also determine whether both associated alleles are necessary for function and delimit the minimal DNA element sufficient for transcriptional activity. An unbiased proteomic approach identifies two transcription factors that specifically bind the A–C haplotype. We then validated these factors by supershift, co-transfection and knock-down analyses. These functional data for the ASD-associated A–C haplotype provide biochemical evidence that *EN2* contributes to ASD risk.

RESULTS

ASD-associated *rs1861972–rs1861973* A–C haplotype increases gene expression

To characterize the ASD-associated *rs1861972–rs1861973* A–C haplotype as a *cis*-regulatory element, we performed luc assays on a series of constructs. These experiments were conducted in mouse cerebellar granule neurons derived from postnatal day 6 (P6) C57BL6/J pups. Cerebellar granule cells are the most abundant neuronal cell type in the brain and because of their small size they can be isolated to near-homogeneity. At P6, *En2* is expressed exclusively in differentiating granule cells. In culture, granule cells exit cell cycle and start to differentiate by ~24 h. By 72 h, the neurons are more differentiated, with a greater number of cells displaying longer neuritic processes (35).
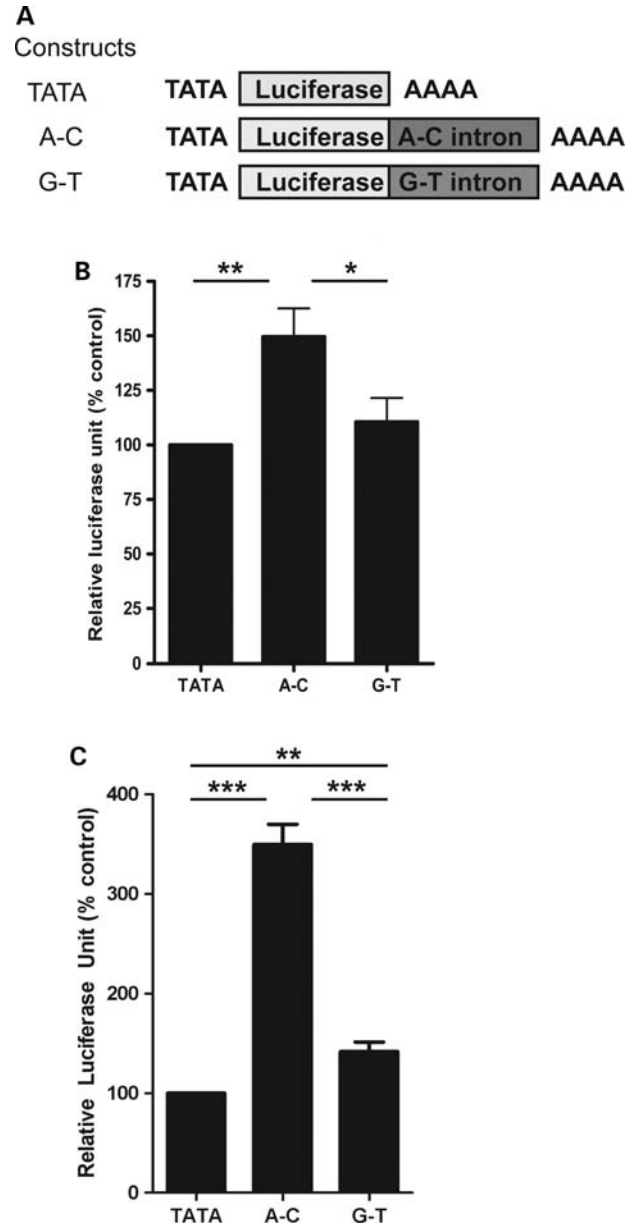
Figure 1. ASD-associated *rs1861972–rs1861973* A–C haplotype increases gene expression. (**A**) Luciferase (luc) constructs used for transfections are diagramed: TATA—pGL3pro vector driven by SV40 minimal promoter; A–C and G–T—pGL3pro vector containing full-length human *EN2* intron with ASD-associated A–C haplotype (A–C) or unassociated G–T haplotype (G–T). The intron was cloned 3′ of luc gene and 5′ of poly A signal, so it is transcribed and spliced as the endogenous gene. (**B**, **C**) Equimolar amount of the three constructs were transiently transfected into P6 mouse cerebellar granule neurons and cultured for 24 h (B) or 72 h (C). Luciferase activities were then measured and normalized to the levels of *Renilla reniformis*. Relative luc units are expressed as percent of TATA control. Note the A–C haplotype function is more pronounced at 72 h. $n = 7$ (B), $n = 10$ (C). *$P < 0.05$, **$P < 0.01$, ***$P < 0.0001$; two-tailed paired Student's *t*-test.

We first tested the luc activities of the full-length A–C and G–T intronic constructs and compared their activity to the intron-less TATA (SV40 minimal promoter containing TATA box sequence) promoter control (Fig. 1A). Equimolar amounts (36) of each construct were electroporated into
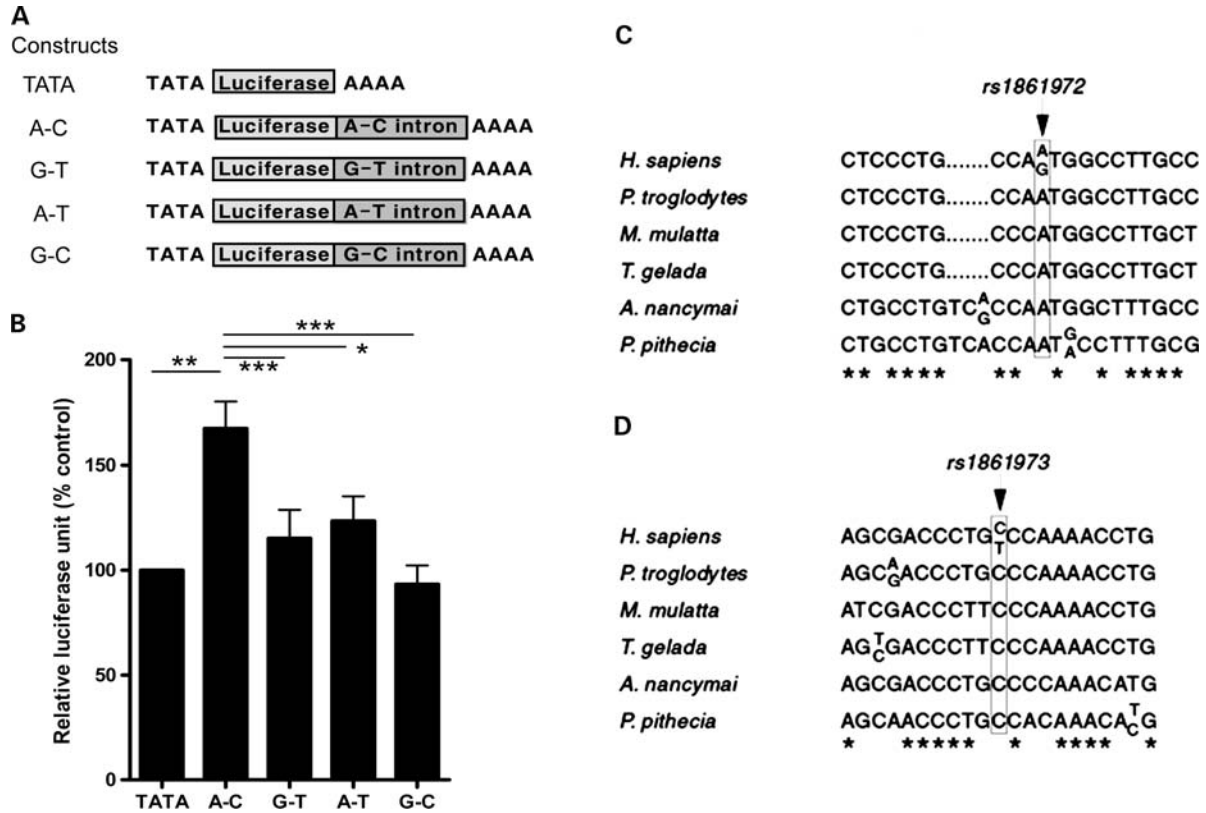
Figure 2. The A–C haplotype function requires both the ASD-associated A and C alleles. (**A**) Luc constructs used for transfections are diagrammed: TATA—pGL3pro vector driven by SV40 minimal promoter; A–C and G–T—pGL3pro vector containing full-length human *EN2* intron with ASD-associated A–C haplotype (A–C) or unassociated G–T haplotype (G–T) as described in Fig. 1; A–T and G–C—pGL3pro vector containing full-length human *EN2* intron with one ASD-associated (A or C) and one unassociated allele (T or G). (**B**) Constructs were transfected and analyzed as previously described. Luc activities were measured after 24 h in culture. Only the ASD-associated A–C haplotype increases gene expression above the TATA control and is significantly different from all the other haplotypes. $n = 6$, $*P < 0.05$, $** P < 0.01$, $*** P < 0.0001$; two-tailed paired Student's *t*-test. (**C, D**) *EN2* intron encompassing *rs1861972* and *rs1861973* was sequenced and compared from five primate species: *P. pithecia*, *A. nancymai*, *T. gelada*, *P. troglodytes* and *M. mulatta.* Sequence alignment was performed for 10-bp 5′ and 10-bp 3′ of *rs1861972* (C) and *rs1861973* (D). Arrows indicate position of *rs1861972* and *rs1861973* in humans and primate species. Perfect matches between all the species are marked with asterisks at the bottom. For heterozygous loci, both alleles are shown. Note that the ASD-associated A–C haplotype is conserved in primate species and the G–T haplotype is only observed in humans. Neither A–T nor G–C haplotype pairs are observed in our analysis. These data suggest that the A–C haplotype evolved together.

primary granule cell cultures, which were grown for 24 h before luc activities were measured. When luc levels were compared with the control, the A–C haplotype resulted in a 50% increase. Luc levels for the A–C haplotype were also significantly higher than the G–T haplotype. The G–T haplotype displayed no significant difference from the intron-less control (Fig. 1B).

We then repeated the same transfections, but allowed the granule cells to differentiate for 72 h. At this time point, the A–C haplotype increased luc levels 250% above the control, whereas the G–T haplotype displayed a less pronounced 41% increase (Fig. 1C). In addition, the difference between the A–C and G–T haplotypes is more significant at 72 h ($P = 0.000001$) than at 24 h ($P = 0.027$) (Fig. 1B, C). These results demonstrate activator function for the A–C haplotype.

A–C haplotype function requires both ASD-associated alleles

Next we examined whether the A–C haplotype function requires both ASD-associated alleles. To investigate this question, two additional mutant constructs were generated that contained only one of the ASD-associated alleles (A–T or G–C haplotype) (Fig. 2A). Transfections were performed as described above and luc levels were measured after 24 h. Consistent with the prior results, the A–C haplotype increased gene expression by 66%. The A–T and G–C haplotypes were not significantly different from the G–T haplotype or from the control and were significantly less active than the A–C haplotype (Fig. 2B). These data demonstrate that both the ASD-associated alleles are required for the A–C haplotype function.

ASD-associated A–C haplotype is conserved in primate species

As both ASD-associated alleles are required for function, we investigated whether the A–C haplotype is conserved during evolution. Vertebrate Multiz Alignment (UCSC Genome Browser, http://genome.ucsc.edu/) determined that the A–C haplotype is not conserved in mouse, rat or chicken, so we turned our attention to primate species. The *EN2* intron was cloned and sequenced from two new world monkeys (*Aotus nancymai* and *Pithecia pithecia*: owl and saki monkeys), two
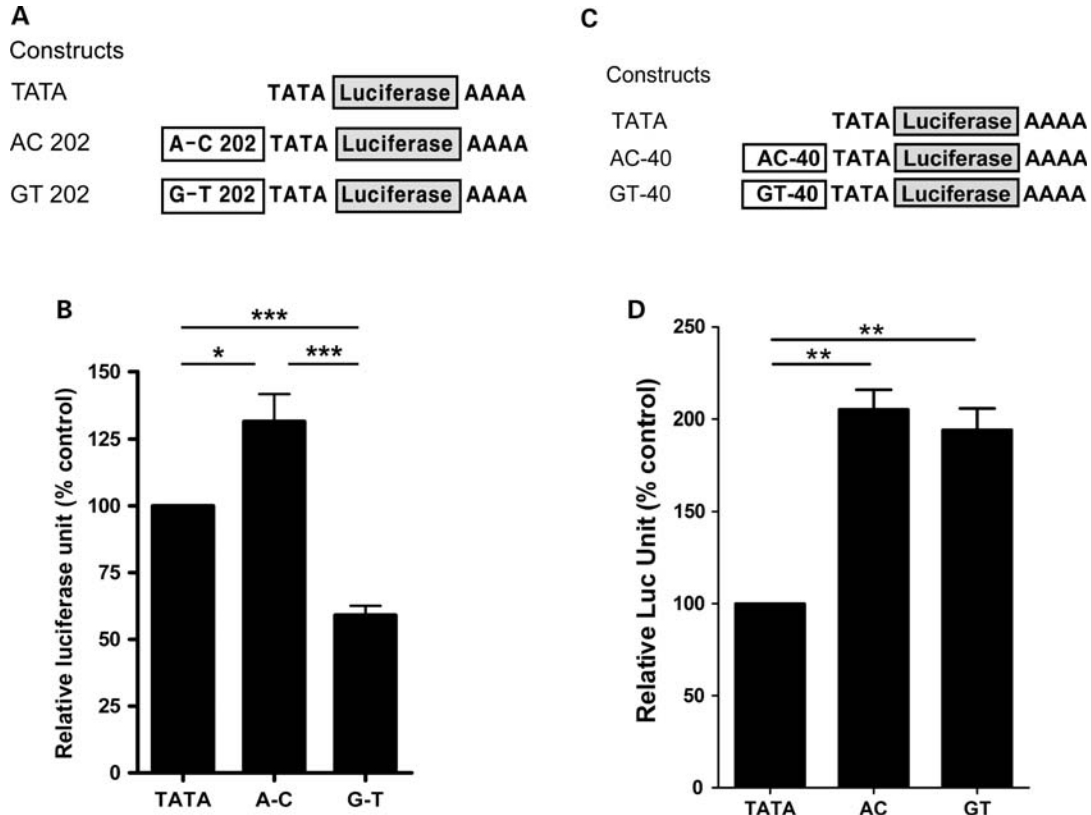
Figure 3. 202 bp encompassing *rs1861972–rs1861973* is sufficient for the A–C haplotype function. (**A**) Luc constructs used for transfections in (**B**) are diagramed: TATA—pGL3pro vector driven by SV40 minimal promoter; A–C 202 and G–T 202—pGL3pro vector containing 202 bp encompassing *rs1861972–rs1861973* with A–C (A–C 202) or G–T (G–T 202) haplotype cloned 5′ of SV40 promoter. (**B**) Constructs were transfected and analyzed as previously described. Luc activities were measured after 24 h in culture. Two-hundred and two base pairs are sufficient to recapitulate the full-length A–C haplotype activator function. $n = 9$, *$P < 0.05$, ***$P < 0.0001$; two-tailed paired Student's *t*-test. (**C**) Luc constructs used for transfections in (**D**) are diagramed: TATA—pGL3pro vector driven by SV40 minimal promoter; AC-40 and GT-40—pGL3pro vector containing 20-bp oligos for *rs1861972*-A and *rs1861973*-C or G and T, which are adjoined 5′ to 3′ and cloned 5′ of SV40 promoter. (**D**) Constructs were transfected and analyzed as previously described. Luc activities were measured after 24 h in culture. Both oligomers result in increased gene expression but failed to show significant differences between ASD-associated and unassociated haplotypes. $n = 4$, **$P < 0.01$; two-tailed paired Student's *t*-test.

old world monkeys (*Macaca mulatta* and *Theropithecus gelada*: rhesus and baboon) and one hominoid species (*Pan troglodytes*: chimpanzee). Sequence comparisons for *rs1861972* and *rs1861973* indicated that the A and C alleles are conserved in all examined primate species (Fig. 2C and D). The G and T alleles are observed only in humans. These data are consistent with the ASD-associated A and C alleles being ancestral. The absence of individuals with an A–T or G–C haplotype also suggests that the A and C alleles have segregated together during evolution and the G and T alleles arose at a similar time. This possibility is consistent with the strong LD observed in the human population ($r^2 = 0.767$ in the combined data set) (26).

Minimal *cis*-element sufficient for the A–C haplotype function

Next, the minimal DNA sequence required for the *cis*-regulatory function of the A–C haplotype was determined, so it could be used as a bait to isolate the protein mediators. In addition, we investigated whether the mechanism is transcriptional because previous data indicated that splicing was not affected significantly (34). *Rs1861972* and *rs1861973* are

situated 150 bp apart from each other in the *EN2* intron. A 202-bp fragment encompassing either the A–C or G–T haplotype was cloned 5′ of the minimal promoter to test transcriptional function, and luc assays were performed as described above (Fig. 3A). At 24 h, the A–C haplotype increased luc levels 31% above the control and the G–T haplotype decreased luc levels 41% below the control. The A–C and the G–T haplotypes were significantly different from each other and the promoter control (Fig. 3B). These results demonstrate that 202 bp is sufficient to recapitulate the A–C haplotype function in two ways: first, the A–C haplotype functions as an activator of gene expression; second, the ASD-associated A–C haplotype is significantly more active than the G–T haplotype. As the 202-bp sequence is not transcribed in these constructs, the A–C haplotype function is likely through a transcriptional mechanism.

We then investigated whether the A–C haplotype could be further broken down into smaller units. This question was addressed by generating 40-bp oligomer constructs of *rs1861972* and *rs1861973* (Fig. 3C). Twenty base-pair region around each allele was chosen based on the common length of the transcription factor binding sites being ∼6–14 bp. Forty-mers with the A–C or G–T haplotype
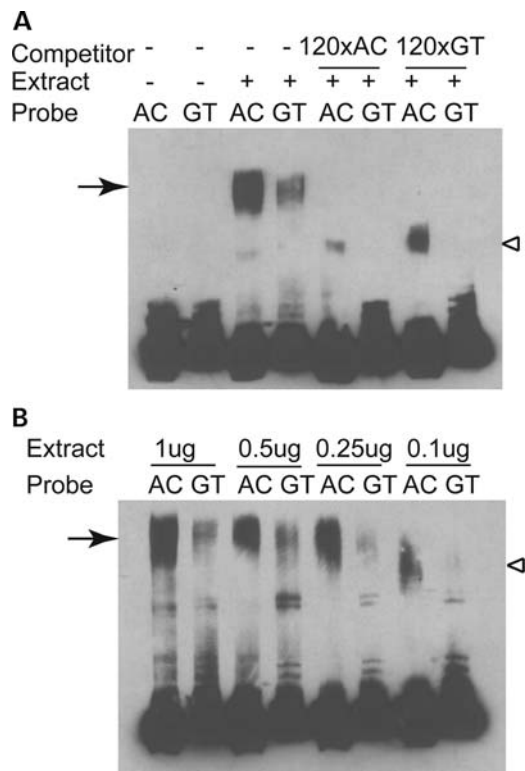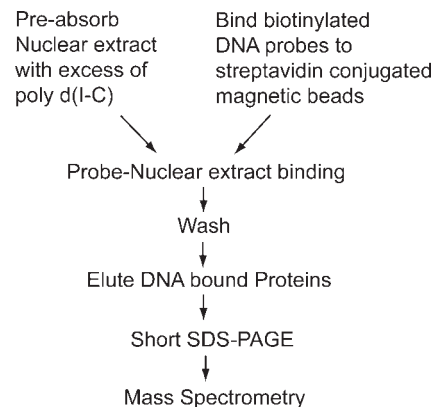
Figure 4. ASD-associated A−C haplotype specifically binds a protein complex. EMSAs were performed using 202-bp probes for the ASD-associated A−C haplotype (AC probe) or unassociated G−T haplotype (GT probe) and nuclear extract from differentiated P6 mouse cerebellar granule neurons. (**A**) Nuclear extract was incubated with biotinylated A−C or G−T probes and run on 5% non-denaturing polyacrylamide gel. Arrow indicates specific binding of the A−C haplotype. For competition analysis, 120 molar excess of unlabeled competitors (120× AC and 120× GT) were added into the reactions. Note that excess competitors not only reduced the probe–protein binding in sequence-specific manner, but also chased the protein–DNA shift to a faster migrating complex (arrowhead). (**B**) Protein titration was carried out using decreasing amount of nuclear extract. Lower protein/probe ratio results in a faster migrating complex similar to the one observed in A (arrowhead).

Figure 5. Proteins specifically binding to the A−C haplotype were identified by affinity purification and mass spectrometry. Protein identification procedure is diagramed. Nuclear extract from differentiated P6 cerebellar granule neurons was pre-absorbed with poly d(I-C), whereas the 202-bp biotinylated DNA probes with A−C, G−T or lambda sequences were bound to streptavidin-conjugated magnetic beads. Nuclear extract and probes were then incubated together in EMSA binding buffer and thoroughly washed. Proteins bound to the probes were eluted and briefly separated on an SDS−PAGE gel. Proteins were excised from the gel and digested with trypsin. Extracted peptides were subjected to LC-MS/MS.

were cloned 5′ of the minimal promoter, and luc assays were performed. At 24 h, both the A−C 40mer and G−T 40mer increased luc levels 105 and 95%, respectively, above the control. However, the A−C and G−T 40mers did not display a significant difference from each other (Fig. 3D). These results demonstrate that 40mer sequences encompassing *rs1861972* and *rs1861973* do not recapitulate the A−C haplotype function. Therefore, we conclude that the 202 bp encompassing *rs1861972* and *rs1861973* is the minimal element necessary and sufficient for the A−C haplotype transcriptional function. This piece of DNA was then used as a bait to identify the proteins that mediate the A−C haplotype function.

A−C haplotype specifically binds a protein complex

To determine whether nuclear proteins specifically bind the A−C haplotype, electrophoretic mobility shift assays (EMSAs) were performed. Cerebellar granule cell nuclear extract was incubated with the 202-bp DNA probes for the A−C or G−T haplotype. A shifted band was detected for the A−C probe, and the same DNA−protein complex was

also observed for the G−T haplotype but at significantly less intensity (Fig. 4A, arrow). For both haplotypes, protein binding was specifically competed by adding 120 molar excess of unlabeled probe. For the A−C haplotype, self-sequence competed better than the G−T haplotype (Fig. 4A, arrowhead, lane 5 versus lane 7), further demonstrating sequence-specific binding.

Competition with unlabeled probes resulted in downward shift of the complex to a faster migrating band (Fig. 4A, arrowhead). This observation led us to speculate that the A−C haplotype probe binds multiple proteins. We reasoned that excessive unlabeled probe resulted in depletion of less abundant protein members, which resulted in some of the complexes to migrate faster. To further investigate this possibility, a protein titration assay was conducted, and similar faster-migrating bands were observed at lower protein concentrations (Fig. 4B, arrowhead). These results determine that a protein complex specifically binds to the A−C haplotype.

Mass spectrometry identified transcription factors specifically binding to the A−C haplotype

To identify members of this protein complex, affinity purification of DNA-bound proteins followed by mass spectrometry was employed (Fig. 5). The 202-bp minimal A−C haplotype sequence and granule cell nuclear extract were used for these experiments. To identify proteins specifically binding to the A−C haplotype, while avoiding false-positives, two control probes were employed. Specifically, the G−T haplotype (202 bp) and a biologically unrelated lambda DNA (∼200 bp)-bound fractions were compared with the A−C (202 bp) haplotype. The method of spectral counting was used to measure the relative abundance of a given protein in different probe-bound fractions (37−40). Briefly, spectral counts for each protein were extracted from mass spectrometry

Table 1. Statistical analysis of transcription factors identified by mass spectrometry

DNA probe Significance	202 bp spanning the A−C versus G−T			202 bp spanning the A−C versus lambda DNA		
	Ratio, A−C/G−T[a] (lower[b] to upper[c] limit)	One-sided, *P*-value for 1-fold difference[d]	Odds ratio[e]	Ratio, A−C/lambda[a] (lower[b] to upper[c] limit)	One-sided, *P*-value for 1-fold difference[d]	Odds ratio[e]
Cux1	5.29 (2.41−11.62)	2.65E-6	1.73	2.85 (1.53−5.3)	0.00047	1.96
Nfib	4.33 (1.33−14.17)	0.0106	1.42	3.25 (1.12−9.47)	0.0245	2.24

[a]Ratio of spectral counts, taken from two independent runs, for a given protein (A−C-bound/G−T-bound fraction or A−C-bound/lambda-bound fraction).
[b]Lower limit of 95% confidence interval for the A−C/G−T or A−C/lambda ratio by Wilson's test.
[c]Upper limit of 95% confidence interval for the A−C/G−T or A−C/lambda ratio by Wilson's test.
[d]One-sided *P*-value for the A−C bound is larger than the negative controls.
[e]Odds ratios are calculated as (spectral counts of a given protein in A−C-bound fraction/total spectral counts of A−C-bound fraction)/(spectral counts of a given protein in G−T-bound fraction/total spectral counts of G−T-bound fraction). Odds ratios for A−C versus lambda were calculated in the same manner.

data and compared between the A−C-bound fraction and the two negative controls using a series of statistical analyses. Candidate proteins were identified based on the following three statistical criteria: (1) the ratio of spectral counts between the A−C and negative control is significantly enriched by the Wilson's test (41), i.e. the lower 95% confidence index is >1; (2) the ratio between the A−C-bound spectral count and negative control is significantly different from 1 (one-sided *P*-value <0.05); and (3) the above two criteria are met for both A−C/G−T and A−C/lambda comparisons. After applying these criteria, seven nuclear proteins were identified (Supplementary Material, Table S1). Among them, two are transcription factors, Cux1 and nuclear factor I/B (Nfib; Table 1). These results demonstrate that Cux1 binds the A−C probe ∼5.29-fold better than the G−T and ∼2.85-fold better than the lambda. Nfib binds the A−C probe ∼4.33-fold better than the G−T and ∼3.25-fold better than the lambda. Interestingly, recognition sequences for both Cux1 and Nfib include CCAAT motifs.

The protein identification procedure was then repeated using different baits before we moved on for validation. As 202 bp is required for the A−C haplotype function, we questioned whether the same proteins could be identified when 20-bp probes encompassing *rs1861972*-A or *rs1861973*-C were used as baits. As negative controls, each reaction was paired with nuclear extracts pre-absorbed with 300 molar excess of unlabeled probe before affinity purification. The same statistical criteria were applied to identify proteins more abundant in A- and C-bound fractions compared with the negative controls. Interestingly, Cux1 and Nfib were not identified, whereas the five other nuclear proteins were detected by either A or C probe (Supplementary Material, Table S1). This finding suggests that only Cux1 and Nfib specifically bind to the 202-bp minimal functional element.

Next, we questioned whether *Cux1* and *Nfib* are co-expressed with *En2* by accessing bioinformatic Web sites. During central nervous system (CNS) development, mouse *Cux1* and *Nfib* transcripts are co-localized with *En2* in the mid-hindbrain junction at E10.5 (VisiGene, http://genome.ucsc.edu/cgi-bin/hgVisiGene). *Cux1* and *Nfib* are also co-expressed with *En2* in the cerebellar primordium and midbrain at E14.5 (GenePaint, http://www.genepaint.org), and in cerebellar granule cells at P6 (42,43). Finally, in the adult CNS, all three genes are restricted to the granule cell layer of cerebellum (Allen Human Brain Atlas, http://www.brain-map.org and

MGI, http://www.informatics.jax.org). Reverse transcription (RT)−polymerase chain reaction (PCR) analysis determined that *Cux1* and *Nfib* are also expressed in cultured granule neurons (data not shown). In addition, human *CUX1* and *NFIB* are co-expressed with *EN2* in the adult cerebellum (Allen Human Brain Atlas). Co-expression of *Cux1* and *Nfib* with *En2* suggests that they are *trans*-acting factors mediating *cis*-regulatory function of the A−C haplotype.

**CUX1 and NFIB bind to the A−C haplotype *in vitro***

To validate our findings from mass spectrometry, we investigated whether human CUX1 and NFIB physically bind to the A−C haplotype. To address this question, supershifts were performed using antibodies against CUX1 and NFIB, the same 202-bp probes, and nuclear extract from human embryonic kidney 293T (HEK293T) cells. We wanted to validate binding in a human cell line, and HEK293T cells were selected as a system for the following two reasons: (1) they are highly transfectable and (2) *CUX1* and *NFIB* are expressed at lower levels compared with cerebellar granule neurons. With the addition of anti-CUX1 antibody, the A−C haplotype complex shifted further upward, indicating the presence of CUX1 protein (Fig. 6A, arrow). Addition of unrelated antibody raised in the same source (rabbit polyclonal) did not result in the same shift, demonstrating specificity. The same results were observed when anti-NFI antibody was added (Fig. 6B, arrow). The G−T haplotype complex also shifted in a similar pattern albeit with lower band intensity, due to weaker binding affinity for both CUX1 and NFIB. EMSAs and supershifts for the rare A−T and G−C haplotypes resulted in either weaker binding to the same complex (A−T) or strong binding to a complex that migrated differently (G−C) (Supplementary Material, Fig. S1). Antibody supershifts for CUX1 and NFIB demonstrate their physical binding to the A−C haplotype *in vitro*, which validates the mass spectrometry findings and indicates that human CUX1 and NFIB bind the A−C haplotype.

CUX1 and NFIB bind to the A−C haplotype at the same time

From the supershift analysis, a majority of the DNA−protein complex was supershifted for both CUX1 and NFIB without much residual unshifted complex being observed. These data suggested that most of the protein−DNA complex contains
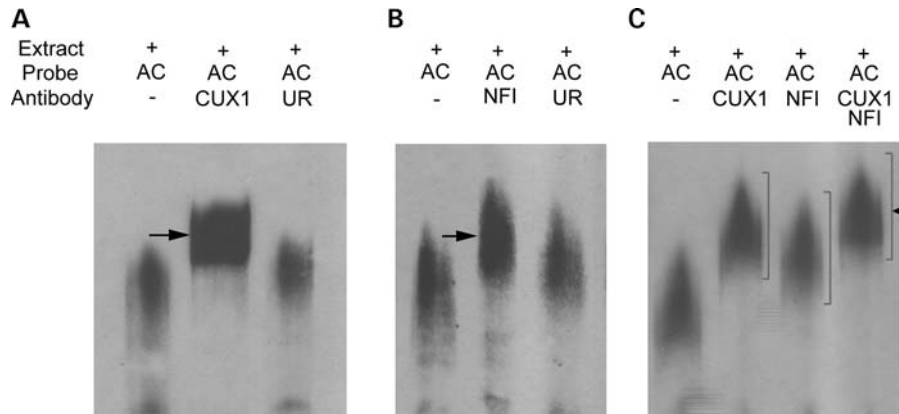
Figure 6. CUX1 and NFIB bind the A–C haplotype together. Supershifts were performed using nuclear extract from HEK293T cells and 202-bp probes encompassing *rs1861972–rs1861973* with A–C haplotype (AC). (**A, B**) Rabbit polyclonal anti-CUX1 antibody (CUX1 in A), rabbit polyclonal anti-NFI antibody (NFI in B) or unrelated rabbit polyclonal antibody (UR) was added to nuclear extract before binding to probes and separated on 4% acrylamide gel. Arrow indicates supershifted band by anti-CUX1 (A) or anti-NFI antibody (B). (**C**) Effect of anti-CUX1 (CUX), anti-NFI (NFI) antibody or of both antibodies (CUX1, NFI) was compared on the same EMSA gel. Brackets highlight the supershifted bands by anti-CUX1, anti-NFI or both antibodies together. Arrowhead indicates higher shift when both anti-CUX1 and anti-NFI antibodies were applied together.
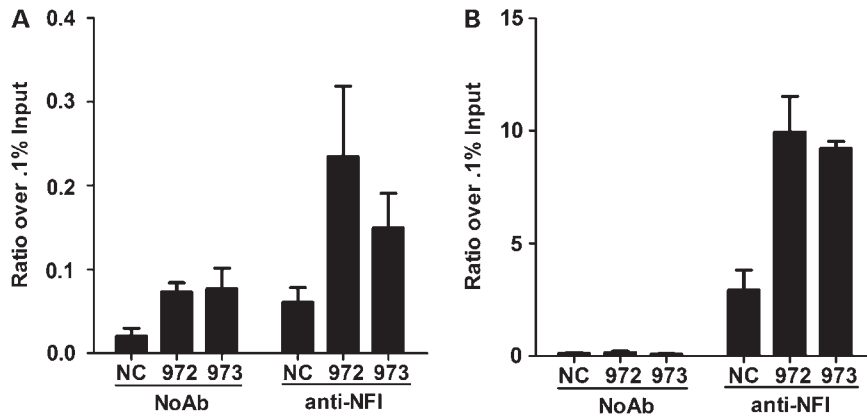
Figure 7. NFIB binds the A–C haplotype *in vivo*. ChIP assays were conducted in human in HEK293T and SH-SY5Y cells. Cells were fixed with 1% formaldehyde and sonicated before immunoprecipitation with rabbit polyclonal anti-NFI antibody (anti-NFI). Immunoprecipitation without antibody (NoAb) was used as negative control. Precipitated DNA was assayed by qPCR to detect the presence of the area encompassing *rs1861972* (972), *rs1861973* (973) or non-specific region (NC) as negative control. Relative levels of binding were normalized to 0.1% of total input DNA. Error bars represent standard error of mean measured from four independent experiments. (**A**) When NFIB was over-expressed in HEK293T cells, binding to the A–C haplotype is observed ∼3-fold higher compared with the two negative controls. (**B**) Endogenous binding of NIFIB was examined in SH-SY5Y cells. Significant enrichment to the *rs1816972* and *rs1861973* was observed compared with no antibody and non-specific region (NC) controls.

CUX1 and NFIB. To investigate whether both proteins bind to the 202 bp A–C probe at the same time, the supershift was performed with anti-CUX1 and anti-NFI antibodies in the same reaction. In the presence of the both antibodies, the complex was shifted even higher (Fig. 6C, arrowhead and brackets) than with anti-CUX1 or anti-NFI alone (Fig. 6C, brackets). These data demonstrate that both CUX1 and NFIB bind the A–C haplotype at the same time.

NFIB binds to endogenous *rs1861972–rs1861973* A–C haplotype

As CUX1 and NFIB bind to 202-bp A–C probe *in vitro*, their binding to the endogenous *rs1861972–rs1861973* haplotype was investigated next. To investigate this question, chromatin immunoprecipitation (ChIP) was conducted using an antibody against NFI in HEK293T cells as well as a medulloblastoma cell line, SH-SY5Y, which is derived from cerebellar granule cells. In HEK293T cells, endogenous binding of NFIB to *rs1861972* and *rs1861973* is ∼4-fold higher than to the negative control region but is marginal (∼1.5-fold difference) compared with no antibody control (data not shown). Indeed, when NFIB is over-expressed, its binding to the A–C haplotype is more prominent displaying ∼3-fold higher levels compared with both no antibody control and negative primer control (Fig. 7A).

When the same experiments were repeated in the human SH-SY5Y neuronal cell line, more significant results were observed. Endogenous binding of NFIB to *rs1861972* and *rs1861973* is >3-fold higher than the negative primer control, and >50-fold higher than the no antibody control (Fig. 7B). Significant results were also observed when NFIB was over-expressed (data not shown). These data demonstrate NFIB binds to endogenous *rs1861972–rs1861973* A–C haplotype.
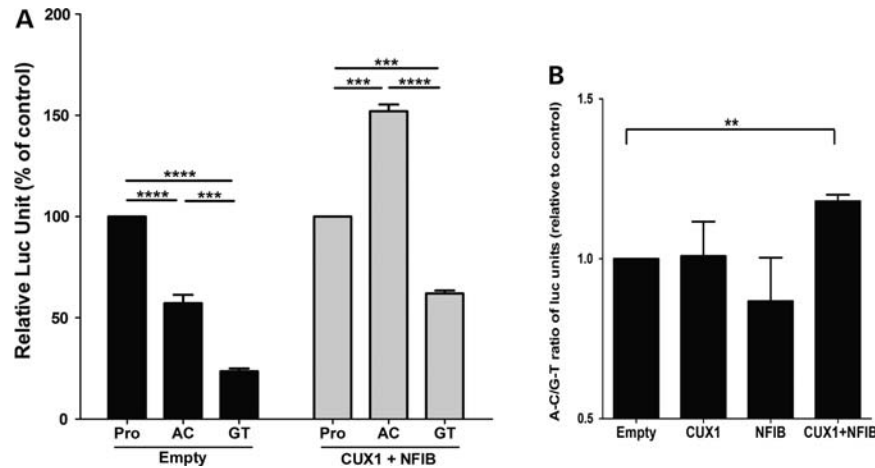
Figure 8. CUX1 and NFIB are sufficient to regulate the A−C haplotype function. Full-length intronic A−C and G−T luc constructs (Fig. 1A) were co-transfected with empty pCMV-SPORT6 vector (Empty) or CUX1 and NFIB expression constructs (CUX1 + NFIB) into HEK293T cells. Luc activities were measured and normalized to the levels of *R. reniformis*. (**A**) Relative luc units are expressed as percent of the promoter control (Pro). Note that CUX1 and NFIB expression is sufficient to convert the A−C haplotype into an activator in HEK293 cells. (**B**) To investigate whether CUX1 and NFIB affected the difference in luc activity for the A−C and G−T haplotypes, A−C/G−T luc ratios were calculated in the absence (Empty) and presence (CUX1, NFIB and CUX1 + NFIB) of over-expression. The difference between the A−C and G−T haplotype was significantly increased when both CUX1 and NFIB are over-expressed. $n = 4$, two-tailed paired Student's *t*-test, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

We also performed ChIP for CUX1. Unfortunately, the available CUX1 antibodies are not amenable for ChIP analysis. CUX1 binding was not significant compared with negative controls at either endogenous levels or when over-expressed. As CUX1 and NFIB bind to the A−C haplotype at the same time *in vitro* and NFIB binds to the same sequence by ChIP, it is likely that CUX1 also binds to the A−C haplotype *in vivo*.

CUX1 and NFIB regulate the A−C haplotype function

As human CUX1 and NFIB bind to the A−C haplotype, we then investigated whether they are sufficient to regulate A−C haplotype function. As we have repeatedly demonstrated, two signatures of the A−C haplotype function are: (1) the A−C haplotype increases gene expression; and (2) the A−C haplotype is significantly more active than the G−T haplotype. So we asked whether CUX1 and NFIB could affect these two characteristics of the A−C haplotype function when both proteins were over-expressed with the full-length A−C or G−T intronic luc constructs in HEK293T cells.

The luc constructs were first transfected alone to HEK293T cells, where *CUX1* and *NFIB* are expressed at low levels endogenously. In these experiments, the A−C haplotype displays significantly higher luc levels than the G−T haplotype. However, both the A−C and G−T haplotypes are significantly lower than the control level (Fig. 8A, empty). So we asked whether CUX1 and NFIB can convert the A−C haplotype to an activator. Indeed, when both CUX1 and NFIB are over-expressed by transfecting full-length cDNA constructs, the A−C haplotype increases luc levels 52% above the control, whereas the G−T haplotype decreases luc levels 38% below the control (Fig. 8A, CUX1 + NFIB). This observation is strikingly reminiscent of the 202-bp A−C haplotype function in granule neurons (Fig. 3B). When CUX1 or NFIB are over-expressed individually, neither protein alone was sufficient to generate the activity of the 202-bp fragment observed in

granule cells, which we used to isolate the proteins. In fact, CUX1 alone converted both A−C and G−T to activators, and NFIB alone further repressed A−C and G−T below the control level (Supplementary Material, Fig. S2). These results demonstrate that both CUX1 and NFIB are sufficient to convert the A−C haplotype to an activator without affecting the repressor activity of the G−T haplotype.

We then investigated whether CUX1 and NFIB contribute to the second signature of the A−C haplotype—a difference in activity between the A−C and G−T haplotypes. To test this idea, we measured the ratio of A−C/G−T luc levels in endogenous HEK293T cells from the above transfections and examined whether this ratio is affected by CUX1 and NFIB over-expression. Indeed, when both CUX1 and NFIB are over-expressed, A−C/G−T ratio is significantly increased (18%, $P < 0.01$) compared with the endogenous control (Fig. 8B). In other words, the difference between the A−C and G−T haplotypes becomes more pronounced when both CUX1 and NFIB are abundant. Over-expression of either CUX1 or NFIB alone does not affect the A−C/G−T ratio (Fig. 8B). Together these data demonstrate that both CUX1 and NFIB are sufficient to mediate the A−C haplotype function.

The A−C haplotype increases endogenous *EN2* levels

To investigate whether the A−C haplotype affects endogenous *EN2* expression, mRNA levels were measured in five human cell lines (SK-N-MC, HEK293T, SH-SY5Y, DAOY and PFSK-1) that are categorized by ATCC as expressing neuronal genes. The cell line *rs1861972−rs1861973* genotype was determined. Two lines are homozygous for the A−C haplotype (AC/AC: SK-N-MC and HEK293T), whereas the other three are heterozygous (AC/GT: SH-SY5Y, DAOY and PFSK-1). qRT-PCR measured endogenous *EN2* levels and a ∼111-fold higher expression was observed in the AC/AC compared with AC/GT cell lines ($n = 3$, $P = 0.0021$) (data not shown).
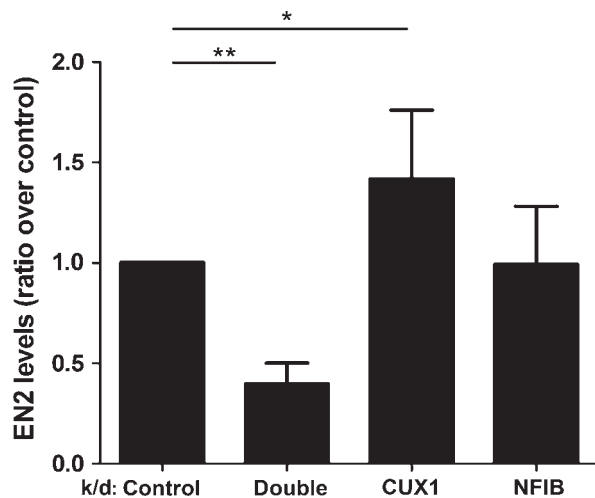
Figure 9. Endogenous *EN2* levels are affected by the CUX1-NFIB double knock-down. To further investigate the regulation of endogenous human *EN2* expression by CUX1 and NFIB, stable single (CUX1, NFIB) or double (CUX1 + NFIB) knock-down (k/d) cell lines were established in HEK293T cells. A non-silencing control was also generated. Endogenous *EN2* mRNA levels were measured by Taqman qRT-PCR for each cell line. *EN2* levels are presented as fold change compared with non-silencing control. $n = 3$, $*P < 0.05$, $**P < 0.01$, *t*-test, two-tailed, paired.

To further investigate whether CUX1 and NFIB regulate endogenous *EN2* levels, mRNA levels were measured in a series of knock-down cell lines. For this purpose, HEK293T cells were chosen because they express high levels of *EN2* and are homozygous for the A–C haplotype. Stable cell lines were established using shRNA constructs targeting CUX1, NFIB, both CUX1 and NFIB or a non-silencing control. Efficiency for each knock-down was validated by qRT-PCR (Supplementary Material, Fig. S3). In CUX1 and NFIB double knock-down cell lines, *EN2* levels were decreased by 60% compared with non-silencing control ($P = 0.0023$) (Fig. 9). For the CUX1 single knock-down, *EN2* levels were slightly elevated ($P = 0.03$), whereas no change was observed in the NFIB single knock-down. Thus, only when CUX1 and NFIB are knocked-down together is endogenous *EN2* expression decreased. Together with our co-transfection and supershift data, these results further support CUX1 and NFIB functioning together as transcriptional activators to regulate *EN2* expression via the A–C haplotype.

DISCUSSION

In this study, we demonstrate that the ASD-associated *rs1861972–rs1861973* A–C haplotype functions as a transcriptional activator. Our initial analysis was conducted in mouse cerebellar neurons, so we could investigate the effect of the A–C haplotype at different stages of neuronal development. When the neurons were allowed to differentiate for 72 h in culture, we observed a further enhancement in activity. This increase is coincident with elevated levels of *En2* in cultured granule cell population at 72 h, suggesting that an A–C haplotype function may mirror endogenous *En2* expression of differentiated granule cells. We also determined that both the ASD-associated A and C alleles are necessary for activity,

and the 202-bp fragment encompassing *rs1861972* and *rs1861973* is sufficient for the A–C haplotype function. We then identified Cux1 and Nfib as transcription factors binding to the 202-bp A–C. Their binding was validated by ChIP, supershifts, co-transfection assays and knock-downs. These studies also indicate that both CUX1 and NFIB bind the A–C haplotype together and both factors are required for the A–C haplotype transcriptional activity. These functional data further support *EN2* as an ASD susceptibility locus and suggest that increased levels of *EN2* contribute to ASD risk.

Our studies have also determined that the activity of the G–T haplotype is context-dependent. Although G–T activity is consistently lower than the A–C haplotype, the G–T haplotype is either non-functional (full-length, 24 h), a weak activator (full-length, 72 h) or a repressor (202 bp). These results suggest that binding of transcription factors to the G–T haplotype is influenced by numerous factors such as position, flanking sequence and cell type.

Importantly, the two ASD-associated alleles create a functional unit as a haplotype. When the rare A–T and G–C haplotypes were tested for function, neither associated allele individually was sufficient for activity. Moreover, when *rs1861972* and *rs1861973* were tested as 40-mer oligomers, A–C and G–T do not display an allelic difference. These data indicate the A–C haplotype function requires (1) both A and C alleles and (2) 150-bp spacing between the two alleles. One explanation for these results is that CUX1 and NFIB do not bind the 20-mers for either ASD-associated allele, but instead requires the associated haplotype with the correct 150-bp spacing. Consistent with this possibility, CUX1 and NFIB were identified by mass spectrometry only when the 202-bp A–C probes were used. A different class of proteins binds to the 20-bp A or C probes. In addition, supershifts demonstrated that the 202-bp G–T and A–T probes bind the CUX1 and NFIB at much lower affinity. Interestingly, the G–C haplotype binds both CUX1 and NFIB but the complex migrates at a different position than the A–C haplotype, suggesting a different protein stoichiometry or composition. Consistent with this possibility, the G–C haplotype is non-functional in luc assays. Finally, *rs1861972* and *rs1861973* are in strong LD (26), suggesting that these SNPs have segregated together during evolution. Our primate re-sequencing data support this possibility. Together our data are consistent with the A–C haplotype functioning as a unit.

The A–C haplotype is not situated in a highly conserved non-coding element (HCNE). In fact, *rs1861972* and *rs1861973* and the sequence flanking the SNPs are not conserved in rodents and other mammalian species (Multiz alignment of 44 vertebrates, UCSC Genome browser). Many developmentally important *cis*-elements reside in HCNEs or other evolutionarily conserved regions of the genome (44). Therefore, a more conventional strategy is to exclude disease-associated polymorphisms in not highly conserved regions for functional study. Although this is a reasonable approach, our study indicates that more recently derived polymorphisms should not be overlooked for functional analysis because they might serve as regulatory elements specific to higher-order species.

Our primate re-sequencing indicates that the A–C haplotype is present in both old and new world monkeys. However, the *rs1861972–rs1861973* haplotype is not conserved in other vertebrate species including mouse, rat and chicken. Thus, the A–C activator sequence has likely evolved prior to primate evolution and could result in increased En2 levels during brain development. Mouse studies indicate that En2 coordinates cerebellar development at multiple stages including proliferation. Although it is well known that the cortex has increased in size during evolution, the cerebellum has undergone a similar parallel expansion especially in the lateral lobes, which form connections with the neocortex. Given En2 regulates cerebellar proliferation, the A–C haplotype may contribute to the increase in cerebellar size. Interestingly, it has been hypothesized that dysregulated growth between the cerebellum and cortex may contribute to ASD (45).

Our current luc analysis used equimolar amount of different constructs. Our previous results demonstrated that the A–C haplotype significantly increases luc levels compared with the G–T haplotype. However, luc levels for both the haplotypes were lower than the intron-less control in granule cells. These previous assays employed equal microgram amounts of the different DNA constructs, which is consistent with ∼80% of published transfections (36). However, a recent report indicated that equimolar DNA amounts provided more accurate assessment of a *cis*-element activity when comparing constructs of different sizes. As our intron-less promoter construct is ∼41% shorter than the full-length A–C or G–T intronic constructs, using the same microgram amount could create a discrepancy in copy numbers of each construct transfected. Our results reiterate the importance of using equimolar amount of constructs when they vary in size and are being compared for transcriptional activities.

To identify *trans*-acting factors that specifically bind to the A–C haplotype, we took an unbiased proteomic approach using affinity purification of DNA-bound proteins followed by mass spectrometry. A more standard methodology is to perform antibody supershifts for candidate proteins based on computer predictions of transcription factor-binding sites. However, for many transcription factors, recognition sites can vary depending on flanking sequence, species and protein-binding partners that are not accurately predicted by bioinformatics. In addition, the bioinformatics analysis typically only distinguishes between transcription factor families, making the identification of specific proteins difficult. Using mass spectrometry, we were able to identify specific proteins, which complemented our *in silico* data. Bioinformatics predicted the binding of C/EBP, NFI and NFY transcription factor families which are comprised of multiple genes (*NFIA*, *B*, *C* and *X*; *C/EBPA*, *B*, *D*, *E*, *G* and *Z*, *NFYA*, *B* and *C*) as well as nine Sp1 members and ∼25 Ets factors (34). It is notable that only a subset of NFI factors (NFIB) was identified by mass spectrometry, and CUX1 had never been predicted based on consensus sequences.

Interestingly, in addition to both ASD-associated alleles being required for transcriptional activity, both CUX1 and NFIB are needed to mediate A–C haplotype function. Only over-expression of both transcription factors in HEK cells was sufficient to recapitulate the two signature features of the A–C haplotype: (1) increased activity compared with the intron-less control and (2) a greater difference between the A–C and G–T haplotypes. Over-expression of either CUX1 or NFIB individually is not sufficient to recreate the granule-cell activity of the A–C haplotype in HEK293T cells. In addition, knock-down of both CUX1 and NFIB was needed to decrease endogenous *EN2* levels. Thus, both associated alleles and both protein mediators are required for the functionality of the A–C haplotype.

CUX1 and *NFIB* are important regulators of brain development. They are co-expressed with *EN2* during CNS development and throughout the adulthood. Mouse Cux1 regulates dendritic branching, spine morphology and synapse formation in cerebral cortex, which contributes to cognitive circuitry. Interestingly, a downstream mediator of this Cux1 function is a mouse ortholog of human FAM9, which is implicated in ASD (46). Mouse Nfi gene family plays a critical role in differentiation of cerebellar granule neurons. In early maturation stage of cerebellar granule neurons, *Nfib* directly regulates Contactin2 (Tag-1) expression (47). Importantly, Contactin2 interacts with the ASD-associated CNTNAP2 (48–50). In addition, *Nfib* is involved in brain connectivity by affecting forebrain commissure formation through midline glial population (51). These data suggest that *CUX1* and *NFIB* may also contribute to ASD susceptibility. It will be worthwhile to investigate whether common and rare variants of *CUX1* and *NFIB* are also associated with ASD.

The A–C haplotype is likely to function in concert with other common and rare variants to affect developmental pathways and to increase ASD risk. Animal studies have established that *En2* regulates connectivity (11,12,52,53), serotonin and norepinephrine systems development (14) and E/I circuit balance (20). All of these developmental processes have been implicated in ASD etiology. Initially *En2* is expressed in all mid-hindbrain cells (E8.5) and then is gradually restricted to differentiated cerebellar granule cells (postnatal through adulthood). As correct spatio-temporal expression of *En2* is essential for CNS development, changes in its levels at critical time points could result in dysfunctions in brain circuitry, alterations in neurotransmitter levels and an imbalance in the ratio of E/I synapses. En2 is best known as transcriptional repressor; however, its secreted form can also promote local protein synthesis in the growth cones by regulating eIF4E-dependent translation initiation, which contributes to retino-tectal connectivity (19,52). Misregulation of local protein synthesis at synapses is extensively characterized in Fragile X Syndrome, a Mendelian disorder accounting for ∼6% of ASD cases (54) and considered to contribute to general ASD etiology (55). Finally, recent human microarray data from laser capture dissections of postmortem samples indicate that *EN2* is also expressed in more anterior brain regions including the cortex, amygdala and hippocampus (brain-map.org). Although this anterior expression is not observed in mouse by *in situ* hybridization analysis, it is possible that human *EN2* may be expressed more broadly and the A–C haplotype could impact *EN2* expression in anterior structures relevant to ASD.

This study provides molecular and biochemical evidence that the ASD-associated A–C haplotype functions as a transcriptional activator. These data further support *EN2* as a susceptibility gene and suggest that increased levels of *EN2*

contribute to ASD risk. Future studies will investigate levels in ASD postmortem samples and whether this increase is sufficient to result in molecular and cell biological changes relevant to ASD.

MATERIALS AND METHODS

Luciferase assays

Luciferase assays were performed as previously described (34) with the following modification. Briefly, C57BL6/J mice were maintained on a 12:12 light:dark cycle as approved by RWJMS IACUC. Cerebellar granule cells were isolated from P6 pups by standard protocols and maintained in Basal Medium Eagle containing 10% horse serum, 5% fetal bovine serum (FBS), 0.9% glucose, 1% penicillin–streptomycin and 1% GlutaMAX^{TM} (Invitrogen) at 35°C under 5% CO_2 following transfection. Equimolar amount (4.6 pmol) of different pGL3-promoter constructs and 300 ng each of phRL-null vector (Promega) were transfected into 5 million granule cells using Mouse Neuron Nucleofector Kit and program A-30 on Amexa Nucleofactor System (Amaxa, Inc.). HEK293T cells were maintained in Dulbecco's modified Eagle medium (DMEM) containing 10% FBS at 37°C under 5% CO_2. Cells were grown to ~85% confluency before equimolar amount (0.145 pmol) of the different pGL3-promoter constructs and 10 ng each of phRL-null vector were transiently transfected with 0.4 µg each of pCMV-SPORT6 vector containing human CUX1 and NFIB full-length cDNA (clone ID 5740343 and 3462726, respectively; Open Biosystems) or 0.8 µg of empty pCMV-SPORT6 vector using Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol. Twenty-four hours or 72 h following transfection, cells were collected and lysed using 1× Promega passive lysis buffer. Luciferase activities were measured using the Veritas^{TM} Microplate Luminometer where 85 µl of Promega luciferase substrate (LARII) and 100 µl of Promega Renilla luciferase substrate (Stop & Glo$^®$) were consecutively added to 30 µl of cell lysates. Luciferase readings were normalized by Renilla luciferase values. Statistical analysis was performed using paired, two-tailed Student's *t*-test.

Primate sequence analysis

Sequence analysis was conducted for five primate species. Genomic DNA was obtained from one individual of the following species: *P. pithecia*, *A. nancymai*, *T. gelada* and *P. troglodytes. M. mulatta* and *P. troglodytes* sequences were obtained from UCSC genome browser (http://genome.ucsc.edu/). The intron of *EN2* from each primate species was amplified using primers recognizing conserved sequence between human, chimpanzee and rhesus in exons 1 and 2 (forward primer: ACTCGGACAGCTCGCAAGC; reverse primer: CGGGTTCTTCTTTGGTTTTCG). The PCR reaction was performed using 0.5 µl of Advantage$^®$ GC genomic polymerase mix (Clontech), 25 ng of genomic DNA, 1.5 M GC Melt, 1.1 mM magnesium acetate, 0.5 µM each of forward and reverse primers and 2 mM each of dNTP in a total volume of 25 µl. PCR cycling conditions were as follows: one cycle at 94°C for 4 min, three cycles each at 94°C for

30 s, 63–58°C for 15 s with 1° decrement every three cycles and 72°C for 4 min, and final 15 cycles of 94°C for 30 s, 58°C for 15 s and 72°C for 4 min. PCR products were separated on 1% agarose gel and ~3.5 kb band was isolated and DNA was purified using QIAquick Gel Extraction Kit (Qiagen). Hundred to two hundred nanograms of each purified PCR product was ligated to TOPO vector (Invitrogen) following the manufacturers' protocol. Four clones for *P. pithecia*, *A. nancymai*, *T. gelada* and *P. troglodytes* were sequenced for both strands using the above reverse primer plus two additional primers TCACCCACTGGAAATCTCC and GGAGTGCTTGGCATTCACCC. Sequences were aligned using multiple alignment tool CLUSTALW (http://align.genome.jp/) version 1.83.

EMSAs and antibody supershifts

EMSAs were performed as previously described (34) with the following modifications. DNA probes were generated as follows. Two hundred and two base pairs encompassing *rs1861972* and *rs1861973* region with the A–C, G–T, A–T or G–C haplotype were amplified from the corresponding full-length intronic luc constructs using the following 5′ biotinylated and high-performance liquid chromatography-purified primers or unlabeled primers for cold probes (forward primer: ACCCAGAGGCGAGGTCCAC; reverse primer: GACCTGCCCCAGGTTTTG). PCR was performed using 1.25 unit Advantage$^®$ GC Genomic LA Polymerase (Clontech), 2× Advantage$^®$ GC-melt LA buffer, 0.4 mM of each dNTP, 0.4 µM of each primer, 100 ng of template DNA and the following cycling conditions: one cycle at 94°C for 4 min; 30 cycles at 94°C for 30 s, 60°C for 30 s and 72°C for 30 s. PCR products were separated on 1% agarose gel and single band was isolated using QIAEX$^®$II Gel Extraction Kit (Qiagen). Resulting probes were sequence-verified. Nuclear extract was prepared from P6 mouse cerebellar granule neurons cultured for 24 h using Panomics nuclear extraction kit (AY2002). For pre-absorption, 1 µg of nuclear extract and 1 µg of poly-d(I-C) (Roche) were incubated with binding buffer (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, pH 7.6, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM ammonium sulfate, 1 mM dithiothreitol (DTT), 30 mM KCl and 0.2% Tween-20) at room temperature for 5 min. Ten nanograms of biotin-labeled probes were then added to a final volume of 10 µl and incubated at 20°C for 30 min. For competition assays, 120-fold molar excess of unlabeled probes were added to the mixture prior to the 30-min incubation. For antibody supershifts, 1–5 µg of rabbit polyclonal anti-NFI antibodies (sc-870X, Santa Cruz), rabbit polyclonal anti-CUX1 antibody (sc-13024X, Santa Cruz, and anti-861, courtesy of Dr. Alain Nepveu) or same amount of unrelated rabbit polyclonal anti-green fluorescent protein (GFP) antibody (Molecular Probes) was incubated with 1 µg of nuclear extract and 4 µl of 5× binding buffer without DTT at 4°C overnight prior to binding to 1 µg of poly-d(I-C). Ten nanograms of probes were then added to the final volume of 20 µl. Protein–DNA complexes were separated on a non-denaturing 4 or 5% acrylamide gel in 0.5× Tris–borate–EDTA buffer until the free probes migrate to the end of the gel. For the supershifts, the gels

were run longer until the shifted bands migrated to the middle of the gel, which allowed for better separation. Gels were then wet-transferred onto a Biodyne Nylon membrane (PALL). Membrane was incubated with streptavidin conjugated with horseradish peroxidase (AY1000, Panomics), washed and mixed with substrate solutions to be exposed to a HyBlotCL™ film (Denville Scientific, Inc.) for chemiluminescence detection.

Affinity purification and mass spectrometry

Affinity purification of proteins specifically binding to the A–C haplotype was conducted as follows. Nuclear extracts and biotinylated DNA probes were prepared as described above for EMSAs. For the lambda control probe, the following biotinylated primers were used to amplify a 197-bp region from lambda DNA (forward primer: TCTATCACCG CAAGGGATAA; reverse primer: ATGAATGGCCTTGTT GATCG). For 20-mer probes, biotinylated oligos of sense and antisense sequences encompassing 10 bp each 5′ and 3′ of *rs1861972*-A and *rs1861973*-C were annealed to generate double-stranded probes. Sequences for the oligos are as follows: (A-20mer sense: CTCCCTGCCAATGGCCTTG CC; A-20mer anti-sense: GGCAAGGCCATTGGCAGGGAG; C-20mer sense: AGCGACCCTGCCCAAAACCTG; C-20mer anti-sense: CAGGTTTTGGGCAGGGTCGCT). Affinity purification was performed as follows. Five hundred nanograms of DNA probes were bound to Dynabeads® MyOne™ Streptavidin T1 (Invitrogen) in 5 mM Tris (pH 7.5), 0.5 mM EDTA and 1 M NaCl at room temperature for 30 min. Sixty micrograms of nuclear extract were pre-incubated with 30 μg of poly d(I-C) in EMSA-binding buffer containing Complete Protease Inhibitor Cocktail (Roche) at room temperature for 45 min. For 20-mer experiments, additional pre-absorption step with 300 molar excess of unlabeled self-probes was included. Pre-incubated nuclear extract was then incubated with DNA probe-bound beads on a rotating platform at room temperature for 30 min. Beads were washed twice with a binding buffer containing 0.1 μg/μl poly d(I-C) and three times with the binding buffer. DNA-bound proteins were eluted with binding buffer containing 1 M KCl at room temperature for 30 min. Sodium dodecyl sulfate (SDS) sample buffer (1% SDS, 40 mM Tris–HCl, pH 6.8, 8% glycerol, 0.8% 2-mercaptoethanol and trace of Bromophenol Blue) was then added to the eluate and loaded on 5% SDS–polyacrylamide gel electrophoresis (PAGE) gel. Electrophoresis was carried out for 10 min at 150 V, so all the proteins entered the gel. The gel was then stained in colloidal Coomassie solution (10% phosphoric acid, 10% ammonium sulfate, 20% methanol and 0.12% Coomassie blue G-250) (56) overnight. About 1 cm of the gel was then excised as a single slice to facilitate subsequent in-gel tryptic digestion. LC-MS/MS experiments were performed using a Dionex U3000 nanoflow chromatography system in line with an LTQ linear ion trap mass spectrometer (ThermoFischer). Extracted peptides from in-gel digest were first solubilized in 0.1% trifluoroacetic acid and loaded on a self-packed 75 μm × 12 cm emitter column packed with 3 μm, 200 Å Magic C18AQ (Michrom Biresources, Inc., Auburn, CA, USA) and eluted with a linear gradient of 2–45% mobile phase B in 90 min (mobile phase A:

0.1% formic acid/water; mobile phase B: 0.1% formic acid/ acetonitrile). Mass spectrometry data were acquired using a data-dependent acquisition procedure with a cyclic series of a full scan followed by MS/MS scans of the 10 most intense peaks with a repeat count of 2 and the dynamic exclusion duration of 30 s. The LC-MS/MS data were searched against mouse ENSEMBL database using an in-house version of the GPM (GPM Extreme, Beavis Informatics Ltd., Winnipeg, Canada) (57) with fixed modification of cysteine by iodoacetamide (+57 Da) and potential modifications for oxidation of methionine (+15.99) and deamination of asparagine (+1 Da). Relative protein abundance was estimated by counting the total number of MS/MS spectra assigned to each protein (39). Only the peptides of log(e) value smaller than −2 were subjected to spectral counting. Statistical analysis was performed using R program as described previously (37,38).

Chromatin immunoprecipitation

ChIP was performed on HEK293T and SH-SY5Y cells under both endogenous and NFIB over-expression conditions. For the over-expression cells were transiently transfected with NFIB full-length cDNA construct (Clone ID 3462726, Open Biosystems) or empty pCMV-SPORT6. Eighty-five percent of confluent cells were fixed in 1% formaldehyde for 10 min. Fixing was stopped by adding glycine at a final concentration of 0.125 M for 5 min. After washing with phosphate-buffered saline, cells were lysed with SDS lysis buffer (1% SDS, 10 mM EDTA and 50 mM Tris, pH 8.0), nuclei were pelleted and resuspended in 500 μl of nuclear lysis buffer (50 mM Tris, pH 7.5, 0.5 M NaCl, 1% NP-40, 1% DOC, 0.1% SDS, 2 mM EDTA). Ten million per milliliter of lysed cell-equivalent nuclei were then sonicated in 500 μl volume in Misonix Sonicator® with Cup Horn (Misonix, Farmingdale, NY, USA) for 3 min at level 2, where each 15-s pulse was broken by 50-s pause. Two-and-a-half million cell-equivalent nuclei were diluted 1–10 in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl and 16.7 mM Tris–HCl, pH 8.0) and anti-NFI antibody was added for immunoprecipitation reaction. For the negative control, no antibody was added to the same amount of cell-equivalent nuclei. Before the addition of antibody, 1% of sheared chromatin was taken from each reaction for total input control. Antibody binding was conducted at 4°C on a rotating platform overnight. The next day, 20 μl of Protein A/G mix magnetic coupled beads was added to the chromatin and incubated at 4°C for 2 h. Magnetic beads were separated and then washed with low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8.0 and 150 mM NaCl), high salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, pH 8.0 and 500 mM NaCl), LiCl (0.25 M LiCl, 1% IGEPAL-CA630, 1% deoxycholic acid, 1 mM EDTA and 10 mM Tris, pH 8.0) and TE (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) wash buffer. Bound chromatin was eluted from the beads by adding elution buffer (1% SDS and 0.1 M NaHCO₃). Input DNA was also mixed with the same elution buffer and from this point on subjected to the same treatment as immunoprecipitated chromatin. To reverse the cross-linking between the DNA and proteins,

NaCl was added to the chromatin to 0.2 M and incubated at 65°C for 6 h. Eluate was then treated with 20 μg of DNase-free RNase A at 37°C for 15 min followed by treatment with 20 μg of Proteinase K at 45°C for 2 h. DNA was then isolated using Promega Wizard Gel and PCR purification kit. Four percent of DNA from each immunoprecipitation and 0.1% of Input DNA was used for qRT-PCR using GoTaq® qPCR Master mix (Promega). To detect the binding of NFIB, 87 bp around *rs1861972* and 114 bp around *rs1861973* were amplified using the following primers. *rs1861972*—forward primer: TCCCTAAAGCCGATTCA-TACA; reverse primer: GGGAAGAAGGGGGCAAG; *rs1861973*—forward primer: CCTTCTGCTCTCCTCCCTCT; reverse primer: GAACCTGACCTGGCCTTCT). As a negative control, 129 bp was amplified from 2 kb upstream of p21 promoter (58) using the following primers—forward primer: CTGTGGCTCTGATTGGCTTT; reverse primer: CTCCTACCATCCCCTTCCTC. Cycling conditions were as follows: one cycle at 94°C for 2 min, 40 cycles at 94°C for 30 s, 60°C for 30 s and 72°C for 40 s, and then the final cycle at 95°C for 15 s, 60°C for 15 s and 95°C for 15 s for primers dissociation curves. Standard curves were generated for each primer set for every experiment to assure the primer efficiencies were similar. Anti-NFI bound or no antibody control fraction was normalized over input DNA levels for each primer set. The whole procedure was repeated four times to obtain the average and standard error of mean.

Generation of CUX1 and NFIB stable knock-downs

Human embryonic kidney cells (HEK293T) were transfected with Expression Arrest GIPZ Lentiviral shRNAmir constructs (Open Biosystems) to knock-down CUX1 and NFIB expression levels. Twenty-four hours prior to transfection, cells were plated at 6×10^5 cells per 6-cm plate. A total of 1.05 nM shRNA construct was used for each single gene knock-down, whereas 0.525 nM of each construct was employed for the double knock-down. As a control, cells were also transfected with 1.05 nM of a non-silencing construct. Transfection was carried out using Lipofectamine 2000 reagent (Invitrogen) following the manufacturer's guidelines. Cells were maintained in a transfection medium for 72 h, and the transfection efficiency was determined by visualizing GFP expression. For selection of stable lines, cells were maintained in a medium containing 2 μg/ml of puromycin for 2 weeks.

EN2 level measurement in human cell lines

Cells were maintained in DMEM (SK-N-MC, HEK293T, SH-SY5Y and DAOY) or RPMI-1640 medium (PFSK-1) and 10% FBS, at 37°C under 5% CO_2. Genomic DNA was isolated using a conventional method and *rs1861972–rs1861973* genotype was determined for each line using Luminex® technology (details in Supplementary Material). For the stable HEK293T knock-downs, the cells were grown as above but supplemented with 2 μg/μl of puromycin. Cells were grown to confluency before total RNA was isolated using TRIzol® reagent (Invitrogen) following the manufacturer's instruction. Single-strand cDNA was synthesized from total RNA and quantified using Taqman® qRT-PCR. Three micrograms of RNA was treated with 2 units of DNaseI (New England Biolabs) in a buffer containing 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$ and 10 mM Tris–HCl, pH 7.6 and RNasin® Ribonuclease Inhibitor (Promega) at 37°C for 30 min. DNaseI was subsequently inactivated by heating at 75°C for 10 min. Single-strand cDNA was then generated using 1 μg of DNaseI-treated RNA and high-capacity cDNA Reverse Transcription Kit (Applied Biosystems) following the manufacturer's instruction. Quantitative PCR was conducted using one-twentieth of total cDNA and Taqman® probe sets for human *EN2* (Hs00171321_m1, fluorescent dye FAM™-labeled) and *GAPDH* internal control (4326317E, fluorescent dye VIC®-labeled) on ABI7900HT (Applied Biosystems). *EN2* level was normalized to endogenous *GAPDH* level by subtracting *GAPDH* Ct from *EN2* Ct (ΔCt). Average ΔCt values were obtained from three replicates of qRT-PCR reaction. ΔΔCt method was applied to calculate relative changes in expression levels. For the cell line analysis, geometric mean was used to combine ΔCt values between two AC/AC cell lines or three AC/GT cell lines. Two-tailed, paired Student's *t*-test (knock-downs) or unequal variance *t*-test (cell lines) was used to test for significance.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Dawson, G. (2010) Recent advances in research on early detection, causes, biology, and treatment of autism spectrum disorders. *Curr. Opin. Neurol.*, **23**, 95–96.
2. Freitag, C.M. (2007) The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry*, **12**, 2–22.
3. Ronald, A., Happe, F., Bolton, P., Butcher, L.M., Price, T.S., Wheelwright, S., Baron-Cohen, S. and Plomin, R. (2006) Genetic

heterogeneity between the three components of the autism spectrum: a twin study. *J. Am. Acad. Child Adolesc. Psychiatry*, **45**, 691–699.

4. Geschwind, D.H. (2009) Advances in autism. *Annu. Rev. Med.*, **60**, 367–380.

5. State, M.W. (2010) The genetics of child psychiatric disorders: focus on autism and Tourette syndrome. *Neuron*, **68**, 254–269.

6. Abrahams, B.S. and Geschwind, D.H. (2008) Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.*, **9**, 341–355.

7. Campbell, D.B., Sutcliffe, J.S., Ebert, P.J., Militerni, R., Bravaccio, C., Trillo, S., Elia, M., Schneider, C., Melmed, R., Sacco, R. *et al.* (2006) A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl Acad. Sci. USA*, **103**, 16834–16839.

8. Davis, C.A., Noble-Topham, S.E., Rossant, J. and Joyner, A.L. (1988) Expression of the homeo box-containing gene En-2 delineates a specific region of the developing mouse brain. *Genes Dev.*, **2**, 361–371.

9. Kuemerle, B., Zanjani, H., Joyner, A. and Herrup, K. (1997) Pattern deformities and cell loss in Engrailed-2 mutant mice suggest two separate patterning events during cerebellar development. *J. Neurosci.*, **17**, 7881–7889.

10. Millen, K.J., Wurst, W., Herrup, K. and Joyner, A.L. (1994) Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development*, **120**, 695–706.

11. Sillitoe, R.V., Stephen, D., Lao, Z. and Joyner, A.L. (2008) Engrailed homeobox genes determine the organization of Purkinje cell sagittal stripe gene expression in the adult cerebellum. *J. Neurosci.*, **28**, 12150–12162.

12. Sillitoe, R.V., Vogel, M.W. and Joyner, A.L. (2010) Engrailed homeobox genes regulate establishment of the cerebellar afferent circuit map. *J. Neurosci.*, **30**, 10015–10024.

13. Vogel, M.W., Ji, Z., Millen, K. and Joyner, A.L. (1996) The Engrailed-2 homeobox gene and patterning of spinocerebellar mossy fiber afferents. *Brain Res. Dev. Brain Res.*, **96**, 210–218.

14. Simon, H.H., Scholz, C. and O'Leary, D.D. (2005) Engrailed genes control developmental fate of serotonergic and noradrenergic neurons in mid- and hindbrain in a gene dose-dependent manner. *Mol. Cell. Neurosci.*, **28**, 96–105.

15. Belmonte, M.K., Allen, G., Beckel-Mitchener, A., Boulanger, L.M., Carper, R.A. and Webb, S.J. (2004) Autism and abnormal development of brain connectivity. *J. Neurosci.*, **24**, 9228–9231.

16. Baader, S.L., Sanlioglu, S., Berrebi, A.S., Parker-Thornburg, J. and Oberdick, J. (1998) Ectopic overexpression of engrailed-2 in cerebellar Purkinje cells causes restricted cell loss and retarded external germinal layer development at lobule junctions. *J. Neurosci.*, **18**, 1763–1773.

17. Joliot, A., Trembleau, A., Raposo, G., Calvet, S., Volovitch, M. and Prochiantz, A. (1997) Association of Engrailed homeoproteins with vesicles presenting caveolae-like properties. *Development*, **124**, 1865–1875.

18. Joliot, A., Maizel, A., Rosenberg, D., Trembleau, A., Dupas, S., Volovitch, M. and Prochiantz, A. (1998) Identification of a signal sequence necessary for the unconventional secretion of Engrailed homeoprotein. *Curr. Biol.*, **8**, 856–863.

19. Wizenmann, A., Brunet, I., Lam, J.S., Sonnier, L., Beurdeley, M., Zarbalis, K., Weisenhorn-Vogt, D., Weinl, C., Dwivedy, A., Joliot, A. *et al.* (2009) Extracellular Engrailed participates in the topographic guidance of retinal axons *in vivo*. *Neuron*, **64**, 355–366.

20. Tripathi, P.P., Sgado, P., Scali, M., Viaggi, C., Casarosa, S., Simon, H.H., Vaglini, F., Corsini, G.U. and Bozzi, Y. (2009) Increased susceptibility to kainic acid-induced seizures in Engrailed-2 knockout mice. *Neuroscience*, **159**, 842–849.

21. Garber, K. (2007) Neuroscience: autism's cause may reside in abnormalities at the synapse. *Science*, **317**, 190–191.

22. Ritvo, E.R., Yuwiler, A., Geller, E., Ornitz, E.M., Saeger, K. and Plotkin, S. (1970) Increased blood serotonin and platelets in early infantile autism. *Arch. Gen. Psychiatry*, **23**, 566–572.

23. Anderson, G.M., Freedman, D.X., Cohen, D.J., Volkmar, F.R., Hoder, E.L., McPhedran, P., Minderaa, R.B., Hansen, C.R. and Young, J.G. (1987) Whole blood serotonin in autistic and normal subjects. *J. Child Psychol. Psychiatry*, **28**, 885–900.

24. Lake, C.R., Ziegler, M.G. and Murphy, D.L. (1977) Increased norepinephrine levels and decreased dopamine-beta-hydroxylase activity in primary autism. *Arch. Gen. Psychiatry*, **34**, 553–556.

25. Cook, E.H. Jr, Leventhal, B.L., Heller, W., Metz, J., Wainwright, M. and Freedman, D.X. (1990) Autistic children and their first-degree relatives:

26. Benayed, R., Gharani, N., Rossman, I., Mancuso, V., Lazar, G., Kamdar, S., Bruse, S.E., Tischfield, S., Smith, B.J., Zimmerman, R.A. *et al.* (2005) Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. *Am. J. Hum. Genet.*, **77**, 851–868.

27. Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L.M. and Millonig, J.H. (2004) Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Mol. Psychiatry*, **9**, 474–484.

28. Brune, C.W., Korvatska, E., Allen-Brady, K., Cook, E.H. Jr, Dawson, G., Devlin, B., Estes, A., Hennelly, M., Hyman, S.L., McMahon, W.M. *et al.* (2008) Heterogeneous association between engrailed-2 and autism in the CPEA network. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 187–193.

29. Petit, E., Herault, J., Martineau, J., Perrot, A., Barthelemy, C., Hameury, L., Sauvage, D., Lelord, G. and Muh, J.P. (1995) Association study with two markers of a human homeogene in infantile autism. *J. Med. Genet.*, **32**, 269–274.

30. Sen, B., Singh, A.S., Sinha, S., Chatterjee, A., Ahmed, S., Ghosh, S. and Usha, R. (2010) Family-based studies indicate association of Engrailed 2 gene with autism in an Indian population. *Genes Brain Behav.*, **9**, 248–255.

31. Wang, L., Jia, M., Yue, W., Tang, F., Qu, M., Ruan, Y., Lu, T., Zhang, H., Yan, H., Liu, J. *et al.* (2008) Association of the ENGRAILED 2 (EN2) gene with autism in Chinese Han population. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **147B**, 434–438.

32. Yang, P., Lung, F.W., Jong, Y.J., Hsieh, H.Y., Liang, C.L. and Juo, S.H. (2008) Association of the homeobox transcription factor gene ENGRAILED 2 with autistic disorder in Chinese children. *Neuropsychobiology*, **57**, 3–8.

33. Yang, P., Shu, B.C., Hallmayer, J.F. and Lung, F.W. (2010) Intronic single nucleotide polymorphisms of engrailed homeobox 2 modulate the disease vulnerability of autism in a Han Chinese population. *Neuropsychobiology*, **62**, 104–115.

34. Benayed, R., Choi, J., Matteson, P.G., Gharani, N., Kamdar, S., Brzustowicz, L.M. and Millonig, J.H. (2009) Autism-associated haplotype affects the regulation of the homeobox gene, ENGRAILED 2. *Biol. Psychiatry*, **66**, 911–917.

35. Manzini, M.C., Ward, M.S., Zhang, Q., Lieberman, M.D. and Mason, C.A. (2006) The stop signal revised: immature cerebellar granule neurons in the external germinal layer arrest pontine mossy fiber growth. *J. Neurosci.*, **26**, 6040–6051.

36. Seelan, R.S., Janckila, A.J., Parthasarathy, R.N. and Casanova, M.F. (2006) The importance of using equimolar DNA for transfection analysis of the 5′ flanking promoter regions of genes. *Anal. Biochem.*, **349**, 306–308.

37. Qian, M., Sleat, D.E., Zheng, H., Moore, D. and Lobel, P. (2008) Proteomics analysis of serum from mutant mice reveals lysosomal proteins selectively transported by each of the two mannose 6-phosphate receptors. *Mol. Cell Proteomics*, **7**, 58–70.

38. Sleat, D.E., Della Valle, M.C., Zheng, H., Moore, D.F. and Lobel, P. (2008) The mannose 6-phosphate glycoprotein proteome. *J. Proteome Res.*, **7**, 3010–3021.

39. Liu, H., Sadygov, R.G. and Yates, J.R. III (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.*, **76**, 4193–4201.

40. Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L. and Samatova, N.F. (2006) Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.*, **5**, 2909–2918.

41. Wilson, E.B. (1927) Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.*, **22**, 209–212.

42. Wang, W., Mullikin-Kilpatrick, D., Crandall, J.E., Gronostajski, R.M., Litwack, E.D. and Kilpatrick, D.L. (2007) Nuclear factor I coordinates multiple phases of cerebellar granule cell development via regulation of cell adhesion molecules. *J. Neurosci.*, **27**, 6115–6127.

43. Wang, W., Stock, R.E., Gronostajski, R.M., Wong, Y.W., Schachner, M. and Kilpatrick, D.L. (2004) A role for nuclear factor I in the intrinsic control of cerebellar granule neuron gene expression. *J. Biol. Chem.*, **279**, 53491–53497.

44. Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K. *et al.* (2005)

Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.*, **3**, e7.

45. Rubenstein, J.L. (2010) Three hypotheses for developmental defects that may underlie some forms of autism spectrum disorder. *Curr. Opin. Neurol.*, **23**, 118–123.

46. Cubelos, B., Sebastian-Serrano, A., Beccari, L., Calcagnotto, M.E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J.M., Bovolenta, P. *et al.* (2010) Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron*, **66**, 523–535.

47. Wang, W., Crandall, J.E., Litwack, E.D., Gronostajski, R.M. and Kilpatrick, D.L. (2010) Targets of the nuclear factor I regulon involved in early and late development of postmitotic cerebellar granule neurons. *J. Neurosci. Res.*, **88**, 258–265.

48. Alarcon, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H. *et al.* (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am. J. Hum. Genet.*, **82**, 150–159.

49. Arking, D.E., Cutler, D.J., Brune, C.W., Teslovich, T.M., West, K., Ikeda, M., Rea, A., Guy, M., Lin, S., Cook, E.H. *et al.* (2008) A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am. J. Hum. Genet.*, **82**, 160–164.

50. Bakkaloglu, B., O'Roak, B.J., Louvi, A., Gupta, A.R., Abelson, J.F., Morgan, T.M., Chawarska, K., Klin, A., Ercan-Sencicek, A.G., Stillman, A.A. *et al.* (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.*, **82**, 165–173.

51. Plachez, C., Lindwall, C., Sunn, N., Piper, M., Moldrich, R.X., Campbell, C.E., Osinski, J.M., Gronostajski, R.M. and Richards, L.J. (2008) Nuclear factor I gene expression in the developing forebrain. *J. Comp. Neurol.*, **508**, 385–401.

52. Brunet, I., Weinl, C., Piper, M., Trembleau, A., Volovitch, M., Harris, W., Prochiantz, A. and Holt, C. (2005) The transcription factor Engrailed-2 guides retinal axons. *Nature*, **438**, 94–98.

53. Cheng, Y., Sudarov, A., Szulc, K.U., Sgaier, S.K., Stephen, D., Turnbull, D.H. and Joyner, A.L. (2010) The Engrailed homeobox genes determine the different foliation patterns in the vermis and hemispheres of the mammalian cerebellum. *Development*, **137**, 519–529.

54. Reddy, K.S. (2005) Cytogenetic abnormalities and fragile-X syndrome in autism spectrum disorder. *BMC Med. Genet.*, **6**, 3.

55. Krueger, D.D. and Bear, M.F. (2011) Toward fulfilling the promise of molecular medicine in fragile X syndrome. *Annu. Rev. Med.*, **62**, 411–429.

56. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G.M., Carnemolla, B., Orecchia, P., Zardi, L. and Righetti, P.G. (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis*, **25**, 1327–1333.

57. Craig, R., Cortens, J.P. and Beavis, R.C. (2004) Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.*, **3**, 1234–1242.

58. Ouellet, S., Vigneault, F., Lessard, M., Leclerc, S., Drouin, R. and Guerin, S.L. (2006) Transcriptional regulation of the cyclin-dependent kinase inhibitor 1A (p21) gene by NFI in proliferating human cells. *Nucleic Acids Res.*, **34**, 6472–6487.