Functional Loss of Semaphorin 3C and/or Semaphorin 3D and Their Epistatic Interaction with Ret Are Critical to Hirschsprung Disease Liability

Qian Jiang,[1,2,14] Stacey Arnold,[1,14] Tiffany Heanue,[3,14] Krishna Praneeth Kilambi,[4] Betty Doan,[1] Ashish Kapoor,[1] Albee Yun Ling,[1] Maria X. Sosa,[1] Moltu Guy,[1] Qingguang Jiang,[5] Grzegorz Burzynski,[1] Kristen West,[1] Seneca Bessling,[1] Paola Griseri,[6] Jeanne Amiel,[7,8] Raquel M. Fernandez,[9,10] Joke B.G.M. Verheij,[11] Robert M.W. Hofstra,[12,13] Salud Borrego,[9,10] Stanislas Lyonnet,[7,8] Isabella Ceccherini,[6] Jeffrey J. Gray,[4] Vassilis Pachnis,[3] Andrew S. McCallion,[1] and Aravinda Chakravarti[1,*]

Innervation of the gut is segmentally lost in Hirschsprung disease (HSCR), a consequence of cell-autonomous and non-autonomous defects in enteric neuronal cell differentiation, proliferation, migration, or survival. Rare, high-penetrance coding variants and common, low-penetrance non-coding variants in 13 genes are known to underlie HSCR risk, with the most frequent variants in the ret proto-oncogene (*RET*). We used a genome-wide association (220 trios) and replication (429 trios) study to reveal a second non-coding variant distal to *RET* and a non-coding allele on chromosome 7 within the class 3 Semaphorin gene cluster. Analysis in *Ret* wild-type and *Ret*-null mice demonstrates specific expression of *Sema3a*, *Sema3c*, and *Sema3d* in the enteric nervous system (ENS). In zebrafish embryos, *sema3* knockdowns show reduction of migratory ENS precursors with complete ablation under conjoint *ret* loss of function. Seven candidate receptors of Sema3 proteins are also expressed within the mouse ENS and their expression is also lost in the ENS of *Ret*-null embryos. Sequencing of *SEMA3A*, *SEMA3C*, and *SEMA3D* in 254 HSCR-affected subjects followed by in silico protein structure modeling and functional analyses identified five disease-associated alleles with loss-of-function defects in semaphorin dimerization and binding to their cognate neuropilin and plexin receptors. Thus, semaphorin 3C/3D signaling is an evolutionarily conserved regulator of ENS development whose dys-regulation is a cause of enteric aganglionosis.

Introduction

The origins of the enteric nervous system (ENS) from the neural crest lineage are hypothesized to be an early evolutionary innovation, present in echinoderms and all chordates, antedating the central nervous system (CNS).[1] Often referred to as the "second brain," the ENS exhibits a similarly wide array of neuronal and glial cells as the CNS and responds to the same neurotransmitters. It is, however, quite distinct from, and can mediate reflexes independently of, the CNS. Physiologically, the ENS controls the critical processes of enzyme secretion, ion transport associated with secretion and absorption, blood flow, and peristalsis in the gastrointestinal (GI) tract. These functions are severely compromised in a diverse group of developmental and acquired pathologies including Hirschsprung disease (MIM 142623, 600155, 613711, 613712, 600156, 606874, 606875, 608462, and 611644), hypoganglionosis, neuronal intestinal dysplasia (MIM 243180 and 601223), chronic intestinal pseudo-obstruction (MIM 300048), and irritable bowel syndrome. These disorders have largely been recognized within the past 60 years and their genetic pathophysiology remains poorly understood.

The most common disorder of the ENS at birth is Hirschsprung disease (HSCR), also known as congenital aganglionosis. It is characterized by a lack of enteric ganglia in the myenteric and submucosal plexuses involving varying lengths of the colon, resulting from incomplete rostral-to-caudal enteric colonization.[1,2] With a population incidence of ~15/100,000 live births in newborns of European ancestry and twice that among Asian newborns, this neurodevelopmental birth defect has classical features of a multifactorial genetic disorder: high heritability (>80%), sex difference in incidence (4.0 male:female), high sibling recurrence risk (250-fold greater than the population), and non-Mendelian inheritance in families.[2] HSCR displays a variable phenotype with recurrence risk depending on gender, familiality, associated developmental anomalies,

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [2]Department of Medical Genetics, Capital Institute of Pediatrics, Beijing 100020, China; [3]Division of Molecular Neurobiology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK; [4]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA; [5]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [6]U.O.C. Genetica Medica, Istituto Gaslini, 16147 Genova, Italy; [7]INSERM UMR 1163, Laboratory of Embryology and Genetics of Congenital Malformations, 75015 Paris, France; [8]Paris Descartes University–Sorbonne Paris Cité, Imagine Institute, 75015 Paris, France; [9]Department of Genetics, Reproduction and Fetal Medicine, Institute of Biomedicine of Seville (IBIS), University Hospital Virgen del Rocío/CSIC/University of Seville, 41013 Seville, Spain; [10]Centre for Biomedical Network Research on Rare Diseases (CIBERER), 41013 Seville, Spain; [11]Department of Medical Genetics, University of Groningen, Groningen 9700, Germany; [12]Department of Clinical Genetics, University of Rotterdam, Erasmus Medical Center, 3015 CE Rotterdam, the Netherlands; [13]Neural Development and Gastroenterology Units, UCL Institute of Child Health, London WC1 NEH, UK
[14]These authors contributed equally to this work
*Correspondence: aravinda@jhmi.edu
http://dx.doi.org/10.1016/j.ajhg.2015.02.014.

and segment length of aganglionosis and is classified as short Hirschsprung (S-HSCR), long Hirschsprung (L-HSCR), or total colonic aganglionosis (TCA) based on the length of the aganglionosis. The reasons for the high heritability and phenotypic variation remain enigmatic, although gene discovery has clarified some genotype-phenotype correlations. Genetic studies have identified rare high-penetrance mutations in 12 genes (*RET* [MIM 164761], *GDNF* [MIM 600837], *NRTN* [MIM 602018], *SOX10* [MIM 602229], *EDNRB* [MIM 131244], *EDN3* [MIM 131242], *ECE1* [MIM 600423], *ZEB2* [MIM 605802], *PHOX2B* [MIM 603851], *KIAA1279* [MIM 609367], *TCF4* [MIM 602272], and *L1CAM* [MIM 308840]), cumulatively explaining only a modest number of cases (<10%) and often involving syndromic forms.[2,3] Surprisingly, significantly greater phenotypic variation is explained by two common low-penetrance non-coding variants at *RET* (rs2435357)[4] and *NRG1* (MIM 142445) (rs4541858),[5] respectively. Of these, rs2435357 disrupts function of a GI-tract-specific enhancer in *RET* intron 1 through abrogation of SOX10 binding, thereby reducing *RET* expression.[6]

HSCR is a model complex disease that has provided numerous insights into the nature of non-Mendelian inheritance. In particular, the *RET* enhancer polymorphism not only confers significant biological susceptibility to the most common form of HSCR, S-HSCR, but also explains three phenotypic features of the disorder: (1) variable risk of disease as a function of gender, familiality, and segment length of aganglionosis, (2) allelic interaction with rare coding mutations, and (3) association of HSCR with other clinical syndromes, including Down syndrome, through non-allelic interactions.[6–8] These are surprisingly pervasive genetic effects for a birth defect that was effectively lethal ~60 years ago, prior to surgical treatment, and suggests the effects of natural selection on the enhancer. These results, and similar genetic properties of a *NRG1* common variant in Chinese HSCR-affected subjects,[5] prompted us to search broadly for additional polymorphisms that could further explain HSCR susceptibility and its associations with the common isolated forms.

Here we demonstrate, through human, mouse, and zebrafish analyses, that class 3 Semaphorins are one such major susceptibility factor in HSCR because (1) a common non-coding variant enhances susceptibility in humans, (2) rare deleterious variants in *SEMA3C* (MIM 602645) and *SEMA3D* (MIM 609907) are more frequent than expected in individuals with HSCR, (3) these variants have in vitro loss-of-function defects in semaphorin dimerization and binding to their cognate neuropilin and plexin receptors, (4) genetic interactions between loss of function of class 3 Semaphorins *and* Ret lead to aganglionosis in zebrafish, and (5) all seven receptors of Sema3 proteins are expressed within the embryonic mouse ENS and their expression is lost in the ENS of *Ret*-null mouse mutant embryos. Thus, class 3 Semaphorin signaling is a critical functional module in ENS development, conserved from zebrafish to humans. These proteins act on RET-positive ENS neurons and uncover an additional axis of genetic defects in HSCR epistatic to RET.

Material and Methods

Affected and Control Samples

Affected persons were classified by segment length of aganglionosis into three classes: S-HSCR (aganglionosis up to the upper sigmoid colon), L-HSCR (aganglionosis up to the splenic flexure), and TCA (total colonic aganglionosis). We used a primary GWAS sample of 220 trios comprising one affected individual and both of his/her parents from the USA, Italy, the Netherlands, Spain, and France, and a replication sample of 429 trios from the same 5 countries representing the phenotypic spectrum of HSCR observed in the population and showing variation by gender, segment length, familiality, presence of additional anomalies, and genomic DNA availability. The primary sample was composed primarily of S-HSCR-affected subjects: 180/40 were male/female; 214/3/0/3 were S-HSCR/L-HSCR/TCA/unknown segment length; 196/24 were simplex/multiplex; and 3 probands had additional anomalies (one with Down syndrome [MIM 190685] and two with Shah-Waardenburg syndrome[2] [MIM 277580, 613265, and 613266]). The replication sample was indifferent to segment length (287/141/1 were male/female/indeterminate; 135/90/83/121 were S-HSCR/L-HSCR/TCA/unknown segment length; 298/129/2 were simplex/multiplex/unknown). To identify sequence variants, we studied a phenotypically broader group of 254 subjects consisting of 205, 34, 14, and 1 individuals from the USA, Italy, the Netherlands, and France, respectively, to include the diversity of segment length of aganglionosis (95 S-HSCR, 71 L-HSCR, 85 TCA, and 3 unknown segment [UN]), sex (173/81 were male/female), familiality (168/86 were simplex/familial), and presence of additional anomalies (236/18 were isolated/had additional anomalies). All study individuals self-identified as being of European ancestry. Details of the samples are in Table S1. In addition, we genotyped 1,054 samples representing control subjects across five continents from the CEPH Human Genome Diversity Panel.[9] Subject ascertainment was conducted with written informed consent approved by the Institutional Review Boards of Johns Hopkins University School of Medicine and all participating institutions.

SNP Genotyping

All members of trios were genotyped on the Affymetrix 500K platform, using 250 ng of genomic DNA digested by the restriction enzymes StyI and NspI, hybridized to arrays, and genotypes called by the CRLMM algorithm:[10] 100 trios each were genotyped with the StyI or NspI arrays and 33 trios on both StyI and NspI arrays to allow improved imputation. For quality control, we set the criteria for SNPs and trios to have genotyping call rates > 80%, allele frequency > 5%, Hardy-Weinberg p value > 10^{-5}, and ≤2 Mendelian errors per family. After genotypes were cleaned, we dropped a total of 13 trios (1 due to unsuccessful PCR amplification; 2 owing to poor genotyping call rates; 10 due to >2 Mendelian inconsistencies) to analyze 220 trios (Table S1). The genotyped data on each trio varied between 250K and 500K markers that we first imputed to a common set of 356,308 SNPs above 5% allele frequency and reliably imputed ($r^2 > 0.3$) across all samples by using the reference data from the HapMap CEU sample[11] and the computer program BEAGLE.[12]

Genotypes on individual SNPs, either for replication or for estimating the global control distribution, were obtained via a fluorogenic 5′ nuclease end-point assay on 10 ng of genomic DNA as template and evaluated on an ABI 7900 (Taqman, Life Technologies). Specific genotypes on the three risk markers (rs2435357, Assay ID: C__16017524_10; rs2506030, Assay ID: C__26742714_10; rs12707682, Assay ID: C__30936238_10) were obtained with TaqMan Human Pre-Designed Assays (Applied Biosystems). PCR amplification was performed with PerfeCTa qPCR SuperMix, ROX (Cat# 95051-100, Quanta BioSciences) on 10 ng DNA templates. Enzyme activation and the initial denaturation cycle was at 95°C for 10 min, followed by 40 cycles of 92°C for 15 s and 60°C for 1 min and a final extension at 60°C for 10 min. Plates were read on an ABI PRISM 7900 sequence detection system and the data analyzed with SDS 2.2 software (Applied Biosystems). All SNPs had call rates greater than 98% (rs2435357, rs2506030, rs12707682 had data completeness of 98.4%, 98.0%, and 99.2%, respectively).

DNA Sequencing

For DNA sequencing, genomic DNA was used as a template for PCR amplification of all coding exons in *SEMA3A*, *SEMA3C*, and *SEMA3D* (Table S3). Amplification products were visualized on 2% agarose gels by electrophoresis and purified with QIAquick PCR purification kit (QIAGEN). Subsequently, all amplicons derived from an individual's DNA sample were pooled in a length-weighted equi-volume ratio before library preparation, emPCR, and pyrosequencing on the 454 GS Junior platform as described.[13]

Statistical Analysis

Genome-wide association studies (GWASs) were first performed on 220 trios comprising the primary sample and 356,308 SNPs by comparing case subjects with pseudo-controls (consisting of the two parental alleles not transmitted to the affected) via the Armitage trend test.[14] We used an initial threshold of 0.0001 to choose 10 SNPs for replication in 429 additional trios and a final genome-wide statistical threshold of 1.4×10^{-7}. For final results, we used HapMap[11] reference data to impute the genotypes of all individuals in >2.5 million SNPs in the primary sample. When necessary, LD blocks and r^2 values were estimated with the HapMap CEU data (r28) via the Tagger program in Haploview 4.2.[15] Associated SNPs were further confirmed in the 1000 Genomes data via the SNP Annotation and Proxy Search (SNAP) tool.[16] The data were also analyzed with the TDT (transmission disequilibrium test) by a maximum likelihood method to estimate the risk-allele transmission frequency (τ) and an allelic odds ratio.[6] Penetrance values were estimated as described with the following parameters: incidence of 15 in 100,000 live births; male:female sex ratio of 4.0; 80:10:10 for proportions of S-HSCR, L-HSCR, and TCA probands; and frequency of multiplex cases as 25%.[6]

Gene Expression Assays

Gene expression in human and mouse tissues and across developmental time points was analyzed with assays from Applied Biosystems and real-time PCR analysis on cDNA panels from Clontech.

RNA In Situ Hybridization

Mouse

In analyses of mouse embryos, section or whole-mount RNA in situ hybridization was performed as described.[17,18] The relevant antisense probes were generated as described[18] via probe templates provided by Dr. David Ginty (Johns Hopkins University, Baltimore, MD): *Sema3A*, cut with EcoRI, polymerized with T7; *Sema3C*, KpnI, T7; and *Sema3D*, XhoI, T3; *Sema3E*, SalI, T7. Additional templates were PlexinA1, cut with BamHI, polymerized with T3; PlexinA3, Xba, T3; Plexin A2, XhoI, T3; PlexinA4, EcoRV, T3 (Sakai and Halloran[19]) (the last two kindly provided by Kevin Mitchell); PlexinD1, NotI, SP6 (Banu et al.[20]) (kindly provided by Jeffrey Macklis); Neuropilin1, SacI, T3 (Sakai and Halloran[19]); and Neuropilin2, XhoI, T3 (Huber et al.[21]). Wild-type embryonic and postnatal tissues were isolated from Parkes (outbred) mice. Mice carrying the null $Ret^{k-/k-}$ mutation and their genotyping have been described earlier.[22,23] The day of the vaginal plug was considered to be E0.5. All animal studies were performed under protocols approved by the MRC National Institute for Medical Research Animal Care and Use Committee (UK) and the Johns Hopkins Institutional Animal Care and Use Committee.

Zebrafish

The translation-blocking morpholinos for *sema3aa* and *sema3ab* have been described previously.[24] The translation-blocking morpholino for *sema3c* had sequence 5′-AAAGAGCATGAATGCTGGA GACGCT-3′, and its control mismatch oligo had sequence 5′-AAA GACCATCAATCCTGCACACGCT-3′ (mismatched residues underlined). The splice-blocking morpholino for *sema3d* has been previously described (3DMO2).[25] The translation blocking morpholino for *sema3e* had sequence 5′-TGAAAGTCCACACACCCCACGC CAT-3′. The translation-blocking morpholino for *ret* is as described.[26] All morpholinos were designed and purchased from Gene Tools. Zebrafish embryos were fixed at 4, 5, or 6 dpf in 4% paraformaldehyde/1× fix buffer (4% sucrose, 0.15 mM $CaCl_2$ in 0.1 M PO_4 buffer) for 2 hr. They were then quickly rinsed and subsequently washed three times for 10 min per wash with 1× PBTx (1× PBS with 0.5% Triton X-100). Three additional water washes, 1 hr per wash, were done for 4 and 5 dpf fish, and six additional water washes, 1 hr per wash, were done for 6 dpf fish. Embryos were pre-incubated in blocking solution (1× PBS, 10% heat inactivated goat serum, 0.5% Triton X-100, 2 mg/ml BSA, 1% DMSO, 0.02% NaN_3) for 1 hr at room temperature and then incubated overnight in blocking solution containing a 1:100 dilution of a 0.2 mg/ml monoclonal (16A11) anti-HuC antibody from Invitrogen (catalog #A-21271). They were then quickly rinsed and subsequently washed three times for 30 min per wash with 1× PBTx, followed by incubation for 2 hr in blocking solution with 1:1,000 dilution of the Alexa Fluor 568-conjugated $F(ab')_2$ fragment of goat anti-mouse IgG (H^+L) secondary antibody from Invitrogen (catalog #A11019).

Determination and Refinement of SEMA3C/D 3D Structures

SEMA3C and SEMA3D homology models were constructed via MODELER[27] by employing one of the monomers from the SEMA3A dimer complex (PDB: 4GZ8: A)[28] as template. Ten homology models were constructed for both SEMA3C and SEMA3D and the model with the best DOPE[29] scores were selected for further studies. Selected SEMA3C/D monomer structures were then duplicated and superimposed on individual chains of the SEMA3A dimer to provide starting structures for docking calculations. One thousand models were generated for both SEMA3C and SEMA3D via RosettaDock[30] and the structures with the best interface score, based on the RosettaDock energy function, were selected. Similarly, the crystal structure of Plexin A2-Semaphorin 3A-Neuropilin-1 (PDB: 4GZA)[28] was used as the starting template

for the generation of Semaphorin-Neuropilin-1 and Semaphorin-Plexin A2 complexes. SEMA3C and SEMA3D models were superimposed on the SEMA3A chain in the complex and employed as the starting structures for RosettaDock calculations. The generated models for Semaphorin-Semaphorin, Semaphorin-Neuropilin, and Semaphorin-Plexin complexes were then used for calculation of the effects of individual substitutions on stability and binding.

In residue substitution calculations by Rosetta, the desired side chain was mutated and optimized by discrete sampling of the residue side-chain conformational space using a backbone-dependent rotamer library.[31] The wild-type and variant structures were scored with the standard Rosetta energy function and the difference in their scores ($\Delta G_{variant} - \Delta G_{wild\text{-}type}$) was noted as the free energy change ($\Delta\Delta G$) of a substitution. Protein-protein interfaces in the variant complexes were then optimized by line minimization of the rigid-body orientation of the complex proteins via the RosettaDock score function. Interface scores ($I\_sc = E_{complex} - \sum E_{partners}$) of the mutant complexes were then evaluated and compared to the interface scores of the wild-type complexes ($\Delta I\_sc = I\_sc_{variant} - I\_sc_{wild\text{-}type}$) for prediction of a substitution's effect on binding.

Site-Directed Mutagenesis and Ligand Preparation

Two *SEMA3C* variants were recreated by site-directed mutagenesis as an N-terminal AP-tagged mouse *Sema3c* cDNA in a pCMV-driven pcDNA1 vector. This construct lacks the native signal sequence of *Sema3c* and the resulting alkaline phosphatase (AP) fusion protein can be expressed as a secreted ligand and used directly as a highly sensitive affinity agent. Similarly, three SEMA3D variants were introduced by site-directed mutagenesis into a mouse *Sema3d* cDNA cloned into the 5′ BspEI/3′ XbaI site of a pAPtag-4 vector. The native signal peptide of *Sema3d* was removed. Mutant strand synthesis reactions were set up according to manufacturers' instructions. PCR products were digested with DpnI (Agilent Technologies) and transformed into MC1061/P3 competent *E. coli* (Cat #: C663-03, Invitrogen). Clones harboring the desired single-nucleotide change were verified by DNA sequencing. Each mutant-encoding vector was transiently introduced into HEK293T cells with Lipofectamine 2000 (Life Technologies). Conditioned medium was harvested 72 hr after transfection and ligand concentration determined as described.[32,33] Equal amounts of AP-Sema-transfected HEK293T cell supernatant was concentrated (with Amicon Ultra-0.5 ml 3K column from Millipore) and subjected to 7.5% SDS-PAGE for immunoblotting with Anti-Placental Alkaline Phosphatase, PLAP (Cat #: 13-2355, American Research Products).

COS7 Cell Binding Assays

AP-ligand binding assays were performed as described.[34] In brief, mouse *Neuropilin1* (Nrp1) or *Neuropilin2* (Nrp2) constructs were overexpressed in COS-7 cells after transfection via Lipofectamine 2000. Forty-eight hours after transfection, cells were incubated with various AP-tagged Sema3C or Sema3D (wild-type or mutant fusion protein) containing media for examining receptor binding capacity. Bound AP activity values were normalized for the fusion protein used and reported as fold change relative to wild-type constructs.

Plotting Routines

Results of GWAS analyses, displayed as Manhattan and QQ plots, used software and public genomic data via available methods.[11,35,36]

Results

A Family-Based Genome-wide Association Study

We studied 220 affected child-parent trios of European ancestry with S-HSCR, because the previously identified *RET* and *NRG1* common variants had the largest genetic effect in this category and they are the commonest form of aganglionosis (~80% of cases).[2] The affected offspring had a sex ratio of 4.5 and had HSCR as the sole phenotype in all but three cases (Table S1). We genotyped a maximum of 500,000 SNPs on an Affymetrix platform and excluded all markers with minor allele frequency $< 5\%$, Hardy-Weinberg p value $< 10^{-5}$, data completeness $< 80\%$, and >2 Mendelian errors. The resulting 356,308 high-quality SNPs common to all samples through imputation had frequencies similar to HapMap[11] reference data and were analyzed for allelic association using both the case versus pseudo-control[14] and the transmission disequilibrium tests.[6] The aim was to identify SNPs that exceed a liberal threshold significance ($p = 10^{-4}$) for replication. We identified nine SNPs at seven independent loci: rs12076751 (chr 1), rs10930239 (chr 2), rs839232 (chr 3), rs9405709 (chr 6), rs1228871, rs12707682, and rs11766001 (chr 7), rs2506030 (chr 10), and rs11073865 (chr 15). We also obtained positive signals at the *NRG1* intronic variant rs4541858 and at numerous SNPs in high linkage disequilibrium (LD) with the common enhancer polymorphism (rs2435357). To replicate the loci, we genotyped a total of 11 SNPs in an independent set of 429 trios of European ancestry representing all segment lengths: the 9 variants listed above plus rs4541858 and rs2435357 (that was not on the array nor could be imputed) (Table S1). We declared replication if $p < 0.005$ (corrected for ten tests not including rs2435357). Next, we used HapMap[11] reference data to impute the genotypes of all individuals in >2.5 million SNPs in the primary sample (Figure S1).

The replication study yielded three statistically significant associations in addition to the well-established functional allele at rs2435357 ($p = 5.73 \times 10^{-34}$) (Table 1). These associations identify a second chromosome 10q pericentromeric locus (rs2506030: replication $p = 2.02 \times 10^{-4}$, total $p = 4.44 \times 10^{-15}$) (Figure 1A) and two variants at chromosome 7q: rs12707682 (replication $p = 1.38 \times 10^{-4}$, total $p = 9.16 \times 10^{-10}$) and rs11766001 (replication $p = 4.24 \times 10^{-4}$, total $p = 5.11 \times 10^{-9}$) (Figure 1C). SNP rs4541858, located in intron 1 of *NRG1* and in complete LD with the Asian-associated *NRG1* SNP rs7835688, displayed significant association in the primary screen ($p = 2.94 \times 10^{-5}$), in view of its earlier discovery but was borderline significant in the replication ($p = 0.041$) and not significant in the total sample by genome-wide criteria ($p = 3.25 \times 10^{-5}$). Inspection of the QQ plot of the imputed GWAS results (Figure S1) suggests that once the effects of the RET and 7q loci are accounted for, no other common polymorphisms are evident. Comparison of marker alleles show higher frequencies in cases in the primary GWAS screen than in the replication sample (Table 1). Because these

Table 1. Genetic Associations for Hirschsprung Disease at *RET*, *SEMA3*, and *NRG1*

SNP	Chr: Position (hg19)	Locus	Risk/ Non-risk Allele	Primary GWAS (220 S-HSCR Trios)			Replication GWAS (429 HSCR Trios)			Combined (629 Trios)
				Risk Allele Frequency			Risk Allele Frequency			
				Case	Control	p	Case	Control	p	p
rs11766001	7: 84,145,202	*SEMA3*	C/A	0.34	0.18	9.35×10^{-7}	0.24	0.17	4.24×10^{-4}	5.11×10^{-9}
rs12707682	7: 84,443,356	*SEMA3*	C/T	0.39	0.23	3.52×10^{-7}	0.32	0.24	1.38×10^{-4}	9.16×10^{-10}
rs4541858	8: 32,410,309	*NRG1*	G/A	0.56	0.42	2.94×10^{-5}	0.48	0.42	4.11×10^{-2}	3.25×10^{-5}
rs2506030	10: 43,447,847	*RET*	G/A	0.63	0.35	2.78×10^{-15}	0.55	0.45	2.02×10^{-4}	4.44×10^{-15}
rs2435357	10: 43,582,056	*RET*	T/C	0.67	0.21	2.83×10^{-33}	0.55	0.22	5.73×10^{-34}	5.21×10^{-64}

For each locus, the SNP, genomic location (chromosome, position [hg19]), risk/non-risk allele, case and pseudo-control allele frequency, and association p value is shown for the primary (220 S-HSCR trios) and replication (429 any segment HSCR trios) samples; only p values for the combined sample is given.

two samples differ in the composition of gender, segment length, and familiality of cases, and prior associations were stronger in S-HSCR-affected simplex males than in other forms,[6] it is unknown whether these small frequency differences are from the "winner's curse" or the composition of cases.

Independent, Common Susceptibility Alleles at *RET* and *SEMA3*

SNP rs2506030 with risk allele G (non-risk allele A) has an allele frequency of 0.63 in cases against a background allele frequency of 0.35 in the primary GWAS sample, a difference that is statistically significant ($p = 2.78 \times 10^{-15}$); in the replication sample these frequencies are 0.55 and 0.45, respectively, a smaller yet significant difference between case and control subjects ($p = 2.02 \times 10^{-4}$) (Table 1). In the combined sample, the genotype data reveal the frequency of the risk allele to be 0.58 in case versus 0.41 in control subjects corresponding to an odds ratio of 2.0; the population penetrance of the three genotypes varies 3.8-fold (1.1×10^{-5} to 4.2×10^{-4}) (Table 2). This susceptibility allele is largely independent of the known *RET* enhancer variant as shown by the fact that the correlation between them is small ($r^2 = 0.23$) in our control chromosomes as well as in the HapMap CEU sample. Nevertheless, rs2506030 has a large allelic effect (odds ratio = 2.0) even in relation to rs2435357 (odds ratio = 5.3) and, indeed, most complex traits. We have previously shown that 8%, 66%, and 13% of all HSCR-affected probands have a rare *RET* coding variant only, the rs2435357 risk variant only, or both; in addition, up to 5% of all probands carry a deletion within *RET*.[6] Therefore, with the detection of rs2506030, <5% of HSCR-affected case subjects are free of any known risk variant, leading us to conclude that *RET* mutations are necessary for HSCR in the vast majority of affected persons.

rs2506030 is located in non-coding DNA ~125 kb upstream of *RET*, proximal to *BMS1* (MIM 611448), a ribosome biogenesis protein (Figure 1A). The variant does not lie within a gene nor do gene annotations of this region lead to suspicion of any gene beyond *RET*. Thus, rs2506030, or SNPs in high LD with it, is a second polymorphic potential regulatory element of *RET*. We searched this genomic locus for all known highly associated ($r^2 \geq 0.8$) common polymorphisms (>10%) and annotated this 15.4 kb region for sequence conservation and candidate fetal intestinal enhancers based on H3K4me1 and H3K27ac histone modifications and DNase I hypersensitive sites (Figure 1B). Only rs2506030 directly overlapped a H3K4me1 and a H3K27ac mark but was not evolutionarily conserved except in primates. We hypothesize that this polymorphism is likely to disrupt an enhancer of *RET*.

The second association was evident at chromosome 7q21.11 at rs12707682 (replication $p = 1.38 \times 10^{-4}$) and rs11766001 (replication $p = 4.24 \times 10^{-4}$). The two markers are spatially distinct in a region of very high LD and the SNPs are highly correlated with one another ($r^2 = 0.50$); we do not consider rs11766001 to be an independent signal (Figure 1C). The peak SNP rs12707682 has risk allele C (non-risk allele T) and an allele frequency of 0.39 in case subjects and 0.23 in control subjects in the primary GWAS sample, a difference that is statistically significant given our screening threshold ($p = 3.52 \times 10^{-7}$) (Table 1). In the replication sample, the genotype data reveal the frequency of the risk allele to be 0.32 in case subjects versus 0.24 in control subjects corresponding to an odds ratio of 1.8; the population penetrance of the three genotypes varies 3.1-fold (1.4×10^{-5} to 4.3×10^{-4}) (Table 2). This region gained interest because it contained a family of intriguing candidate genes, four related members of class 3 Semaphorins that are critical to neuronal development.[37] The sentinel SNP was located between the genes encoding Semaphorin 3A (*SEMA3A* [MIM 603961]) and 3D (*SEMA3D* [MIM 609907]), ~1.16 mb upstream of Semaphorin 3E (*SEMA3E* [MIM 608116]), and at a considerable distance of ~3.89 Mb upstream of Semaphorin 3C (*SEMA3C* [MIM 602645]).

We searched this large candidate genomic region for one or more variants that might explain the association. We identified a 139 kb segment that contained all common polymorphisms (>10%) highly associated ($r^2 > 0.8$) with rs12707682 and annotated this region for sequence conservation and candidate fetal intestinal enhancers based

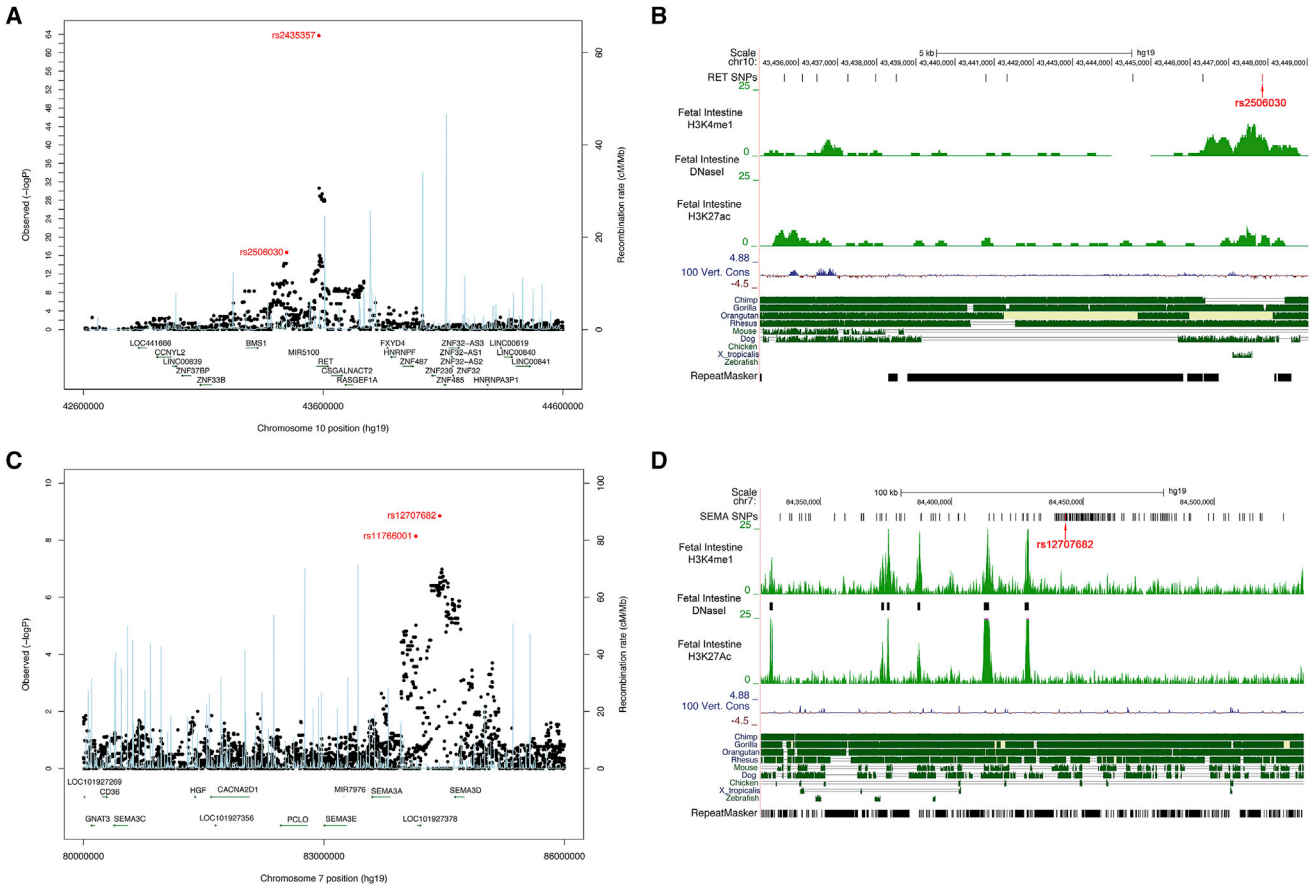

Figure 1. Genetic Associations at the *RET* and *SEMA3* Loci

(A) Black dots represent $-\log_{10}$ p values (y axis, left) for all SNPs studied plotted against their hg19 genomic location (with recombination rate plotted in light blue (y axis, right). The leading variants in the combined data are highlighted in red. *RET* and additional genes near the peak SNP rs2506030 are depicted at the bottom of the window with arrows indicating the direction of transcription; the highest peak is at the previously described enhancer variant rs2435357; the new peak is at rs2506030.

(B) The 15.4 kb genomic region of high ($r^2 > 0.8$) linkage disequilibrium around SNP rs2506030 showing locations of all known common (>10%) SNPs, sequence conservation, and enhancer marks from human male fetal gut (DNase I hypersensitive sites, H3K4me1, H3K27ac).

(C) The four *SEMA3* genes near the peak SNPs rs12707682 and rs11766001, with legend as in (A).

(D) The 219 kb genomic region of high ($r^2 > 0.8$) linkage disequilibrium around rs12707682 is depicted, with legend as in (A).

on H3K4me1 and H3K27ac histone modifications and DNase I hypersensitive sites (Figure 1D). We did not detect any overlap between these polymorphisms and any regulatory feature despite our suspicion that rs12707682 or its surrogates probably disrupt function of a non-coding regulatory element in the GI tract. Fine mapping of this region in additional cases and mapping regulatory elements at other GI developmental time-points might be required for an answer.

The risk alleles at *RET* (rs2435357 and rs2506030) and at *SEMA3* (rs12707682) are common in the general European ancestry population with allele frequencies of 24%, 39%, and 26%, respectively, in the HapMap CEU samples[6,11] and with odds ratios of 5.3, 2.0, and 1.8, respectively, and therefore have significant impact on disease risk. rs2435357 is significantly associated with the more common attributes of HSCR: male gender, short segment length, and simplex familiality;[6] this is also the case for the *NRG1* variant in Asians.[5] Consequently, we asked whether the same was true for the rs2506030 and rs12707682 variants. These analyses demonstrate statistically significant differences in risk according to gender (higher in males: 3.51×10^{-4}), segment length (higher in S-HSCR: 1.21×10^{-6}), and familiality (higher in simplex cases: 1.51×10^{-2}) for *RET* rs2506030 but no differences for *SEMA3* rs12707682 ($p > 0.42$) (Table S2). These differences can be translated into estimates of population-level penetrance by genotype to show that the rs2506030 effect is largely through risk-allele homozygotes, whereas non-significant rs12707682 risks are larger for both homozygotes and heterozygotes. In other words, the genotype-phenotype correlation between HSCR is so far demonstrable only for both *RET* risk polymorphisms.

Functions of Class 3 Semaphorins in Mice and Zebrafish

The C risk allele at rs12707682 is widespread in humanity: we genotyped control samples from the Human Genome

Table 2. Population-Level Genotype Penetrance at Variants *RET* and *SEMA3* in the Combined Sample of 649 Trios

Genotype	GWAS and Replication Trios		
	No. of Cases	No. of Controls	Disease Penetrance
	RET rs2506030 (n = 586)		
AA	109	207	1.1×10^{-4}
AG	272	273	1.9×10^{-4}
GG	205	106	4.2×10^{-4}
Risk allele (G) frequency	0.58	0.41	–
Allelic odds ratio	2.0		
	SEMA rs12707682 (n = 626)		
TT	266	367	1.4×10^{-4}
CT	288	225	2.6×10^{-4}
CC	72	34	4.3×10^{-4}
Risk allele (C) frequency	0.35	0.23	–
Allelic odds ratio	1.8		

The observed numbers of genotypes in cases and pseudo-controls are shown with penetrance calculated under Hardy-Weinberg expectations with pseudo-control allele frequencies and assuming a population incidence of 15/100,000 births. The allelic odds ratio is estimated from the case and pseudo-control allele frequencies.

Diversity Panel to find that the risk allele is absent only in Oceania (Figure S2). The risk allele frequency is higher in Asians than in Europeans, analogous to rs2435357, and proportional to their respective disease incidence.[2,4] However, unlike rs2435357, rs12707682 is as frequent in Africa as in Asia. Consequently, given this global risk, we wished to identify which, if any, of the class 3 Semaphorins can lead to HSCR.

Semaphorins are transmembrane, secreted, or GPI-linked proteins known to be involved in neuronal migration, proliferation, survival, or axonal guidance.[37] We first asked which of *SEMA3A*, *SEMA3C*, *SEMA3D*, and *SEMA3E* are expressed in the human digestive system? We analyzed gene expression via real-time PCR in fetal and adult mouse and human tissues, in comparison to *RET* expression and after normalization to *GAPDH* (MIM 138400). Importantly, signals for all four class 3 Semaphorins are present in gut tissues, although the strongest signals are from *SEMA3A* and *SEMA3C* and the weakest from *SEMA3D* (Figure S3). All four genes produced a similar pattern of expression in the GI tract with none mimicking *RET*. Similarly, no clear temporal or spatial mimic was evident from analysis of the remaining tissue panels. To determine the in vivo spatial and temporal expression patterns during GI tract development, we performed RNA in situ hybridization of *Sema3A*, *Sema3C*, *Sema3D*, and *Sema3E* in the developing mouse gut relative to *Ret*. *Ret* is expressed in enteric neural crest-derived cells (ENCCs) during phases of ENCC migration and colonization of the developing gut and is prominent at E11.5 (Figure 2A, *Ret* panel) and persists in the ENS during later stages where it is visualized in the myenteric plexus of the intestine, which is situated within the outer muscular layers of the gut (Figure 2A, *Ret* panel). At postnatal stages, expression of *Ret* persists in the myenteric plexus (Figure 2A, *Ret* panel, white arrowheads) and is also seen in the newly forming sub-mucosal plexus (Figure 2A, *Ret* panel, yellow arrowheads). Specific expression of candidate genes within the ENS can be verified by using the $Ret^{k-/k-}$ mouse strain as shown by the fact that the ENS is absent from the intestine of these embryos.[38,39]

We find that *Sema3A* is expressed within the developing gut (Figure 2A, *Sema3A* panel, asterisk) but not within areas that correspond to the developing ENS. In agreement with this observation, expression of *Sema3A* is unchanged when comparing wild-type and $Ret^{k-/k-}$ intestines (Figure 2B, *Sema3A* panel). Expression of *Sema3C* is detected within a small population of ENS cells in the E15.5 gut (Figure 2A, *Sema3C* panel, arrowheads) but cannot be detected at the earlier or later time points (Figure 2A, *Sema3C* panel). ENS expression of *Sema3C* at E15.5 is further confirmed by the fact that expression is absent from the $Ret^{k-/k-}$ intestine (Figure 2B, *Sema3C* panel). *Sema3D* is also expressed within a subset of ENCCs and can be detected within the E11.5 gut (Figure 2A, *Sema3D* panel, arrowheads) and persists within a subset of ENS at E15.5 and P3 (Figure 2A, *Sema3D* panel, arrowheads). ENS expression of *Sema3D* is also confirmed by the absence of expression in the $Ret^{k-/k-}$ intestine (Figure 2B, *Sema3D* panel). Finally, *Sema3E* expression is detected within the ENS only at P3 (Figure 2A, *Sema3E* panel, arrowheads) and cannot be detected at the earlier stages (Figure 2A, *Sema3E* panel). Consequently, it is absent from the wild-type and $Ret^{k-/k-}$ embryos at E15.5 (Figure 2B, *Sema3E* panel). These observations, in comparison to overlapping temporal and spatial localization with *Ret*, suggest that *Sema3d* is the best candidate for a role in HSCR with a probable role for *Sema3c* as well. Although *Sema3a* is not expressed in the ENS, it remains a HSCR candidate because it could affect migration of ENCCs in the developing gut.[40]

To confirm the role of class 3 semaphorins in ENCC migration, we used morpholino-based gene expression knockdowns in zebrafish to assess whether loss of *sema3c* and *sema3d* function also produces aganglionosis. First, knockdown of *ret* in zebrafish embryos leads to a loss of ENS neurons at 4–6 dpf, similar to that seen in the $Ret^{k-/k-}$ mouse embryos (Figures 3A, 3D, and 3E).[26] Two orthologs for *SEMA3A* exist in zebrafish and, therefore, morpholinos for both were co-injected for analysis of ENS expression. Use of 4.5 ng of each translation blocker (one targeting *sema3aa* and the other targeting *sema3ab*) failed to produce a loss of neurons and was indistinguishable from un-injected controls; similarly, injection of 4.2 ng of a translation blocker for the sole *SEMA3E* ortholog, *sema3e*, had no apparent effect on the ENS population of the zebrafish gut (data not shown). Conversely, injection of morpholinos for *sema3d* or *sema3c* produced

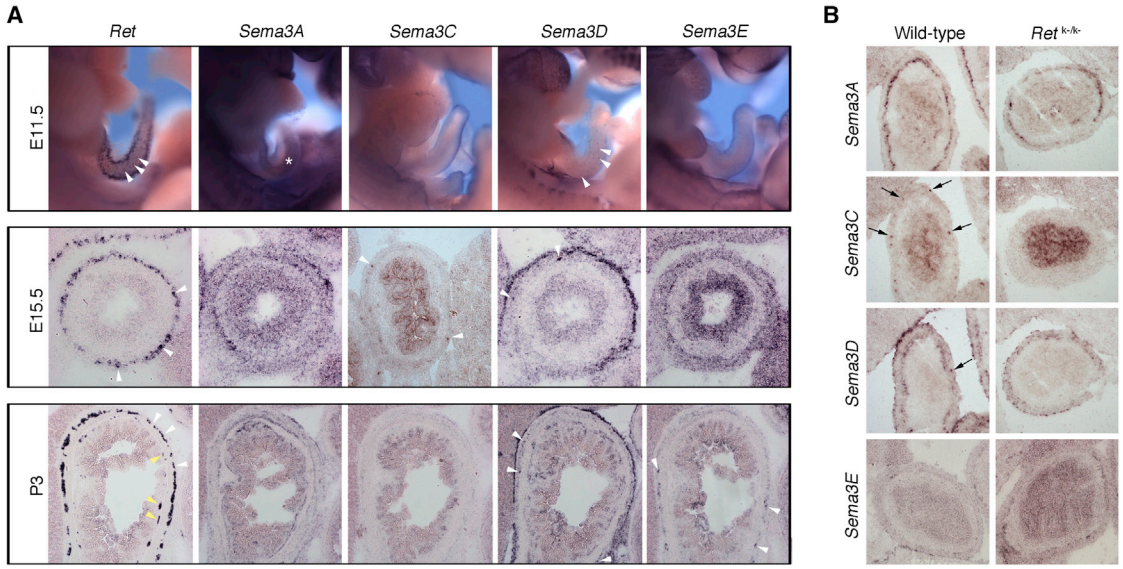

Figure 2. Expression of *Sema3A*, *Sema3C*, *Sema3D*, and *Sema3E* in the Developing Mouse Gut
(A) RNA in situ hybridization analysis of *Sema3A*, *Sema3C*, *Sema3D*, and *Sema3E* in the developing mouse gastrointestinal tract relative to *Ret*. Expression analysis is shown for *Ret*, *Sema3A*, *Sema3C*, *Sema3D*, and *Sema3E* in the developing gut at stage E11.5, in E15.5 intestine sections, and in P3 intestine sections.
(B) *Sema3C* and *Sema3D* gene expression patterns in wild-type and $Ret^{k-/k-}$ mutant embryos. Comparisons of *Sema3A*, *Sema3C*, *Sema3D*, and *Sema3E* on intestinal sections from E15.5 wild-type and $Ret^{k-/k-}$ mutant embryos are shown.

profound effects on the zebrafish GI tract (5 dpf shown but effects seen at all time points). A splice-blocking morpholino for the sole *SEMA3D* ortholog *sema3d* had a modest effect on gut innervation at a dosage of 2.1 ng (Figure 3B) but a severe impact at twice this dosage (Figure 3C); a similar dose effect was seen for a translation-blocking *ret* morpholino at 1 versus 2 ng (Figures 3D and 3E). A translation-blocking morpholino (*sema3c*) for the sole *SEMA3C* ortholog had a strong effect on gut innervation even at a low dosage of 1.1 ng (Figure 3F). As a control, the highest tested dosages of both *sema3c* and *sema3d* were independently co-injected with 1 ng of a morpholino targeting *p53* in an effort to counter any nonspecific apoptosis stemming from cytotoxicity.[41] Neither co-injection reversed the loss of enteric neurons (data not shown).

The gene expression studies in *Ret* wild-type and null mouse embryos suggest an ENS localization for *Sema3d/c* and, thus, their wild-type gene expression is contingent on wild-type *Ret* function. We have previously shown, in a HSCR mouse model, that reduced dosages of *Ret* and *Ednrb* mutations that independently yield no enteric phenotype do produce aganglionosis when combined in an oligogenic mutant mouse.[22,23] Consequently, we co-injected morpholinos for either *sema3c* or *sema3d* together with a *ret* morpholino in zebrafish embryos to test for an analogous genetic interaction between apparently independent pathways. Co-injection of the lower, mildly detrimental dosages of *sema3d* and *ret* abrogated gut innervation (Figures 3B, 3D, and 3G). However, co-injection of the potent *sema3c* morpholino at a lower *ret* dosage had no further effect on the ENS population than that mediated by *sema3c* alone (Figures 3D, 3F, and 3H): co-injection with 2 ng of *ret* was required to eliminate gut innervation (Figure 3I). These data suggest epistatic interactions between loss of function of *sema3c/d* and *ret* in zebrafish GI tract development. Given its more profound interaction with *ret*, *sema3d* again appears to be the most likely HSCR candidate with a smaller effect from *sema3c*. These observations are consistent with the mouse gene expression data (Figure 2).

Sequence Variants of *SEMA3* in HSCR

To assess the role of class 3 Semaphorins in HSCR, beyond the genetic association, we performed targeted next-generation sequencing of the gut-expressed *SEMA3A*, *SEMA3C*, and *SEMA3D* genes in 254 diverse HSCR-affected case subjects of European ancestry and identified 12 non-synonymous variants (Table 3, Table S3, Figure S4). Among these, five variants in SEMA3C (p.Ser329Gly, p.Val337Met) and SEMA3D (p.His424Gln, p.Val457Ile, p.Pro615Thr) were selected as being likely causal based on our mouse and zebrafish results and their locations at highly conserved amino acid residues with their seven orthologs and seven paralogs (SEMA3A through SEMA3G) (Figures S5 and S6). In addition, these variants were clustered in either the Sema or the Ig-like domain, both demonstrated to be important for Sema structure and function, and were located in or close to their protein interfaces, important for Sema homo-dimerization and binding to their receptors (Figures S4 and S6). We have laid greater weight on structural conservation of class 3 Semaphorins rather than features predicted from genomic sequence alone. These biochemical features are, not exclusively, correlated with the variants' conservation (PhyloP) and predicted

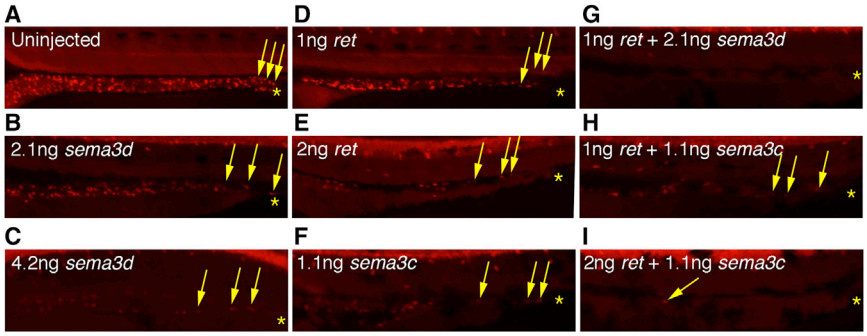

Figure 3. Synergistic Effects of Expression Knockdown of *sema3c/sema3d* and *ret* during Zebrafish Embryogenesis

Morpholino co-knock-down of *ret* with individual *sema3* orthologs to test for individual and epistatic effects on the ENS was measured at 5 dpf (days post fertilization) by immunofluorescence using the HuC primary antibody and an Alexa Fluor 568-tagged secondary antibody to observe the extent of colonization by ENS neurons. Asterisks denote posterior end of gut tube, and arrows denote position of the most posterior HuC-positive enteric neurons. The uninjected control shows complete colonization up to the cloacal end (A). The splice-blocking morpholino against *sema3d* had a modest effect on gut innervation at a dosage of 2.1 ng (B) but a severe impact at twice this dosage (C). A similar effect can be seen for a translation-blocking morpholino against *ret* at 1 (D) versus 2 (E) ng. Finally, a translation-blocking morpholino against *sema3c* has a strong effect on gut innervation even at a low dosage of 1.1 ng (F). Combination of morpholinos at the lower doses of *sema3d* (2.1 ng) and *ret* (1 ng) eliminates gut innervation (G) but required a higher dosage of *ret* (2 ng) when combined with *sema3c* (1.1 ng) to eliminate gut innervation (compare H and I).

deleterious effect (Condel) (Table 3). The five variants are rare in public databases (NHLBI Exome Sequencing Project, Exome Variant Server) and are generally over-represented in case subjects as compared to control subjects. The total frequency of the five variants is 4.7%, 1.7%, and 1.4% in S-HSCR-affected (n = 95), L-HSCR/TCA-affected (n = 156), and EVS control (n = 4,300) subjects, respectively (Table 3). The ~3-fold frequency difference between S-HSCR and L-HSCR/TCA is modestly statistically significant ($p = 0.039$) and suggests that *SEMA3C/D* effects might be primarily restricted to S-HSCR-affected case subjects, the group in whom the genetic association was first detected.

The reported zebrafish studies have shown genetic interaction between loss of function of *Ret* and loss of function of *Sema3c/d*: we asked whether this effect is recapitulated in HSCR. Thus, we studied the risk alleles at rs2435357, rs2506030, and rs12707682 and all 12 identified rare coding variants at *SEMA3A*, *SEMA3C*, and *SEMA3D* in the 35 HSCR-affected subjects who harbored them (Table S4). We note three features. First, HSCR-affected subjects have a higher allele frequency of the three common alleles even in the subset with coding variants (Table S5). Second, rs2435357 frequency is larger for subjects with class 3 semaphorin variants than in those without. Third, individuals with rare class 3 semaphorin coding variants tend to have specific phenotypic features: S-HSCR cases are largely simplex, display no apparent parent-of-origin effect in *SEMA3* variant inheritance (six inherited from mother versus eight from father), and 25% frequency of deleterious *RET* coding mutations (4 versus 12 cases have/do not have coding mutations). In contrast, half of the L-HSCR/TCA-affected subjects have positive familial history (9/18), a trend toward maternal inheritance of *SEMA3* variants (11 from mother versus 4 from father), and a 53% frequency of deleterious *RET* coding mutations (9 versus 8 cases have/do not have coding variants). These results are all consistent with genetic interaction between *RET* and *SEMA3*. However, none of these comparisons are statistically significant but they are trends that need to be investigated in much larger samples.

Functional Analysis of *SEMA3C/3D* Variants in HSCR

We investigated how the non-synonymous variants identified might influence class 3 Semaphorin functions, primarily their effect on the stability and binding of various SEMA3C and SEMA3D complexes. First, we examined the predicted effects of each variant on semaphorin dimerization or binding to plexins and neuropilins, by using computational homology modeling, docking, and mutation algorithms based on published 3D structures (Table 4, Figure 4).[28] We estimated ΔI_sc and ΔΔG values that signify the effects of the substitutions on binding and stability of Semaphorin complexes. We find that both p.Ser329Gly and p.Val337Met substitutions in SEMA3C are more proximal to the Semaphorin-Semaphorin interface and, hence, do not directly affect Semaphorin-Neuropilin and Semaphorin-Plexin interfaces. The p.Ser329Gly variant is found to destabilize the Semaphorin-Semaphorin interface as evidenced by a ΔI_sc of 1.158. Also, as demonstrated by the ΔΔG values, p.Ser329Gly mildly decreases whereas p.Val337Met highly decreases SEMA stability, thereby potentially indirectly influencing its binding. Similarly, SEMA3D p.His424Gln and p.Pro615Thr have non-zero ΔI_sc values, confirming their location at the Semaphorin-Semaphorin interface, possibly decreasing the binding affinity. In addition, p.His424Gln and p.Val457Ile are found to decrease Semaphorin stability, whereas p.Pro615Thr is found to increase Semaphorin stability. Finally, p.His424Gln is the only substitution predicted to have a destabilizing effect on the Semaphorin-Plexin interface.

These predictions suggested that cellular functional analyses of these sequence variants would be instructive. We used transient transfections of wild-type and mutant SEMA3C/D fusion protein constructs into HEK293T cells to measure concentrations of the secreted Sema ligands via colorimetric assays, which were then confirmed by

Table 3. Features of 12 *SEMA3A*, *SEMA3C*, and *SEMA3D* Missense Variants in 254 HSCR-Affected Subjects

Sequence Variants						Variant Allele Frequency (%)		
Gene & Exon	Location (hg19)	dbSNP ID	Substitution	Conservation Score (PhyloP)	Predicted Genetic Effect (Condel)	S-HSCR (n = 95)	L-HSCR & TCA (n = 156)	EVS (n = 4,300)
SEMA3A E2	chr7: 83,764,220	–	p.Ser54Gly	4.48	deleterious	0.0	0.7	–
SEMA3A E11	chr7: 83,634,712	rs147436181	p.Val435Ile	4.00	deleterious	1.1	1.0	1.4
SEMA3C E9	chr7: 80,430,074	rs199738991	p.Ser329Gly	2.68	neutral	0.5	0.0	0.035
SEMA3C E10	chr7: 80,427,530	rs1527482	p.Val337Met	4.55	deleterious	2.1	1.4	0.49
SEMA3D E2	chr7: 84,727,240	rs147821428	p.Ser65Pro	−0.08	neutral	2.2	3.7	2.3
SEMA3D E6	chr7: 84,694,794	rs148644955	p.Thr222Ala	0.67	neutral	1.1	0.0	0.081
SEMA3D E11	chr7: 84,651,849	rs141893504	p.His424Gln	0.15	deleterious	0.5	0.0	0.52
SEMA3D E11	chr7: 84,651,752	rs142496795	p.Val457Ile	6.18	deleterious	0.5	0.0	0.012
SEMA3D E16	chr7: 84,636,183	rs117730916	p.Pro615Thr	4.60	deleterious	1.1	0.3	0.31
SEMA3D E16	chr7: 84,636,125	rs370785183	p.Arg634Gln	1.86	neutral	0.5	0.0	0.023
SEMA3D E17	chr7: 84,629,063	rs34138570	p.Ile676Thr	4.77	neutral	0.5	0.0	0.21
SEMA3D E17	chr7: 84,628,989	rs7800072	p.Lys701Gln	0.84	neutral	34.0	38.0	32.7

The table lists gene/exon name, location (hg19), dbSNP ID, amino acid substitution, conservation score (PhyloP vertebrate 46-way comparison), predicted deleterious effect (Condel), and variant allele frequency in S-HSCR-affected subjects (n = 95), L-HSCR/TCA-affected subjects (n = 156), and unaffected adult controls (EVS, n = 4,300). All comparisons are for individuals of European ancestry.

immunoblotting for each of the variant proteins. Overall, the five HSCR variants showed marked reductions in protein secretion ($n = 5$, one-way ANOVA, $p < 0.001$) (Figures 5A and 5B). However, the effects were not uniform and were significant for p.Val337Met in SEMA3C (Figure 5A) and for p.His424Gln and p.Val457Ile in SEMA3D (Figure 5B) with fold changes (relative to the wild-type construct) of 0.45, 0.62, and 0.87, respectively. To further verify these fusion protein abundance levels, equal amounts of alkaline phosphatase (AP)-SEMA-transfected HEK293T cells supernatant was concentrated (~25-fold) and subjected to 7.5% SDS-PAGE. The immunoblotting analyses confirmed the abundance level changes of each of these variant proteins. To rule out the possibility of differential transfection rates of the different constructs leading to fusion protein level fluctuations, we also performed double transfections (with AP-SEMA and FLAG-Hoxb7) in HEK293T cells and performed immunoblotting of the cell lysate. The exposure of the same film to FLAG-antibody and actin-antibody detected similar amounts of the corresponding molecule, across the different parallel transfections, showing transfection rates to be comparable (Figure 5C).

Finally, we sought to test the role of the five substitutions on the ability of SEMA3C or SEMA3D to function as a ligand. To determine whether each variant fusion protein could bind Neuropilin, the high-affinity obligatory binding partner in the Semaphorin holoreceptor complex, AP-SEMA (wild-type or variant)-containing media (quantified by measuring AP activity) was applied to COS-7 cells overexpressing mouse Neuropilin 1 or 2 (Nrp1, Nrp2). We tested all five variants: Sema3C p.Ser329Gly did not show any effect in ligand concentration and we observed reduced binding affinity of the cognate SEMA protein to both Nrp1 and Nrp2, but to different extents (Figures 5D–5G). Specifically, p.Ser329Gly and p.Val337Met in SEMA3C exhibited fold changes in binding affinity to Nrp1 of 1.12 and 0.37 (Figure 5D); p.His424Gln, p.Val457Ile, and p.Pro615Thr in SEMA3D showed fold changes of 0.56, 0.51, and 0.66, respectively (Figure 5E). With respect to Nrp2, these cognate SEMA proteins displayed binding affinity changes as of 1.18, 0.37, 0.76, 0.65, and 0.85, respectively, relative to the wild-type construct; results shown are averages and SEM of three independent experiments (one-way ANOVA, $^{***}p < 0.001$, $^{**}p < 0.01$).

Receptors of Sema3 Proteins Are HSCR Candidates

Vertebrate Sema3 proteins are known to interact with five specific Plexins (PlexinA1, PlexinA2, PlexinA3, PlexinA4, and PlexinD1) and both Neuropilin1 and Neuropilin2. To determine which of these putative Sema3 receptor components are candidates for Sema3c and Sema3d signaling in the ENS, we examined the expression patterns of these genes within the embryonic mouse ENS relative to *Ret*. In cross-sections of the E15.5 gut, expression of *Ret* is observed in a nearly complete ring in the outer layers of the intestine (Figure 6A). High-magnification view of this myenteric region shows clear expression of *Ret* in individual ENS neurons (Figure 6B, *Ret* panel). In $Ret^{k-/k-}$ gut sections, which lack ENS cells, *Ret* expression is absent (Figure 6B, *Ret* panel). Analysis of Plexin and Neuropilin expression patterns (Figures 6B and 6C; Wild-type panels) shows that most of the known Sema3 receptor components, *PlexinA1-A4* and *Neuropilin1-2 (Nrp1, 2)*, are

Table 4. Predicted Effects of Substitutions on the Stability and Binding of Semaphorin Complexes

	Sema-Sema Complex		Sema-Neuropilin Complex		Sema-Plexin Complex	
Substitution	**ΔI_sc**	**ΔΔG**	**ΔI_sc**	**ΔΔG**	**ΔI_sc**	**ΔΔG**
Sema3C: p.Ser329Gly	1.158	5.986	0	1.974	0	1.974
Sema3C: p.Val337Met	0	30.256	0	15.067	0	15.067
Sema3D: p.His424Gln	0.005	5.038	0	2.420	0.023	2.592
Sema3D: p.Val457Ile	0	0.613	0	0.305	0	0.305
Sema3D: p.Pro615Thr	0.002	−92.941	0	−46.458	0	−46.458

The change in the free energy upon substitution ($\Delta\Delta G = \Delta G_{variant} - \Delta G_{wild\text{-}type}$) calculated via the standard Rosetta energy function estimates the variant's effect on stability of the Semaphorin complexes. The change in the interface score upon substitution ($\Delta I\_sc = I\_sc_{variant} - I\_sc_{wild\text{-}type}$) predicts the variant's effect on binding between semaphorins and neuropilins or plexins, respectively. All free energies are listed in Rosetta Energy Units (REU).

expressed in the ENS. An absence of corresponding expression patterns in the *Ret*$^{k-/k-}$ gut enables us to confirm that the wild-type expression domains are within the ENS (Figures 6B and 6C; *Ret*$^{k-/k-}$ panels). In contrast, although *PlexinD1* is expressed broadly within the gut, it does not appear to be expressed in the ENS (Figure 6C; Wild-type panel). Consistent with this observation, the pattern of *PlexinD1* appears unchanged in *Ret*$^{k-/k-}$ gut sections (Figure 6C; *Ret*$^{k-/k-}$ panel). We observe *Nrp2* expression within the ENS in contrast to a previously published study gut.[40] We note, however, that *Nrp2* expression in that study used a gene-trap reporter, rather than direct detection of RNA or protein, and that independent analysis from the GenePaint and Eurexpress gene expression databases also show ENS expression of *Nrp2*. Given the fact that Semaphorin receptor components are expressed in the ENS, they too are candidate genes for HSCR.

Discussion

Major mutations leading to HSCR are invariably in genes, expressed in neuroblasts or mesenchymal cells, involved in the earliest stages of fate determination of ENCCs and their subsequent survival, proliferation, migration, and differentiation.[1–3,42] These include components of the RET signaling pathway, affecting their migration down the GI tract via a GDNF gradient, and the EDNRB signaling pathway, modulating migration as well as proliferation through EDN3-mediated inhibition of differentiation. Regulating these genes are three major transcription factors, SOX10, PHOX2B, and ZEB2, influencing the timing, migration, and survival of ENCCs.[42] It is, therefore, unsurprising that deleterious heterozygous mutations in these genes have high penetrance for HSCR or result in syndromic forms, being early developmental blocks in the ENS. What is remarkable is that a large majority (79%) of HSCR-affected individuals have a low-penetrance *RET* enhancer mutation.[6] Studies described here have now identified a second such *RET* variant, suggesting that reduced *RET* function is necessary but might not be sufficient for aganglionosis. Therefore, we hypothesize that clinical disease occurs only when pathways additional to RET also have compromised function. We have previously shown that coordinate loss of function of *RET* and *EDNRB* contributes to human disease and recapitulates aganglionosis in a mouse model.[22,23] We now demonstrate that loss-of-function genetic defects in *SEMA3C/D* are an additional major pathway in HSCR and ENS development.

Class 3 Semaphorins are secreted glycoproteins that signal through neuropilin and plexin receptors and contain a ~400 amino acid "Sema," a cysteine-rich PSI (plexins, semaphorins, and integrins), and a C2 type immunoglobulin (IG) domain.[43] Although initially characterized as neuronal guidance cues, signaling of semaphorin ligands is now known to be major contributors to homeostasis and morphogenesis of many tissues and is widely studied for its role in cell migration, proliferation, and survival, neuronal connectivity, and cardiac development.[44–46] Sema3s are the only vertebrate-secreted semaphorins and consist of seven family members, SEMA3A through SEMA3G. Their obligate co-receptors are Neuropilin1 or Neuropilin2, which, together with a Plexin receptor (PlexinA1 through PlexinA4, PlexinD1), form an active holoreceptor complex for most Sema3s to transduce biochemical responses in distinct neuronal subtypes.[47] The studies described here clearly demonstrate that Sema3s, and perhaps their cognate receptors, have significant roles in ENS development.

We have provided three pieces of information that demonstrate a specific role of *SEMA3C/D* in HSCR: (1) mapping of a common susceptibility allele adjacent to *SEMA3D* and identification of rare hypomorphic mutations in affected individuals, (2) their gene expression in the human, mouse, and zebrafish GI tract and specifically the ENS, and (3) the synergistic effect of *Ret* and *Sema3c/d* loss of function on aganglionosis in model systems. A previous RNA microarray study compared gene expression in wild-type and *Ret*$^{k-/k-}$ mouse GI tracts and identified three ENS-localized genes that are semaphorin signaling components: *Crmp1*, *Dpysl3 (Crmp4)*, and *Nrp1*.[48] These data add to the importance of SEMA3D signaling in the ENS since in zebrafish, *sema3d* has been shown to mediate repulsive cues for axonal guidance through binding with *nrp1a*, an

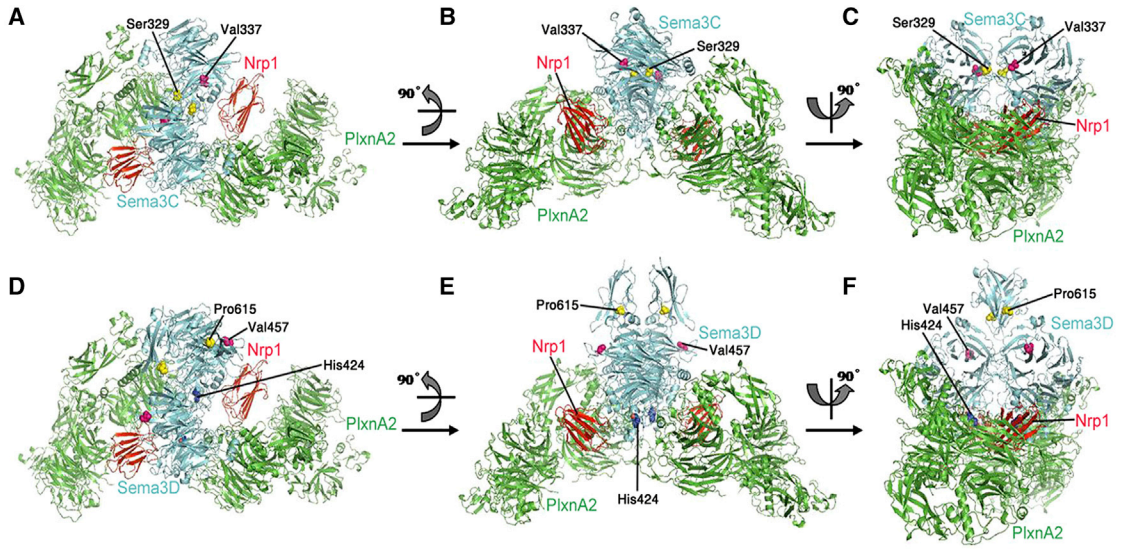

Figure 4. In Silico Modeling of Semaphorin3 with Their Neuropilin and Plexin Complexes
Models of the Sema3C/Neuropilin1/PlexinA2 (A, B, C) and Sema3D/Neuropilin1/PlexinA2 (D, E, F) complexes in top view (A, D), side view (B, E), and front view (C, F) are shown, with the Sema3 protomers colored in cyan, Neuropilin1 in golden brown, and PlexinA2 in green. The five residues variant in HSCR-affected subjects are colored (yellow for Ser329 and magenta for Val337 in Sema3C; dark blue for His424, magenta for Val457, and yellow for Pro615 in Sema3D) for visualization.

ortholog of *Nrp1*,[49] and acts synergistically with *dpysl3* to promote peripheral axon outgrowth.[50] Other class 3 Semaphorins might also have a role in the GI tract; *Sema3a* has previously been shown to be expressed in the chick rectal wall,[51] and in the mouse hindgut mesenchyme.[40] In contrast, our results do not confirm a role for *Sema3a* or *Sema3e* in the ENS but they might have indirect roles. Mice lacking *Sema3a* display premature entry of sacral enteric neural precursors (ENPs) into the distal hindgut, indicating a potential repulsive role. However, the contribution of sacral ENPs to the ENS is small; the largest contribution comes from vagal ENPs, and although *Sema3a* is highly expressed in the mouse caecum and migrating vagal ENPs express *Nrp1*, there is no effect on the timing or magnitude of ENS colonization by this population in *Sema3a*$^{-/-}$ mice.[40] The possibility exists, to the contrary, that it is *Sema3d* that exerts an influence on vagal ENPs via *Nrp1*. The *Sema3d* knockout mouse has yet to be examined for ENS defects and our genetic interaction data suggest that these defects might manifest in *Sema3d* mutant mice only with concomitant *Ret* loss of function.

The mechanism of action of *Sema3c/d* in the ENS is unknown but both have been implicated in patterning defects during heart development, a not-infrequent feature of HSCR.[2,3] First, *Sema3c* mutant mice die within hours after birth from interruption of the aortic arch and improper septation of the cardiac outflow tract; *Sema3d* mutant mice have anomalous pulmonary venous connections and a SEMA3D p.Pro615Thr variant, identical to one we identified in this study, has been reported in a case with total anomalous pulmonary venous connection.[52,53] In both instances, wild-type semaphorins appear to establish boundaries for stereotypical neural crest cell migration in the developing heart. We hypothesize that an identical function is served in GI tract development. One attractive hypothesis is that class 3 Semaphorins are synthesized and secreted into the gut mesenchyme and form an appropriate gradient across the gut radius. Because ENCCs in the embryonic gut express plexins and neuropilins, binding of Sema3c/Sema3d to the receptor complex could function as a signal to promote the migratory behavior of neurons, as observed for other semaphorin family members.[54] These axon guidance cues might also exert long-range regulation of neural connectivity and synaptic specificity.[55]

Beyond deciphering the developmental mechanisms that are compromised in HSCR by class 3 Semaphorin mutations, numerous genetic questions remain. First, what are the molecular identities of the polymorphic *SEMA3* and *RET* susceptibility alleles? Are they transcriptional enhancers, like the intron 1 *RET* element, and if so which transcriptional regulators bind them? It is most likely that the risk allele at rs2506030 affects RET. But, does the risk allele at rs12707682 control expression of *SEMA3D* ~182 kb away or *SEMA3C* that is ~4 Mb away, or both? Regulatory *cis*-elements routinely act over significant distances ($\geq$100 kb) from their cognate promoters, with several operating over distances ~1 Mb.[56] Consequently, gene expression studies in the GI tract of HSCR-affected individuals with known genotypes of rs12707682 are needed to resolve this question. Because the defect is during development, providing these answers will require studies in cellular and animal models of this variant. Finally, our human genetic data do predict some of its properties that are relevant to the genetic counseling of HSCR: the risk allele is more prevalent in males than

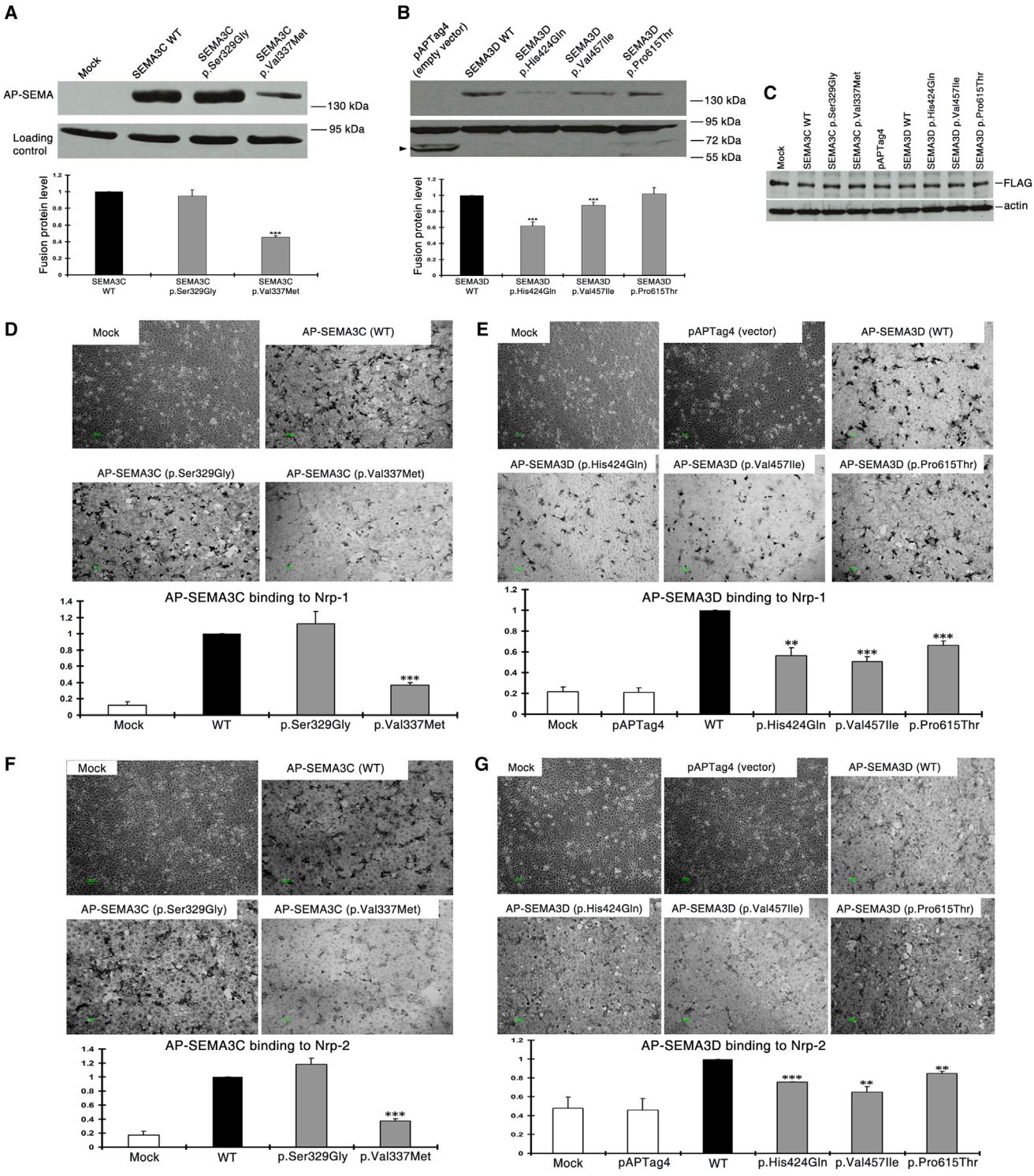

Figure 5. AP-SEMA Fusion Protein Accumulation and Receptor Binding

(A–C) Secreted SEMA ligand concentrations were determined by measuring alkaline phosphatase (AP) activity for SEMA3C (A) and SEMA3D (B); results are shown as fold change relative to the wild-type construct and error bars represent SEM. The upper bands in (A) (for SEMA3C) and (B) (for SEMA3D) represent the AP-SEMA fusion protein of predicted molecular weight (MW) of ~130 kDa; the lower bands of ~90 kDa are suspected to be an abundant albumin within the medium recognized by the antibody and used as loading control. An additional band (arrowhead) was detected in only the pAPTag4 transfected supernatant at ~63 kDa, indicating the expression of AP (MW 67 kDa) upon empty vector transfection. Mock represents conditioned medium from untransfected HEK293T cells; pAPTag4 represents the empty vector transfection. To test the transfection rate, double-transfection with AP-SEMA and FLAG-Hoxb7 on HEK293T cells followed by immunoblotting is presented in (C).

(legend continued on next page)

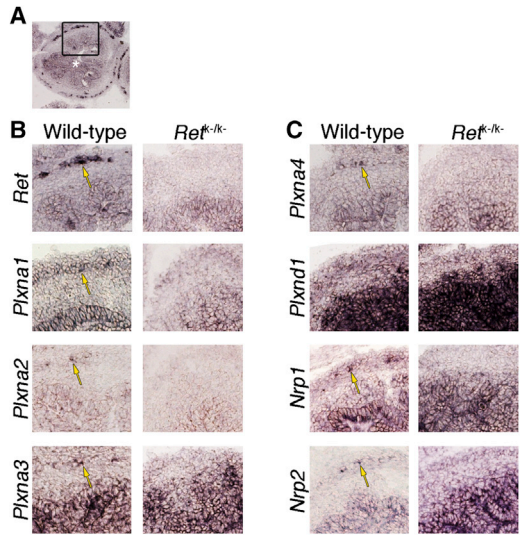

Figure 6. Analysis of Gene Expression of *PlexinA1*, *Plexin A2*, *PlexinA3*, *Plexin A4*, *PlexinD1*, *Neuropilin1*, and *Neuropilin2* Relative to *Ret* in the E15.5 Mouse Wild-Type and *Ret* Mutant Gut

RNA in situ hybridization analysis was performed for all probes on E15.5 intestine tissue cross sections from both wild-type embryos and $Ret^{k-/k-}$ mutant embryos. *Ret* is expressed in enteric neurons in the myenteric plexus in the outer regions of the E15.5 intestine (A, lumen denoted by asterisk). High-magnification view of boxed area of (A) is shown in (B), revealing individual *Ret*-expressing enteric neurons in the myenteric plexus (B, *Ret* panel, arrow). *Ret* expression is absent in corresponding cross-sections in $Ret^{k-/k-}$ intestine tissue, which lack ENS neurons (B, *Ret* panel). *PlexinA1* is expressed in population of ENS cells (B, *Plxna1* panel, arrow), and this expression is absent in $Ret^{k-/k-}$ gut tissues (B, *Plxna1* panel), thus confirming that the expression observed in wild-type tissue corresponds to ENS expression. Expression of *PlexinA2*, *PlexinA3*, and *PlexinA4* is also observed in ENS cells (B and C, *Plxna2*, *Plxna3*, *Plxna4* panels, arrows), and expression is not detectable in $Ret^{k-/k-}$ gut tissues (B and C, *Plxna2*, *Plxna3*, *Plxna4* panels). *PlexinD1* has widespread expression within the gut, but not, apparently, within the ENS (C, *Plxnd1* panel). Consistent with this observation, the expression profile appears unchanged in $Ret^{k-/k-}$ gut tissues (C, *Plxnd1* panel). Populations of the ENS do express *Neuropilin1* and *Neuropilin2* (C, *Nrp1*, *Nrp2* panels, arrows), and this expression is not observed in $Ret^{k-/k-}$ gut tissues (C, *Nrp1*, *Nrp2* panels).

females, simplex than multiplex cases, and in S-HSCR as compared L-HSCR/TCA, although none of these effects are significant. This difference between the phenotypic associations of rs12707682 and rs2435757 could be due either to a smaller effect, and consequent lower statistical power, of the *SEMA3* variants or to an effect restricted to only some genes (e.g., *RET*) owing to their molecular properties (e.g., sex differences in gene expression). We now know of four such non-coding but highly polymorphic alleles, two at *RET* and one each at *NRG1* and *SEMA3C/D*, the latter two interacting with *RET*. This is surprising for a rare multifactorial developmental disorder and begs the question of how many such common HSCR susceptibility alleles exist. We suspect that a focused search of ENS-localized and GI-tract mesenchymal genes can identify additional common variants that, together with rare deleterious alleles, will complete the molecular disease architecture of HSCR.

Accession Numbers

The NCBI dbSNP accession number for the novel variant listed in Table 3 is ss1582838823, scheduled to appear in the next dbSNP build (B144), planned for fall 2015.

Supplemental Data

Supplemental Data include six figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.02.014.

Acknowledgments

We wish to thank our subjects, their families, referring physicians, and genetic counselors for participating in these studies, and we are grateful to Courtney Berrios, Julie Albertus, Maura Kenton, and Jennifer (Scott) Bubb for family ascertainment and genetic counseling; Priyanka Nandakumar, Alexis Rhea, Ankit Rakha, and Carl Kashuk for technical assistance; Drs. David Ginty (Johns Hopkins University, Baltimore; Sema3a/c/d/e), Kevin Mitchell (University of Dublin, Ireland; PlexinA2/A4), and Jeffrey Macklis (Harvard University, Cambridge; PlexinD1) for reagents; and Dr. Ian Shephard (Emory University, Atlanta) for protocols and technical advice on zebrafish analysis. We are also thankful to Drs. David Ginty and Alex Kolodkin for discussions and comments on this paper. The studies here were supported by the following grants: MERIT NIH Award HD28088 to A.C.; NIH GM71648 to A.S.M.; National Natural Science Foundation of China (No. 81300266), Beijing Natural Science Foundation (No. 7142029), and Beijing Excellent Scientist Fund (No. 2013D003034000007) to Qian Jiang; MRC grants to T.H. and V.P.; Italian Ministry of Health ("Cinque per mille," Ricerca Corrente to the Gaslini Institute and Ricerca Finalizzata # RF-2010-2314356) to I.C.; ANR-10-IAHU-01 grant to J.A. and S.L.; 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. Borrego; and NWO (no. 901-04-225) Bernoulle Foundation and Ubbo Emmius Foundation to R.M.W.H. The arrays were a gift from Affymetrix. A.C. was a member of the Advisory Board of Affymetrix (2000–2013), a potential conflict of interest managed by the policies of the Johns Hopkins School of Medicine.

Received: December 2, 2014
Accepted: February 20, 2015
Published: April 2, 2015

(D and E) AP-SEMA binding affinity with Neuropilin1. COS-7 cells transfected with Nrp1 incubated with AP-SEMA3C- (D) or AP-Sema3D- (E) containing medium. Enzymatic detection of binding indicates that all ligands (except for empty vector) bind to the transfected cells but to different extents (scale bars represent 50 μm). Quantitative measurements of AP-SEMA binding to Nrp-1-expressing COS-7 cells were conducted. Results shown are bound AP activity averages and SEM of three independent experiments.
(F and G) AP-SEMA binding affinity with Neuropilin2. Legend is the same as in (D) and (E).

Web Resources

The URLs for data presented herein are as follows:

CEPH Human Genome Diversity Panel, http://www.cephb.fr/en/hgdp_panel.php
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
International HapMap Project, release 22, http://hapmap.ncbi.nlm.nih.gov/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/

References

1. Gershon, M.D. (1998). The Second Brain (Harper & Co.).
2. Chakravarti, A., and Lyonnet, S. (2001). Hirschsprung disease. In The Metabolic and Molecular Bases of Inherited Disease, Eighth Edition, C.R. Scriver, A.L. Beaudet, D. Valle, W.S. Sly, B. Childs, K. Kinzler, and B. Vogelstein, eds. (McGraw-Hill), pp. 6231–6255.
3. Alves, M.M., Sribudiani, Y., Brouwer, R.W., Amiel, J., Antiñolo, G., Borrego, S., Ceccherini, I., Chakravarti, A., Fernández, R.M., Garcia-Barcelo, M.M., et al. (2013). Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. Dev. Biol. *382*, 320–329.
4. Emison, E.S., McCallion, A.S., Kashuk, C.S., Bush, R.T., Grice, E., 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.
5. Garcia-Barcelo, M.M., Tang, C.S., Ngan, E.S., Lui, V.C., Chen, Y., So, M.T., Leon, T.Y., Miao, X.P., Shum, C.K., Liu, F.Q., et al. (2009). Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. Proc. Natl. Acad. Sci. USA *106*, 2694–2699.
6. Emison, E.S., Garcia-Barcelo, M., Grice, E.A., Lantieri, F., Amiel, J., Burzynski, G., Fernandez, R.M., Hao, L., Kashuk, C., West, K., et al. (2010). Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. Am. J. Hum. Genet. *87*, 60–74.
7. de Pontual, L., Pelet, A., Clement-Ziza, M., Trochet, D., Antonarakis, S.E., Attie-Bitach, T., Beales, P.L., Blouin, J.L., Dastot-Le Moal, F., Dollfus, H., et al. (2007). Epistatic interactions with a common hypomorphic RET allele in syndromic Hirschsprung disease. Hum. Mutat. *28*, 790–796.
8. 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.
9. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., et al. (2002). A human genome diversity cell line panel. Science *296*, 261–262.
10. Lin, S., Carvalho, B., Cutler, D.J., Arking, D.E., Chakravarti, A., and Irizarry, R.A. (2008). Validation and extension of an empirical Bayes method for SNP calling on Affymetrix microarrays. Genome Biol. *9*, R63.
11. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al.; International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
12. Browning, S.R., and Browning, B.L. (2010). High-resolution detection of identity by descent in unrelated individuals. Am. J. Hum. Genet. *86*, 526–539.
13. Jiang, Q., Turner, T., Sosa, M.X., Rakha, A., Arnold, S., and Chakravarti, A. (2012). Rapid and efficient human mutation detection using a bench-top next-generation DNA sequencer. Hum. Mutat. *33*, 281–289.
14. Sham, P.C., and Curtis, D. (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Ann. Hum. Genet. *59*, 323–336.
15. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263–265.
16. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
17. Taraviras, S., Marcos-Gutierrez, C.V., Durbec, P., Jani, H., Grigoriou, M., Sukumaran, M., Wang, L.C., Hynes, M., Raisman, G., and Pachnis, V. (1999). Signalling by the RET receptor tyrosine kinase and its role in the development of the mammalian enteric nervous system. Development *126*, 2785–2797.
18. Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.
19. Sakai, J.A., and Halloran, M.C. (2006). Semaphorin 3d guides laterality of retinal ganglion cell projections in zebrafish. Development *133*, 1035–1044.
20. Banu, N., Teichman, J., Dunlap-Brown, M., Villegas, G., and Tufro, A. (2006). Semaphorin 3C regulates endothelial cell function by increasing integrin activity. FASEB J. *20*, 2150–2152.
21. Huber, A.B., Kania, A., Tran, T.S., Gu, C., De Marco Garcia, N., Lieberam, I., Johnson, D., Jessell, T.M., Ginty, D.D., and Kolodkin, A.L. (2005). Distinct roles for secreted semaphorin signaling in spinal motor axon guidance. Neuron *48*, 949–964.
22. 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.
23. Carrasquillo, M.M., McCallion, A.S., Puffenberger, E.G., Kashuk, C.S., Nouri, N., and Chakravarti, A. (2002). Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. Nat. Genet. *32*, 237–244.
24. Yu, H.H., and Moens, C.B. (2005). Semaphorin signaling guides cranial neural crest cell migration in zebrafish. Dev. Biol. *280*, 373–385.
25. Brown, C.B., Feiner, L., Lu, M.M., Li, J., Ma, X., Webber, A.L., Jia, L., Raper, J.A., and Epstein, J.A. (2001). PlexinA2 and semaphorin signaling during cardiac neural crest development. Development *128*, 3071–3080.
26. Heanue, T.A., and Pachnis, V. (2008). Ret isoform function and marker gene expression in the enteric nervous system is conserved across diverse vertebrate species. Mech. Dev. *125*, 687–699.
27. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2007).

Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein Sci. *Chapter 2*, 9.
28. Janssen, B.J., Malinauskas, T., Weir, G.A., Cader, M.Z., Siebold, C., and Jones, E.Y. (2012). Neuropilins lock secreted semaphorins onto plexins in a ternary signaling complex. Nat. Struct. Mol. Biol. *19*, 1293–1299.
29. Shen, M.Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Sci. *15*, 2507–2524.
30. Chaudhury, S., Berrondo, M., Weitzner, B.D., Muthu, P., Bergman, H., and Gray, J.J. (2011). Benchmarking and analysis of protein docking performance in Rosetta v3.2. PLoS ONE *6*, e22477.
31. Dunbrack, R.L., Jr., and Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. *6*, 1661–1681.
32. Flanagan, J.G., and Cheng, H.J. (2000). Alkaline phosphatase fusion proteins for molecular characterization and cloning of receptors and their ligands. Methods Enzymol. *327*, 198–210.
33. Flanagan, J.G., and Leder, P. (1990). The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. Cell *63*, 185–194.
34. Giger, R.J., Urquhart, E.R., Gillespie, S.K., Levengood, D.V., Ginty, D.D., and Kolodkin, A.L. (1998). Neuropilin-2 is a receptor for semaphorin IV: insight into the structural basis of receptor function and specificity. Neuron *21*, 1079–1092.
35. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv, http://dx.doi.org/10.1101/005165.
36. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.
37. Kruger, R.P., Aurandt, J., and Guan, K.L. (2005). Semaphorins command cells to move. Nat. Rev. Mol. Cell Biol. *6*, 789–800.
38. Durbec, P.L., Larsson-Blomberg, L.B., Schuchardt, A., Costantini, F., and Pachnis, V. (1996). Common origin and developmental dependence on c-ret of subsets of enteric and sympathetic neuroblasts. Development *122*, 349–358.
39. 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.
40. Anderson, R.B., Bergner, A.J., Taniguchi, M., Fujisawa, H., Forrai, A., Robb, L., and Young, H.M. (2007). Effects of different regions of the developing gut on the migration of enteric neural crest-derived cells: a role for Sema3A, but not Sema3F. Dev. Biol. *305*, 287–299.
41. Robu, M.E., Larson, J.D., Nasevicius, A., Beiraghi, S., Brenner, C., Farber, S.A., and Ekker, S.C. (2007). p53 activation by knockdown technologies. PLoS Genet. *3*, e78.
42. Heanue, T.A., and Pachnis, V. (2007). Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. Nat. Rev. Neurosci. *8*, 466–479.
43. Kolodkin, A.L., Matthes, D.J., and Goodman, C.S. (1993). The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. Cell *75*, 1389–1399.
44. Kolodkin, A.L., and Tessier-Lavigne, M. (2011). Mechanisms and molecules of neuronal wiring: a primer. Cold Spring Harb. Perspect. Biol. *3* http://dx.doi.org/10.1101/cshperspect.a001727.
45. Lwigale, P.Y., and Bronner-Fraser, M. (2009). Semaphorin3A/neuropilin-1 signaling acts as a molecular switch regulating neural crest migration during cornea development. Dev. Biol. *336*, 257–265.
46. Scholl, A.M., and Kirby, M.L. (2009). Signals controlling neural crest contributions to the heart. Wiley Interdiscip Rev Syst Biol Med *1*, 220–227.
47. Tran, T.S., Kolodkin, A.L., and Bharadwaj, R. (2007). Semaphorin regulation of cellular morphology. Annu. Rev. Cell Dev. Biol. *23*, 263–292.
48. Heanue, T.A., and Pachnis, V. (2006). Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. Proc. Natl. Acad. Sci. USA *103*, 6919–6924.
49. Wolman, M.A., Liu, Y., Tawarayama, H., Shoji, W., and Halloran, M.C. (2004). Repulsion and attraction of axons by semaphorin3D are mediated by different neuropilins in vivo. J. Neurosci. *24*, 8428–8435.
50. Tanaka, H., Nojima, Y., Shoji, W., Sato, M., Nakayama, R., Ohshima, T., and Okamoto, H. (2011). Islet1 selectively promotes peripheral axon outgrowth in Rohon-Beard primary sensory neurons. Dev. Dyn. *240*, 9–22.
51. Shepherd, I.T., and Raper, J.A. (1999). Collapsin-1/semaphorin D is a repellent for chick ganglion of Remak axons. Dev. Biol. *212*, 42–53.
52. Feiner, L., Webber, A.L., Brown, C.B., Lu, M.M., Jia, L., Feinstein, P., Mombaerts, P., Epstein, J.A., and Raper, J.A. (2001). Targeted disruption of semaphorin 3C leads to persistent truncus arteriosus and aortic arch interruption. Development *128*, 3061–3070.
53. Degenhardt, K., Singh, M.K., Aghajanian, H., Massera, D., Wang, Q., Li, J., Li, L., Choi, C., Yzaguirre, A.D., Francey, L.J., et al. (2013). Semaphorin 3d signaling defects are associated with anomalous pulmonary venous connections. Nat. Med. *19*, 760–765.
54. Osborne, N.J., Begbie, J., Chilton, J.K., Schmidt, H., and Eickholt, B.J. (2005). Semaphorin/neuropilin signaling influences the positioning of migratory neural crest cells within the hindbrain region of the chick. Dev. Dyn. *232*, 939–949.
55. Carrillo, R.A., Olsen, D.P., Yoon, K.S., and Keshishian, H. (2010). Presynaptic activity and CaMKII modulate retrograde semaphorin signaling and synaptic refinement. Neuron *68*, 32–44.
56. Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.