

# Mapping of sudden infant death with dysgenesis of the testes syndrome (SIDDT) by a SNP genome scan and identification of *TSPYL* loss of function

Erik G. Puffenberger<sup>\*†</sup>, Diane Hu-Lince<sup>\*†</sup>, Jennifer M. Parod<sup>\*†</sup>, David W. Craig<sup>‡</sup>, Seth E. Dobrin<sup>‡</sup>, Andrew R. Conway<sup>§</sup>, Elizabeth A. Donarum<sup>¶</sup>, Kevin A. Strauss<sup>\*</sup>, Travis Dunckley<sup>‡</sup>, Javier F. Cardenas<sup>‡</sup>, Kara R. Melmed<sup>‡</sup>, Courtney A. Wright<sup>‡</sup>, Winnie Liang<sup>‡</sup>, Phillip Stafford<sup>‡</sup>, C. Robert Flynn<sup>||</sup>, D. Holmes Morton<sup>†</sup>, and Dietrich A. Stephan<sup>\*\*</sup>

<sup>\*</sup>Clinic for Special Children, Strasburg, PA 17579; <sup>†</sup>Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ 85004; <sup>§</sup>Silicon Genetics, Redwood City, CA 94063; <sup>¶</sup>Department of Neurodevelopmental Genetics, Barrow Neurological Institute, Phoenix, AZ 85012; and <sup>||</sup>Arizona BioDesign Institute and Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85248

Edited by Albert de la Chapelle, Ohio State University, Columbus, OH, and approved June 15, 2004 (received for review February 19, 2004)

**We have identified a lethal phenotype characterized by sudden infant death (from cardiac and respiratory arrest) with dysgenesis of the testes in males [Online Mendelian Inheritance in Man (OMIM) accession no. 608800]. Twenty-one affected individuals with this autosomal recessive syndrome were ascertained in nine separate sibships among the Old Order Amish. High-density single-nucleotide polymorphism (SNP) genotyping arrays containing 11,555 single-nucleotide polymorphisms evenly distributed across the human genome were used to map the disease locus. A genome-wide autozygosity scan localized the disease gene to a 3.6-Mb interval on chromosome 6q22.1-q22.31. This interval contained 27 genes, including two testis-specific Y-like genes (*TSPYL* and *TSPYL4*) of unknown function. Sequence analysis of the *TSPYL* gene in affected individuals identified a homozygous frameshift mutation (457\_458insG) at codon 153, resulting in truncation of translation at codon 169. Truncation leads to loss of a peptide domain with strong homology to the nucleosome assembly protein family. GFP-fusion expression constructs were constructed and illustrated loss of nuclear localization of truncated *TSPYL*, suggesting loss of a nuclear localization patch in addition to loss of the nucleosome assembly domain. These results shed light on the pathogenesis of a disorder of sexual differentiation and brainstem-mediated sudden death, as well as give insight into a mechanism of transcriptional regulation.**

Over two generations, nine families from the Belleville Amish Community have lost 21 infants to a recently discovered disorder we have entitled sudden infant death with dysgenesis of the testes syndrome (SIDDT; Fig. 1) [Online Mendelian Inheritance in Man (OMIM) accession no. 608800]. The condition is not seen in the Lancaster County Old Order Amish population. Three of these infants have been cared for at the Clinic for Special Children. Most previous cases were studied extensively at regional medical centers; however, no diagnostic tests were available, and clinical recognition of the syndrome was difficult, particularly in affected females. Caretakers say that they can often recognize affected infants at birth by the unusual sound of their cry, which is a staccato sound, similar to the cry of a goat (see *Materials and Methods* for a complete description of the syndrome).

Homozygosity mapping to identify disease genes in autosomal recessive disorders common in founder populations by using traditional methods has often been hampered by microsatellite marker density (1, 2). Typically,  $\approx 400$  microsatellite markers are used in such linkage studies spaced at an average 10-centimorgan density throughout the genome. Single-nucleotide polymorphisms (SNP) are present in the genome at an average density of  $\approx 1$  per 1,300 bp and hold enormous potential as a high-density high-throughput genotyping strategy for disease gene mapping (3). Information content is a function of marker heterozygosity, distance between markers, and pedigree structure. Although

each individual biallelic SNP is less informative than a single microsatellite marker in most cases (e.g., average heterozygosity of 0.37 on Affymetrix 10K Array vs. 0.72 on the Center for Inherited Disease Research database), the greater number of SNPs in aggregate leads to higher information content at any particular point in the genome (4). The Affymetrix 10K Array assay requires only 250 ng of DNA and generates 11,555 SNP genotypes with an average resolution of one SNP every 210 kb (5). This new genotyping platform has a throughput 100-fold greater than microsatellite genotyping and an accuracy of  $>99.5\%$  with automated allele calling. The high density and information content of this genotyping platform make it ideal for localizing small regions of homozygosity.

Blinded validation studies were done to verify that gene mapping could be accomplished using the SNP arrays in small disease pedigrees with known map location. The gene for each disease was previously mapped and the causative mutation identified. In each case, we were able to reproduce the mapping results. A software analysis package entitled VARIA (Silicon Genetics) was developed to handle the large amount of data inherent to these assays and to correctly localize disease-carrying loci from markers that are in linkage disequilibrium with one another. The genome-wide linkage scan conducted on the multiplex SIDDT pedigree rapidly and unambiguously mapped this disorder to 6q22 with a location score of 8.11 [maximum 2-point logarithm of odds (LOD) of 2.41] in a 3.6-Mb interval. Sequencing of two candidate genes in the region identified a nonsense mutation in the testis-specific Y-like gene (*TSPYL*). Functional validation was performed through construction of GFP-fusion proteins of both the full-length and truncated *TSPYL* proteins to investigate the effect of truncation on cellular localization.

## Materials and Methods

**Subjects and Samples.** DNA samples used in mapping and sequencing studies of SIDDT were acquired by the Clinic for Special Children, with informed consent. Peripheral blood was collected from four affected individuals, their parents, siblings, and extended family members. Samples from maple syrup urine

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

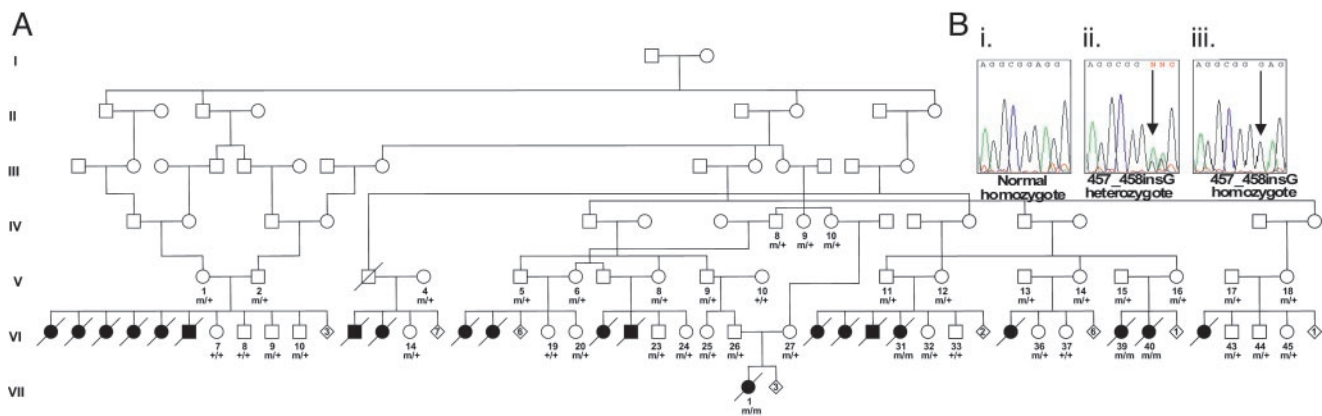
Abbreviations: SIDDT, sudden infant death with dysgenesis of the testes syndrome; *TSPYL*, testis-specific Y-like; SNP, single-nucleotide polymorphism; NAP, nucleosome assembly protein; LOD, logarithm of odds.

Data deposition: The disorder reported in this paper has been deposited in the Online Mendelian Inheritance in Man (OMIM) database (accession no. 608800).

<sup>†</sup>E.G.P., D.H.-L., and J.M.P. contributed equally to this work.

<sup>\*\*</sup>To whom correspondence should be addressed. E-mail: dstephan@tgen.org.

© 2004 by The National Academy of Sciences of the USA



**Fig. 1.** SIDDIT in a consanguineous Old Order Amish pedigree. (A) *TSPYL* 457\_458insG mutation status is indicated for available pedigree members (m denotes 457\_458insG, whereas + represents wild-type sequence). (B) Sequencing of *TSPYL* reveals a homozygous single base-pair insertion (457\_458insG) in SIDDIT syndrome patients. (i) *TSPYL* sequence from a normal control. (ii) 457\_458insG heterozygote. (iii) 457\_458insG homozygote.

disease, Crigler–Najjar syndrome, and congenital nephrotic syndrome families were previously collected with informed consent and described elsewhere (1, 2). Genealogies for all families in the four diseases described herein were prepared through interviews and several published family histories. Additional research was performed at the Lancaster Mennonite Historical Society (Lancaster, PA).

**Clinical Description of the Condition.** Infants with SIDDIT appear normal at birth, develop signs of viscerotonic dysfunction early in life, and die before 12 months of age of abrupt cardiorespiratory arrest. One infant died in the hospital while awake and attached to a cardiac-respiratory monitor. The monitor showed simultaneous asystole and cessation of respiratory efforts. Another infant died suddenly and unexpectedly at home, with no premonitory signs. In both cases, neuropathological examinations were done by experts in sudden infant death syndrome. Brain and peripheral nerves were normal. Specifically, there was no dysplasia or inflammation of the brainstem and no pathology of cervical anterior horn cells or lower motor neurons of the hypoglossal nerve. The absence of anterior horn cell pathology contrasts with signs of progressive cranio-cervical and upper thoracic motor unit dysfunction in affected patients. These signs can include tongue fasciculations, ocular palsies, symmetric weakness of the facial nerve, and diminished upper extremity reflexes.

Signs of abnormal autonomic and visceral nerve regulation manifest within the first months of life and include neonatal bradycardia, hypothermia, severe gastroesophageal reflux, laryngospasm, bronchospasm, and abnormal cardiorespiratory patterns during sleep. All affected infants have a hyperactive startle reflex that can be provoked by loud noise, bright light, movement, or tactile stimulation (e.g., bathing). The reflex is unresponsive to clonazepam. Affected newborns are difficult to feed and may require nasogastric tube alimentation. Infant VII-1 was evaluated twice, at 4 and 7 weeks of age, for feeding and respiratory problems. At age 4 weeks, she presented with poor growth, persistent stridor, and episodes of laryngospasm and cyanosis. A barium swallow excluded tracheoesophageal fistula. A five-channel pneumogram and pH probe showed esophageal pH < 4 for 46% of study (normal < 6%), with central, obstructive, and mixed apneas. Direct laryngoscopy revealed arytenoid and postcricoid supraglottic erythema. She had persistent bilateral perihilar consolidations on serial chest radiographs.

This infant underwent gastrostomy tube placement and fundoplication, which prevented acid reflux and provided airway protection, but she continued to have life-threatening apnea and

obstructive breathing. When a five-channel pneumogram and pH probe was repeated at age 7 weeks, attempted passage of the probe through both left and right nares provoked severe bradycardia and cyanosis. This 23-h study showed normal lower esophageal pH, frequent isolated bradycardias, extreme background heart rate variability (range 52–250 beats per minute), and multiple bradycardic events associated with disorganized breathing or prolonged apnea. A salivagram showed normal transit of sublingual tracer from mouth to stomach, with no aspiration. A cranial MRI was normal and electroencephalogram showed normal waking, drowsiness, and slow-wave sleep electrical patterns. Despite exclusive gastrostomy tube feeding, growth remained poor, and she died suddenly before her first birthday.

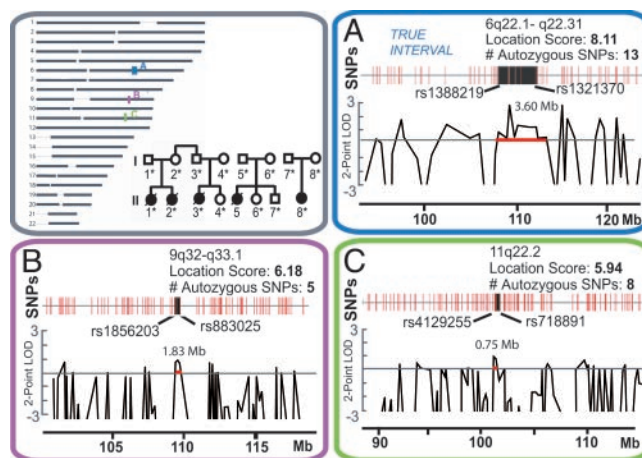
Genotypic males with SIDDIT syndrome have fetal testicular dysgenesis and ambiguous genitalia and can be mistaken for females. In two genetically male but externally phenotypic female infants with the syndrome, intraabdominal testes were examined at autopsy. They were dysplastic, with tortuous vascularity and an arrest of cell development at an early stage. Both Leydig and Sertoli cells were present but diminished in number and nested in rudimentary fibrotic cords. Normal regression of Mullerian structures in males indicates that Sertoli cells secrete Mullerian inhibiting hormone. However, Leydig cells fail to produce the testosterone and dihydrotestosterone throughout fetal life necessary to sustain growth and development of the external male organ. This can be detected postnatally as an abnormal response to human chorionic gonadotropin challenge. Development of male genitalia arrests at variable embryologic stages. At birth, some males were thought to be females on the basis of external genitalia, but other male infants had fusion and rugation of the gonadal sac and partial development of the penile shaft. Variable maturation of male genitalia indicates that Leydig cell failure can occur at different times throughout fetal development.

Female sexual development is normal, as is neonatal hypothalamic and ovarian function. Patient VII-1, a genotypic and phenotypic female, had normal internal and external genitalia and normal postpartum elevation of serum estradiol, luteinizing hormone (LH), and follicle-stimulating hormone (FSH). Adrenal and gonadal pathways of steroid biosynthesis were studied extensively in patient VI-31 and found to be normal. In this child, pituitary and hypothalamic functions were also normal with regard to serum thyroid, corticotropin (ACTH), cortisol, LH/FSH, and prolactin levels.

Additional endocrinological and biochemical studies were done on SIDDIT syndrome patients at various stages of investi-

gation. Plasma sterol and organic analyses by gas chromatography mass spectrometry, serum acylcarnitines by tandem mass spectrometry, steroid immunoassays for congenital adrenal hyperplasias, and plasma amino acid analyses yielded normal results. Cerebrospinal fluid amino acids and neurotransmitter metabolites were normal in patient VII-1. Chromosomes were done in many cases, both to establish genotypic sex and to detect structural aberrations; no abnormalities were found. PCR and gene sequencing excluded spinal muscular atrophy (SMN1 exon 7 deletion). An assay was done for the triplet repeat expansion of the testosterone receptor gene, which in adult males causes testicular failure and progressive bulbular weakness (Kennedy syndrome), and results were normal.

**Genotyping.** DNA from whole blood was isolated by using the PUREGENE DNA Isolation Kit (Gentra Systems). SNP genotyping was done by using the GeneChip Mapping 10K Array and Assay Kit (Affymetrix). The genotyping protocol was modified from Kennedy *et al.* (5). All incubations were done by using a Tetrad thermal cycler (MJ Research, Cambridge, MA). Internal positive and negative controls were performed in parallel by using the supplied genomic DNA (Affymetrix). Two hundred fifty nanograms of double-stranded genomic DNA was digested with *Xba*I (New England Biolabs) for 2 h at 37°C followed by heat inactivation for 20 min at 70°C. Digested DNA was then incubated with a 0.25 M *Xba* adapter (Affymetrix) and DNA ligase (New England Biolabs) in standard ligation buffer (New England Biolabs) for 2 h at 16°C followed by heat inactivation for 20 min at 70°C. Ligated products were amplified in quadruplicate by using 0.5 M of the supplied generic primer *Xba*I (Affymetrix) in PCR Buffer II (Applied Biosystems) with 2.5 mM MgCl<sub>2</sub>/2.5 mM dNTPs/10 units of AmpliTaq Gold polymerase (Applied Biosystems) under the following PCR conditions: 95°C for 5 min, followed by 35 cycles (95°C for 20 sec, 59°C for 15 sec, and 72°C for 15 sec) and a final extension at 72°C for 7 min. Fragments in the 250- to 1,000-bp size range are preferentially amplified under these conditions. PCR products were purified with the QIAquick PCR purification kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations, with the exception of the elution procedure. DNA from each of four PCR replicate samples was bound to separate columns and washed. The eluant collected from column 1 was used to elute the remaining three columns in series. The final purified product is the combination of four purified PCR product samples. Eighteen to twenty micrograms of purified PCR products were fragmented with 0.24 units of the supplied GeneChip Fragmentation Reagent (Affymetrix) for 30 min at 37°C followed by a heat inactivation for 15 min at 95°C. Samples were then labeled with 102 units of terminal deoxytransferase (Affymetrix) and 0.143 mM of supplied GeneChip DNA Labeling Reagent in TdT buffer (Affymetrix) for 2 h at 37°C followed by 15 min at 95°C. After end-labeling, the fragments were hybridized to a GeneChip Human Mapping 10K Array for 16–18 h at 48°C while rotating at 60 rpm. Microarrays were washed by using the Fluidics Station 450 (Affymetrix) in 0.6 × SSPE (sodium chloride, sodium phosphate, and EDTA), followed by a three-step staining protocol. We incubated the arrays first with 10 μg/ml streptavidin (Pierce), washed with 6× SSPE and incubated with 5 μg/ml biotinylated antistreptavidin (Vector Laboratories) and 10 μg/ml streptavidin–phycoerythrin conjugate (Molecular Probes), and finally washed with 6× SSPE per the manufacturer's recommended times. Microarrays were scanned by using the GeneChip Scanner 3000 according to the manufacturer's protocol (Affymetrix). Data acquisition was performed by using the GeneChip GCOS software. Initial data analysis was done by using the GeneChip DNA ANALYSIS software and then exported to VARIA (Silicon Genetics) for analysis. All raw genotype data and call statistics for the four disorders are in Tables 1–4, which are



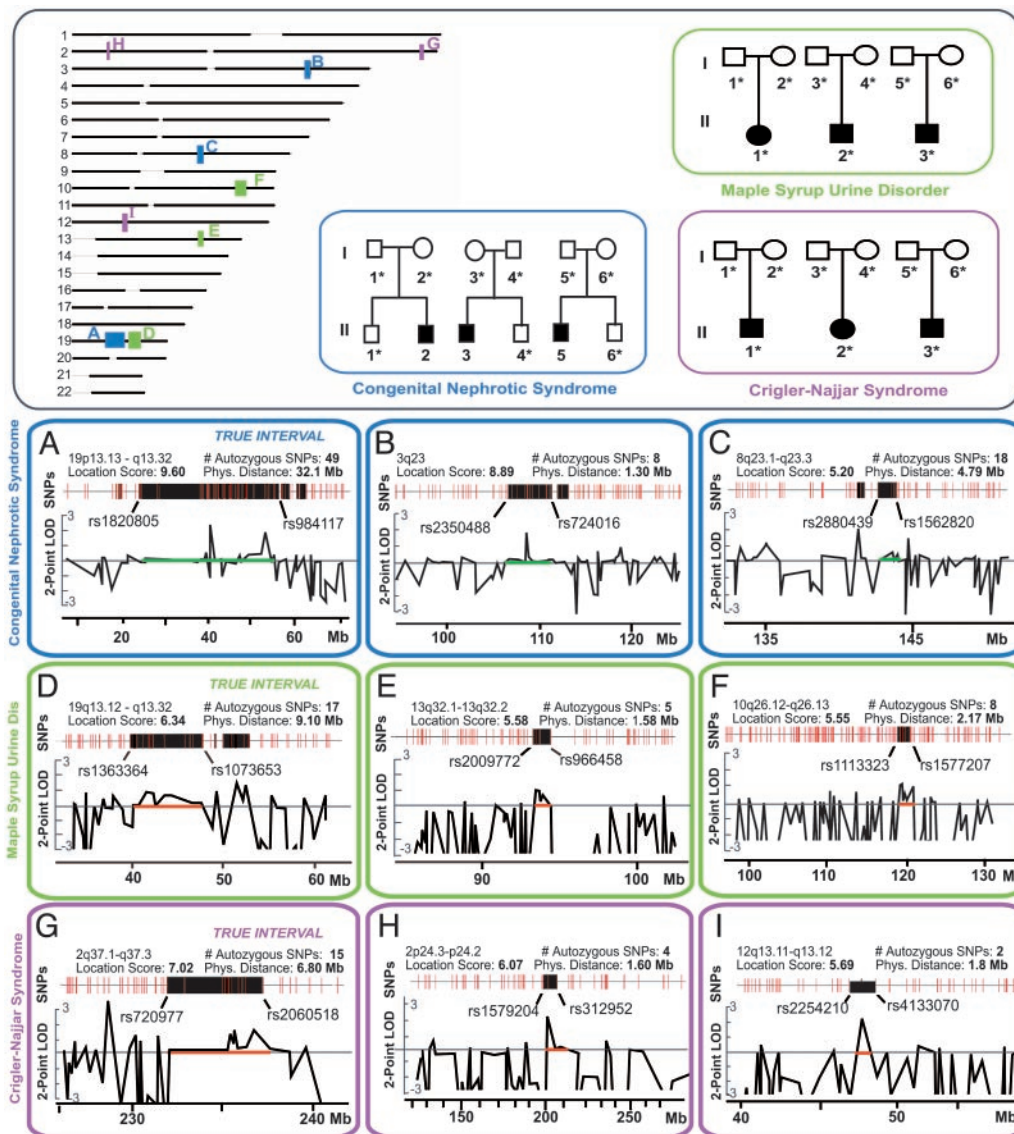
**Fig. 2.** SNP genotyping reveals a 1.1-centimorgan region (3.6 Mb) on chromosome 6q22.1-q22.31 that contains the *SIDDT* gene. Asterisks denote individuals who were genotyped. The three regions across the genome with the highest location scores are ranked by descending score. The four affected individuals are homozygous for 13 adjacent SNPs at the 6q22 locus. Horizontal red bars indicate the autozygous region (with physical distance illustrated), and two-point LOD scores for each SNP are plotted to illustrate that information content of biallelic markers in small pedigrees alone make mapping difficult. The region is bounded by SNPs rs1388219 and rs1321370. The proximal and distal transmitted haplotypes flanking the homozygous region were not diverse, whereas there was little haplotype sharing in intervals B and C outside the autozygous segment, indicating these were false-positive regions. The two regions with the next best location scores had small numbers of contiguous homozygous SNPs, smaller homozygous intervals, and little sharing outside the homozygous interval.

published as supporting information on the PNAS web site. The overall performance statistics of the genotyping arrays over the 485,310 genotypes in 43 individuals (including the *SIDDT* pedigree), yielded average call rates of 81.2% and average signal detection rates of 91.43%.

**Linkage Mapping.** Data analysis was done by using the VARIA software package developed by Silicon Genetics. SNP positions came from dbSNP build 115 and National Center for Biotechnology Information build 34v2 of the human genome. VARIA searches for genomic regions that are identical by descent between all affected individuals and assumes no mutation heterogeneity within affected individuals. Several assumptions were made for the analysis, including a genotype error rate of 1% and Hardy–Weinberg equilibrium. Details on the generation of “location scores” can be found at <http://www.silicongenetics.com/Support/autozygosity.pdf> (referred to as “regional LOD scores” in the pdf; ref. 6). Briefly, the two-point LOD scores across the autozygous region are summed to produce the location score, accounting for size of the shared interval as well as the magnitude of allele frequency differences within the interval. One hundred seventy Old Order Amish Control chromosomes (including six untransmitted chromosomes from the *SIDDT* parents and 38 from the three mapping validation sibships) were used to estimate SNP allele frequencies. The three regions of the genome having the highest location scores are illustrated in Figs. 2 and 3.

**Sequencing.** Mutation analysis was performed for the *TSPYL* and *TSPYL4* genes in the linked region on chromosome 6q22. The exon of each target gene was amplified by using specific primers and 50 ng of genomic DNA from affected and unaffected family members.

The *TSPYL* gene lacks introns and contains a coding region of



**Fig. 3.** Accurate disease gene localization using the Affymetrix GeneChip Mapping 10K Assay Kit and the Silicon Genetics VARIA software package. Mapping of the congenital nephrotic syndrome locus, maple syrup urine disease locus, and the Crigler–Najjar syndrome locus was accomplished by using minimal numbers of individuals in each pedigree. The three highest location scores across the genome are indicated for each disorder in descending order. The location score combined with physical distance was found to be the best predictor of the correct region in each case. No affected individuals were genotyped in the NPHS1 pedigree, yet it was still possible to detect the correct locus, and the inferred region of autozygosity is indicated by a green horizontal bar.

1,314 bases (GenBank accession no. AL050331). The mRNA is  $\approx 3,200$  bases in length (GenBank XM371844), and the mature TSPYL protein is 437 aa. Primer sequences for TSPYL amplification were 5'-AGATCTCCAGTCTGACGACAC-3' (forward) and 5'-AGGAAACAGGGTGCAGAAAAGT-3' (reverse). TSPYL sequencing primers (in addition to the forward and reverse primers, above) were TSPYL1036-F: 5'-GGCCGAGTGGTGTCTCTTTCTA-3'; TSPYL618-F: 5'-GGAGGATAGATTGAGGAGGAG-3'; TSPYL237-F: 5'-TACTCCCAGATCCGAGTTGTT-3'. In addition to the 457.458insG mutation, two polymorphic variants were incidentally detected while sequencing control samples: a known nonsynonymous SNP 541G/A (A181T) and a unique in-frame short tandem repeat [523(GTG)<sub>2</sub>-3] that codes for either two or three adjacent valine residues at positions 175–177 in the peptide. The 541A allele had a frequency of 7.8%, whereas the 523(GTG)<sub>2</sub> allele had a frequency of 30.2% on control chromosomes. Primer sequences for PCR amplification of TSPYL4 were 5'-AAACTCCCCTCCAGACTGAC-3' (forward) and

5'-CACAAATGCAGAAAAGCATGAAG-3' (reverse). TSPYL4 sequencing primers (in addition to the forward and reverse primers, above) were TSPYL4C277-F: 5'-ACACAGGTGATGGCGAACACAG-3'; TSPYL4C776-F: 5'-CCATCGATCAAGAGTTGTCAAA-3'; TSPYL4C1165-F: 5'-CAGGCTCATATCCACAGAAC-3'; and TSPYL4C889-R: 5'-TAATGAACTTCTGCCTGCAT-3'.

PCR products were purified by using QIAquick columns (Qiagen) and then sequenced using the BigDye Terminator cycle sequencing protocol (Applied Biosystems). Extension products were size-fractionated on an Applied Biosystems 310 Genetic Analyzer. Sequences were compared to normal sequence for each gene using GenBank AL050331, which contains both genes. Population-based control samples were sequenced in an identical fashion.

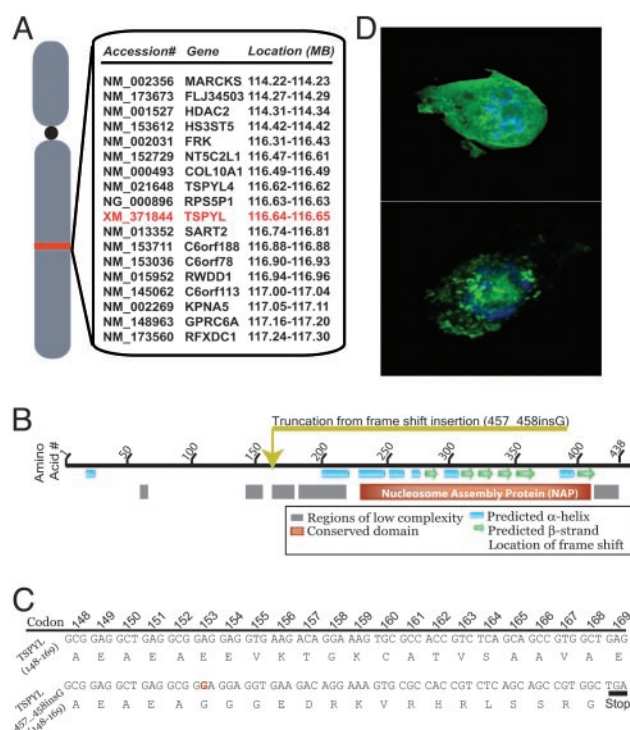
**Functional Validation.** Full-length and truncated TSPYL cDNAs were amplified by using gene-specific primers for the full-length:

forward, 5'-CACCATGAGCGGCCTGGATGGGGTCAA-GAGG-3'; reverse, 5'-TAGCTCGAGACCAGACTGGAAC-CCAAAGGGCCTGGGGATC-3'; and the truncated: identical forward to above; reverse, 5'-TAGCTCGAGTGGCGGCT-GTCTCTACCTCC-3' versions from genomic DNA using the Expand High Fidelity PCR System (Roche, Indianapolis). Cycling conditions were 94°C for 2 min and then 9 cycles (94°C for 15 sec, 68°C for 1 min, and 68°C for 105 sec) followed by 94°C for 15 sec and then 24 cycles (68°C for 1 min, 72°C for 3 min) and then 72°C for 12 min. PCR products were cloned directionally into the pENTR-D-TOPO Gateway vector (Invitrogen) and then into the Gateway destination vector pcDNA-DEST47 (Invitrogen) containing a C-terminal GFP tag. All clones were sequence verified. HeLa<sub>F2</sub> cells (10<sup>5</sup>) were grown in DMEM supplemented with 10% FBS (GIBCO/BRL). The HeLa<sub>F2</sub> cells were transfected with 4 μg of the *TSPYL* full-length or truncated GFP constructs, respectively, using Lipofectin Reagent (Invitrogen) according to the manufacturer's recommendations. Twenty-four hours post-transfection, subconfluent HeLa<sub>F2</sub> cells were washed three times with PBS. The transiently transformed cells were fixed in 2% formaldehyde and permeabilized in 0.1% Triton X-100. Nuclear localization was confirmed by using a 4',6-diamidino-2-phenylindole dihydrochloride stain from Molecular Probes. Images were captured with FITC and UV filter sets and a ×100 oil immersion objective in conjunction with a Leica (Deerfield, IL) TCS-NT confocal microscope.

## Results and Discussion

We first wanted to verify that statistically significant and unambiguous disease gene mapping could be accomplished with a very small number of affected sibships and a dense genotyping platform. These blinded validation studies used three small disease-mapping sets from the Old Order Amish and Old Order Mennonite populations of southeastern Pennsylvania, namely, maple syrup urine disease (1, 7), Crigler–Najjar syndrome (1, 8), and congenital nephrotic syndrome (1, 2). The gene for each disease had been previously identified. Genome-wide linkage scans were performed and regions of autozygosity identified through a location score statistic. Heterozygous markers flanking the identical by descent homozygous segment in affecteds define the minimal genetic interval. In each of the three test cases, the interval with the largest location score correctly identified the true gene location (Fig. 3). The location score is not a formal statistic, but a summation of two-point LOD data across a region of homozygosity. Equal marker density across physical distances is assumed, which is a flaw of the statistic, but this issue should resolve as marker spacing stabilizes in future array iterations.

The SIDDT mapping panel was comprised of four affected individuals (Fig. 1, VI-31, VI-39, VI-40, and VII-1) and their parents from three sibships. The region that received the highest two-point LOD score (2.41; Fig. 2) and the highest location score (8.11; Fig. 2) was on chromosome 6q22.1-q22.31 [Genome build 34v2, National Center for Biotechnology Information (NCBI)]. The homozygous segment spanned 3.60 Mb, corresponding to 1.1 centimorgans. Examination of the minimal shared region in affecteds, which was flanked by SNPs rs1388219 and rs1321370, revealed 27 known and hypothetical genes based on both the NCBI and Celera annotations, 18 of which are characterized (Fig. 4). The SIDDT phenotype includes testicular dysgenesis; thus genes with known or inferred function related to sexual differentiation and testicular development were sought. Two genes were possible candidates based on sequence homology to the testis-specific gene *TSPY* at chromosome Yp11: the paralogs testis-specific Y-encoded like genes *TSPYL* and *TSPYL4*. Unlike *TSPY*, which has testis-restricted expression, *TSPYL* is expressed in the brain and testes (as well as other tissues).



**Fig. 4.** Mutation of the *TSPYL* gene causes SIDDT. (A) The gene content of the interval contained 18 characterized and 9 hypothetical genes (not shown). The genomic interval surrounding *TSPYL* and *TSPYL4* (25-kb proximal) contains only two documented (CA)<sub>n</sub> repeats, neither of which is in the standard Applied Biosystems linkage panel; thus, this locus might have been missed by using the small DNA sample set and a standard microsatellite marker screen. (B) Individuals with SIDDT syndrome were found to have a *TSPYL* frameshift mutation at codon 153 causing cessation of translation at codon 169. The predicted secondary structure shows a series of α-helices and β-strands in the NAP homologous region. (C) The primary amino acid sequence. We speculate that the 5' domain may be a DNA-binding domain or may interact with regulatory complexes within the chromatin. (D) Subcellular localization of *TSPYL* is altered through loss of the 3' portion of the peptide (Upper, truncated *TSPYL*; Lower, full-length *TSPYL*). Loss of the NAP functional domain probably directly affects the ability of *TSPYL* to shuttle histones from the cytoplasm to the nucleus but also disrupts the nuclear localization signal on the tertiary surface of the peptide. *TSPYL* localizes to the nucleus with a punctate staining pattern, whereas the truncated form illustrates diffuse cytoplasmic staining. The nucleus is counterstained with 4',6-diamidino-2-phenylindole dihydrochloride.

Complete sequencing of the *TSPYL* gene in an affected individual (VI-40) revealed a homozygous single base insertion, 457-458insG (Fig. 4). This variant causes premature truncation of the protein at codon 169. This change was not seen as a polymorphic variant in GenBank. Sequencing of all 42 individuals in the SIDDT pedigree revealed that all affected individuals were homozygous for the change, all parents of affected individuals were heterozygous, and no unaffected siblings were homozygous for the change. Fifty-eight Old Order Amish controls were genotyped for the insertion ( $n = 116$  chromosomes). Most of these samples were from the Lancaster County Old Order Amish population; however, eight controls were available for study from the Juniata and Mifflin County Old Order Amish. None of the Lancaster County Old Order Amish carried the variant, but four heterozygotes were detected from the Mifflin and Juniata County Old Order Amish. Although the sample size is small ( $n = 16$  chromosomes), this suggests that the *TSPYL* 457-458insG variant has an especially high carrier frequency in this genetic isolate. A subset of sudden infant death syndrome and/or ambiguous genitalia in these populations could be caused

by TSPYL loss of function. A second gene family member, *TSPYL4*, was found 25 kb proximal to *TSPYL*, but direct sequencing of an affected individual identified no coding variants. The truncated *TSPYL* exhibited inappropriate subcellular targeting in an *in vitro* assay (Fig. 4), providing additional evidence that *TSPYL* is causative of SIDDT.

Aspects of the molecular function of *TSPYL* can be gleaned from its primary sequence (Fig. 4). The *TSPYL* nucleosome assembly protein (NAP) domain is preceded by a region of low complexity with no known functional motifs. NAPs are a family of proteins that function as chaperones, shuttling histones from the cytosol to the nucleosome. Studies in yeast show that NAPs are critical to nucleosome assembly, mitotic progression, and chromatin formation (9–11). NAP domains may also function as transcription factors during embryogenesis. During the development of *Xenopus* embryos, the NAP-containing protein NAPIL is broadly expressed at its highest level during development of hematopoietic tissue. When NAPIL is overexpressed, genes involved in tissue development are up-regulated, specifically *GATA-2*, a gene essential for hematopoiesis (11). Although we focused on *TSPYL* based on its similarity to *TSPY*, the knowledge that genotypic females are affected suggests that *TSPYL*, unlike *TSPY*, is expressed in other tissues (12, 13). *TSPYL* may play a role in development by altering regulation of specific developmental genes and contributing to region-specific chromatin remodeling. Several data sets within the Gene Expression Omnibus (GEO) illustrate that *TSPYL* is highly and almost exclusively (with respect to adult structures) expressed in

the fetal brain (GEO no. GSM14799). In addition, *TSPYL* has been shown to be negatively regulated in the hippocampus in a linear dose-dependent fashion by corticosteroids (GEO no. GSM12543), sensitively negatively regulated by JNK2 (GEO no. GSM1514), and positively regulated by testosterone (GEO no. GSM6733). Elucidation of *TSPYL* function may shed light on fundamental aspects of embryogenesis of the human nervous and reproductive systems through a previously not described signaling mechanism. More important, studies of *TSPYL* expression and function in the developing brain may provide new insight into the genetic basis of apnea, dysphagia, cardiac arrests, and sudden unexplained deaths in infancy. Present clinical evidence suggests that in SIDDT, sudden death may result from dysregulation of the autonomic brainstem systems that control cardiac and pulmonary protective reflexes (14–16). The lethal event may be profound vagally mediated laryngobronchospasm or asystole.

We thank the Old Order Amish families who participated in the research and the Old Order Amish community for their willingness to participate in research studies, Robert Wells at Affymetrix, Inc., and G. Terry Sharrer at the Smithsonian Institution for facilitating this work, Drs. Allen Ettinger and Bruce Lidston for bringing this disorder to our attention, Phanindra Tangirala for posting data on the Web, and John Pearson for bioinformatics support. This work was supported in part by National Institute of Neurological Disorders and Stroke Grant 1U24NS04357-01 (to D.A.S.) and by National Research Service Award Fellowship 7F32 NS43932-02 (to D.H.-L.).

1. Puffenberger, E. G. (2003) *Am. J. Med. Genet.* **121C**, 18–31.
2. Bolk, S., Puffenberger, E. G., Hudson, J., Morton, D. H. & Chakravarti, A. (1999) *Am. J. Hum. Genet.* **65**, 1785–1790.
3. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., *et al.* (1999) *Nat. Genet.* **22**, 231–238.
4. Matsuzaki, H., Loi, H., Dong, S., Tsai, Y., Fang, J., Law, J., Di, X., Liu, W., Yang, G., Liu, G., *et al.* (2004) *Genome Res.* **14**, 414–425.
5. Kennedy, G. C., Matsuzaki, H., Dong, S., Liu, W. M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., *et al.* (2003) *Nat. Biotechnol.* **21**, 1233–1237.
6. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. (2001) *Nat. Genet.* **29**, 229–232.
7. Zhang, B., Kuntz, M. J., Goodwin, G. W., Edenberg, H. J., Crabb, D. W. & Harris, R. A. (1989) *Ann. N.Y. Acad. Sci.* **573**, 130–136.
8. Kadakol, A., Ghosh, S. S., Sappal, B. S., Sharma, G., Chowdhury, J. R. & Chowdhury, N. R. (2000) *Hum. Mutat.* **16**, 297–306.
9. Ishimi, Y. & Kikuchi, A. (1991) *J. Biol. Chem.* **266**, 7025–7029.
10. Mosammaparast, N., Ewart, C. S. & Pemberton, L. F. (2002) *EMBO J.* **21**, 6527–6538.
11. Steer, W. M., Abu-Daya, A., Brickwood, S. J., Mumford, K. L., Jordanares, N., Mitchell, J., Robinson, C., Thorne, A. W. & Guille, M. J. (2003) *Mech. Dev.* **120**, 1045–1057.
12. Vogel, T., Dittrich, O., Mehraein, Y., Dechend, F., Schnieders, F. & Schmidtke, J. (1998) *Cytogenet. Cell Genet.* **81**, 265–270.
13. Schnieders, F., Dork, T., Arnemann, J., Vogel, T., Werner, M. & Schmidtke, J. (1996) *Hum. Mol. Genet.* **5**, 1801–1807.
14. Lucet, V., de Bethmann, O. & Denjoy, I. (2000) *Biol. Neonate* **78**, 1–7.
15. Canning, B. J. & Mazzone, S. B. (2003) *Am. J. Med.* **115**, 45S–8S.
16. Kinney, H. C., Filiano, J. J. & White, W. F. (2001) *J. Neuropathol. Exp. Neurol.* **60**, 228–247.