HHS Public Access

Author manuscript

J Pediatr Surg. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:

J Pediatr Surg. 2021 December ; 56(12): 2286–2294. doi:10.1016/j.jpedsurg.2021.04.010.

Multiple, independent, common variants at *RET*, *SEMA3* and *NRG1* gut enhancers specify Hirschsprung disease risk in European ancestry subjects

Ashish Kapoor[a,b,*], **Priyanka Nandakumar**[a], **Dallas R. Auer**[a], **Maria X. Sosa**[a], **Holly Ross**[a], **Juli Bollinger**[a], **Jia Yan**[a], **Courtney Berrios**[a], **The Hirschsprung Disease Research Collaborative (HDRC)**[**], **Aravinda Chakravarti**[a,c,*]

[a]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA;

[b]Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX 77030, USA;

[c]Center for Human Genetics and Genomics, New York University School of Medicine, New York, NY 10016, USA.

Abstract

Purpose: Hirschsprung disease (HSCR) is a developmental disorder of the enteric nervous system (ENS) characterized by congenital aganglionosis arising from coding variants in ENS genes causing partial or total loss-of-function. Low-penetrance, common, noncoding variants at *RET*, *SEMA3* and *NRG1* loci are also associated with HSCR, with small-to-moderate loss of gene expression mediated through sequence variants in *cis*-regulatory elements (CRE) as another causal mechanism. Since these latter variants are common, many individuals carry multiple risk variants. However, the extent and combinatorial effects of *all* putative CRE variants within and across these loci on HSCR is unknown.

Methods: Using 583 HSCR subjects, one of the largest samples of European ancestry studied, and genotyping 56 tag variants, we evaluated association of *all* common variants overlapping putative gut CREs and fine-mapped causal variants at *RET*, *SEMA3* and *NRG1*.

Results: We demonstrate that 28 and 8 tag variants, several of which are genetically independent, overlap putative-enhancers at the *RET* and *SEMA3* loci, respectively, as well as two fine-mapped tag variants at the *NRG1* locus, are significantly associated with HSCR. Importantly, disease risk

[*]**Corresponding author contacts:** Aravinda Chakravarti, Ph.D., Center for Human Genetics and Genomics, New York University School of Medicine, 435 East 30th Street SB Room 802/3, New York, NY 10016, USA, (T) 212-263-8023, aravinda.chakravarti@nyulangone.org; Ashish Kapoor, Ph.D., Center for Human Genetics, Institute of Molecular Medicine, University of Texas Health Science Center at Houston, 1825 Pressler Street, Room SRB 530C, Houston, TX 77030, USA, (T) 713-500-2439, (F) 713-500-2477, ashish.kapoor@uth.tmc.edu.

[**]A list of authors and their affiliations appear at the end of the paper.

Conflicts of Interest

The authors have no conflicts of interest to report.

increases with increasing numbers of risk alleles from multiple variants within and across these loci, varying >25-fold across individuals.

Conclusion: This increasing allele number-dependent risk, we hypothesize, arises from HSCR-relevant ENS cells sensing the reduced gene expression at multiple ENS genes since their developmental effects are integrated through gene regulatory networks.

Keywords

Hirschsprung disease; enhancer; polymorphism; genetic risk

1. Introduction

Hirschsprung disease (HSCR) is a developmental disorder of the enteric nervous system (ENS), characterized by loss of enteric ganglia in the myenteric and submucosal plexuses of the gut [1]. Failure of the precursor enteric neural crest cells to proliferate, differentiate, migrate and/or colonize the gastrointestinal tract leads to HSCR, the most common (15 per 100,000 live births [2]) cause of functional intestinal obstruction in neonates and infants [3]. The ensuing aganglionosis is caudal to rostral and based on the extent of aganglionosis, patients can be classified into short-segment HSCR (S-HSCR; aganglionosis limited up to the upper sigmoid colon, ~80% of cases), long-segment HSCR (L-HSCR; aganglionosis up to splenic flexure and beyond, ~15% of cases) or total colonic aganglionosis (TCA; aganglionosis of the entire large intestine, ~5% of cases) forms [1]. Nearly 80% of cases have aganglionosis only; in the remainder, HSCR co-occurs with multiple congenital anomalies, specific syndromes and/or chromosomal variants [1,4]. The critical features of HSCR are its high heritability (>80%), complex inheritance pattern, sex bias (4:1 affected male: female) and high recurrence risk to siblings and other relatives, influenced by sex, segment length of aganglionosis, familiality and the presence of additional anomalies [2]. An unsolved problem is to explain these diverse phenotypic features and associations with a single developmental mechanism.

Genetic studies in HSCR have uncovered rare, high-penetrance, coding variants in 14 genes [4–6] and common, low-penetrance, noncoding variants near *RET* [6–8], *NRG1* [9] and the *SEMA3* gene cluster [6]. The pathogenic coding alleles at these critical ENS genes are partial or total loss-of-function alleles in heterozygotes, and collectively lead to a population attributable risk (PAR) of ~18.2% [10]. In contrast, the polymorphic noncoding risk variants at *RET* and *SEMA3* lead to a much larger PAR of ~37.7% [10]. Although these noncoding variants at *RET* and *SEMA3* individually confer low-to-moderate risk (odds ratio, OR: 1.6–3.9), collectively, they contribute to a 30-fold *risk variation* across individuals depending on risk allele dosage [11]. While reduced *RET* expression mediated by noncoding regulatory variants is the mechanism at *RET*, the causal molecular basis of associations at *NRG1* and *SEMA3* remain unexplored. Functional studies at the *RET* locus, have proven that three risk variants lead to reduced *RET* expression by disrupting SOX10, GATA2 and RARB binding at three functionally distinct, but synergistically acting, *RET* gut enhancers or *cis*-regulatory elements (CREs) [8,12]. Although these three *RET* CRE variants are genetically independent (based on linkage disequilibrium (LD) in controls) they have synergistic effects on risk [11,12].

Given the widespread nature of multiple putative CREs for human genes, uncovered by the ENCODE [13] and NIH RoadMap Epigenomics [14] Projects, and supported by studies in model systems [15], along with extensive common noncoding variation in humans [16,17], it is *a priori* likely that *multiple CRE variants* at any given disease-associated locus can affect gene expression and thereby modulate disease risk, as we have shown for *RET* [12]. When multiple CRE variants affect a target gene's expression in *cis* there may be 'haplotype effects' independent of the LD between risk variants [12]. Thus, if larger numbers of variants lead to more extreme gene expression changes, and concomitantly higher risk, then affected individuals must be enriched for multiple risk variants. Indeed, there is considerable evidence, from prior segregation [2], linkage [18,19], and association [11] studies, that multiple risk variants define HSCR in *individuals*. In other words, mechanistically, association loci are likely to have multiple causal CRE variants, within and between loci, modulating gene expression and disease risk [20].

In this study, we have attempted to identify multiple causal CRE variants by evaluating association of *all* common noncoding variants overlapping HSCR-relevant putative enhancers at the *RET* and *SEMA3* loci with HSCR risk, in a large sample of European ancestry cases. We also include all fine-mapped variants at the *NRG1* locus [21], associated with HSCR risk in Asian ancestry subjects, to reassess whether this association also exists in this larger European ancestry sample. We indeed demonstrate that multiple, statistically independent, putative enhancer-overlapping variants at the *RET* and *SEMA3* loci, and fine-mapped variants at the *NRG1* locus, are significantly associated with HSCR risk. We also demonstrate that this risk increases as a logistic function with increasing risk allele dosage from multiple variants within and across these loci, suggesting that enteric neurons can quantitatively assess gene expression status across these, and presumably other, ENS genes. We hypothesize that this integration occurs through gene regulatory networks (GRNs).

2. Materials and Methods

2.1 Patient samples:

We analyzed HSCR patients and their family members ascertained from two sources: ongoing family studies from this (Chakravarti) laboratory (AC-HSCR) and by the Hirschsprung Disease Research Collaborative (HDRC; Appendix S1). The AC-HSCR collection includes participants recruited from referrals by practicing physicians, genetic counselors, family members and from self-referral through a study website and support group postings (n=461 families). The HDRC is a multi-institutional study that recruits participants from newly diagnosed cases from pediatric surgery and clinical centers at HDRC institutions (n=276 families). The majority of AC-HSCR and HDRC participants are of self-described European ancestry.

2.2 Ethics statement:

All individuals were ascertained with written informed consent approved by the Institutional Review Board of the enrolling institution. This study is based on samples approved by the Institutional Review Board of the Johns Hopkins School of Medicine (March 31, 2018).

2.3 Variant selection:

For the *RET* (Figure S1) and *SEMA3* (Figure S2) loci, we defined the target regions based on HSCR genome-wide association study (GWAS) signals [6] and flanking recombination hotspots from HapMap [16]. Within these regions we selected common (minor allele frequency (MAF) >10%) variants observed in 1000 Genomes [17] non-Finnish European (NFE) ancestry (n=404) controls that overlapped putative gut enhancer marks from public epigenomic databases [13,14], as described earlier [12] (Dataset S1). Due to lack of prior evidence for association at *NRG1* locus in European ancestry HSCR subjects, we did not perform the extensive association screen but selected all 17 fine-mapped variants reported to be associated with HSCR in Asian ancestry subjects (Dataset S1) [21]. All variants selected at a locus fall within the same topologically-associated domain as the underlying risk gene (*RET*, *SEMA3D* and *NRG1*) [22]. Using 1000 Genomes [17] NFE ancestry genotype data and the software Tagger [23], 42, 11 and 7 tag variants at the *RET*, *SEMA3* and *NRG1* loci, respectively were selected with a LD r^2 cutoff of 0.5. We added the top two GWAS hits at *SEMA3* locus (rs11766001 and rs12707682) [6] for a total of 62 variants that underwent multiplex genotyping assay design (Dataset S1).

2.4 Multiplex genotyping assay:

All steps of genotyping, including the design of custom multiplexed assays using iPLEX chemistry, were performed following the manufacturer's recommendations (Agena Bioscience, San Diego, CA, USA). Variants that failed *in silico* assay design or failed in genotyping of reference DNA samples were replaced with a best proxy variant from the filtered sets of variants when available; if not, we used the best replacement from the larger sets of common variants (based on LD) to generate two multiplexed assay pools for 61 tag variants (Dataset S1).

2.5 Variant genotyping quality control (QC):

Genotyping of 2,205 unique DNA samples was carried out for each multiplexed pool. After sample and variant QC (Supplementary Methods, Dataset S1, Table S1, S2), we combined variants across the two pools (31 and 25 variants in pools #1 and #2, respectively) to generate genotype data for 56 variants in 1,959 samples from 724 families. Genotypes at each variant from all unrelated HSCR probands (n=583) were tested for Hardy-Weinberg Equilibrium (HWE): none were significant ($P<1.0\times10^{-6}$) except rs2435357 ($P=2.69\times10^{-19}$; Table S3), which was expected owing to its high population-level association with HSCR [7,8]. None of the HWE tests using genotypes from 1000 Genomes [17] NFE ancestry subjects were statistically significant (Table S3).

2.6 Statistical genetic analyses:

For control allele frequencies, we used publicly available data from the Genome Aggregation Database (gnomAD) [24], based on whole-genome sequencing of NFE ancestry subjects. These allele frequencies, determined through a different technology than our genotyping, were highly concordant to our pseudo-control data from HSCR trios (Figure S3). All HSCR cases with complete information on sex, segment length of aganglionosis and familiality, were classified into 8 risk categories, scored by known HSCR risk factors

(score of 0 each for male, short segment and simplex versus a score of 1 each for female, long segment/TCA and multiplex), and eventually summed into risk scores between 0–3 (Table S4) [25]. Population-level disease penetrance/prevalence for each risk allele bin was estimated using Bayes' theorem with the observed background control frequency and a disease incidence of 15 per 100,000 live births [8]. Polygenic risk scores (PRS) were calculated as the sum of risk alleles at HSCR-associated variants weighted by $\log_{10}$-transformed OR from case-control tests (AC-HSCR and HDRC combined cases vs. NFE controls). Statistical significance thresholds for single variant-based tests were adjusted for multiple tests using the Bonferroni correction for 56 tests at $P<8.9\times10^{-4}$. All *P*-values reported are two-tailed.

3. Results

3.1 HSCR subtypes differ between cohorts

We used two HSCR sample cohorts in this study: AC-HSCR and HDRC. AC-HSCR, collected through this laboratory, is likely to over-sample high-risk families while HDRC collected at clinical centers at first surgical encounter is likely to be representative of HSCR in the general population (Appendix S1). To assess if ascertainment differences can influence our analysis (see below), we compared the proportions of HSCR subtypes by sex, segment length (S-HSCR versus L-HSCR/TCA) and familiality in probands across the two cohorts (Table 1). Ignoring samples with unknown status, there was no significant difference between the cohorts for sex (71% vs. 76% males in AC-HSCR and HDRC, respectively; $P=0.22$). However, HDRC was significantly enriched for S-HSCR and simplex cases as compared to AC-HSCR (73% vs. 60% S-HSCR, $P=0.007$; 85% vs. 66% simplex, $P=2.1\times10^{-6}$ in HDRC and AC-HSCR, respectively). This confirms the expectation that AC-HSCR was enriched for high-risk families, i.e. those with L-HSCR/TCA and with affected relatives. These data also suggest that, in unselected HSCR cases the frequency of males, L-HSCR/TCA and positive family history are 76%, 27% and 15%, respectively; in AC-HSCR these are 71%, 40% and 34%, respectively, and except for sex are considerably higher. We also classified each proband into 8 risk categories based on known HSCR risk factors [25] and four risk classes (scores 0–3; Table S4). Comparisons of risk scores 0, 1 and 2+3 (combined due to small sample sizes) between the two cohorts showed a highly significant difference ($P=3.3\times10^{-4}$), with groups 0 and 2+3 being more abundant in HDRC and AC-HSCR, respectively (Table S5). Thus, the two samples may reveal different genotype-phenotype correlations dependent on causal variants enriched in high- versus low-risk probands; also, the presence of causal association in one sample does not guarantee it in the other [8].

3.2 Widespread genetic association of RET, SEMA3 and NRG1 variants with HSCR

Previous studies have established that common noncoding variants at *RET*, *SEMA3* and *NRG1* are significantly associated with HSCR risk [6–9]. However, these studies were from gene discovery laboratories who may have over-sampled probands from high-risk families or suffered from the 'winners curse'. This is relevant because our first discovered genetic association had differing phenotypic features depending on HSCR risk factors [7,8,11]. Our goal here, for the *RET* and *SEMA3* loci, was therefore to assess the numbers and

degree of genetic associations at functionally relevant CREs, in HSCR cases including randomly ascertained new samples (HDRC), and to test for the presence of independent signals within each locus. We also wished to reassess the HSCR association [9] at *NRG1* in European ancestry subjects. We genotyped a total of 2,205 subjects across 737 families from AC-HSCR and HDRC for 61 tag variants (MAF>10%), which after QC (Dataset S1, Tables S1–S3) reduced to genotype data from 56 tag variants in 1,959 subjects (701 affected) from 724 families (583 probands). From these, data from 583 independent HSCR cases were used for association analyses (Table 1). Of the 375 AC-HSCR independent cases, 320 have been evaluated earlier for association with limited set of risk variants at *RET*, *SEMA3* and *NRG1*, and are used here for direct comparisons to the HDRC samples [11,12].

We used population-based case-control analysis and compared allele frequencies between HSCR cases and NFE ancestry controls (Table 2). At *RET*, 27 of 37 tag variants showed significant association with HSCR, including three previously published independent associations at rs2435357 (60% vs. 27%, OR=4.0, $P=1.1\times10^{-123}$) [7,8,11], rs2506030 (56% vs. 39% at rs788261, OR=2.0, $P=2.8\times10^{-30}$, $r^2=0.98$ with 2506030) [6,11] and rs7069590 (84% vs. 77%, OR=1.5, $P=1.8\times10^{-7}$) [12]. Among all *RET* variants evaluated, the most significant association remained at rs2435357. The ORs observed here at rs2435357, rs2506030 and rs7069590 are very similar to that previously reported in European ancestry HSCR subjects, in overlapping samples from AC-HSCR and by others (ORs range: 3.9–6.7, 1.8–2.3 and 1.5–1.7 for rs2435357, rs2506030 and rs7069590, respectively; Table S6). At *SEMA3*, 8 of 12 tag variants showed significant association with HSCR, including two previously published GWAS hits at rs12707682 (31% vs. 23%, OR=1.6, $P=2.6\times10^{-11}$) and rs11766001 (19% vs. 14%, OR=1.5, $P=1.3\times10^{-6}$) [6,11]. The ORs observed here at rs12707682 and rs11766001 are also very similar to that reported in European ancestry HSCR subjects, in overlapping samples from AC-HSCR and by others (ORs range: 1.3–1.8 and 1.4–2.3 for rs12707682 and rs11766001, respectively; Table S6). The most significant association at *SEMA3* was, however, at rs1228877 (45% vs. 33%, OR=1.6, $P=2.6\times10^{-16}$). At *NRG1* locus, 2 variants showed significant association with HSCR, including one of the Asian ancestry HSCR GWAS hits at rs16879552 (5% vs. 3%, OR=1.9, $P=2.7\times10^{-6}$) [9], a new result in European ancestry HSCR subjects [6,11,26]. This new observation, previously not significant ($P=0.4$) in overlapping samples from AC-HSCR [11] (Table S6), is driven by the HDRC samples (see below). The most significant association at *NRG1* was, however, at rs16879576 (5% vs. 2%, OR=2.3, $P=7.2\times10^{-9}$), also driven by the HDRC samples (see below). We also calculated parent-to-child transmission rates for all tag variants in 314 HSCR trios and observed over-transmission of risk alleles at 31 variants ($P<0.05$) (Table S7).

We calculated pairwise LD (r^2) between all tag variants at each of the three loci in HSCR cases, as well as the 1000 Genomes [17] NFE controls, to assess the underlying LD structure (Figures S4–S6) and identify associations at independent genetic variants. Based on these results, several new and independent association signals at *RET* (Figure S4) and *SEMA3* (Figure S5) are evident. However, the two *NRG1* variants rs16879552 and rs16879576 are in moderate LD ($r^2=0.53/0.73$ in controls/cases) (Figure S6) and may represent a singular causal association. Unlike the *RET* locus, *SEMA3* and *NRG1* loci have been largely associated with HSCR risk in European versus Asian ancestry subjects, respectively. Having

detected new significant associations at *NRG1* variants in European ancestry subjects here, we assessed if allele frequency differences between the two ancestries, in addition to ascertainment differences (see below), could explain these observed population-specific associations. Indeed, in general, allele frequencies for tag variants, including the GWAS hits at *SEMA3* and *NRG1* loci are quite different between NFE and East Asian gnomAD controls [24] (Figure S7).

Given the ascertainment differences in AC-HSCR versus HDRC, we performed association analysis for all tag variants in them separately (Table S8). At *SEMA3*, the risk allele at rs12707682 was at higher frequency and led to an increased OR in HDRC (23%/29%/36% allele frequency in controls/AC-HSCR/HDRC with significant ORs of 1.4 (95% CI: 1.17–1.63) vs. 1.9 (95% CI: 1.55–2.33) in AC-HSCR and HDRC, respectively). However, within the same locus, significant risks at rs11766001 and rs1228937 were only evident in AC-HSCR (14%/21%/16% and 20%/28%/23% allele frequency in controls/AC-HSCR/HDRC with ORs of 1.7 (95% CI: 1.38–2.00) vs. 1.1 (95% CI: 0.87–1.50) and 1.6 (95% CI: 1.32–1.84) vs. 1.2 (95% CI: 0.98–1.55) in AC-HSCR and HDRC for rs11766001 and rs1228937, respectively). The confidence intervals for the above three *SEMA3* variants overlap and so the estimates are not significantly different between the two cohorts. At *NRG1*, significant risks at rs16879552 and rs16879576 were only observed in HDRC (3%/4%/8% and 2%/3%/7% allele frequency in controls/AC-HSCR/HDRC with ORs of 1.3 (95% CI: 0.85–1.88) vs. 3.1 (95% CI: 2.18–4.48) and 1.7 (95% CI: 1.09–2.51) vs. 3.6 (95% CI: 2.44–5.36) in AC-HSCR and HDRC for rs16879552 and rs16879576, respectively), in accord with results from our previous study of overlapping AC-HSCR cases [11] (Table S6). At the *RET* locus, there were several variants with differential effects on HSCR risk between AC-HSCR and HDRC, including five with significant risk in AC-HSCR cases only (rs3758514, rs788243, rs7069590, rs12764797 and rs2472737), five with significant risk in HDRC cases only (rs3004255, rs4949062, rs3026693, rs1800860 and rs10793423), and at least five with substantially increased risk in HDRC as compared to AC-HSCR (rs2435357, rs2506024, rs2505515, rs2505513 and rs7074964) (Table S8). Overall, these results demonstrate that generally we detect a greater magnitude of associations in HDRC as compared to AC-HSCR, i.e., in samples enriched for simplex S-HSCR cases. For further analyses, any variant with statistically significant association in AC-HSCR, HDRC or AC-HSCR and HDRC combined cases was considered to be associated with HSCR risk.

We expect common variants to contribute differential risk across HSCR subtypes and, depending on the proportions of HSCR subtypes across cohorts, to overall risk. To evaluate this expectation we classified all HSCR subjects (AC-HSCR and HDRC combined) into four risk score groups (0, 1, 2, and 3) based on HSCR subtypes, as outlined above (Table S4) and performed case-control association tests for all variants across the three loci. Figures S8, S9 and S10 show the odds ratios and confidence intervals for each variant at *SEMA3*, *NRG1* and *RET* locus, respectively, by risk score 0, 1 and 2+3 (combined due to small sample size). As expected, common variants associated with overall HSCR risk contribute the largest risk to group 0, followed by group 1, with largely no significant risk to the 2+3 group. Unsurprisingly, the variants with no significant effects shows no such trend.

3.3 Risk of HSCR is additive across variants and genes

Having observed multiple single variant associations at *RET* (28 variants; Tables 2, S8; Figure S11), *SEMA3* (8 variants; Tables 2, S8; Figure S12) and *NRG1* (2 variants; Tables 2, S8), we next assessed their cumulative effects, within and across loci, on HSCR risk (Tables 3, S9–S10; Figure 1, S13). Here, we analyzed the two collections as one sample. For within locus analysis, given the rarity of risk alleles at the significant *NRG1* variants rs16879552 and rs16879576, we did not assess their effect. Risk alleles across all significant variants within *RET* and *SEMA3* loci were counted in each individual with neighboring classes combined to generate five non-overlapping bins to reduce numbers of tests and to have similar sample sizes across bins in controls (Tables S9, S10). Similarly, for analysis across loci, risk alleles at all significant variants were counted in each individual and combined to generate five non-overlapping bins (Table 3).

At *RET*, the total risk allele dosage was significantly associated with HSCR risk (Table S9; χ^2=134.7, P=3.9×10^{-28}; Figure S13) with the first three bins showing significant protection (OR=0.3/0.4/0.5 and P=3.2×10^{-8}/2.8×10^{-6}/6.7×10^{-5}, for 17–23, 24–26 and 27–29 risk alleles, respectively) and the highest risk alleles bin having significantly increased disease risk (OR=4.9, P=2.8×10^{-26} for 35–52 risk alleles). The logistic function describing the relationship between HSCR risk and *RET* risk allele counts explains 80% of risk variation (Figure S13). We estimated the population-level penetrance (population probability of being affected given a genotype), which ranged from 4.9 cases to 40.0 cases for the lowest (17–23 risk alleles) and the highest (35–52 risk alleles) bins, respectively (Table S9). These values translate to a population incidence difference ranging from ~1/20,000 to ~1/2,500 live births. At *SEMA3*, although there was a trend for an analogous association with risk allele dosage (Table S10; χ^2=11.9, P=0.018; Figure S13), only two risk alleles bins (1–3 risk alleles, OR=0.7, P=0.014; 8–9 risk alleles, OR=1.5, P=0.019) showed a trend towards statistical significance. The logistic function describing the relationship between HSCR risk and *SEMA3* risk allele counts explains 78% of risk variation (Figure S13). Counting risk alleles across the three loci led to an expected association between total risk allele dosage and HSCR risk (Table 3; χ^2=156.4, P=8.5×10^{-33}; Figure 1), which, however, was more significant when compared to risk from the *RET* locus alone (Table S9; χ^2=134.7, P=3.9×10^{-28}; Figure S13). The risk for HSCR from these variants increased from 0.2 to 5.4, a 26-fold change, as the bin size increased from 19–29 to 43–66 risk alleles. Also, the logistic function describing the relationship between HSCR risk and risk allele counts across the three loci explains 92% of risk variation, more than explained from *RET* alone (Figures 1, S13). The estimated population-level penetrance ranged from 3.7 cases to 45.4 cases for the lowest (19–29 risk alleles) to the highest (43–66 risk alleles) bins, respectively (Table 3). These values translate to a population incidence difference ranging from ~1/27,000 to ~1/2,200 live births. Collectively, these results show a clear additive effect of *RET* risk variants on HSCR risk and provide evidence for minor additional effects from *SEMA3* and *NRG1* risk variants.

Next, we wanted to assess how the combined (polygenic) risk from all the 38 significantly associated variants (8 *SEMA3*, 2 *NRG1* and 28 *RET* variants) varied across HSCR subjects. PRS for each HSCR subject was calculated and varied from 5.12 to 16.04 (median=10.66)

and from 6.47 to 16.46 (median=11.28) in AC-HSCR and HDRC cohorts, respectively. Although the PRS variation followed normal distributions in both cohorts with mean values of 10.60 and 11.19 for AC-HSCR and HDRC, respectively, there was indeed a small yet significant difference in the two sample means (z=2.63; P=0.009; Figure 2). We also calculated PRS after LD r^2-based pruning of significantly associated variants at each locus to remove moderately correlated variants (r^2>0.2) leaving 3, 1 and 11 variants at *SEMA3*, *NRG1* and *RET* locus, respectively. PRS varied from 1.63 to 6.65 (median=4.55) and from 1.87 to 7.06 (median=4.94) for in AC-HSCR and HDRC cohorts, respectively. As observed above with the PRS distribution based on all associated variants, the two cohorts show small but significant difference in polygenic risk distribution with relatively higher risk in the HRDC cohort (mean of 4.48 and 4.77 for AC-HSCR and HDRC, respectively; z=2.71; P=0.007; Figure S14). We expect the difference in composition of HSCR subtypes between the two cohorts (Table 1) to be the major factor leading to demonstrable variation in polygenic risk.

4. Discussion

Evidence in the literature [6–9,11,26,27], and the far more extensive data reported here (Table 2, Table S8), more than adequately confirms the role of multiple CRE polymorphisms at *RET*, *SEMA3* and *NRG1* on HSCR susceptibility in European ancestry subjects. In comparison with a *trans*-ethnic meta-analysis of GWAS for HSCR [28], the top *RET* variant here, rs2435357 is in high LD with the meta-analysis lead variant rs2505998 (r^2=0.97 in Non-Finnish 1000 Genome European ancestry subjects), thus likely representing the same association signal. At *NRG1*, the top two associated variants here, rs16879552 and rs16879576 are not in LD with rs7005606, the meta-analysis lead variant (r^2 of 0.024 and 0.011 in Non-Finnish 1000 Genome European ancestry subjects, respectively), thus likely representing a different association signal. In addition, rs7835688, which is in complete LD with rs7005606 (r^2=1 in Non-Finnish 1000 Genome European ancestry subjects), is not associated with HSCR in our study. At *SEMA3*, the lead variant rs80227144 from the *trans*-ethnic meta-analysis is in weak LD with the best tagging variant here, rs11766001 (r^2=0.23 in Non-Finnish 1000 Genome European ancestry subjects) that is associated with HSCR in our study. The top two associated variants here, rs1228877 and rs12707682, are in very low LD with rs80227144 (r^2 0.072 and 0.149 in Non-Finnish 1000 Genome European ancestry subjects, respectively), thus likely representing a different association signal.

Whether these specific variants disrupt enhancer function to dysregulate these genes remain to be experimentally demonstrated, as we have previously shown for rs2435357, rs2506030, and rs7069590 at *RET* [12]. However, the most parsimonious hypothesis is that these gut enhancer variants reduce gene expression at *RET*, *SEMA3* and *NRG1* to increase the risk of aganglionosis, with risk increasing as a logistic function of the numbers of risk alleles (Figure 1). A first question is whether marked expression reduction at *RET* alone leads to HSCR given its large impact relative to *SEMA3* and *NRG1*? This is unlikely because disease also occurs in non-risk allele homozygotes at *RET*, albeit at a lower rate. We have recently demonstrated [10], following years of indirect evidence from segregation [2], linkage [18,19] and association studies [11] that individual patients harbor multiple rare risk coding variants across different genes; here, we show the same feature with common CRE

risk variants. This prompts the second question: how does an enteric neuron 'count' risk variants, or recognize expression change, across diverse genes to reflect its potential for aganglionosis?

The mechanical and technical aspects of mapping variants to disease, long recognized through GWAS of complex traits, does not, however, clarify the underlying 'counting' mechanism even when we know that disease results from gene expression changes. For understanding disease pathophysiology, we need to understand the degree of expression change required for aganglionosis to result. Clearly, single variants, the units of mapping, show disease association. However, this does not imply that the variant alone leads to disease, indeed, it likely does not, because its genetic effect is usually small. Further, even 50% gene expression reduction at *RET*, in heterozygotes for null coding alleles, does not necessarily lead to a clinical phenotype [8]. Thus, there must be >50% expression reduction at each gene before disease results, a threshold unlikely to be reached by single variants or even single genes and their enhancers. Consequently, multiple functional variants act together to have a sufficient effect on gene expression to reach this threshold; the more extreme the threshold the larger the genetic effect whether from one or many variants.

The genetic consequences of this hypothesis is that these risk variants are 'counted' by enteric neurons. One possibility is that multiple risk variants physically exist on a single chromosome in *cis* to exert a significant effect on gene expression, irrespective of whether these variants are in LD or not. Indeed, the causal genetic variants may be correlated, leading to the suspicion that they do not impart independent effects; it is precisely this dependence (accumulation) that leads to a 'jackpot' effect on gene expression by that haplotype. In other words, LD changes the frequency of this variant haplotype but not its impact on the phenotype.

How then are the effects counted across genes? Current data suggests that the *RET*, *SEMA3* and *NRG1* effects on HSCR are additive, not unlikely given their small effect sizes. As we have shown for *RET* [12], developmental genes function within a GRN and the effects of a set of variants on one gene affects the expression of other network genes. Thus, combinations of alleles at these three genes may impart greater risk through affecting all genes in the network. For ENS development, the known GRN regulates both *RET* and *EDNRB*, so that all variants that reduce their expression levels significantly will lead to aganglionosis. Proving this hypothesis using human genetic data alone is going to be arduous because HSCR is a rare disorder and difficult to collect thousands of samples. In contrast, mouse models with enhancer variants at *Ret*, *Sema3* and *Nrg1* can be used to test the contention that even with variants that reduce gene expression at developmentally important disease genes, disease only results from specific combinations of genotypes across these loci depending on the architecture of the GRN and three-dimensional chromatin contacts. Such a hypothesis may be sufficient to explain the need for multiple genes in a disease, multiple variants in these genes, small genetic effects of disease-associated alleles, the familiality of the disorder and its reduced penetrance.

This work also shows that the nature of the patients ascertained in a study can significantly impact the variants and genes identified. Past studies had a greater focus on the analysis

of rare variants in coding genes, which were successful in multiplex, severe and syndromic families. However, identification of regulatory defects in HSCR have succeeded mostly in simplex, short segment non-syndromic cases. Since the latter are far more common than the former, exploring the regulatory hypothesis in HSCR through GWAS is likely to be more fruitful for explaining disease risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank the numerous patients, their families, referring physicians, nurses and genetic counselors who have contributed to these studies, and Erick Kaufmann, Jennifer (Scott) Bubb, Maura Kenton, Julie Albertus and Magan Trottier for family ascertainment, family studies and genetic counseling across the years. The studies reported here were partially supported by a grant from the US National Institutes of Health (MERIT Award HD28088 to A.C.).

Funding

The studies reported here were partially supported by a grant from the US National Institutes of Health (MERIT Award HD28088 to A.C.).

Hirschsprung Disease Research Collaborative (HDRC)

Corresponding author

Aravinda Chakravarti[1,2,#]

Steering Committee

Courtney Berrios[1,3,#]

Aravinda Chakravarti[1,2,#]

Philip Frykman[4]

Cheryl Gariepy[5]

Raj Kapur[6]

Jacob Langer[7]

Sample Collection

Jeffrey Avansino[8] (Principal Investigator), Courtney Berrios[1,3,#], Amina Bhatia[9], Andrea Bischoff[10] (Principal Investigator), Aravinda Chakravarti[1,2,#] (Principal Investigator), Nicole Chandler[11],*** (Principal Investigator), Robert Cina[12] (Principal Investigator), Daniel DeUgarte[13] (Principal Investigator), James Dunn[13] (Principal Investigator), Megan Durham[9] (Principal Investigator) Jason Frischer[14],*** (Principal Investigator), Philip Frykman[4] (Principal Investigator), Samir Gadepalli[15],*** (Principal Investigator), Cheryl Gariepy[5] (Principal Investigator), Ankush Gosain[16] (Principal Investigator), Gunadi[17] (Principal

Investigator), Raj Kapur[6] (Principal Investigator), Jacob Langer[7] (Principal Investigator), Monica Lopez[18] (Principal Investigator), Lisa McMahon[19] (Principal Investigator), Suyin Lum Min[20],*** (Principal Investigator), Hector Monforte[11],***, Isam Nasr[21], Dorothy Rocourt[22] (Principal Investigator), David Rodeberg[23] (Principal Investigator), Michael Rollins[24] (Principal Investigator), Robert Russell[25] (Principal Investigator), Payam Saadai[26] (Principal Investigator), Lois Sayrs[19], Donald Shaul[27],*** (Principal Investigator), Douglas Tamura[28] (Principal Investigator), Richard Weiss[29] (Principal Investigator), Jia Yan[1,2,#]

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [2]Center for Human Genetics and Genomics, New York University School of Medicine, New York, NY 10016, USA; [3]Childrens's Mercy Research Institute, Kansas City, MO 64108, USA; [4]Department of Pediatric Surgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; [5]Department of Gastroenterology, Hepatology & Nutrition, Nationwide Children's Hospital, Columbus, OH 43205, USA; [6]Department of Laboratories Pathology, Seattle Children's Hospital, Seattle, WA 98105, USA; [7]Department of General and Thoracic Surgery, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada; [8]Department of Pediatric Surgery, Seattle Children's Hospital, Seattle, WA 98105, USA; [9]Children's Healthcare of Atlanta, Emory University School of Medicine, Atlanta, GA 30322, USA; [10]Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO 80045, USA; [11]All Children's Hospital, St. Petersburg, FL 33701, USA; [12]Department of Surgery, Medical University of South Carolina College of Medicine, Charleston, SC 29425, USA; [13]Department of Surgery, University of California Los Angeles School of Medicine, Los Angeles, CA 90095, USA; [14]Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH 45229, USA; [15]Pediatirc Surgery, Michigan Medicine, Ann Arbor, MI 48109, USA; [16]Le Bonheur Children's Hospital, University of Tennessee Health Science Center, Memphis, TN 38103, USA; [17] Pediatric Surgery Division, Department of Surgery, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/Dr. Sardjito Hospital, Yogyakarta 55281, Indonesia; [18]Pediatric Surgery, Texas Children's Hospital, Houston, TX 77030, USA; [19]Department of Surgery, Phoenix Children's Hospital, Phoenix, AZ 85016, USA; [20]The Children's Hospital of Winnipeg, University of Manitoba, Winnipeg, MB R3A 1S1, Canada; [21]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [22]Pediatric Surgery, Penn State Health Children's Hospital, Hershey, PA 17033, USA; [23]Department of Surgery, East Carolina University School of Medicine, Greenville, NC 27834, USA; [24]Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT 84132, USA; [25]Department of Surgery, University of Alabama at Birmingham School of Medicine, Birmingham, AL 35233, USA; [26]Department of Surgery, University of California, Davis School of Medicine, Sacramento, CA 95817, USA; [27]Kaiser Permanente Southern California, Los Angeles, CA 90027, USA; [28]Valley Children's Hospital, Madera, CA 93636, USA; [29]Pediatric Surgery, Connecticut Children's Medical Center, Hartford, CT 06106, USA.

***Site no longer part of the HDRC

#Author of the manuscript; everyone else is a contributor

List of abbreviations

HSCR	Hirschsprung disease
ENS	enteric nervous system
S-HSCR	short-segment HSCR
L-HSCR	long-segment HSCR
TCA	total colonic aganglionosis
PAR	population attributable risk
OR	odds ratio
CREs	*cis*-regulatory elements
LD	linkage disequilibrium
GRNs	gene regulatory networks
AC-HSCR	Chakravarti laboratory HSCR
HDRC	Hirschsprung Disease Research Collaborative
GWAS	genome-wide association study
MAF	minor allele frequency
NFE	non-Finnish European
HWE	Hardy-Weinberg Equilibrium
gnomAD	Genome Aggregation Database
PRS	Polygenic risk scores

References

[1]. Chakravarti A, Lyonnet S Hirschsprung Disease. In Scriver CR, Beaudet AL, Valle D, et al. editors. The Metabolic and Molecular Bases of Inherited Disease (8th edition), New York, McGraw-Hill; 2001, p. 6231–6255.

[2]. Badner JA, Sieber WK, Garver KL, et al. A genetic study of Hirschsprung disease. Am.J.Hum.Genet 1990;46:568–80. [PubMed: 2309705]

[3]. Parisi MA, Kapur RP Genetics of Hirschsprung disease. Curr.Opin.Pediatr 2000;12:610–7. [PubMed: 11106284]

[4]. 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. [PubMed: 17965226]

[5]. Alves MM, Sribudiani Y, Brouwer RW, et al. Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. Dev.Biol 2013;382:320–9. [PubMed: 23707863]

[6]. 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. [PubMed: 25839327]

[7]. 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. [PubMed: 15829955]

[8]. 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. [PubMed: 20598273]

[9]. 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. [PubMed: 19196962]

[10]. Tilghman JM, Ling AY, Turner TN, et al. Molecular Genetic Anatomy and Risk Profile of Hirschsprung's Disease. N.Engl.J.Med 2019;380:1421–32. [PubMed: 30970187]

[11]. 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. [PubMed: 25666438]

[12]. Chatterjee S, Kapoor A, Akiyama JA, et al. Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. Cell 2016;167:355,368.e10. [PubMed: 27693352]

[13]. ENCODE Project Consortium, Bernstein BE, Birney E, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74. [PubMed: 22955616]

[14]. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–30. [PubMed: 25693563]

[15]. Segal E, Raveh-Sadka T, Schroeder M, et al. Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature 2008;451:535–40. [PubMed: 18172436]

[16]. International HapMap Consortium A haplotype map of the human genome. Nature 2005;437:1299–320. [PubMed: 16255080]

[17]. 1000 Genomes Project Consortium A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73. [PubMed: 20981092]

[18]. Bolk S, Pelet A, Hofstra RM, et al. A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. Proc.Natl.Acad.Sci.U.S.A 2000;97:268–73. [PubMed: 10618407]

[19]. Gabriel SB, Salomon R, Pelet A, et al. Segregation at three loci explains familial and population risk in Hirschsprung disease. Nat.Genet 2002;31:89–93. [PubMed: 11953745]

[20]. Kapoor A, Lee D, Zhu L, et al. Multiple SCN5A variant enhancers modulate its cardiac gene expression and the QT interval. Proc.Natl.Acad.Sci.U.S.A 2019;116:10636–45. [PubMed: 31068470]

[21]. Tang CS, Tang WK, So MT, et al. Fine mapping of the NRG1 Hirschsprung's disease locus. PLoS One 2011;6:e16181. [PubMed: 21283760]

[22]. Rao SS, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159:1665–80. [PubMed: 25497547]

[23]. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5. [PubMed: 15297300]

[24]. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91. [PubMed: 27535533]

[25]. Turner TN, Sharma K, Oh EC, et al. Loss of delta-catenin function in severe autism. Nature 2015;520:51–6. [PubMed: 25807484]

[26]. Fadista J, Lund M, Skotte L, et al. Genome-wide association study of Hirschsprung disease detects a novel low-frequency variant at the RET locus. Eur.J.Hum.Genet 2018;26:561–9. [PubMed: 29379196]

[27]. Li Q, Zhang Z, Diao M, et al. Cumulative Risk Impact of RET, SEMA3, and NRG1 Polymorphisms Associated With Hirschsprung Disease in Han Chinese. J.Pediatr.Gastroenterol.Nutr 2017;64:385–90. [PubMed: 27203398]

[28]. Tang CS, Gui H, Kapoor A, et al. Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease. Hum.Mol.Genet 2016;25:5265–75. [PubMed: 27702942]

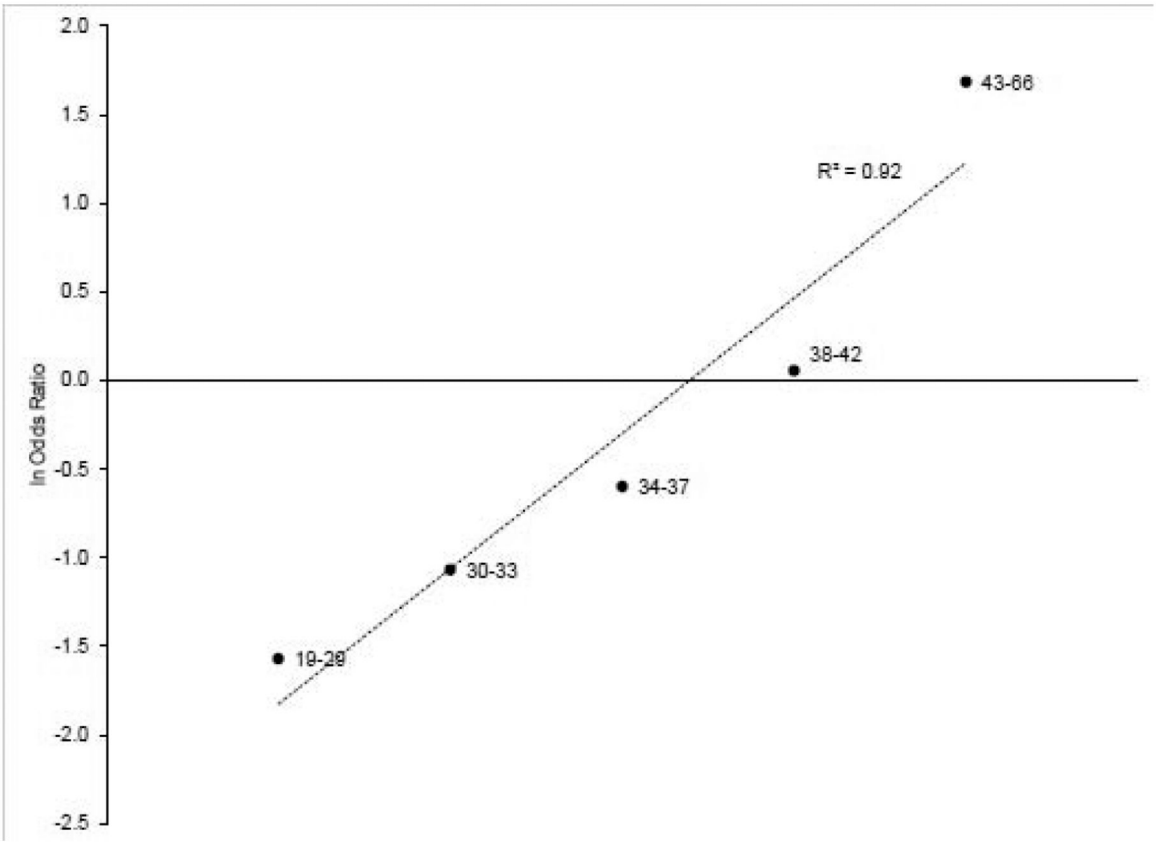

Figure 1: Logistic function relating Hirschsprung disease risk with the total number of risk alleles at *RET, SEMA3* and *NRG1* loci.

The logarithm of odds ratios (Y-axis) for binned risk allele counts (X-axis) based on significant variants across the three loci are shown along with the best linear fit (dashed line).

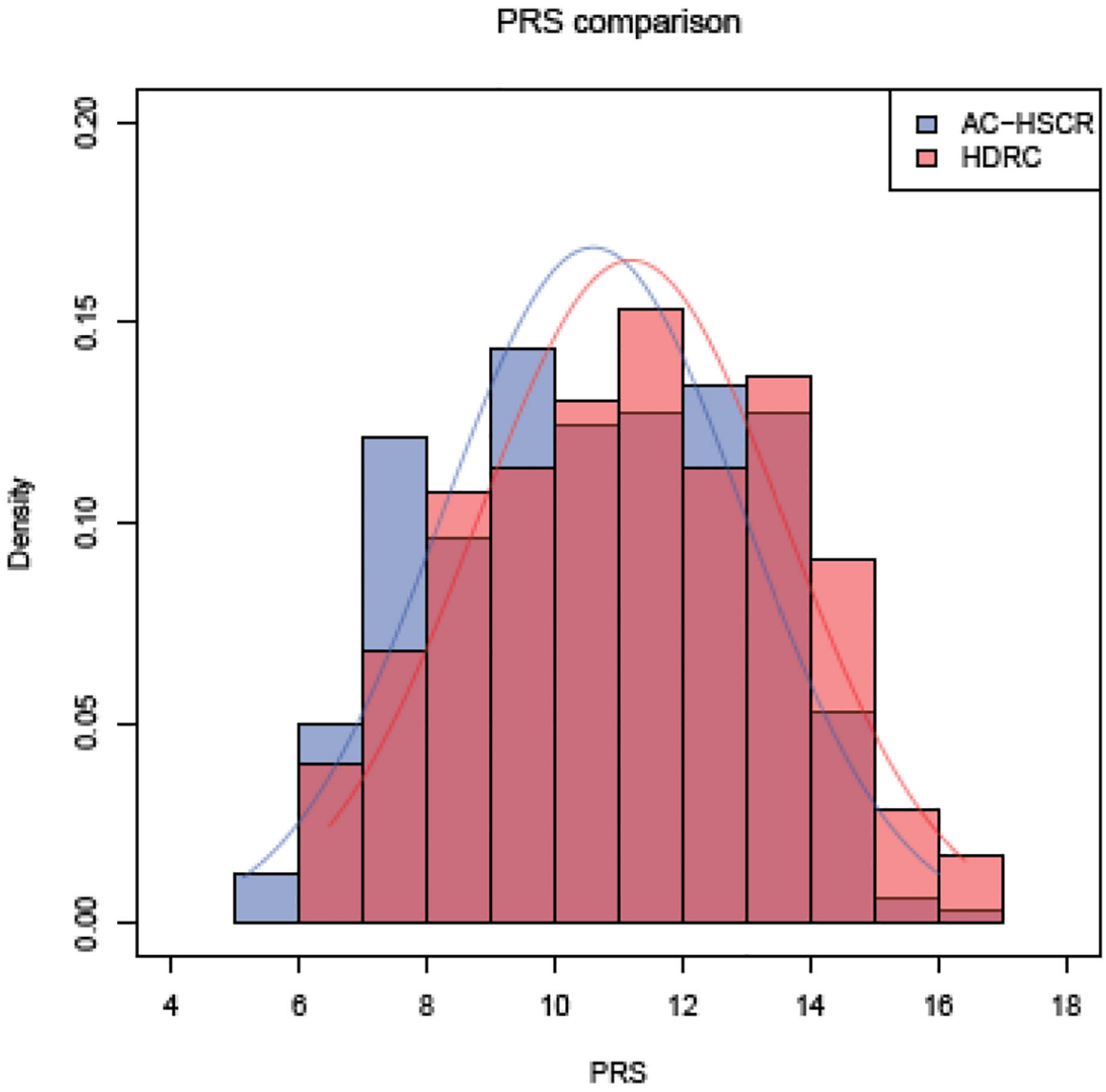

Figure 2: Polygenic risk score distributions.

Empirical frequency distributions of polygenic risk scores (PRS) in AC-HSCR (blue) and HDRC (red) HSCR subjects are plotted along with their corresponding fitted normal distributions.

Table 1:

Distribution of HSCR probands from the AC-HSCR and HDRC studies by sex, segment length of aganglionosis and familiality.

Category	Combined Count (%)	AC-HSCR Count (%)	HDRC Count (%)	Test of frequency difference	
All	**583 (100)**	**375 (100)**	**208 (100)**		
				χ^2	*P*
Sex					
Male	425 (72.9)	267 (71.2)	158 (76.0)	1.54[a]	2.15×10^{-1}
Female	158 (27.1)	108 (28.8)	50 (24.0)		
Segment length					
Short	274 (47.0)	162 (43.2)	112 (53.8)		
Long	57 (9.8)	42 (11.2)	15 (7.2)	7.22[b]	7.23×10^{-3}
TCA	91 (15.6)	65 (17.3)	26 (12.5)		
Indeterminate	161 (27.6)	106 (28.3)	55 (26.4)		
Familiality					
Simplex	425 (72.9)	249 (66.4)	176 (84.6)	22.47[c]	2.14×10^{-6}
Multiplex	158 (27.1)	126 (33.6)	32 (15.4)		

[a]Comparison between males and females,

[b]short and long+TCA cases, and

[c]simplex and multiplex cases, respectively.

Table 2:

Case-control association tests for tag variants at *SEMA3*, *NRG1* and *RET* in HSCR.

Gene Locus	Variant	Coded allele	Allele frequency (case/control)	Odds ratio (95% CI)	*P*
SEMA3	rs10266471	C	0.83/0.78	1.37 (1.17–1.60)	**9.98×10^{-5}**
SEMA3	rs1228877	G	0.45/0.33	1.65 (1.46–1.86)	**2.60×10^{-16}**
SEMA3	rs55737059	A	0.90/0.89	1.04 (0.85–1.26)	7.26×10^{-1}
SEMA3	rs1228937	C	0.26/0.20	1.44 (1.25–1.65)	**1.95×10^{-7}**
SEMA3	rs11766001	C	0.19/0.14	1.46 (1.25–1.71)	**1.31×10^{-6}**
SEMA3	rs2190050	T	0.22/0.19	1.18 (1.02–1.36)	2.80×10^{-2}
SEMA3	rs35092834	T	0.73/0.66	1.44 (1.26–1.64)	**1.27×10^{-7}**
SEMA3	rs12707682	C	0.31/0.23	1.56 (1.37–1.77)	**2.57×10^{-11}**
SEMA3	rs7811476	G	0.32/0.30	1.06 (0.93–1.21)	3.65×10^{-1}
SEMA3	rs6468008	G	0.58/0.51	1.34 (1.19–1.52)	**1.49×10^{-6}**
SEMA3	rs17560655	C	0.18/0.12	1.59 (1.36–1.86)	**6.96×10^{-9}**
SEMA3	rs6974210	G	0.87/0.86	1.04 (0.87–1.24)	6.69×10^{-1}
NRG1	rs16879425	C	0.12/0.09	1.28 (1.06–1.54)	9.53×10^{-3}
NRG1	rs10954845	G	0.24/0.20	1.26 (1.10–1.45)	1.20×10^{-3}
NRG1	rs4422736	T	0.07/0.05	1.34 (1.06–1.69)	1.39×10^{-2}
NRG1	rs7826312	C	0.59/0.58	1.03 (0.91–1.16)	6.46×10^{-1}
NRG1	rs16879552	T	0.05/0.03	1.90 (1.45–2.50)	**2.66×10^{-6}**
NRG1	rs7835688	C	0.50/0.47	1.14 (1.01–1.28)	3.72×10^{-2}
NRG1	rs16879576	A	0.05/0.02	2.34 (1.74–3.14)	**7.17×10^{-9}**
RET	rs11239767	T	0.82/0.81	1.07 (0.92–1.25’	3.89×10^{-1}
RET	rs12570505	T	0.89/0.89	1.04 (0.86–1.26)	6.82×10^{-1}
RET	rs3758514	T	0.65/0.60	1.20 (1.06–1.?)	4.73×10^{-3}
RET	rs61844837	C	0.94/0.91	1.61 (1.25–2.06)	**1.54×10^{-4}**
RET	rs2796552	A	0.73/0.70	1.15 (1.01–1.32)	3.81×10^{-2}
RET	rs788243	G	0.57/0.49	1.40 (1.24–1.57)	**5.74×10^{-8}**
RET	rs2796575	A	0.87/0.84	1.2 ‘ (1.06–1.52)	8.29×10^{-3}
RET	rs788274	A	0.27/0.13	2.50 (2.17–2.87)	**3.87×10^{-40}**
RET	rs788263	G	0.56/0.39	1.99 (1.77–2.25)	**4.93×10^{-30}**
RET	rs788261	C	0.56/0.39	2.00 (1.77–2.25)	**2.83×10^{-30}**
RET	rs788260	A	0.56/0.39	1.99 (1.76–2.25)	**5.86×10^{-30}**
RET	rs3004255	T	0.47/0.40	1.31 (1.17–1.48)	**7.61×10^{-6}**
RET	rs1935646	A	0.29/0.26	1.14 (1.00–1.30)	5.43×10^{-2}
RET	rs11819501	C	0.93/0.90	1.36 (1.08–1.70)	7.71×10^{-3}
RET	rs2488279	T	0.. 8/0.20	1.56 (1.37–1.79)	**3.70×10^{-11}**
RET	rs57621833	T	0.77/0.73	1.24 (1.07–1.42)	3.12×10^{-3}

Gene Locus	Variant	Coded allele	Allele frequency (case/control)	Odds ratio (95% CI)	*P*
RET	rs1547930	G	0.83/0.76	1.58 (1.35–1.85)	**1.05×10^{-8}**
RET	rs4949062	C	0.73/0.67	1.34 (1.17–1.53)	**1.77×10^{-5}**
RET	rs2115816	A	0.40/0.36	1.18 (1.04–1.33)	7.78×10^{-3}
RET	rs3026693	A	0.21/0.16	1.41 (1.21–1.63)	**6.45×10^{-6}**
RET	rs706_ 590	T	0.84/0.77	1.53 (1.30–1.79)	**1.79×10^{-7}**
RET	rs243535 ‘	T	0.60/0.27	4.02 (3.56–4.55)	**1.13×10^{-123}**
RET	rs12247456	G	0.80/0.68	1.84 (1.59–2.14)	**1.04×10^{-16}**
RET	rs73937	G	0.79/0.68	1.83 (1.58–2.12)	**2.32×10^{-16}**
RET	rs2506024	A	0.80/0.60	2.68 (2.31–3.11)	**9.76×10^{-42}**
RET	rs78146028	A	0.95/0.89	2.30 (1.77–3.00)	**1.55×10^{-10}**
RET	rs12764797	G	0.95/0.91	2.05 (1.55–2.70)	**2.06×10^{-7}**
RET	rs2505515	G	0.88/0.75	2.46 (2.05–2.95)	**8.38×10^{-24}**
RET	rs1800860	G	0.74/0.68	1.33 (1.16–1.53)	**3.63×10^{-5}**
RET	rs11238441	C	0.91/0.82	2.41 (1.95–2.97)	**3.21×10^{-17}**
RET	rs2472737	A	0.29/0.22	1.40 (1.23–1.60)	**5.62×10^{-7}**
RET	rs2505513	C	0.39/0.25	1.90 (1.67–2.15)	**1.95×10^{-24}**
RET	rs10793423	G	0.52/0.44	1.37 (1.21–1.54)	**2.97×10^{-7}**
RET	rs869184	T	0.69/0.53	1.93 (1.70–2.20)	**1.16×10^{-24}**
RET	rs953999	G	0.80/0.68	1.94 (1.67–2.25)	**6.13×10^{-19}**
RET	rs72783255	C	0.89/0.86	1.28 (1.06–1.54)	1.07×10^{-2}
RET	rs7074964	A	0.42/0.29	1.79 (1.59–2.03)	**3.88×10^{-21}**

The tag variants at each locus are listed in genomic order. The frequency of the coded allele (higher frequency in HSCR) at each variant in cases and controls, odds ratio with 95% confidence interval (CI) and the statistical significance of association (*P*) are provided. Multiple test-corrected significant *P*-values are in bold.

Table 3:

Odds ratio and risk of HSCR as a function of the number of risk-increasing variants at *RET*, *SEMA3* and *NRG1*.

Risk allele counts	Number (%) of cases	Number (%) of controls	Odds ratio (95% CI)	*P*	Penetrance (x10^5)
19–29	25 (5.0)	82 (20.3)	**0.21** (0.13–0.33)	5.64×10^{-11}	3.7
30–33	44 (8.9)	89 (22.0)	**0.34** (0.23–0.51)	7.32×10^{-8}	6.0
34–37	66 (13.3)	88 (21.8)	**0.55** (0.39–0.78)	8.28×10^{-4}	9.1
38–42	94 (18.9)	73 (18.1)	1.06 (0.75–1.48)	7.46×10^{-1}	15.7
43–66	268 (53.9)	72 (17.8)	**5.40** (3.96–7.36)	1.54×10^{-26}	45.4

Overall χ^2=156.43; P=8.52×10^{-33}

The numbers of cases and controls classified by the number of risk alleles at *RET*, *SEMA3* and *NRG1* significant variants, odds ratio with 95% confidence interval (CI), statistical significance of association (*P*) and estimated population-level penetrance (prevalence) are provided. Statistically significant values are in bold. The penetrance value shown is the expected number of cases in 100,000 live births of that class assuming a population average of 15 cases per 100,000 live births.