ARTICLE

Differential Contributions of Rare and Common, Coding and Noncoding Ret Mutations to Multifactorial Hirschsprung Disease Liability

Eileen Sproat Emison,[1] Merce Garcia-Barcelo,[2] Elizabeth A. Grice,[1] Francesca Lantieri,[3,4] Jeanne Amiel,[5,6] Grzegorz Burzynski,[7] Raquel M. Fernandez,[8,9] Li Hao,[1] Carl Kashuk,[1] Kristen West,[1] Xiaoping Miao,[2] Paul K.H. Tam,[2] Paola Griseri,[4] Isabella Ceccherini,[4] Anna Pelet,[5,6] Anne-Sophie Jannot,[5,6,10] Loic de Pontual,[5,6] Alexandra Henrion-Caude,[5,6] Stanislas Lyonnet,[5,6] Joke B.G.M. Verheij,[7] Robert M.W. Hofstra,[7] Guillermo Antiñolo,[8,9] Salud Borrego,[8,9] Andrew S. McCallion,[1] and Aravinda Chakravarti[1,*]

The major gene for Hirschsprung disease (HSCR) encodes the receptor tyrosine kinase RET. In a study of 690 European- and 192 Chinese-descent probands and their parents or controls, we demonstrate the ubiquity of a >4-fold susceptibility from a C→T allele (*rs2435357*: $p = 3.9 \times 10^{-43}$ in European ancestry; $p = 1.1 \times 10^{-21}$ in Chinese samples) that probably arose once within the intronic *RET* enhancer MCS+9.7. With in vitro assays, we now show that the T variant disrupts a SOX10 binding site within MCS+9.7 that compromises RET transactivation. The T allele, with a control frequency of 20%–30%/47% and case frequency of 54%–62%/88% in European/Chinese-ancestry individuals, is involved in all forms of HSCR. It is marginally associated with proband gender ($p = 0.13$) and significantly so with length of aganglionosis ($p = 7.6 \times 10^{-5}$) and familiality ($p = 6.2 \times 10^{-4}$). The enhancer variant is more frequent in the common forms of male, short-segment, and simplex families whereas multiple, rare, coding mutations are the norm in the less common and more severe forms of female, long-segment, and multiplex families. The T variant also increases penetrance in patients with rare *RET* coding mutations. Thus, both rare and common mutations, individually and together, make contributions to the risk of HSCR. The distribution of *RET* variants in diverse HSCR patients suggests a "cellular-recessive" genetic model where both *RET* alleles' function is compromised. The *RET* allelic series, and its genotype-phenotype correlations, shows that success in variant identification in complex disorders may strongly depend on which patients are studied.

Introduction

The hallmark of Mendelian disease is that the underlying mutations have the same, recognizable inheritance pattern and effect (loss or gain of function) across families, independent of genetic background. Different families often have different molecular mutations: allelic diversity is high and usually maintained by mutation-selection balance of individually rare mutations. In human families segregating such a disease, the recurrence risks to relatives depend only on the degree of relationship to the proband, and trivially on the frequency of the mutant allele in the population, thus making genetic counseling simple and pragmatic. At the other extreme, numerous, if not most, human diseases have complex, multifactorial inheritance where the causation arises from the small allelic differences at many genes in an equal and additive manner.[1] Complex disorders are now thought to involve significant unequal and nonadditive genetic effects as well. For these phenotypes, recurrence risk is not a property of any single gene but rather an emergent property of the ensemble genotypes at many loci.

Cedric Carter was the first to outline several major and unique genetic features of multifactorial diseases (traits) that distinguish them from Mendelian ones.[2] First, the underlying disease allele is polymorphic. Second, recurrence risks in relatives of a proband depend inversely on population incidence, increasing with greater severity and being greater for a less frequently affected class. Finally, recurrence risks vary across families, even for the same genetic relationship. Thus, genetic counseling is much more complex and dependent on the degree of familial aggregation.[3,4] Carter explained these "non-Mendelian" features by using the additive multifactorial model of genetic effects that produces a normal distribution of liability (genotype and exposure) in the population, a distribution that shifts higher along the liability scale in relatives of a proband, with affected individuals having liability beyond a threshold.[2] Although the multifactorial model is well established, its molecular properties remain unclear, specifically how the ensemble of multilocus genotypes lead to the corresponding non-Mendelian phenotypic features. Indeed, both the "common disease-common

[1]Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [2]Department of Surgery, The University of Hong Kong, Pokfulam, Hong Kong; [3]Department of Health Science, Biostatistics Unit, University of Genova, Genova 16126, Italy; [4]Laboratorio di Genetica Molecolare, Istituto Gaslini, Genova 16147, Italy; [5]Département de Génétique, Université Paris Descartes, Faculté de Médecine, Paris 75006, France; [6]INSERM U-781, AP-HP Hôpital Necker-Enfants Malades, Paris 78743, France; [7]Department of Medical Genetics, University of Groningen, Groningen 9700, The Netherlands; [8]Unidad de Gestión Clínica de Genética, Reproducción y Medicina Fetal, Hospitales Universitarios Virgen del Rocío, Seville 41013, Spain; [9]CIBER de Enfermedades Raras (CIBERER), ISCIII, Seville 41013, Spain; [10]INSERM U-535, "Génétique épidémiologique et structure des populations humaines," Villejuif F-94817, France
*Correspondence: aravinda@jhmi.edu
DOI 10.1016/j.ajhg.2010.06.007.

variant" and the "common disease-rare variant" hypotheses may be true for specific mutant alleles but need not be general rules for all complex disease genes and their alleles.[5,6] To explore how the actual allelic diversity of a gene impacts the phenotypic diversity and non-Mendelian features of a complex disease, we examined the mutational diversity and phenotypic consequences of the *RET* (MIM 164761) tyrosine kinase in Hirschsprung disease (HSCR [MIM 142623]). We demonstrate that both rare and common mutations, coding and noncoding, within the same gene contribute to HSCR in unique and specific ways that lead to recognizable genotype-phenotype associations. Consequently, for a complex disease, the patients we choose to study can uncover either rare or common variants and may be one major reason for recent debates on the frequency of mutant alleles and their genetic effects, and the success or failure of replication of genetic findings.

Hirschsprung disease (HSCR), or congenital intestinal aganglionosis, is a developmental defect associated with the lack of intramural ganglion cells in the myenteric and submucosal plexuses along varying segments of the gastrointestinal tract: it is classified into short-segment (S-HSCR; aganglionosis up to the upper sigmoid colon), long-segment (L-HSCR; aganglionosis up to the splenic flexure and beyond), and total colonic aganglionosis (TCA) forms. The disorder shows marked differences in incidence with rates of 15 and 28 cases per 100,000 live births among Europeans and Asians, respectively.[3] HSCR has all the hallmarks of a classical multifactorial disorder displaying high heritability (81%–100% depending on the sex of the proband and affected sibling), large sex difference (sex ratio 3.9), high sibling recurrence risk (200-fold greater than the population), and non-Mendelian inheritance in families.[4] Indeed, HSCR was one of Carter's seminal examples of multifactorial inheritance.[7]

Positional cloning and candidate gene analysis in syndromic and familial cases have identified 11 genes (*RET, GDNF* [MIM 600837], *NRTN* [MIM 602018], *SOX10* [MIM 602229], *EDNRB* [MIM 131244], *EDN3* [MIM 131242], *ECE1* [MIM 600423], *ZFHX1B* [MIM 605802], *PHOX2B* [MIM 603851], *TCF4* [MIM 602272], and *KIAA1279* [MIM 609367]) with mutations in HSCR; of these, the tyrosine kinase *RET* is the major gene with >80% of all known mutations.[3,8] Coding or splice junction mutations at these genes have between 50% and 70% penetrance, show sex dependence, and are rare (<1%) in the population. These mutations have been identified in ~50% of familial (mostly L-HSCR, TCA) and up to 20% of sporadic (mostly S-HSCR) cases and cumulatively explain ~0.1% of the heritability. Consequently, although the identified genes have led to a deep understanding of the genetic basis of HSCR and are important to specific families, they are not a significant explanation of its incidence. In contrast, a common (24% frequency in Europeans) *RET* susceptibility allele within an enhancer in intron 1 impacts all forms of HSCR and explains ~2% (1.14% in females, 2.63% in males) of the heritability or 20-fold greater than the contribution by rare coding variants.[9] The vast majority (80%) of HSCR occurs as S-HSCR with multifactorial inheritance, a recurrence risk of 4%, and a sex ratio of 5.5.[4] Thus, the typical HSCR case is isolated (nonsyndromic and negative family history) and the few multiplex families observed are largely comprised of affected siblings. Genome-wide linkage studies have identified three loci (*RET* and two yet unidentified genes at 3p12 and 19q21) that are together necessary and sufficient to explain the segregation and recurrence risk of S-HSCR.[10] Other modifier genes have also been mapped.[8] However, *RET* is the only gene known to play a major role in *all* forms of HSCR susceptibility, and a quantitative and unbiased study of its allelic spectrum provides clues on the role of both rare and common sequence variants in the complex inheritance of HSCR, particularly its relationships to disease subtypes that are correlated with risk, namely, gender, segment length, and familiality.

We studied 882 probands and their 1478 first-degree relatives from US, European, and Chinese families to examine the role of *RET* genetic variation with respect to HSCR and its subtypes and performed two types of molecular analyses: sequencing of the protein coding sequence (CDS) and haplotype analysis of single nucleotide polymorphisms (SNPs) across the entire 61.8 kilobase (kb) *RET* locus. We unequivocally demonstrate the significantly different genetic properties of rare versus common variants, their complementary effects on disease severity and prevalence, and the sensitivity of recurrence risk to correlated risk factors. Our conclusions have significant implications for disease *gene* versus disease *variant/mutation* discovery and interpretations of the genetic architecture of complex traits.

Material and Methods

Patient Collection and DNA Samples

Each center ascertained patients and their families under informed consent and with ethical guidelines of their respective home institutions. HSCR probands were identified either from a survey of surgical/clinical records from hospitals or large physician practices or from referrals to each center by their physicians, genetic counselors, or family members: therefore, there is the likelihood of biased representation of specific disease subtypes at each center. The phenotypic diagnosis of HSCR and the segment length involved was based on surgery, pathology, or other medical records, with affected individuals being classified as total colonic aganglionosis (TCA), long-segment (L-HSCR; aganglionosis of the splenic flexure and beyond), or short-segment (S-HSCR; aganglionosis distal to the splenic flexure and usually involving the sigmoid colon) HSCR. A random collection of HSCR patients, without regard to segment length, gender, or familiality, with or without other associated anomalies, and with or without demonstrated coding sequence *RET* mutations, were studied. Each investigator provided DNA samples and phenotypic data from probands and their families for central genetic and statistical analysis at Johns Hopkins University except that the Chinese case-control samples were genotyped in Hong Kong. In total, 586 father-mother-proband trios (133 from France, 127 from Spain,

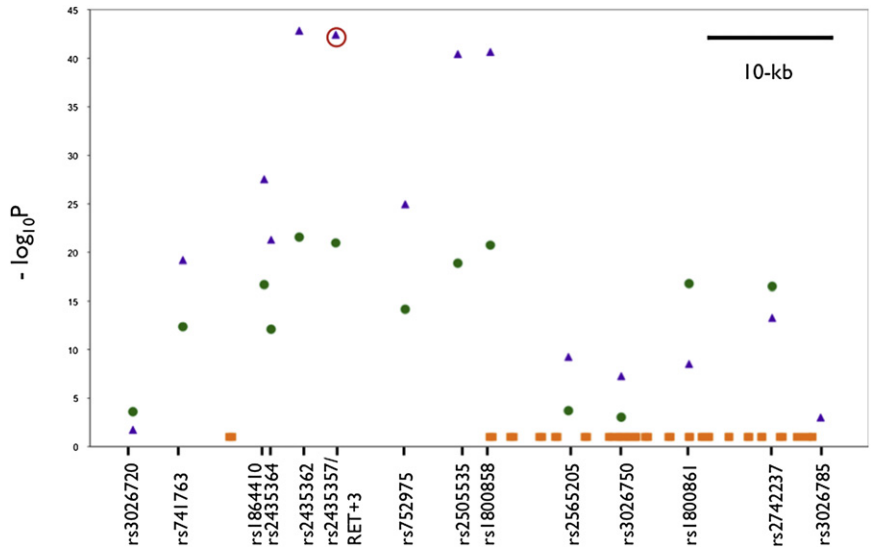

Figure 1. Genetic Associations between *RET* and HSCR in Two Populations
Transmission disequilibrium test (TDT) of polymorphisms across the *RET* locus. -logP values for each of 14 SNPs, with purple triangles representing genotypes from individuals of European descent (Italy, France, Netherlands, Spain, and US) and green circles those of Chinese descent (trio and case-control p values were combined via Fisher's method), are shown on the *y* axis against the physical locations of SNPs with respect to *RET* exons (orange). The circled SNP *rs2435357* segregates a functional hypomorphic enhancer mutation; *rs3026785* is monomorphic in the Chinese.

120 from USA, 125 from Netherlands, 81 from Italy) and 104 parent-proband duos (27 from France, 14 from Spain, 47 from USA, 4 from Netherlands, 12 from Italy) of European descent were genotyped; 94 parent-child samples (49 trios and 45 duos) together with 98 cases and 168 controls from Hong Kong were also examined. The distribution of gender, familiality, and segment length for the complete set of 882 probands is provided in Table S1 available online.

SNP Genotyping

A total of 2672 samples were genotyped but after sample and data cleaning, 882 probands together with 1260 of their parents (24 affected), 218 siblings (65 affected sibs), and 64 other relatives (22 affected), for a total of 2424 samples, were analyzed (Table S2). The DNA samples were shipped as lyophilized genomic DNA and were resuspended in water prior to genotyping. All liquid handling steps were completed on a Tecan Genesis Workstation. Each of 14 SNPs (Table S3) was genotyped with Taqman technology (Applied Biosystems, Foster City, CA). The SNPs were chosen to physically span the entire locus (Figure 1; Table S3) so that copy number mutations could be detected, to span the two major LD blocks at *RET* so that the genetic association signal could be detected anywhere within the gene and based on our prior knowledge of SNP assay performance. In brief, genotyping reactions were cycled on MJ tetrads and endpoint reads performed on the ABI Prism 7900HT with the SDS2.2 software for allelic discrimination. Genotypes were called only when samples were within a clearly identified cluster. One hundred and twenty clinical samples and four CEPH families were genotyped in duplicate: from a single discordant call among 3114 replicate genotypes, we estimated our error rate to be 0.032%. Information on the genotyping performance of the SNPs studied in the trios is shown in Table S3. Genotyping in "clinical" or nonreference samples can expose several different types of "errors:" sample misidentification (laboratory errors or disputed parentage), DNA degradation, genotyping errors, and copy number variants and mutations. All of these families have been used in previous HSCR genome-wide linkage analyses, making disputed paternity very unlikely; furthermore, no trio demonstrated apparent parental exclusions at multiple markers except for nine families that probably harbor de novo deletion mutations (see main text). This inference was based on finding SNPs with genotypes in parents well within clusters but with offspring genotypes as apparent homozygotes and/or outside clusters; the claims were buttressed with the finding of multiple SNPs showing Mendelian inconsistencies or being consistent with deletion from a single parental origin. To address DNA sample quality issues, we observed that missing genotypes were nonrandomly distributed across samples: ~30% of samples had any missing data, the majority (62%) of which had only one missing genotype and 97% of which had six or fewer missing genotypes; we chose to eliminate from analysis the ~3% of samples that showed seven or more missing genotypes (Figure S1). Genotyping errors remain a small possibility in these data; to reduce this impact, we eliminated all Mendelian inconsistencies to observe a genotyping call rate of more than 95% with the exception of *rs1864410*. To assess genotype quality, we tested the Hardy-Weinberg equilibrium on all parents of trios; as shown in Table S3, none of the SNPs were statistically significant given multiple corrections (14 tests: $p < 0.0036$). The 98 cases and 168 controls from Hong Kong were genotyped with a multiplexed assay with Sequenom (San Diego, CA) technology; all SNPs, except *rs3026785*, were polymorphic and clustered well.

DNA Sequencing

DNA sequencing to reveal coding sequence (CDS) mutations in the 20 *RET* exons was performed by PCR amplification followed by automated DNA sequencing on an ABI 3100 or 3730XL; the Netherlands samples were prescreened by SSCP (single strand conformational polymorphism) and DGGE (density gradient gel electrophoresis) while the Italian samples were prescreened by DHPLC (denaturing high performance liquid chromatography) prior to DNA sequencing. For this study, we distinguished disease-causing mutations from neutral variation if the alteration was absent from 200 control chromosomes from the respective local population and (1) introduced a premature stop codon, (2) altered a conserved splice junction, or (3) was statistically predicted to be deleterious or had been previously demonstrated to affect RET protein function.[11] We identified a total of 75 unique (10 nonsense or frame shift, 48 missense, and 15 splice junction, 2 noncoding) mutations in 85 families (16 French, 9 Spanish, 20 Dutch, 17 Italian, 17 Chinese, 6 US families).

Association Data Analysis

From the genotypes of trios, and separately for the Chinese case-control samples, haplotypes were estimated with PHASE v2.1[12] and classified as being transmitted (Tr) or untransmitted (Un) to HSCR offspring. Allele frequencies were estimated separately for Tr and Un chromosomes by gene counting and tests of Hardy-Weinberg performed by a standard χ^2 test with one degree of freedom (df) based on the four classes AA, AB, BA, and BB for alleles A and B on Tr and Un chromosomes, respectively. This tactic was necessary because there was prior knowledge of association between *RET* SNPs and HSCR and failure to account for this known allelic difference on Tr/Un chromosomes would falsely inflate the χ^2 statistic. Population recombination rates were estimated for Tr and Un chromosomes separately with complete trios and the simple hot spot model implemented in PHASE v2.1. Estimates of linkage disequilibrium (LD) on Tr and Un chromosomes were calculated and visualized with GOLD[13] (Figure S2). All association tests between *RET* genotypes and HSCR subtypes were performed with standard contingency χ^2 tests with 1 or 2 df, as appropriate. The transmission disequilibrium test[14] (TDT) was performed on parent-child trio genotypes via a maximum likelihood method[9] that also estimated the transmission (segregation) ratio τ. Estimation of disease penetrance by *RET rs2435357* genotype was performed as previously described[9] by using the inverse Bayes' rule on genotypes of probands with an estimated HSCR incidence of and a *rs2435357* background T allele frequency of 15 cases per 100,000 live births and 24%, and 28 cases per 100,000 live births and 47%, for those with European and Chinese descent, respectively.

Under the null hypothesis of no LD, the TDT has a χ^2 distribution with one degree of freedom (df). Under the alternative LD hypothesis, the value of the statistic can be used to quantify some properties of the genetic locus contributing to the complex disease; for a sample of n transmissions, the expected value of the test statistic is $n(2\tau - 1)$,[2] where τ is the transmission frequency; under random association, $\tau = 1/2$. Thus, τ can be estimated from parent-offspring trio genotype data.

Assume a complex disease autosomal locus with genotypes *DD*, *Dd*, and *dd*, population frequencies p^2, $2pq$, and q^2, and marginal penetrance values of f_0, f_1, and f_2, respectively; assume that d is the mutant allele and D the wild-type. Assuming contributions from other genes and environmental factors, the penetrance values are neither zero nor unity and ordered as $f_2 \geq f_1 \geq f_0$. The marginal penetrance of the alleles D and d are then $f_D = pf_0 + qf_1$ and $f_d = pf_1 + qf_2$ with total disease incidence $f = pf_D + qf_d$. Consequently, under single ascertainment of probands, the transmission ratio is:

$$\tau = f_d/(f_D + f_d)$$

$$= \gamma/(1 + \gamma),$$

where γ is the penetrance (risk) ratio $f_d/f_D \geq 1$. The precise interpretation of γ depends on whether allelic action at the locus is additive or multiplicative; in the simple multiplicative risk model, γ is also the allelic risk ratio but under an additive model it is an allele frequency-weighted sum of individual penetrance values. Note also that γ and τ can be estimated from the penetrances f_0, f_1, and f_2 by assuming a specific incidence (f) and allele frequency q only. In contrast, for case-control data, the frequency of the susceptibility allele among cases (q') is related to the risk ratio as $\gamma = q'p/qp'$ where $p' = 1 - q'$. We interpret γ as the penetrance risk ratio or the relative allelic (genetic) effect.

Estimating Haplotype Frequencies of Rare and Common RET Mutations

We denote the collection of all coding and enhancer mutations by m and e, respectively, with the wild-type states being denoted as $+$. Consequently, four "haplotypes" can be distinguished $++$, $e+$, $+m$, and em with frequencies y_1, y_2, y_3, and y_4, respectively. Thus, individuals who can have either of the four features of harboring no mutations (00), enhancer variant only (10), CDS variant only (01), and both variants (11) have frequencies (denoted by f):

$$f_{00} = y_1^2,$$

$$f_{10} = (y_1 + y_2)^2 - y_1^2,$$

$$f_{01} = (y_1 + y_3)^2 - y_1^2,$$

$$f_{11} = 1 - f_{00} - f_{10} - f_{01}.$$

Because the frequencies of the four classes 00, 10, 01, and 11 are observed, the maximum likelihood estimates of the haplotype frequencies are:

$$y_1 = \sqrt{f_{00}},$$

$$y_2 = \sqrt{(f_{00} + f_{10})} - \sqrt{f_{00}},$$

$$y_3 = \sqrt{(f_{00} + f_{01})} - \sqrt{f_{00}},$$

$$y_4 = 1 - y_{00} - y_{10} - y_{01}.$$

Enhancer Function Analysis

Luciferase Assays

Patient DNAs genotyped as homozygous for the mutant *rs2435357* (T) and wild-type (C) alleles were amplified, sequenced to verify their composition, and cloned into the Gateway pDONR221 entry vector in accordance with the manufacturer's protocol (Promega, Madison, WI). Subsequently, *RET* MCS sequences were subcloned into a SmaI site in a Gateway modified pGL3 (Promega, Madison, WI) firefly luciferase vector containing an SV40 promoter and a complete firefly luciferase open reading frame. These constructs were designated pDSma_*RET*_MCS+9.7 (mut and WT) and pDSma_RET_MCS+5.1/+9.7 (mut and WT). Control plasmids (Promega, Madison, WI) containing only the SV40 promoter and luciferase open reading frame (pDSma_promoter), or including an SV40 control enhancer with the above promoter and luciferase open reading frame (pDSma_control), served as experimental controls. Expression values were normalized against promoter only construct (pD*SMA*_promoter) expression.

The neuroblastoma cell line (Neuro-2a, ATCC# CCL-131) and the epithelial adenocarcinoma cell line (HeLa, ATCC# CCL-2) were cultured according to ATCC protocols (ATCC, Manassas, VA). Approximately 10^6 cells were cotransfected (Lipofectamine Plus, Invitrogen, Carlsbad, CA) with 0.4 μg of the appropriate pDSma firefly luciferase plasmid and 0.01 μg phRL-SV40 control Renilla luciferase plasmid. Dual Luciferase Assays (Promega, Madison, WI) were performed in accordance with the manufacturer's instructions. Luciferase activity was assayed 24 hr after transfection (Monolight 2010, Analytical Luminescence Laboratories,

CA) in triplicate and were consistent upon repetition. Statistical significance was determined with a two-tailed t test assuming unequal variances.

cDNA Transactivation Assays

pCMV_Sox10 (gift of W. Pavan, NHGRI) cDNA was cotransfected with pDMA_MCS+9.7. A total of 1 μg of cDNA was transfected per well in a 24-well plate. Cells were harvested 48 hr after transfection. Transfections and luciferase assays were conducted as described above. Deletion constructs were generated with the Stratagene QuikChange II Site-Directed Mutagenesis Kit according to manufacturer's protocol.

Chromatin Immunoprecipitation Assays

A modified protocol based on the EZ ChIP Chromatin Immunoprecipitation kit (Upstate; #17-371) was followed. N2A cells were transfected with HA-tagged Sox10 cDNA as described above. Cells were harvested and formaldehyde cross-linked 24 hr after transfection. Kit instructions and reagents were utilized for cell lysis, DNA sonication, immunoprecitation with the HA antibody, elution of protein/DNA complexes, cross-link reversal, and DNA purification. 2 μl of purified DNA was added to triplicate real-time quantitative PCR reactions performed on an ABI 7900 (Applied Biosystems, Foster City, CA) with SYBR-Green reagents according to the EZ-ChIP protocol. Negative controls included cells that were not transfected with the HA-tagged cDNA (no transfection), a sample where no antibody was added to immunoprecipitate the protein-DNA complex (no antibody), and a sample where water was the input (mock).

Results

A Polymorphic Enhancer Variant at *rs2435357* Is Associated with HSCR in the European and Chinese Populations

We have previously demonstrated, in a small sample of 126 parent-child US HSCR trios with differing segment lengths, that a common polymorphism *rs2435357* (previously termed *RET+3*:C→T, with ancestral allele C and derived mutation allele T) in a noncoding enhancer element (MCS+9.7: Multispecies Conserved Sequence located 9.7 kb from the ATG) in intron 1 of *RET* is a HSCR susceptibility allele.[9,15] The genotypes of control individuals from multiple populations across the world (Human Genome Diversity Panel) showed that the *rs2435357* minor allele had an average frequency of 24% in individuals with European ancestry and 45% in individuals with Asian ancestry, but was virtually absent (<1%) in Africa.[9] Although specific *RET* SNP associations with HSCR has been shown in selected European[16,17] and Chinese[18] samples, we conducted comprehensive mapping across the entire *RET* locus in affected individuals from Europe and China to identify significant genetic effects via a common set of SNPs.

We examined 14 SNPs, including the enhancer polymorphism *rs2435357*, spanning the 61.8 kb of *RET* to extract information on common variation in 519 European parent-child trios. The SNPs selected, their locations, genetic properties, and genotyping quality control statistics, together with the samples genotyped, are provided in Tables S1–S3 and Figure S1. As shown in Figure 1, there is highly significant evidence of association of HSCR (as a combined category) via the transmission disequilibrium test (TDT[14]) with all 14 *RET* SNPs. The peak associations are at *rs2435357* ($p = 3.9 \times 10^{-43}$) and its intron 1 neighbor *rs2435362* ($p = 1.5 \times 10^{-43}$) with which it shows high linkage disequilibrium (LD: $r^2 = 0.94$ on US and European untransmitted chromosomes). We also performed a TDT analysis on 94 trios and duos sampled in Hong Kong: we obtained an identical pattern of peak significant associations at *rs2435357* ($p = 3.9 \times 10^{-5}$) and *rs2435362* ($p = 2.4 \times 10^{-6}$) (Figure 1). In addition, we also examined *rs2435357* in 98 HSCR cases and 168 controls from Hong Kong: the T allele had a frequency of 0.43 (137/322) in controls and 0.83 (156/188) in cases ($p = 5.1 \times 10^{-19}$). The combined p value for TDT and case-control for these Chinese samples is lowest at 1.1×10^{-21} at *rs2435357*, confirming the replication of the association from TDT analysis. Overall, the T allele at *rs2435357* is common among HSCR patients in the USA (49% versus 26% in controls), in Europe (54%–62% versus 20%–30% in controls), and in Hong Kong (88% versus 47% in controls) (Table S4).

The Enhancer Variant at *rs2435357* Disrupts SOX10 Binding and RET *trans*-Activation In Vitro

Our previous functional analysis demonstrated that MCS+9.7 is a gut transcriptional enhancer that shows reduced activity with the T allele;[9] we show here that this reduced activity is due to a reduced ability of SOX10 to bind to MCS+9.7 and transactivate *RET* in T-bearing rather than C-bearing alleles. SOX10 has previously been demonstrated to directly activate *RET* transcription through interaction with PAX3 at the *RET* promoter.[19,20] We have here identified two putative SOX10 transcription factor binding sites (TFBS) within MCS+9.7 (Figure 2A). The SOX10 core binding sequence, 5′A/TA/TCAAT/A3′, has been identified and functionally verified within promoters of several SOX10-responsive genes, including *RET*.[19] In wild-type *RET* MCS+9.7, the more 5′ putative SOX10 TFBS (SOX10-BS1) exactly matches the core binding sequence[21–24] and lies 58 nucleotides 5′ to *rs2435357*. SOX10-BS1 is conserved only in primates as seen in sequence alignments (Figure 2A). The second putative TFBS (SOX10-BS2) encompasses *rs2435357*. The putative site, 5′TACA**C/T**GG3′ (bolded C is the ancestral allele), is not an exact match to the consensus binding site; however, a similar sequence (5′TACACAG3′) in the promoter of the myelin protein zero (*MPZ*) gene has been shown to bind SOX10.[25,26] The human sequence of SOX10-BS2 is conserved in the primates and pig; cow and carnivore sequences exactly match the *MPZ* SOX10 TFBS, and rat and mouse sequences are closely conserved (Figure 2A).

To determine whether SOX10 could modulate MCS+9.7 activity, we undertook a series of transactivation assays in HeLa cells; MCS+9.7 does not function as an enhancer in this epithelial cell line and it does not express SOX10.[15] HeLa cells were cotransfected with a construct

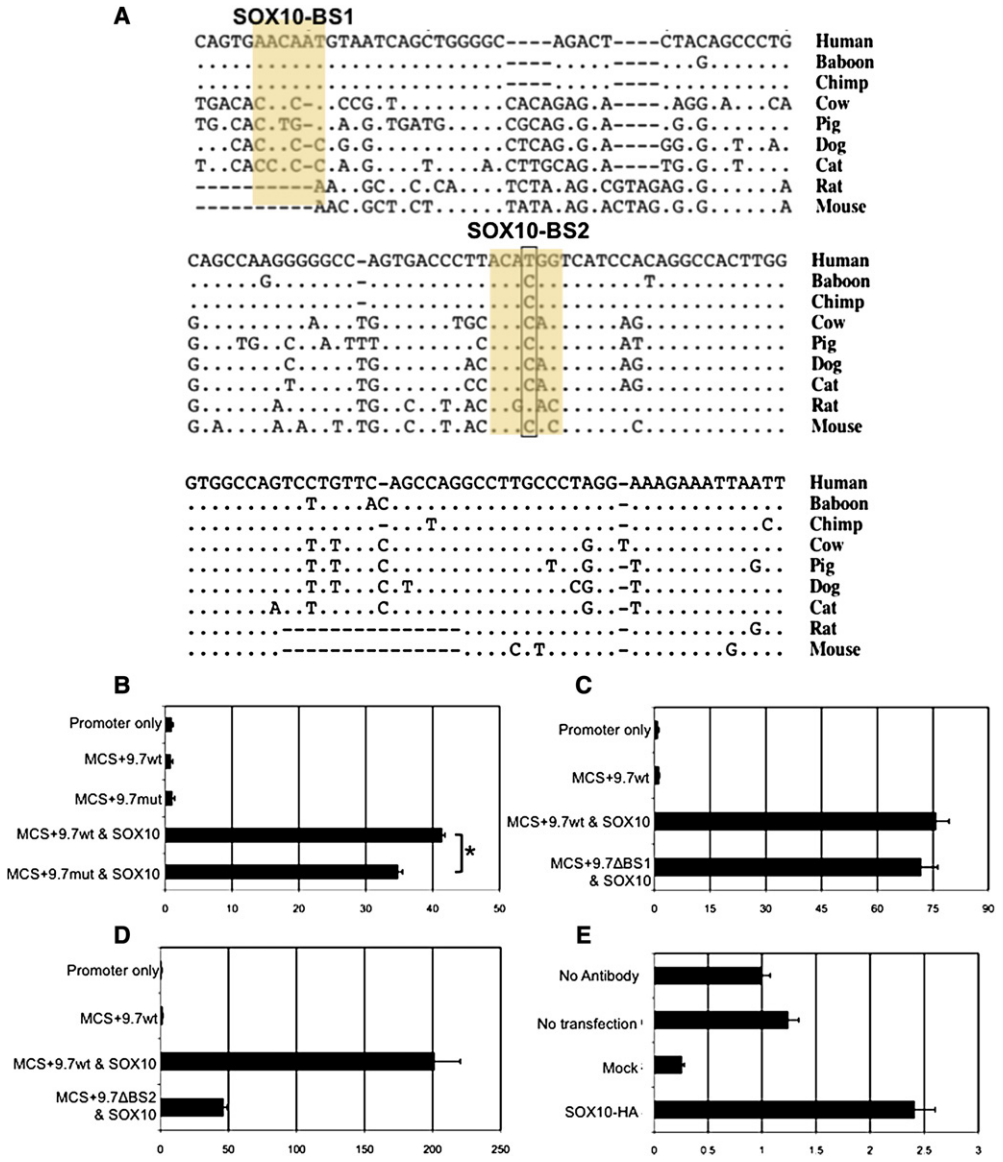

Figure 2. SOX10 Transactivation and Binding to MCS+9.7

(A) Location and conservation of putative SOX10 transcription factor binding sites (TFBS). Human genomic sequence is aligned with the orthologous sequence interval in eight different mammals via multi-PIPmaker. The putative SOX10 TFBS (SOX10-BS1 and SOX10-BS2) are highlighted in orange. The HSCR-associated variant *rs2435357* is boxed in black.

(B) Reporter expression in HeLa cells of MCS+9.7 when cotransfected with *SOX10* cDNA. Mutant (mut) and wild-type (WT) correspond to nucleotides T and C, respectively. Expression difference between mut and WT constructs is significant $p < 0.01$ by the two-tailed unpaired t test as indicated by the asterisk; error bars report standard errors of triplicate quantitative PCR reactions here and in (C)–(E).

(C and D) Deletion of MCS+9.7 putative SOX10 binding sites reduces luciferase reporter expression in vitro. Expression in HeLa cells of altered MCS+9.7 luciferase reporter constructs when cotransfected with SOX10. MCS+9.7ΔBS1 and MCS+9.7ΔBS2 correspond to the deletions of the two putative SOX10 TFBS delineated in (A). Luciferase expression values for MCS+9.7ΔBS1 are shown in (C). Luciferase expression values for MCS+9.7ΔBS2 are shown in (D).

(E) SOX10 physically associates with MCS+9.7 as demonstrated by chromatin immunoprecipitation. *x* axis is fold enrichment normalized to the no antibody control. Also included are a negative transfection control and a mock (no sample) control.

expressing SOX10 under the control of the CMV promoter (pCMV_Sox10) and the luciferase reporter vector pDSMA_MCS+9.7, containing MCS+9.7 in context of the SV40 minimal promoter. Cotransfection with pDSMA_MCS+9.7 and pCMV_Sox10 significantly increased luciferase expression from the pDSMA_MCS+9.7 vector (Figure 2B) and suggests a role for SOX10 in *RET* transcriptional modulation through MCS+9.7.

To evaluate the potential pathological significance of *rs2435357T*, we cotransfected pCMV_Sox10 with both 5′-TACA**T**GG-3′ and 5′-TACA**C**GG-3′ allele-bearing forms in pDSMA_MCS+9.7. Transfection with MCS+9.7 containing the T allele results in a statistically significant, but small, decrease of SOX10-dependent luciferase expression (Figure 2B). The magnitude of this effect is not entirely unexpected given the hypomorphic nature of the mutant

Table 1. Single Ancestral Origin of the *rs2435357 RET* Enhancer Variant

Haplotype														European		Chinese			
														Tr (%)	Un (%)	Tr (%)	Un (%)	Cases (%)	Controls (%)
T	G	A	A	A	**T**	A	G	A	C	A	T	G	T	28.2	10.2	7.2	3.1	5.1	3.3
T	G	A	A	A	**T**	A	G	A	C	G	G	C	T	24.8	11.4	77.3	42.7	70.4	36.0
T	G	C	A	C	**C**	A	A	G	C	G	G	G	C	1.4	4.7	0	0	0.5	0.0
T	G	C	A	C	**C**	G	A	G	C	A	T	G	T	1.4	4.5	3.1	8.3	2.0	6.5
T	G	C	A	C	**C**	A	A	G	G	A	T	G	T	3.8	6.9	0	0	0.0	0.6
C	G	C	A	C	**C**	A	A	G	G	A	T	G	T	8.2	18.4	1	7.3	2.0	6.3
C	G	C	A	C	**C**	A	A	G	G	G	G	C	T	1	1.8	2.1	4.2	0.0	0.3
C	G	C	A	C	**C**	A	A	G	C	A	T	G	T	4.3	2.6	0	1	0.0	0.3
T	C	C	G	C	**C**	G	A	G	C	A	T	G	T	20.2	31.2	5.2	28.1	11.2	27.7
Other Haplotypes														6.8	8.4	4.1	5.2	9.0	19.0

Haplotypes for 14 SNPs (ordered as in Figure 1 and Table S3) were separately estimated from all available trios of European and Chinese ancestry, as well as from the Chinese case-control samples. Transmitted (Tr) and untransmitted (Un) haplotypes are shown with *rs2435357* (C: wild-type and T: mutant) in bold. The first two haplotypes show an elevated frequency on Tr versus Un chromosomes and are the only haplotypes known to contain the mutant allele.

allele.[9] To determine whether either SOX10 TFBS is required for the transactivation of *RET* through MCS+9.7, we individually deleted and tested each site from pDSMA_MCS+9.7. Deletion of a 6 bp core sequence encompassing SOX10-BS1 did not result in a significant diminution of luciferase activity compared with the wild-type construct (Figure 2C). By contrast, deletion of SOX10-BS2 showed a significant 4-fold decrease in luciferase activity compared with the wild-type (Figure 2D). These data are consistent with a critical role for SOX10-BS2 binding of SOX10 in *RET* activation via MCS+9.7 and are thus consistent with its potential role in HSCR pathogenesis.

To directly confirm that SOX10 enhances reporter expression by physical association with MCS+9.7, we undertook chromatin immunoprecipitation (ChIP) analysis in a biologically relevant neuroblastoma cell line (Neuro2A). In brief, Neuro2A cells were transfected and incubated with hemagglutinin (HA)-tagged SOX10 cDNA and ChIP was performed with an antibody to the HA tag; binding to *RET* MCS+9.7 was detected by real-time quantitative PCR. Our results indicate that binding of HA-tagged SOX10 protein to MCS+9.7 was enriched compared to the no antibody control, the no transfection control, and the mock IP (Figure 2E). These data indicate that SOX10 physically interacts with MCS+9.7, consistent with the above data and a potentially important role played by SOX10 in *RET* activation via MCS+9.7. These data also demonstrate that *rs2435357,* and not *rs2435362,* is the susceptibility factor.

The Enhancer Mutation at *rs2435357* Has a Likely Single Worldwide Origin

The high frequency of the *rs2435357* T allele among European and Chinese HSCR cases does not distinguish between the possibilities of a single versus recurrent mutation for this susceptibility allele. To address this question, we constructed 14 SNP haplotypes on chromosomes transmitted (Tr) and untransmitted (Un) to probands from the European and Chinese trio genotypes and the Chinese case-control samples. These results, shown in Table 1, demonstrate that in both populations the mutant allele exist on two identical haplotypes. One disease haplotype (TGAAA**T**AGACATGT: mutation in bold; allelic differences underlined) has a frequency of 28.2% on Tr chromosomes in Europeans but is 5.1%–7.2% in the Chinese, with background (Un) rates of 10.2% and 3.1%–3.3%, respectively. We designate this as the ancestral haplotype based on the higher ATG frequency on Un chromosomes in both Europeans and Chinese. A second disease haplotype (TGAAA**T**AGACGGCT) has a similar (24.8%) frequency in Europeans (Un: 11.4%) but a considerably higher prevalence (70.4%–77.3%) in the Chinese (Un: 36.0%–42.7%). Noting the location of the allelic difference between these two disease haplotypes at the 3′ end (GGC versus ATG), and their frequency difference, the most parsimonious hypothesis is that the mutation occurred once on the TGAAA**T**AGACATGT haplotype and subsequently recombined to create the TGAAA**T**AGACGGCT haplotype.[27]

This hypothesis is buttressed by our studies of the recombination pattern within *RET*. We used the haplotype data to estimate LD (with the normalized statistic D′) for the US, European, and Chinese samples. The results show extensive LD across the 61.8 kb locus with generally greater association on Tr than on Un chromosomes, as expected, given the younger age of mutant allele-bearing chromosomes (Figure S2). Moreover, at least in the US and European samples, there is a sharp decline in LD across the region demarcated by the intron 5 SNP *rs2565205* and the intron 8 SNP *rs3026750*; this feature is absent in the Chinese samples. We estimated the population recombination rate ρ ($\rho = 4N_e r$, where N_e is the effective population

Table 2. Genetic Effect of the *RET* Enhancer Mutation at *rs2435357* by HSCR Subtypes

Feature	Transmitted Counts		τ	χ^2_1	γ	p Value
	T*	C*				
Gender						
Male	397	69	0.85	230.9	5.8	3.8×10^{-52}
Female	138	34	0.80	62.9	4.1	2.2×10^{-15}
Segment length						
Short	318	45	0.88	205.3	7.1	1.5×10^{-46}
Long	82	10	0.89	56.3	8.2	6.2×10^{-14}
Total colonic aganglionosis	50	23	0.68	10.0	2.2	1.6×10^{-3}
Unknown	85	25	0.77	32.7	3.4	1.1×10^{-8}
Familiality						
Simplex	413	63	0.87	257.4	6.6	6.3×10^{-58}
Multiplex	122	40	0.75	41.5	3.0	1.2×10^{-10}
All Probands	535	103	0.84	292.5	5.3	1.4×10^{-65}

Transmission disequilibrium tests of all available trios classified by gender, segment length of aganglionosis, and familiality. The numbers of transmitted C (wild-type) and T (mutant) alleles from informative parents, the estimated transmission (segregation) ratio (τ), the allelic risk ratio (γ), and the significance of association (p value) are provided.

size and *r* the recombination rate in Morgans/10 kb) that is expected to be 4. As observed in Figure S3, the *RET* locus is a recombination cold spot (relative to the genome) but almost all the recombination that occurs at this locus does so in the interval between *rs2565205* and *rs3026750* at 2–3 times the genomic average. This feature can explain the origin of the new disease haplotype TGAAA**T**AG ACGGCT by recombination within this hotspot. We hypothesize that the recombinant disease haplotype occurred prior to the differentiation of these two populations based on its high background frequency in both Europeans and Chinese and it is the putative recombinant TGAAA**T**AGACGGCT haplotype that has expanded in Asia.

Although parsimony suggests that the two disease haplotypes have a common origin, one ancestral and one recombinant, it is also possible that both disease haplotypes arose independently in Europe and both migrated to Asia but TGAAA**T**AGACGGCT rose to a greater frequency than TGAAA**T**AGACATGT by genetic drift or selection.

Penetrance at *rs2435357* Varies with Hirschsprung Disease Subtypes

The *RET* disease haplotypes shows an approximately 2-fold difference on Tr versus Un chromosomes with the higher background (Un) frequency leading to a 2-fold higher incidence on the T chromosomes in the Chinese as compared to the US/European samples, and a concomitant 2-fold increase in HSCR disease incidence (28 and 15 cases per 100,000 live births, respectively). The lower frequency of *rs2435357T* in Africa also predicts the lower disease incidence in Africans.[9] Thus, irrespective of the contributions of other genes to HSCR, the *rs2435357* susceptibility allele frequency difference is one explanation of the disease prevalence difference across populations.

We next tested whether the genetic susceptibility at *rs243537* is similar across three features known to be associated with overall HSCR risk, namely, gender, familiality (simplex versus multiplex), and segment length variation (S-HSCR, L-HSCR, and TCA).[3] The three features signify two different types of effects: gender is a part of the "genetic background" and can thus *modify* the mutation effect whereas familiality and segment length are outcomes probably *modified* by the mutation. We first performed the TDT across all 568 European, US, and Chinese trios and 124 duos with complete data on all three subtypes. The results in Table 2 imply three major conclusions. First, *rs2435357* has an effect on *all* forms and subtypes of HSCR. Second, within each subtype there are biological differences between classes. Thus, there are moderate but nonsignificant differences in transmission frequency by gender ($p = 0.13$) but large significant differences in transmission frequency by segment length ($p = 7.6 \times 10^{-5}$, ignoring unknown segment length) and familiality ($p = 6.2 \times 10^{-4}$). Third, estimation of the relative allelic or genetic effect γ shows that the mutational effect is *directly* related to class frequency: the greater the abundance of the class, the greater the genetic effect of *rs2435357*. Thus, for the enhancer susceptibility allele, the genetic effects are contrary to the Carter prediction for all three features. Under mutation-selection balance, the mutational effect on a phenotype is *inversely* proportional to its frequency. Thus, the above results may be a novel feature of multifactorial disease common variants.

To confirm these findings, we directly estimated the disease genotype effects for both the European and

Table 3. Prevalence of *rs2435357 RET* Genotypes by HSCR Subtypes

	Chinese		European						
Genotype	Male	Female	Male	Female	Short	Long	TCA	Simplex	Multiplex
CC	8	2	10	3	13	2	3	9	5
CT	17	6	**22**	6	**32**	4	5	**20**	7
TT	**154**	**35**	**178**	**33**	**285**	**35**	**20**	**175**	**39**

The prevalence (population penetrance) is shown as the number of cases per 100,000 live births, with both trios and case-control samples, with values in excess of the background rate in bold. We assumed a population background rate of 15 cases and 28 cases per 100,000 live births in European ancestry and Chinese populations, respectively. The Chinese samples were not used for segment length and familiality analyses because the overwhelming majority of samples were short segment (173/192 cases) and simplex (190/192 cases).

Chinese samples.[9] We estimated the population probability of being affected given each genotype (population penetrance or prevalence), by segment length, gender, and familiality for the European ancestry sample, but only by gender for the Chinese sample because there was negligible variation in the other features (173 S-HSCR, 6 L-HSCR, 13 TCA; 190 simplex, 2 multiplex) (Table 3). These results demonstrate three major features not evident from Table 2. The estimated penetrance/prevalence values are shown per 100,000 live births in comparison to the population average of 15 per 100,000 European live births and 28 per 100,000 Chinese live births. (Observed counts of affecteds by these categories show significant differences in risk for all three risk factors in the European sample but not for the Chinese sample where the sample size is smaller [Table S5].) First, the estimated population penetrance/prevalence values, within each category, vary by two orders of magnitude with some genotypes elevating risk 20- to 40-fold above the average. The greatest risk is for TT homozygotes across all subclasses (male versus female, short versus long versus TCA, multiplex versus simplex), but the only elevated risks for CT heterozygotes are in the commonest male/S-HSCR/simplex subclass. In other words, the effect of *rs2435357T* is largely recessive but with some effect in heterozygotes only in the most common categories. Second, the penetrance or prevalence or risk, within each disease subtype, is amplified 3- to 4-fold for the most frequent class (male, S-HSCR, simplex) in the population. This relationship is unexpected for Mendelian rare mutations and has not been previously observed for a multifactorial phenotype. Third, although the data are limited, penetrance of *rs2435357T* in the Chinese is similar to that in the Europeans. Thus, the allele frequency of the susceptibility allele is higher in the Chinese but the genetic effect is near identical to that in the European sample. Consequently, the HSCR prevalence difference in these two populations is highly correlated with the allele frequency or haplotype difference.

These results suggest that the *rs2435357* genetic effect is not constant but significantly modifiable by gender and other factors that regulate segment length and familiality although additional, larger, and independent studies will be needed to state this definitively. All of these subtypes are themselves dependent on additional genes and developmental processes, so our result for HSCR is a clear example of the quantitative effects of genetic background on mutant allele/disease expression.

Rare Coding and Polymorphic RET Mutations Have Complementary Effects in HSCR

More than 100 *RET* mutations, generally rare, missense changes segregating with disease in families and absent in unaffected controls,[8] unlike the common enhancer polymorphism,[9] are known in familial and syndromic HSCR patients. To study the contribution of all variants across the allelic spectrum and for a diverse phenotypic set of probands, we performed a comprehensive variation (mutation) detection study. The distributions of these probands according to whether they harbor a rare coding (CDS) mutation only, the common enhancer variant (CT or TT) only, neither, or both are presented in Figure 3. The genotyping and sequencing was complete for 577 probands and, among these, the numbers were 73 (13%), 382 (66%), 45 (8%), and 77 (13%) cases with no mutations, enhancer variant only, CDS variant only, and both disease allelic types, respectively.

The frequency distributions of these four mutational classes can be classified according to the three disease subtypes we examined before (Table S6). The first noticeable feature is that 13% (range: 9%–18%) of probands have no identifiable *RET* mutation, with the general trend that *total* mutation frequency is *lower* in the less common and more heritable classes (females, TCA). Our mutation detection protocols were comprehensive but we cannot ignore the possibility of additional undetected *RET* mutations, particularly additional regulatory variants. In any case, the identifiable fraction of *RET* mutations is 87% and shows its major genetic effect in HSCR. The *RET*-negative cases may have mutations at other genes[10] or could have HSCR from environmental, stochastic, or epigenetic effects that reduce *RET* function. The second feature of these data is the preponderance of the enhancer mutation in all categories and its higher frequency in the commoner risk classes (male, S-HSCR, simplex): the enhancer mutation is directly proportional to the risk factor class frequency but inversely proportional to its penetrance and thus recurrence risk. These analyses suggest that CDS and enhancer *RET* mutations have quite distinct, complementary, and

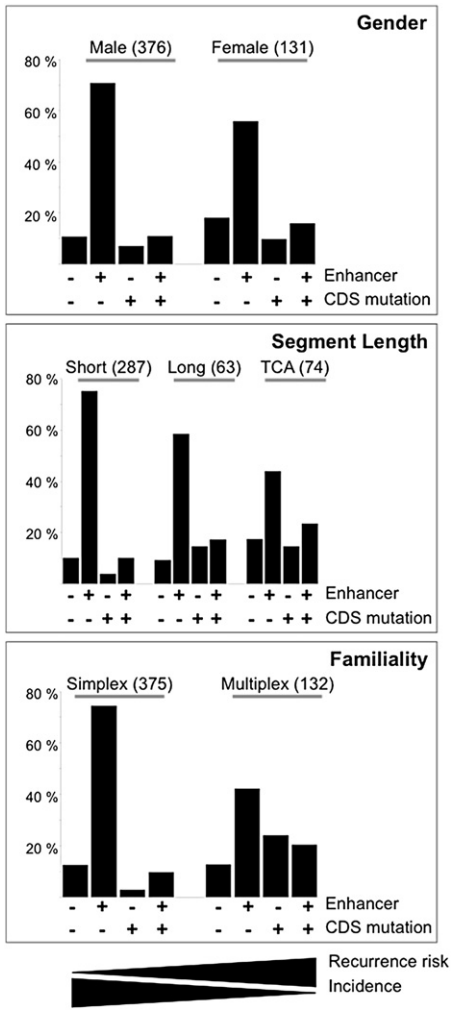

Figure 3. Proportion of Patients with Rare and Common Mutations Classified by Gender, Segment Length, and Familiality
Frequencies of *RET* coding sequence (CDS) and enhancer mutation distribution in probands within HSCR subtypes. In each case, the percent of patients with no identified mutation, enhancer mutation only, CDS mutation only, and both enhancer and CDS mutations are shown (plus and minus signs indicate presence/absence) together with the respective sample size in each subgroup. A qualitative scale, demonstrating the direction of increase (decrease) in recurrence risk (incidence) of the categories from left to right, is shown.

even opposite, genetic properties with regard to frequency, genetic effect, recurrence risk, and frequency of a disease subtype class. Indeed, the frequency of the *rs2435357T* allele is 27%, 52%, 60%, and 66% among CDS mutation-positive multiplex, CDS mutation-negative multiplex, CDS mutation-positive simplex, and CDS mutation-negative simplex probands, respectively. Modeling of the haplotype frequencies of these two types of mutations, based on the observation in cases estimates the frequencies of haplotypes containing no mutations, enhancer mutation only, CDS mutation only, and both mutations as 0.356, 0.532, 0.096, and 0.015, respectively. This suggests that almost all enhancer alleles exist with CDS mutations in *trans* and not *cis*. Moreover, the frequency of the enhancer variant among those without a CDS mutation is $0.532/(0.356 + 0.532) = 0.60$ whereas the frequency among those with a CDS mutation is $0.015/(0.096 + 0.015) = 0.14$ and smaller than the background frequency of 0.24. This negative association between the occurrence of coding mutations and the enhancer variant is highly significant ($\chi^2_1 = 26.84$, $p = 2.2 \times 10^{-7}$). This suggests that association studies of *RET* would be much less powerful in multiplex (usually CDS mutation-positive) as compared to simplex (CDS mutation-negative) cases.

The mutation distribution in Figure 3 also shows that only a minority of patients have *RET* CDS mutations *alone*. Moreover, the frequency of those with both a cds and the enhancer mutations is greater than those with a single CDS mutation in six of the seven subclasses; they are about equal in multiplex families where the enhancer has no additional effect. Among all cases, the numbers of probands who harbor the enhancer variant is 84% (382/455) among those without a CDS mutation and 63% (77/102) among those with a CDS mutation, a difference that shows marginal significance ($p = 0.04$). Thus, the enhancer mutation also exerts an effect in patients with a coding mutation but the effect is attenuated relative to the primary deleterious effect of a structural *RET* mutation. This interaction between rare and common variants within the same gene is analogous to that observed at *CFH* (MIM 134370) in age-related macular degeneration (MIM 610698)[28] and suggests a recessive effect. Additional types of mutations in *RET* undoubtedly exist. For example, we have used the genotype data to search for trios in which the failure of Mendelian segregation of *RET* SNPs is consistent with de novo deletions. Figure 4 shows nine families with evidence of deletions supported by multiple markers and occurring in families with diverse phenotypes. Of these, six cases arose in the maternal germline and one in the paternal germline (two were indeterminate). These data support the existence of deletions in a small fraction (~1%) of all cases and are unlikely to compromise our interpretations greatly.

Discussion

Despite the great interest in complex disease genetics and the success of genome-wide association studies (GWAS) for gene mapping, we still have a meager understanding of the specific genes, the allelic spectrum of susceptibility variants, and the genetic mechanisms that underlie these phenotypes.[29] The basic inheritance model for complex diseases is multifactorial inheritance,[1,2] and the large numbers of variants of small additive effects across alleles and loci identified in GWAS supports this model. However, this is a statistical model adequate for explaining some features of inheritance, such as the contribution of specific genes to heritability or the correlations between or recurrence risk to relatives, but not a molecular model of gene action that is necessary for understanding the genetic

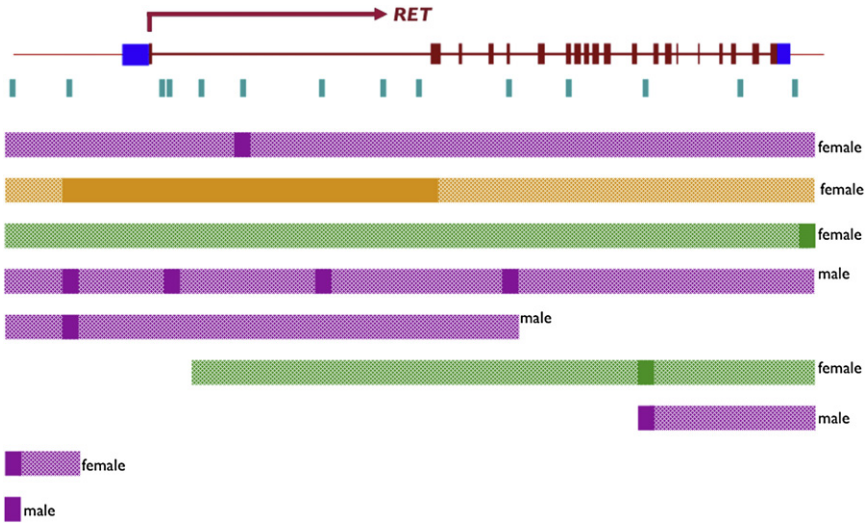

Figure 4. Mendelian Errors Compatible with Gene Deletion
Nine affected individuals with patterns of Mendelian inconsistencies suggestive of deletion are shown (each row represents a proband). *RET* exons (brown) and genotyped SNPs (blue) are shown across the top, in scale. For each proband, the longest *RET* region consistent with a deletion is represented: SNPs that are Mendelian inconsistent between parents and children are shown in solid color; SNPs that are Mendelian consistent but for which the child is homozygous are stippled in color. The entire genomic region is shown for illustrative purposes but only the SNPs were examined. Putative deletions occurring in the maternal and paternal germlines, or those with uncertain origin, are colored in purple, orange, and green, respectively. All deletions, but two, involve multiple exons.

properties of the component genes. HSCR, and the studies described here, remains a compelling "model disease" for elucidating the latter molecular properties and connecting them to the former statistical properties.

In this study, we have replicated our prior discovery of a common, noncoding enhancer mutation (*rs2435357*) associated with HSCR in US samples to a larger European sample, discovered a similar association in Chinese samples, identified the molecular basis of the susceptibility, identified interactions between numerous phenotypic subclasses and *RET* variants, and quantified the relative contributions of rare and common *RET* variants to HSCR. We have specifically demonstrated that (1) the functional susceptibility at the *RET* enhancer MCS+9.7 arises from the failure of SOX10 to activate *RET* transcription; (2) the susceptibility T allele probably arose as a single mutation after humans had left Africa and expanded in both range and frequency toward Asia; (3) the European (24%) and Chinese (47%) T allele frequency is significantly elevated in HSCR cases and sufficient to explain the prevalence difference between these two populations; (4) *rs2435357* is a genetic susceptibility factor in all phenotypic forms of HSCR but with quantitatively different effects; (5) the penetrance of the T allele is higher in individuals who have the more *common* subtypes (male, S-HSCR, and simplex cases) with lower recurrence risk in relatives, whereas coding sequence mutations are more abundant in the less frequent subtypes (female, L-HSCR/TCA, and multiplex cases) with higher recurrence risk in relatives (this negative association is highly significant at $p = 2.2 \times 10^{-7}$) ; and (6) *rs2435357* also has a genetic effect in those harboring rare coding mutations. Additional *RET* coding sequence polymorphisms exist,[11] but assessing their specific contribution, beyond their correlated effect resulting from LD with *rs2435357*, will require functional assays for each. Nevertheless, the existing features satisfy many of the properties of the multifactorial model at the phenotypic level (recurrence risk inversely related to incidence, increased with greater clinical severity, and greater from the less frequently affected class) for the rare coding mutations but are violated by the common noncoding variant at MCS+9.7. Our analyses show that different subclasses of the phenotype have different mutational profiles and, once recognized, improve genotype-phenotype correlations and enhance the proportion of the risk that the *RET* gene explains. Nevertheless, there must be additional variation unlinked to *RET* that also impacts the observed differences in HSCR risk by gender, familiality, and segment length.[8,10]

Genetic counseling for recurrence of multifactorial diseases involve quoting a small, fixed risk of 3% to all consultands, but our *RET* results strongly suggest that *both* phenotypic features and family history will need to be accounted for. This is clearly possible for the total phenotype given family history, segment length, gender,[4] rare coding mutations (that are incompletely penetrant dominant and recessives), and syndromic cases (based on the molecular karyotype). Incorporating the effects of common variants such as *rs2435357* that have large relative risks is possible (Table 3) but will make little *quantitative* change to the *absolute* risks we have published before[4] because this is only one of several genes, each of small effect. Consequently, one needs well-designed outcome studies to assess how such risks may be understood by consultands and how it might change their attitudes.

A second significant finding from our analyses is the "inconstant and modifiable"[30] nature of the common enhancer variant. Unlike Mendelian HSCR variants, the enhancer variant does not have fixed genetic effects on HSCR but are contextual and depend on genetic and other host factors and disease subtypes (gender, segment length, familiality). We do not know the specific factors that lead to these properties but they probably depend on both random (environmental effects, stochastic factors) and nonrandom (interaction effects through other genes, such as those on chromosome 21[31] or epigenetic modification

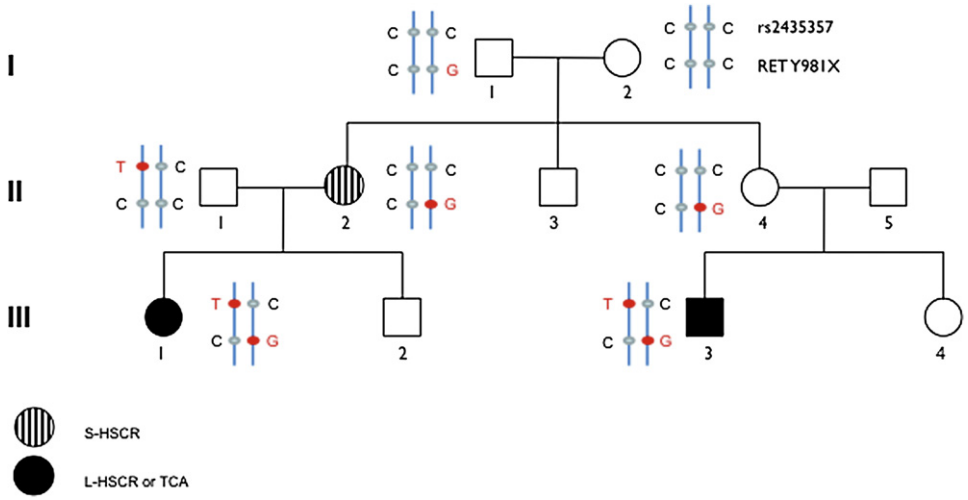

Figure 5. Segregation of Both a Rare and Common Mutation in a HSCR Pedigree
A family segregating S-HSCR, L-HSCR, and TCA with two *RET* mutations, a rare nonsense mutation (Y981X) and the common enhancer variant (*rs2435357*), is shown with genotypes of relevant individuals only. The Y981X mutation is inherited by all affected individuals from the paternal grandfather (I-I) and is present in two unaffected individuals (I-I and II-4). However, the severe forms of L-HSCR and TCA in the grandchildren (III-I and III-3) also harbor the common enhancer variant inherited instead from the individuals marrying into this family (II-I and II-5). This represents a case of allelic penetrance modification of a nonsense mutation.

of *RET*) factors. Nevertheless, our study shows that which patients we choose for a genetic study may strongly determine the *types* of genetic variants we uncover. Thus, GWAS of multiplex cases biased with female, L-HSCR/TCA patients may have low power for detecting common variants whereas studies of simplex cases biased with male, S-HSCR patients will have inadequate power for detecting rare variants.[9,32,33] These differences are usually attributed to *frequency* of the disease variant but at its basis are strongly affected by the phenotype *subtypes* we choose to study. Thus, for molecular understanding of a complex disease, a broad gamut of phenotypes covering familial and isolated cases needs to be examined to assess the complete contributions of genes and variants to heritability.

The debate on the frequency of disease alleles in common disorders often treats common and rare alleles as two distinct classes. However, disease outcome and penetrance depends on the genotype and can include both types of alleles. Consider the HSCR pedigree shown in Figure 5 that segregates HSCR in an incompletely penetrant autosomal-dominant manner with variable expressivity (II-2 has S-HSCR, III-1 has L-HSCR, and III-3 has TCA) but, at the sequence level, has been demonstrated to segregate both a rare nonsense mutation (Y981X) and the enhancer susceptibility allele at *rs2435357*. The severely affected individuals harbor both mutations whereas carriers of Y981X only are either mildly affected (II-2) or unaffected (I-1, II-4). Here, *rs2435357T* (inherited from the two males who married into the family, II-1 and II-5) modifies the penetrance of the Y981X allele in *trans* (interallelic interaction) and suggests that *rs2435357T* in *cis* should have no additional effect on the null allele. This hypothesis needs to be tested in a larger sample of cases with *cis*- versus *trans*-haplotypes harboring a coding mutation. These data also suggest that the *RET* genetic effect in the severe cases is more recessive than dominant and is the likely cause of the small but perceptible effect of *rs2435357* in patients with rare coding mutations.

The two classes of *RET* mutations affect distinct genetic aspects of RET: rare changes that reduce RET signaling (loss of function in any of the several functional domains within RET) and a common variant that reduces *RET* transcription. The former variants produce a *mutant* protein with compromised function whereas the latter variant produces reduced amounts of a *wild-type* protein. Specifically, how these two different biological alterations, sometimes in the same patient (Figure 5), lead to HSCR is unknown. Given that the majority of susceptibility alleles mapped by GWAS appear to be similar to the *rs2435357* rather than the coding variants, it is possible that their genetic properties may be different from those we have classically come to expect from studies of rare Mendelian phenotypes. The complementary genetic effects of these two types of mutations is a key finding of our studies and, hopefully, is a model for other complex diseases because it is known in other cases such as *LDLR* (MIM 606945) variants and LDL cholesterol. In other words, many diseases may be disorders of protein quantity rather than quality.[34,35]

Despite the findings of this study and their implications, we require additional independent replications of these findings. Such new studies could take additional notice of randomizing patient ascertainment, particularly with respect to subtypes and improved diagnosis utilizing clinical-radiological-pathological data and from additional populations. In addition, higher-resolution genetic analysis of *RET* together with marker data across the genome

could add greater support to the results. With multiple common and rare variants, we could start to address how molecular genotypes can modify the recurrence risk in individual cases.

Recent HSCR studies provide some new insights into the genetic mechanisms involved. The early discovery of *RET* deletions in HSCR patients[36] suggested that RET haploinsufficiency was the genetic cause of HSCR; however, in the mouse, *RET* null mutations[37] do not display aganglionosis except in homozygotes, suggesting complete loss of function as the primary cause. Although originally reconciled as human-mouse developmental differences, four types of information now suggest that these are not real differences: (1) we have previously shown that compound *Ret* null heterozygote-*Ednrb* hypomorphic homozygote genotypes in the mouse display full penetrance and phenotypic features similar to humans (variation in segment length and sex ratio alteration);[38] (2) all *RET* mutations in HSCR, be they in the coding sequence leading to a mutant protein[39] that fails to appropriately transduce the signal from its ligand GDNF or from reduced transcriptional activity of *RET* affected through the binding of SOX10 at the *MCS*+9.7 enhancer,[15,40] are partial or complete genetic loss-of-function mutations; (3) the effects of the *rs2435357T* enhancer mutation is 5- to 10-fold greater in TT homozygotes than CT heterozygotes and the enhancer mutation is at elevated frequency even in cases with a coding mutation; and (4) interacting proteins, such as overexpressed SOD1, can reduce RET dimerization[41] and consequent loss of function. These results jointly suggest that HSCR may result only when >50% of *RET* function has been lost irrespective of whether this loss arises from structural or regulatory mutations within the gene or from interactions with other susceptibility genes.[8,31] The studies presented here show that 87% of the *RET* alleles in patients have mutations, which is more than what a heterozygote effect demands and close to recessive genetic action. We suspect, but cannot yet prove, that the remaining 13% also harbor alleles with compromised function whether they are compromised by the actions of interacting genes or epigenetic markings or random environmental factors. In other words, *RET* mutations may be necessary for HSCR pathogenesis so that *RET* is a rate-limiting step in enteric ganglion cell development.

We hypothesize that HSCR is a cellular recessive, analogous (but not genetically identical) to the childhood cancer retinoblastoma,[42] and requires two "hits" to *RET*. There are many similarities in patterns between these two disorders with respect to transmission (familiality versus laterality), expressivity and severity (segment length versus number of tumor foci), and inheritance (syndromic/isolated cases and Mendelian versus complex inheritance). Such a model can also potentially account for why the penetrance of mutations is not complete and why additional genes may be required. For retinoblastoma, additional mutations, all somatically acquired, select for growth of the tumor whereas in HSCR, additional mutations (genetic evidence suggesting that many of these are inherited) are those that fail to rescue the early *RET* loss of function. We can begin to appreciate why individuals with the wild-type genotypes at *RET* can be affected with HSCR. Some of these could simply be the effect of other genes or could arise from nonheritable variation. Cell biology and modeling studies have shown the crucial importance of genetic control of precursor neural crest cell number during development.[43] We believe that perturbing this cell number, by any mechanism, is essential to HSCR and a major explanation of the variation in penetrance by *RET* mutations. Thus, a search for mutations that control precursor neural crest cell number may be key to understanding HSCR.

Supplemental Data

Supplemental Information includes three figures and six tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We wish to thank the numerous patients, their families, and the referring physicians across the world that have participated in studies of Hirschsprung disease in our laboratories over the past 20 years, and the numerous laboratory members who have contributed to discussion of these results. This study is a contribution from The International Hirschsprung Disease Consortium. Grant support is as follows: (USA) National Institutes of Health, HD28088 to A.C. and GM71648 to A.S.M.; and the "Holes for Hirschsprung" fundraiser to A.C.; (Italy) Italian Telethon (GGP04257) and European "E-Rare" program—Istituto Superiore di Sanità (Italy) to I.C.; (France) ANR (Eranet and Maladies Rares Grants) and FRM (Fondation pour la Recherche Médicale) to S.L.; (Hong Kong) HKU 765407 from the Hong Kong Research Grant Council to M.G.-B. and P.K.H.T.; (Netherlands) NWO (no. 901-04-225) Bernoulle Foundation and Ubbo Emmius Foundation to R.M.W.H.; and (Spain) Fondo de Investigación Sanitaria, Spain (PI070080), European E-Rare program (PI071315), and Consejeria de Educación Ciencia Y Empresa de la Junta de Andalucía (CTS2590) to S.B. Author contributions are as follows. (USA) A.C. designed the study and ascertained families; E.S.E. and K.W. performed all genotyping; L.H. and E.S.E. performed DNA sequence analysis; E.A.G. and A.S.M. designed and performed the functional enhancer biological analyses; C.K., E.S.E., and A.C. conducted statistical analysis of the data; and E.S.E., A.S.M., and A.C. participated in data interpretation and writing the manuscript. We gratefully acknowledge the work of Erick Kaufmann, Jennifer (Scott) Bubb, Maura Kenton, Julie Albertus, and Courtney Nichols in family ascertainment. (Italy) F.L. and I.C. ascertained families; F.L. conducted statistical analysis of data; and F.L., P.G., and I.C. participated in data interpretation and editing the manuscript. (France) J.A., A.P., and L.d.P. performed DNA sequencing; A.-S.J. conducted statistical analyses of data; and J.A. and S.L. participated in the study design, data interpretation, and ascertainment of families. (Hong Kong) M.G.-B. and X.M. performed DNA sequencing and genotyping of the sporadic Chinese HSCR patients and control samples; M.G.-B. conducted statistical

analysis of data, and together with P.K.H.T., data interpretation; and P.K.H.T. ascertained all Chinese HSCR patients. (Netherlands) G.B. performed all genotyping; R.M.W.H. participated in data interpretation and manuscript editing; and J.B.G.M.V. was involved in ascertaining families. (Spain) S.B. and G.A. ascertained Spanish HSCR families; R.M.F. performed DNA sequencing for mutation detection; and R.M.F., G.A., and S.B. participated in manuscript editing.

Received: November 30, 2009
Revised: May 22, 2010
Accepted: June 11, 2010
Published online: July 1, 2010

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/SNP
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

References

1. Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. *52*, 399–433.
2. Carter, C.O. (1969). Genetics of common disorders. Br. Med. Bull. *25*, 52–57.
3. Chakravarti, A., and Lyonnet, S. (2001). Hirschsprung Disease. In The Metabolic and Molecular Bases of Inherited Disease, 8th edition, C.R. Scriver, A.L. Beaudet, D. Valle, W.S. Sly, B. Childs, K. Kinzler, and B. Vogelstein, eds. (NY: McGraw-Hill), pp. 6231–6255.
4. Badner, J.A., Sieber, W.K., Garver, K.L., and Chakravarti, A. (1990). A genetic study of Hirschsprung disease. Am. J. Hum. Genet. *46*, 568–580.
5. Lander, E.S. (1996). The new genomics: Global views of biology. Science *274*, 536–539.
6. Pritchard, J.K. (2001). Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. *69*, 124–137.
7. Bodian, M., and Carter, C. (1963). A family study of Hirschsprung disease. Ann. Hum. Genet. *26*, 261–277.
8. Amiel, J., Sproat-Emison, E., Garcia-Barcelo, M., Lantieri, F., Burzynski, G., Borrego, S., Pelet, A., Arnold, S., Miao, X., Griseri, P., et al; Hirschsprung Disease Consortium. (2008). Hirschsprung disease, associated syndromes and genetics: A review. J. Med. Genet. *45*, 1–14.
9. Emison, E.S., McCallion, A.S., Kashuk, C.S., Bush, R.T., Grice, E.A., Lin, S., Portnoy, M.E., Cutler, D.J., Green, E.D., and Chakravarti, A. (2005). A common sex-dependent mutation in a *RET* enhancer underlies Hirschsprung disease risk. Nature *434*, 857–863.
10. Gabriel, S.B., Salomon, R., Pelet, A., Angrist, M., Amiel, J., Fornage, M., Attié-Bitach, T., Olson, J.M., Hofstra, R., Buys, C., et al. (2002). Segregation at three loci explains familial and population risk in Hirschsprung disease. Nat. Genet. *31*, 89–93.
11. Kashuk, C.S., Stone, E.A., Grice, E.A., Portnoy, M.E., Green, E.D., Sidow, A., Chakravarti, A., and McCallion, A.S. (2005). Phenotype-genotype correlation in Hirschsprung disease is illuminated by comparative analysis of the RET protein sequence. Proc. Natl. Acad. Sci. USA *102*, 8949–8954.
12. Stephens, M., Smith, N.J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. *68*, 978–989.
13. Abecasis, G.R., and Cookson, W.O. (2000). GOLD—Graphical overview of linkage disequilibrium. Bioinformatics *16*, 182–183.
14. Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. *52*, 506–516.
15. Grice, E.A., Rochelle, E.S., Green, E.D., Chakravarti, A., and McCallion, A.S. (2005). Evaluation of the *RET* regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. Hum. Mol. Genet. *14*, 3837–3845.
16. Fitze, G., Schreiber, M., Kuhlisch, E., Schackert, H.K., and Roesner, D. (1999). Association of RET protooncogene codon 45 polymorphism with Hirschsprung disease. Am. J. Hum. Genet. *65*, 1469–1473.
17. Borrego, S., Sáez, M.E., Ruiz, A., Gimm, O., López-Alonso, M., Antiñolo, G., and Eng, C. (1999). Specific polymorphisms in the RET proto-oncogene are over-represented in patients with Hirschsprung disease and may represent loci modifying phenotypic expression. J. Med. Genet. *36*, 771–774.
18. Garcia-Barceló, M.M., Sham, M.H., Lui, V.C., Chen, B.L., Song, Y.Q., Lee, W.S., Yung, S.K., Romeo, G., and Tam, P.K. (2003). Chinese patients with sporadic Hirschsprung's disease are predominantly represented by a single RET haplotype. J. Med. Genet. *40*, e122.
19. Lang, D., Chen, F., Milewski, R., Li, J., Lu, M.M., and Epstein, J.A. (2000). Pax3 is required for enteric ganglia formation and functions with Sox10 to modulate expression of c-ret. J. Clin. Invest. *106*, 963–971.
20. Lang, D., and Epstein, J.A. (2003). Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. Hum. Mol. Genet. *12*, 937–945.
21. Denny, P., Swift, S., Connor, F., and Ashworth, A. (1992). An SRY-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein. EMBO J. *11*, 3705–3712.
22. Harley, V.R., Lovell-Badge, R., and Goodfellow, P.N. (1994). Definition of a consensus DNA binding site for SRY. Nucleic Acids Res. *22*, 1500–1501.
23. Kanai, Y., Kanai-Azuma, M., Noce, T., Saido, T.C., Shiroishi, T., Hayashi, Y., and Yazaki, K. (1996). Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region, and their differential expression in mouse spermatogenesis. J. Cell Biol. *133*, 667–681.
24. Mertin, S., McDowall, S.G., and Harley, V.R. (1999). The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic Acids Res. *27*, 1359–1364.
25. Peirano, R.I., Goerich, D.E., Riethmacher, D., and Wegner, M. (2000). Protein zero gene expression is regulated by the glial transcription factor Sox10. Mol. Cell. Biol. *20*, 3198–3209.
26. Peirano, R.I., and Wegner, M. (2000). The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. Nucleic Acids Res. *28*, 3047–3055.
27. Garcia-Barcelo, M.M., Ganster, R.W., Lui, V.C., Leon, T.Y., So, M.T., Lau, A.M., Fu, M., Sham, M.H., Knight, J., Zannini, M.S., et al. (2005). TTF-1 and RET promoter SNPs: regulation

of RET transcription in Hirschsprung's disease. Hum. Mol. Genet. *14*, 191–204.
28. Boon, C.J., Klevering, B.J., Hoyng, C.B., Zonneveld-Vrieling, M.N., Nabuurs, S.B., Blokland, E., Cremers, F.P., and den Hollander, A.I. (2008). Basal laminar drusen caused by compound heterozygous variants in the CFH gene. Am. J. Hum. Genet. *82*, 516–523.
29. Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science *322*, 881–888.
30. Altenburg, E., and Muller, H.J. (1920). The genetic basis of truncate wing, an inconstant and modifiable character in *Drosophila*. Genetics *5*, 1–59.
31. Arnold, S., Pelet, A., Amiel, J., Borrego, S., Hofstra, R., Tam, P., Ceccherini, I., Lyonnet, S., Sherman, S., and Chakravarti, A. (2009). Interaction between a chromosome 10 RET enhancer and chromosome 21 in the Down syndrome-Hirschsprung disease association. Hum. Mutat. *30*, 771–775.
32. Romeo, G., Ronchetto, P., Luo, Y., Barone, V., Seri, M., Ceccherini, I., Pasini, B., Bocciardi, R., Lerone, M., Kääriäinen, H., et al. (1994). Point mutations affecting the tyrosine kinase domain of the RET proto-oncogene in Hirschsprung's disease. Nature *367*, 377–378.
33. Edery, P., Lyonnet, S., Mulligan, L.M., Pelet, A., Dow, E., Abel, L., Holder, S., Nihoul-Fékété, C., Ponder, B.A., and Munnich, A. (1994). Mutations of the RET proto-oncogene in Hirschsprung's disease. Nature *367*, 378–380.
34. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA *106*, 9362–9367.
35. Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. *10*, 451–481.
36. Puliti, A., Covone, A.E., Bicocchi, M.P., Bolino, A., Lerone, M., Martucciello, G., Jasonni, V., and Romeo, G. (1993). Deleted and normal chromosome 10 homologs from a patient with Hirschsprung disease isolated in two cell hybrids through enrichment by immunomagnetic selection. Cytogenet. Cell Genet. *63*, 102–106.
37. Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F., and Pachnis, V. (1994). Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. Nature *367*, 380–383.
38. McCallion, A.S., Stames, E., Conlon, R.A., and Chakravarti, A. (2003). Phenotype variation in two-locus mouse models of Hirschsprung disease: Tissue-specific interaction between Ret and Ednrb. Proc. Natl. Acad. Sci. USA *100*, 1826–1831.
39. Pasini, B., Borrello, M.G., Greco, A., Bongarzone, I., Luo, Y., Mondellini, P., Alberti, L., Miranda, C., Arighi, E., Bocciardi, R., et al. (1995). Loss of function effect of RET mutations causing Hirschsprung disease. Nat. Genet. *10*, 35–40.
40. Griseri, P., Bachetti, T., Puppo, F., Lantieri, F., Ravazzolo, R., Devoto, M., and Ceccherini, I. (2005). A common haplotype at the 5′ end of the RET proto-oncogene, overrepresented in Hirschsprung patients, is associated with reduced gene expression. Hum. Mutat. *25*, 189–195.
41. Kato, M., Iwashita, T., Takeda, K., Akhand, A.A., Liu, W., Yoshihara, M., Asai, N., Suzuki, H., Takahashi, M., and Nakashima, I. (2000). Ultraviolet light induces redox reaction-mediated dimerization and superactivation of oncogenic Ret tyrosine kinases. Mol. Biol. Cell *11*, 93–101.
42. Knudson, A.G. Jr. (1971). Mutation and cancer: Statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA *68*, 820–823.
43. Simpson, M.J., Zhang, D.C., Mariani, M., Landman, K.A., and Newgreen, D.F. (2007). Cell proliferation drives neural crest cell invasion of the intestine. Dev. Biol. *302*, 553–568.