

## ORIGINAL ARTICLE

# Molecular Genetic Anatomy and Risk Profile of Hirschsprung's Disease

Joseph M. Tilghman, B.S., Albee Y. Ling, B.S., Tychele N. Turner, Ph.D.,  
 Maria X. Sosa, M.S., Niklas Krumm, M.D., Ph.D., Sumantra Chatterjee, Ph.D.,  
 Ashish Kapoor, Ph.D., Bradley P. Coe, Ph.D., Khanh-Dung H. Nguyen, Ph.D.,  
 Namrata Gupta, Ph.D., Stacey Gabriel, Ph.D., Evan E. Eichler, Ph.D.,  
 Courtney Berrios, M.S., and Aravinda Chakravarti, Ph.D.

## ABSTRACT

**BACKGROUND**

Hirschsprung's disease, or congenital aganglionosis, is a developmental disorder of the enteric nervous system and is the most common cause of intestinal obstruction in neonates and infants. The disease has more than 80% heritability, including significant associations with rare and common sequence variants in genes related to the enteric nervous system, as well as with monogenic and chromosomal syndromes.

**METHODS**

We genotyped and exome-sequenced samples from 190 patients with Hirschsprung's disease to quantify the genetic burden in patients with this condition. DNA sequence variants, large copy-number variants, and karyotype variants in probands were considered to be pathogenic when they were significantly associated with Hirschsprung's disease or another neurodevelopmental disorder. Novel genes were confirmed by functional studies in the mouse and human embryonic gut and in zebrafish embryos.

**RESULTS**

The presence of five or more variants in four noncoding elements defined a widespread risk of Hirschsprung's disease (48.4% of patients and 17.1% of controls; odds ratio, 4.54; 95% confidence interval [CI], 3.19 to 6.46). Rare coding variants in 24 genes that play roles in enteric neural-crest cell fate, 7 of which were novel, were also common (34.7% of patients and 5.0% of controls) and conferred a much greater risk than noncoding variants (odds ratio, 10.02; 95% CI, 6.45 to 15.58). Large copy-number variants, which were present in fewer patients (11.4%, as compared with 0.2% of controls), conferred the highest risk (odds ratio, 63.07; 95% CI, 36.75 to 108.25). At least one identifiable genetic risk factor was found in 72.1% of the patients, and at least 48.4% of patients had a structural or regulatory deficiency in the gene encoding receptor tyrosine kinase (*RET*). For individual patients, the estimated risk of Hirschsprung's disease ranged from 5.33 cases per 100,000 live births (approximately 1 per 18,800) to 8.38 per 1000 live births (approximately 1 per 120).

**CONCLUSIONS**

Among the patients in our study, Hirschsprung's disease arose from common noncoding variants, rare coding variants, and copy-number variants affecting genes involved in enteric neural-crest cell fate that exacerbate the widespread genetic susceptibility associated with *RET*. For individual patients, the genotype-specific odds ratios varied by a factor of approximately 67, which provides a basis for risk stratification and genetic counseling. (Funded by the National Institutes of Health.)

From the Center for Complex Disease Genomics, Johns Hopkins University School of Medicine, Baltimore (J.M.T., A.Y.L., T.N.T., M.X.S., S.C., A.K., K.-D.H.N., C.B., A.C.); the Department of Genome Sciences, University of Washington School of Medicine (T.N.T., N.K., A.K., B.P.C., E.E.E.), and the Howard Hughes Medical Institute, University of Washington (E.E.E.) — both in Seattle; and Broad Institute of Harvard and MIT, Cambridge, MA (K.-D.H.N., N.G., S.G.). Address reprint requests to Dr. Chakravarti at the Center for Human Genetics and Genomics, New York University School of Medicine, 435 E. 30th St., Rm. 802/3, New York, NY 10016, or at aravinda.chakravarti@nyulangone.org.

N Engl J Med 2019;380:1421-32.

DOI: 10.1056/NEJMoa1706594

Copyright © 2019 Massachusetts Medical Society.

**H**IRSCHSPRUNG'S DISEASE IS CHARACTERIZED by the lack of ganglia in the myenteric and submucosal plexuses of the gut. It is a "model" complex disorder because it exemplifies multifactorial inheritance and yet has been molecularly tractable.<sup>1,3</sup> The disease (with an incidence of 15 cases per 100,000 live births) is characterized by high heritability (>80%) and marked sex differences (male:female ratio, 4:1).<sup>2</sup> Patients have aganglioneurosis affecting bowel segments of variable length, as a result of incomplete rostral-to-caudal enteric neuronal colonization; on the basis of these segment lengths, the condition is classified as short, long, or total colonic aganglioneurosis (see the Supplementary Appendix, available with the full text of this article at NEJM.org). Approximately 18% of patients have multiple anomalies, some with specific syndromes; approximately 12% have major chromosomal variants.<sup>2,3</sup> Features of Hirschsprung's disease include its high (3 to 17%) sibling recurrence risk (i.e., the risk of being born with the disease, given that one full sibling is affected) and the variation in risk according to sex, segment length, and familiarity.<sup>2</sup>

Hirschsprung's disease has multifactorial causes, although no environmental causes are known.<sup>4,5</sup> Complex segregation analyses have refined this view by showing genetic heterogeneity according to the extent of aganglioneurosis. The long form is characterized by autosomal dominant inheritance and the short form by recessive or multifactorial inheritance, and the variants associated with both forms have incomplete penetrance.<sup>6</sup> This finding led to the discovery of 17 genes with approximately 500 rare disease-associated coding variants, chiefly the genes encoding the receptor tyrosine kinase *RET* and the G-protein-coupled receptor *EDNRB* (Table S1 in the Supplementary Appendix).<sup>3,7-13</sup> Four noncoding variants, individually conferring moderate risks (odds ratio, 1.6 to 3.9) but together conferring risk that can vary by as much as a factor of 30 with increasing risk-allele dosage,<sup>13</sup> are genetic modifiers of Hirschsprung's disease.<sup>11,14</sup> These data suggest widespread and variable genetic susceptibility to the disease from multiple genes, reflected in the differing presentations and recurrence risks among relatives.

We suspected that, in contrast to the genetic risk factors for other complex diseases, many genetic risk factors make individually large contributions to the risk of Hirschsprung's disease.

We undertook genotyping, exome-sequencing, and functional assays to study pathogenic alleles in a set of patients with Hirschsprung's disease with representative phenotypes. Beyond studying known genes and identifying new ones, we investigated the variation in risk according to the type of pathogenic allele, the contribution of each type of allele to Hirschsprung's disease in the general population, and the distribution of these types of alleles across phenotypes. Our primary goal was to enable genetic stratification of patients in order to determine how genetic susceptibility manifests in clinical disease and its penetrance. Such genetic stratification could be used to determine whether postsurgical outcome — for example, continued bowel dysfunction and enterocolitis, which is reported in 30 to 50% of patients<sup>15,16</sup> — is related to genotype.

## METHODS

### PARTICIPANTS AND GENOMEWIDE ANALYSES

We conducted exome sequencing of samples from 190 patients of European ancestry and 47 of their affected relatives (7 parents, 12 children, 17 siblings, and 11 second-degree relatives) with diverse phenotypes. The control sample used in exome sequencing consisted of publicly available, ancestry-matched exome data on 740 samples from the 1000 Genomes Project and the National Institute of Mental Health Repository. For the analysis of common noncoding variant studies, we used a different set of 627 control samples that were genotyped in our laboratory: 404 from the 1000 Genomes Project and an additional 223 "pseudo-controls" (generated from the chromosomes not transmitted to the affected child in 254 parent-child trios<sup>13</sup>). For the analysis of copy-number variants, we used a third control set of 19,584 adults of European ancestry.<sup>17</sup>

Sequence variants, genotypes and their frequencies at single-nucleotide variants, small (<50 bp) insertions or deletions, and copy-number variants were identified and annotated.<sup>17-21</sup> Single-nucleotide polymorphism (SNP) arrays were used to validate copy-number variant calls. Genotype calls at transcription-enhancer variants were from our previous studies involving the same study population.<sup>13,22</sup> The sample ascertainment and analysis methods are described in detail in the Methods section, Tables S1 through S5, and Figures S1 through S3 in the Supplementary Appendix.

**PATHOGENIC ALLELES, GENES, AND LOCI**

For assessing the effect of common noncoding variants, we used four disease-associated SNPs — rs2435357, rs7069590, and rs2506030 in *RET* and rs11766001 in the *SEMA3* gene cluster.<sup>13,22</sup> We have previously shown that the *RET* noncoding variants are located within transcription enhancers bound by the transcription factors RARB, GATA2, and SOX10; these variants lead to reduced *RET* expression and an elevated risk of Hirschsprung's disease.<sup>22</sup> Although the causality of the rs11766001 polymorphism in the *SEMA3* locus is unproven, considerable data support causality or a strong association with a causal variant in *SEMA3C* or *SEMA3D*, which have been shown to be necessary for gut innervation.<sup>12,13</sup> Coding pathogenic alleles at each gene were defined as nonsense or missense changes in codons encoding amino acids that are conserved (with respect to their position in the oligopeptide) across species, splice-site single-nucleotide variants, and all coding insertion–deletion variants with a frequency of 5% or less. These definitions gave acceptable levels of true and false positives at known Hirschsprung's disease genes (Table S1 in the Supplementary Appendix).<sup>23</sup> Disease-associated coding variants can have incomplete penetrance and be present in controls; therefore, we identified Hirschsprung's disease–associated genes as those that had a greater number of unique pathogenic alleles in patients than in controls (Fig. S4 in the Supplementary Appendix). We assessed large copy-number variants (deletions of more than 500 kb and duplications of more than 1 Mb) with a frequency of less than 1% among controls to determine whether they were significantly enriched among patients or had previously been found to be associated with a developmental disorder (Tables S8 and S9 and Fig. S5 in the Supplementary Appendix).<sup>17,24</sup> Additional details are provided in the Methods section in the Supplementary Appendix.

To assess the role of a gene in Hirschsprung's disease, we first used reverse-transcriptase polymerase chain reaction (RT-PCR) to assess its RNA expression in the human embryonic gut at Carnegie stage 22, by which time gut neurogenesis is complete (Fig. S6 in the Supplementary Appendix). Second, we tested gene expression by RNA sequencing and RT-PCR in the developing mouse gut during the equivalent developmental period (embryonic day 10.5) (Fig. S6 in the Sup-

plementary Appendix). Third, we used morpholinos (antisense oligonucleotides) to knock down gene expression in zebrafish embryos at 6 days after fertilization and enumerated the enteric neurons colonizing the gut relative to controls (Fig. S7 in the Supplementary Appendix).<sup>12</sup>

The sequence data generated as part of this study are available in the National Center for Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP) under accession number phs000497. The RNA sequencing data used in this study are accessible under NCBI Gene Expression Omnibus (GEO) accession number GSE99232.

**STATISTICAL ANALYSIS**

Population-level risks were estimated for groups of pathogenic alleles, genes, or loci with the use of odds ratios with significance thresholds (corrected for multiple testing) and 95% confidence intervals.<sup>13,25</sup> The odds ratios were converted to estimated population penetrance (equivalent to the population incidence or risk) with Bayes' theorem, under the assumption of an incidence of 15 cases per 100,000 European-ancestry live births.<sup>11</sup> Allele frequencies among controls were obtained from a variety of public resources<sup>17,21,26</sup> to estimate the population attributable risk. Additional details are provided in the Methods section in the Supplementary Appendix.

**RESULTS****COMMON REGULATORY VARIANTS AND RISK**

Four common transcription-enhancer variants were associated with a moderate risk of Hirschsprung's disease in our sample of 190 patients and 627 controls (Table S2 in the Supplementary Appendix).<sup>9-13</sup> The frequency of these variants allowed us to estimate their total effect according to dosage in reference to persons with one allele (none had zero alleles): a risk of Hirschsprung's disease (odds ratio >1) is evident only with three or more alleles (Table 1), but, in view of multiple comparisons, the risk was considered significant only when at least five risk alleles were present (odds ratio, 4.54; 95% confidence interval [CI], 3.19 to 6.46;  $P=1.22\times 10^{-16}$ ) (Table 2). Thus, the population risk of Hirschsprung's disease varies by a factor of 24, from approximately 1 case per 19,100 live births (0 or 1 risk allele) to 1 case per 710 live births (seven or eight risk alleles) ac-

**Table 1.** Population Risk of Hirschsprung's Disease as a Function of *RET* and *SEMA3* Noncoding Risk-Allele Dosage.\*

No. of Risk Alleles	Patients (N = 186) <i>number (percent)</i>	Controls (N = 627) <i>number (percent)</i>	Odds Ratio (95% CI)	P Value
1	9 (4.8)	87 (13.9)	1.00 (Reference)	—
2	13 (7.0)	137 (21.9)	0.90 (0.38–2.13)	$8.24 \times 10^{-1}$
3	40 (21.5)	172 (27.4)	2.16 (1.03–4.52)	$4.28 \times 10^{-2}$
4	34 (18.3)	124 (19.8)	2.55 (1.20–5.42)	$1.51 \times 10^{-2}$
5	35 (18.8)	84 (13.4)	3.87 (1.81–8.29)†	$3.01 \times 10^{-4}$
6	41 (22.0)	18 (2.9)	20.66 (8.84–48.31)†	$6.03 \times 10^{-15}$
7 or 8	14 (7.5)	5 (0.8)	24.28 (7.68–76.74)†	$1.72 \times 10^{-8}$

\* A total of 186 patients and 627 controls (404 samples from persons of European ancestry [excluding Finns] from the 1000 Genomes Project and 223 pseudo-controls (generated from the chromosomes not transmitted to the affected child from 254 parent–child trios with complete genotype data) were classified according to the number of Hirschsprung's disease risk alleles present at *RET* single-nucleotide polymorphisms (SNPs) rs2435357, rs7069590, and rs2506030 and *SEMA3* SNP rs11766001. All patients had at least one risk allele.

† The odds ratio indicated a significant association (calculated with a two-sided Fisher's exact test) at a family-wise error rate of 0.05, after correction for performing six tests.

cording to enhancer risk-allele dosage, which shows the wide differences in basal susceptibility to Hirschsprung's disease.

#### RISK ASSOCIATED WITH RARE CODING VARIANTS

We first tested whether coding pathogenic alleles, as we defined them, for the 17 known Hirschsprung's disease genes statistically discriminated patients from controls (Table S1 in the Supplementary Appendix). As compared with the 29 pathogenic alleles found in 71 (9.6%) of 740 controls, 36 pathogenic alleles were found in 41 (21.6%) of 190 patients, a percentage 2.25 times as high ( $P = 5.97 \times 10^{-6}$ ) (Table S6 in the Supplementary Appendix), which indicates a higher burden of pathogenic alleles in patients. Furthermore, the pathogenic alleles that were found in patients had a significantly lower mean frequency in an external reference population, the Exome Aggregation Consortium database<sup>21</sup> (ExAC), than did the pathogenic alleles found in controls ( $5.58 \times 10^{-4}$  vs.  $1.11 \times 10^{-3}$ ,  $P = 2.14 \times 10^{-5}$ ) (Table S6 in the Supplementary Appendix), which indicates that the rare coding changes observed in patients have been subject to greater purifying selection than those observed in controls. That is, even though pathogenic alleles in both patients and controls met our definition of pathogenicity, when we compared the frequency of each set (variants in the patients being one set and variants in the con-

trols the other) with the frequency of the specific variants of each set in persons in the ExAC database, those of the patient set were less frequent in the ExAC database than were those in the control set.

To assess the enrichment of pathogenic alleles for each gene, we estimated the probability (P value) of finding as many or a greater number of distinct pathogenic alleles in patients, restricting our analysis to 15,963 single-nucleotide variants in 4027 genes for which there was at least one identified pathogenic allele in both patients and controls. We identified 3 genes, *EDNRB*, *ADAMTS17*, and *ACSS2*, that exceeded the significance threshold of  $1.24 \times 10^{-5}$  (5% significance across 4027 genes) (Fig. S4 in the Supplementary Appendix). More broadly, at a P value threshold of 0.001, we found 10 genes instead of the expected 4 ( $P = 1.3 \times 10^{-3}$ ) (Table 3). We performed functional tests on these 10 genes to distinguish false from true candidates.

The top 10 genes had a minimum of 4 pathogenic alleles each and included both of the major genes, *RET* and *EDNRB*. We also found evidence of 7 novel Hirschsprung's disease genes — *ACSS2*, *ADAMTS17*, *ENO3*, *FAM213A*, *SH3PXD2A*, *SLC27A4*, and *UBR4* — on the basis of both an excess of pathogenic alleles and enteric nervous system gene expression in humans and mice during enterogenesis; assays in zebrafish further confirmed *ACSS2*,



**Table 2. Distribution of Hirschsprung's Disease Risk According to the Molecular Class of Risk Alleles.**

Risk-Allele Class	Independent Genes or Loci	Frequency		Odds Ratio (95% CI)*	P Value	Population Attributable Risk†
		Patients	Controls			
	number	percent	percent			percent
Transcription enhancers, common variants, known‡	2	48.4	17.1	4.54 (3.19–6.46)	1.22×10 <sup>-16</sup>	37.7
Coding genes, rare variants§						
Known and novel loci	24	34.7	5.0	10.02 (6.45–15.58)	3.41×10 <sup>-25</sup>	31.1
Known loci	17	21.6	3.9	6.70 (4.06–11.04)	9.65×10 <sup>-14</sup>	18.2
CNVs, rare variants¶						
Known and novel loci	9	11.4	0.2	63.07 (36.75–108.25)	4.19×10 <sup>-51</sup>	11.3
Known loci	1	5.9	0.1	73.69 (34.97–155.29)	1.23×10 <sup>-29</sup>	9.1

\* Although the differences were not significant, the odds ratios in males were consistently larger than those in females (Table S10 in the Supplementary Appendix).

† The combined population attributable risk for all three classes of pathogenic alleles, under the assumption of independent effects, is 61.9% for all 24 known and novel loci and 53.7% for the 18 known loci.

‡ Five or more common disease variants were observed in 90 of 186 patients and 107 of 627 controls.

§ Rare coding sequence variants were identified in 66 of 190 patients and in 37 of 740 controls.

¶ The copy-number variants (CNV) that were considered to be pathogenic, as reported in this table and in all our other analyses of risk, were clinically identified alterations (e.g., trisomy 21 or 22q deletion) or deletions of more than 500 kb or duplication of more than 1000 kb, with a frequency of less than 1% among controls, that had previously been significantly associated with a developmental disorder. CNVs were identified in 21 of 185 patients and in 40 of 19,584 controls.

|| The only copy-number variant that was previously known to be associated with Hirschsprung's disease was trisomy 21.

*ENO3*, *SH3PXD2A*, and *UBR4* (Fig. S6 in the Supplementary Appendix). The 7 novel genes harbored 39 distinct pathogenic alleles occurring in 40 patients (21.1%), as compared with 23 distinct pathogenic alleles occurring in 28 controls (3.8%) ( $P=3.46\times 10^{-16}$ ). Of the 39 pathogenic alleles in patients, only 6 were identified in 8 controls (1.1%). When all 24 Hirschsprung's disease genes were considered, we identified 75 unique pathogenic alleles occurring in 34.7% of patients (66 of 190), a percentage significantly higher than the 5.0% observed among controls (37 of 740; odds ratio, 10.02; 95% CI, 6.45 to 15.58;  $P=3.41\times 10^{-25}$ ) (Table 2). The mean allele frequencies of the pathogenic alleles in patients and controls in the ExAC database are  $4.22\times 10^{-4}$  and  $8.26\times 10^{-4}$ , respectively, a difference similar in magnitude to the difference we observed for alleles in the 17 previously known Hirschsprung's disease genes. The causality of these variants is further confirmed by higher-than-expected genotype concordance between probands with coding pathogenic alleles and their affected relatives ( $P=0.005$ ) (Tables S6 and S7 and the Methods section in the Supplementary Appendix).

#### PATHWAYS AND FUNCTIONAL GROUPS

Owing to genetic heterogeneity and chance fluctuations, the overall contribution of pathways to Hirschsprung's disease can be estimated more accurately than that of individual genes (Table S7 in the Supplementary Appendix). In Hirschsprung's disease, the RET and EDNRB signaling pathways play major roles with strong epistatic interactions.<sup>7,8,11</sup> Thus, we considered members of the RET (*GDNF*, *NRTN*, *GFRA1*, and *RET*) and EDNRB (*ECE1*, *EDN3*, and *EDNRB*) signaling modules for burden analysis. A third pathway, also epistatic to RET, involves the class 3 semaphorins and their receptors: here we consider only *SEMA3C* and *SEMA3D* because of their association with Hirschsprung's disease.<sup>12,13</sup> A fourth class consists of the transcription-factor genes (*SOX10*, *ZEB2*, *PHOX2B*, and *TCF4*) that are critical to the early development of the enteric nervous system and harbor rare coding variants that cause Hirschsprung's disease-associated syndromes (Table S1 in the Supplementary Appendix). We considered two additional categories: other known genes (*KIF1BP*, *LICAM*, *IKBKAP*, and *NRG1*)<sup>3,10</sup> and the seven novel genes identified in this study.

**Table 3.** Genes with an Excess of Rare Coding Pathogenic Alleles in Hirschsprung's Disease.

Gene	No. of Distinct Pathogenic Alleles (N = 190)		P Value	Embryonic Intestinal Gene Expression Present*		Zebrafish Cell Migratory Defect Present†
	Observed	Expected		Human	Mouse	
<i>EDNRB</i> ‡	6	0.22	$1.35 \times 10^{-7}$	Yes	Yes	Yes
<i>ADAMTS17</i> ‡	5	0.23	$4.45 \times 10^{-6}$	Yes	Yes	NT
<i>ACSS2</i> ‡	6	0.45	$8.08 \times 10^{-6}$	Yes	Yes	Yes
<i>RET</i>	7	0.99	$7.73 \times 10^{-5}$	Yes	Yes	Yes
<i>SLC27A4</i>	4	0.22	$8.48 \times 10^{-5}$	Yes	Yes	No
<i>SH3PXD2A</i>	4	0.22	$8.88 \times 10^{-5}$	Yes	Yes	Yes
<i>MMAA</i>	4	0.23	$9.26 \times 10^{-5}$	No	No	No
<i>ENO3</i>	5	0.45	$1.03 \times 10^{-4}$	Yes	Yes	Yes
<i>FAM213A</i>	4	0.40	$7.51 \times 10^{-4}$	Yes	Yes	No
<i>UBR4</i>	11	3.47	$9.52 \times 10^{-4}$	Yes	Yes	Yes

\* The human embryonic gut samples were evaluated at Carnegie stage 22, and mouse embryonic gut samples were evaluated at embryonic day 10.5.

† NT indicates that the gene was not tested owing to a lack of an identifiable zebrafish orthologue.

‡ The number of distinct pathogenic alleles was significant after multiple-test correction for 4027 genes.

We compared the total numbers of pathogenic-allele genotypes in each of these six classes or pathways among the 66 variant-positive patients with their corresponding frequencies among controls (Table S7 in the Supplementary Appendix). Genes encoding members of the *EDNRB* pathway (odds ratio, 69.03; 95% CI, 8.68 to 547.92), transcription-factor genes (odds ratio, 35.73; 95% CI, 4.15 to 307.72), and novel genes (odds ratio, 23.2; 95% CI, 11.04 to 48.72) had the largest risk effects, followed by genes encoding members of the *RET* pathway (odds ratio, 16.03; 95% CI, 5.21 to 49.28) and *SEMA3C* and *SEMA3D* (odds ratio, 2.65; 95% CI, 1.25 to 5.60). Other known genes (odds ratio, 3.15; 95% CI, 1.22 to 8.09) also made measurable risk contributions, but with an order of magnitude smaller effect. These risk rankings were reflected in the inverse contributions of these classes to the total risk of Hirschsprung's disease. Pathogenic alleles causing greater risk probably have higher penetrance and are therefore selected against with greater intensity. If so, the abundant coding variants in genes of the *RET* pathway have lower penetrance than coding variants in the genes of the *EDNRB* pathway, the genes encoding transcription factors, and the novel genes.

These data also indicate that *RET* has a smaller coding-variant risk burden than previously be-

lieved: 6.3% of the patients (12 patients) had *RET* coding pathogenic alleles, in contrast to approximately 50% from the older data.<sup>2,7</sup> This difference could arise from differing definitions of pathogenicity or from the preponderance of familial and severe cases in earlier studies. Nevertheless, *RET* regulatory pathogenic alleles, which have even lower penetrance than coding pathogenic alleles,<sup>9</sup> were prevalent and, together with *RET* coding variants, conferred substantial risk in 92 of 190 patients (48.4%); this finding highlights the fact that reduced *RET* expression is the predominant cause of Hirschsprung's disease. Moreover, coding or noncoding (in the case of *RET* transcription-enhancer variants) pathogenic alleles in affecting genes that encode members of the *RET* regulatory network,<sup>22</sup> which is made up of *RET*, its transcription factors (*RARB*, *GATA2*, and *SOX10*), its ligands (*GDNF* and *NRTN*), and its coreceptor (*GFRA1*), were found in 120 of our patients (63.2%). In contrast, genes of the *EDNRB* pathway contributed to only 8 cases (4.2%).

#### FREQUENCY OF COPY-NUMBER VARIANTS IN HIRSCHSPRUNG'S DISEASE

Of the 190 patients, 17 (8.9%) had syndromic presentations or known major chromosomal variants (Table 4). To detect subkaryotypic changes, we examined the exome data to identify large copy-

**Table 4. Karyotypes and Large CNVs in Hirschsprung's Disease.**

Karyotype or CNV	Size*	Syndrome Present	Detection Method†	CNVs‡		P Value
				Patients (N = 185)	Controls (N = 19,584)	
	<i>kilobases</i>				<i>number</i>	
<b>Recurrent variant</b>						
Free and mosaic trisomy 21§	47,710	Yes	Karyotyping (9 of 11)	11	17¶	6.68×10 <sup>-16</sup>
16p11.2del§	740–985	Yes (1 of 3), no (2 of 3)	Karyotyping (1 of 3), exome sequencing (3 of 3)	3	12	3.38×10 <sup>-4</sup>
1q21.1dup	509–1,185	No	Exome sequencing (3 of 3)	3	27	2.72×10 <sup>-3</sup>
1q21.1del§	1,425	Yes	Exome sequencing	1	6	6.37×10 <sup>-2</sup>
22q11.2del§	8,000	Yes	Karyotyping	1	0	9.36×10 <sup>-3</sup>
Tetrasomy 22q§	1,447	Yes	Karyotyping	1	0	9.36×10 <sup>-3</sup>
17p11.2dup (CMT1A)§	1,835	No	Exome sequencing	1	5	5.49×10 <sup>-2</sup>
<b>Nonrecurrent variant</b>						
47,XX,+der(15)t(4:15)§	7,768 (chr 4); 3,800 (chr 15)	Yes	Karyotyping, exome sequencing	1	0	9.36×10 <sup>-3</sup>
1p33del	582	No	Exome sequencing	1	0	9.36×10 <sup>-3</sup>
12p13.31del	554	Yes	Exome sequencing	1	0	9.36×10 <sup>-3</sup>
13q21.33-q31.1del§**	14,356	Yes	Karyotyping	1	0	9.36×10 <sup>-3</sup>
2q21.2-q22.2del§	8,847	Yes	Exome sequencing	1	0	9.36×10 <sup>-3</sup>
8p23.3del	579	Yes	Exome sequencing	1	0	9.36×10 <sup>-3</sup>
2p25.3dup	1,377	No	Exome sequencing	1	1	1.86×10 <sup>-2</sup>
7q21.12dup	1,498	No	Exome sequencing	1	11	1.06×10 <sup>-1</sup>
10q24.3-q26.13inv	25,600	Yes	Karyotyping	1	—	—

\* The estimated smallest region determined on the basis of karyotype, exome sequencing, or SNP array data is shown. The abbreviation chr denotes chromosome.

† Variants were detected by karyotyping, exome sequencing, or both, with validation by SNP array, including in two patients with trisomy 21 for whom we did not have a submitted karyotype.

‡ The control data are from the study by Coe et al.<sup>17</sup>; however, the numbers for the variant 47,XX,+der(15)t(4:15) are for the two duplications at the translocation site, not for the translocation, and the control numbers were not available and not expected for the 10q24.3-q26.13 inversion.

§ The variant had previous evidence of pathogenicity in another developmental disorder.

¶ The number was estimated from population data (Table S8 in the Supplementary Appendix).

|| The P value was significant after multiple-test correction.

\*\* The 13q21.33-q31.1del CNV deletes *EDNRB*.

number variants. In total, we identified 16 distinct copy-number variants; 14 of these variants (and their loci) were not previously known to be associated with Hirschsprung's disease (Table 4). We assessed the pathogenicity of each variant on the basis of its enrichment in patients or their association with a known developmental disorder to identify 9 chromosomal variants and copy-number variants in 11.4% of patients (21 of 185), with a corresponding frequency of 0.2% (40 of 19,584) in controls, a highly significant effect

(odds ratio, 63.07; 95% CI, 36.75 to 108.25;  $P=4.19\times 10^{-51}$ ) (Tables 2 and 4, and Table S9 in the Supplementary Appendix).<sup>2,3,6</sup>

Of the 21 instances of pathogenic chromosomal variants in patients, 18 (86%) were recurrent and 3 were nonrecurrent, and 18 were in patients with syndromic presentations (Table S9 in the Supplementary Appendix). The most frequent (11 variants, 52%) recurrent finding was trisomy 21, but the other 7 occurred at well-known loci for other genomic disorders. The elevated fre-

quency of trisomy 21 among patients with Hirschsprung's disease (odds ratio, 73.69; 95% CI, 34.97 to 155.29;  $P=1.23\times 10^{-29}$ ) (Table 2) is not surprising, given previous observations.<sup>14</sup> However, the 16p11.2del copy-number variant, which is usually associated with autism,<sup>27</sup> is also significantly enriched (odds ratio, 30.03; 95% CI, 9.70 to 92.97;  $P=3.62\times 10^{-9}$ ). Overall, the 9.7% frequency of patients with Hirschsprung's disease who have recurrent chromosomal variants is significantly higher than the expected frequency (odds ratio, 53.30; 95% CI, 30.30 to 93.76;  $P=2.60\times 10^{-43}$ ). These recurrent chromosomal changes are known to be associated with intellectual disability, autism, neurodevelopmental delay, epilepsy, and Charcot-Marie-Tooth disease type 1A,<sup>27,28</sup> perhaps owing to pathways common to the enteric and central nervous systems. The three nonrecurrent variants, one of which deletes *EDNRB*, were unique, and all occurred in patients with syndromic presentations (Table 4).

#### DISTRIBUTION OF DIVERSE PATHOGENIC ALLELES

Pathogenic alleles in at least 32 genes and loci contribute to Hirschsprung's disease: rare coding variants in 24 genes, common noncoding variants at four sites within 2 loci, and large copy-number variants and chromosomal anomalies in at least 8 additional loci (not including 13q21.33-q31.1del, which overlaps *EDNRB*). The common noncoding risk genotypes (five or more risk alleles), rare coding variants, and copy-number variants occur at decreasing (by orders of magnitude) frequencies in the general population — 17.1%, 5.0%, and 0.2% — but with increasing odds ratios of 4.54, 10.02, and 63.07, respectively (Table 2). In consequence, all three variant classes make major contributions to the risk of Hirschsprung's disease, with population attributable risks of 37.7%, 31.1%, and 11.3%, respectively, and a total attributable fraction of 61.9%. In addition, although the differences are not significant, the odds ratios among males are consistently higher than those among females (Table S10 in the Supplementary Appendix). Thus, the sex effect in Hirschsprung's disease is not caused by a specific gene or variant but is a property of the disorder. We conclude that, first, even in this rare disorder, common variants are responsible for the majority of cases of Hirschsprung's disease, despite their individually lower risks, because of their high population prevalence. Second, the total risk from all rare

coding pathogenic alleles (which have a much higher penetrance) is also high but is differentially spread over 24 genes. Third, the population risk from copy-number variants is the lowest, spread over the effects of 9 loci but with a majority contribution from trisomy 21. These risks from both known and novel genes and loci are almost certainly overestimates owing to the “winner's curse.” Consequently, we reestimated the risks, taking into consideration only the well-established risk factors and genes known before this study, and we found the same pattern: these variant classes occur at frequencies of 17.1%, 3.9%, and 0.1% in the general population, but with increasing risks — odds ratios of 4.54, 6.70, and 73.69, respectively (Table 2). These three categories contribute to the population attributable risks of 37.7%, 18.2%, and 9.1%, respectively, or a total attributable fraction of 53.7%.

Finally, we quantified the risk associated with combinations of pathogenic alleles (Table S11 in the Supplementary Appendix).<sup>29</sup> We classified each patient's total burden of pathogenic alleles according to sex, segment length, familiarity, and the presence or absence of additional anomalies; we pooled all patients with copy-number variants into one class, given the low frequency of this type of variant. The results showed three cardinal features (Table 5). First, genetic risk factors of any type were identifiable in 72.1% of patients, and patients harbored various combinations of different types of pathogenic alleles, all in significant excess relative to controls. Second, each of the three variant classes (five or more common noncoding variants, rare coding variants, and copy-number variants) were present in substantial percentages of diagnoses (48.4%, 34.7%, and 11.4%, respectively) (Table 2). One, two, or three different classes of molecular lesion were present in 51.9%, 18.4%, and 1.7% of patients, respectively — roughly their expected frequencies — with no evidence of interaction, a finding consistent with multifactorial expectations, although the statistical power for such detection is probably low (Table 5, and Table S11 in the Supplementary Appendix). Third, the genotype-specific odds ratios for Hirschsprung's disease, estimated in reference to the class with no identifiable genetic risk factor, vary by a factor of 67 and increase with the pathogenic allele burden. These data allow us to estimate the absolute risk of Hirschsprung's disease, given a person's geno-



type. Persons with no identifiable risk factors have an estimated population risk of 5.33 per 100,000 (approximately 1 per 18,800), a low risk of disease. At the other extreme, persons with both common enhancer risk genotypes and rare coding variants and those with copy-number variants have substantial estimated risks of 2.85 per 1000 (approximately 1 per 350) and 8.38 per 1000 (approximately 1 per 120), respectively.

We did not detect any significant genotype–phenotype associations with respect to sex, segment length, familiarity, or syndromic status. However, patients with a copy-number variant and patients with both a common transcription-enhancer risk genotype and a rare coding variant — the two classes with the highest relative risks — are characterized by an excess representation of males and of nonfamilial cases. The sex ratio in classes with no evident pathogenic alleles or those with rare coding single-nucleotide variants only is approximately 1. This latter class is most often seen in persons with an affected relative (familial disease), which suggests that most segregating pathogenic alleles in affected families are rare coding variants. There was also a greater tendency for Hirschsprung's disease to be syndromic among patients in higher risk classes than among those in lower risk classes.

## DISCUSSION

Hirschsprung's disease can arise both from low-penetrance genetic disorders<sup>2,6–8</sup> and from high-penetrance monogenic syndromes.<sup>2,3</sup> Risk prediction and genetic counseling therefore depend on family history, risk factors (sex and segment length), and targeted assessment for syndromic features.<sup>6</sup> Thus, in a small subset of patients, classical genetic testing of *RET*, *EDNRB*, and genes that are associated with syndromes may be informative. The results reported here, however, suggest that widespread genomic analyses may be useful for clinical research and improved risk stratification.

Hirschsprung's disease is usually an isolated condition and unassociated with family history. However, genetic causal factors can be identified in approximately 72% of cases, for which molecular class, frequency, and disease risk can be quantified on the basis of sequence data alone, explaining between 53.7% and 61.9% of population attributable risk. Approximately 21% of pa-

tients have multiple risk factors, with the genotype-specific incidence increasing by a factor of more than 100 (risk ranging from approximately 1 in 18,800 to 1 in 120) as the number of genotypic risk factors increases from zero to three. Therefore, we have sufficient quantification of disease risk according to genotype to address questions of underlying causes and genetic architecture and to provide genetic counseling for the highest-risk 21% of patients and their relatives.

We have made considerable strides in understanding the functional basis of Hirschsprung's disease. The majority of the 32 genes and loci are known to have roles in the development of the enteric nervous system. In contrast, the majority of patients (63.2%) have identifiable pathogenic alleles only within the known *RET* regulatory network, which lead to decreased *RET* signaling. The *RET* effect is potentially even larger, affecting 78.9% of patients, because an additional 5.8% of patients harbor pathogenic alleles in *UBR4*, a novel E3 ligase gene identified in this study and a candidate for *RET* signal termination; 5.9% of patients have trisomy 21, which results in an elevated dosage of *SOD1*, encoding a negative regulator of *RET*<sup>14</sup>; and 4.2% of cases involve *EDNRB*, which is *SOX10*-regulated.<sup>30,31</sup> Thus, genetic testing of at least the *RET* regulatory network is warranted for risk stratification.

In order to understand the biology of enteric nervous system cell proliferation, migration, colonization, and neuronal specialization, it is important to understand the steps subsequent to the transition and differentiation of enteric nervous system cells, such as the likely axonal guidance functions of *SEMA3C* and *SEMA3D*.<sup>12</sup> The seven novel genes identified here, all of which are expressed in the human gut at the appropriate developmental stages, probably control some aspects of axonal guidance, cell proliferation, and local inflammation (Table S12 in the Supplementary Appendix). We hypothesize that screening the genes regulating these processes in early gut development will further resolve the remaining approximately 40% of Hirschsprung's disease risk.

A continuing challenge in the study of Hirschsprung's disease is to understand the cellular mechanisms underlying the disease. Whether we consider the persons with the highest (1 in 120) or lowest (1 in 18,800) risk, the absolute risk of disease is still small. What are the cellular events that trigger or prevent aganglionosis, given a par-

**Table 5. Distribution of Hirschsprung's Disease Cases by Genetic Risk Profile and Population Effects.**

	Risk Class**	Patients				Odds Ratio (95% CI)†	Estimated Incidence‡	Male Sex	Short Segment§	Simplex¶	Nonsyndromic
		Coding**	CNV††	Observed‡‡ no. (%)	Expected§§ no.						
Noncoding											
	–	–	–	50 (28)	140.69	1.00 (Reference)	5.33×10 <sup>–5</sup>	26 (52)	17 (46)	28 (56)	38 (76)
	+	–	–	53 (30)	29.02	5.07 (2.93–8.76)	2.74×10 <sup>–4</sup>	35 (66)	24 (57)	36 (68)	46 (87)
	–	+	–	27 (15)	7.41	9.73 (4.22–22.39)	5.47×10 <sup>–4</sup>	14 (52)	9 (41)	14 (52)	17 (63)
	+	+	–	29 (16)	1.53	40.68 (10.84–152.67)	2.85×10 <sup>–3</sup>	24 (83)	14 (58)	24 (83)	20 (69)
	–	–	+								
	+	–	+								
	–	+	+	20 (11)	0.36	66.80 (11.44–389.96)	8.38×10 <sup>–3</sup>	29 (81)	18 (58)	30 (83)	21 (58)
	+	+	+								

\* Each risk class was defined on the basis of the combination of types of pathogenic variants present (+) and absent (-).

† Odds ratios were calculated with patients with Hirschsprung's disease who had no detectable disease-associated variants used as the reference group.

‡ Population incidences were calculated by assuming a total incidence of 15 cases of Hirschsprung's disease per 100,000 live births. The observed numbers of the variant classes (50, 53, 27, 29, and 20) are close to the expected numbers estimated from random association of disease variants (53.44, 50.12, 28.40, 26.64, and 20.41;  $\chi^2 = 0.67$ ; 1 df;  $P = 0.41$ ).

§ Hirschsprung's disease is classified as short, long, or total on the basis of the length of the length of bowel segment affected.

¶ A simplex case is one in which the patient has no known family members with the disease.

|| Category includes patients with five or more common noncoding risk alleles at *RET* (rs2435357, rs2506030, and rs7069590) and *SEMA3D* (rs11766001).

\*\* A coding variant was considered to be present if the patient had at least one rare, deleterious variant found in any of the 24 Hirschsprung's disease susceptibility genes.

†† The CNVs that were considered to be pathogenic, as reported in this table and in all our other analyses of risk, were clinically identified alterations (e.g., trisomy 21 or 22q deletion) or deletions of more than 500 kb or duplication of more than 1000 kb, with a frequency of less than 1% among controls, that had previously been associated with a developmental disorder. All patients with CNVs were pooled into one class, given the low frequency of this type of variant.

‡‡ Percentages were calculated with 179 used as the denominator (i.e., the patients with complete data for all three mutation classes).

§§ Expected values are calculated from the frequencies of each risk class among controls (see Table 2).

ticular genotype? A part of the answer is the existence of very rare de novo gene mutations affecting the enteric nervous system,<sup>32</sup> which require larger cohorts of trios for the detection of an association. However, the incomplete penetrance of most Hirschsprung's disease variants implies that stochastic, environmental, or epigenetic factors must be important.

In our study, we found that the risk of the complex phenotype that is Hirschsprung's disease stemmed from a combination of variants in numerous genes and different classes of genetic variants: noncoding variants, single-nucleotide variants and copy-number variants, and both rare and common variants. Despite the current thinking in human medical genetics, most of the risk of Hirschsprung's disease arose from a common widespread genetic susceptibility, on top of which

rare coding and rarer copy-number variants exacerbated the risk. Despite this molecular diversity, the implicated genes clustered, on the basis of their known function, into gene regulatory networks (which, in Hirschsprung's disease, regulate the transition from enteric neural-crest cells to enteric neuroblasts, axonal guidance, and neuroblast proliferation), a model that may be relevant to the understanding of other complex disorders.

Supported by grants from the National Institutes of Health (MERIT award R37 HD28088 to Dr. Chakravarti, R01 MH101221 to Dr. Eichler, and U54 HG003067 to Dr. Gabriel). Dr. Eichler is a Howard Hughes Medical Institute investigator.

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

We thank the numerous patients and their family members, physicians, and genetic counselors who have contributed to these studies over the years; and Erick Kaufmann, Jennifer (Scott) Bubb, Sue Lewis, Maura Kenton, and Julie Albertus for family ascertainment and genetic counseling.

## REFERENCES

1. Carter CO. Genetics of common disorders. *Br Med Bull* 1969;25:52-7.
2. Chakravarti A, Lyonnet S. Hirschsprung disease. In: Scriver CR, Beaudet AL, Valle D, et al., eds. *The metabolic and molecular bases of inherited disease*. 8th ed. Vol. 4. New York: McGraw-Hill, 2001: 6231-55.
3. Amiel J, Sproat-Emison E, Garcia-Barcelo M, et al. Hirschsprung disease, associated syndromes and genetics: a review. *J Med Genet* 2008;45:1-14.
4. Bodian M, Carter OO. A family study of Hirschsprung's disease. *Ann Hum Genet* 1963;26:261-77.
5. Passarge E. The genetics of Hirschsprung's disease: evidence for heterogeneous etiology and a study of sixty-three families. *N Engl J Med* 1967;276: 138-43.
6. Badner JA, Sieber WK, Garver KL, Chakravarti A. A genetic study of Hirschsprung disease. *Am J Hum Genet* 1990;46:568-80.
7. Ederly P, Lyonnet S, Mulligan LM, et al. Mutations of the RET proto-oncogene in Hirschsprung's disease. *Nature* 1994; 367:378-80.
8. Puffenberger EG, Hosoda K, Washington SS, et al. A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* 1994; 79:1257-66.
9. Emison ES, McCallion AS, Kashuk CS, et al. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* 2005;434:857-63.
10. Garcia-Barcelo MM, Tang CS, Ngan ES, et al. Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc Natl Acad Sci U S A* 2009;106:2694-9.
11. Emison ES, Garcia-Barcelo M, Grice EA, et al. Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. *Am J Hum Genet* 2010; 87:60-74.
12. Jiang Q, Arnold S, Heanue T, et al. Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet* 2015; 96:581-96.
13. Kapoor A, Jiang Q, Chatterjee S, et al. Population variation in total genetic risk of Hirschsprung disease from common RET, SEMA3 and NRG1 susceptibility polymorphisms. *Hum Mol Genet* 2015;24:2997-3003.
14. Arnold S, Pelet A, Amiel J, et al. Interaction between a chromosome 10 RET enhancer and chromosome 21 in the Down syndrome-Hirschsprung disease association. *Hum Mutat* 2009;30:771-5.
15. Dasgupta R, Langer JC. Evaluation and management of persistent problems after surgery for Hirschsprung disease in a child. *J Pediatr Gastroenterol Nutr* 2008; 46:13-9.
16. Menezes M, Corbally M, Puri P. Long-term results of bowel function after treatment for Hirschsprung's disease: a 29-year review. *Pediatr Surg Int* 2006;22:987-90.
17. Coe BP, Witherspoon K, Rosenfeld JA, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet* 2014;46:1063-71.
18. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491-8.
19. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res* 2012;22:1525-32.
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38(16):e164.
21. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536:285-91.
22. Chatterjee S, Kapoor A, Akiyama JA, et al. Enhancer variants synergistically drive dysfunction of a gene regulatory network in Hirschsprung disease. *Cell* 2016;167(2):355-368.e10.
23. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;133:1-9.
24. Itsara A, Cooper GM, Baker C, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 2009;84:148-61.
25. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
26. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015;526:68-74.
27. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011;1380:42-77.

28. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* 1991;66:219-32.
29. Angrist M, Bolk S, Halushka M, Lapchak PA, Chakravarti A. Germline mutations in glial cell line-derived neurotrophic factor (GDNF) and RET in a Hirschsprung disease patient. *Nat Genet* 1996;14:341-4.
30. Carrasquillo MM, McCallion AS, Puffenberger EG, Kashuk CS, Nouri N, Chakravarti A. Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. *Nat Genet* 2002; 32:237-44.
31. Stanchina L, Baral V, Robert F, et al. Interactions between Sox10, Edn3 and Ednrb during enteric nervous system and melanocyte development. *Dev Biol* 2006; 295:232-49.
32. Gui H, Schriemer D, Cheng WW, et al. Whole exome sequencing coupled with unbiased functional analysis reveals new Hirschsprung disease genes. *Genome Biol* 2017;18:48.

Copyright © 2019 Massachusetts Medical Society.

**JOURNAL ARCHIVE AT NEJM.ORG**

Every article published by the *Journal* is now available at NEJM.org, beginning with the first article published in January 1812. The entire archive is fully searchable, and browsing of titles and tables of contents is easy and available to all. Individual subscribers are entitled to free 24-hour access to 50 archive articles per year. Access to content in the archive is available on a per-article basis and is also being provided through many institutional subscriptions.