Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE

C Nava1,2,3,4, F Lamari5,28, D Héron1,6,7,8,28, C Mignot6,8,4, A Rastetter1,2,3, B Keren9, D Cohen10,11, A Faudet1, M Huguet17,18,19, S Caillet4, B Leheup20, M Leboyer21,22,23,24, C Gillberg25, R Delorme26, T Bourgeron17,18,19, A Brice1,2,3,4 and C Deplienne1,2,3,27

Introduction

Autism spectrum disorders (ASD) constitute a common but heterogeneous group of neuropsychological disorders characterized by impairment of social interactions and communication, stereotyped behaviors and restricted interests. Although they are probably the most heritable of psychiatric conditions, with a concordance rate of 80–90% in monozygotic twins versus 10–20% in dizygotic twins, few genes have been reliably associated with ASD.1–3 Recent studies have highlighted the vast heterogeneity and complexity of the genetics of these disorders. All mutations or copy number variants (CNVs) associated so far with ASD have been rare, with minor allele frequencies <1%. A few de novo or inherited CNVs, some of which are recurrent, such as duplications of the 15q11-q13 or 7q11.23 and deletions of 16p11.2 regions, were shown to confer a highly penetrant risk of autism.4–7

More recently, de novo mutations in various highly interconnected genes were shown to contribute to ASD, suggesting that abnormalities in different genes could converge to alter common pathways.8–10 Abnormalities in at least two pathways were repeatedly related to ASD: the first includes...
mutations in \textit{TSC1/TSC2}, \textit{NF1} or \textit{PTEN} in the mTOR (mammalian target of rapamycin) pathway; the second is illustrated by mutations in \textit{NLGN3–4}, \textit{SHANK1–3} and \textit{NRXN1}, all of which encode synaptic proteins.\textsuperscript{11–13}

A striking feature of ASD is the excess of affected males, with a sex-ratio disequilibrium of 4:1 that reaches 10:1 in patients with normal cognitive abilities.\textsuperscript{14} This suggests that genes located on sex chromosomes contribute to the etiology of ASD or that the penetrance of autistic traits depends on sex determinants such as hormones. In favor of the first hypothesis, mutations in \textit{NLGN4X} and \textit{NLGN3} on chromosome X have been identified in a few families with ASD.\textsuperscript{15} Additionally, the analysis of all or selected genes located on the X chromosome successfully identified new candidate genes for intellectual disability (ID),\textsuperscript{16} ASD and schizophrenia.\textsuperscript{17} Interestingly, the risk of recurrence of ASD is significantly increased in families with two affected sibs, reaching 32\% and more when both affected subjects are males.\textsuperscript{18} This suggests that highly penetrant forms of ASD with autosomal recessive or X-linked inheritance have been overlooked.

To test the hypothesis that yet undiscovered X-linked genes are associated with highly penetrant forms of ASD, we selected 12 unrelated families with at least two affected males compatible with X-linked inheritance and analyzed all of the coding regions on the X chromosome.

\textbf{Material and methods}

\textbf{Patients.} The entire exome of chromosome X was sequenced in 12 families with two affected males with ASD or ID compatible with X-linked inheritance, recruited from the ‘Centre de Référence Déficiences Intellectuelles de causes rares’ (Pitié-Salpêtrière Hospital; Supplementary Figure S1). Index cases were evaluated by specialized geneticists and pediatric neurologists and/or child psychiatrists. Patients were assessed with the Autism Diagnostic Interview-Revised. Nine index cases had autism with ID and three had Asperger syndrome or high-functioning autism based on DSM IV-TR (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision) criteria. Clinical features of the index cases and the affected relatives are detailed in Supplementary Clinical Table. Normal results had Asperger syndrome or high-functioning autism based on Autism Diagnostic Interview-Revised. Nine index cases had autism with ID and three had Asperger syndrome or high-functioning autism based on DSM IV-TR (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision) criteria. Clinical features of the index cases and the affected relatives are detailed in Supplementary Clinical Table. Normal results were previously obtained by karyotyping, searches for fragile-X syndrome, microarray analysis (CytoSNP-12, Illumina, San Diego, CA, USA) and sequencing of \textit{NLGN3–4X} and \textit{SHANK3} when appropriate, as well as metabolic screening (including at least creatine and guanidinoacetate analysis).

For \textit{TMLHE} screening, a cohort of 161 patients (134 patients with autism and ID and 27 patients with Asperger syndrome) recruited at the Pitié-Salpêtrière Hospital (Centre de référence Déficiences Intellectuelles de causes rares or Centre référent diagnostic autisme, Paris, France) and 340 patients from the PARIS (Paris Autism Research International sib pair) study (including 194 patients with autism and ID and 59 patients with Asperger syndrome) were included, for a total of 501 unrelated male patients with ASD. In addition, 765 healthy male controls from North Africa (\(n = 320\)), Europe (\(n = 350\)) and Lebanon (\(n = 95\)) were included to test the new variants.

The analysis of microrearrangements in \textit{TMLHE} included 178 additional patients with ASD from the PARIS studies previously included in Autism Genome Project (AGP, http://www.autismgenome.org/)\textsuperscript{12} and 896 healthy male individuals. The control groups included 371 European male subjects from La Pitié-Salpêtrière hospital, 142 from other European laboratories and 383 control individuals from the Study on Addiction Genetics and Environment (\(n = 371\)) and HapMap CEPH Utah (\(n = 12\)) series.\textsuperscript{7} Raw intensities and genotypes were obtained from NHGRI-dbGaP (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1). The sub-set of control data set used in the specific CNV analyses in this paper is composed of control samples that passed all quality control filters (Log R ratios s.d. = 0.27; B allele frequency s.d. = 0.13; Call Rate >0.99). Informed written consent was obtained from each individual or his/her parents before blood sampling. All experiments were performed in accordance with French guidelines and rules.

\textbf{Next-generation sequencing.} Next-generation sequencing was performed at Integragen SA (Evry, France). Regions of the X chromosome corresponding to coding and 3′–5′ untranslated region (UTR) sequences were captured from genomic DNA using a custom Agilent SureSelect Target Enrichment methodology (Agilent, Santa Clara, CA, USA) with the biotinylated oligonucleotide probe library, followed by paired-end 75b massively parallel sequencing on Illumina GAIIx (Illumina). For detailed explanations of the process, see Gnirke \textit{et al.}\textsuperscript{19} Sequence capture, enrichment and elution were performed precisely according to the manufacturer’s instructions and protocols. Briefly, 3\(\mu\)g of each genomic DNA was fragmented by sonication to yield fragments of 150–200 bp and then purified. Paired-end adapter oligonucleotides from Illumina were ligated on repaired A-tailed fragments, then purified and enriched by six PCR cycles. Then 500 ng of these purified libraries were hybridized to the SureSelect oligo probe capture library for 24 h. After hybridization, washing and elution, the eluted fraction was PCR-amplified for 10–12 cycles, purified and quantified by quantitative PCR to obtain sufficient DNA template for downstream applications. Each eluted-enriched DNA sample was sequenced on an Illumina GAIIx as paired-end 75b reads. Image analysis and base calling was performed using Real Time Analysis Pipeline version 1.9 (Illumina) with default parameters.

\textbf{Bioinformatics analysis.} Sequencing data was analyzed according to the Illumina pipeline (CASAVA1.7) and aligned with the Human reference genome (Hg19) using the ELANDv2 algorithm. Variant annotation (RefSeq gene annotation) identification of known polymorphisms (referenced in dB SNP or 1000 Genome) and analysis of the position and consequences of the variants (for example, exonic, intronic, silent and nonsense), was determined with an in-house pipeline from the positions included in the bait coordinates. The frequencies (in the homozygous or heterozygous state) were determined from all exomes sequenced at Integragen and from exome results provided by HapMap. Results per sample were obtained in tabulated text files, and coverage/depth statistical analyses were performed for each
bait. The 13,464 single nucleotide polymorphisms (SNPs) and 1532 indels in the 12 male index cases included 1467 SNPs (10.9%) and 590 indels (38.5%) predicted to be at the heterozygous state. Eleven of these variants were tested by Sanger sequencing and proved to be false-positives. Further analysis focused, therefore, on variants predicted to be hemizygous. A total of 171 hemizygous variants were tested and confirmed by Sanger sequencing. The strategy used for selecting potentially pathogenic variants is detailed in Figure 1. They: (i) were located in chromosome X regions common to the two affected sibs (4331 SNPs and 299 indels in 11 families); (ii) had with a minor allele frequency <1% in dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/), Exome variant server (http://evs.gs.washington.edu/EVS/) and 29 other exomes; (iii) were found in genes expressed in brain according to the Unigene (http://www.ncbi.nlm.nih.gov/unigene) or Uniprot (http://www.uniprot.org/) databases; and (iv) were predicted to have an impact on the gene or the protein (nonsense variants, missense variants, predicted at least once in silico to be deleterious, and synonymous, intronic or 5–3'UTR variants with possible effects on splice sites or promoters using Alamutv2.1/AlamutHT). For variants present in at least two index cases, only those segregating in all the affected members of all families were retained. Mutation interpretation and amino-acid conservation in orthologs and paralogs were assessed using the Alamutv2.1/AlamutHT softwares (Interactive Biosoftware, Rouen, France). Prediction of pathogenicity was assessed using PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (scale-invariant feature transform) (http://sift.bii.a-star.edu.sg/), Mutpred (http://mutpred.mutdb.org/) and SNPs&GO (http://snps-and-go.biocomp.unibo.it/snps-and-go/). Frequencies were compared with the Fisher's exact test.

High-density SNP arrays. Index cases and affected relatives were screened using Illumina cytoSNP-12 arrays to search for CNVs and identify regions on the X chromosome shared by the affected sibs. Illumina microarray experiments were automated and performed at the P3S platform (Pitié-Salpêtrière Hospital), according to the manufacturer's specifications (Illumina, San Diego, CA, USA). Image acquisition was performed using a BeadArray Reader (Illumina). Image data analysis and automated genotype calling was performed using GenomeStudio2011.1 (Illumina). Genomic positions were based on the UCSC and Ensembl Genome Browsers. Genotypes on chromosome X were analyzed in affected relatives of each family to identify shared X regions (with the exception of family 12 for which SNP microarray data were unavailable for the affected uncle). Shared regions were defined as identical genotypes spanning at least 2 Mb.

For the analysis of TMLHE microrearrangements, control individuals (n=525) and ASD cases (n=356) were genotyped using Illumina Human 1M-single BeadChip arrays. Samples were processed using the manufacturer’s recommended protocol, and BeadChips were scanned on the Illumina BeadArray Reader using default settings. Analysis and intra-chip normalization were performed using Illumina’s BeadStudio software v3.3.7, with a GenCall cutoff of 0.1. The quality-control criteria were selected: the Array with call rate >95%; standard deviation for log R ratio values in the autosomes <0.35; and standard deviation of the B Allele frequency values (that is, allelic ratios within the 0.25–0.75 ranges) >0.13. For the samples that passed the above SNP and intensity quality-control filters, we used the QuantSNP20 and visualizer SnapPeep CNV calling algorithms. The required data for CNV analysis, that is, within-sample normalized fluorescence (that is, X and Y normalized values), between-sample normalized fluorescence (that is, Log R ratios and B allele frequency values) and genotypes for each sample, were exported directly from Illumina’s Beadstudio software. We excluded CNVs when they failed stringent quality control criteria: <5 consecutive probes covering 1 kb of sequence were merged using outside probe boundaries (that is, union of the CNVs) and low confidence score log Bayes factor <15.

Sanger sequencing. Specific primer pairs were designed to amplify 182 variants detected by next-generation sequencing. In addition, eight primer pairs were designed to amplify the coding exons and adjacent intron–exon boundaries of the TMLHE gene. Primer sequences are provided in supplemental data. Forward and reverse sequence reactions were performed with the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Life Technologies Corporation, Carlsbad, CA, USA) using the same primers. G50-purified sequence products were run on an ABI 3730 automated sequencer (Applied Biosystems) and data were analyzed with Seqscape v2.6 software (Applied Biosystems). The mutation nomenclature is based on the TMLHE cDNA reference sequence (NM_018196.3).

Quantitative multiplex PCR and long-range PCR. The presence of a deletion of exon 2 in TMLHE was tested in the 501 patients with ASD and 371 male control individuals by quantitative multiplex PCR. One hundred and seventy-eight patients with ASD were also screened using Illumina Human 1M-single BeadChip arrays with the same results. Two primer pairs were used in the quantitative multiplex assay: one specific of exon 2 of TMLHE (final concentration: 0.5 µM) and one amplifying exon 2 of GPR128 (final concentration: 0.08 µM). PCR conditions were as follows: 96 °C 5 min, 20 cycles: 94 °C 30 s, 60 °C – 0.5 °C/cycle 30 s, 72 °C 40 s, and 15 cycles: 94 °C 30 s, 50 °C 30 s, 72 °C 40 s, followed by 7 min at 72 °C. PCR products were quantified on a Caliper LabChip system (Caliper Life Sciences, Hopkinton, MA, USA). In addition, the presence of the deletion of exon 2 in TMLHE was confirmed by long-range PCR using the SequalPrep Long PCR kit (Invitrogen, Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer’s recommendations. The PCR conditions were as follows: 2 min at 94 °C, 10 cycles: 94 °C 10 s, 58 °C 30 s, 68 °C 18 min (1 min kb⁻¹), 25 cycles: of 94 °C 10 s, 58 °C 30 s, 68 °C 18 min (1 min kb⁻¹)+20 s/cycle, followed by 5 min at 72 °C. Frequencies were compared with the Fisher’s exact test.

Cell culture and mRNA experiments. Lymphoblasts from the two affected brothers and the mother of family 9 were isolated from peripheral blood cells using standard procedures. Fibroblasts were taken from skin cells of the affected
brothers. Lymphoblastic cells and fibroblasts were pre-
treated, or not, overnight with 10 μg ml⁻¹ emetin, an inhibitor of
nonsense-mediated decay.

Total RNA from lymphoblasts and fibroblasts was isolated
using the Qiagen RNeasy Mini kit (Invitrogen). cDNAs were
synthesized using the SuperScript III First-Strand Kit (Invitrogen).
The reverse-transcribed TMLHE cDNA was amplified and sequenced using specific primers located in exons 2 (forward) and 4 (reverse). The PCR products were run on 2% agarose gels. TMLHE mRNA was quantified using the Qiagen QuantiTect primer assays for
TMLHE of (15 s at 95 °C and 30 s at 60 °C) was run in triplicate on a Lightcycler 480 (Roche, Applied
Sciences, Penzberg, Germany). Forty-five two-step cycles
were performed and analyzed using Lightcycler 480 software release 1.5.0. Relative
abundance was calculated using the formula

\[ r = \frac{TMLHE}{\text{control gene}} \]

where \( \Delta \Delta C_t = (C_t \text{ Gene tested} - C_t \text{ control genes}) - TMLHE - (C_t \text{ Gene tested} - C_t \text{ control genes}) \) PPIA.

**Chromatography and mass spectrometry.** A mixture of internal standard solution was prepared by dissolving
30 μM of carnitine-(N-methyl-d3) (Cambridge Isotopes Lab, Andover, MA, USA) and 15 μM of ε-N-trimethyl-([13C3])-L-
lysine (Sigma-Aldrich, St Louis, MO, USA) in methanol. In
all, 20 μL of internal standard mixture was added to 30 μL of plasma or urine after mixing, 100 μL of methanol was added and mixed with vortex for protein precipitation. The mixture
was incubated in ice for 15 min and centrifuged at 15000 g for
10 minutes at +4 °C. The supernatant was transferred into the vials and 5 μL were injected into tandem LCMS/MS system. Calibration curves were performed by serial dilution of stock solution containing 46.8 μM of l-carnitine and 20 μM of ε-N-trimethyllysine (Sigma-Aldrich) in methanol. An Acquity UPLC (ultra performance liquid chromatography) chromatographic system equipped with a BEH C18 RP column (1.7 μm, 50 mm × 2) maintained at 45 °C was coupled to a TQD (tandem quadrupole detector) MS/MS system
(Waters, Guyancourt, France) and used as an LCMS/MS system
(liquid chromatography–mass spectrometry/mass spectrometry)
for trimethyllysine (TML) and carnitine measurement.

The mobile phases were: eluent A, ultrapure water; eluent B, acetonitrile. The elution gradient was as follows: flow rate 0.8 ml min⁻¹, 0–1 min, 0% A; 1–1.2, 0–100% A; 1.2–2 min, 100%; 2–2.2, 100–0% A; 2.2–5, column equilibration with 100% B. The detector was used in multiple-reaction monitoring to detect the transition of a specific precursor to daughter ions 189.1/84.1 and 162.2/103.1 for N⁵-trimethyllysine and carnitine, respectively.

**Dietary assessment.** In order to determine prospective and retrospective dietary intakes, the patient and their parents had an interview with a dietitian. A dietary questionnaire listing all the food and beverages consumed by the patient for a period of 3 days was completed by the parents. Nutritional intakes were estimated with the DSMS software. Carnitine intakes were calculated based on the reference table of the Linus Pauling Institute Oregon State University.

**Table 1** Summary of all SNPs and indels detected by exome analysis of chromosome X in the index cases of families 1–12 and those located in chromosome X regions shared by families 1–11

| Family | Nb SNPs (novel) | Nb indels (novel) | Linked regions (Mb) | SNPs in linked regions | Indels in linked regions |
|--------|-----------------|------------------|---------------------|------------------------|-------------------------|
|        |                 |                  |                     | Total Nonsense (novel) | Missense (novel) | Synonymous (novel) | Non-coding (novel) | Total Exons (novel) | Non-coding (novel) |
| Family 1 | 1233 (54)       | 108 (45)         | 109                 | 828 1 (0)               | 122 (3) 137 (3) | 568 (23)          | 77 5 (1) | 72 (30) |
| Family 2 | 1039 (40)       | 82 (37)          | 37                  | 369 1 (0)               | 59 (3) 48 (1)   | 261 (8)           | 19 0       | 19 (6)  |
| Family 3 | 871 (16)        | 56 (22)          | 33                  | 251 1 (0)               | 36 (0) 40 (1)   | 174 (2)           | 15 0       | 15 (9)  |
| Family 4 | 861 (27)        | 67 (23)          | 57                  | 340 1 (0)               | 72 (4) 58 (1)   | 209 (2)           | 21 2       | 19 (8)  |
| Family 5 | 892 (30)        | 81 (34)          | 76                  | 412 1 (0)               | 57 (1) 66 (0)   | 288 (15)          | 36 2       | 34 (14) |
| Family 6 | 989 (33)        | 83 (37)          | 27                  | 319 0                   | 54 (2) 46 (2)   | 219 (7)           | 28 3       | 25 (14) |
| Family 7 | 829 (18)        | 66 (27)          | 36                  | 200 1 (0)               | 37 (1) 41 (0)   | 121 (2)           | 14 2       | 12 (6)  |
| Family 8 | 1459 (63)       | 113 (60)         | 42                  | 614 1 (0)               | 71 (3) 101 (1)  | 441 (22)          | 42 4       | 38 (24) |
| Family 9 | 1098 (35)       | 82 (33)          | 81                  | 656 1 (1)               | 93 (2) 80 (2)   | 472 (21)          | 23 1       | 22 (7)  |
| Family 10 | 835 (31)       | 63 (24)          | 15                  | 77 0                    | 6 (0) 10 (1)   | 61 (2)            | 7 0        | 7 (1)   |
| Family 11 | 905 (24)        | 73 (32)          | 53                  | 265 3 (0)               | 38 (1) 38 (1)   | 186 (6)           | 17 1       | 16 (5)  |

Abbreviation: SNP, single-nucleotide polymorphism.
Linked regions and the total number of SNPs and indels in linked regions are indicated in bold.

**Results**

**Chromosome X exome sequencing.** The coding regions of all genes on chromosome X, including 5′–3′ UTRs, were sequenced in 12 index cases using next-generation sequencing (pedigrees are shown in Supplementary Figure S1). A mean number of 1000 SNPs (829–1459) and 78 indels (56–113; Table 1 and Supplementary Table SA and SB) were identified per patient. In parallel, chromosome X regions common to the affected sibs, ranging from 15 to 109 Mb, were identified in 11 families using Illumina cytoSNP-12 arrays (Supplementary Figure S2). The mean numbers of SNPs and indels located in shared X regions were 394 (77–828) and 27 (7–77) per patient, respectively. Further analysis focused on variants that were (i) absent or rare in databases (minor allele frequency <1%), (ii) located in genes expressed in brain and (iii) predicted to be deleterious (Figure 1). Thirty-eight possibly deleterious variants
(mean number per family: 3.2 range: 0–9), all confirmed by Sanger sequencing, were detected (Table 2). Analysis of matched control populations excluded two variants that had a frequency >1%. The variants were present in 15/22 asymptomatic male relatives. Altogether, these results identified 36 rare, possibly deleterious variants in 33 different genes in 9 families.

In two families, the variants were in genes previously implicated in ID.21–24 Both variants, c.2904_2906del/p.Ser969del in PHF8 and c.2849T>A/p.Val950Asp in HUWE1, affect amino-acids that were highly conserved during evolution and were not found in a large control population or reported in Hapmap, 1000 Genomes and the Exome Variant Server (Figure 2). Alternatively, mutations in PHF8 causing a loss-of-function were previously identified in patients with ID and cleft lip/palate,21–23 the p.Ser969del variant segregated with high-functioning autism without other clinical features in family 8. In the index case of family 4, p.Val950Asp in HUWE1 was predicted to be deleterious by SIFT, Polyphen-2 and Mutpred algorithms. Surprisingly, this variant was not found in the proband’s brother, who was less severely affected, and turned out to have occurred de novo in the proband. We hypothesized that Val950Asp in HUWE1 contributed to a genetically complex disorder by increasing the severity of the phenotype. Alternatively, the phenotype of the brother could have had a different etiology. These results illustrate the complexity of inheritance in ASD, in which a combination of rare inherited and de novo events can contribute to the disorder.6–10

Among the remaining variants, a nonsense mutation (c.229C>T/p.Arg77X) in TMLHE, encoding e-N-trimethyllysine hydroxylase, the enzyme catalyzing the first step of carnitine biosynthesis from TML, segregated with autism and moderate ID in family 9 (Figures 3a and b). The p.Arg77X mutation was absent from 508 healthy male controls and databases. This study identified several additional rare variants that might contribute to ASD such as c.521C>A/p.Ala174Asp (family 1) in ODZ1, which encodes teneurin-1, a transmembrane protein expressed in the developing central nervous system that might have a role in neuronal connectivity.26 In addition, 11 variants located in introns or UTRs of genes expressed in the brain and predicted to have a possible effect on gene expression were found in five families. Five of these genes are involved in axon guidance (PLXNA3, PLXNB3, KAL1) or neurotransmission (SYN1, GABRE), four regulate transcription, splicing or translation (TXLNG, TSPYL2, AFF2, involved in brain development, and RBM3, involved in RNA processing, regulation of translation and production of miRNA);26 the remaining genes are involved in ubiquitination (KLHL13) or in protein transport (BCAP31).

Finally, no rare variants meeting the criteria defined in Material and methods were found in three families (families 3, 7 and 10), suggesting that genetic factors in these families are possibly located in unexplored regions of the X chromosome, on autosomes, or were present with a frequency >1%.

Screening of TMLHE and functional consequences of mutations. To investigate the effect of the p.Arg77X mutation at the mRNA level, we performed quantitative reverse transcriptase–PCR analysis in lymphoblasts and fibroblasts from the two affected brothers of family 9 and their mother. The mutated mRNA was significantly downregulated in the cells of the affected sibs. Pretreatment of the cells with emetin restored the mRNA levels, indicating that this downregulation corresponded to the degradation of the mutated mRNA by nonsense-mediated decay (Figure 3e). Interestingly, two mRNA isoforms were detectable in the
Table 2. Summary of rare, possibly deleterious variants located in shared regions on chromosome X

| Index case | Position (Mb) | Gene   | NM_number | Nucleotide change | Amino-acid change | Mutation type | rs number (dbSNP 135) | rsMAF(dbSNP) | GRANtham distance | SIFT | PolyPhen2 | Mutpred | SNP&GO | Controls | Variant found in healthy man in the family | Causative status |
|-----------|--------------|--------|-----------|------------------|------------------|---------------|----------------------|--------------|------------------|------|-----------|---------|--------|----------|------------------------------------------|------------------|
| Family 1  | 3228527      | MXRA5  | NM_015419 | c.7987G>T        | p.Gly2663Trp     | Missense      | rs143264549          | 0.002        | 1.08             | neutral     | 0/0.95 | NA       | possible |          |          |                                  | NA               |
| Family 2  | 17750042     | NHS    | NM_198270 | c.4351G>C        | p.Ala1451Pro     | Missense      | rs139106444          | 0.003        | 1.03             | neutral     | 0/0.33 | NA       | possible |          |          |                                  | unlikely         |
| Family 3  | 3240238      | MXRA5  | NM_015419 | c.3488G>A        | p.Arg1163His     | Missense      | rs139106444          | 0.004        | 1.00             | neutral     | 0/0.16 | NA       | possible |          |          |                                  | NA               |
| Family 4  | 151092546    | MAGEA4 | NM_001079858 | c.670C>T          | p.Val208Phe     | Missense      | rs143264549          | 0.002        | 1.08             | neutral     | 0/0.95 | NA       | possible |          |          |                                  | NA               |
| Family 5  | 54956008     | TRO    | NM_001039705 | c.2851G>A       | p.Gly951Ser     | Missense      | rs143264549          | 0.002        | 1.08             | neutral     | 0/0.95 | NA       | possible |          |          |                                  | NA               |
| Family 6  | 151123143    | GABRE  | NM_004961 | c.30G>T          | p.Asp546Asn     | Missense      | rs143264549          | 0.002        | 1.08             | neutral     | 0/0.95 | NA       | possible |          |          |                                  | NA               |

Abbreviation: MAF, minor allele frequency; SIFT, scale-invariant feature transform; UTR, untranslated region.

This list includes variants that are rare or absent from databases (MAFs < 1%), are found in genes expressed in brain and have a possible impact on gene or protein function (nonsense, missense, predicted to be deleterious by at least one bioinformatics algorithm or in an intron, 5'UTR or 3'UTR with a predicted effect on a splice site or promoter). Mutpred: score < 0.5: benign; 0.5–0.75: possibly disease-associated; and > 0.75: probably disease-associated. Grantham distance: conservative (0–50), moderately conservative (51–100), moderately radical (101–150) or radical (> 151).
patients and their heterozygous mother, one with a premature termination codon in exon 3 and one missing exon 3, which restored the reading frame (Figure 3c). Exon 3 skipping could have resulted from nonsense-associated altered splicing, a mechanism alternative to nonsense-mediated decay.27

To confirm that mutations in TMLHE are associated with ASD, we screened its coding sequence in 501 male probands with ASD. Two missense substitutions (c.730G>C/p.Asp244His and c.1107G>T/p.Glu369Asp) were found in two additional unrelated patients with ASD but not in 330 controls and not reported in databases. Aspartic acid 244 and glutamic acid 369 are highly conserved in other species (Figure 3d). Aspartic acid 244 is predicted to bind the 2-oxoglutarate cofactor and constitutes one of the three key residues of the catalytic core,28 suggesting
a complete loss-of-function of the protein with this substitution.

To further analyze the consequences of the TMLHE mutations, we assayed carnitine and TML, the precursor of carnitine biosynthesis, in plasma and urine from the brothers with the p.Arg77X mutation and the patient with the p.Asp244His variant for whom biological samples were available. Carnitine was slightly but not significantly decreased in the plasma of the patients; the values remained in the normal range (Figure 3f). By contrast, mass spectrometry revealed a significant 2–3-fold increase in the TML precursor in the plasma of all the three patients with the Arg77X and Asp244His mutations compared with controls (Figure 3g). Carnitine intake, estimated from a 3-day dietary recall questionnaire, was normal in all patients (estimated at 55, 80 and 96 mg j−1, respectively).

Very few variants in TMLHE predicted to have deleterious effects are present in genetic databases (HapMap, 1000 Genomes); in particular, no nonsense mutations have been identified and only 18 non-synonymous variants have been reported on >8700 X chromosomes in the Exome Variant Server. Furthermore, non-synonymous variants are far more frequent in females than in males (n = 22/3381 versus n = 3/1998, if we exclude the Asn235Thr variant specific of the African population), indicating that variations in TMLHE are not well tolerated and are liable to be pathogenic. Mutations in TMLHE were found in patients with autism and ID but not in patients with Asperger syndrome. If we exclude patients with Asperger syndrome, the difference between point mutations in TMLHE in male patients with autism and mental retardation and male individuals from Exome variant server is significant (P = 0.05).

Interestingly, Celestino-Soper et al.29 have recently reported that the deletion of exon 2 of TMLHE is a CNV present in ~1/350 males and associated with ASD with a low penetrance. To compare the frequency of this CNV in healthy individuals and patients with ASD, we specifically assessed its presence in 896 unrelated healthy male controls and 691 patients with ASD by quantitative multiplex assay or Illumina Human 1M-single BeadChip arrays (see Material and methods). This study revealed the presence of one deletion of exon 2 of TMLHE in a single subject out of 896 male controls, whereas it was present in 3 out of 691 male patients with ASD (P = 0.3). Interestingly, one of the three patients with the deletion of exon 2 had an affected brother who did not carry the deletion. In addition, a duplication encompassing exon 1 of TMLHE, absent from 525 healthy controls, was detected by Illumina Human 1M-single BeadChip arrays in an additional patient with ASD (P = 0.4). Altogether, these results suggest that deficits in TMLHE could be rarer than previously reported. These deficits seem to be more frequent in patients with ASD, suggesting that they constitute susceptibility factors for autism, although the difference did not reach significance for microarrangements in TMLHE.

Discussion

This study focused on rare variants on chromosome X in multiplex families with ASD compatible with X-linked inheritance. Altogether, this study identified 36 possibly deleterious variants in 33 genes, including PHF8, HUWE1 and TMLHE. Variants in genes common to at least two families were exceptional, confirming that ASD is genetically highly heterogeneous.

The X chromosome contains the largest number of genes expressed in the brain.30 For this reason, mutations causing monogenic forms of ID have been identified in numerous genes on chromosome X.31 Interestingly, almost all genes involved in ASD, such as NLGN3/4X on chromosome X or SHANK3 on chromosome 22qter, are also mutated in patients with ID without autistic features.11,31 In this study, variants in two genes were previously implicated in ID: PHF8, which encodes a histone lysine demethylase that regulates rRNA synthesis32 and retinoic acid-induced neuronal differentiation,33 and HUWE1, which encodes an E3 ubiquitin-protein ligase that controls neural differentiation and proliferation by catalyzing the polyubiquitination and degradation of the N-Myc oncoprotein.34,35 Missense mutations and microduplications encompassing HUWE1 were identified in a few large families with moderate-to-severe ID,24 whereas nonsense, truncating and one missense mutations in PHF8 were previously reported in patients with ID and cleft lip/palate.21–23 Interestingly, in a previous report, two brothers with a deletion encompassing PHF8 and two nearby genes (FAM120C and WNK3) also had autistic features.36 In our study, both variants alter highly conserved amino acids in the proteins and are predicted to be deleterious; the p.Ser969del variant in PHF8 segregated in the two affected brothers from family 8 and the p.Val950Asp in HUWE1 occurred de novo in the index case of family 4 who is more severely affected than his brother. The p.Glu441Lys in PPP1R3F was also present in the two affected brothers of family 4 but was inherited from their unaffected maternal grandfather. Interestingly, another missense variant (p.Phe245Leu) in PPP1R3F was reported in a patient with Asperger syndrome.17 This observation suggests that the variants identified in family 4 act as risk factors for ASD when associated with other deleterious variants on the X chromosome or autosomes. Further studies are needed to confirm the roles of these genes in ASD.

Among the variants detected in this study, there was a single nonsense mutation in TMLHE that segregated with ASD in the two affected brothers of family 9. TMLHE is located at the far end of the long arm of chromosome X (Xq28) and encodes c-N-trimethyllysine hydroxylase, the enzyme that catalyzes the first of the four steps of endogenous carnitine biosynthesis. Screening of additional male patients with ASD identified two missense variants predicted to be deleterious in unrelated sporadic patients. The p.Arg77X and p.Asp244His mutations were associated with a significant increase of TML, the substrate of TMLHE and precursor of carnitine biosynthesis, in the plasma of the patients, confirming that they lead to loss of TMLHE function and deficit of endogenous carnitine biosynthesis. However, although carnitine was mildly decreased, it remained in the normal range in the plasma and urine of the patients. This result is not surprising as, in humans, the carnitine pool mainly comes from food intake. A small pool (~25%) of carnitine is also synthesized in liver, kidney and brain, but the precise role of endogenous
Figure 3  Identification of TMLHE mutations in three families. (a) Pedigrees and segregation analysis of the TMLHE mutations in families 9, PED-804 and AU-205. The arrows indicate the index cases. (b) Sequence electropherograms of the mutations at the hemizygous state in the index cases (835–03 in family 9, 804–03 and 205–03) and the affected brother of family 9 (835–04), and at the heterozygous state in the mothers (835–02, 804–02 and 205–02). (c) Sequence analysis confirmed that the long isoform contains the premature termination codon in exon 3 and that exon 3 was skipped in the short isoform, probably as a consequence of nonsense-associated alternative splicing. (d) Alignment of the region flanking the two missense variants in orthologous proteins showing the conservation of the altered amino acids. (e) Quantification of TMLHE mRNA expression in fibroblasts (F) and lymphoblasts (L) from members of family 9 by quantitative real-time PCR, using primer pairs in exons 7 and 8. TMLHE mRNA was expressed 10 times less in patients compared with healthy controls (green bars). Overnight treatment with 10 μg ml⁻¹ emetin (blue bars), an inhibitor of nonsense-mediated decay, restored the expression of the TMLHE mRNA. (f) Assay of free carnitine by UPLC chromatographic and TQD mass spectrometry in the plasma of patients. (g) Assay of trimethyllysine (TML) by UPLC chromatographic and TQD mass spectrometry in the plasma of patients.
synthesis in these tissues remains unknown.\textsuperscript{28} Carnitine is an essential metabolite in all animal species as well as in numerous microorganisms and plants. In mammalian cells, carnitine is present as free carnitine and acylcarnitines, including acetylcarnitine.\textsuperscript{37,38} The main role of carnitine is to transport activated long-chain fatty acids across the inner mitochondrial membrane for \( \beta \)-oxidation. But additional neuroprotective, neuromodulatory and neurotrophic roles have also been suggested.\textsuperscript{39–42} In particular, carnitine is an antioxidant and that might protect mitochondria from oxidative stress.

Another recent study supports the observation that the loss-of-function of \( \text{TMLHE} \) is associated with ASD. The deletion of exon 2 of \( \text{TMLHE} \), originally identified in ASD male patients,\textsuperscript{43} was shown to be a CNV that is present in male controls at a frequency of \( \sim 1/350 \). The first coding exon of \( \text{TMLHE} \), exon 2, encodes the signal peptide necessary to address the protein to mitochondria. Deletion of this exon causes enzyme deficiency and impairs endogenous carnitine biosynthesis, as observed for p.Arg77X and p.Asp244His mutations. Remarkably, 6 out of 7 sib pairs with ASD were concordant for the deletion of exon 2, whereas the frequency of this CNV was only slightly, not significantly, increased in sporadic ASD patient. The authors concluded that the deletion of exon 2 of \( \text{TMLHE} \) is a risk factor for nondysmorphic ASD, with a penetrance estimated at 2–4%.\textsuperscript{29} In this study, the deletion of exon 2 of \( \text{TMLHE} \) was present in three male patients with ASD out of 691 and only in one male control out of 896, indicating that this CNV could even be rarer than \( \sim 1/350 \). Although CNVs altering \( \text{TMLHE} \) tend to be more frequent in patients with ASD than in controls, the difference of frequencies between patients and controls was not significant. Yet, this finding is concordant with the results of the previous study in which significance between patients and controls was not achieved either.\textsuperscript{29} Interestingly, in our case, the only sib pair with ASD was discordant for the deletion of exon 2 of \( \text{TMLHE} \) and all the other patients were sporadic cases. However, association studies are not appropriate for rare variants, especially in the case of a high genetic heterogeneity such as in ASD. The arguments suggesting that a loss-of-function of \( \text{TMLHE} \) could contribute to autism are: (i) the demonstration that mutations have functional biological impact; (ii) the segregation of the mutations with the disorder in the families; and (iii) the identification of different rare functional mutations in the same gene in unrelated individuals.\textsuperscript{44} Taking these arguments into account, our results confirm that the deficiency of \( \text{TMLHE} \) likely contributes to ASD, probably in association with other genetic or non-genetic abnormalities, and reinforces the view that point mutations could also be identified in patients with ASD. The penetrance of the point mutations remains to be determined but might be higher than previously anticipated for deletion of exon 2.

The mechanism by which \( \text{TMLHE} \) deficiency leads to ASD remains unclear. The increase in TMLH could be toxic at some stage during brain development or interfere with the establishment of normal neuronal networks. Alternatively, the deficiency of carnitine itself or one of the three intermediates of endogenous carnitine biosynthesis (HTML, TMABA or \( \gamma \)-BB, Supplementary Figure S4) might be deleterious for brain development, alone or combined with a deficit in carnitine intake. To test the hypothesis that patients with p.Arg77X and p.Asp244His had a deficiency of carnitine intake in addition of the \( \text{TMLHE} \) mutation, we assessed their dietary intake over several days. The calculated carnitine intake was normal in all the three patients but reflects only current carnitine intake; a lack of carnitine during specific antenatal or neonatal periods cannot be ruled out. A retrospective questionnaire revealed that patient PED-804–03 (with the p.Asp244His mutation) refused to eat meat around 12 months of age. However, young children frequently refuse meat, particularly those with ASD. Interestingly, low levels of carnitine and acetylcarnitines and altered brain fatty acid metabolism were reported in subjects with ASD.\textsuperscript{45–47} However, if a systemic carnitine deficiency constitutes a risk factor for ASD, infants with low meat intake would have a higher risk of developing ASD. This is unlikely as the prevalence of autism is not notably increased in vegetarians or populations with low meat intake. In addition, carnitine and related metabolites are often abnormal in patients without ASD. These arguments suggest that low levels of carnitine and acetylcarnitines in ASD patients could reflect a more complex deficit. Further studies are, therefore, needed to decipher the real role of carnitine and \( \text{TMLHE} \) deficiency in ASD.

Recent studies have emphasized the role of carnitine in promoting social interactions in animal models. Desert (\( \text{Schistocerca gregaria} \)) and migratory locusts (\( \text{Locusta Migratoria} \)) reversibly change between two phenotypes (solitarious and gregarious) that differ in bodily appearance, physiology, brain size and organization, and behavior. At low population density, locusts in the solitarious phase avoid their congeners; when the population increases, locusts become gregarious and aggregate in migratory swarms. The genomes of the two forms of locust are equivalent. The transformation is driven only by epigenetic regulation.\textsuperscript{38} Interestingly, carnitine was recently shown to constitute a key regulatory metabolite in the phase transition in the migratory locust.\textsuperscript{49} Remarkably, the only other metabolite that is known to regulate phase transition in desert locusts is serotonin.\textsuperscript{50} The link between carnitine and serotonin is unclear, but carnitine has been proposed to act as a neuromodulator in the animal central nervous system; in addition, acylcarnitines could promote the biosynthesis and release of neurotransmitters, including dopamine and melatonin.\textsuperscript{51} Although these results were obtained in species very distant from humans, they support the hypothesis that carnitine has a conserved role in socialization during evolution and offers potentially novel insights into the complex role of carnitine and its derivatives acetylcarnitine and acylcarnitines in the brain.

Finally, \( \text{TMLHE} \) could also have other yet unknown functions. An isofrom has been reported in which exon 2 is spliced out and an alternative start codon in exon 3 is used.\textsuperscript{42} This isofrom lacks the mitochondrial targeting signal and probably localizes in another cellular compartment. Contrary to the deletion of exon 2, which affects only the mitochondrial isofrom,\textsuperscript{29} the mutations reported in this study are predicted to affect other isoforms as well. Interestingly, the TMLHE protein was reported to interact with nuclear complex p130/RBL2 that regulates gene expression, supporting the hypothesis that \( \text{TMLHE} \) has another cellular localization and function.\textsuperscript{52} The mechanism by which the \( \text{TMLHE} \) deficiency leads to ASD remains to be further characterized.
Altogether, our results confirm that a TMLHE deficiency is associated with ASD and support the hypothesis that rare variants on the X chromosome are involved in the etiology of ASD and contribute to the sex-ratio disequilibrium characteristic of these disorders.

Conflict of interest
The authors declare no conflict of interest.

Acknowledgements. We thank the patients for their participation in the study, the P3S platform and the genotyping and sequencing platform of the ICMP for technical assistance and the DNA and cell bank of CRICM for DNA extraction and cell culture. We also thank Dr Merle Ruberg for critical reading of the manuscript and Pr André Mégarbane for kindly providing DNA of controls individuals from Lebanon. This study was financially supported by GIS-Maladies Rares, Fondation de France, ERA-NEURON EHIUFAUTISM, INSERM and AP-HP.

1. Freitag CM, Steai W, Klucka SM, Duketis E, Waltes R. Genetics of autistic disorders: a review and clinical implications. Eur Child Adolesc Psychiatry 2010; 19: 169–173.
2. Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry 2010; 167: 1387–1383.
3. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 2011; 68: 1095–1102.
4. Sanders SJ, Ercan-Sencikel AG, Hus V, Luo R, Murtha MT, Moreno-Duca L et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011; 70: 863–885.
5. Depeine C, Moreno-Duca L, Heron D, Boutelle D, Gennetier A, Delorme R et al. Screening for genomic rearrangements and methylation abnormalities of the 15q11.2-q13 region in autism spectrum disorders. Biol Psychiatry 2009; 66: 349–359.
6. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick LI et al. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 2009; 5: e1000356.
7. Pinto D, Pamhgenata AT, Kie L, Arrey N, Merico D, Regan R et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466: 368–372.
8. Naxle MK, Kou Y, Liu L, Malayan A, Samocha KE, Sabo A et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485: 242–245.
9. O’Reak BJ, Vives L, Gitirajan S, Karakoc E, Krumm N, Coe BP et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; 485: 246–250.
10. Sanders SJ, Murtha MT, Gupta AR, Murdock JD, Rabouss MJ, Wilsey AJ et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; 485: 237–241.

Translation

Altogether, our results confirm that a TMLHE deficiency is associated with ASD and support the hypothesis that rare variants on the X chromosome are involved in the etiology of ASD and contribute to the sex-ratio disequilibrium characteristic of these disorders.

Conflict of interest
The authors declare no conflict of interest.

Acknowledgements. We thank the patients for their participation in the study, the P3S platform and the genotyping and sequencing platform of the ICMP for technical assistance and the DNA and cell bank of CRICM for DNA extraction and cell culture. We also thank Dr Merle Ruberg for critical reading of the manuscript and Pr André Mégarbane for kindly providing DNA of controls individuals from Lebanon. This study was financially supported by GIS-Maladies Rares, Fondation de France, ERA-NEURON EHIUFAUTISM, INSERM and AP-HP.

1. Freitag CM, Steai W, Klucka SM, Duketis E, Waltes R. Genetics of autistic disorders: a review and clinical implications. Eur Child Adolesc Psychiatry 2010; 19: 169–173.
2. Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry 2010; 167: 1387–1383.
3. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 2011; 68: 1095–1102.
4. Sanders SJ, Ercan-Sencikel AG, Hus V, Luo R, Murtha MT, Moreno-Duca L et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011; 70: 863–885.
5. Depeine C, Moreno-Duca L, Heron D, Boutelle D, Gennetier A, Delorme R et al. Screening for genomic rearrangements and methylation abnormalities of the 15q11.2-q13 region in autism spectrum disorders. Biol Psychiatry 2009; 66: 349–359.
6. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick LI et al. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 2009; 5: e1000356.
7. Pinto D, Pamhgenata AT, Kie L, Arrey N, Merico D, Regan R et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466: 368–372.
8. Naxle MK, Kou Y, Liu L, Malayan A, Samocha KE, Sabo A et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485: 242–245.
9. O’Reak BJ, Vives L, Gitirajan S, Karakoc E, Krumm N, Coe BP et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; 485: 246–250.
10. Sanders SJ, Murtha MT, Gupta AR, Murdock JD, Rabouss MJ, Wilsey AJ et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; 485: 237–241.

Translation

Altogether, our results confirm that a TMLHE deficiency is associated with ASD and support the hypothesis that rare variants on the X chromosome are involved in the etiology of ASD and contribute to the sex-ratio disequilibrium characteristic of these disorders.

Conflict of interest
The authors declare no conflict of interest.

Acknowledgements. We thank the patients for their participation in the study, the P3S platform and the genotyping and sequencing platform of the ICMP for technical assistance and the DNA and cell bank of CRICM for DNA extraction and cell culture. We also thank Dr Merle Ruberg for critical reading of the manuscript and Pr André Mégarbane for kindly providing DNA of controls individuals from Lebanon. This study was financially supported by GIS-Maladies Rares, Fondation de France, ERA-NEURON EHIUFAUTISM, INSERM and AP-HP.
49. Wu R, Wu Z, Wang X, Yang P, Yu D, Zhao C et al. Metabolomic analysis reveals that carnitines are key regulatory metabolites in phase transition of the locusts. Proc Natl Acad Sci USA 2012; 109: 3259–3263.

50. Anstey ML, Rogers SM, Ott SR, Burrows M, Simpson SJ. Serotonin mediates behavioral gregation underlying swarm formation in desert locusts. Science (New York, NY) 2009; 323: 627–630.

51. Scafidi S, Fiskum G, Lindauer SL, Bamford P, Shi D, Hopkins I et al. Metabolism of acetyl-L-carnitine for energy and neurotransmitter synthesis in the immature rat brain. J Neurochem 2010; 114: 820–831.

52. Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S et al. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. Mol Cell 2007; 26: 539–551.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)