ORIGINAL ARTICLE

The Genetic Architecture of Congenital Diarrhea and Enteropathy

Zeenat Gaibee, M.D.,[1] Neil Warner, Ph.D.,[2] Katlynn Bugda Gwilt, Ph.D.,[3] Wenjuan Li, Ph.D.,[2] Rei Guan, Ph.D.,[2] Michael Yourshaw, Ph.D.,[4]* Ryder Whittaker Hawkins, Ph.D.,[2] Christiane Zorbas, Ph.D.,[5] Jonathan St.-Germain, Ph.D.,[6] Mahdi Tabatabaie, B.Sc.,[6] Suli Mao, B.Sc.,[7] Vered Pinsk, M.D.,[8]* Baruch Yerushalmi, M.D.,[8] Lee-kai Wang, Ph.D.,[4] Stanley F. Nelson, M.D.,[4] Laura Wozniak, M.D.,[9] Dror S. Shouval, M.D.,[10] Manar Matar, M.D.,[10] Amit Assa, M.D.,[11] Nathaniel Frost, B.Sc.,[2] Lissette Jimenez, M.D.,[3] Sari Acra, M.D., M.P.H.,[12] Thomas Walters, M.D.,[1] Stephen Mouat, M.D.,[13] Michael Li, B.C.S.,[14] Denis L.J. Lafontaine, Ph.D.,[5] Matthew Tyska, Ph.D.,[7] Brian Raught, Ph.D.,[6] Yaron Avitzur, M.D.,[1] Wayne I. Lencer, M.D.,[3] James R. Goldenring, M.D., Ph.D.,[7,15] Martín G. Martín, M.D., M.P.P.,[16] Jay R. Thiagarajah, M.D., Ph.D.,[3] and Aleixo M. Muise, M.D., Ph.D.[1,2,17]

ABSTRACT

BACKGROUND

Next-generation sequencing has enabled precision therapeutic approaches that have improved the lives of children with rare diseases. Congenital diarrhea and enteropathies (CODEs) are associated with high morbidity and mortality. Although treatment of these disorders is largely supportive, emerging targeted therapies based on genetic diagnoses include specific diets, pharmacologic treatments, and surgical interventions.

METHODS

We analyzed the exomes or genomes of infants with suspected monogenic congenital diarrheal disorders. Using cell and zebrafish models, we tested the effects of variants in newly implicated genes.

RESULTS

In our case series of 129 infant probands with suspected monogenic congenital diarrheal disorders, we identified causal variants, including a new founder *NEUROG3* variant, in 62 infants (48%). Using cell and zebrafish models, we also uncovered and functionally characterized three novel genes associated with CODEs: *GRWD1*, *MYO1A*, and *MON1A*.

CONCLUSIONS

We have characterized the broad genetic architecture of CODE disorders in a large case series of patients and identified three novel genes associated with CODEs. (Funded by the National Institutes of Health and others.)

Author affiliations are listed at the end of the article. Dr. Muise can be contacted at aleixo.muise@utoronto.ca or at the Hospital for Sick Children, 555 University Ave., Toronto, ON M5G 1X8, Canada; Dr. Thiagarajah can be contacted at jay.thiagarajah@childrens.harvard.edu or at Boston Children's Hospital, EN605, 300 Longwood Ave., Boston, MA 02115; Dr. Martín can be contacted at mmartin@mednet.ucla.edu or at the David Geffen School of Medicine at UCLA, 650 Charles E. Young Dr. South, CHS Rm. 67-200H, Los Angeles, CA 90095.

*Deceased.

Drs. Gaibee, Warner, Bugda Gwilt, W. Li, Guan, Avitzur, Lencer, Goldenring, Martín, Thiagarajah, and Muise contributed equally to this article.

N Engl J Med 2025;392:1297-309.
DOI: 10.1056/NEJMoa2405333

Congenital diarrhea and enteropathies (CODEs) are a group of rare disorders that primarily affect the function of intestinal epithelial cells, leading to infantile-onset diarrhea and poor growth. Molecular defects in CODEs can be classified into six categories: epithelial trafficking and polarity, immune-cell-regulation, nutrient and electrolyte transport, enteroendocrine-cell development, nutrient metabolism, and other.[1,2] CODEs are associated with substantial morbidity and mortality. Patients often receive lifelong fluid and nutritional management.[3] Genetic causes include pathogenic variants in *MYO5B* (microvillus inclusion disease),[4] *EPCAM* (tufting enteropathy[5]), *NEUROG3* (enteric anendocrinosis[6]), *DGAT1* (protein-losing enteropathy[7]), and *SLC9A3* (congenital sodium diarrhea[8]). Treatment options are currently limited. However, an understanding of some of the genetic causes of CODEs has led to targeted therapies such as dietary treatments[3] and the development of preclinical pharmacologic treatments,[2] such as leflunomide[9] for the treatment of intestinal epithelial defects observed in patients with TTC7A defiency.[10] One study examined 1005 patients with monogenic inflammatory bowel disease,[11] and small single-center case series have examined 6 patients and 24 patients with congenital diarrheal disorders.[12,13] Here, we report a multicenter study that aimed to systematically examine genes that underpin CODE disorders.

METHODS

PATIENT POPULATION

Infants had a clinical diagnosis of congenital diarrhea, defined as chronic diarrhea lasting more than 2 weeks and starting in infancy with no attributed anatomical, infectious, or allergic cause. (The article by Thiagarajah et al.[1] discusses specific types of diarrhea based on stool osmolality, including osmotic, secretory, and mixed.) They were treated locally at the Hospital for Sick Children (HSC) in Toronto, Vanderbilt University Medical Center (VUMC), Boston Children's Hospital (BCH), and the University of California, Los Angeles (UCLA), and were referred from across the United States to VUMC, BCH, and UCLA and from several international institutions to BCH and UCLA. Families who previously consented to genetic testing and further research studies were included in the analysis as part of the Pediatric Congenital Diarrhea and Enteropathies Consortium (PediCODE; www.pedicode.org); ethics approval involved a single National Institutes of Health protocol for VUMC, BCH, and UCLA in the United States and a research ethics board for the HSC in Canada. All the families provided written informed consent to participate in the study; at least one parent or guardian provided written informed consent for the participation of their children. The demographic characteristics of the study population are provided in Table S1 in Supplementary Appendix 1, available with the full text of this article at NEJM.org. Individual patients with CODEs or their families were previously sequenced, and raw reads were aligned as described in the Methods section in Supplementary Appendix 1. Principal component analysis of the case series showed a well-distributed ancestral diversity (Fig. S1). Quality-control measures, cohort description, and terms are outlined in Supplementary Appendix 1, including Tables S2 through S6.

GENETIC ANALYSIS

An overview of the genetic analysis of the case series and identification of monogenic genes associated with CODEs is provided in Figure 1A. Briefly, GEMINI (short for "genome mining") was used to identify candidate variants on the basis of known inheritance models of previously identified genes that, when variant, are known to cause CODEs and are listed in the Online Mendelian Inheritance in Man (OMIM) database[14] (Table S7 in Supplementary Appendix 1). We first searched exome data for rare variants (Genome Aggregation Database[15] allele frequency, <0.01) and damaging variants (Combined Annotation Dependent Depletion [CADD][16] score, >20) in genes previously associated with CODEs. We then carried out manual filtering on the basis of confirmatory inheritance pattern, segregation, previous annotation in the ClinVar database, concurrence with clinical features associated with phenotypes of known genetic disease, and overall pathogenicity based on the American College of Medical Genetics and Genomics–Association for Molecular Pathology (ACMG–AMP) classification.[17] An additional manual screening of the remaining cases was carried out to determine ClinVar-annotated pathogenesis and to search for novel genes associated with CODEs. Cases were considered to be solved when an

ACMG–AMP pathogenic or likely pathogenic classification was identified in a gene associated with CODEs in the OMIM that fit the case description and expected inheritance pattern.

FUNCTIONAL ANALYSIS

Functional studies to investigate novel candidate genes included the use of proximity-dependent biotin identification (BioID), cell-based assays, and the creation of novel zebrafish models (see the Methods section in Supplementary Appendix 1). Briefly, BioID was used to identify alterations in interactions between proteins encoded by variants in candidate genes that may be associated with CODEs, as compared with their wild-type counterparts with known binding-partner proteins. High-confidence interactors were defined as those with a Bayesian false discovery rate of 0.01 or less. Studies in zebrafish were carried out at the Zebrafish Genetics and Disease Models Facility at the HSC, with approval from the Animal Care Committee. Candidate genes in zebrafish were knocked out with the use of a clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 system (Figs. S2 and S3 in Supplementary Appendix 1). Student's t-test was used to compare the means of two groups of zebrafish. RNA was extracted and sequenced from pooled wild-type or variant zebrafish, and differential expression analysis was carried out with the use of the DESeq2 package in R software. Several cell lines, including CACO-2_{BBE} (monolayers), HeLa, MDCK-FcRn, and HT-29 cells, were used for functional studies; statistical analyses of these studies are described in Figures 2, 3, and 4 and in Supplementary Appendix 1.

RESULTS

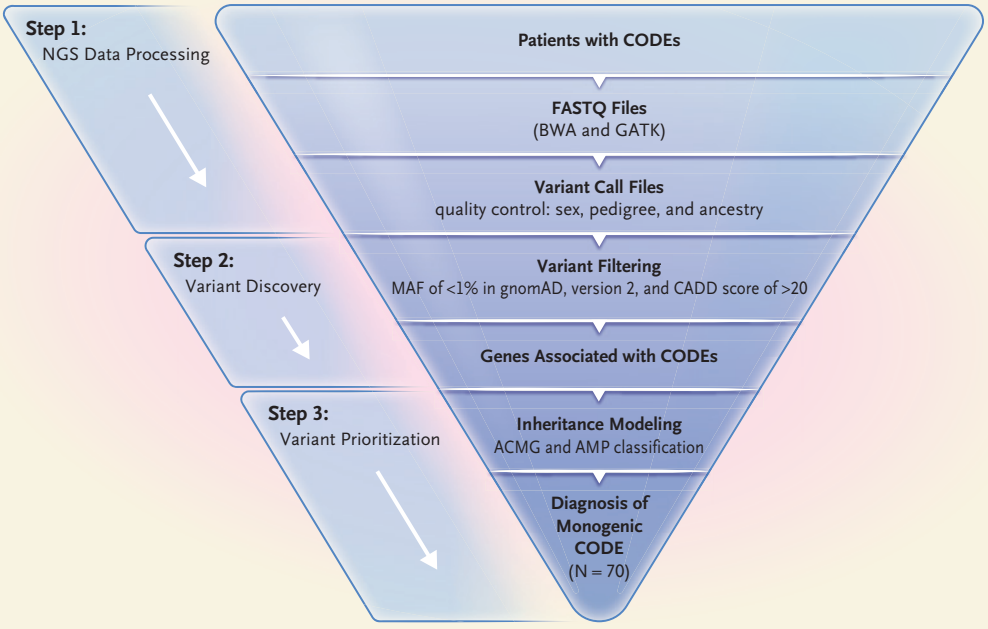
GENETIC ARCHITECTURE OF CODES

We analyzed next-generation sequencing data obtained from PediCODE sites. Overall, 139 infants (including 10 sibling pairs) who presented with CODEs and 182 parents and siblings without CODEs were included in the analysis (Table S1 in Supplementary Appendix 1). Among the 129 families included in the analysis, exome sequencing was performed in at least one family member (other than the proband) in 98 families, to assist with making a genetic diagnosis (Table S2 in Supplementary Appendix 1). A total of 14 probands (1 sibling pair) originated from consanguineous parents, as suggested by a relationship coefficient of 0.1 or more (Table S3 in Supplementary Appendix 1).

In 62 of 129 probands, we identified causal variants in 1 of 24 known monogenic genes associated with CODEs (Fig. 1B and Supplementary Appendix 2). Of these 62 probands, 58 (94%) had autosomal recessive disorders, and 4 (6%) had X-linked disorders. Cases in nearly half the probands were linked to genes involved in epithelial trafficking and polarity, including *EPCAM* (12 cases), *MYO5B* (8 cases), *TTC7A* (4 cases), and *SKIV2L* (3 cases). Other variants in genes associated with CODEs that were detected in this analysis included those in *SLC9A3* (5 cases), *DGAT1* (5 cases), *XIAP* (4 cases), and *NEUROG3* (3 cases). Eight of the 10 sibling pairs who presented with CODEs harbored the same pathogenic variant as the proband (Fig. 1B). Collectively, variants in these 8 genes caused disease in 52 of 70 identified cases (74%) in 62 probands and 8 siblings. Seven infants from 6 unrelated families with tufting enteropathy carried a previously reported founder variant identified in the Arabian Gulf, *EPCAM* Q167Pfs*21.[18] Four infants with enteric anendocrinosis from 3 unrelated Bedouin families (each with an exome sequencing–calculated relationship coefficient of <0.1) had a novel variant in *NEUROG3* (Q137R).

Three probands had variants that were not discoverable through exome sequencing; genome sequencing or Sanger sequencing was required to identify the causal variant in these probands, including Proband 68, who had an *XIAP* deletion. Genome sequencing also identified a *PERCC1* variant in Proband 57, a recently identified gene needed for enteroendocrine cell function that was previously unannotated and, therefore, not captured by exome sequencing.[19] Proband 5 had an intronic splice donor site variant in *SLC9A3* that was missed owing to poor coverage of exon 8 (Fig. S4 in Supplementary Appendix 1). Furthermore, examination of median sequencing depth showed that several genes had poor coverage of exon 1, as has been noted in previous studies. Notably, several exons in *MYO5B*, the second most common cause (when pathogenically variant) of CODEs in this study, had poor sequence coverage, which suggests that deeper sequencing of *MYO5B* should be considered if there is a high suspicion for microvillus inclusion disease and negative results from *MYO5B* exon sequencing.

A Genetic Analysis



B Genes Associated with CODEs

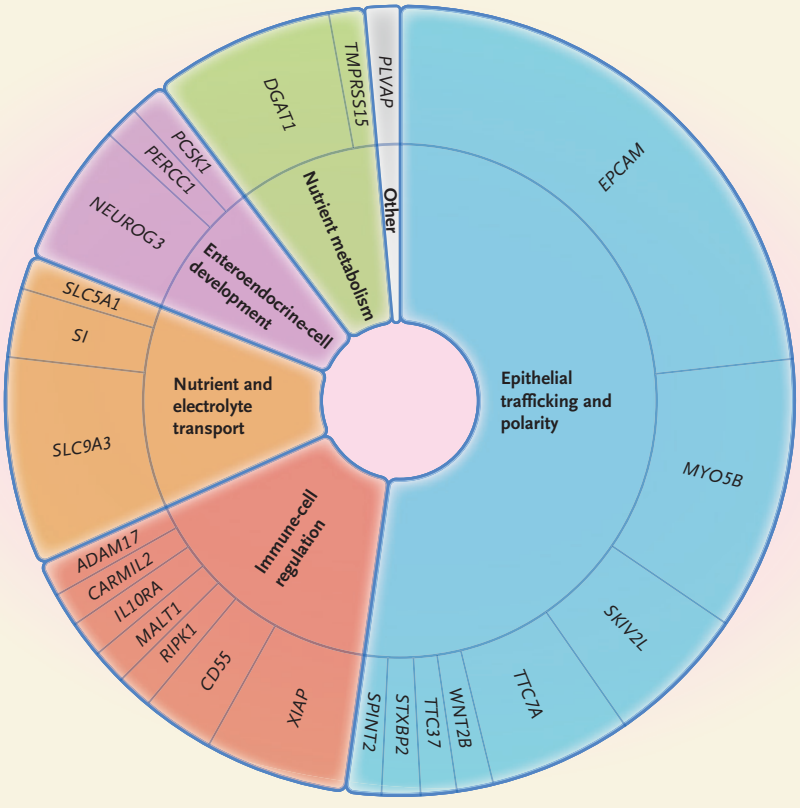


Figure 1 (facing page). The Genetic Architecture of Congenital Diarrhea and Enteropathies (CODEs).
In Panel A, a patient flow diagram shows DNA sequencing and single bioinformatics pipeline and variant annotation with quality-control processes. FASTQ files are a text-based format for storing both a biologic sequence and its corresponding quality scores. The peddy tool was used for quality control by confirming expected family relationships (e.g., parents and siblings), inferring sample sex, and estimating ancestry with the use of genetic markers. ACMG denotes American College of Medical Genetics and Genomics, AMP Association for Molecular Pathology, BWA Burrows–Wheeler Aligner, CADD Combined Annotation Dependent Depletion, GATK Genome Analysis Toolkit, gnomAD Genome Aggregation Database, MAF minor allele frequency, and NGS next-generation sequencing. Panel B shows genes associated with CODEs that were identified in patients included in analyses by the Pediatric Congenital Diarrhea and Enteropathies Consortium (PediCODE).

Further interrogation of the exome data from patients with negative exome sequencing analysis revealed three novel candidate genes that may be associated with CODEs on the basis of genetic heritability, population frequency, CADD score, known gene function, or available animal models. Details are provided in Table S8 in Supplementary Appendix 1.

GENETIC AND FUNCTIONAL SIGNIFICANCE OF *GRWD1* VARIANTS

The first candidate gene, *GRWD1*, was identified in a sibling pair (brother and sister) who presented with diffuse arterial hypoplasia and diarrhea within the first years of life that was treated with parenteral nutrition. Exome sequence analysis revealed rare and damaging *GRWD1* variants: a maternally inherited 19:48451128 A/G variant (p.H307R; CADD score, 25) and a paternally inherited 19:48452786 G/T variant (p.V368F; CADD score, 28).

Glutamate-rich WD40 repeat–containing protein 1 (GRWD1) is a highly conserved member of the WD40 protein family (Fig. 2A). It is needed in late steps of the assembly of large ribosomal subunits[20] and is a regulator of p53.[21] Because knockout of *grwd1* was lethal in zebrafish, we generated first-generation (F0) mosaic *grwd1* mutant zebrafish using CRISPR-Cas9 editing (crispant *grwd1*) that survived until 15 days after fertilization (Fig. S2 in Supplementary Appendix 1). Although crispant *grwd1* larvae at 8 days after fertilization had normal body length, they had significantly reduced gut length (Fig. 2B) with disrupted intestinal architecture, including enlarged, rounded goblet cells, disorganized enterocytes, and irregular gut lumen surface, as compared with the intestines of wild-type control fish (Fig. 2C). The goblet-cell area was greater in crispant *grwd1* zebrafish than in control fish. Differential expression analysis and pathway analysis of whole *grwd1* crispant larvae at 8 days after fertilization revealed perturbation in several ribosomal and p53-regulated genes (Fig. S5).

Dysfunction of ribosome biogenesis triggers nucleolar surveillance,[22] a regulatory loop in which unassembled ribosomal proteins form a complex with the ubiquitin E3 ligase Hdm2, which in turn leads to p53 stabilization and increased expression of the target genes of p53.[23] Differential RNA expression analysis of whole *grwd1* crispant larvae at 8 days after fertilization revealed perturbation in several ribosomal and p53-regulated genes, a finding consistent with the activation of nucleolar surveillance.

To characterize the GRWD1 protein variants associated with CODEs, we conducted BioID (Fig. 2D and Supplementary Appendix 3)[24] of the GRWD1 protein and the missense variants H307R and V368F in human HEK293 Flp-In T-REx cells. BioID of GRWD1 identified a high-confidence interaction with a single ribosomal protein, RPL3, and with multiple ribosome biogenesis factors that are important for nucleolar steps in the assembly of the large subunit.[25] Both H307R and V368F variant proteins showed decreased interactions with RPL3 and ribosome assembly factors and increased interactions with prefoldin and CCT family protein folding chaperones. In budding yeast and *Caenorhabditis elegans*, the GRWD1 orthologues rrb1 (yeast) and GRWD-1 (*C. elegans*) are reported to act as chaperones for rpl3[26,27] and thus are critical to the formation and function of the ribosomal peptidyl transferase center.[28] Decreased RPL3 binding for the GRWD1 variants associated with CODEs was confirmed in an orthogonal affinity purification assay (Fig. 2E). Finally, Flag-tagged versions of the H307R and V368F GRWD1 variants were predominantly localized to the cytoplasm, and the wild-type protein was primarily localized to the nucleus in transfected HeLa cells (Fig. 2F). The GRWD1-variant

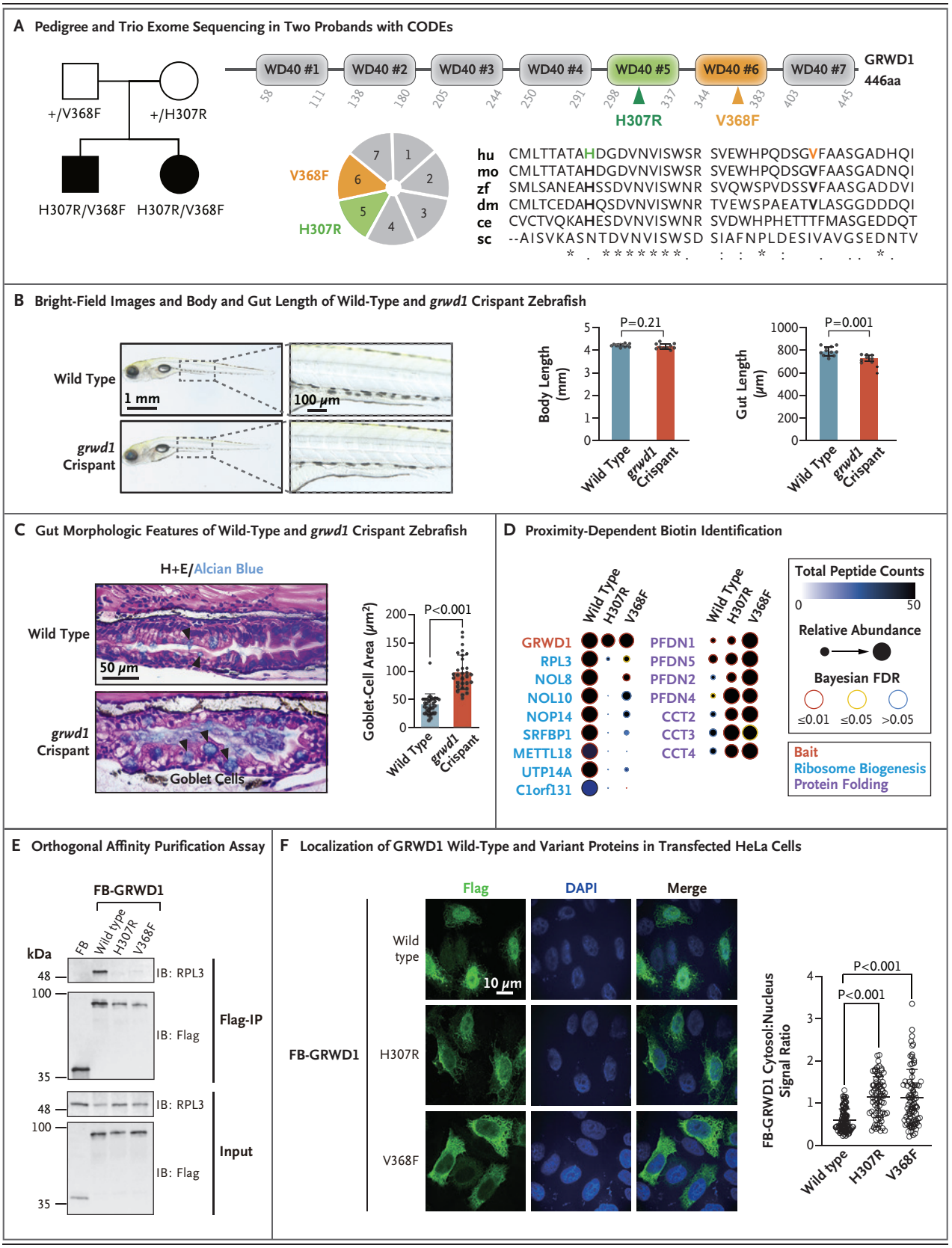

A Pedigree and Trio Exome Sequencing in Two Probands with CODEs
+/V368F
+/H307R
H307R/V368F
H307R/V368F
WD40 #1
WD40 #2
WD40 #3
WD40 #4
WD40 #5
WD40 #6
WD40 #7
GRWD1
446aa
58
111
138
180
205
244
250
291
298
337
344
383
403
445
H307R
V368F
V368F
H307R
hu CMLTTATAHDGDVNVISWSR SVEWHPQDSGVFAASGADHQI
mo CMLTTATAHDGDVNVISWSR SVEWHPQDSGVFAASGADNQI
zf SMLSANEAHSSDVNVISWNR SVQWSPVDSSVFAASGADDVI
dm CMLTCEDAHQSDVNVISWNR TVEWSPAEATVLASGGDDDQI
ce CVCTVQKAHESDVNVISWNR SVDWHPHETTTFMASGEDDQT
sc --AISVKASNTDVNVISWSD SIAFNPLDESIVAVGSEDNTV
B Bright-Field Images and Body and Gut Length of Wild-Type and grwd1 Crispant Zebrafish
Wild Type
grwd1 Crispant
1 mm
100 μm
P=0.21
Body Length (mm)
P=0.001
Gut Length (μm)
Wild Type
grwd1 Crispant
C Gut Morphologic Features of Wild-Type and grwd1 Crispant Zebrafish
H+E/Alcian Blue
Wild Type
grwd1 Crispant
50 μm
Goblet Cells
P<0.001
Goblet-Cell Area (μm²)
D Proximity-Dependent Biotin Identification
Wild Type
H307R
V368F
GRWD1
RPL3
NOL8
NOL10
NOP14
SRFBP1
METTL18
UTP14A
C1orf131
PFDN1
PFDN5
PFDN2
PFDN4
CCT2
CCT3
CCT4
Total Peptide Counts
0
50
Relative Abundance
Bayesian FDR
≤0.01
≤0.05
>0.05
Bait
Ribosome Biogenesis
Protein Folding
E Orthogonal Affinity Purification Assay
FB-GRWD1
kDa
FB
Wild type
H307R
V368F
48
100
35
IB: RPL3
IB: Flag
Flag-IP
Input
F Localization of GRWD1 Wild-Type and Variant Proteins in Transfected HeLa Cells
Flag
DAPI
Merge
Wild type
H307R
V368F
FB-GRWD1
10 μm
FB-GRWD1 Cytosol:Nucleus Signal Ratio
P<0.001
P<0.001

Figure 2 (facing page). Genetic and Functional Studies of *GRWD1* Variants.

Panel A, left, shows the pedigree of two probands with CODEs who had biallelic variants in *GRWD1* that were identified by trio exome sequencing. Squares represent male family members, circles female family members, and solid symbols the probands. Panel A, right, includes a diagram of the seven WD40 repeats of GRWD1 and amino acid conservation across species of the two variants identified in the two probands (human [hu], mouse [mo], zebrafish [zf], *Drosophila melanogaster* [dm], *Caenorhabditis elegans* [ce], and *Saccharomyces cerevisiae* [sc]). Alignments were generated with the use of Clustal Omega; asterisks indicate conserved residues, colons strongly similar residues, and periods weakly similar residues. **Panel B**, left, shows representative bright-field images of wild-type zebrafish (N=12) and *grwd1* crispant zebrafish (N=10) at 8 days after fertilization. Panel B, right, shows bar graphs of body length and gut length of the zebrafish, measured with the use of Fiji software. **Panel C** shows gut morphologic features of wild-type and *grwd1* crispant zebrafish. At left are representative images of wild-type zebafish (N=6) and *grwd1* crispant zebrafish (N=6) at 8 days after fertilization with hematoxylin and eosin (H+E) and Alcian blue staining. At right is a bar graph of goblet-cell area, as measured with the use of the Fiji selection tool to draw the region of interest (30 goblet cells per fish). For the bar graphs in Panels B and C, statistical differences were determined with the use of Student's t-test; dots indicate individual values, the height of the shaded bars indicates mean values, and I bars indicate standard deviations. **Panel D** shows a ProHits-viz dot plot depicting proximity-dependent biotin identification (BioID) data for select proximity interactors of the wild-type and CODE variant GRWD1 proteins. Dot size indicates the relative abundance of each interacting partner detected across each of the three "bait" proteins. Dot shade indicates total peptide counts detected for each interactor, and the shade of the ring surrounding each dot indicates the confidence level (Bayesian false discovery rate [FDR]) of each indicated bait–prey interaction. Interactors were grouped according to functional annotation, as indicated. For statistical analysis, a Bayesian FDR was assigned to identified proteins with the use of SAINT (version 3.6.1; 20 BirA*Flag-only controls compressed to 4). **Panel E** shows the results of an orthogonal affinity purification assay. HEK293 T cells were transfected with either Flag-BirA* alone (FB) or a Flag-BirA*–tagged GRWD1 protein (FB-GRWD1; wild type or CODE variant, as indicated). Flag-tagged proteins were precipitated with Flag agarose affinity gel, and eluates were analyzed by means of Western blotting, with the use of antibodies directed against the Flag epitope or the RPL3 protein. Data are representative of three independent experiments. IB denotes immunoblot, and IP immunoprecipitation. **Panel F** shows the localization of GRWD1 proteins in transfected HeLa cells. HeLa cells expressing FB-GRWD1 wild-type, H307R, or V368F proteins were fixed and stained for Flag epitope (green) and 4′,6-diamidine-2-phenylindole (DAPI) (blue). The GRWD1 cytosol:nucleus signal ratio was calculated with the use of Volocity software. Statistical differences were determined with the use of Student's t-test. The long horizontal lines represent means from three independent experiments, with more than 70 cells analyzed, and the shorter bars represent standard deviations.

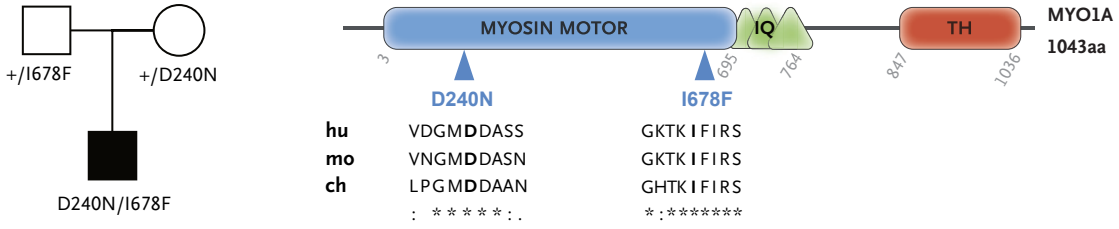
proteins were mislocalized and had defective RPL3 chaperone activity, which resulted in defective ribosomal biogenesis, and our *grwd1* crispant zebrafish model showed abnormalities in intestinal goblet cells. We conclude that *GRWD1* deficiency leads to defective ribosomal biogenesis that may result in systemic disease.

GENETIC AND FUNCTIONAL SIGNIFICANCE OF *MYO1A* VARIANTS

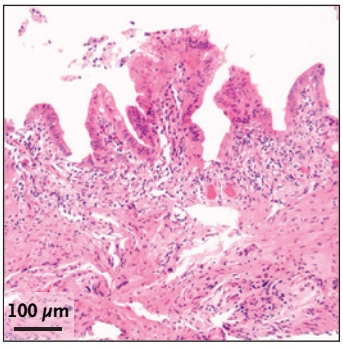
The second candidate gene, *MYO1A*, was identified in a male infant who presented at 2 months of age with an abrupt onset of diarrhea peaking at 150 ml per kilogram of body weight per day and a 15% weight loss. With total parenteral nutrition, the diarrhea continued. At 12 months, his stool remained liquid, averaging 20 to 50 ml per kilogram per day. Rare biallelic variants in *MYO1A* were identified, including a paternally inherited 12:57037571 T/A variant (p.I678F; CADD score, 25) and a maternally inherited 12:57044132 C/T variant (p.D240N; CADD score, 24) (Fig. 3A). Myosin-1A (MYO1A) is an actin-based monomeric motor protein whose expression is limited to the intestinal tract, where it localizes almost exclusively to, and is one of the most abundant proteins in, the brush border. It is critical for normal enterocyte brush-border function.[29,30] *Myo1a* knockout mice show multiple defects in the intestinal brush border, including extensive membrane herniations and fused microvilli.[31] These alterations are similar to those observed in the proband's duodenal biopsy at 4 months of age, which showed closely packed enterocytes, some of which had a rounded teardrop shape (Fig. 3B), and a reduction of MYO1A at the brush border (Fig. 3C). Similarly, when overexpressed in CACO-2_{BBE} cells, the MYO1A I678F and D240N variants both showed aberrant localization as compared with a wild-type construct, which showed strong enrichment in apical microvilli as expected (Fig. 3D, 3E, and 3F).

The proband had severe diarrhea that resolved by his second birthday, an outcome consistent with the phenotype observed in *Myo1a* knockout mice that survive as a result of redistribution to the brush border of other, compensatory class I

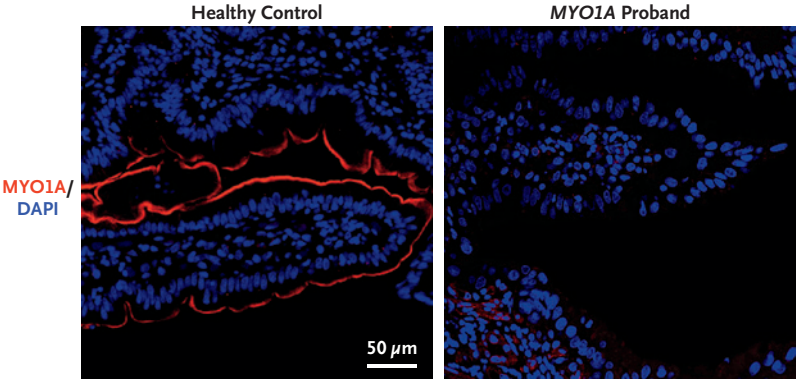
A Pedigree and Trio Exome Sequencing in a Proband with CODE



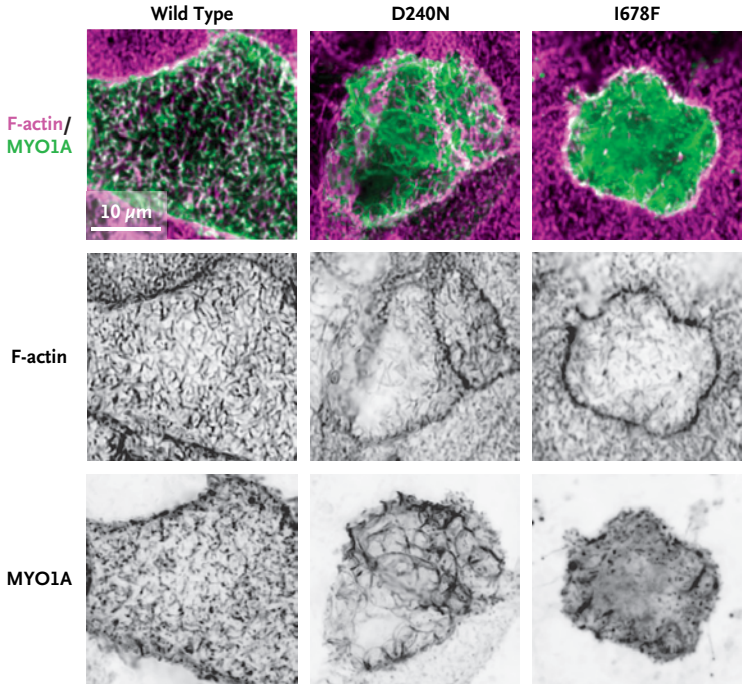
B Hematoxylin and Eosin Staining of Intestinal-Biopsy Specimen



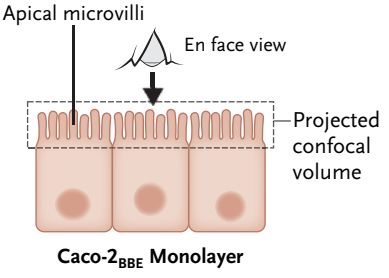
C Immunofluorescence Images



D Maximum Intensity Projections of Confocal Volumes



E Orientation of Image Planes



F Colonization with Apical Microvilli

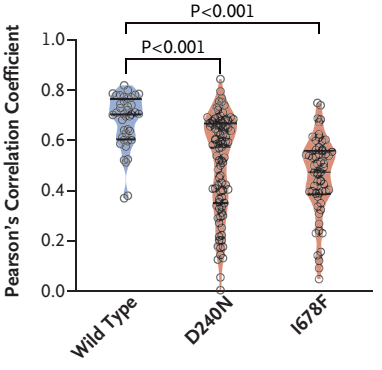


Figure 3 (facing page). Genetic and Functional Studies of *MYO1A* Variants.

Panel A, left, shows the pedigree of a proband with CODE who had biallelic variants in *MYO1A* that were identified by trio exome sequencing. Squares represent male family members, circle the mother, and solid symbol the proband. Panel A, right, includes a schematic illustration of the protein domain architecture, which highlights the amino acid change and its conservation across species (human [hu], mouse [mo], and chicken [ch]). Alignments were generated with the use of Clustal Omega; asterisks indicate conserved residues, colons strongly similar residues, and period weakly similar residues. IQ denotes IQ motif, and TH tail homology. **Panel B** shows hematoxylin and eosin staining of an intestinal-biopsy specimen from the proband. **Panel C** includes immunofluorescence images of intestinal tissue from a healthy control and the proband showing MYO1A (red) and DAPI (blue). **Panel D** shows maximum-intensity projections of confocal volumes indicating localization of wild-type, D240N, and I678F variants of enhanced green fluorescent protein (EGFP)–MYO1A (green) in CACO-2_{BBE} cells, fixed and stained with phalloidin to highlight filamentous actin (F-actin) (magenta). The top row shows two-channel merge images; inverted single-channel images for EGFP and phalloidin are shown beneath each merge. **Panel E** includes a diagram depicting orientation of the image planes shown in Panel D relative to the CACO-2_{BBE} monolayer and the volume sampled during confocal imaging. **Panel F** shows Pearson's correlation coefficients calculated between green (MYO1A construct) and magenta (F-actin) channels on a per-cell basis; this value reflects the extent of colocalization between each expressed protein and the microvillar actin cytoskeleton. Each point represents a measurement from a single cell (40 cells for wild type, 87 for D240N, and 58 for I678F). For each set of data, the middle horizontal line indicates the median and the top and bottom lines the interquartile range. Statistical differences were determined with the use of a Kruskal–Wallis analysis-of-variance test.

myosins, MYO1C and MYO1D.[31] The patient's *MYO1A* variants seemed to result in the mislocalization from the microvilli of MYO1A.

GENETIC AND FUNCTIONAL SIGNIFICANCE OF *MON1A* VARIANT

The third candidate gene, *MON1A*, was identified in a female newborn who presented with diarrhea on day 3 after birth that developed into profound diarrhea with vomiting and failure to thrive before 1 year of age. Exome sequence analysis of the consanguineous family identified a homozygous 3:49911685 G/A *MON1A* variant (p.R249C; CADD score, 27) (Fig. 4A). MON1 homologue A (MON1A) is a binding partner of CCZ1,[32] and together, they make up a guanine nucleotide exchange factor (GEF) complex[33] that supports RAB7A function in late endocytic trafficking.[34]

As expected, BioID of MON1A identified CCZ1B, along with several proteins that had been identified in a previous RAB7A BioID analysis.[35] While maintaining the interaction with CCZ1B, the MON1A R249C variant showed notably reduced interactions with several RAB7A endocytic trafficking regulators (Fig. 4B and Supplementary Appendix 4). In findings consistent with a defect in endosome sorting caused by the disruption of MON1A–CCZ1 complex function,[32] immunohistochemical analysis of a duodenal-biopsy specimen from the proband showed that NHE3 and EZRIN were mislocalized from the apical brush border and showed substantially fewer RAB7-positive vesicles than those observed in biopsy specimens from a healthy control (Fig. 4C). Overexpression of the R249C MON1A variant resulted in decreased size and number of RAB7-positive vesicles, effects that could be reversed by complementation with MON1A (Fig. 4D). Functional studies showed that MON1A is necessary for endolysosome formation and acidification (Fig. 4E); similarly, polarized trafficking was altered by R249C MON1A with reduced bidirectional epithelial transcytosis of IgG in MDCK-FcRn cells (Fig. 4F). Finally, *mon1a* knockout (–/–) zebrafish (Fig. S3 in Supplementary Appendix 1) did not show any obvious developmental defects but had multiple intestinal abnormalities, including morphologically atypical goblet cells with higher mucin secretion and lower activity of lysosome-rich enterocytes than wild-type zebrafish (Fig. 4G and Fig. S6). The R249C MON1A variant thus appears to be defective in its roles in endosomal sorting and RAB7-dependent endosome maturation in the gut, leading to diarrhea.

DISCUSSION

We have characterized the broad genetic architecture of CODE disorders, highlighting the genes most commonly responsible and establishing the

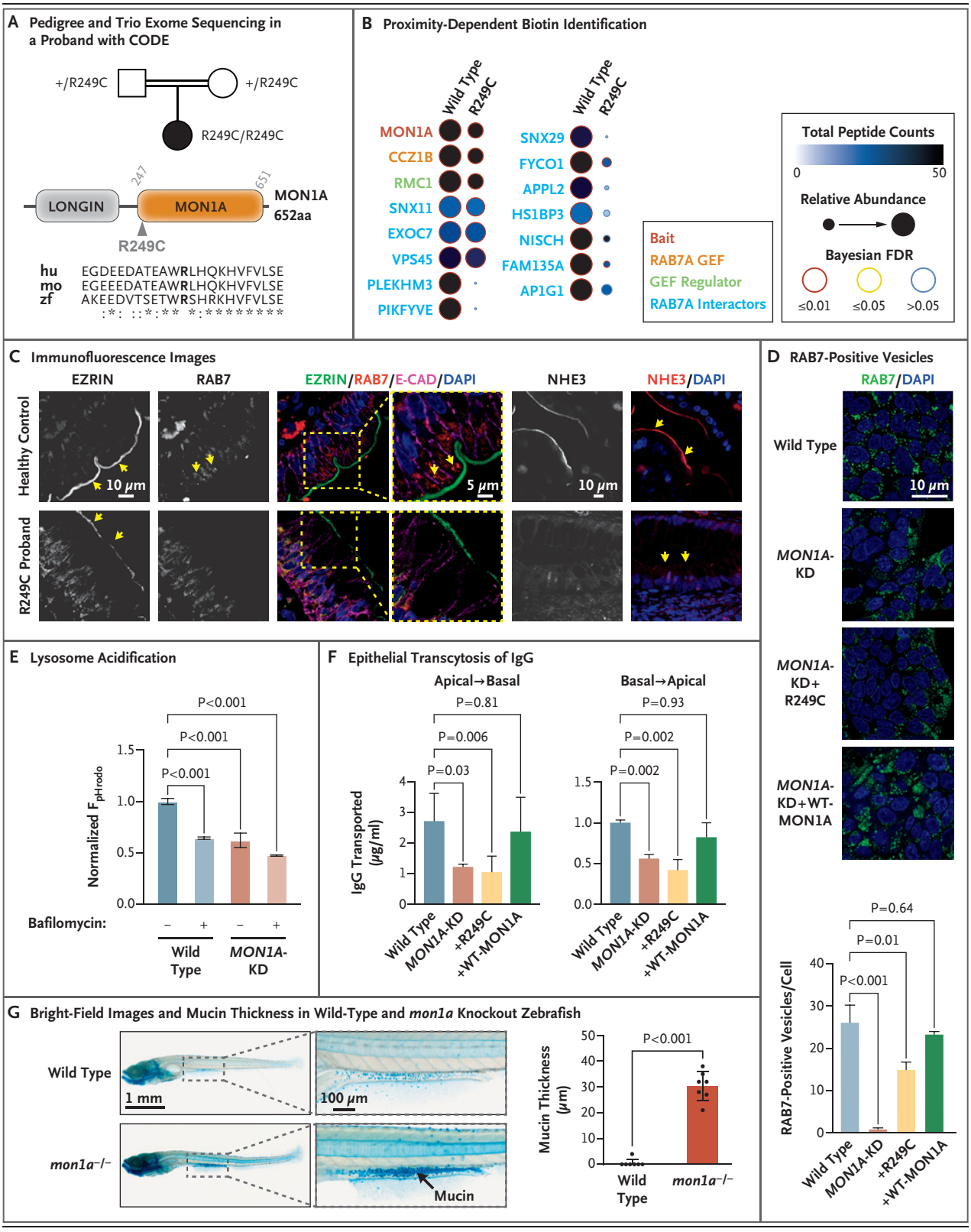

A Pedigree and Trio Exome Sequencing in a Proband with CODE
+/R249C
+/R249C
R249C/R249C
247
651
LONGIN
MON1A
MON1A 652aa
R249C
hu EGDEEDATEAWRLHQKHVFVLSE
mo EGEEEDATEAWRLHQKHVFVLSE
zf AKEEDVTSETWRSHRKHVFVLSE
B Proximity-Dependent Biotin Identification
Wild Type
R249C
MON1A
CCZ1B
RMC1
SNX11
EXOC7
VPS45
PLEKHM3
PIKFYVE
SNX29
FYCO1
APPL2
HS1BP3
NISCH
FAM135A
AP1G1
Bait
RAB7A GEF
GEF Regulator
RAB7A Interactors
Total Peptide Counts
0
50
Relative Abundance
Bayesian FDR
≤0.01
≤0.05
>0.05
C Immunofluorescence Images
EZRIN
RAB7
EZRIN/RAB7/E-CAD/DAPI
NHE3
NHE3/DAPI
Healthy Control
R249C Proband
10 μm
5 μm
D RAB7-Positive Vesicles
RAB7/DAPI
Wild Type
MON1A-KD
MON1A-KD+R249C
MON1A-KD+WT-MON1A
E Lysosome Acidification
Normalized F pHrodo
P<0.001
P<0.001
P<0.001
Bafilomycin:
Wild Type
MON1A-KD
F Epithelial Transcytosis of IgG
Apical→Basal
Basal→Apical
IgG Transported (μg/ml)
P=0.81
P=0.006
P=0.03
P=0.93
P=0.002
P=0.002
Wild Type
MON1A-KD
+R249C
+WT-MON1A
G Bright-Field Images and Mucin Thickness in Wild-Type and mon1a Knockout Zebrafish
Wild Type
mon1a−/−
1 mm
100 μm
Mucin
Mucin Thickness (μm)
P<0.001
Wild Type
mon1a−/−
RAB7-Positive Vesicles/Cell
P=0.64
P=0.01
P<0.001
Wild Type
MON1A-KD
+R249C
+WT-MON1A

Figure 4 (facing page). Genetic and Functional Studies of *MON1A* Variants.

Panel A, left, shows the pedigree of a proband with CODE who had biallelic variants in *MON1A* that were identified by trio exome sequencing. The square represents the father, circles female family members, double bars consanguinity, and solid symbol the proband. Panel A, right, includes a schematic illustration of the protein domain architecture, highlighting the amino acid change and its conservation across species (human [hu], mouse [mo], and zebrafish [zf]). Alignments were generated with the use of Clustal Omega; asterisks indicates conserved residues, and colons strongly similar residues. **Panel B** shows a ProHits-viz dot plot depicting BioID data for select proximity interactors of the wild-type and CODE variant MON1A proteins, as described in Figure 2. **Panel C**, left and middle, includes immunofluorescence images of intestinal tissue from a healthy control and the proband showing EZRIN (grayscale and green), RAB7 (grayscale and red), E-cadherin (E-CAD) (magenta), and DAPI (blue). The arrows show reduced EZRIN at the apical brush border and loss of RAB7-positive vesicles in the proband. Panel C, right, includes immunofluorescence images of tissue from a healthy control and the proband showing NHE3 (grayscale and red). The arrows show that NHE3 is localized on the apical brush border in the healthy control and loss of brush-border localization in the proband. **Panel D**, top, shows representative images of HT-29 cells with wild-type *MON1A* (wild type), knockdown of *MON1A* (*MON1A*-KD), *MON1A*-KD with expression of patient-variant R249C MON1A protein (*MON1A*-KD+R249C), or *MON1A*-KD with expression of wild-type MON1A protein (*MON1A*-KD+WT-MON1A). Panel D, bottom, shows summary graphs of the size and count of RAB7-positive vesicles per cell. **Panel E** shows lysosome acidification in HT-29 cells with the use of pHrodo-EGF staining of wild-type *MON1A* or *MON1A*-KD with or without bafilomycin. Data are presented as the proportion relative to wild-type *MON1A* without bafilomycin. **Panel F** shows FcRn-dependent IgG concentrations crossing the epithelium either from the apical surface to the basolateral surface (left) or from the basolateral surface to the apical surface (right) in MDCK-FcRn cells with wild-type *MON1A*, knockdown of *MON1A*, or knockdown of *MON1A* with expression of R249C MON1A or wild-type MON1A protein. For the bar graphs in Panels D, E, and F, statistical differences were determined with the use of ordinary one-way analysis of variance with multiple comparisons; the height of the shaded bars indicates the mean of three experiments (Panels D and E) or four experiments (Panel F), and error bars indicate standard errors. **Panel G**, left, shows representative bright-field images of wild-type zebrafish (N=60) and *mon1a* knockout zebrafish (N=56) at 9 days after fertilization. A total of 12.5% of *mon1a*$^{-/-}$ mutants showed high mucin secretion on Alcian blue staining. Blue dots and areas indicate goblet cells and mucins in the gut. Panel G, right, shows a bar graph of mucin thickness, measured in 60 of 60 wild-type zebrafish and 7 of 56 *mon1a*$^{-/-}$ zebrafish. The latter group was characterized by a phenotype of increased mucin thickness. The statistical difference was determined with the use of Student's t-test. Dots indicate individual values, the height of the shaded bars indicates mean values, and I bars indicate standard deviations.

landscape of variants and heritability, including a potentially new Bedouin founder *NEUROG3* variant. The diagnostic yield of approximately 50% was higher than the yield of 4% observed in monogenic inflammatory bowel disease[11] and other monogenic disorders[36] and similar to the high yield observed in some studies of neuromuscular disorders.[37]

We identified causal variants in half the cases of CODEs that were diagnosed in tertiary hospitals. It is possible that some variants in genes associated with CODEs escaped identification by exome sequence analysis. This lack of identification seems particularly likely in infants with autosomal recessive disease, in whom only one variant was identified. The "missing variant" might have been detectable through genome sequencing or through obtaining better sequence coverage when exome sequencing is performed.

It is likely that many persons have rare or private genetic variants in novel genes. We also uncovered three novel candidate genes that may be associated with CODEs and provided functional validation showing that the variants alter protein function in relevant intestinal models that meet the criteria for reporting.[38] We used BioID to further support pathogenic classification of these novel genes associated with CODEs by identifying dysregulated protein interactions in relevant intestinal pathways that are disrupted with candidate variants. However, additional cases are required to further understand the pathogenesis of CODEs and define genotype–phenotype correlations. Overall, we demonstrate the usefulness of next-generation sequencing as a powerful tool to identify known and novel pathogenic variants that cause congenital diarrhea.

Supported by the National Institutes of Health High Impact, Interdisciplinary Science in NIDDK Research Areas (grants RC2DK118640 and RC2DK122532), the Leona M. and Harry B. Helmsley Charitable Trust, the Canadian Institutes of Health Research (Foundation Grant FDN#408445), Canada Research Chairs, the Lassonde Family Precision IBD Initiative, the California

Center for Rare Diseases at the UCLA Institute for Precision Health, and the UCLA Clinical Genomics Center.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank all the patients and their families for participating in this study and the coordinators and clinicians caring for these patients; the staff of the Centre for Applied Genomics facilities at the Hospital for Sick Children (HSC), Toronto; Xiucheng Cui and Jason Burgess in the Zebrafish Genetics and Disease Models Facility, HSC, for the generation of zebrafish mutant lines and genotyping; and the staff of the High-Performance Computing Facility, Centre for Computational Medicine, HSC, for data processing and data analysis This research was enabled in part by digital infrastructure provided by Compute Ontario (www.computeontario.ca) and the Digital Research Alliance of Canada (www.alliancecan.ca).

AUTHOR INFORMATION

[1] Division of Gastroenterology, Hepatology, and Nutrition, Hospital for Sick Children, Toronto; [2] Cell Biology Program, Research Institute, Hospital for Sick Children, Toronto; [3] Division of Gastroenterology, Hepatology, and Nutrition, Boston Children's Hospital and Harvard Medical School, Boston; [4] Department of Human Genetics, University of California, Los Angeles, Los Angeles; [5] RNA Molecular Biology, Fonds de la Recherche Scientifique, Université Libre de Bruxelles, BioPark Campus, Gosselies, Belgium; [6] Princess Margaret Cancer Centre, University Health Network, and Department of Medical Biophysics, University of Toronto, Toronto; [7] Department of Cell and Developmental Biology, Epithelial Biology Center, Vanderbilt University School of Medicine, Nashville; [8] Division of Pediatrics, Pediatric Gastroenterology Unit, Soroka University Medical Center and Faculty of Health Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel; [9] Department of Pediatrics, Cedars–Sinai Medical Center, Los Angeles; [10] Institute of Gastroenterology, Nutrition, and Liver Diseases, Schneider Children's Medical Center of Israel, Petach Tikva, Israel; [11] Juliet Keidan Institute of Pediatric Gastroenterology and Nutrition, Shaare Zedek Medical Center, Hebrew University of Jerusalem, Jerusalem; [12] Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Vanderbilt University Medical Center, Nashville; [13] Department of Paediatric Gastroenterology and Hepatology, Starship Children's Health, Te Toka Tumai Auckland, Auckland, New Zealand; [14] Center for Computational Medicine, Research Institute, Hospital for Sick Children, Toronto; [15] Department of Surgery, Epithelial Biology Center, Vanderbilt University School of Medicine, Nashville; [16] Department of Pediatrics, Division of Gastroenterology and Nutrition, Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Mattel Children's Hospital, and the David Geffen School of Medicine, University of California, Los Angeles, Los Angeles; [17] Department of Paediatrics and Biochemistry, Institute of Medical Science, University of Toronto, Toronto.

REFERENCES

1. Thiagarajah JR, Kamin DS, Acra S, et al. Advances in evaluation of chronic diarrhea in infants. Gastroenterology 2018;154:2045-2059.e6.
2. Kaji I, Thiagarajah JR, Goldenring JR. Modeling the cell biology of monogenetic intestinal epithelial disorders. J Cell Biol 2024;223(7):e202310118.
3. Avitzur Y, Jimenez L, Martincevic I, et al. Diet management in congenital diarrheas and enteropathies — general concepts and disease-specific approach, a narrative review. Am J Clin Nutr 2024;120:17-33.
4. Müller T, Hess MW, Schiefermeier N, et al. MYO5B mutations cause microvillus inclusion disease and disrupt epithelial cell polarity. Nat Genet 2008;40:1163-5.
5. Sivagnanam M, Mueller JL, Lee H, et al. Identification of EpCAM as the gene for congenital tufting enteropathy. Gastroenterology 2008;135:429-37.
6. Wang J, Cortina G, Wu SV, et al. Mutant neurogenin-3 in congenital malabsorptive diarrhea. N Engl J Med 2006;355:270-80.
7. Haas JT, Winter HS, Lim E, et al. DGAT1 mutation is linked to a congenital diarrheal disorder. J Clin Invest 2012;122:4680-4.
8. Janecke AR, Heinz-Erian P, Yin J, et al. Reduced sodium/proton exchanger NHE3 activity causes congenital sodium diarrhea. Hum Mol Genet 2015;24:6614-23.
9. Jardine S, Anderson S, Babcock S, et al. Drug screen identifies leflunomide for treatment of inflammatory bowel disease caused by TTC7A deficiency. Gastroenterology 2020;158:1000-15.
10. Avitzur Y, Guo C, Mastropaolo LA, et al. Mutations in tetratricopeptide repeat domain 7A result in a severe form of very early onset inflammatory bowel disease. Gastroenterology 2014;146:1028-39.
11. Crowley E, Warner N, Pan J, et al. Prevalence and clinical features of inflammatory bowel diseases associated with monogenic variants, identified by whole-exome sequencing in 1000 children at a single center. Gastroenterology 2020;158:2208-20.
12. Cakir M, Sag E, Guven B, et al. Early onset congenital diarrheas; single center experience. Pediatr Neonatol 2021;62:612-9.
13. Sharma SS, Sankaranarayanan S, Kumar VH, Kumar NC, Sundaram CS. Congenital diarrheal disorders in neonates: a single-center experience. Indian Pediatr 2021;58:1096-7.
14. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet 2007;80:588-604.
15. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-43.
16. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47:D1:D886-D894.
17. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405-24.
18. Salomon J, Espinosa-Parrilla Y, Goulet O, et al. A founder effect at the EPCAM locus in congenital tufting Enteropathy in the Arabic Gulf. Eur J Med Genet 2011;54:319-22.
19. Oz-Levi D, Olender T, Bar-Joseph I, et al. Noncoding deletions reveal a gene that is critical for intestinal function. Nature 2019;571:107-11.
20. Gratenstein K, Heggestad AD, Fortun J, Notterpek L, Pestov DG, Fletcher BS. The WD-repeat protein GRWD1: potential roles in myeloid differentiation and ribosome biogenesis. Genomics 2005;85:762-73.
21. Kayama K, Watanabe S, Takafuji T, et al. GRWD1 negatively regulates p53 via the RPL11-MDM2 pathway and promotes tumorigenesis. EMBO Rep 2017;18:123-37.
22. Nicolas E, Parisot P, Pinto-Monteiro C, de Walque R, De Vleeschouwer C, Lafontaine DL. Involvement of human ribosomal proteins in nucleolar structure and p53-dependent nucleolar stress. Nat Commun 2016;7:11390.
23. Sloan KE, Bohnsack MT, Watkins NJ. The 5S RNP couples p53 homeostasis to ribosome biogenesis and nucleolar stress. Cell Rep 2013;5:237-47.
24. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol 2012;196:801-10.

25. Tafforeau L, Zorbas C, Langhendries JL, et al. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. Mol Cell 2013;51:539-51.
26. Iouk TL, Aitchison JD, Maguire S, Wozniak RW. Rrb1p, a yeast nuclear WD-repeat protein involved in the regulation of ribosome biosynthesis. Mol Cell Biol 2001;21:1260-71.
27. Pillet B, Méndez-Godoy A, Murat G, et al. Dedicated chaperones coordinate cotranslational regulation of ribosomal protein production with ribosome assembly to preserve proteostasis. Elife 2022;11:e74255.
28. Meskauskas A, Dinman JD. Ribosomal protein L3: gatekeeper to the A site. Mol Cell 2007;25:877-88.
29. Mazerik JN, Tyska MJ. Myosin-1A targets to microvilli using multiple membrane binding motifs in the tail homology 1 (TH1) domain. J Biol Chem 2012;287:13104-15.
30. Tyska MJ, Mooseker MS. MYO1A (brush border myosin I) dynamics in the brush border of LLC-PK1-CL4 cells. Biophys J 2002;82:1869-83.
31. Tyska MJ, Mackey AT, Huang J-D, Copeland NG, Jenkins NA, Mooseker MS. Myosin-1a is critical for normal brush border structure and composition. Mol Biol Cell 2005;16:2443-57.
32. Nelms B, Dalomba NF, Lencer W. A targeted RNAi screen identifies factors affecting diverse stages of receptor-mediated transcytosis. J Cell Biol 2017;216:511-25.
33. Nordmann M, Cabrera M, Perz A, et al. The Mon1-Ccz1 complex is the GEF of the late endosomal Rab7 homolog Ypt7. Curr Biol 2010;20:1654-9.
34. Gao J, Langemeyer L, Kümmel D, Reggiori F, Ungermann C. Molecular mechanism to target the endosomal Mon1-Ccz1 GEF complex to the pre-autophagosomal structure. Elife 2018;7:e31145.
35. Yan B-R, Li T, Coyaud E, et al. C5orf51 is a component of the MON1-CCZ1 complex and controls RAB7A localization and stability during mitophagy. Autophagy 2022;18:829-40.
36. Tammimies K, Marshall CR, Walker S, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. JAMA 2015;314:895-903.
37. Ghaoui R, Cooper ST, Lek M, et al. Use of whole-exome sequencing for diagnosis of limb-girdle muscular dystrophy: outcomes and lessons learned. JAMA Neurol 2015;72:1424-32.
38. Casanova J-L, Conley ME, Seligman SJ, Abel L, Notarangelo LD. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. J Exp Med 2014;211:2137-49.