Autism and ultraconserved non-coding sequence on chromosome 7q

Esther Richler, Jennifer G. Reichert, Joseph D. Buxbaum and Lynne Alison McInnes

Objective **Autism has been linked to a broad region on chromosome 7q that contains a large number of genes involved in transcription and development. This region is also enriched for ultraconserved non-coding elements, defined as human–rodent sequences that are 100% aligned over $\geq$ 200 base pairs, which have a high likelihood of being functional. Therefore, as only a few rare coding variants have been detected in the autism candidate genes on 7q examined to date, we decided to screen these ultraconserved elements for possible autism susceptibility alleles.**

Methods **We used denaturing high-performance liquid chromatography, and DNA sequencing, to perform variant detection in a total of 146 cases with autism, 96 from the Autism Genetic Resource Exchange and 50 from the Central Valley of Costa Rica, as well as 124 controls from the Polymorphism Discovery Resource Panel. We screened 10 consecutive ultraconserved elements in, or flanking, the genes DLX5/6, AUTS2 and FOXP2 on chromosome 7q.**

Results **Although we did find several rare variants in autism cases that were not present in controls, we also observed rare variants present in controls and not cases. The most common variant occurred in controls at a frequency of 3.3%. Interestingly, two ultraconserved elements each harbored three independent variants and one ultraconserved element harbored two independent variants, suggesting that ultraconservation is maintained chiefly by a decreased tendency toward fixation, rather than a significantly lower mutation rate.**

Conclusions **Our results show that these sequences are unlikely to harbor major autism susceptibility alleles.**
***Psychiatr Genet* 16:19–23 © 2006 Lippincott Williams & Wilkins.**

Psychiatric Genetics 2006, **16:**19–23

Keywords: autistic disorder, conserved sequence, evolution, genetic variation, molecular, nucleic acid, polymorphism, regulatory sequences, single nucleotide

Department of Psychiatry, Mount Sinai School of Medicine, New York, USA

Correspondence and requests for reprints to Lynne Alison McInnes, MD., One Gustave L. Levy Place Box 1229, New York, NY 10029, USA
Tel: +1 212 241 8177; fax: +1 212 241 2053;
e-mail: Alison.mcinnes@mssm.edu

Sponsorship: This work was supported by NINDS R01 043540 (to L.A.M) and the Seaver Autism Research Center and Cure Autism Now (to J.D.B), as well as by the National Institute of Mental Health through a STAART Grant (U54 MH 066673) (to J.D.B.).

Received 14 September 2005 **Accepted** 1 November 2005

Introduction

Autism is a complex genetic disorder characterized by deficits in verbal and non-verbal communication as well as impaired socialization and persistent repetitive behaviors. Convergent findings from several independent genome-wide linkage studies of autism, as well as follow-up studies in expanded samples, show evidence of linkage to autism in overlapping regions covering 7q21–7q34 (Folstein and Rosen-Sheidley, 2001). Additionally, multiple chromosomal abnormalities interrupting chromosome 7q have been identified in cases with autism or autistic features (The Autism Chromosome Rearrangement Database http://projects.tcag.ca/autism/) although no coding mutations were found in either of the two genes disrupted by breakpoints, AUTS2 at 7q11.2 (Sultana *et al.*, 2002) and ST7 at 7q31 (Vincent *et al.*, 2000), in the small number of cases screened.

The coding sequence of several other excellent candidate genes in the chromosome 7 linkage region have been screened for mutations by various groups. Although common variants have not been found, a small number of rare non-synonymous variants that could be true susceptibility alleles have been identified in WNT2 (Wassink *et al.*, 2001), CPA1 and CPA5 (Bonora *et al.*, 2002), RELN (Bonora *et al.*, 2003), LAMB1, CUTL1 and PTPRZ1 (Bonora *et al.*, 2005) and the DLX5/6 bigene pair (John L. Rubenstein, unpublished data). The DLX5/6 gene pair findings are especially interesting as these genes have also been implicated in Rett syndrome (Horike *et al.*, 2005). For example, methyl-CpG binding protein 2 (Mecp2) is thought to bind a silent chromatin loop structure associated with the DLX5/6 genes to regulate their expression, and DLX5 appears to lose its maternal imprinting status in Mecp2 knock-out mice (Horike *et al.*, 2005).

Another candidate gene in the chromosome 7 autism linkage region, FOXP2, has been intensively investigated as it is responsible for a severe speech and language

0955-8829 © 2006 Lippincott Williams & Wilkins

phenotype (Lai *et al.*, 2001) and a nonsense mutation in exon 7 was recently discovered in two affected siblings with verbal dyspraxia (Macdermot *et al.*, 2005). No unequivocal disease mutations, however, have been identified in multiple different screens utilizing a combined total of 308 individual cases of autism and a greater number of controls (Newbury *et al.*, 2002; Wassink *et al.*, 2002; Gauthier *et al.*, 2003; Li *et al.*, 2005). In fact, little variation has been observed in the coding sequence of candidate genes for the major psychiatric disorders studied to date, so it is very possible that at least some common disease susceptibility alleles lie in the non-coding sequence. Therefore, we examined comparative genomic strategies to identify novel regulatory elements in the chromosome 7q region linked to autism.

Comparative genomics is widely used to identify unknown functional elements in the genome on the basis of the theory that DNA sequence important for an organism is likely to evolve less rapidly than a non-functional sequence (Kimura, 1983). Initially, a human–mouse conservation threshold of 70% sequence identity over at least 100 base pairs was arbitrarily used to screen for potential regulatory sequences; however, there are $\sim$1.3 million such sequences in the human genome and many of these likely represent neutrally evolving DNA (Ovcharenko *et al.*, 2004). Therefore, to increase the likelihood of finding a functional non-coding sequence while reducing the number of spurious elements, investigators have utilized more evolutionarily distant species for comparison, such as the pufferfish *Fugu Rubripes* (Woolfe *et al.*, 2005), or have increased the stringency of human–rodent alignments (Bejerano *et al.*, 2004). For instance, the former group performed a whole human–*Fugu* genome comparison and identified 1400 conserved non-coding elements, many of which overlap with 21 experimentally verified transcription enhancers in or near genes involved in development (listed in Ovcharenko *et al.*, 2004).

On the other hand, Bejerano *et al.* (2004) ranked and filtered human, mouse and rat alignments, from genomic regions exhibiting synteny between these species, by the extent of uninterrupted identity over $\geq$200 base pairs and found 481 such ultraconserved (UC) elements. Two hundred and fifty-six of these are non-coding, either intronic ($n = 100$) or intergenic. A table of these UC sequences can be found in their supplementary data (http://www.cse.ucsc.edu/$\sim$jill/ultra.html). These human–rodent UC elements are also not randomly distributed. The non-coding UC elements tend to cluster near genes involved with transcription and development, and are enriched in regions linked to autism on chromosomes 7q, 2q and X. This is not surprising in that autism is a neurodevelopmental disorder and might well be associated with genes influencing brain development.

Although one might expect to find only rare variants in UC elements, the relationship between the strength of sequence conservation and natural selection is not straightforward. For instance, a recent study demonstrated a genome-wide relaxation of selective constraint in the primate lineage, which may have resulted from a reduced effective population size in comparison with rodents (Kryukov *et al.*, 2005). Additionally, this effect was most prominent in conserved non-coding sequence, as opposed to protein coding sequence, and the authors suggest that this is because relatively weaker selection is operating on the former. To our knowledge, no reports of systematic variation detection in UC elements have yet been published; however, we reviewed all the UC elements using the RefSeq track of the UCSC genome annotation database (Kent *et al.*, 2002) Human Genome May 2004 (hg17) assembly and found variants in 19, some of which are highly polymorphic with the frequency of the minor allele ranging up to 0.488. Therefore, as it was at least theoretically possible that we might find an allele associated with autism in $\geq$5% of our sample, and because UC elements are highly likely to be functional, we chose to screen 10 consecutive non-coding UC elements distributed across AUTS2 (UC218–219), the DLX5/6 bigene pair (UC220–221) and FOXP2 (UC222–227) for alleles associated with autism. Notably, UC221 overlaps a known enhancer that regulates the expression of the DLX5/6 homeobox genes in the ventral forebrain (Zerucha *et al.*, 2000).

Methods

Study population

We examined 96 unrelated autism cases, all members of multiplex families, from the Autism Genetic Resource Exchange (AGRE; http://www.agre.org) and 50 cases, related by no fewer than six generations, from a genetically isolated founder population in the Central Valley of Costa Rica (CVCR) (McInnes *et al.*, 2005). Written, informed consent was obtained from all study participants and participating family members. Collection and use of the Costa Rican samples was approved by the Bioethics Committee at the Hospital Nacional de Ninos. We used 124 Caucasians from the Polymorphism Discovery Resource Panel (Collins *et al.*, 1998) to examine the extent of normal genetic variation in these amplicons. We did not have controls for the Costa Rican population at the time, but screened these samples in case we found variants that might be shared by individuals with distant common ancestors, lending support for the association of the variant with autism.

Power to detect variants at a frequency of 1 or 5%

The probability of detecting a variant with allele frequency p in a sample of n individuals was calculated as $1-(1-p)^{2n}$. Power to detect a variant with a frequency of 0.05 in both the 124 control individuals and the 96 AGRE cases was $\sim$100% while power to detect a

variant with a frequency of 0.01 was 92 and 85%, respectively.

Denaturing high-performance liquid chromatography

Primers to UC regions were designed using primer3 software (Rozen and Skaletsky, 2000). The melting temperature profile of the primers was determined using Transgenomic Navigator software, and GC clamps were added as needed. Genomic DNA was amplified using Hotstar Taq (Qiagen, Valencia, California, USA) and a touchdown polymerase chain reaction protocol.

AUTS2 UC218 L – cttgttgtgaaacactagctctaatac R – cacatgacaatcaaaaccatgac

AUTS2 UC219 L – cccatttttaaatcatgcacag R – aacacacagtaatgaatcaggagaa

DLX5 UC220 L – ttccacctcccaaaataaacc R – cgaatcgttggaggactgac

DLX5 UC221A L – tcgcattgttacattaggaaaa R – GTATTAAAAGTGGAAGAAAATTACAGG

DLX5 UC221B L – cgcccgccgcCATTCTCTTAAATGCAGCCATAA R – taattcagcaaagccccact

FOXP2 UC222 L – ATAGGGATTTGTGCTGGAATGT R – TTTAGAGGAAACAAAAGGCAAAA

FOXP2 UC223 L – AAACTCCAGTGGTTATCTGTTTTTG R – CCACAGAATCTTGCCGAAGT

FOXP2 UC224 L – GTAAGGAGGCCTAATTGAGCTATG R – CACAAGATTGCATTTTCTCTTCC

FOXP2 H-M L – tttgcattcaatcctttcaca R – aataaaaattacaattacagccatctt

FOXP2 UC225 L – CACTTGGCCTGACATATAGAGTTG R – TCAGGCATGAAATGGTCATAAGT

FOXP2 CNE_A L – tcccataaatgcacagattcc R – atgcctgagcagaggaaaag

FOXP2 UC226 L – ATAGCAGTCCAAAAATCGTGTG R – GTGACTCCAATTGAAGGTGTTG

FOXP2 CNE_B L – gcaattcaaccaccattcct R – tttcactgctctccaaaacaaa

FOXP2 UC227 L – GACAGAATAATTGTCTCATAAATCCA R – CAGATATCGCTAGTTCCAAAGG

Denaturing high-performance liquid chromatography analysis was performed using the WAVE Nucleic Acid Fragment Analysis System (Transgenomics, San Diego, California, USA) according to the manufacturer's protocol. Samples showing a deviant elution profile were directly sequenced using the ABI PRISM BigDye system (Foster City, California, USA) and either an ABI 377 or an ABI 3700 automated sequencer.

Results and discussion

Although we did find several variants in autism cases that were not present in controls, we also observed variants present in controls and not in cases. Moreover, all variants detected were rare and the most common variant occurred in controls (UC226) at a frequency of 3.3%. Table 1 summarizes our results. We found 13 novel independent variants in seven UC elements; two elements had three variants and one element had two variants. The variant in UC218 that was present in an individual from all three samples may be *de novo* in the CVCR case, although we cannot rule out a non-paternity that escaped us. Variants in the UC elements did not appear to be correlated with their position in the element (within 20 base pairs of the ends) or the depth of conservation of the element. For completeness, we also screened three non-coding elements within FOXP2 conserved at 100% between the human and the mouse (but not rat) for ≥ 200 base pairs. Two of these were human–*Fugu* conserved non-coding elements (CNE_A and CNE_B) previously defined by Woolfe *et al.* (2005) and had no variants. We found one variant in an AGRE case, but not the affected sibling, in the third 356 base pair human–mouse element within FOXP2.

Despite lack of evidence for an association with autism, the variants in UC221 and UC226 are of theoretical interest for students of population genetics and comparative genomics. The UC221 element, harboring three variants, overlaps an enhancer governing transcription of the DLX5/6 gene pair *and* the edges of two mutually alternatively spliced exons of a brain-expressed hypothetical protein FLJ34048 coded for on the opposite strand. The positions of the three variants are highlighted in Fig. 1. Two minimal seven nucleotide motifs conserved with zebrafish, 156i and 156ii, are necessary for protein binding and functional activity (Zerucha *et al.*, 2000). A single-nucleotide substitution in either binding site greatly reduces transcription in transient transfection assays, although the first sequence, 156i, is responsible for most of this activity. The same targeted point mutation in the first binding site (nucleotide 4 of 7) also increases non-specific protein binding in a gel mobility shift assay while mutations in both sites abolish protein binding. The variant that we observed in the AGRE case lies in nucleotide 1 of the first

Table 1 Summary of variant detection in UC elements

UC element	Length (base pairs)	Variant	CVCR ($n=50$)	AGRE ($n=96$)	DPR ($n=124$)	DPR total (n)	Conserved with *Fugu* (%/base pairs)	Transmission
AUTS2								
UC218	286	47 A/G	1	1	1	150	61/273	*De novo* in CVCR?
UC219	210	7 C/T	0	0	1	148	0	
		61 T/G	1	0	1			
DLX6/5								
UC220	394	221 A/G	2	0	1	147	47/257	
UC221	501	61 C/G	0	0	(1)	126	0	
		129 A/G	0	1	0			Paternal; not in affected sib
		156 C/T	1	0	0			Paternal
FOXP2								
UC222	201	**159 A/C**	**1**	**0**	**1**	**146**	**84/201**	
UC223	268	**148 C/T**	**0**	**0**	**1**	**124**	**87/267**	
UC224	295	253 A/G	1	0	0	150	0	Not maternal; father not available
H–M[a] 100%	356	202 G/A	0	1	0	124		Maternal; not in affected sib
UC225	201	–	0	0	0	138	69/201	
		+28 A/G	0	1	0			Maternal; shared by both affected sibs
CNE_A[b]	222	–	0	0	0	140	95/192	
UC226	205	65 C/T	1	0	0	150	60/127	Maternal
		69 C/A	1	1	5			
		121 T/A	0	0	1			
CNE_B[c]	244	–	0	0	0	138	90.4/196	
UC227	231	–	0	0	0	124	71/231	

The variant column notes the nature of the variant and its position in the UC element. The CVCR, AGRE and DPR columns contain the number of variants found in the cases of autism from Costa Rica, the AGRE consortium and the first $n=124$ DPR controls, respectively. DPR total indicates the total number of DPR samples screened for that UC, although only one variant in UC221 (61 C/G) was found in cases numbered above 124 (DPR DNA was limiting so we could not amplify all 150 DPR samples for every UC). The *Fugu* column gives the percentage conservation of the UC element with *Fugu* over the number of base pairs conserved. The transmission column gives details regarding the parental origin of the variant, if the parents were available, and transmission to affected sibs. UC, ultraconserved; CVCR, Central Valley of Costa Rica; AGRE, Autism Genetic Resource Exchange; DPR, Polymorphism Discovery Resource Panel.

[a]H–M (chr7:113,665,296–113,665,651) indicates that the element is 100% conserved in the human and mouse, but not rat, in this case for 356 base pairs.

[b]CNE A (chr7:113,736,786–113,737,007).

[c]CNE B (chr7:113,856,310–113,856,553) are not UCs as they are 100% conserved in the mouse and human, but not rat. As they overlap with highly conserved human–*Fugu* CNEs (Woolfe *et al*., 2005), however, we considered them likely to harbor regulatory sequence and included these two non-UC elements in our screen. The bold UC elements contain variants at bases that are conserved down to *Fugu*.

Fig. 1

[GTAGCTCCCCAGGATCAATTCTGAACAAAGCCTCCAGCTGCAGTGCCATCCAATTTGAAGC] AGACATTGGGGACA
ATTTAAGGTTTTTATCCACAAGAAGGTTTTTTTCCATTCTCTTAAATGCAGCC**ATAATTAG**AGTAATTTTTCATGTA
GCCCGCTGATTACAGCGTTTTTACCGTCAAAG**ATAATTAC** [CTGTAATTTTCTTCCACTTTTAATACTAAAAAGCCA
TCTTTATTTAGATTCAGGAACAGGAAAGGCGAAACAAAAGAGGGAAATTATTCTGTTATTCATACACAAATTGCAGA
GACGTAGGAC] CTAAAATTGAAAATTAACCAAAATTATAATGCTG

UC 221 overlapping the DLX5/6 enhancer. Bold sequences represent DLX protein binding sites 156i and 156ii (Zerucha *et al*., 2000). Underlined nucleotides are conserved with zebrafish. This element is not conserved with *Fugu*. The first highlighted 'C' is in coding sequence on the opposite strand and is changed to a 'G' in a Polymorphism Discovery Resource Panel case. The second highlighted 'A' is changed to a 'G' in an autism case from the AGRE sample and lies in the first nucleotide of the 156i protein binding site. This variant is present in an unaffected father but not in an affected sibling and unaffected mother. The third highlighted, paternally inherited 'C' is changed to a 'T' in an autism case from Costa Rica. Multiple spliced ESTs support the existence of two alternatively spliced exons on the opposite strand (flanked by brackets).

seven-nucleotide sequence element and might be expected to alter function. Even if it did alter function, however, if these genes are imprinted in human brain as suggested by Horike *et al*. (2005), there might be no phenotypic effect because the variant is inherited from the father. Speculation regarding the phenotypic impact aside, identification of three rare variants in this element, and in UC226, suggests that ultra-conservation is maintained chiefly by a decreased tendency toward fixation, rather than a significantly lower mutation rate, consistent with the model discussed above.

In summary, we did not find variants unequivocally predisposing to autism in any of the 10 consecutive non-coding UC elements lying within or adjacent to candidate genes for autism on chromosome 7q. Further studies are needed to explore the extent, and functional impact, of variation in UC non-coding elements in relationship to either common or rare diseases.

Acknowledgement

We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Cure Autism Now and is supported, in part, by grant MH64547 from the National Institute of Mental Health to Daniel H. Geschwind (PI). We also wish to thank Kathryn Castelle for her gracious assistance with this manuscript.

References

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, *et al.* (2004). Ultraconserved elements in the human genome. *Science* **304**: 1321–1325.

Bonora E, Bacchelli E, Levy ER, Blasi F, Marlow A, Monaco AP, *et al.* (2002). Mutation screening and imprinting analysis of four candidate genes for autism in the 7q32 region. *Mol Psychiatry* **7**:289–301.

Bonora E, Beyer KS, Lamb JA, Parr JR, Klauck SM, Benner A, *et al.* (2003). Analysis of reelin as a candidate gene for autism. *Mol Psychiatry* **8**:885–892.

Bonora E, Lamb JA, Barnby G, Sykes N, Moberly T, Beyer KS, *et al.* (2005). Mutation screening and association analysis of six candidate genes for autism on chromosome 7q. *Eur J Hum Genet* **13**:198–207.

Collins FS, Brooks LD, Chakravarti A (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* **8**:1229–1231.

Folstein SE, Rosen-Sheidley B (2001). Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* **2**:943–955.

Gauthier J, Joober R, Mottron L, Laurent S, Fuchs M, De Kimpe V, *et al.* (2003). Mutation screening of FOXP2 in individuals diagnosed with autistic disorder. *Am J Med Genet* **118A**:172–175.

Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* **37**:31–40.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, *et al.* (2002). The human genome browser at UCSC. *Genome Res* **12**:996–1006.

Kimura M (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.

Kryukov GV, Schmidt S, Sunyaev S (2005). Small fitness effect of mutations in highly conserved non-coding regions. *Hum Mol Genet* **14**:2221–2229.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**:519–523.

Li H, Yamagata T, Mori M, Momoi MY (2005). Absence of causative mutations and presence of autism-related allele in FOXP2 in Japanese autistic patients. *Brain Dev* **27**:207–210.

Macdermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, *et al.* (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet* **76**:1074–1080.

McInnes LA, Gonzalez PJ, Manghi ER, Esquivel M, Monge S, Delgado MF, *et al.* (2005). A genetic study of autism in Costa Rica: multiple variables affecting IQ scores observed in a preliminary sample of autistic cases. *BMC Psychiatry* **5**:15.

Newbury DF, Bonora E, Lamb JA, Fisher SE, Lai CS, Baird G, *et al.* (2002). FOXP2 is not a major susceptibility gene for autism or specific language impairment. *Am J Hum Genet* **70**:1318–1327.

Ovcharenko I, Stubbs L, Loots GG (2004). Interpreting mammalian evolution using Fugu genome comparisons. *Genomics* **84**:890–895.

Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**:365–386.

Sultana R, Yu CE, Yu J, Munson J, Chen D, Hua W, *et al.* (2002). Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics* **80**:129–134.

Vincent JB, Herbrick JA, Gurling HM, Bolton PF, Roberts W, Scherer SW (2000). Identification of a novel gene on chromosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual. *Am J Hum Genet* **67**: 510–514.

Wassink TH, Piven J, Vieland VJ, Huang J, Swiderski RE, Pietila J, *et al.* (2001). Evidence supporting WNT2 as an autism susceptibility gene. *Am J Med Genet* **105**:406–413.

Wassink TH, Piven J, Vieland VJ, Pietila J, Goedken RJ, Folstein SE, *et al.* (2002). Evaluation of FOXP2 as an autism susceptibility gene. *Am J Med Genet* **114**:566–569.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, *et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* **3**:e7.

Zerucha T, Stuhmer T, Hatch G, Park BK, Long Q, Yu G, *et al.* (2000). A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J Neurosci* **20**:709–721.