

## ORIGINAL RESEARCH ARTICLE

Examination of *AVPR1a* as an autism susceptibility geneTH Wassink<sup>1</sup>, J Piven<sup>2</sup>, VJ Vieland<sup>1,3</sup>, J Pietila<sup>4</sup>, RJ Goedken<sup>3</sup>, SE Folstein<sup>5</sup> and VC Sheffield<sup>4</sup>

<sup>1</sup>Department of Psychiatry, University of Iowa College of Medicine, Iowa City, IA, USA; <sup>2</sup>Neurodevelopmental Disorders Research Center and Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; <sup>3</sup>Department of Biostatistics, Division of Statistical Genetics, University of Iowa College of Public Health, Iowa City, IA, USA; <sup>4</sup>Department of Pediatrics and the Howard Hughes Medical Institute, University of Iowa College of Medicine, Iowa City, IA, USA; <sup>5</sup>Department of Psychiatry, Tufts University School of Medicine, Boston, MA, USA

**Impaired reciprocal social interaction is one of the core features of autism. While its determinants are complex, one biomolecular pathway that clearly influences social behavior is the arginine-vasopressin (AVP) system. The behavioral effects of AVP are mediated through the AVP receptor 1a (AVPR1a), making the *AVPR1a* gene a reasonable candidate for autism susceptibility. We tested the gene's contribution to autism by screening its exons in 125 independent autistic probands and genotyping two promoter polymorphisms in 65 autism affected sibling pair (ASP) families. While we found no nonconservative coding sequence changes, we did identify evidence of linkage and of linkage disequilibrium. These results were most pronounced in a subset of the ASP families with relatively less severe impairment of language. Thus, though we did not demonstrate a disease-causing variant in the coding sequence, numerous nontraditional disease-causing genetic abnormalities are known to exist that would escape detection by traditional gene screening methods. Given the emerging biological, animal model, and now genetic data, *AVPR1a* and genes in the AVP system remain strong candidates for involvement in autism susceptibility and deserve continued scrutiny.**

*Molecular Psychiatry* (2004) 9, 968–972. doi:10.1038/sj.mp.4001503  
Published online 6 April 2004

**Keywords:** candidate gene; linkage; linkage disequilibrium; reciprocal social interaction; language

Autism is comprised of deficits in three domains: (1) speech and language, (2) reciprocal social interaction, and (3) repetitive, stereotyped motor, and behavioral activity. Autism also has a genetic basis, with estimates of heritability approaching 90%.<sup>1,2</sup> While the genetics of autism are complex and remain to be disentangled, it may be that each symptom domain has independent genetic determinants; supportive evidence for this hypothesis exists, for example, in the domain of speech and language.<sup>3,4</sup> One biomolecular pathway that clearly influences social behavior, and is therefore worthy of consideration in autism, is the arginine-vasopressin (AVP) system. The hormone AVP has been shown, across numerous organisms, to influence such varied social traits as vocal modulation,<sup>5</sup> sexual and parenting behavior,<sup>6</sup> mate affiliation and attachment,<sup>7</sup> aggression,<sup>8</sup> and social recognition.<sup>9</sup> The effects of AVP tend to be more pronounced in males, as opposed to the related hormone, oxytocin, whose effects tend to be stronger in females.<sup>7,10</sup>

The behavioral influence of AVP is mediated through the AVP receptor 1a (AVPR1a), as demonstrated most

clearly by studies comparing two related rodent species: montane and prairie voles.<sup>10</sup> Prairie voles are monogamous, highly social, and exhibit biparental care, while montane voles are nonmonogamous, comparatively asocial, and exhibit no parental care. Central administration of AVP in prairie voles induces prosocial behaviors, but has no effect on montane voles. Within prairie voles, this effect is limited to males, with oxytocin regulating female social behaviors.<sup>7,10</sup> The differential interspecies influence of AVP may be related to differences in the CNS tissue distribution of AVPR1a, which is weighted more heavily toward reward pathways in the prairie vole. These differences may, in turn, arise from a structural variant of the *AVPR1a* gene that appears to function as a tissue-specific promoter. While no interspecies coding sequence differences exist, the 5' promoter region of *AVPR1a* in the prairie vole has a 428 base pair sequence that is not found in the montane vole. Transgenic insertion of *AVPR1a* with this 5' fragment into mice induces social behavior and responsivity to centrally administered AVP resembling that of the prairie vole.<sup>11</sup> Based on these data, genes involved in the human AVP system, including *AVPR1a*, are reasonable candidates for autism susceptibility genes.

Further piquing our interest in *AVPR1a* were recent analyses of our linkage data. We previously

Correspondence: TH Wassink, MD, University of Iowa College of Medicine, Psychiatry Research/MEB, Iowa City, IA 52242, USA. E-mail: thomas-wassink@uiowa.edu  
Received 09 October 2002; revised 13 January 2003; accepted 25 February 2003

performed a genome-wide linkage screen in a sample of autism affected sibling pair (ASP) families, with loci on chromosomes 7q and 13q emerging as the strongest regions of interest.<sup>12</sup> We subsequently divided the sample into two groups—*language impaired vs language normal*—based on the degree of language impairment in the probands. The purpose of this was to reduce the apparent heterogeneity of autism, with the hypothesis that stronger linkage signals might arise from more homogeneously defined sets of families. We reanalyzed the 7q and 13q data, and found that virtually all of the linkage signal at both loci arose from the language-impaired families.<sup>4</sup> This split of these LOD scores was consistent with data from other sources at these loci,<sup>3,13</sup> supporting the efficacy of this subgrouping approach.

We have since analyzed all of our linkage data in this manner, and one interesting result arises on the region of chromosome 12q that contains the *AVPR1a* gene. *AVPR1a* is located on chromosome 12q at ~75 cM. In our total sample, under a dominant model of transmission, our strongest signal on chromosome 12 is a multipoint heterogeneity LOD of 0.85 at 78.1 cM (unpublished data). In the subgroup analyses, this value increases to 2.0 in the language normal group, while dropping off to nearly zero in the language-impaired group. Thus, it is plausible that this split of the LOD score might arise from autism-related variants of the *AVPR1a* gene, which would be expected to contribute to social aspects of autism as opposed to language impairment, and is therefore only revealed when the language-impaired subgroup is removed from the analyses.

The only published examination of *AVPR1a* in autism found suggestive evidence of linkage disequilibrium (LD); this study did not, however, report subgroup-based analyses.<sup>14</sup> The *AVPR1a* gene itself has 1472 nucleotides contained within two exons that are separated by a 2.2 kb intron; the gene codes for a 418 amino-acid protein with seven putative transmembrane domains. The mRNA contains a 5'-untranslated region (UTR) that is ~2 kb long, a 1 kb 3'-UTR, and the transcription start site is 1973 bp upstream from the translation start site.<sup>15</sup> Thibonnier *et al*<sup>16</sup> identified four nearby microsatellite repeats, three of which are in the 5'-flanking region. To test the role of *AVPR1a* and its 5' promoter region in autism, we examined the two closest 5' markers (the same markers tested by Kim *et al*) for linkage and LD in 65 autism ASP families and screened the entire coding sequence for variants in 125 independent autistic probands.

## Materials and methods

### *Ascertainment and patient sample*

All autistic probands and their families were ascertained and diagnosed through the Collaborative Linkage Study of Autism under a previously described protocol.<sup>12</sup> Briefly, families were recruited from three regions of the United States (Midwest, New England, and mid-Atlantic) through four clinical

data collection sites: the University of Iowa, Tufts University-New England Medical Center, Johns Hopkins University, and the University of North Carolina. All probands were at least 3 years old and were assessed with the Autism Diagnostic Interview-Revised (ADI-R)<sup>17</sup> and the Autism Diagnostic Observation Schedule (ADOS or the more recent ADOS-G<sup>18</sup>). ASP and trio families were recruited. All probands were required to meet ADI and ADOS criteria for autism. Probands were excluded if they had fragile X syndrome (based on fragile X DNA testing), a known chromosomal abnormality, or any other neurological or medical condition suspected to be associated with autism. The sample available for this study included 65 autism ASP families and other families from which an additional 60 independent autistic probands were derived. All individuals (or, when appropriate, their guardians) provided written, informed consent for participation in this study.

### *Mutation screening*

DNA was extracted from whole blood using standard procedures. Intron/exon boundaries were identified using the NCBI Locus Link utility (<http://www.ncbi.nlm.nih.gov/LocusLink/>). Exons, including flanking splice-junction sequences, were screened for coding sequence variants in each independent proband ( $n = 125$ ) using single-strand conformational polymorphism (SSCP) analysis on amplicons of 250 nucleotides or less (primers available upon request). SSCP in our laboratory has a sensitivity of 90% in detecting nucleotide changes when compared with direct sequencing. PCR of amplicons was performed with 20 ng of genomic DNA amplified in a reaction mixture containing 1.0  $\mu$ l of PCR buffer (100 mM Tris-HCl (pH8.8), 500 mM KCl, 15 mM MgCl<sub>2</sub>, 0.01% gelatin (w/v)), 200  $\mu$ M each of dATP, dCTP, dGTP, and dTTP, 2.5 pM of each primer, and 0.05 U of *Taq* DNA polymerase, increased to a final volume of 10.0  $\mu$ l with water. Samples were initially denatured at 94°C for 3 min, followed by 40 cycles of 94°C for 30 s, 54°C or 58°C for 30 s, and 72°C for 30 s. PCR products were electrophoresed on 6% nondenaturing polyacrylamide gels at 20 W for approximately 3 h at ambient temperature. The gels were then silver stained using standard protocols.<sup>19</sup>

Amplicons containing SSCP shifts were forward and reverse sequenced on an Applied Biosystems model 377 automated sequencer (ABI, Foster City, CA, USA) using dye terminator chemistry. Sequence data were analyzed using the Sequencher 3.1 gene analysis computer program (Gene Codes, Ann Arbor, MI, USA). Sequence variants detected in this manner were then sequenced in all available family members from the probands' families of origin and were also assessed in a similar manner (SSCP screening followed by direct sequencing of variants) in a comparison group of 160 nonautistic individuals of similar ethnicity.

### Linkage and LD

Testing for linkage and LD was performed for two polymorphisms in the 65 ASP families. The first polymorphism, named RS3, was a complex (CT)<sub>4</sub>-TT-(CT)<sub>8</sub>-(GT)<sub>24</sub> repeat located 3625 bp upstream of the transcription start site and amplified with primers TCCTGTAGAGATGTAAGTGC (forward) and gttctt-TCTGGAAGAGACTTAGATGG (reverse). The second, named RS1, was a (GATA)<sub>n</sub> repeat 553 bp upstream of the transcription start site, amplified with primers AGGGACTGGTTCTACAATCTGC (forward) and ACCTCTCAAGTTATGTTGGTGG (reverse). PCR was performed using the protocol described above, and products were electrophoresed on 6% polyacrylamide genotyping gels at 60 W for approximately 2 h. Gels were read blindly with respect to sample status by two independent raters. Homozygotes for each polymorphism were direct sequenced to determine allele sizes; alleles are reported as base pairs to enable comparison with Kim *et al.*<sup>14</sup>

Two-point heterogeneity LOD scores were calculated using the MLINK and HOMOG subroutines of the LINKAGE package<sup>20</sup> under one simple recessive model (disease allele frequency = 0.10, penetrance = 50%) and one simple dominant model (disease allele frequency of 0.04, penetrance = 40%). Multipoint heterogeneity LOD scores based on the two markers simultaneously (using the same recessive and dominant models) and the transmission disequilibrium test (TDT) statistic were calculated using Genehunter 2.1.<sup>20</sup> Both affected siblings were used in computing the TDT, and nominal significance levels were calculated for each allele.

For any test where an allele was differentially transmitted at a 0.05 level of significance, we also used GeneHunter's PERM1 routine as an additional estimate of the significance of the observed data. This test uses a permutation method that: (1) creates a new data set wherein each pair of transmitted/untransmitted alleles are arbitrarily reversed at  $P=0.50$ , (2) calculates the TDT statistic for this new data set, and (3) repeats this procedure a specified number of times (1000 in our case). The test then reports the number of replicates that produced a higher maximum  $\chi^2$  than the observed maximum  $\chi^2$ .<sup>21</sup>

In addition to examining our entire ASP sample in this manner, we split the families based on their language characteristics, as described above, and examined both resultant subgroups for linkage and LD. Families were classified as language impaired if neither proband had developed phrase speech by 36 months, and otherwise as language normal.<sup>22</sup> In all, 43 families were thus classified as language impaired and the remaining 22 as language normal. Furthermore, in the initial genome-wide screen, parental phenotypes were classified as unknown. For the language-related analyses, however, information with regard to parental language phenotype was available, which had been gathered by direct questioning. Thus for these analyses, parents with probable or definite delayed onset of speech, trouble learning to read, or

persistent trouble with spelling were classified as language impaired, and otherwise as language normal.

### Results

Screening of the coding sequence revealed only one variant, a conservative T→C substitution 408 nucleotides downstream from the start codon, leaving the amino-acid F136 unchanged. This SNP is a common variant identified as rs1042615 in dbSNP. Transmission of alleles from heterozygous parents for the two polymorphisms are shown in Tables 1 and 2. No evidence for linkage was found in either the pooled or language-impaired samples; all two- and multipoint LOD scores in those groups were <0.5 (data not shown). Modest evidence for linkage emerged, however, from the language normal group under the recessive model of inheritance, with two-point LOD scores of 0.68 ( $\alpha=0.44$ ,  $\theta=0.00$ ) for RS1 and 1.28 ( $\alpha=1.00$ ,  $\theta=0.14$ ) for RS3, and a maximum multipoint HLOD of 1.2 near RS3. The strongest evidence of LD also arose from the language normal families. In this group, two alleles from RS1 and one from RS3 were transmitted differently than expected by chance, with the permutation procedure supporting the significance of these results (Tables 1 and 2).

### Discussion

We have examined whether the gene *AVPR1a* contributes to autism susceptibility in our sample of families. Our coding sequence screening detected one conservative variant, indicating that differences in *AVPR1a* at the amino-acid level are unlikely to confer genetic vulnerability to autism. This supports the study of Kim *et al.*,<sup>14</sup> who detected a number of additional rare variants (these may have been either undetected in our screening or absent from our

**Table 1** TDT of RS1

Allele	Language impaired (n = 43 families)			Language normal (n = 22 families)			All (n = 65 families)		
	T	NT	$\chi^2$	T	NT	$\chi^2$	T	NT	$\chi^2$
308	2	2	0.00	1	1	0.00	3	3	0.00
312	11	20	2.61	3	12	5.40 <sup>a</sup>	14	32	7.04 <sup>c</sup>
316	36	29	0.75	15	18	0.27	51	47	0.16
320	28	31	0.15	19	9	3.57 <sup>b</sup>	47	40	0.56
324	16	14	0.13	5	5	0.00	21	19	0.10
328	18	15	0.27	5	3	0.50	23	18	0.61
332	0	2	2.00	0	0	0.00	0	2	2.00
336	6	4	0.40	2	2	0.00	8	6	0.29
				23 of 1000 replicates had a max $\chi^2 > 5.40$			30 of 1000 replicates had a max $\chi^2 > 7.04$		

<sup>a</sup> $P=0.02$ ,  $df=1$ .

<sup>b</sup> $P=0.06$ ,  $df=1$ .

<sup>c</sup> $P=0.008$ ,  $df=1$ .

**Table 2** TDT of RS3

Allele	Language-impaired subgroup (n = 43 families)			Language normal subgroup (n = 22 families)			All families (n = 65 families)		
	T	NT	$\chi^2$	T	NT	$\chi^2$	T	NT	$\chi^2$
318	2	2	0.00	0	0	0.00	2	2	0.00
322	1	1	0.00	0	0	0.00	1	1	0.00
324	0	4	4.00	0	0	0.00	0	4	4.00
326	0	0	0.00	0	2	2.00	0	2	2.00
328	5	7	0.33	13	3	6.25 <sup>a</sup>	18	10	2.29
330	16	15	0.03	5	1	2.67	21	16	0.68
332	30	32	0.06	19	26	1.09	49	58	0.76
334	24	19	0.58	11	14	0.36	35	33	0.06
336	15	13	0.14	11	9	0.20	26	22	0.33
338	17	22	0.64	5	5	0.00	22	27	0.51
340	4	6	0.40	1	5	2.67	5	11	2.25
342	3	3	0.00	1	1	0.00	4	4	0.00
344	10	4	2.57	5	3	0.50	15	7	2.91
346	2	2	0.00	2	4	0.67	4	6	0.40
348	3	2	0.20	0	0	0.00	3	2	0.20

40 of 1000 replicates had a max  $\chi^2 > 6.25$

<sup>a</sup> $P = 0.01$ ,  $df = 1$ .

sample), but nothing appeared to contribute substantially to autism susceptibility. Our analyses of the two 5' polymorphisms, however, suggest that variability outside the coding sequence might be involved; this evidence, as in our genome-wide study, arises primarily from families whose probands have a less severe impairment of language. In the language normal families, the marker RS3 demonstrated a two-point LOD score of 1.28 and a maximum multipoint HLOD of 1.2. The TDT analyses suggest the possibility that disequilibrium exists at both polymorphisms in these families. Allele 312 of the marker RS1 gave the strongest evidence for LD, although it was under instead of overtransmitted to affecteds, indicating that it may provide or be associated with a protective, as opposed to detrimental, effect. Furthermore, due to the suggestive linkage at RS3, we also ran the TDT for this marker in the language normal families using only one randomly chosen affected sibling per family, and found results consistent with the two-sibling analyses (see Table 3).

These findings must be interpreted, however, in light of a number of limitations. First, the alleles showing nonrandom transmission in our sample are different from the alleles highlighted in the Kim *et al*<sup>14</sup> study of AVPR1a, indicating either a different genetic background between the two samples or spurious false positives. It is also of note, however, that the Kim *et al* report did not incorporate language-based subgroup analyses, which provided the context for the emergence of our findings. Second, by performing the subgroup analyses, we introduce more

**Table 3** TDT of RS3 in language normal families using only one affected child per family

Allele	Language normal (n = 22 families)		
	T	NT	$\chi^2$
318	0	0	0.00
322	0	0	0.00
324	0	0	0.00
326	0	1	1.00
328	7	1	4.50 <sup>a</sup>
330	2	1	0.33
332	10	13	0.39
334	5	8	0.69
336	5	5	0
338	2	3	0.2
340	1	2	0.33
342	1	0	1
344	3	1	1
346	1	2	0.33
348	0	0	0.00

76 of 1000 replicates had a max  $\chi^2 > 4.50$

<sup>a</sup> $P = 0.03$ ,  $df = 1$ .

tests, leading to an increased probability of false-positive results. Lastly, we acknowledge that grouping our families based on their language characteristics is not an ideal approach for testing a gene hypothesized to be related to social function. Our reasons for this are two-fold: first, our measures of language function are reliable, valid, and at a more advanced stage of development than our measures of social function; and second, the removal of the language-impaired families will presumably result in greater homogeneity of the remaining language normal families. Therefore, given the emerging biological, animal model, and now genetic data, AVPR1a and genes in the AVP system remain strong candidates for involvement in autism susceptibility and deserve continued scrutiny.

## References

- Bailey A, Le CA, Gottesman I, Bolton P, Simonoff E, Yuzda E *et al*. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25**: 63–77.
- Folstein S. Twin and adoption studies in child and adolescent psychiatric disorders. *Curr Opin Pediatr* 1996; **8**: 339–347.
- Tomblin JB, Nishimura C, Zhang X, Murray JC. Association of developmental language impairment with loci at 7q3. *Am J Hum Genet (Suppl)* 1998; **63**: A312.
- CLSA. Incorporating language phenotypes strengthens evidence of linkage to autism. *Am J Med Genet* 2001; **105**: 539–547.
- Winslow J, Insel TR. Vasopressin modulates male squirrel monkeys' behavior during social separation. *Eur J Pharmacol* 1991; **200**: 95–101.
- Wang ZX, Liu Y, Young LJ, Insel TR. Hypothalamic vasopressin gene expression increases in both males and females postpartum in a biparental rodent. *J Neuroendocrinol* 2000; **12**: 111–120.
- Winslow JT, Hastings N, Carter CS, Harbaugh CR, Insel TR. A role for central vasopressin in pair bonding in monogamous prairie voles. *Nature* 1993; **365**: 545–548.
- Coccaro EF, Kavoussi RJ, Hauger RL, Cooper TB, Ferris CF. Cerebrospinal fluid vasopressin levels: correlates with aggression

- and serotonin function in personality-disordered subjects. *Arch Gen Psychiatry* 1998; **55**: 708–714.
- 9 Bluthé RM, Dantzer R. Chronic intracerebral infusions of vasopressin and vasopressin antagonist modulate social recognition in rat. *Brain Res* 1992; **572**: 261–264.
  - 10 Insel TR, Young LJ. Neuropeptides and the evolution of social behavior. *Curr Opin Neurobiol* 2000; **10**: 784–789.
  - 11 Young LJ, Nilsen R, Waymire KG, MacGregor GR, Insel TR. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* 1999; **400**: 766–768.
  - 12 CLSA. An autosomal genomic screen for autism. *Am J Med Genet* 1999; **88**: 609–615.
  - 13 Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P *et al*. A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet* 2002; **71**: 45–55.
  - 14 Kim SJ, Young LJ, Gonen D, Veenstra-VanderWeele J, Courchesne R, Courchesne E *et al*. Transmission disequilibrium testing of arginine vasopressin receptor 1A (AVPR1A) polymorphisms in autism. *Mol Psychiatry* 2002; **7**: 503–507.
  - 15 Thibonnier M, Auzan C, Madhun Z, Wilkins P, Berti-Mattera L, Clauser E. Molecular cloning, sequencing, and functional expression of a cDNA encoding the human V1a vasopressin receptor. *J Biol Chem* 1994; **269**: 3304–3310.
  - 16 Thibonnier M, Graves MK, Wagner MS, Chatelain N, Soubrier F, Corvol P *et al*. Study of V(1)-vascular vasopressin receptor gene microsatellite polymorphisms in human essential hypertension. *J Mol Cell Cardiol* 2000; **32**: 557–564.
  - 17 Lord C, Pickles A, McLennan J, Rutter M, Bregman J, Folstein S *et al*. Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *J Autism Dev Disord* 1997; **27**: 501–517.
  - 18 Lord C, Risi S, Lambrecht L, Cook Jr EH, Leventhal BL, DiLavore PC *et al*. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 2000; **30**: 205–223.
  - 19 Bassam BJ, Caetano-Anolles G, Gresshoff PM. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 1991; **196**: 80–83.
  - 20 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
  - 21 Markianos K, Daly MJ, Kruglyak L. Efficient multipoint linkage analysis through reduction of inheritance space. *Am J Hum Genet* 2001; **68**: 963–977.
  - 22 Folstein SE. Autism LOD on chromosome 7q was increased subsetting the sample on language acquisition (abstract). *Am J Hum Genet* 2000; **67**: 944.