Electronical Supplementary Document

Methods: Genomic DNA was isolated from peripheral blood leukocytes by standard procedures. Whole exome target enrichment was performed with 0.5 µg of genomic DNA from patient 2 and with the Agilent SureSelectHuman All Exon 60 Mb Capture kit (Agilent Technologies, Santa Clara, CA), and 2 x 150 bp paired-end sequences were produced using an Illumina Hi-Seq4000 platform. Variants were identified in genes known to cause forms of congenital diarrheas using the SeattleSeq Annotation server (http://snp.gs.washington.edu/SeattleSeqAnnotation138/), and were filtered for autosomal recessive mode of inheritance, predicted effect on protein expression, and allele frequency of <0.005 in the Exome Aggregation Consortium (http://exac.broadinstitute.org/) database. The identified missense variant was evaluated in silico to estimate pathogenicity by PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2), CADD (http://cadd.gs.washington.edu/score), and MutationTaster (http://www.mutationtaster.org/). A copy-number detection in targeted NGS data was performed using panelcn.MOPS (https://ml.jku.at/software/panelcnmops/). Sanger sequencing was used to determine the segregation of the EPCAM variant detected by exome sequencing. EPCAM variant designations are based on the NCBI reference sequence NM_002354.2.