Published in final edited form as:

Gastroenterology. 2008 August ; 135(2): 429–437. doi:10.1053/j.gastro.2008.05.036.

Identification of EpCAM as the Gene for Congenital Tufting Enteropathy

Mamata Sivagnanam^{1,2}, James L. Mueller^{3,4}, Hane Lee⁵, Zugen Chen⁵, Stanley F. Nelson⁵, Dan Turner⁶, Stanley H. Zlotkin⁶, Paul B. Pencharz⁶, Bo-Yee Ngan⁷, Ondrej Libiger⁸, Nicholas J. Schork⁸, Joel E. Lavine^{1,2}, Sharon Taylor^{1,2}, Robert O. Newbury⁹, Richard D. Kolodner^{3,10}, and Hal M. Hoffman^{2,3,4,10,11}

¹Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

²Department of Pediatrics, Rady Children's Hospital, 3020 Children's Way, San Diego, CA 92123

³Ludwig Institute of Cancer Research, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0669, USA

⁴Department of Pediatrics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

⁵Department of Human Genetics, University of California Los Angeles, University of California, Los Angeles, CA 90095, USA

⁶Division of Pediatric Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children, University of Toronto, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8

⁷Division of Pathology, Dept of Pediatric Lab Medicine, The Hospital for Sick Children, University of Toronto, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8

⁸Scripps Genomic Medicine, Scripps Health, La Jolla, California, USA; Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California, USA; Center for Human Genetics and Genomics, University of California San Diego, 10550 North Torrey Pines Road, La Jolla, CA, 92037 USA

⁹Department of Pathology, Rady Children's Hospital, 3020 Children's Way, San Diego, CA 92123

¹⁰Department of Medicine, Cellular and Molecular Medicine, and Moore's-UCSD Cancer Center, University of California San Diego, La Jolla, CA 92093-0669

¹¹Division of Rheumatology, Allergy and Immunology, Department of Medicine, and Moore's-UCSD Cancer Center University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0635, USA

Abstract

Correspondence: Mamata Sivagnanam, Division of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, University of California San Diego, 9500 Gilman Drive, MC 0669, La Jolla, CA 92093, USA, mengineer@ucsd.edu, phone 619-543-7544, fax 619-543-7537.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of Interest: None

Financial Disclosures: None

Background and Aims—Congenital Tufting Enteropathy (CTE) is a rare autosomal recessive diarrheal disorder presenting in the neonatal period. CTE is characterized by intestinal epithelial cell dysplasia leading to severe malabsorption and significant morbidity and mortality. The pathogenesis and genetics of this disorder are not well understood. The objective of this study was to identify the gene responsible for CTE.

Methods—A family with 2 children affected with CTE was identified. The affected children are double second cousins providing significant statistical power for linkage. Using Affymetrix 50K Single Nucleotide Polymorphism (SNP) chips, genotyping was performed on only two patients and one unaffected sibling. Direct DNA sequencing of candidate genes, RT-PCR, immunohistochemistry, and Western blotting were performed on specimens from patients and controls.

Results—SNP homozygosity mapping identified a unique 6.5 MB haplotype of homozygous SNPs on chromosome 2p21 where approximately 40 genes are located. Direct sequencing of genes in this region revealed homozygous G > A substitution at the donor splice site of exon 4 in *Epithelial Cell Adhesion Molecule* (EpCAM) of affected patients. RT-PCR of duodenal tissue demonstrated a novel alternative splice form with deletion of exon 4 in affected patients. Immuno-histochemistry and Western blot of patient intestinal tissue revealed decreased expression of EpCAM. Direct sequencing of *EpCAM* from two additional unrelated patients revealed novel mutations in the gene.

Conclusions—Mutations in the gene for EpCAM are responsible for Congenital Tufting Enteropathy. This information will be used to gain further insight into the molecular mechanisms of this disease.

Diarrhea is a major cause of neonatal death in the developing world. Although most diarrheal diseases are infectious or inflammatory in origin, the study of intrinsic intestinal diseases of infancy can provide a better understanding of the mechanisms of more common diarrheal diseases. Congenital tufting enteropathy (CTE) is a rare inherited intractable diarrhea of infancy characterized by villus atrophy and absence of inflammation. CTE presents in the first few months of life with chronic watery diarrhea and impaired growth¹. Most affected individuals are dependent on parenteral nutrition to acquire adequate caloric and fluid intake and allow for normal growth and development. This disease persists throughout life and imparts significant morbidity and mortality¹. Severe electrolyte imbalances can present early in the neonatal period, often before parents and physicians recognize any problem. Prolonged parenteral therapy brings inevitable complications such as liver disease, bacteremia, vascular complications², and poor quality of life³. Although small bowel transplant is a therapeutic option⁴, it carries its own risks, with 3-year survival rates for recipients after intestinal transplant approaching 30%⁵.

Since its initial description in 1994⁶, several case reports of CTE patients have been published^{7–9}, but little is known about its incidence or pathogenesis. By some accounts the incidence is estimated at 1/50,000 -100,000 live births in Western Europe¹⁰. Many patients are likely not recognized because survival is dependent on immediate aggressive therapy, and diagnosis requires comprehensive pathologic evaluation for confirmation. The inheritance pattern of this disorder in reported kindreds suggests autosomal recessive inheritance, but no formal genetic studies have been published.

The diagnosis of CTE is made by recognition of villus changes of the epithelium of the small intestine. Findings include total or partial villus atrophy and crypt hyperplasia without evidence of inflammation¹. Focal epithelial tufts are characteristically found in the duodenum and jejunum⁸. These tufts are composed of enterocytes with rounding of the plasma membrane resulting in teardrop like configuration (Fig. 1A). Pathologic studies have demonstrated differences in desmosomes as well as alterations in the distribution of the $\alpha 2/\beta 1$ -integrin adhesion molecule subunit¹¹. Other histologic studies have reported changes in extracellular

matrix such as reduced laminin expression in the intestinal crypts¹². Changes reported in integrins and laminins suggest that dysfunctional epithelial cell interactions and adhesion play a role in the pathogenesis of CTE. Intestinal features resembling CTE are seen in a knockout mouse in which the gene encoding the transcription factor Elf3 is disrupted¹³. However, variations in this gene have not been reported in CTE patients. Given the neonatal abnormalities in the intestine associated with CTE, understanding the genetic basis for this disease would be expected to provide important insights into the development and biology of the intestine.

Although CTE was described more than 10 years ago, the pathogenesis of CTE remains poorly understood due to the severity and rarity of the disease. Elucidating the genetic basis for rare conditions such as CTE has been difficult without large cohorts of patients and kindreds. The recent evolution in the field of human genetics in the past 10 years has enhanced the approaches for identifying human disease genes. In this study, we exploit powerful modern genomic analysis techniques to identify the gene responsible for CTE using only two affected patients from the same family.

Materials and Methods

Patients

Informed consent of the subjects/parents was obtained according to the local Institutional Review Board guidelines. Five patients were recruited, two from the same family and three from unrelated kindreds. The gene was identified in one large Mexican-American kindred (Pedigree 1) with two male affected subjects (P1 and P2). The other three subjects from The Hospital for Sick Children (Toronto, ON) were recruited to replicate the results (P3, P4 and P5). In all five, diagnosis of CTE was confirmed by typical clinical presentation and multiple endoscopic duodenal biopsies with characteristic histology. Patient P1 presented at 7 weeks of life with diarrhea and failure to thrive. Patient P2 presented at 4 weeks of age with dehydration, diarrhea, and failure to thrive. Both P1 and P2 were unable to sustain normal growth on enteral feeds. Therefore, supplemental parenteral nutrition was initiated and remains their predominant source of nutrition. Patient P3 presented with diarrhea at one month. Initial biopsies were inconclusive but repeated histological examination at one year of age was consistent with CTE. Patient P4 presented with failure to thrive at 3 weeks of age. Intestinal biopsy was performed at two months of age confirming a diagnosis of CTE. Patient P5 was symptomatic from 3 weeks of age. Biopsies at 4 months and 7 months showed villus atrophy and tufts. All five patients remain on parenteral nutrition for at least 40% of their caloric needs (Table 1).

Available unaffected subjects (parents/siblings) were also recruited and after informed consent, blood samples (Gentra Systems) or saliva samples (Oragene) were collected and genomic DNA was extracted. In addition, available endoscopic duodenal tissue from affected subjects and age matched normal controls was obtained. Controls include those without pathology (N1–N4) as well as those with inflammatory bowel disease (N5, N6). Anonymous North American normal control DNA was obtained from Coriell Cell Repositories (M450PDR). Mexican American controls without congenital diarrhea were obtained from Jeanette McCarthy at San Diego State University.

Linkage Mapping

SNP mapping was performed using the 50K genotyping chip according to the Affymetrix Gene Chip Mapping Assay Manual (Affymetrix Santa Clara, CA) at the UCLA DNA Microarray Facility. For linkage analysis, the SNP results generated by Affymetrix GeneChip® DNA Analysis Software were converted to different formats by scripts written in the Nelson Lab for each program used (available on request). The identity by descent (IBD) mapping algorithm written by Merriman et al.¹⁴ was used for homozygous block and IBD detection. A

conservative error rate of 1% was used to allow the software to tolerate possible genotyping errors. MAPMAKER/HOMOZ version 0.9¹⁵ was used to calculate the multipoint logarithmic odds (LOD) score for the putative susceptibility haplotype of the region of interest between markers rs2166746 and rs4364055. Based on the allele frequencies of 235 markers, we conservatively assumed a rare disease allele frequency of 0.01 in the general population, a disease frequency of 0.001, and a penetrance of 99%. Since the genetic relationship between the first generations of the two families was likely to be further than siblings, we conservatively calculate the LOD score by assuming that one great grandparent on each side of the family were siblings not documented in the pedigree.

Candidate Gene Evaluation

A list of candidate genes present in the genomic region identified by linkage analysis was generated using University of California-Santa Cruz Genome Bioinformatics Site resources¹⁶. There were 38 known genes including 8 encoding hypothetical proteins in this region. As CTE is a disease of the gastrointestinal tract, known genes showing expression in the duodenum, ileum, jejunum and colon were initially selected as priority candidate genes for screening¹⁶. Primers were designed to amplify the coding exons of the genes of interest including intron/exon boundaries (Primer3). Direct sequencing of PCR amplified products of DNA obtained from affected and available unaffected subjects and anonymous normal North American and Mexican American controls was performed using a 3730xl DNA analyzer (ABI).

Duodenal RNA expression

After informed consent of all the individuals, fresh-frozen duodenal biopsies were collected from available affected patient P2 and an age-matched control N1. Tissue homogenization and RNA isolation were performed according to Trizol protocol (Invitrogen). cDNA was generated using Multiscribe™ TaqMan reverse transcriptase (Applied Biosystems) and used as template for PCR with primers in 5' and 3' untranslated regions of *EpCAM*. Full-length wild-type and mutant *EpCAM* cDNAs were cloned into TOPO 2.1 TA and the clones were confirmed by sequencing. The inserts were then excised with restriction enzymes and analyzed by electrophoresis through a 1% agarose gel using Hyperladder IV (Bioline) as DNA size standards.

Immunohistochemistry

Immunofluorescent staining of available formaldehyde fixed, paraffin-embedded duodenal biopsy tissue was performed. Two samples from each affected patient, P1 and P2, and 1 sample from each patient P3, P4, and P5 were stained along with 2 age matched normal controls (N2, N3) and one patient with inflammatory bowel disease (N5). Paraffin-embedded tissue sections with 5 µm thickness were mounted on a glass slide and allowed to dry. Slides were deparaffinized, quenched with 3% H₂O₂, immersed in citric acid buffer, processed in a microwave oven at 95C for 10 min and blocked with TSA buffer. Mouse monoclonal anti-*EpCAM* antibody (clone 323/A3 Abcam, Cambridge, MA) was applied at a dilution of 1:50 overnight. Fluorescent secondary antibody (mouse IgG) at a dilution of 1/200 was applied. Slides were washed and mounted with VECTASHIELD HardSet Mounting Medium with DAPI. Isotype controls lacking primary antibody as well as those lacking secondary antibody were performed. Immunofluorescent staining (primary clone 323/A3, secondary FITC-Jackson ImmunoResearch Laboratories, INC. West Grove, PA) in wild-type and mutant (deletion of exon 4) *EpCAM* transfected 293 cells (FuGENE) revealed similar intensity staining of *EpCAM* in cells expressing both forms confirming the presence of the antibody epitope in the mutant form.

Western Blotting

One piece of flash frozen duodenal tissue collected from patient P2, two normal controls (N1, N4) and one patient with inflammatory bowel disease (N6) were collected and ground in Kontes tissue grinder (Fisher 885451-0020) with 100 μ l of complete sample buffer (50mM Tris pH 7.8, 50mM NaCl, 0.1% NP40, 5mM EDTA, 10% glycerol, 1 tablet Complete Mini Protease Inhibitor Cocktail Tablet, Roche Applied Science). Cell lysate from 293 cells transfected with wild-type EpCAM, mutant EpCAM (lacking exon 4) and not transfected cells were used as further controls. Thirty μ g of whole cell protein sample were mixed with 30 μ l of loading buffer before separation by a gel (Criterion Pre-CAST gel, BioRad). After completion of electrophoresis, samples were transferred to polyvinylidene difluoride membrane filter (Immuno-Blot PVDF membrane, Bio-Rad). The transferred samples were incubated overnight with each antibody: sc-25308 to EpCAM (1:500, mouse monoclonal antibody), 311-1k1 to EpCAM (1:500, mouse monoclonal), EP700Y to E-Cadherin (1:10,000, rabbit monoclonal) and Actin (1:30,000, mouse monoclonal antibody). The second antibodies (1:2000, ECLTM, Anti Mouse IgG, GE Healthcare) for sc-25308, 311-1k1 and (1:2000, ECLTM, Anti Rabbit IgG, GE Healthcare) for EP700Y were incubated for 60 minutes, rinsed and then incubated with enhanced chemiluminescence Western blotting detection reagent (ECL+Plus, GE Healthcare) for 1 minute. The membrane was exposed to X-ray film for 0.5–10 min.

Results

We identified a kindred of Mexican-American descent (Pedigree 1; Fig. 2A) that includes 2 boys presenting with congenital diarrhea. Duodenal biopsies revealed severe villus blunting and epithelial tufts consistent with a diagnosis of CTE (Fig. 1B). There was no reported consanguinity in the kindred, but the 2 affected children are double second cousins creating a unique genetic relationship (Fig. 2A).

A panel of over 50,000 SNPs (Affymetrix 50K) was typed on 3 individuals including 2 affected subjects and one unaffected sibling. The data from this analysis resulted in high information content across the genome. The content percentages for the three assays were 95.37%, 98.08% and 99.74%, and the distribution of the genotype calls (AA, AB, and BB) was comparable across the dataset. Error rates for this study could not be directly measured, but are likely below 0.5% based on published experience with a comparable platform¹⁷.

Since the ancestors were not affected, we searched for shared homozygous segments in two of the affected individuals, based on the hypothesis that the two affected double second cousins are indeed the offspring of consanguineous matings, not revealed by the pedigree (Fig. 2A), and the likely hypothesis that the disease is autosomal recessive. To support the hypothesis of an undocumented inbreeding loop, we noted that there was strong evidence of inbreeding as each individual had more than one homozygous block that was significantly larger than the blocks seen in a control outbred population (p -value $< 10^{-20}$) of 75 individuals. Thus, each affected child's genome was consistent with a distant consanguineous mating, but clearly more distant than a first cousin mating. Furthermore, the 2 affected second cousins shared more haplotypes than would be expected (19.5%) between outbred double second cousins (~12%) suggesting a close relationship. Thus, directly from observing individuals' genotypes, we were able to determine that some combination of great grandparents of these individuals must be related to each other creating an inbreeding loop. We thus searched for regions of homozygosity in the two affected subjects, and identified a single such interval of 6.5Mb on chromosome 2, spanning 46504175-53011057, with a common haplotype in both affected individuals (Fig. 2B). This region (represented by a peak) was the largest area of shared homozygosity across the whole genome between the two affected subjects. The unaffected control sibling was heterozygous over this interval (Fig. 2C).

To estimate the significance of the linkage, we used MAPMAKER/HOMOZ to calculate the multipoint LOD score for the putative susceptibility haplotype. We assumed that the parents are third cousins since the relationship between the first generations of the two families was estimated to be further than siblings. We therefore performed the analysis under an autosomal recessive inheritance model with a rare population frequency of the allele (0.01) to determine a cumulative LOD score of 4.7. We thus consider the linkage to be of genome-wide significance.

On the basis of the physical interval, a list of 38 known genes was generated of which 12 were expressed in the intestine. We prioritized candidates for direct sequencing using information on gene function available in public databases¹⁶. We identified a homozygous G>A substitution in the affected patients at the donor splice site (c.491+1G>A) of exon 4 of *EpCAM* (Fig. 2D). The parents and unaffected sibling were found to be heterozygous for this variant consistent with autosomal recessive inheritance. This mutation was not found in any of 400 healthy controls, 200 normal controls representing the racial cross section of North America and another 200 Mexican-American controls.

Direct sequencing of all 9 exons of *EpCAM* from the three additional patients (P3, P4, and P5) revealed novel mutations in DNA from P3 and P4. A homozygous exon 4 acceptor splice site mutation G>A was identified in DNA from Patient P3 (c.427-1G>A). A heterozygous c.200G>A substitution within exon 3 resulting in a missense mutation predicted to cause a cysteine to tyrosine change at position 66 (C66Y) was found in patient P4 (Table 1). The mutations found in patients P3 and P4 were not found more than 170 North American control DNAs.

We isolated RNA from duodenal tissue from P2 and an age-matched normal control. Studies using RT-PCR and primers constructed in the 5' and 3' untranslated regions revealed a slightly smaller PCR product derived from the cDNA of affected duodenal tissue. Direct sequencing (data not shown) of these products revealed a novel alternative splice form in the affected patients that results in the complete deletion of exon 4 (66 base pairs) from *EpCAM* mRNA (Fig. 3B and 3D).

To further study the effect of this *EpCAM* variant on protein expression in the intestine, fluorescent immunohistochemical (IHC) staining was performed on formaldehyde-fixed, paraffin embedded duodenal biopsy tissue from all 5 patients and age matched controls using antibody 323/A3 (the epitope for this mAb maps to the first EGF-like domain of *EpCAM* which is encoded by exon 2) to *EpCAM*. Epithelial *EpCAM* staining was absent or markedly decreased in tissue from all 5 affected subjects; staining was normal from age-matched unaffected subjects and one control patient with inflammatory bowel disease. Photomicrographs shown are representative of multiple biopsies obtained at different times from affected patients P1–P5, normal control duodenal tissue, and isotype control (Fig. 3E).

Using *EpCAM* antibodies sc-25308 (epitope amino acids 24–93), 311-1k1 (epitope amino acids 141–219), and E144 (epitope C-terminus), *EpCAM* expression was found to be significantly decreased in intestinal tissue from patient P1 with CTE, compared with 2 normal and 1 IBD patient. Presence of intestinal epithelium was confirmed with similar E-Cadherin expression in all patient specimens (mAB EP700Y), and equivalent total protein was confirmed by similar Actin expression (Fig 3F). 293 cells transfected with wild-type and mutant *EpCAM* showed detectable bands corresponding to *EpCAM* using mAB sc-25308 (Fig 3F) and E144 (data not shown), but not 311-1k1 consistent with its epitope near the deleted exon (Fig 3F). E-Cadherin was also confirmed in all three cell line controls upon overexposure of the blot (data not shown).

Discussion

Congenital Tufting Enteropathy is a rare autosomal recessive diarrheal disorder presenting in the neonatal period with significant morbidity and mortality. Using a family with 2 children affected with CTE, Single Nucleotide Polymorphism genotyping was performed revealing a unique 6.5 MB haplotype of homozygous SNPs on chromosome 2p21. Direct sequencing of genes in this region revealed homozygous G > A substitution at the donor splice site of exon 4 in *Epithelial Cell Adhesion Molecule* of affected patients. RT-PCR of duodenal tissue demonstrated a novel alternative splice form with deletion of exon 4 in affected patients. Immunohistochemistry and Western blot of patient intestinal tissue revealed decreased expression of EpCAM. Direct sequencing of *EpCAM* from two additional unrelated patients revealed 2 additional mutations in the gene. The identification of *EpCAM* as the gene responsible for CTE will not only improve the diagnosis of this congenital diarrhea, but it is an important step in the understanding of the underlying pathophysiology and mechanisms involved in normal and abnormal intestinal morphogenesis and differentiation.

This study highlights the power of modern genetic technology to identify disease genes associated with rare diseases using a small number of affected patients. SNP genotyping allows for dense whole genome analysis and identification of linkage that was previously impossible. Recently, several large genome wide association studies using this methodology have provided clues to the genetic basis of common diseases, such as diabetes mellitus, breast cancer and Crohn's disease, using large cohorts of patients and controls^{18, 19}. However, this technology has not been widely applied to the study of rare disease genes. Traditional microsatellite marker mapping has been used in the mapping of rare disease genes, even in relatively small inbred families²⁰. Here, we have applied SNP technology for the identification of a gene responsible for CTE by analysis of only two distantly related affected patients with a unique relationship. The use of SNPs allows for a high likelihood for the identification of linkage.

Like most cell adhesion molecules, the primary function of EpCAM appears to be cell-cell interaction. This is supported by studies with L929 fibroblasts which are normally incapable of cellular adhesion, but form multicellular aggregates of cells when expressing EpCAM, suggesting involvement in homotypic cell-cell interactions²¹. EpCAM is known to recruit intracellular α -actinin to the sites of homophilic contacts²². EpCAM also co-localizes with E-cadherin in the areas of cell-cell junctions and directly associates with claudin-7, a tight junction protein²³.

Using RT-PCR we demonstrate a 66 bp in-frame deletion in *EpCAM* small intestinal mRNA in the affected patients only. This variation does not result in a frameshift; therefore translation of the C terminal portion of the EpCAM protein is not likely to be affected. Lack of immunofluorescent staining and lower levels EpCAM on Western Blot in affected patient tissue suggests the splice site mutation affects protein expression in the intestinal tissue. The function of the domain coded for by exon 4 is not known (Fig. 3C), but its deletion could affect protein stability, localization, or function. This may occur by altering post-translational modifications such as proteolytic cleavage, homophilic adhesion (mediated by the EGF domains), or transmembrane domain anchoring of EpCAM to intestinal epithelial cell membranes. Mutations in *EpCAM* may disrupt its association with α -actinin, claudin-7 or E-Cadherin leading to mucosal integrity breakage and intestinal failure seen in CTE. Interestingly, deletion of a small portion of the extracellular domain of EpCAM completely abolishes the interaction of the intracellular domain with α -actinin²⁴. It is possible that deletion of exon 4 has a similar consequence.

Cell adhesion molecules have also received much attention for their morphoregulatory roles in development and influence in specifying cell fate in numerous tissues²⁵. In fact, EpCAM

is primarily known for its potential role in tumorigenesis resulting from increased expression on the cell surface of human carcinoma cells including tumors of the gastrointestinal system, breast, thyroid, and kidney²² and therefore it is being studied as a target for cancer therapy²⁶.

We speculate that EpCAM plays a role in normal intestinal development, as described in the pancreas²⁷, resulting in the pathologic findings observed in CTE. EpCAM function may be important for the development of the crypt villus axis, where epithelial cells originate from stem cells in the crypt and migrate distally to the tip of the villus prior to shedding. Mechanisms that lead to apoptosis of intestinal epithelial cells are still not completely clear, but it is plausible that dysregulation of apoptosis in intestinal epithelial cells may play a pathogenic role in diseases such as intestinal dysplasia and carcinomas^{28, 29}.

Many families reported with CTE are consanguineous or follow a pattern consistent with autosomal recessive inheritance^{2, 9}, as was demonstrated in our initial family. However, analysis of patient P4 revealed a heterozygous missense mutation in an exon coding for an important extracellular functional domain (EGF2), and no additional coding mutations were identified. In this patient, it is possible that CTE is transmitted in an autosomal dominant fashion. Alternatively, compound heterozygosity with a second mutation in an unsequenced noncoding region is also possible.

Residual gut function and longevity vary among CTE patients. In addition, some patients have associated malformations including punctated keratitis, choanal atresia, and esophageal atresia^{10, 30}. The clinical phenotype spectrum in CTE may be explained by different mutations in the same gene or mutations in other genes. An *EpCAM* mutation was not found in one of our patients suggesting genetic heterogeneity. Interestingly, this patient (P5) is currently 20 years old and his survival suggests less severe disease. It is also possible that sequence variants within the promoter or intronic regions of the *EpCAM* gene could contribute to disease development. Genes that code for proteins with significant homology, similar function, or that form complexes with *EpCAM* are also candidate genes for further study.

The identification of this CTE mutation will improve our understanding of this disorder, and offer new research directions in this field. Furthermore, our findings should elucidate the essential role of adhesion molecules in the development of the gastrointestinal system. This approach illustrates the utility of using unique kindreds and powerful new genetic technology to better characterize difficult to study rare diseases.

Acknowledgements

Grant Support: This work was supported in part by an NIH T32DK07202 and American College of Gastroenterology Clinical Research Award.

References

1. Sherman PM, Mitchell DJ, Cutz E. Neonatal enteropathies: defining the causes of protracted diarrhea of infancy. *J Pediatr Gastroenterol Nutr* 2004 38;:16–26. [PubMed: 14676590]
2. Gambarara M, Diamanti A, Ferretti F, et al. Intractable diarrhea of infancy with congenital intestinal mucosa abnormalities: outcome of four cases. *Transplant Proc* 2003;35:3052–3053. [PubMed: 14697977]
3. Jeppesen PB, Langholz E, Mortensen PB. Quality of life in patients receiving home parenteral nutrition. *Gut* 1999;44:844–852. [PubMed: 10323888]
4. Paramesh AS, Fishbein T, Tschernia A, et al. Isolated small bowel transplantation for tufting enteropathy. *J Pediatr Gastroenterol Nutr* 2003;36:138–140. [PubMed: 12500010]

5. Kato T. Intestinal and Multivisceral Transplantation. *World Journal of Surgery* 2002;26:226–237. [PubMed: 11865353]
6. Reifen RM, Cutz E, Griffiths AM, et al. Tufting enteropathy: a newly recognized clinicopathological entity associated with refractory diarrhea in infants. *J Pediatr Gastroenterol Nutr* 1994;18:379–385. [PubMed: 8057225]
7. Cameron DJ, Barnes GL. Successful pregnancy outcome in tufting enteropathy. *J Pediatr Gastroenterol Nutr* 2003;36:158. [PubMed: 12500014]
8. Beck NS, Kang IS, Suh YL. Protracted diarrhea: results of the five-year survey in a tertiary hospital in Korea. *J Korean Med Sci* 2001;16:736–741. [PubMed: 11748354]
9. Goulet OJ, Brousse N, Canioni D, et al. Syndrome of intractable diarrhoea with persistent villous atrophy in early childhood: a clinicopathological survey of 47 cases. *J Pediatr Gastroenterol Nutr* 1998;26:151–161. [PubMed: 9481629]
10. Goulet O, Salomon J, Ruemele F, et al. Intestinal epithelial dysplasia (tufting enteropathy). *Orphanet J Rare Dis* 2007;2
11. Patey N, Scoazec JY, Cuenod-Jabri B, et al. Distribution of cell adhesion molecules in infants with intestinal epithelial dysplasia (tufting enteropathy). *Gastroenterology* 1997;113:833–843. [PubMed: 9287975]
12. Teller IC, Beaulieu JF. Interactions between laminin and epithelial cells in intestinal health and disease. *Expert Rev Mol Med* 2001;2001:1–18. [PubMed: 14585148]
13. Ng AY, Waring P, Ristevski S, et al. Inactivation of the transcription factor Elf3 in mice results in dysmorphogenesis and altered differentiation of intestinal epithelium. *Gastroenterology* 2002;122:1455–1466. [PubMed: 11984530]
14. Cha Y, Lee H, Jen JC, et al. Episodic vertical oscillopsia with progressive gait ataxia: Clinical description of a new episodic syndrome and evidence of linkage to chromosome 13q. *J Neurol Neurosurg Psychiatry*. 2007
15. Kruglyak L, Daly MJ, Lander ES. Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. *Am J Hum Genet* 1995;56:519–527. [PubMed: 7847388]
16. <http://www.genome.ucsc.edu>
17. Lee H, Jen JC, Wang H, et al. A genome-wide linkage scan of familial benign recurrent vertigo: linkage to 22q12 with evidence of heterogeneity. *Human Molecular Genetics* 2006;15:251–258. [PubMed: 16330481]
18. Carlson CS, Eberle MA, Kruglyak L, et al. Mapping complex disease loci in whole-genome association studies. *Nature* 2004;429:446–452. [PubMed: 15164069]
19. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–1093. [PubMed: 17529967]
20. Boerlum AD, Balslev T, Haagerup A, et al. A new locus for Seckel syndrome on chromosome 18 p 11. 31-q 11. 2. *European Journal of Human Genetics* 2001;9:753–757. [PubMed: 11781686]
21. Litvinov SV, Velders MP, Bakker HA, et al. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J Cell Biol* 1994;125:437–446. [PubMed: 8163559]
22. Balzar M, Winter MJ, de Boer CJ, et al. The biology of the 17-1A antigen (Ep-CAM). *J Mol Med* 1999;77:699–712. [PubMed: 10606205]
23. Ladwein M, Pape UF, Schmidt DS, et al. The cell-cell adhesion molecule EpCAM interacts directly with the tight junction protein claudin-7. *Exp Cell Res* 2005;309:345–357. [PubMed: 16054130]
24. Balzar M, Briaire-de Bruijn IH, Rees-Bakker HA, et al. Epidermal growth factor-like repeats mediate lateral and reciprocal interactions of Ep-CAM molecules in homophilic adhesions. *Mol Cell Biol* 2001;21:2570–2580. [PubMed: 11259604]
25. Takeichi M. Morphogenetic roles of classic cadherins. *Curr. Opin. Cell Biol* 1995;7:619–627. [PubMed: 8573335]
26. Adkins JC, Spencer CM. Edrecolomab (monoclonal antibody 17-1A). *Drugs* 1998;56:619–626. [PubMed: 9806108]
27. Trelstad RL. *The Role of Extracellular Matrix in Development*. AR Liss. 1984
28. Jones BA, Gores GJ. *Physiology and pathophysiology of apoptosis in epithelial cells of the liver, pancreas, and intestine: the American Physiological Society*. 1997

29. Damiano JS, Cress AE, Hazlehurst LA, et al. Cell Adhesion Mediated Drug Resistance (CAM-DR): Role of Integrins and Resistance to Apoptosis in Human Myeloma Cell Lines. *Blood* 1999;93:1658–1667. [PubMed: 10029595]
30. Bird LM, Sivagnanam M, Taylor S, et al. A new syndrome of tufting enteropathy and choanal atresia, with ophthalmologic, hematologic and hair abnormalities. *Clinical Dysmorphology* 2007;16:211–221. [PubMed: 17786112]
31. We would like to thank the patients and their families for participation in studies. We thank Lynne Bird, Jorrit Enserink, Kim Barrett and R. S. for helpful discussions and Amir Misaghi, Scott Anderson, Justin Anderson, Tikvah Hayes and Rahsaan Taylor for technical assistance. We thank Jeannette McCarthy for providing control specimens. This work was supported in part by an NIH T32DK07202 and American College of Gastroenterology Clinical Research Award

Abbreviations

CTE, Congenital Tufting Enteropathy; EpCAM, Epithelial Cell Adhesion Molecule.

Fig. 1. Schematic of duodenal mucosa showing histology of (A) Normal intestinal villus and (B) Congenital tufting enteropathy villus with crowded epithelial cells forming tufts, villus atrophy. (C) H&E stained duodenal tissue (20X magnification) from affected patients (P1,P2) exhibiting tufting and crowding of epithelial cells.

Fig. 2. (A) Pedigree 1: family with congenital tufting enteropathy including affected patients (P1, P2) and unaffected siblings (U1, U2). Generations are indicated by Roman numerals on the left. Affecteds (P1, P2) are double second cousins. (B) Homozygosity Mapping: Genome-wide homozygosity intervals are plotted for individuals P1 and P2. The y-axis indicates the length of the intervals and the x-axis shows relative chromosomal position with chromosomes indicated across the top of each panel. The homozygous blocks shared between P1 and P2 are shown in the bottom bar graph (P1 & P2). The interval identified as homozygous in both individuals is highlighted with a gray box bar across all three panels (C) Haplotype Structure of the Pedigree. Homozygous blocks on chromosome 2p21 that are shared by the two affected

descendants (P1, P2) are shown as black bars with genotypes at the flanking ends shown with SNP IDs from dbSNP. 220 SNPs in the middle of each block are not shown for simplicity as all 220 SNPs are homozygous and common in both affected individuals. (D) *EpCAM* exon 4 donor splice site mutation in Pedigree 1: Nucleotides across x-axis with mutation in bold. Coding nucleotides in capital letters and intronic nucleotides in lower case. Homozygous mutant affected patients (P1, P2), Heterozygous unaffected sibling (U2), Normal Control (C1)

Fig. 3.

(A) Genomic region surrounding *EpCAM* with exons labeled and CTE patient (P1, P2, P3, P4) mutations noted. Nucleotide and amino acid coordinates of mutations are shown assuming the A of the ATG codon is nucleotide 1. (B) Schematic representation of wild type and mutant RNA with variant splicing of Exon 4 as seen in patients P1 and P2 (C) *EpCAM* protein with Epidermal Growth Factor domains (EGF I and II) and transmembrane region (TM). Dashed box represents area coded by Exon 4. Predicted location of cysteine to tyrosine missense mutation at aa position 66 (C66Y) found in patient P4 (D) RT-PCR products of *EpCAM* from cDNA of control (WT) and affected patient (MUT) duodenal tissue demonstrating smaller product size of mutant (E) Fluorescent Immunohistochemistry of duodenal biopsies with 323/

A3 EpCAM antibody (green) wild type (WT), mutant (MUT) and isotype control (IC). Demonstrating minimal non-specific staining of affected patient tissue (MUT) and epithelial predominance unaffected patient tissue (WT) (F) Patient Tissue Western Blot Analysis demonstrating decreased protein expression of EpCAM (mAB sc-25308 and 311-1k1) in CTE affected patient (P2) as compared with normal (N1, N4) and IBD patient (N6), equal levels of E-Cadherin (mAB EP700y) and Actin. 293 cell western blot analysis of 293 cells transfected with wild-type EpCAM (WT), mutant EpCAM (lacking exon 4) (MUT) and non-transfected cells (NT). Detectable bands corresponding to EpCAM using mAB sc-25308, but not 311-1k1 consistent with its epitope near the deleted exon 4.

Table 1

Phenotype and Mutations found in CTE Patients

Nucleotide and amino acid coordinates are given assuming the A of the ATG translation initiation codon is nucleotide 1. The normal control frequency refers to the number of times an indicated nucleotide change was found in the indicated number of control DNA samples analyzed. Parenteral nutrition requirement reflects percentage of total caloric intake provided parenterally.

Patient	Ethnicity	Age at Diagnosis	Current Age	Known Consanguinity	Variant in EpCAM	Normal Control Frequency
P1	Mexican American	14 months	3.5 yrs	None*	c.491+1G>A	0/200
P2	Mexican American	11 months	2.5 yrs	None*	c.491+1G>A	0/200
P3	Native Canadian	1 year	12 yrs	Great Grandparents were siblings	c.427-1G>A	0/171
P4	Russian	2 months	5 yrs	None	c. 200G>A	0/171
P5	English/Italian	4 months	20 yrs	None	None	N/A