# Supporting Information

## Celestino-Soper et al. 10.1073/pnas.1120210109

### SI Materials and Methods

**Sources of Human Samples.** For SSC, AGRE, and NIMH samples, DNA derived from LCLs was obtained from the Rutgers University Cell and DNA Repository. Blood from individuals from Nashville, TN, was used for establishment of LCLs, and DNA was extracted with Puregene chemistry on a Qiagen Autopure. Blood from adult normal white individuals from the BPR was used for the establishment of LCLs. For samples from the South Carolina Autism Project (SCAP), DNA was extracted from LCLs. For local samples from Houston, TX, DNA was extracted from blood. LCL-derived DNA was used for array CGH and PCR assay, unless otherwise specified. DNA for samples from the laboratories of three of the authors (S.W.S., C.B., and J.D.B.) was prepared as described previously (1).

For PCR analysis of exon 2 of *TMLHE*, we studied 1,887 male probands and 2,101 unaffected fathers from the SSC. The SSC enrolls young simplex cases of autism as described elsewhere (2). In addition, we studied simplex and multiplex (male-male) probands and siblings, as tabulated in Table 1, from the AGRE (http://research.agre.org/), the National Institute of Mental Health Human Genetics Initiative (NIMH-HGI) (www.nimhgenetics.org), local sources (Nashville, TN); the laboratories four of the authors (S.W.S, C.B., J.D.B., and M.E.H.), and the Paris Autism Research International Sibpair study to expand the data for exon 2 deletion. We tested additional unaffected autism fathers as listed in Table 1. We tested 897 male controls from the NIMH-HGI (https://www.nimhgenetics.org/nimh_human_genetics_initiative/); 80 autism males from the SCAP; 24 autism males from a collection in Houston, TX; 36 unaffected fathers from Houston, TX; and 49 control white males from the BPR. The laboratory of one of the authors (M.E.H.) analyzed 3,018 male controls from the Wellcome Trust Case–Control Consortium for the presence or absence of exon 2 of *TMLHE* with a genome-wide human SNP array 6.0 (Affymetrix). All individuals tested are enumerated in Table 1.

**Lymphoblast Cultures and Expression Analysis.** LCLs from AGRE and SSC individuals were obtained from the Rutgers University Cell and DNA Repository. An LCL from NA12003 (GM12003) (3) was obtained from the Coriell Institute for Medical Research (http://ccr.coriell.org). LCLs from adult normal white individuals from the BPR were obtained from the John W. Belmont and Tissue Culture Core laboratories, Department of Molecular and Human Genetics, Baylor College of Medicine. Cell pellets were frozen in Houston and shipped on dry ice to Amsterdam for enzyme assays.

LCLs were cultured in RPMI medium 1640 with L-glutamine (Lonza), 10% (vol/vol) FBS (Thermo Fisher Scientific), and 1% penicillin/streptomycin solution (Thermo Fisher Scientific). Cultures were incubated at 37 °C in a 5% carbon dioxide water-jacketed incubator. Cultures were grown in T75 tissue culture flasks (Greiner Bio-One) for 5 d in 40–50 mL of medium. For RNA extraction, 50 mL of the confluent culture was spun down for 10 min at 1,000 rpm in a Beckman-Coulter Allegra X-22 Centrifuge, the medium was aspirated off, and the pellet was frozen at −80 °C. For protein assays, 50 mL of the confluent culture was spun down for 10 min at 1,000 rpm, the medium was aspirated off, and the pellet was washed twice in 15 mL of sterile 1× PBS without calcium and magnesium (Cellgro; Mediatech, Inc.) following centrifugation for 10 min at 1,000 rpm. The pellet was resuspended in 1 mL of PBS, transferred to a 1.5-mL Eppendorf tube, and spun for 3 min at 1,000 rpm. The liquid was decanted, and the pellet was frozen at −80 °C.

RNA extraction from lymphoblast pellets was performed using the miRNeasy Mini Kit (Qiagen) following the manufacturer's instructions. For RT-PCR studies, 1 μg of RNA was converted to cDNA using the SuperScript III First Strand kit (Invitrogen) with 50 ng/μL random hexamers following the manufacturer's instructions. For quantitative RT-PCR studies, 100–500 ng of RNA was converted to cDNA using the High Capacity RNA-to-cDNA Master Kit (part no. 43907778; Applied Biosystems) following the manufacturer's instructions. The cDNA product (20 μL) was diluted fourfold and used for quantitative RT-PCR for each assay in triplicate.

Quantitative RT-PCR analyses of the levels of gene expression for *TMLHE* exons were performed using TaqMan Gene Expression Assays (Applied Biosystems): exon boundaries 1–2 (Hs00942999_m1) and 5–6 (Hs00379460_m1). Expression levels were endogenously normalized using TaqMan Gene Expression Assays for beta actin (ACTB) (Hs00357333_g1). Each reaction contained 1 μL of cDNA, 1× assay (primer mix), 1× TaqMan Universal PCR Mastermix (part no. 4324018; Applied Biosystems), and $H_2O$ in a total volume of 10 μL. The quantitative RT-PCR assay was performed in 384-well plates using the Applied Biosystems 7900HT Real-Time PCR System via standard cycling conditions: 95 °C for 10 min and 40 cycles of 95 °C for 15 s and 60 °C for 1 min. The data were analyzed using SDS RQ Manager software (Applied Biosystems) using the ΔΔCt protocol, after normalizing the expression level to the endogenous controls (ΔΔCt refers to data normalized endogenously to ACTB; and each assay was then normalized to HI0688).

PCR assay to detect exon 2 skipping in LCL cDNA of AGRE AU 0177 individuals was performed using the FastStart Taq DNA Polymerase, dNTPack (Roche Diagnostics). Briefly, 50–100 ng of cDNA was used in a 25-μL reaction that also contained 0.25 μM primers (Integrated DNA Technologies, Inc.) (Table S7), 200 μM each dNTP, 1 unit of FastStart Taq DNA Polymerase, and 10× PCR buffer containing magnesium chloride. The PCR assay was performed with the following reaction conditions: 95 °C for 6 min; 40 cycles of 95 °C for 30 s, 54 °C for 30 s, and 72 °C for 1 min; and a final extension at 72 °C for 7 min.

**Genomic DNA Analysis.** PCR assay to amplify *TMLHE* exons 1–8 was performed using the FastStart Taq DNA Polymerase, dNTPack. Briefly, 50–100 ng of genomic DNA was used in a 25-μL reaction that also contained 0.25 μM primers (Table S7), 200 μM each dNTP, 1 unit of FastStart Taq DNA Polymerase, and 10× PCR buffer containing magnesium chloride. The PCR assay was performed with the following reaction conditions: 95 °C for 6 min; 40 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min; and a final extension at 72 °C for 7 min. The PCR products were purified and sequenced using the standard protocol at a commercial service laboratory (GeneWiz, Inc.).

A long-range PCR assay to confirm the *TMLHE* deletion and to sequence the junction in several individuals was performed using the LA PCR kit (Takara Bio, Inc.). Briefly, 50–100 ng of genomic DNA was used in a 25-μL reaction that also contained 0.5 μM primers (Integrated DNA Technologies, Inc.) (Table S7), 400 μM each dNTP, 1.25 units of LA Taq (Takara Bio, Inc.), and 10× LA PCR buffer (Takara Bio, Inc.). The PCR assay was performed with the following reaction conditions: 94 °C for 1 min, 35 cycles of 94 °C for 30 s and 68 °C for 30–60 s per expected kilobases of extended DNA, and 72 °C for 10 min. PCR products were purified from the agarose gel using the Wizard SV Gel and PCR Clean-Up System (Promega) and sent

for nucleotide sequencing by Sanger di-deoxynucleotide sequencing (Macrogen USA).

**PCR Assay to Screen for Deletions.** A PCR assay to detect *TMLHE* exon 2 deletions using a multiplex assay with the addition of internal control primers that amplified exon 10 of *SLC22A5* was performed using the FastStart Taq DNA Polymerase, dNTPack. Briefly, 50–100 ng of genomic DNA was used in a 25-μL reaction that also contained 0.25 μM *TMLHE* primers and 0.15 μM *SLC22A5* primers (Integrated DNA Technologies, Inc.) (Table S7), 200 μM of each dNTP, 1 unit of FastStart Taq DNA Polymerase, 10× PCR buffer containing magnesium chloride, and 5.0 μL of GC-rich buffer. The PCR assay was performed with the following reaction conditions: 95 °C for 5 min and 10 cycles of 94 °C for 45 s, 65 °C for 45 s (with decrease of 1 °C per cycle), and 72 °C for 2 min. This was followed by 30 cycles at 94 °C for 45 s, 55 °C for 45 s, and 72 °C for 2 min. A final extension at 72 °C for 7 min was also performed.

**Phenotypic Analyses.** For all SSC cases, verbal IQ (VIQ), nonverbal IQ (NVIQ), and full-scale IQ (FSIQ) scores were derived from either the Differential Abilities Scales—Second Edition or the Mullen Scales of Early Learning. For all AU cases (AGRE), VIQ scores represent standard scores from the Peabody Picture Vocabulary Test, whereas NVIQ scores represent standard scores from the Raven Colored Progressive Matrices. Language level represents overall language abilities from the Autism Diagnostic Interview—Revised, where 0 = functional use of at least three-word phrases on a daily basis, 1 = speech used daily but not meeting criteria for a 0 code, and 2 = fewer than five words total or speech not used daily.

**Linkage Analyses.** Nonparametric linkage analysis was performed using samples from the AGRE data center, which were genotyped at the Children's Hospital of Philadelphia using the Illumina Infinium HumanHap 550 BeadChip. Only those families with at least two autistic male children with no known chromosomal abnormalities were included in the analysis. A total of 411 families consisting of 1,961 individuals with available genotype data were analyzed. Because it was hypothesized that *TMLHE* (Xq28) could be potentially involved in the genetic etiology of autism, we analyzed all available SNPs within the 2-megabase region that encompassed the *TMLHE* gene and all available SNP marker loci within the pseudoautosomal region for the *VAMP7* [Xq28 (pseudoautosomal region)] gene. Quality control of genotype data was performed using PEDCHECK (4) to identify Mendelian inconsistencies and MERLIN (5) to detect occurrences of double recombination events over short genetic distances, which are most likely attributable to genotyping error. Genotype frequencies were estimated from the AGRE data, but because all founders were genotyped, these estimated frequencies had no impact on the resulting nonparametric linkage (NPL) and LOD scores. Genetic map distances according to the Rutgers combined linkage-physical map of the human genome Build 37 version (6) were used to carry out the multipoint analysis. For markers that are not on the Rutgers map, the physical map position from the human reference sequence (Build 37) was used to interpolate the genetic map positions. Multipoint nonparametric linkage analysis was performed using MERLIN for the gene regions located within the pseudoautosomal region (*VAMP7*), and MINX, a version of MERLIN that was developed specifically to perform linkage analysis of the X chromosome, was used to analyze the *TMLHE* gene region. The results of the nonparametric analysis are based on the NPL-all scoring function (7), which takes into consideration within-pair as well as between-pair allele sharing for a given pedigree. The NLP scores and LOD scores were both derived using Kong and Cox exponential and linear models (8).


1. Pinto D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466:368–372.
2. Fischbach GD, Lord C (2010) The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron* 68:192–195.
3. McCarroll SA, et al.; International HapMap Consortium (2006) Common deletion polymorphisms in the human genome. *Nat Genet* 38:86–92.
4. O'Connell JR, Weeks DE (1998) PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266.
5. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—Rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101.
6. Matise TC, et al. (2007) A second-generation combined linkage physical map of the human genome. *Genome Res* 17:1783–1786.
7. Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. *Biometrics* 50(1):118–127.
8. Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179–1188.

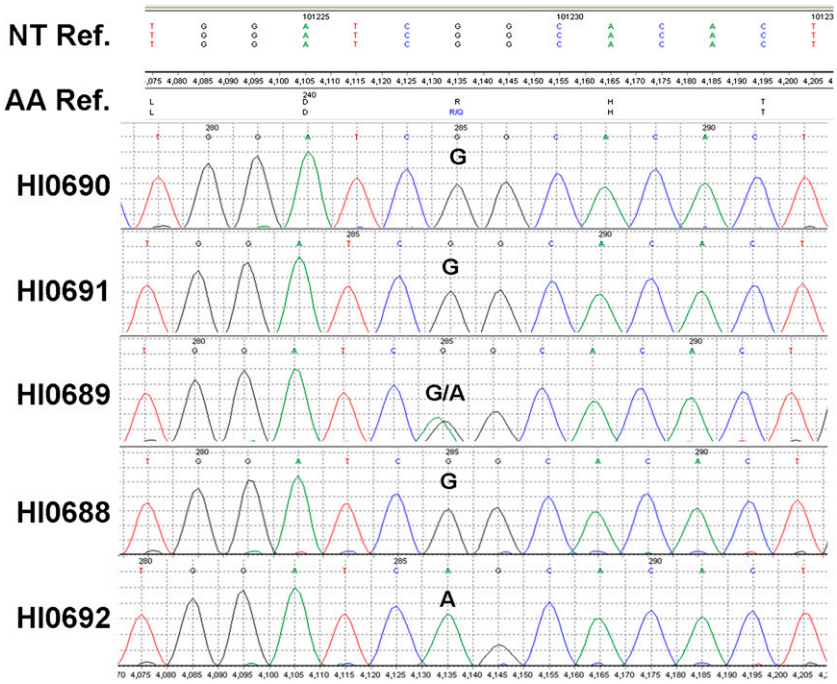


**Fig. S1.** Sequencing for the R241Q mutation in AGRE family AU 0177. The mother (HI0689) and her unaffected son (HI0692) from AGRE family AU 0177 have an R241Q (G-to-A nucleotide) missense mutation. The two affected males (HI0690 and HI0691) and the father (HI0688) do not have the mutation. AA Ref., amino acid sequence of reference; HI, AGRE individuals; NT Ref., nucleotide sequence of reference.

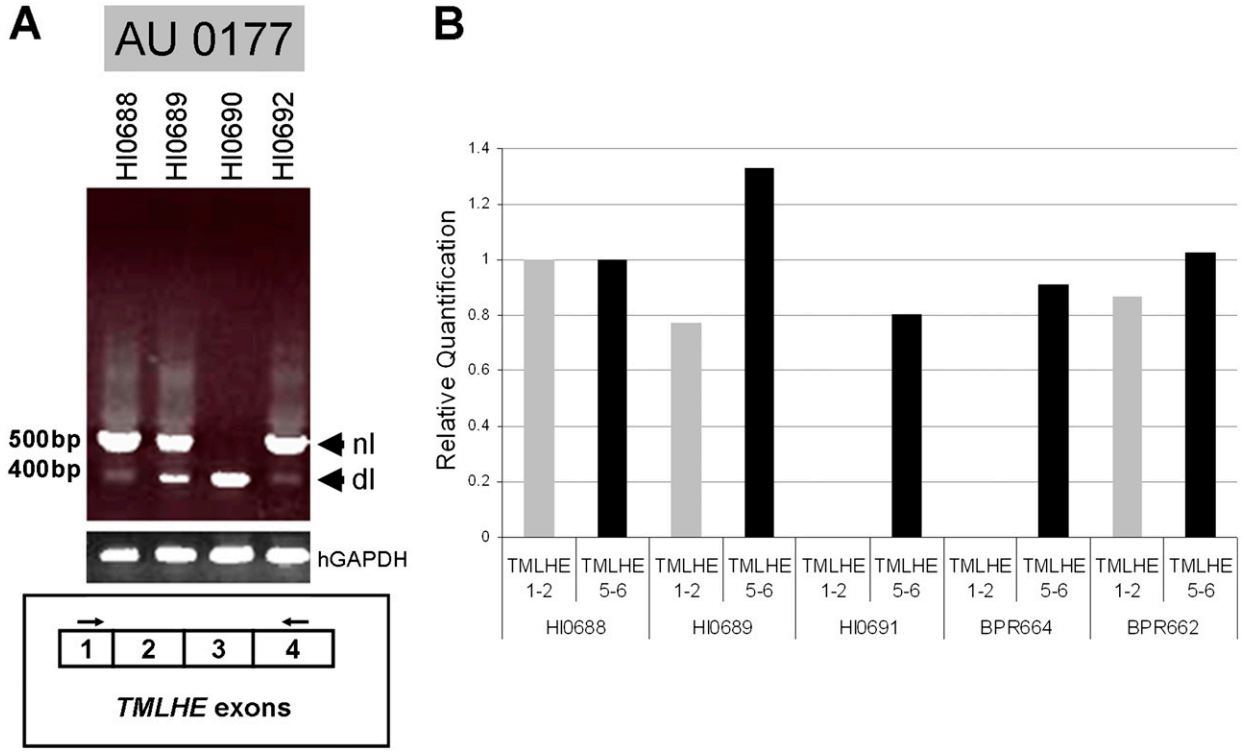


**Fig. S2.** *TMLHE* exon 2 deletions produce stable mRNA. (*A*) RT-PCR assay from LCLs of AGRE family AU 0177 showed the presence of a strongly expressed transcript containing exon 2 in all cases except for HI0690 (affected), who has deletion of exon 2. HI0689 (mother) is heterozygous for the deletion. A low level of exon 2 skipping is seen in the unaffected individuals (HI0688 and HI0692). Arrows represent primers. (*B*) Quantitative RT-PCR assay from LCLs of AGRE family AU 0177 and from BPR individuals normalized to β-actin (endogenous control) and unaffected father HI0688. All samples show detection of the exon 5–6 product. Individuals with hemizygous exon 2 deletion (HI0691 and BPR664) show undetectable *TMLHE* exon 1–2 product, compared with HI0688. The vertical axis shows the relative quantification or ΔΔCt (*SI Materials and Methods*), and the horizontal axis shows the *TMLHE* primer assay and samples tested. dl, deletion, hGAPDH, human GAPDH control; nl, normal.

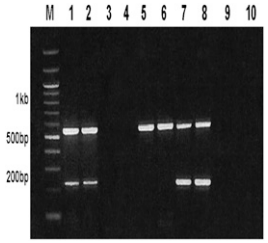


**Fig. S3.** Multiplex PCR assay for detection of *TMLHE* exon 2 deletion. (*Upper band*) *SLC22A5* exon 10 primers. (*Lower band*) *TMLHE* exon 2 primers (*SI Materials and Methods*). Lanes 1 and 2 show DNAs with no *TMLHE* exon 2 deletion. Lanes 3 and 4 show DNAs of poor DNA quality that fail both PCR reactions. Lanes 5 and 6 show DNAs of individuals with *TMLHE* exon 2 deletion. Lanes 7 and 8 show DNAs with no *TMLHE* exon 2 deletion. Lanes 9 and 10 show negative controls (water). M, marker.

**Table S1. *TMLHE* exon 2 deletions identified in this study**

| Identification | Intron 1 deletion | Coordinates (hg19) | Size, bp |
|---|---|---|---|
| Simplex probands | | | |
| SSC 11000.p1 | No | ChrX:154770740-154784574 | 13,834* |
| SSC 11144.p1 | No | ChrX:154770452-154784444 | 13,992 |
| SSC 11229.p1 | Yes | ChrX:154770326-154785986 | 15,660 |
| SSC 11680.p1 | No | ChrX:154762300-154778227 | 15,927* |
| SSC 13489.p1 | Yes | ChrX:154772351-154784444 | 12,093 |
| SSC 13928.p1 | Yes | ChrX:154734433-154794037 | 59,604 |
| SSC 14243.p1 | No | ChrX:154772351-154784444 | 12,094 |
| CANAGEN 157936 | Yes | ChrX:154772351-154784444 | 12,094 |
| CANAGEN 138147 | Yes | ChrX:154772351-154778095 | 5,745 |
| CANAGEN 85278 | Yes | ChrX:154772351-154778095 | 5,745 |
| Multiplex probands | | | |
| HI0181_AU0305 | Yes | ChrX:154772351-154779283 | 6,932 |
| HI0690_AU0177 | Yes | ChrX:154771166-154778207 | 7,041* |
| HI0841_AU0585 | Yes | ChrX:154771801-154778190 | 6,389* |
| HI3170_AU1031 | Yes | ChrX:154772351-154778095 | 5,744 |
| HI3689_AU1163 | No | ChrX:154762244-154778095 | 15,851 |
| NIMH-152-MM0057-3 | Yes | ChrX:154772351-154784444 | 12,093 |
| NIMH-63-8701 | Yes | ChrX:154771801-154778190 | 6,389* |
| Unaffected controls | | | |
| BPR664 | Yes | ChrX:154772351-154784444 | 12,093 |
| NA12003 | Yes | ChrX:154771321-154781153 | 9,832* |
| NIMH 150-10442 | Yes | ChrX:154771801-154778190 | 6,389* |
| NIMH 150-11126 | Yes | ChrX:154771801-154778190 | 6,389* |
| NIMH 150-12165 | Yes | ChrX:154771801-154778190 | 6,389* |
| SSC 11088_Fa | Yes | ChrX:154771801-154778190 | 6,389* |
| SSC 12414_Fa | Yes | ChrX:154771653-154785478 | 13,825* |
| SSC 12762_Fa | Yes | ChrX:154772351-154778095 | 5,744 |
| SSC 13700_Fa | Yes | ChrX:154770937-154778210 | 7,273* |
| SSC 13985_Fa | Yes | ChrX:154771801-154778190 | 6,389* |
| SSC 13988_Fa | Yes | ChrX:154772351-154778095 | 5,744 |
| SSC 14217_Fa | Yes | ChrX:154772351-154784444 | 12,094 |

CANAGEN, Canadian Genetic; ChrX, chromosome X; Fa, father; HI and AU, DNA and family identifiers for AGRE participants; NA, HapMap North American family identifier; p1, proband; NIMH, NIMH participant; SSC, SSC participant.

*Breakpoints determined by sequencing; other breakpoints based on array data.

**Table S2. Deletion breakpoints defined by PCR assay and sequencing**

| Identifier | Sequence | Features | Repeats at breakpoint | |
|---|---|---|---|---|
| | | | Left | Right |
| SSC 11000.p1 | L: **TGCACGTGTACC**CTAGAA<br>S: **TGCACGTGTACCC**CAGAA<br>R: TGCACATGTACC**CCAGAA** | 7-bp microhomology | LINE (L1PA7) | LINE (L1PA6) |
| NA12003 | L: **TACACTCCCAC**CAACACTGT<br>S: **TACACTCCCACCAACAGTG**<br>R: TACATTCCCAC**CAACAGTGT** | 11-bp microhomology | LINE (L1PA7) | LINE (L1MC) |
| SSC 12414.fa | L: **TGCA**...**ACCT**...CAGTCTAC<br>S: **TGCA**...**ACCT**...**CAATCTAT**<br>R: TCCA...ACCT...**CAATCTAT** | 64-bp homology | LINE (L1PA7) | LINE (L1PA6) |
| SSC 11680.p1 | L: **GTTCTTGTC**CTAGGGTGT<br>S: **GTTCTTGTCCTACTAAAA**<br>R: ACCCTGTCT**CTACTAAAA** | 3-bp microhomology | None | SINE (AluSg) |
| NIMH 05C39576A, NIMH 04C34199A, NIMH 04C28495A, NIMH-63-8701, HI0841_AU0585, SSC 13985.fa, SSC 11088.fa | L: **CTGAATGGT**ATTGCCTAG<br>S: **CTGAATGGTTTGAGACCA**<br>R: GTCAGCAGT**TTGAGACCA** | No microhomology | LINE (L1PA7) | SINE (AluSg) |
| SSC 13700.fa | L: **CTCCCACTT**ATGAGTGAG<br>S: **CTCCCACTTATGGTGAAA**<br>R: CTGACTAAC**ATGGTGAAA** | 3-bp microhomology | LINE (L1PA7) | LINE (L1M5) |
| AU 0177 (HI0690 and HI0691) | L: **TAAACA**TATGTGTG<br>S: **TAAACA**AGAGTTCGAGACC AGCATGGCC**AACATGGT**<br>R: CTGACT**AACATGGT** | No microhomology; extra nucleotides | LINE (L1PA7) | SINE (AluSg) |

Bold nucleotides are derived from junction sequence; underlined nucleotides represent microhomology segments; nucleotides in blue font represent genomic sequences not present in the junction sequence; nucleotides in pink font represent a short interspersed element (SINE). fa, father; AU and HI, DNA and family identifiers for AGRE participants; L, left breakpoint region; LINE, long interspersed element; NA, HapMap North American family identifier; NIMH, NIMH participant; p1, proband; R, right breakpoint region; S, fragment of sequenced PCR product containing breakpoints; SSC, SSC participant; STR, simple tandem repeat.

**Table S3. Nonsynonymous point mutations in *TMLHE* exons 1–8**

| Collection | DNA ID | Nucleotide | Protein | Inheritance |
|---|---|---|---|---|
| Autism families | | | | |
| SSC | 13793.p1 | 154754266 G > A | R70H | Maternal |
| AGRE AU-0177 | Mother/unaffected half-brother | 154741370 G > A | R241Q | Unknown/maternal |
| SSC | 12353.p1 | 154736695 G > A | E287K | Maternal (also in unaffected sister) |
| SSC | 12041.p1 | 154722107 G > A | V349I | Unknown |
| Control males | | | | |
| NIMH | 04C26633A | 154754255 G > A | V74I | Unknown |

Results are from sequencing genomic DNA from 98 AGRE male probands from male-male sib pairs, 536 SSC male probands, and 443 NIMH male controls. One male proband (13793.p1) was identified by analysis of carnitine metabolites in plasma. AU, DNA identifier for AGRE participant; NIMH, NIMH participant; SSC, SSC participant.

**Table S4. Phenotypic information on affected males with deletion of exon 2**

| Individual | VIQ/NVIQ/FSIQ | Age, y* | Regression | Seizures | Language level† | Comments |
|---|---|---|---|---|---|---|
| Simplex probands | | | | | | |
| SSC 11000.p1 | 50/78/65 | 9.2 | No | No | 0 | ADOS Mod 2 at age 9.2 y |
| SSC 11144.p1 | 99/127/111 | 10.8 | No | No | 0 | ADOS Mod 3 at age 10.9 y |
| SSC 11229.p1 | 56/68/63 | 9.4 | No | No | 0 | ADOS Mod 3 at age 9.4 y |
| SSC 13489.p1 | 94/99/97 | 7.5 | No | No | 0 | ADOS Mod 3 at age 7.5 y |
| SSC 13793.p1 (R70H) | 48/56/52 | 4.2 | Yes | No | 1 | ADOS Mod 1 at age 4.2 y |
| SSC 13928.p1 | 27/53/38 | 6.3 | No | No | 1 | ADOS Mod 1 at age 6.3 y |
| SSC 14243.p1 | 61/74/70 | 6.3 | Yes | No | 0 | ADOS Mod 2 at age 6.3 y |
| CANAGEN 157936 | 89/109/102 | 5 | No | No | 0 | ADOS Mod 3 at age 5 y |
| CANAGEN 138147 | 121/—/115 | 11.7 | No | No | 0 | ADOS at age 8.3 y |
| CANAGEN 85278 | Unavailable | | | | | |
| Multiplex families | | | | | | |
| HI0691 (AU 0177-03) | Unavailable | 4.9 | Yes | Yes | 2 | ADOS Mod 1 at age 7.3 |
| HI0690 (AU 0177-04) | 92/—/— | 2.9 | No | No | 2 | ADOS Mod 2 at age 5.2 |
| HI3170 (AU 1031-01) | 111/122/— | 7.8 | No | — | 0 | ADOS Mod 3 at age 7.8 |
| HI 3165 (AU 1031-303) | —/untestable/— | 4.3 | No | — | 2 | ADOS Mod 1 at age 4.3 |
| HI3689 (AU 1163-304) | —/88/— | 8.4 | No | — | 0 | ADOS Mod 2 at age 8.2 |
| HI3691 (AU 1163-305) | —/110/— | 6.9 | No | — | 1 | ADOS Mod 1 at age 6.9 |
| HI0842 (AU 0585-03) | 92/80/— | 4.5 | No | — | 0 | ADOS Mod 3 at age 5.7 |
| HI0841 (AU 0585-04) | —/untestable/— | 3.2 | No | No | 2 | ADOS Mod 1 at age 4.4 |
| HI0181 (AU 0305-3) | 121/114/— | 7.7 | No | No | 0 | ADOS Mod 3 at age 11.1 |
| HI0182 (AU 0305-4) | —/untestable/— | 6 | No | No | 2 | ADOS Mod 1 at age 9.4 |
| NIMH-152-MM0057-3 | 158/128/143 | 6 | No | No | 0 | ADOS Mod 4 at age 17 |
| NIMH-152-MM0057-4 | 87/120/104 | 3 | No | No | 1 | Premodular ADOS at age 4 |
| NIMH-63-701 | —/44/— | 4 | No | — | 2 | — |

ADOS Mod, Autism Diagnostic Observation Schedule modified; HI and AU, DNA and family identifiers for AGRE participants; NIMH, NIMH participant; p1, proband; SSC, SSC participant; FSIQ, full-scale IQ; —, data not available or not collected.

*Age for probands/cases = age (in years) at the time of the Autism Diagnostic Interview—Revised (ADI-R) evaluation.

†Language level represents overall language abilities from the ADI-R, where 0 = functional use of at least three-word phrases on a daily basis, 1 = speech used daily but not meeting criteria for a 0 code, and 2 = fewer than five words total or speech not used daily.

**Table S5. Phenotypic information on control males with deletion of exon 2**

| Fathers/controls | Highest education | Age, y* | BAP-Q | SRS-ARV | FHI-I | Comments |
|---|---|---|---|---|---|---|
| SSC 11088_Fa | BA/BS | 48.7 | 4.75/2.33/4.00/3.69 | 15 | 0 | — |
| SSC 12414_Fa | BA/BS | 39.1 | 1.17/2.58/1.17/1.64 | 18 | 0 | — |
| SSC 12762_Fa | Graduate/ professional | 40 | 3.17/1.25/4.00/2.81 | 19 | 0 | — |
| SSC 13700_Fa | Associate | 40.3 | 3.25/1.92/2.33/2.50 | 37 | 0 | — |
| SSC 13985_Fa | BA/BS | 49.8 | 1.50/1.67/2.50/1.89 | 36 | 2 | — |
| SSC 13988_Fa | Associate | 35.3 | 1.58/2.92/2.42/2.31 | 11 | 0 | — |
| SSC 14217_Fa | Some high school | 32.3 | 1.0/1.17/1.5/1.22 | 9 | 3 | — |
| NIMH 150-10442 | MA/MS | 60 | — | — | — | Alcoholism, maximum BMI = 38.7 |
| NIMH 150-11126 | Associate | 52 | — | — | — | Prescription sedative/ painkiller dependence, maximum BMI = 21.0 |
| NIMH 150-12165 | Some college | 63 | — | — | — | Maximum BMI = 35.3 |
| NA12003 | — | 97 | — | — | — | — |
| BPR664 | — | — | — | — | — | — |

BAP-Q, Broad Autism Phenotype Questionnaire; BMI, body mass index; Fa, father; FHI-I, Interviewer's Impression of Interviewee form (International Molecular Genetic Study of Autism Consortium, total score); NA, HapMap North American family identifier; NIMH, NIMH participant; SRS-ARV, Social Responsiveness Scale—Adult Research Version (total score); SSC, SCC participant; —, data not available or not collected.
*Age for fathers/controls = age (in years) at the time of the Autism Diagnostic Interview—Revised (ADI-R) evaluation.

**Table S6. Calculation of penetrance based on hypothetical population sample of 2 million***

| | Females | Males | 1 in 350 control males deleted | 1 in 250 autism males deleted | 1 in 150 autism males deleted |
|---|---|---|---|---|---|
| Normal | 996,000 | 984,000 | 2,811 | | |
| Autism | 4,000 | 16,000 | | 64 | 107 |
| Penetrance | | | | $\frac{64}{(2811+64)} = 2.2\%$ | $\frac{107}{(2811+107)} = 3.7\%$ |

*Assume equal number of males and females, assume 1 in 100 incidence of autism, assume 4:1 male/female ratio, assume 1 in 350 control males are deleted for *TMLHE* exon 2, and assume either 1 in 250 or 1 in 150 autism males deleted.

**Table S7. Primers for several assays**

| Primer name | Individuals | Gene/exon | Assay | Primer sequence (5′— 3′) |
|---|---|---|---|---|
| *TMLHE* Af | SSC 11000_P1 | *TMLHE* 2 | LR PCR | CTACAGAAACAGAGAGATGGCAGTATCACAG |
| *TMLHE* Ar | SSC 11000_P1 | *TMLHE* 2 | LR PCR | GTGGTGATTCCTCAAGGATCTACAACTAGAAT |
| *TMLHE* Bf | AGRE AU0177 | *TMLHE* 2 | LR PCR | TTTCTCCTAATGCTATCCCTCCCCTAGACC |
| *TMLHE* Br | AGRE AU0177 | *TMLHE* 2 | LR PCR | GATGTGCAAGACCTCTATACTTGCTATTGTCA |
| *TMLHE* Cf | SSC 11680_P1 | *TMLHE* 2 | LR PCR | GTCTTGAGCTTTACAGGAGAGAAGCAGAATG |
| *TMLHE* Cr | SSC 11680_P1 | *TMLHE* 2 | LR PCR | GATGTGCAAGACCTCTATACTTGCTATTGTCA |
| *TMLHE* Df | SSC 12414_Fa | *TMLHE* 2 | LR PCR | CTCCAGAAAGCACCCCTACAGACATAGAGA |
| *TMLHE* Dr | SSC 12414_Fa | *TMLHE* 2 | LR PCR | CTCTCATCTTCTAGAGACTGCCTGCCTTAC |
| *TMLHE* Ef | NA12003 | *TMLHE* 2 | LR PCR | TTTCTCCTAATGCTATCCCTCCCCTAGACC |
| *TMLHE* Er | NA12003 | *TMLHE* 2 | LR PCR | GCACTTGAGGGAGGAAGATAAATGGAGTTAAC |
| *TMLHE* Ff | NIMH 05C39576A, NIMH 04C34199A, NIMH 04C28495A, NIMH-63-8701, HI0841_AU0585, SSC 13985.fa, SSC 11088.fa, SSC13700.fa | *TMLHE* 2 | LR PCR | TTTCTCCTAATGCTATCCCTCCCCTAGACC |
| *TMLHE* Fr | Same as *TMLHE* Ff | *TMLHE* 2 | LR PCR | GATGTGCAAGACCTCTATACTTGCTATTGTCA |
| *TMLHE* 1F | | *TMLHE* 1 | PCR | GTAAAACGACGGCCAGTAACAGAAGCTCCTCCAGTGC |
| *TMLHE* 1R | | *TMLHE* 1 | PCR | CAGGAAACAGCTATGACCGCCAGGGTTAAGGTGAGTA |
| *TMLHE* 2Fc | | *TMLHE* 2 | PCR | GTAAAACGACGGCCAGTCAATTTCTCCTTTGTTTAATCCTC |
| *TMLHE* 2Rc | | *TMLHE* 2 | PCR | CAGGAAACAGCTATGACGGTCCTGTCAACTCCACAAAA |
| *TMLHE* 3F | | *TMLHE* 3 | PCR | GTAAAACGACGGCCAGTCCCCTCAACCCTTATTGCTT |
| *TMLHE* 3R | | *TMLHE* 3 | PCR | CAGGAAACAGCTATGACCTTGCCATGTTCCCTGTCTT |
| *TMLHE* 4F | | *TMLHE* 4 | PCR | GTAAAACGACGGCCAGTTCATGCATACCACTCAAACCA |
| *TMLHE* 4R | | *TMLHE* 4 | PCR | CAGGAAACAGCTATGACCGAGGGTGACCATACCTGTT |
| *TMLHE* 5F | | *TMLHE* 5 | PCR | GTAAAACGACGGCCAGTTCCACAGTTTTCTCCCCAAC |
| *TMLHE* 5R | | *TMLHE* 5 | PCR | CAGGAAACAGCTATGACCTCCCCAGAAGAAGATGCTG |
| *TMLHE* 6F | | *TMLHE* 6 | PCR | GTAAAACGACGGCCAGTCAGGGCCCAGCACATATAA |
| *TMLHE* 6R | | *TMLHE* 6 | PCR | CAGGAAACAGCTATGACTTCTTTGGTTGGTGGCTATTT |
| *TMLHE* 7F | | *TMLHE* 7 | PCR | GTAAAACGACGGCCAGTACCTCAGCTTGGTGAGGAGA |
| *TMLHE* 7R | | *TMLHE* 7 | PCR | CAGGAAACAGCTATGACGATGAGGCTTGCTTCCAGAG |
| *TMLHE* 8F | | *TMLHE* 8 | PCR | GTAAAACGACGGCCAGTAACAGAAGCTCCTCCAGTGC |
| *TMLHE* 8R | | *TMLHE* 8 | PCR | CAGGAAACAGCTATGACTTTGTGGGATGGGATAAGGA |
| *TMLHE* rt1Fb | | *TMLHE1* | RT-PCR | CTCCTTGCTTGGAATTCTGC |
| *TMLHE* rt4R | | *TMLHE4* | RT-PCR | TTCAGTCCCTCGTTGGTTTC |
| *TMLHE* ex2F1 | | *TMLHE* 2 | Multiplex PCR | CACCTACACAGCAGGCTTCA |
| *TMLHE* ex2R1 | | *TMLHE* 2 | Multiplex PCR | TATGTGCTGCCAAAGCAAAC |
| *OCNT2* ex10F | | *SLC22A5* 10 | Multiplex PCR | CTGCTGCAGGATTCTCTTCC |
| *OCNT2* ex10R | | *SLC22A5* 10 | Multiplex PCR | TGCTTCTCACATTCCTCCAG |

Fa or fa, father; HI and AU, DNA and family identifiers for AGRE participants; LR, long-range; NA, HapMap North American family identifier; P1, proband; SSC, SSC participant; STA.