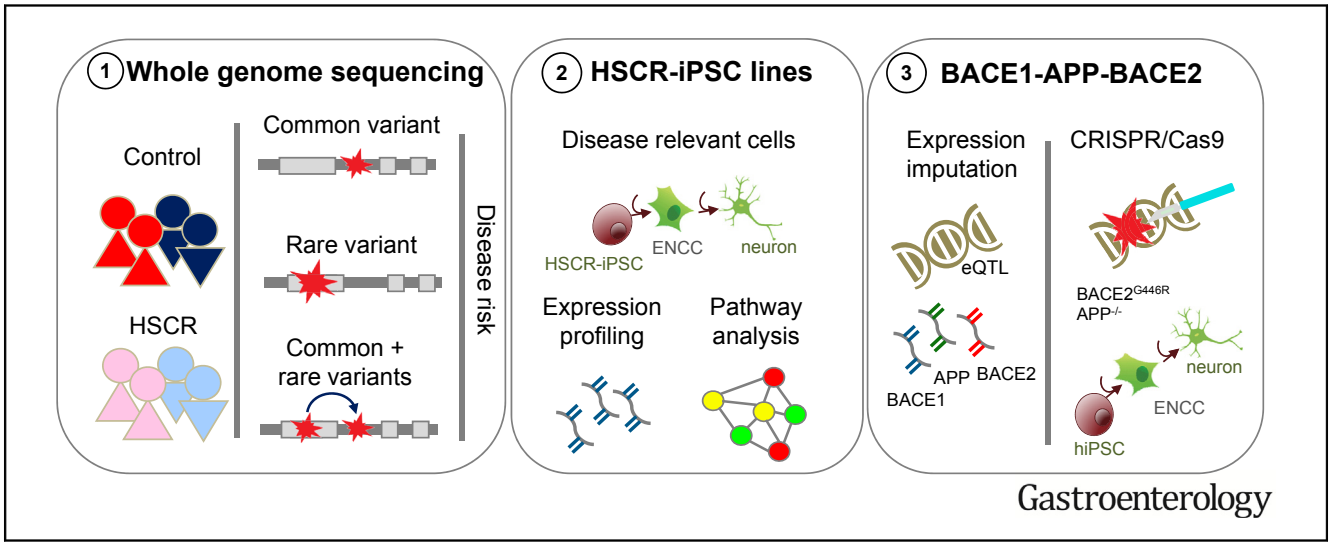
Gastroenterology 2018;155:1908–1922

Identification of Genes Associated With Hirschsprung Disease, Based on Whole-Genome Sequence Analysis, and Potential Effects on Enteric Nervous System Development

Clara Sze-man Tang,[1,6,*] **Peng Li**,[1,*] Frank Pui-Ling Lai,[1,6] Alexander Xi Fu,[7] Sin-Ting Lau,[1,6] Man Ting So,[1] Kathy Nga-Chu Lui,[1] Zhixin Li,[1,6] Xuehan Zhuang,[1] Michelle Yu,[1] Xuelai Liu,[8] Ngoc D. Ngo,[9] Xiaoping Miao,[10] Xi Zhang,[11] Bin Yi,[11] Shaotao Tang,[11] Xiaobing Sun,[12] Furen Zhang,[13] Hong Liu,[13] Qiji Liu,[14] Ruizhong Zhang,[15] Hualong Wang,[16] Liuming Huang,[17] Xiao Dong,[18] Jinfa Tou,[19] Kathryn Song-Eng Cheah,[4] Wanling Yang,[5] Zhenwei Yuan,[20] Kevin Yuk-lap Yip,[7] Pak-Chung Sham,[2,3] Paul Kwang-Hang Tam,[1] Maria-Mercè Garcia-Barcelo,[1] and Elly Sau-Wai Ngan[1]

[1]Department of Surgery, Li Ka Shing Faculty of Medicine, University of Hong Kong, Pokfulam, Hong Kong, China; [2]Department of Psychiatry, Li Ka Shing Faculty of Medicine, University of Hong Kong, Pokfulam, Hong Kong, China; [3]Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, University of Hong Kong, Pokfulam, Hong Kong, China; [4]School of Biological Sciences, Li Ka Shing Faculty of Medicine, University of Hong Kong, Pokfulam, Hong Kong, China; [5]Department of Pediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Pokfulam, Hong Kong, China; [6]Dr Li Dak-Sum Research Centre, The University of Hong Kong, Pokfulam, Hong Kong, China; [7]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China; [8]Hebei Medical University Second Hospital, Shijiazhuang, Hebei, China; [9]National Hospital of Pediatrics, Ha Noi, Viet Nam; [10]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; [11]Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; [12]Department of Paediatric Surgery, Shandong Medical University, Shandong, China; [13]Shandong Provincial Institute of Dermatology and Venereology, Shandong Academy of Medical Sciences, Jinan, Shandong, China; [14]The Key Laboratory for Experimental Teratology of the Ministry of Education, Shandong University School of Medicine, Jinan, Shandong, China; [15]Guangzhou Women and Children's Medical Center, Guangzhou, Guangdong, China; [16]Changchun Children's Hospital, Changchun, Jilin, China; [17]Bayi Children's Hospital, General Hospital of Beijing Military Region, Beijing, China; [18]Shenzhen Children's Hospital, Shenzhen, Guangdong, China; [19]Zhejiang Children's Hospital, Hangzhou, Zhejiang, China; and [20]Department of Paediatric Surgery, Shengjing Hospital, China Medical University, Shenyang, China



See editorial on page 1681.

BACKGROUND & AIMS: Hirschsprung disease, or congenital aganglionosis, is believed to be oligogenic—that is, caused by multiple genetic factors. We performed whole-genome sequence analyses of patients with Hirschsprung disease to identify genetic factors that contribute to disease development and analyzed the functional effects of these variants. **METHODS:** We performed whole-genome sequence analyses of 443 patients with short-segment disease, recruited from hospitals in China and Vietnam, and 493 ethnically matched

individuals without Hirschsprung disease (controls). We performed genome-wide association analyses and gene-based rare-variant burden tests to identify rare and common disease–associated variants and study their interactions. We obtained induced pluripotent stem cell (iPSC) lines from 4 patients with Hirschsprung disease and 2 control individuals, and we used these to generate enteric neural crest cells for transcriptomic analyses. We assessed the neuronal lineage differentiation capability of iPSC-derived enteric neural crest cells using an in vitro differentiation assay. **RESULTS:** We identified 4 susceptibility loci, including 1 in the phospholipase D1 gene (*PLD1*) ($P = 7.4 \times 10^{-7}$). The patients had a significant excess of rare protein-altering variants in genes previously associated with Hirschsprung disease and in the β-secretase 2 gene (*BACE2*) ($P = 2.9 \times 10^{-6}$). The epistatic effects of common and rare variants across these loci provided a sensitized background that increased risk for the disease. In studies of the iPSCs, we observed common and distinct pathways associated with variants in *RET* that affect risk. In functional assays, we found variants in *BACE2* to protect enteric neurons from apoptosis. We propose that alterations in BACE1 signaling via amyloid β precursor protein and BACE2 contribute to pathogenesis of Hirschsprung disease. **CONCLUSIONS:** In whole-genome sequence analyses of patients with Hirschsprung disease, we identified rare and common variants associated with disease risk. Using iPSC cells, we discovered some functional effects of these variants.

Keywords: Genetics; Amyloid Beta; Enteric Nervous System; CRISPR/Cas9.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Hirschsprung disease (HSCR) is a complex congenital disease characterized by absence of nerve cells in the distal colon.

NEW FINDINGS

In whole-genome sequence analyses of patients with Hirschsprung disease, the authors identified rare and common variants associated with disease risk. Using iPSC cells, they discovered functional effects of these variants.

LIMITATIONS

A limited number of cases (443) and control (493) were included in the WGS analysis. Only an in vitro hPSC-based model was used for functional studies.

IMPACT

The integration of jointly analyzed rare and common variants with the use of hiPSC-based model represents a powerful approach for studying oligogenic diseases.

Hirschsprung disease (HSCR), or congenital aganglionosis, is a highly heritable oligogenic disorder with significant phenotypic variability. The incidence rate of HSCR varies by population and is highest among Asians (2.8/10,000 live births).[1] Patients are classified according to the extent of aganglionosis into 3 main types: short-segment HSCR (S-HSCR) (80%), long-segment HSCR (L-HSCR) (15%), and total colonic aganglionosis (TCA) (5%). These subtypes are believed to differ in genetic architecture. L-HSCR/TCA is mostly autosomal dominant, although with incomplete penetrance, whereas S-HSCR follows a complex, non-Mendelian inheritance pattern.[2] Differential contribution of common and rare variants in the major gene, *RET*, has been suggested as one of the underlying factors for such differences.[3] The genetic effect of the common HSCR-associated enhancer variant (rs2435357), which decreases the expression of *RET*, is directly proportional to the subtype prevalence, that is, a larger effect in S-HSCR in males. On the contrary, the frequency of the *RET* coding mutations correlates positively with disease severity.

Our previous meta-analysis of genome-wide association studies estimated that common variants together account for a small proportion of heritability estimated from family studies.[4] Rare variants might therefore contribute significantly to the missing heritability. Thus far, most of the genetic analyses on HSCR focused on assessing the contribution of *RET* and other genes known to participate in the development of the enteric nervous system (ENS) in syndromic, familial, or more severe forms of HSCR (ie, L-HSCR and TCA). To identify novel HSCR genes and explore the oligogenic nature of the disease, we carried out a high-coverage whole-genome sequencing (WGS) study of the most common S-HSCR subtype and tested for association of both common and, more importantly, rare variants with HSCR. By integrating the human pluripotent stem cell–based model, we further defined new biological pathways underlying the HSCR pathogenesis.

Materials and Methods

Patients

The discovery cohort comprised 464 patients with sporadic S-HSCR and 498 control individuals. Patients had been recruited at hospitals in China (n = 341) and Hanoi, Vietnam (n = 102). After quality control (detailed in the Supplementary Materials and Methods), a total of 443 S-HSCR patients and 493 control individuals remained for genetic analyses. The follow-up cohort included 534 ethnically matched controls subjected to Sanger sequencing for *BACE2*. Informed consent was obtained from all participants, and the study was approved by the

***Authors share co-first authorship.**

Abbreviations used in this paper: **Aβ, amyloid β; APP, amyloid β precursor protein; BACE2, β-site amyloid β precursor protein–cleaving enzyme; CASQ2, calsequestrin 2; CI, confidence interval; DEG, differentially expressed gene; ENCC, enteric neural crest cell; ENS, enteric nervous system; GO, gene ontology; hiPSC, human induced pluripotent stem cell; HSCR, Hirschsprung disease; L-HSCR, long-segment Hirschsprung disease; MAF, minor allele frequency; OR, odds ratio; PLD1, phospholipase D1; S-HSCR, short-segment Hirschsprung disease; TCA, total colonic aganglionosis; URV, ultra-rare variant; WGS, whole-genome sequencing.**

Most current article

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2018.09.012

institutional review board of the University of Hong Kong and the Hospital Authority (UW 13-225).

WGS and Variant Calling

WGS was performed on all samples using Illumina HiSeq X Ten to a mean coverage of ×30. Sequence reads were then processed according to Genome Analysis Toolkit, version 3.4, best practices recommendations[5] (see Supplementary Methods).

Variant Annotation

Annotation was done using KGGseq (http://grass.cgs.hku.hk/limx/kggseq/) for protein function against the UCSC RefGene database, pathogenicity, and population frequencies. We defined protein-truncating variants as those that lead to (1) gain of the stop codon, (2) frameshift and (3) alteration of the essential splice sites. Damaging variants include all protein-truncating variants and missense or in-frame variants predicted to be deleterious by KGGseq. Benign variants are missense variants or in-frame variants predicted benign by KGGseq. Finally, protein-altering variants comprise both damaging and benign variants. Rare variants are those whose minor allele frequency (MAF) is <0.01 in public databases. Ultra-rare variants (URVs) are defined as a singleton variant, that is, one that appeared only once in our whole data set, not present in dbSNP138 or public databases (see Supplementary Methods).

Known Genes of ENS Development and Their Interactome

Genes with mutations reported to cause colonic aganglionosis in mutant mice according to the Mouse Genomics database were considered to be known ENS genes. ENS interactome was defined by genes encoding proteins that show protein–protein interaction with known ENS genes (see Supplementary Methods).

Copy Number Variants

Overlapping copy number variants for ENS genes were detected using 4 different yet complementary software programs to maximize accuracy.

Gene-Based and Gene Set–Based Burden Test for Rare Variants

For the set of known ENS genes, we first assessed the enrichment of (1) damaging and (2) all rare protein-altering variants collectively in patients compared with control individuals.

RET Common and Rare Variants Epistasis

To assess if the effects of rare *RET* protein-alternating variants changed with the dosage of common HSCR-associated risk alleles (T for rs2435357 and A for rs9282834), we stratified samples into 3 groups: carrying zero, 1, or at least 2 common HSCR-associated risk alleles. Samples were further subdivided into 3 subgroups (totaling $9 = 3 \times 3$ combinations), according to the presence of mutations and their predicted pathogenicity (damaging and benign).

RET Haplotype Configurations

To determine if the rare *RET* protein-alternating variants occur in *cis* or *trans* with the enhancer variant (rs2435357), we performed read-aware phasing, as described in the Supplementary Methods.

Imputation of Expression Using PrediXcan

To impute the gene expression of *BACE2, BACE1*, and *APP*, we considered 2 tissue models (each with >300 samples): (1) the neural (tibial nerve, 361 individuals) and (2) whole blood (369 individuals) (see Supplementary Methods).

Human Induced Pluripotent Stem Cells (hiPSCs)

Two control (IMR90 and UE02302) and 4 HSCR (3 S-HSCR and 1 TCA) hiPSC lines were used to generate enteric neural crest cells (ENCCs) and ENS neurons. IMR90 iPSC (clone 2) was purchased from WiCell Research Institute, UE02302 was a gift from Dr Guangjin Pan (Guangzhou Institutes of Biomedicine and Health, China),[6] and the HSCR-iPSC lines were generated as previously described.[7] $BACE2^{-/-}$, $BACE2^{G446R}$, and $BACE2^{G446R}APP^{-/-}$ mutant hiPSC lines were derived from this control line (Supplementary Methods and Supplementary Table 1). All the control and diseased hiPSC lines were cultured on Matrigel (BD Biosciences, 354234)-coated plates with the defined medium mTeSR1 (Stemcell Technologies, 05850), and the culture medium was changed daily.

Neural Crest Induction

hiPSCs were plated on Matrigel-coated plates (10^5 cells cm^{-2}) in ES cell medium containing 10 ng/mL FGF2 (PeproTech, 100-18B). Differentiation was initiated by replacing the ES medium with KSR medium containing LDN193189, SB431542, and CHIR99021 and then gradually switching to N2 medium and caudalized with 1 μmol/L retinoic acid, as described previously.[7] ENCCs were then enriched using $p75^{NTR}$ and HNK-1 antibodies.

Fluorescence-Activated Cell Sorter and Flow Cytometry Analysis

For flow cytometry analysis or cell sorting, the cells were dissociated with Accutase and labeled with anti-human antibodies (Supplementary Table 2). The labeled cells were detected using a FACSCalibur instrument. Isotype-matched antibodies were used as controls. FlowJo, version 8.2, (Tree Star, Inc) was used to analyze the flow data.

In Vitro Differentiation of ENCCs to Enteric Neurons

ENCCs (3×10^4) from the 10-day induction protocol were seeded as droplets on poly-ornithine/laminin/fibronectin–coated dishes in N2 medium containing 10 ng/mL FGF2 and 3 μmol/L CHIR99021. After 24 hours, N2 medium was replaced by the neuronal differentiation medium: N2 medium containing GDNF, ascorbic acid, brain-derived neurotrophic factor, nerve growth factor, neurotrophin-3, and cyclic adenosine monophosphate. Cells were cultured in the neuronal differentiation

medium up to 30 days, and the culture medium was changed every 2 days. ENS neurons at differentiation day 30 were fixed for immunocytochemistry analyzes or harvest using Accutase for RNA sequencing and Western blot analysis.

Immunofluorescence Analysis

For immunofluorescence, the cells were fixed with 4% paraformaldehyde in phosphate-buffered saline at room temperature for 30 minutes, blocked, and incubated in primary antibody solutions and host-appropriate secondary antibody. Cells were then counterstained with mounting medium with 4′,6-diamidino-2-phenylindole (DAKO) and photographed using a Carl Zeiss confocal microscope (LSM 800). Quantitative image analysis of differentiated neuronal cultures was done with ImageJ plugins (National Institutes of Health). A minimum of 4,000 cells were analyzed per sample. Percentages of neuronal cells were measured over the total number of cells (4′,6-diamidino-2-phenylindole) and the values reported in bar charts represent the mean ± standard error of the mean.

Cell Culture, Transfection, and Immunoblotting

The 293FT cell line was used to analyze the biological impacts of *BACE2* variants in amyloid β precursor protein (APP) processing and BACE2 membrane localization. The 293FT cells were cultured in DMEM medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin, at 37°C in 5% CO_2, and the culture medium was changed every other day. For transfection, approximately 1 million cells were seeded onto 6-well plates (Nunc) 24 hours before transfection. Green-fluorescent protein–tagged APP, together with FLAG-tagged wild-type or mutant β-site amyloid β precursor protein–cleaving enzyme (BACE2) were overexpressed in the 293FT cell line by transfection using FuGENE HD Transfection Reagent (Promega) according to the transfection protocol. Two days after transfection, the cells were collected and lysed with protein lysis buffer. For the membrane and cytosolic protein fractionation, 293FT cells overexpressing FLAG-tagged wild-type BACE2, S442F, or G446R BACE2 were collected 48 hours after transfection. Membrane and cytosolic proteins were extracted using the Mem-PER plus membrane protein extraction kit (no. 89842, Thermo Fisher Scientific) according to the manufacturer's protocol. Next, 20 μg of total protein from cell lysates was separated on 12% sodium dodecyl sulfate–polyacrylamide gels and blotted with the corresponding primary antibodies as listed in Supplementary Table 2. The same membranes were stripped and hybridized with anti-β-actin monoclonal antibody (Millipore, MAB 1501) as a protein-loading control. All blots were incubated with secondary horseradish peroxidase-conjugated secondary antibodies (DAKO).

Statistical Analysis

Statistical significance was determined by the 2-sided unpaired Student t test or 1-way analysis of variance using GraphPad Prism 7 (GraphPad Software). P values are indicated by asterisks in the figures. Differences among groups of $P < .05$ were considered statistically significant. All experiments were replicated at least 3 times, and data are shown as means with standard error of mean or standard derivation.

Experimental details are available in the Supplementary Methods.

Results

New Loci Showing Association With S-HSCR

The discovery WGS analysis included 443 S-HSCR patients and 493 ethnically matched control individuals in whom a total of 36.7 million autosomal variants (33.4 million single nucleotide variants and 3.3 million insertions/deletions) were called by the Genome Analysis Toolkit. On average, we identified 3.8 million single-nucleotide variants and 0.5 million insertions/deletions per individual. Most of these variants (77.5%) were novel or rare (MAF < 1%).

We first performed genome-wide association analysis on common and low-frequency biallelic variants (MAF > 1%, n = 7,224,040). The analysis showed the strongest association of 328 variants with HSCR ($P < 5 \times 10^{-8}$), all of which mapped to the known disease susceptibility loci of *RET* and *NRG1* (Figure 1*A*, *upper panel*). Four new loci were identified in this study with moderate association ($P < 1 \times 10^{-6}$) with the disease, which includes 2 intergenic (1 between *LINC01518* and *LOC283028* on 10q11.21 and another between *SLC4A7* and *EOMES* on 3p24.1) and 2 intronic variants on *PLD1* and calsequestrin 2 (*CASQ2*) (Figure 1*A*, *lower panel* and Supplementary Table 3).

Increased Burden of Rare Variants Among S-HSCR

Considering the contribution of protein-truncating (stop-gain, splicing, or frameshift) URVs in L-HSCR,[8] we assessed whether these types of disruptive variants also contributed significantly to S-HSCR disease risk. Among the 936 WGS samples, a total of 4985 protein-truncating URVs were detected. Specially, we observed a significant enrichment of these disruptive URVs in S-HSCR patients only in highly constrained genes, that is, genes having high probability of being intolerant to loss of function (Exome Aggregation Consortium pLI > 0.9; odds ratio [OR], 1.28; 95% confidence interval [CI], 1.06–1.53; $P = 8.6 \times 10^{-3}$) but not in unconstrained genes or for synonymous URVs (Figure 1*B*). Such overrepresentation was primarily detected for URVs that can elicit nonsense-mediated decay ($P = 7.6 \times 10^{-3}$ vs $P = .57$ for those that escape from nonsense-mediated decay). The effect remained significant after excluding the protein-truncating URVs in the known HSCR genes (5 URVs in *RET* and 2 URVs in *ZEB2* only present in patients) in which rare protein-truncating variants in these genes were reported to cause more Mendelian forms of HSCR[9] (OR, 1.24; 95% CI, 1.03–1.49; $P = .021$). Consistently, the number of patients harboring at least 1 protein-truncating URVs was 9% higher than the number of control individuals (42.7% S-HSCR patients vs 33.7% control individuals).

Next, we focused on the known ENS-associated genes (n = 10 with minor allele count > 5) that, when mutated, can recapitulate the colonic aganglionic feature of HSCR in

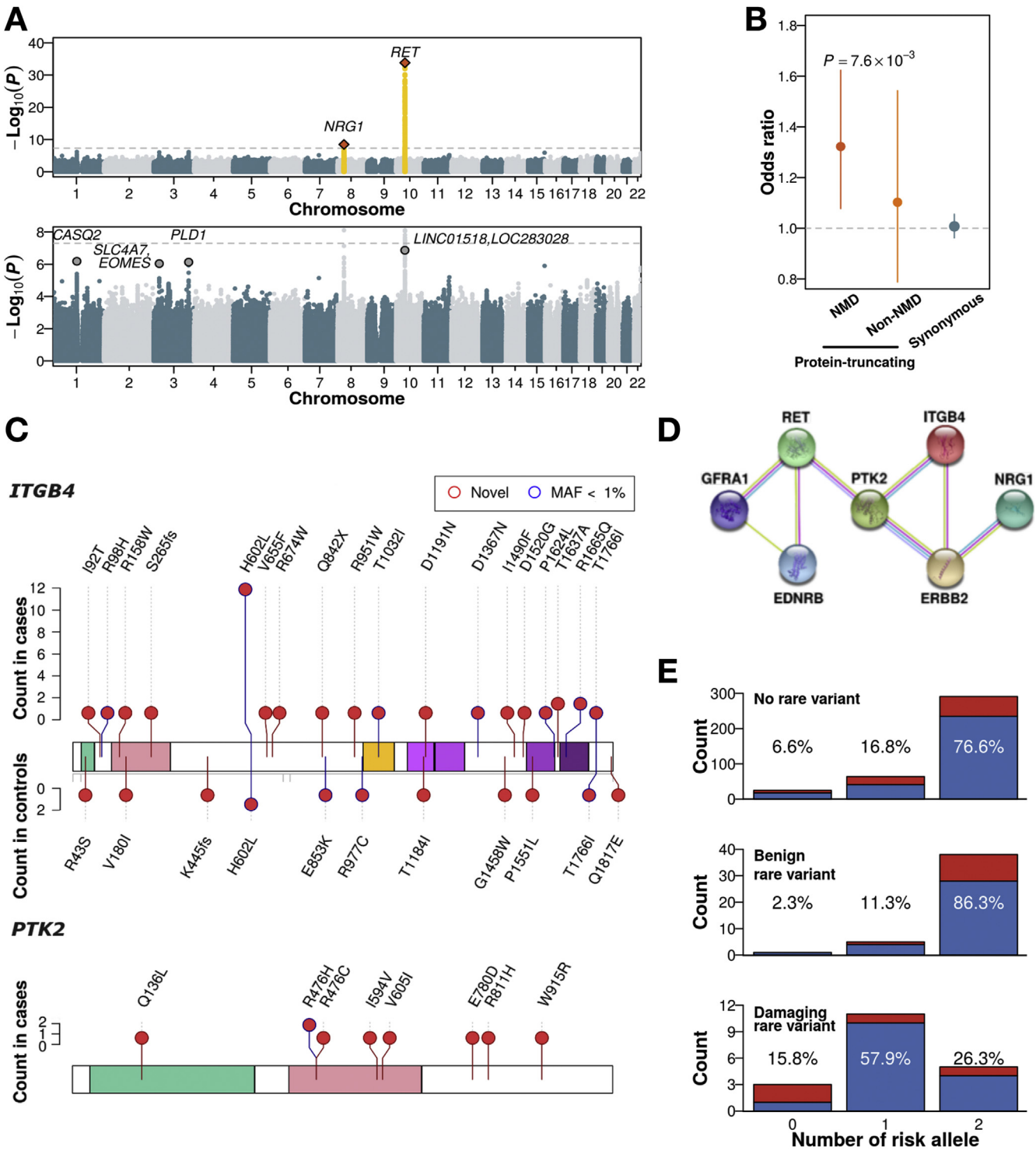


Figure 1. Common and rare variant association analysis on WGS data. (*A*) Manhattan plot of genome-wide association analysis results for variants with MAF > 1%. (*Upper panel*) SNPs within loci (±1 megabase of top SNPs) passing genome-wide significance (*horizontal dashed line*) ($P < 5 \times 10^{-8}$) are labeled in *yellow*, and the top SNP of each region is labeled separately in orange diamond. (*Lower panel*) Magnified plot for moderately associated variants. SNPs showing marginal association ($5 \times 10^{-8} \leq P < 1 \times 10^{-6}$) are denoted by *dark grey circles*. (*B*) Association between protein-truncating and synonymous URVs and HSCR. Protein-truncating variants are further classified by the in silico prediction of the ability to trigger nonsense mediated decay (NMD). (*C*) Profiles of rare damaging variants in *ITGB4* and *PTK2*. Rare variants are represented by lollipops, and line color of the lollipop represents the frequency in the ExAC database. Counts of alternative allele in patients (*top panel*) and control individuals (*bottom panel*) are shown. (*D*) Protein interaction based on STRING database between significant genes from common variant analysis and rare variant analysis on ENS genes and their interactome. (*E*) Counts of S-HSCR patients with no (*top*), benign (*middle*), and damaging (*bottom*) *RET* rare variants across various levels of predisposition of *RET* common variants. *Blue bar*, male; *red bar*, female. ExAC, Exome Aggregation Consortium; NMD, nonsense-mediated decay; SNP, single-nucleotide polymorphism.

Table 1. Rare Variant (SNV and Indels) Gene-Based Burden Tests for Genes Displaying Hirschsprung Phenotype in Knockout Mice

	Count of Damaging Rare Variants				Count of All Protein-Altering Rare Variants			
Gene	Patients	Control Individuals	OR (95% CI)[a]	*P*[b]	Patients	Control Individuals	OR (95% CI)	*P*
RET	19	11	1.96 (0.92,4.17)	.075	63	29	2.65 (1.67–4.20)	**1.2×10^{-5}**
EDNRB	13	0	Not applicable	**1.9×10^{-4}**	16	4	4.58 (1.52–13.81)	**5.5×10^{-3}**
ERBB2	12	3	4.55 (1.27–16.22)	**7.7×10^{-3}**	21	19	1.24 (0.66–2.34)	.446
GFRA1	6	1	6.76 (0.81–65.33)	**.019**	7	4	1.96 (0.57–6.75)	.158
AEBP2	3	7	0.47 (0.12–1.84)	.254	3	8	0.41 (0.11–1.57)	.134
IHH	3	6	0.55 (0.14–2.23)	.415	3	6	0.55 (0.14–2.23)	.415
ITGB1	4	6	0.74 (0.21–2.64)	.767	4	6	0.74 (0.21–2.64)	.767
GDNF	6	6	1.11 (0.36–3.48)	.769	7	6	1.30 (0.43–3.91)	.769
SOX10	3	2	—	—	3	4	0.83 (0.19–3.74)	.930
ZEB2	5	4	1.40 (0.37–5.23)	.555	6	7	0.95 (0.32–2.86)	.950

NOTE. Significant gene-based associations (false discovery rate (FDR)<0.01) are bolded. indel, insertion/deletion.
[a]OR refers to odds ratio compared by 2 × 2 table without adjustment by principal components.
[b]Gene-based *P*-valued for CMC burden test (score test) with 3 principal components as covariates.

mutant mice according to the Mouse Genomics Informatics database. Our data irrefutably showed that S-HSCR patients had a significantly increased burden of rare variants in known ENS genes (SKAT-O set-based $P = 3.9 \times 10^{-4}$ for damaging and $P = 7.8 \times 10^{-5}$ for all protein-altering rare-variants) (Table 1). In particular, all rare protein-altering variants in *RET* (Combined and Multivariate Collapsing [CMC] gene-based $P = 1.2 \times 10^{-5}$) (Supplementary Table 4) and damaging rare variants in *EDNRB* ($P = 1.9 \times 10^{-4}$), *ERBB2* ($P = 7.7 \times 10^{-3}$), and *GFRA1* ($P = .019$) were overrepresented in S-HSCR cases (false discovery rate, <0.1), thus confirming the role of the 3 already known pathways involved in ENS development. Three large case-unique deletions (from 245 kilobases to 16.7 megabases) encompassed *EDNRB* (Supplementary Table 5 and Supplementary Figure 1).

ENS Genes and Their Interactome

To investigate the oligogenic nature of HSCR, we explored if rare variants in other genes within the same pathways with the 3 aforementioned significant ENS genes also increased risk, either synergistically or independently, for HSCR. We identified 13 S-HSCR patients with double hits of rare variants among the known ENS genes (all rare protein-altering variants for *RET* and rare damaging variants for other ENS genes) and none in control individuals (Supplementary Table 6). Half of them are female (n = 7), and nearly all have rare variants affecting multiple known ENS pathways. Although our previous sequencing study of L-HSCR trios showed that multiple rare variants could occur in genes within the same protein–protein interacting network, we next tested for rare variants burden among the interacting partners of the 3 major HSCR genes, that is, *RET*,

Table 2. HSCR-Associated Variants Include the Common Variant rs2435357 in Intron 1 of *RET* and Low-Frequency Missense Variant rs9282834 Encoding RET D489N

Common Variant Risk Allelic Dosage[a]	Rare Variants		Count of S-HSCR (%)	OR (95% CI)	*P*
	Damaging[b]	Benign			
≥2	+	–	5 (1.1)	17.57 (1.91–161.41)	**.01**
	–	+	38 (8.6)	24.17 (9.63–60.64)	**1.2×10^{-11}**
	–	–	291 (65.7)	10.19 (6.25–16.63)	**1.5×10^{-20}**
1	+	–	11 (2.5)	7.12 (2.48–20.40)	**2.6×10^{-4}**
	–	+	5 (1.1)	2.99 (0.86–10.36)	.08
	–	–	64 (14.4)	1.21 (0.72–2.03)	.48
0	+	–	3 (0.7)	4.50 (0.84–24.19)	.08
	–	+	1 (0.2)	0.99 (0.10–9.39)	.99
	–	–	25 (5.6)	1	—

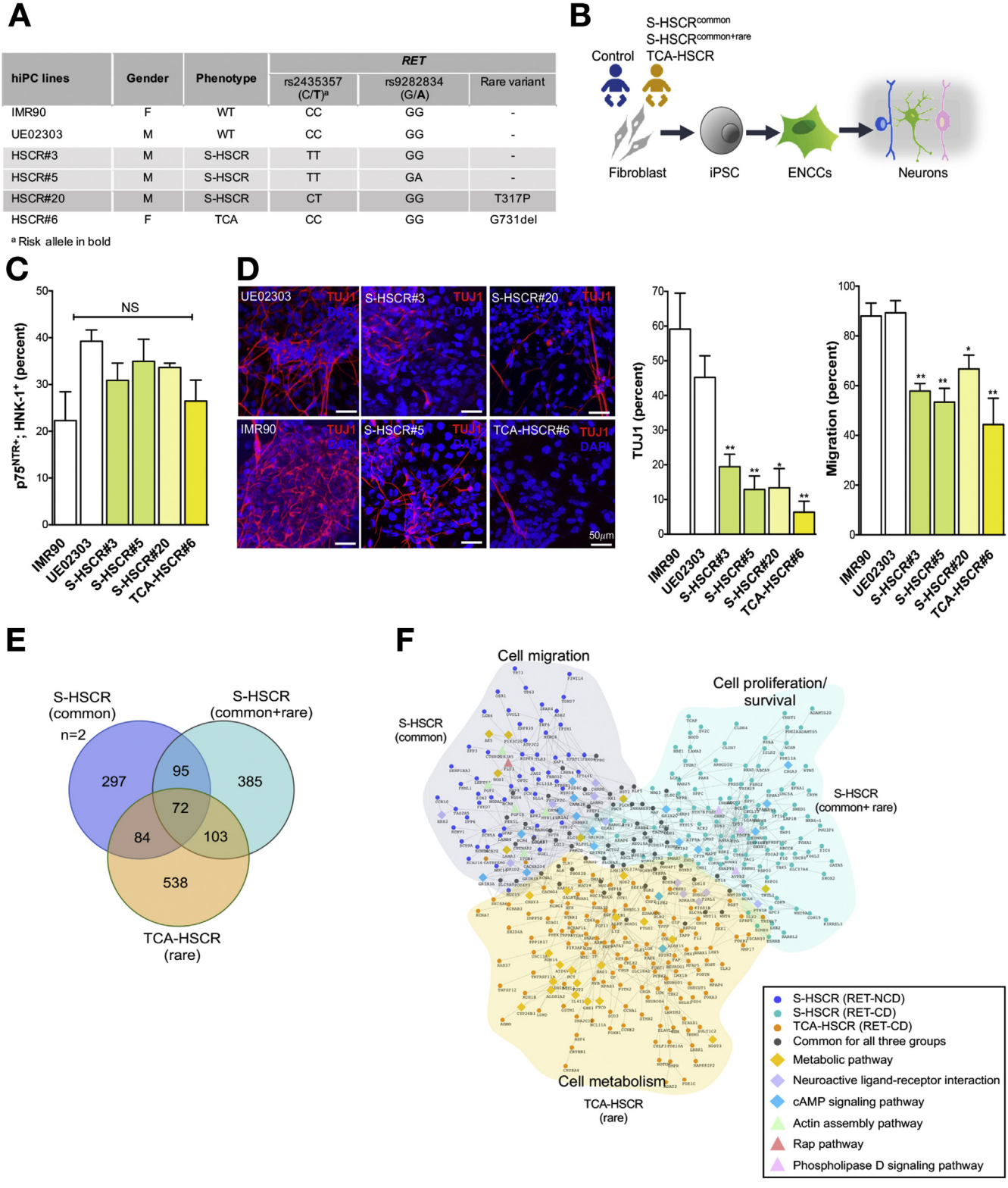
NOTE. Significant associations are bolded.
[a]≥2 indicates more than 2 risk alleles.
[b]Rare variants are classified into damaging and benign according to in silico prediction.

EDNRB, and *ERBB2*. Among the 87 interactome genes tested, 2 genes encoding the ERBB2-interacting partners, *ITGB4* ($P = 1.04 \times 10^{-3}$) and *PTK2* (also interacts with RET and ITGB4; $P = 1.39 \times 10^{-3}$), were also significantly enriched with damaging rare variants (false discovery rate, 0.1) (Figure 1*C* and *D* and Supplementary Table 7).

A

hiPC lines	Gender	Phenotype	*RET*		
			rs2435357 (C/**T**)[a]	rs9282834 (G/**A**)	Rare variant
IMR90	F	WT	CC	GG	-
UE02303	M	WT	CC	GG	-
HSCR#3	M	S-HSCR	TT	GG	-
HSCR#5	M	S-HSCR	TT	GA	-
HSCR#20	M	S-HSCR	CT	GG	T317P
HSCR#6	F	TCA	CC	GG	G731del

[a] Risk allele in bold



The RET-Sensitized Genetic Background in HSCR

Unlike *EDNRB* and *ERBB2*, many of the variants identified in *RET* were predicted to be benign, suggesting that the disease-associated rare variants in *RET* are not necessarily deleterious per se. To test the hypothesis, we assessed the risk conferred by *RET* rare variants across various levels of genetic predisposition conferred by the presence of common HSCR-associated *RET* risk alleles (T allele for rs2435357 and A allele for rs9282834).[4] Overall, we detected significant increased risk for individuals with 2 or more *RET* common risk alleles and for individuals with 1 common and 1 damaging rare variants (Table 2). *RET* variants predicted to be benign were predominately enriched in patients who had at least 2 *RET* risk alleles, further increasing their risk to S-HSCR at least 2-fold in addition to the effect of the common high-risk allele (dosage-specific OR, 2.34; 95% CI, 1.01–5.44; $P_{\text{dosage}} = .049$). On the other hand, nearly 58% of patients carrying *RET*-damaging variants have a single common-risk allele occurring mostly in *trans* compared with <17% for other patients carrying no or benign rare variants (Figure 1*E*). For these patients with a double hit of common and rare variants, the damaging *RET* variants conferred a nearly 5-fold increase in the risk of S-HSCR (dosage-specific OR, 4.82; 95% CI, 1.82–12.77; $P_{\text{dosage}} = 1.6 \times 10^{-3}$). Interestingly, 89% of these damaging *RET* variants are missense changes, whereas 2 out of 3 S-HSCR cases without any common risk allele but only rare damaging *RET* variant predisposition carries protein-truncating variants. This possibly reflects a different genetic architecture, such that the common variant has low impact on penetrance of rare variants in case of haploinsufficiency induced by these truncating mutations. Disregarding the truncating variants, no increase in disease risk was detected for individuals carrying only a single *RET* missense mutation or a common risk allele.

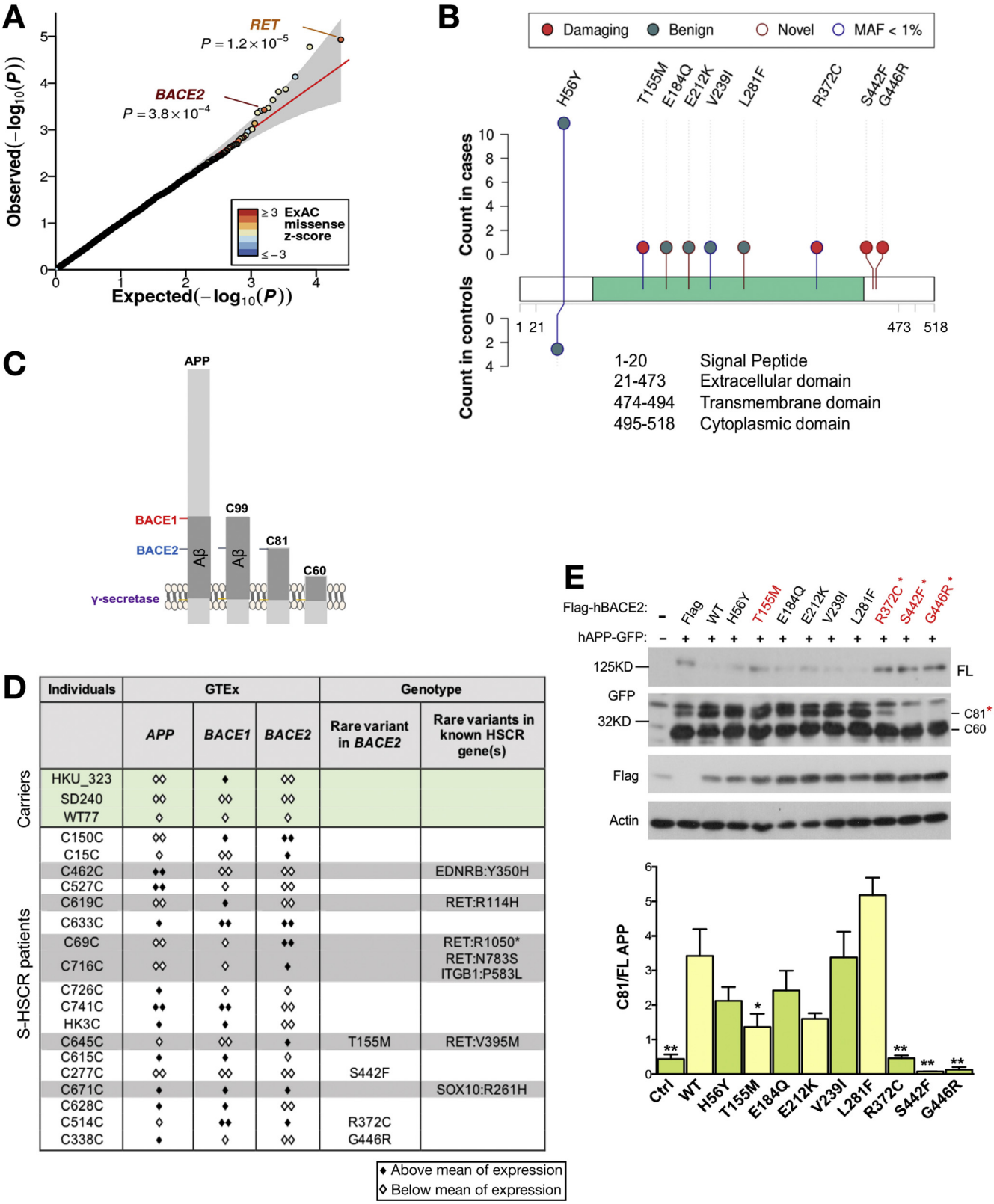
Dysregulation of Common and Distinctive Biological Pathways in HSCR Patients With Different RET-Sensitized Genetic Backgrounds

We further made use of hiPSC-based disease model[7] to elucidate the underlying biological pathways associated with the *RET* variants. A total of 6 hiPSC lines were used for comparison, including 2 control hiPSC lines derived from 2 healthy individuals (1 female and 1 male) carrying non-risk *RET* alleles (CC) in rs2435357; 2 S-HSCR patients carrying 2 common high-risk alleles (TT) in rs2435357; 1 S-HSCR patient harboring a rare damaging variant in *RET*, encoding T317P and a risk *RET* allele (CT in rs2435357); and 1 TCA-HSCR patient with an in-frame deletion rare variant in *RET* (G731del), representing the 3 major subclasses of *RET* variants identified in the HSCR patients (Figure 2*A*). All the hiPSC lines were derived from the fibroblasts of healthy individuals (control) or HSCR patients, and the corresponding genotype in each hiPSC line was confirmed by Sanger sequencing. Using these hiPSC lines and our established in vitro differentiation protocol,[7] ENCCs carrying the exact genetic makeup of the patients were obtained (Figure 2*B*). The neuronal differentiation and migration capabilities of ENCCs derived from each hiPSC line were assessed using in vitro differentiation and scratch assays as described.[7] Consistently, the control and HSCR-iPSC lines could generate comparable yields of ENCCs as marked by the 2 neural crest markers ($p75^{NTR}$ and HNK-1) (Figure 2*C*), but all the HSCR-ENCCs were less competent to make neurons and migrate than those of controls (Figure 2*D*). In particular, the differentiation defect of ENCCs derived from the TCA-HSCR (HSCR 6) iPSC were found relatively more severe than those from S-HSCR ENCCs (HSCR 3, HSCR 5, and HSCR 20), as monitored based on the formation of neuron-like processes and the expression of the pan-neuronal marker TUJ1.

We then sequenced the transcriptomes of the ENCCs derived from the control and HSCR-iPSC lines to delineate the biological pathways underlying these cellular defects. To identify the most relevant pathways attributed by the genetic lesions with the relatively small sample size, we regrouped the S-HSCR samples according to their variant type in the *RET* gene: control with low-risk *RET* alleles (IMR90 and UE02303); $\text{S-HSCR}^{\text{common+rare}}$ with a common high-risk allele and a damaging rare variant in *RET* (HSCR 20), and $\text{S-HSCR}^{\text{common}}$ with 2 common *RET* risk alleles (HSCR 3 and HSCR 5). A TCA-HSCR with an in-frame deletion rare variant in *RET* (HSCR 6) was also included for comparison. Cells derived from 2 independent hiPSC lines in the control and S-HSCR (common) groups were considered as replicates. From this analysis, we identified 655 and 548 differentially expressed genes (DEGs) in the $\text{S-HSCR}^{\text{common+rare}}$ and $\text{S-HSCR}^{\text{common}}$ groups, respectively, compared

Figure 2. Common and distinctive transcriptomic profiles of HSCR-iPSC–derived ENCCs. (*A*) Overview of healthy and diseased hiPSC lines used for the functional analyses and RNA sequencing. (*B*) Schematics show the generation of the control and HSCR-iPSC ENCCs and the hiPSC-derived enteric neurons. (*C*) Bar chart shows the ENCC yield from each hiPSC line (mean percentage of $p75^{NTR+}$ HNK-1^{+} cells ± SEM from 4–6 independent experiments). (*D*) FACS-enriched $p75^{NTR+}$ HNK-1^{+} ENCCs were directed to the neuronal lineage. hiPSC-derived enteric neurons were detected based on the expression of a pan-neuronal marker (TUJ1). Scratch assays were performed to measure the migratory ability of iPSC-derived ENCCs and the wound closure measured as the percentage of scar width over time (18 hours). The bar charts show mean ± SEM from 3–7 independent experiments. $^{*}P < .05$, $^{**}P < .001$ indicate significant difference from the healthy control individuals. (*E*) Transcriptional profiles of $p75^{NTR+}$HNK-1^{+} ENCCs derived from the control and HSCR-hiPSC lines were obtained by RNA sequencing. The Venn diagram shows DEGs commonly and uniquely found in $\text{S-HSCR}^{\text{common+rare}}$, $\text{S-HSCR}^{\text{common}}$ and TCA-HSCR groups compared with the control ENCCs. (*F*) GeneMANIA shows genes (network nodes) in the association networks created using the DEGs. Genes belonging to different GO terms with the involved biological processes indicated. FACS, fluorescence-activated cell sorter; NS, not significant; SEM, standard error of the mean.

with the control group (log2 fold change ≥ 1.5; adjusted $P <$ .05), where 167 genes were dysregulated in both groups (Figure 2*E* and Supplementary Table 8). When we included the TCA-HSCR line for comparison, we found that DEGs exhibit extensive interactions within each group and among the 3 groups (TCA-HSCR, S-HSCR$^{common+rare}$, and S-HSCR^{common}). For instance, the pathways related to the RET-GDNF signaling pathway (neuroactive ligand-receptor



	Individuals	GTEx			Genotype	
		APP	*BACE1*	*BACE2*	Rare variant in *BACE2*	Rare variants in known HSCR gene(s)
Carriers	HKU_323	◊◊	♦	◊◊		
	SD240	◊◊	◊◊	◊◊		
	WT77	◊	◊	◊		
S-HSCR patients	C150C	◊◊	♦	♦♦		
	C15C	◊	◊◊	♦		
	C462C	♦♦	◊◊	◊◊		EDNRB:Y350H
	C527C	♦♦	◊	◊◊		
	C619C	◊◊	♦	◊◊		RET:R114H
	C633C	♦	♦♦	♦♦		
	C69C	◊◊	◊	♦♦		RET:R1050*
	C716C	◊◊	◊	♦		RET:N783S ITGB1:P583L
	C726C	♦	◊	◊		
	C741C	♦♦	♦♦	◊◊		
	HK3C	♦	♦	◊◊		
	C645C	◊	◊◊	♦	T155M	RET:V395M
	C615C	♦	♦	◊		
	C277C	◊◊	◊◊	◊◊	S442F	
	C671C	♦	♦	♦		SOX10:R261H
	C628C	♦	♦	◊◊		
	C514C	◊	♦♦	♦	R372C	
	C338C	♦	◊	◊◊	G446R	

♦ Above mean of expression
◊ Below mean of expression

interaction and cyclic adenosine monophosphate signaling pathways), and the metabolic pathway Gene Ontology (GO) terms were enriched in all groups. Additional DEGs belonging to PLD1 signaling pathway GO terms (mediating cell growth/survival and differentiation) were identified in the S-HSCR$^{common+rare}$ group, whereas GO terms related to cell migration (eg, Rap1 signaling pathway and actin assembly molecules) were enriched in the S-HSCR^{common} group (Figure 2*F*). These data suggest that RET-GDNF signaling is the main signaling pathway implicated in HSCR pathogenesis, and dysregulation of distinctive biological pathways in addition to the common pathways is likely involved in these 3 subgroups of HSCR patients, which may account for disease severity and/or phenotypes.

Beyond ENS Genes: Association of BACE2 With HSCR

Using *RET* as a model, we tested for association of all genes expressed in the ENCC (n = 11,898) considering all rare protein-alternating variants. In addition to *RET*, *BACE2* (OR, 7.3; 95% CI, 2.2–24.9; $P = 3.8 \times 10^{-4}$) was considered to be intolerant to missense changes (Exome Aggregation Consortium missense *z*-score > 1.96) among the top HSCR-associated genes (Figure 3*A*). To validate this finding, we Sanger-sequenced an independent set of 534 ethnically matched control individuals. The combined analysis of WGS and Sanger sequencing confidently established the association between *BACE2* and HSCR ($P = 2.9 \times 10^{-6}$), which surpassed multiple testing association of 11,898 genes (Supplementary Table 9).

BACE2 is a homolog of *BACE1,* and it encodes a transmembrane aspartyl protease, β-secretase 2. A total of 9 rare protein-alternating variants, all validated by Sanger sequencing, were detected in this gene, affecting 18 HSCR patients and 3 control individuals. Eight variants were found only in patients. Six of these 8 variants mapped to the peptidase domain, potentially affecting its protease function. In silico prediction suggested that variants encoding T155M, R372C, S442F, and G446R may alter protein function (damaging). The remaining variant, encoding H56Y, falls into the prodomain that assists in protein folding, and it was found in both S-HSCR patients and control individuals (Figure 3*B*).

BACE2 has a broad substrate specificity similar to BACE1.[10–12] APP is highly expressed in nerve tissue, and it contains recognition sites for BACE1 and BACE2 (Figure 3*C*).[13,14] BACE1 and γ-secretase cleave APP sequentially and produce amyloid β (Aβ) and C60, whereas BACE2 cleaves APP in the Aβ region and eventually generates C81 fragment and prevents Aβ formation. Accumulation of Aβ induces neuronal death, representing the underlying cause of Alzheimer disease.[10,11,14,15] To explore the potential implication of BACE1-APP-BACE2 pathway in HSCR pathogenesis, we first examined how the expression of *BACE1*, *BACE2*, and *APP* may confer risk of HSCR. The relative expression levels of these genes in the individuals who carry *BACE2* rare variants and noncarriers were imputed based on the Genotype-Tissue Expression (GTEx) neural tissues (tibial nerve) or whole blood models using PrediXcan.[16] The relative expression levels of these genes in the individuals carrying *BACE2* variant(s) were compared with their mean expression level in the control group (493 individuals). All 3 of these genes were predicted to be expressed at relatively lower levels in the 3 non-HSCR carriers. On the other hand, among the 18 S-HSCR patients, most (1) exhibited higher imputed expression levels in at least 1 of these genes, (2) carried additional mutation(s) in known HSCR gene(s), and/or (3) harbored 1 of the 4 damaging variants in *BACE2* (Figure 3*D*). These observations poised us to further investigate how the *BACE2* variants alter APP cleavage. To this end, we overexpressed wild-type or mutant BACE2 (Flag-tagged) together with green-fluorescent protein–tagged APP in a human embryonic kidney cell line (293FT). The protease activity of BACE2 was then monitored based on the ratios of C81 fragment to the full-length APP (FL) using Western blot analysis. As shown in Figure 3*E*, the 4 rare variants encoding T155M, R372C, S442F, and G446R significantly reduced the APP processing activity of BACE2. In particular, S442F and G446R substitutions almost completely abolished the protease activity of BACE2. Both of these 2 rare variants are residing adjacent to the transmembrane domain of BACE2, and that may interfere with the membrane docking of BACE2 (Supplementary Figure 2).

BACE2 Is Crucial for Neuronal Survival, but It Is Not Required for the Derivation of ENCCs From hiPSC

To elucidate how the *BACE2* variants interrupt the ENS development and eventually lead to HSCR disease, we used hiPSC to generate ENCCs (Figure 4*A*) and then subsequently directed hiPSC-derived ENCCs to enteric neurons, modeling the progressive differentiation processes that occur during ENS development.[7] To this end, we generated 2 isogenic

Figure 3. Identification of novel rare variants in *BACE2*. (*A*) Q-Q plot of rare-variant association analysis. (*B*) Mutation profile of *BACE2*. *Red circle*, putative deleterious variant. Counts of alternative allele in patients (*top*) and control individuals (*bottom*) are shown. (*C*) Schematic illustration of the cleavage sites for BACE1, BACE2, and γ-secretase on amyloid precursor protein (APP) and the leaving C-terminal fragment after cleavage by BACE1 (Aβ), BACE2 (C81), or γ-secretase (C60). (*D*) Table summarizes the expression quantitative loci –predicted expression levels of *BACE1*, *BACE2*, and *APP* in individuals carrying *BACE2* rare variant (3 control individuals and 18 HSCR patients), relative to the mean expression level in the control group (493 individuals). (*E*) Western blot analysis: Flag-tagged wild-type (WT) and BACE2 mutants were overexpressed in the 293FT human cell line. The full-length and cleaved APP (GFP-tagged) were detected. C81 fragment represents BACE2-cleaved APP, as marked by an asterisk. The enzymatic cleavage ability of BACE2 was analyzed based on the level of C81 fragment over the full length (FL) APP, and the quantitative results are shown in the bar chart (mean values ± SEM across 3 independent experiments). GFP, green-fluorescent protein; SEM, standard error of the mean.

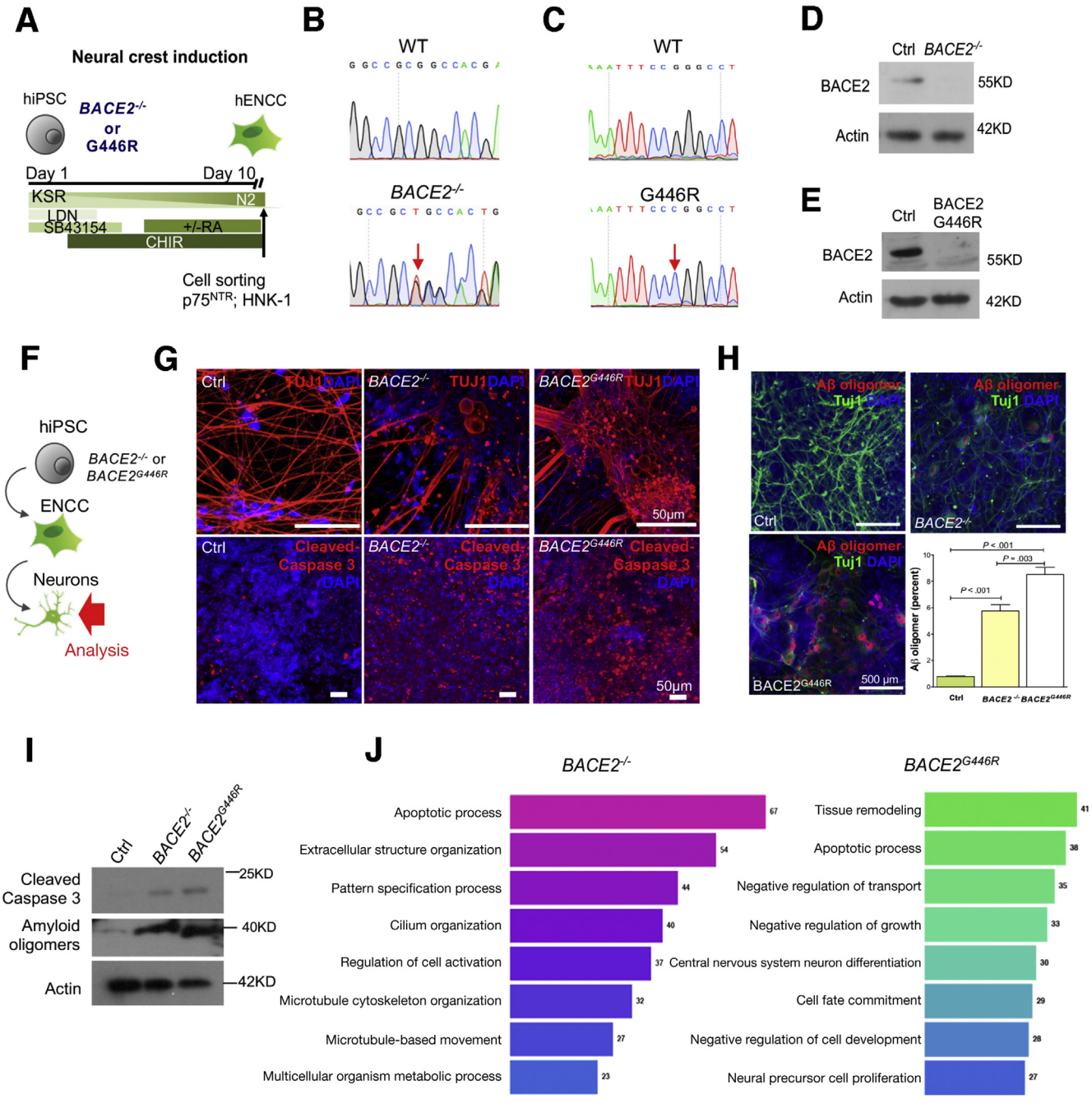


Figure 4. *BACE2* deficiency promotes Aβ accumulation and induces apoptosis in hiPSC-derived ENS neurons. (*A*) Stepwise differentiation protocol for the generation of ENCCs from hiPSC. Homozygous deletion and *BACE2*^{G446R} mutant hiPSC lines were generated using CRISPR/Cas9. Sanger sequencing confirms (*B*) the deletion and (*C*) G>C substitution in *BACE2* locus of hiPSC lines. (*D*, *E*) Western blot analyses of protein lysates from wild type (Ctrl), *BACE2*$^{-/-}$ and *BACE2*^{G446R} hiPSC-derived ENCCs. Actin was used as a loading control. (*F*) Schematic shows the generation of ENS neurons from the hiPSC-derived ENCCs. Immunocytochemistry analyses of the neurons derived from the control group, *BACE2*$^{-/-}$ or *BACE2*^{G446R} hiPSC at day 30 with (*G*) TUJ1 or cleaved caspase 3 antibody (*red*); (*H*) TUJ1 (*green*) and Aβ oligomer (*red*), counterstained with DAPI (*blue*). Percentages of ENS neurons with Aβ oligomer accumulation over the total number of TUJ1^{+} neurons in each group are shown in the bar chart (mean ± SEM from 3 independent experiments). (*I*) Western blot analyses of protein lysates from wild type (Ctrl), *BACE2*$^{-/-}$, and *BACE2*^{G446R} hiPSC-derived ENS neurons at day 30. Actin was used as a loading control. (*J*) GO terms enriched in the *BACE2*$^{-/-}$ and *BACE2*^{G446R} hiPSC-derived ENS neurons. Ctrl, control; DAPI, 4′,6-diamidino-2-phenylindole; SEM, standard error of the mean.

hiPSC lines using clustered regularly interspaced palindromic repeats (ie, CRISPR)/CRISPR-associated protein 9 (ie, Cas9)–mediated genome editing technology: 1 carried a *BACE2*-null mutation ($BACE2^{-/-}$) (Figure 4*B*), and the other harbored a *BACE2*-damaging rare variant ($BACE2^{G446R}$) (Figure 4*C*) (Supplementary Figure 3). Western blotting with BACE2 antibody against the C-terminus of BACE2 was performed to confirm the absence of wild-type BACE2 in ENCCs derived from the $BACE2^{-/-}$ (Figure 4*D*) and $BACE2^{G446R}$ (Figure 4*E*) hiPSC lines. $BACE2^{-/-}$ and $BACE2^{G446R}$ hiPSC lines could efficiently generate ENCCs with comparable yield, seen in the control hiPSC line, as measured by flow cytometry using antibodies against HNK-1 and $p75^{NTR}$ (Supplementary Figure 4*A*). The majority of ENCCs derived from these hiPSC lines were co-expressing the 2 key ENCC markers, RET and SOX10 (Supplementary Figure 4), suggesting that BACE2 is not required for ENCC derivation from hiPSC. Although the yield of ENCCs was not affected, we found that $BACE2^{-/-}$ or $BACE2^{G446R}$ ENCCs migrate significantly faster than the control cells, as illustrated in the scratch assays (Supplementary Figure 5). This is concordant with a previous study, which showed that knocking down *BACE2* in $EDNRB^{-/-}$ hiPSC-derived ENCCs can improve the cell migration.[17] Nevertheless, the enhanced migratory ability of ENCCs is unlikely the mechanism underlying HSCR pathogenesis.

Expression quantitative loci prediction suggests that the relative expression levels of *BACE1*, *BACE2*, and *APP* may affect HSCR susceptibility. Indeed, dynamic expression of these proteins was observed in hiPSC, hiPSC-derived ENCCs, and their neuronal derivatives (ENS neurons), representing the 3 key developmental windows of the ENS. Western blot analysis showed that BACE1 is expressed in both hiPSC-derived ENCCs and ENS neuron but not in hiPSCs. A low level of BACE2, on the other hand, was detected in hiPSCs, and its expression levels were elevated over time when hiPSCs were differentiating into ENCCs and then to ENS neurons, with the highest expression in the ENS neurons. APP expression levels were comparable in hiPSCs and ENCCs, but were was remarkably reduced in ENS neurons, inversely correlated to the BACE2 level (Supplementary Figure 6). Co-expression of these 3 proteins in the enteric ganglion of human colon was consistently observed based on the immunohistochemistry data available in The Human Protein Atlas database (Supplementary Figure 7), supporting the causal role of these molecules in ENS development. Therefore, we further examined how the HSCR-associated rare variants in *BACE2* might affect endogenous APP processing in ENS neurons. We directed the control, $BACE2^{-/-}$, or $BACE2^{G446R}$ ENCCs to the neuronal lineage by culturing them in neuronal differentiation medium for 30 days (Figure 4*F*). Immunocytochemistry analyses showed that both the control and *BACE2* mutant ENCCs can give rise to neurons efficiently, expressing the pan-neuronal marker β-III-tubulin (TUJ1). By day 30 of neuronal differentiation, $BACE2^{-/-}$ and $BACE2^{G446R}$ ENS neurons underwent obvious morphologic changes, with enlarged cell bodies and granules. Subsequent immunocytochemistry using antibody against cleaved caspase 3 further showed that many of the $BACE2^{-/-}$ and $BACE2^{G446R}$ ENS neurons are undergoing apoptosis and expressing high levels of cleaved caspase 3 (Figure 4*G*). Accumulation of Aβ oligomers was observed in the tightly arranged $BACE2^{-/-}$ and $BACE2^{G446R}$ ENS neurons, but such intensively stained granules were not found in the control cells carrying wild-type *BACE2* (Figure 4*H*). $BACE2^{G446R}$ ENS neurons consistently exhibited a more severe phenotype than $BACE2^{-/-}$ ENS neurons, suggesting a potential dominant negative effect of $BACE2^{G446R}$. The elevated level of cleaved caspase 3 and amyloid oligomers in $BACE2^{-/-}$ and $BACE2^{G446R}$ ENS neurons was further confirmed with Western blotting (Figure 4*I*). Genome-wide gene expression profiles of the control and mutant ENS neurons were obtained by performing bulk RNA sequencing. We identified 2604 DEGs in $BACE2^{G446R}$ ENS neurons, representing 1044 up-regulated and 1560 down-regulated genes, whereas 1353 and 1875 genes were up- and down-regulated, respectively, in $BACE2^{-/-}$ ENS neurons compared with the control (hiPSC-ENS neurons) (log2 fold change $\geq$ 1.5; adjusted $P < .05$) (Supplementary Table 10 and Supplementary Figure 8). GO enrichment analysis showed enrichment of apoptosis GO terms in both $BACE2^{-/-}$ and $BACE2^{G446R}$ ENS neurons (Figure 4*J*), further supporting the idea that BACE2 deficiency may induce apoptosis of ENS neurons.

To directly show the involvement of APP in the neuronal death induced by $BACE2^{G446R}$, we knocked out *APP* in the $BACE2^{G446R}$ hiPSC line to generate a double-mutant line ($BACE2^{G446R}APP^{-/-}$) using CRISPR/Cas9 (Figure 5*A* and Supplementary Figure 9). Deletion of *APP* significantly improved survival of hiPSC-derived enteric neurons, as evidenced by reduced activated caspase 3 staining (Figure 5*B* and *D*). As expected, APP deletion eliminated Aβ oligomer, as illustrated by immunocytochemistry (Figure 5*C*) and Western blot analyses (Figure 5*D*). In summary, our data show that BACE2 can protect the ENS neurons from undergoing apoptosis by properly processing APP and preventing the Aβ accumulation (Figure 5*E*).

Discussion

Our WGS analysis of 443 S-HSCR patients and 493 ethnically matched control individuals provides a glimpse into the genetic architecture of HSCR, in particular the short-segment type, which has long been regarded as a paradigm for the study of oligogenic and complex diseases. The functional follow-up of the most representative genetic features using an hiPSC-based model has led to the discovery of a new HSCR disease mechanism and helped dissect the effects of the interplay among common and rare variants on the phenotype. This study presents a new paradigm not only for the understanding oligogenic disorders but also for the implementation of new therapeutic avenues for these diseases.

In our genetic study using a population-based association approach, not only did we confirm the roles of *RET* and *EDNRB* as the 2 major HSCR genes, but we also firmly established *ERBB-NRG* as a core HSCR pathway with significant contribution of both common (*NRG1*) and rare

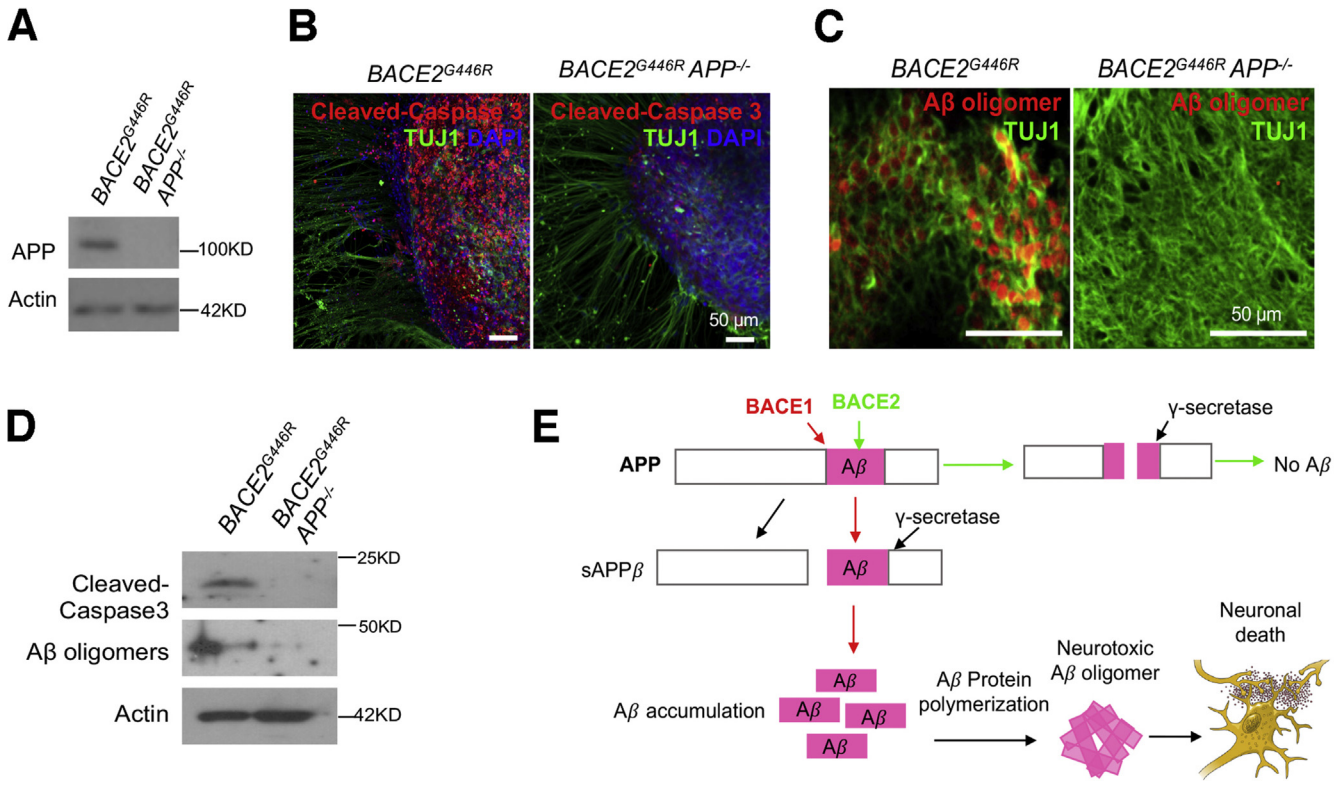


Figure 5. Deletion of *APP* improves the survival of *BACE2*^{G446R} hiPSC-derived ENS neurons. (*A*) Western blot analysis of protein lysates from *BACE2*^{G446R} and *BACE2*^{G446R}/*APP*$^{-/-}$ hiPSC-derived ENCCs confirms the complete knockout of *APP*. Actin was used as a loading control. Immunocytochemistry of ENS neurons derived from *BACE2*^{G446R} and *BACE2*^{G446R}/*APP*$^{-/-}$ hiPSC at day3 0 with (*B*) TUJ1 and cleaved caspase 3 antibodies countered stained with DAPI and (*C*) Aβ oligomer. (*D*) Western blot analyses of protein lysates from *BACE2*^{G446R} and *BACE2*^{G446R}/*APP*$^{-/-}$ hPSC derived neurons at day 30. Actin was used as a loading control. (*E*) Schematic illustrates the generation and clearance of the Aβ oligomers by BACE1 and BACE2, respectively. DAPI, 4′,6-diamidino-2-phenylindole; hPSC, human pluripotent stem cell.

(*ERBB2*) variants. Further exploration of the oligogenic inheritance with the ERBB2 interactome yielded novel candidate genes, *ITGB4* and *PTK2,* involved in focal adhesion. *ITGB4,* encoding β4-integrin, is a homologue of a known ENS-associated gene, *ITGB1*. Both β1- and β4-integrins are important receptors that mediate adhesion to the extracellular matrix through interaction with extracellular matrix proteins such as laminin and collagen. It has been shown that focal adhesion kinase encoded by *PTK2* can be activated by β4-integrin. In addition, focal adhesion kinase can mediate RET signaling by direct binding,[18] which might provide the functional convergence between the ERBB2 and RET pathways. The unique occurrence of double hits across multiple known HSCR pathways corroborated our observation that altered coordination or regulation of the core ENS pathways and their core interactome (or regulome) might be disease causing. Scrutinizing these core interactors with an integrative multi-omics analysis, including genetics, cell-type specific transcriptomics, and regulome, will prove to be useful in understanding the variable expressivity and disease mechanisms of complex, particularly oligogenic, disorders.

In fact, a genetic model that takes into account the interaction between the common regulatory, rare damaging, and benign *RET* variants provided a better fit than considering them additively. We found that the modifying effect of the *RET* common risk allele on disease risk depends on the predicted pathogenicity of the coding variants. In the latter, the *RET* risk alleles reducing gene expression occurred mostly in *trans* in HSCR patients, which might increase the penetrance of these pathogenic rare coding variants by reducing the amount of functional transcripts and thereby explain the observed variable penetrance. Such an epistatic effect not only supports a sensitized genetic background for *RET* but also suggests that the effects of rare variants in any gene should not be considered in isolation from the polygenic background. Inclusion of benign, rare missense changes might therefore be more powerful in identifying novel disease-susceptibility genes for complex polygenic diseases.

Our population-based rare variant association study identified *BACE2* as a novel HSCR gene. Unlike BACE1, the physiologic functions of BACE2 are not well characterized. *Bace2*-knockout mice display an overall healthy phenotype.[19] Here, we used hiPSC to model the ENS development in vitro and defined a novel role of BACE2 in the hiPSC-derived enteric neurons, where BACE2 abolishes Aβ production and prevents the accumulation of amyloid to protect neurons from undergoing apoptosis. More important, our data also indicated that the BACE1-APP-BACE2 pathway

would be another key pathway underlying HSCR pathogenesis. Similar to other HSCR pathways, double hits interrupting more than 1 gene within this pathway are likely required to cause the disease. It may not be limited to these 3 core genes (*BACE1*, *BACE2*, and *APP*) but might include other genes involved in Aβ production or processing. For instance, PLD1 is involved in Aβ production through regulating APP metabolism and trafficking.[20,21] Concordantly, our whole-genome association analysis has identified *PLD1* as a new susceptibility locus for HSCR, reinforcing the contribution of this new pathway in the HSCR pathogenesis.

An emerging view is that the genetic complexity of HSCR is substantially larger than expected, with likely epistasis of rare and common variants both within and between loci. Here, we showed the effect of rare variants among various ENS genes, particularly BACE1-APP-BACE2 pathway genes, and that some of these rare variants are modified by the genetic background sensitized by common regulatory variants such as *RET*. Whether a hit is from protein-altering rare or common-variant regulating gene expression, 2 or more hits are required to influence disease risk. Such epistasis highlights the importance of the underlying genetic background when assessing a patient's disease risk for complex disorders. Instead of testing the burden of rare variants and the association of common variants separately, integration of statistical methods that jointly consider both and are weighted on the effect or frequency and the use of hiPSC-based model might be advantageous. Last but not least, given the variable expressivity of HSCR, it can be challenging to decide which children should proceed with rectal biopsy for disease diagnosis. The genetic findings in our study can contribute to the model that incorporates genetic risk in HSCR risk prediction. For example, those with *RET* or *ZEB2* URVs, *EDNRB*-damaging mutations, or a double hit of damaging mutations in known ENS genes can be considered as having very high genetic risk. Similarly, polygenic risk of other rare variants, common variants, and their interactions can be modeled together in risk prediction. Because HSCR has long been regarded as an example of oligogenic disease, catalyzing the search for HSCR genes beyond the family approach may have profound implications for other rare disorders.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2018.09.012.

References

1. 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.
2. Badner JA, Sieber WK, Garver KL, et al. A genetic study of Hirschsprung disease. Am J Hum Genet 1990;46:568–580.
3. Emison ES, Garcia-Barcelo M, Grice EA, et al. Differential contributions of rare and common, coding and non-coding Ret mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet 2010;87:60–74.
4. Tang CS, Gui H, Kapoor A, et al. Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease. Hum Mol Genet 2016;25:5265–5275.
5. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–498.
6. Xue Y, Cai X, Wang L, et al. Generating a non-integrating human induced pluripotent stem cell bank from urine-derived cells. PLoS One 2013;8:e70573.
7. **Lai FP**, **Lau ST**, Wong JK, et al. Correction of Hirschsprung-Associated Mutations in Human Induced Pluripotent Stem Cells Via Clustered Regularly Interspaced Short Palindromic Repeats/Cas9, Restores Neural Crest Cell Function. Gastroenterology 2017;153:139–153.
8. **Gui H, Schriemer D, Cheng WW**, et al. Whole exome sequencing coupled with unbiased functional analysis reveals new Hirschsprung disease genes. Genome Biol 2017;18:48.
9. Tam PK, Garcia-Barcelo M. Genetic basis of Hirschsprung's disease. Pediatr Surg Int 2009;25:543–558.
10. Farzan M, Schnitzler CE, Vasilieva N, et al. BACE2, a beta-secretase homolog, cleaves at the beta site and within the amyloid-beta region of the amyloid-beta precursor protein. Proc Natl Acad Sci U S A 2000;97:9712–9717.
11. Vassar R, Kuhn PH, Haass C, et al. Function, therapeutic potential and cell biology of BACE proteases: current status and future prospects. J Neurochem 2014;130:4–28.
12. Yan R. Physiological functions of the beta-site amyloid precursor protein cleaving enzyme 1 and 2. Front Mol Neurosci 2017;10:97.
13. **Bergstrom P, Agholme L**, Nazir FH, et al. Amyloid precursor protein expression and processing are differentially regulated during cortical neuron differentiation. Sci Rep 2016;6:29200.
14. **Wang S, Bolos M**, Clark R, et al. Amyloid beta precursor protein regulates neuron survival and maturation in the adult mouse brain. Mol Cell Neurosci 2016;77:21–33.
15. Sun X, He G, Song W. BACE2, as a novel APP theta-secretase, is not responsible for the pathogenesis of Alzheimer's disease in Down syndrome. FASEB J 2006;20:1369–1376.
16. **Gamazon ER, Wheeler HE, Shah KP**, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091–1098.
17. Fattahi F, Steinbeck JA, Kriks S, et al. Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. Nature 2016;531:105–109.
18. Plaza-Menacho I, Morandi A, Mologni L, et al. Focal adhesion kinase (FAK) binds RET kinase via its FERM domain, priming a direct and reciprocal RET-FAK

transactivation mechanism. J Biol Chem 2011; 286:17292–17302.
19. **Dominguez D, Tournoy J, Hartmann D**, et al. Phenotypic and biochemical analyses of BACE1- and BACE2-deficient mice. J Biol Chem 2005;280: 30797–30806.
20. Cai D, Netzer WJ, Zhong M, et al. Presenilin-1 uses phospholipase D1 as a negative regulator of beta-amyloid formation. Proc Natl Acad Sci U S A 2006; 103:1941–1946.
21. Cai D, Zhong M, Wang R, et al. Phospholipase D1 corrects impaired betaAPP trafficking and neurite outgrowth in familial Alzheimer's disease-linked presenilin-1 mutant neurons. Proc Natl Acad Sci U S A 2006; 103:1936–1940.

Author names in bold designate shared co-first authorship.

Received May 28, 2018. Accepted September 5, 2018.

Reprint requests
Address requests for reprints to: Elly Sau-Wai Ngan, PhD, Department of Surgery, University of Hong Kong, Pokfulam, Faculty of Medicine Building, 21 Sassoon Road, Hong Kong, SAR, China. e-mail: engan@hku.hk; fax: 852 3917 9621; or Maria-Mercedes Garcia-Barcelo, Department of Surgery, University of Hong Kong, Pokfulam, Faculty of Medicine Building, 21 Sassoon Road, Hong Kong, SAR, China. e-mail: mmgarcia@hku.hk; fax: 852 3917 9621.

Acknowledgments
We are grateful to the numerous patients, their families, and referring physicians who have participated in these studies in our laboratories and to the numerous members of our laboratories for their valuable contributions over many years.

Author contributions: Peng Li, Frank Pui-Ling Lai, and Sing-Ting Lau performed functional studies. Clara Sze-man Tang, Alexander Xi Fu, Xuehan Zhuang, Kevin Yuk-Lap Yip, and Zhixin Li performed genetics and bioinformatics analyses. Man-Ting So did the genetic sample preparation and Sanger sequencing. Paul Kwang-Hang Tam, and Michelle Yu provided clinical information, together with Xuelai Liu, Ngoc D. Ngo, Xioping Miao, Xi Zhang, Bin Yi, Shaotao Tang, Xiobing Sun, Furen Zhang, Hong Liu, Qiji Liu, Ruizhong Zhang, Hualong Wang, Liuming Huang, Xiao Dong, Jinfa Tou, Kathy S.E. Cheah, Wanling Yang, and Zhenwei Yuan, who contributed to the clinical recruitment of patients and control individuals. Elly S.W. Ngan, Pak Chaung Sham, and Maria-Mercedes Garcia-Barcelo supervised the project and prepared the manuscript.

Conflicts of interest
The authors disclose no conflicts.

Funding
The work described in this article was substantially supported by Theme-Based Research Scheme (grant no. T12C-714/14-R.), University-Industry Collaboration Programme (UICP) (UIM/299), General Research Fund (HKU17109215 to Elly Sau-Wai Ngan and HKU17119514 to Paul Kwang-Hang Tam), Health Medical Research Fund (HMRF 02131866 and 01121516 to Maria-Mercedes Garcia-Barcelo, 04151966 to Clara Sze-man Tang, and 03143236 to Elly Sau-Wai Ngan), and Seed Fund for Basic Research of the University of Hong Kong (grant no. 201606159005 to Clara Sze-man Tang).

Supplementary Methods

Patients

The discovery cohort comprised 464 S-HSCR patients and 498 control individuals analyzed by WGS. All patients analyzed were diagnoised with sporadic HSCR, with no known family history of HSCR, and were recruited at hospitals in China (n = 341) and Hanoi, Vietnam (n = 102). Among these, 165 S-HSCR cases were analyzed for rare coding variants in RET,[1] and 98 S-HSCR patients were genotyped by Affymetrix 500K SNP array for genome-wide association analysis.[2,3] To avoid cofounding effects due to population stratification, control individuals were recruited from the same or nearby cities to match with the subpopulations of the patients. Samples failing heterozygosity, sex concordance, duplication, and relatedness were excluded. We further removed ethnic outliers as indicated by principal components analysis (PCA). After quality control, a total of 443 S-HSCR patients and 493 control individuals remained. PCA of the resulting genetic data indicated that patients and control individuals were well matched for ethnic origin (Supplementary Figure 10). Demographic information of these samples is summarized in Supplementary Table 11. The follow-up cohort included 534 ethnically matched control individuals subject to Sanger sequencing. Informed consent was obtained from all participants, and the study was approved by the institutional review board of the University of Hong Kong and the Hospital Authority (UW 13-225).

WGS and Variant Calling

All samples were whole-genome sequenced using Illumina HiSeq X Ten at Macrogen, Inc, (Korea) to a mean coverage of ×30. Raw sequence reads were first aligned to human reference genome (hg19) using Burrows–Wheeler Aligner.[4] After alignment, we achieved an average mapped read depth of ×26–×64 (median, ×36.6), and the breadth of coverage was highly comparable between patients and control individuals (Supplementary Figure 11). Aligned reads were then processed according to Genome Analysis Toolkit (GATK), version 3.4 best practices recommendations.[5] In brief, Picard was used for duplicate removal, and GATK was used for indel realignment and base quality score recalibration. Variants, both single nucleotide variants (SNVs) and insertion/deletion (indels), were called by GATK HaplotypeCaller. Variant-based quality control was initially carried out using GATK variant quality score recalibration. We selected the truth sensitivity tranches of 99.6% and 99.1% for SNVs and indels, respectively. Additionally, we applied genotype-level quality control using KGGseq,[6] which set low quality genotypes, that is, those with genotype quality <20 or covered by fewer than 8 reads (read depth [DP] <8) to missing to avoid false positive and negative calls. Finally, variants with call rate <0.9 or those violating Hardy–Weinberg equilibrium ($P < 1 \times 10^{-5}$) were removed. This resulted in a final call set of 33.4 million SNVs and 3.3 million indels, the majority of which are novel (61.5% for SNVs and 68.7% for indels). Transition-to-transversion ratio for novel SNVs is 2.10 across the whole genome and 3.00 when considering only the exome. A genotype concordance of 99.96% between duplicates was achieved, which showed the high quality of our variant call set for subsequent rare variant burden test.

Variant Annotation

The variant call set was annotated using KGGseq for protein function against the RefGene gene annotations, pathogenicity (eg, PolyPhen2, SIFT, and CADD), and population frequencies (e.g., 1000 Genomes Project phase 3, Exome Aggregation Consortium [ExAC, version 0.2] and National Heart, Lung, and Blood Institute [NHLBI] Exome Sequencing Project databases. We defined protein-truncating variants as variants that lead to (1) gain of the stop codon, (2) frameshift, and (3) alteration of the essential splice sites. Damaging variants include all protein-truncating variants, as well as missense variants or in-frame variants predicted to be deleterious by KGGseq, whereas benign variants are missense variants or in-frame variants predicted to be benign by KGGseq. Finally, protein-altering variants comprise both damaging and benign variants. Rare variants are those whose minor allele frequency (MAF) is <0.01 in any of the following public databases: 1000 Genomes Project phase 3; ExAC, version 0.2; and NHLBI Exome Sequencing Project databases.

Association Between URVs and HSCR

URV was defined as a singleton variant, that is, one that appeared only once in our whole data set, not present in dbSNP138, and not seen in public databases (1000 Genomes Project phase 3; ExAC, version 0.2, and NHLBI Exome Sequencing Project databases). We adopted the definition of *constrained gene* as described previously,[7] that is, intolerant of loss of function (ExAC pLI ≥ 0.9, n = 3488). Protein-truncating URVs are further divided into 2 types, one that can elicit nonsense-mediated decay (NMD) and another that escapes NMD (non-NMD). Non-NMD URVs are protein-truncating variants in which predicted stop codon occurs in the last exon or in the last 50 base pairs of the penultimate exon. For each type of URV, the URV score of each individual was constructed by summing the corresponding number of singleton URVs in genes expressed in ENCCs. Association tests between protein-truncating URVs, synonymous URVs, and HSCR were carried out by regressing the HSCR disease status on the corresponding URV score while adjusting for the first 3 principal components. To avoid confounding due to technical artefacts, the number of synonymous URVs was added as a covariate when assessing the impact of protein-truncating URVs (NMD and non-NMD).

Known Genes of ENS Development and Their Interactome

Genes displaying colonic aganglionosis and Hirschsprung phenotype in mutant mice according to the Mouse

Genomics Informatics were considered as known ENS genes. ENS interactome was defined by genes encoding proteins that show protein-protein interaction with known ENS genes in the InWeb database.

Copy Number Variants (CNVs) Overlapping the Known ENS Genes

For the known ENS genes, we detected overlapping CNVs using 4 different yet complementary software programs: CNVnator,[8] Seeksv,[9] DELLY,[10] and LUMPY,[11] to maximize the accuracy. Default parameters were used in all software except for CNVnator, in which the bin size used to partition the genome was set to 50 base pairs (bp). Only those CNVs (1) larger than 50 bp, (2) called by at least 3 software programs, and (3) supported by at least 10 soft-clip reads were selected for downstream analysis. We used BEDTools[12] to calculate the overlap of CNVs among individuals and across regions. CNVs with >50% of length overlapping the centromere or short repeat regions were excluded. *Case-unique CNVs* were defined as CNVs with <50% reciprocal overlap against all CNVs found in the WGS control samples and those documented in the Database of Genomic Variants[13] and in the population controls of DECIPHER[14] at the known ENS loci.

Gene-Based and Gene Set–Based Burden Test for Rare Variants

For the set of known ENS genes, we first assessed the enrichment of (1) damaging and (2) all rare protein-altering variants collectively in patient compared with control samples. Gene-based and gene set–based association tests, adjusted for the first 3 principal components from the aforementioned PCA, were carried out using combined multivariate and collapsing test[15] and SKAT-O, respectively, by rvtests.[16] Genes and gene sets passing false discovery rate < 0.1 were considered to be significant.

To test for exome-wide burden of other genes expressed in hiPSC-induced enteric neural crest cells (ENCC) from IMR90 (see "Neural Crest Induction" section), we performed gene-based association test for all rare protein-altering variants, regardless of the in silico deleterious prediction. Genes with association $P < 4.0 \times 10^{-6}$, equivalent to multiple testing of 11,898 ENCC-expressed genes, were considered to be significantly associated with HSCR. For the *BACE2* association where no genome-wide genotype data was available for the follow-up Sanger-sequenced samples, the analysis was adjusted for country (China or Vietnam) and sample subpopulation (Northern or Southern China).

Epistasis Between RET Common and Rare Variants

To decipher the genetic architecture of S-HSCR, we assessed whether the effects of rare *RET* protein-alternating variants varied with the dosage of common HSCR-associated risk alleles (T for rs2435357 and A for rs9282834). Samples were stratified into 3 groups carrying 0, 1 or at least 2 common HSCR-associated risk alleles. Samples were further subdivided into t3hree subgroups (totaling $9 = 3 \times 3$ combinations), according to the presence of mutations and their predicted pathogenicity (damaging or benign). Next, we computed the odds ratio for individuals with each combination relative to the baselines, that is, not carrying any common risk allele or rare variants. For dosage-specific odds ratio, groups of samples carrying no mutation under the same common risk allele dosage were used as at baseline.

Haplotype Configurations of RET Common and Rare Variants

To determine if the rare *RET* protein-alternating variants occurs in *cis* or *trans* with the common regulatory enhancer variant (rs2435357), we performed read-aware phasing using SHAPEIT2.[17] Briefly, phase informative reads spanning at least 2 heterozygous sites were obtained from the BAM files and were used to phase the common and rare variants in the genotype data in a VCF file.

Imputation of Expression Using PrediXcan

To impute the gene expressions of *BACE2*, *BACE1*, and *APP*, we considered 2 tissue models (each with >300 samples): (1) the neural (tibial nerve, 361 individuals) and (2) whole blood (369 individuals), built from expression quantitative loci (eQTLs) of the GTEx database (v6p release) using PrediXcan.[18] Because no linear model on tibial nerve (false discovery rate < 5%) was available in PredictDB for *BACE1* and *APP*, we examined whether whole-blood tissue is a good surrogate for unmeasured ENCCs. Tissue-specific regulatory potentials for eQTLs included in the GTEx whole-blood models were checked against ChromHMM-based chromatin segmentation prediction from ROADMAP. Transcription factor binding motifs affected by the eQTLs were predicted using the R package motifbreakR[19] based on its processed HOCOMOCO transcription factor–binding motif database. Although many of selected eQTLs overlapped with neural (brain and neuronal progenitors) or blood sample enhancers or promoters in ROADMAP and were predicted to strongly affect motifs of transcription factors important in ENS development (eg, FOXD3 and GLI1) (Supplementary Table 12 and 13), GTEx whole-blood models were then used to impute the expression of *BACE1* and *APP*. For *BACE2*, tibial nerve tissue model was used for imputation, because it provides a better fit than whole-blood tissue model ($P = 2.6 \times 10^{-14}$ for tibial nerve vs $P = 3.4 \times 10^{-5}$ for whole-blood regarding the correlation between predicted and observed expression). Imputed gene expressions were normalized using the mean expression and standard deviation of all 493 WGS control samples.

Human Induced Pluripotent Stem Cells

Two control hiPSC lines (IMR90 and UE02302-hiPSC) were used to generate ENCCs and ENS neurons. IMR90 iPSC (clone 2) was purchased from WiCell Research Institute, and UE02302 was a gift from Dr Guangjin Pan (Guangzhou

Institutes of Biomedicine and Health, China).[20] *BACE2*$^{-/-}$, *BACE2*^{G446R},and *BACE2*^{G446R}*APP*$^{-/-}$ mutant hiPSC lines were derived from this control line. All the control and mutant hiPSC lines were cultured on Matrigel (BD Biosciences, 354234)-coated plates with the defined medium mTeSR1 (StemCell Technologies, 05850), and the culture medium was changed daily.

Plasmid Constructions

The high-fidelity Cas9 plasmid pSpCas9(BB)-2A-GFP (PX458) and human codon-optimized Cas9 expression plasmid Cas9D10A-2A-GFP were purchased from Addgene (nos. 48138 and 44720, respectively). The guide RNAs (gRNAs) targeting the human *BACE2* or *APP* genomic regions with PAM targets 19 base pairs were designed using the CRISPR design website (http://crispr.mit.edu/). For generation of gRNA expression construction, the gRNA expression vector (Addgene no. 41824) was linearized with AflII, and the gene-specific gRNA targeting sequence was incorporated into gRNA plasmid using Gibson assembly (New England Biolab, catalog no. E2611L) according to manufacturer's protocol. All the gRNA oligonucleotide sequences are listed in Supplementary Table 1.

An expression plasmid containing human *BACE2* full-length complementary DNA (NM_012105.3) was purchased from Sino Biological Inc. (HG10783-NF). Human APP-GFP plasmid was purchased from Addgene (no. 69924). HSCR-associated mutations were introduced into *BACE2* expression construct using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent), with specific primers as listed in Supplementary Table 1 according to the manufacturer's protocol. DNA sequences and mutations were confirmed by Sanger sequencing.

Generation of BACE2-*Mutant or* BACE2/APP-*Knockout hiPSC Lines*

CRISPR-Cas9^{D10A} nickase-based genome edit system was used to generate the *BACE2*- and *APP*-knockout hiPSC lines, as previously described.[21] Two single guide RNAs targeting exon 1 of the *BACE2* gene or exon 3 of the *APP* gene locus were created according the gRNA cloning protocol described. Four million hiPSCs were transfected with gRNA constructs and a green fluorescent protein (GFP)–fused Cas9^{D10A} nickase expression plasmid by electroporation using Nuclofector transfection kit (Lonza, VPH-5022). After transfection, cells were seeded on Matrigel-coated plate with mTeSR1 medium for 48 hours, and then GFP-expressing cells were sorted into Matrigel-coated 96-well plates by fluorescence-activated cell sorting (FACS) to get a single cell. A single colony was formed around 7–14 days, and then the single colony was passaged twice using ReLeSR (Stemcell Technologies, 05872) according to the manufacturer's protocol. Subsequently, the genomic DNA was isolated, and the targeted region of the *BACE2* or *APP* gene was amplified by polymerase chain reaction using the primers listed in Supplementary Table 1. The polymerase chain reaction products were directly sequenced after being treated with exonuclease I and shrimp alkaline phosphatase. Colonies carrying bi-allelic nonsense mutations were expanded and subjected to functional studies.

To generate the *BACE2*^{G446R}-mutant hiPSC line, spCas9-eGFP and single guide RNA targeting the *BACE2* mutation site (1336G>C), together with a single-stranded oligo DNA nucleotide were co-transfected into the control hiPSC (UE020302).[22] Mutant clones were selected and validated using Sanger sequencing, as described.

Neural Crest Induction

Control or mutant hiPSCs were seeded on Matrigel-coated plates (10^5 cells/cm^2) in iPSC medium containing 10 ng/mL FGF2 (PeproTech, 100-18B) and 10 μmol/L Y-27632; at this stage, the cells were marked as day 0. Differentiation was initiated by replacing iPSC medium with KSR medium, containing knockout Dulbecco's modified Eagle medium (DMEM) plus 15% KSR (Life Technologies, 10828-028), NEAA (Life Technologies, 11140-050), L-glutamine (Life Technologies, 25030-081), β-mercaptoethanol (Life Technologies, 21985-023), LDN193189 (100 nmol/L, Stemgent), and SB431542 (10 μmol/L, Tocris). After differentiation, the KSR medium was then gradually changed to N2 medium at day 4 by increasing N2 from 25% to 75% from days 4 to 9, as described previously.[23] The N2 medium contains Neural basal medium (Life Technologies, 22103-049): DMEM/F12 (Life Technologies, 10565-018; 1:1), N2 supplement (Life Technologies, 17502-048), B27 supplement (Life Technologies, 17504-044), and insulin (Life Technologies, 12585-014). To induce ENCC differentiation, different small-molecule combinations were used to treat cells with LDN193189 (from day 0 to day 3), SB431542 (from day 0 to day 4), 3 μmol/L CHIR99021 (Tocris Bioscience, 4423) (from day 2 to day 10), and 1 μmol/L retinoic acid (from day 6 to day 9). The differentiated cells are sorted at day 10 after staining using p75^{NTR}, and HNK-1 antibodies as described.[23–26]

FACS and Flow Cytometry Analysis

To analyze the ENCC yield from the control and mutant hiPSC lines, the 10-day–differentiated cells were dissociated with Accutase (Innovative Cell Technologies, AT104) and then incubated with anti-human antibodies including APC-HNK-1 (BD Pharmingen, 560845) and FITC-p75^{NTR} (Miltenyi Biotec, 130-091-917) for 30–45 minutes on ice. To stain for PE-RET (Neuromics, FC15018), the cells were fixed in 4% paraformaldehyde for 10 minutes at room temperature and permeabilized using 0.1% (weight/volume) Saponin solution, then washed and blocked in phosphate-buffered saline (PBS) with 2% fetal bovine serum (FBS). The cells were then stained with antibodies for 30–45 minutes on ice. Approximately 10^6 cells were stained, and labeled cells were detected using a FACSAriaIII (BD Immunocytometry Systems). Isotype-matched antibodies were used as controls. FlowJo, version 8.2 (Tree Star), was used to analyze flow data.

For cell sorting, HNK-1/p75^{NTR} stained cells were washed and resuspended in PBS with 2% FBS. The HNK-1and p75^{NTR} double-positive cells were enriched using

FACS (BD FACSAria III Cell Sorter). The HNK-1 and $p75^{NTR}$ double-positive cells were gated and sorted using the 4-way purity mode, and the purity of sorted cells was >96% and evaluated by flow cytometry. The sorted neural crest cells were collected for immunostaining or subsequent experiments. A list of primary antibodies and working dilutions is provided in Supplementary Table 2.

In Vitro Differentiation of ENCCs to ENS Neurons

Approximately 40,000 FACS-enriched ENCCs were seeded as droplets on poly-ornithine/laminin/fibronectin–coated 24-well plates in N2 medium containing 10 ng/mL FGF2, 3 μmol/L CHIR99021, and 10 μmol/L Y-27632. After 24 hours, N2 medium was replaced by neuronal differentiation medium: N2 medium containing BDNF (10 ng/mL; PeproTech, 450-01), GDNF (10 ng/mL; PeproTech, 450-10), ascorbic acid (200 μmol/L; Sigma, A4034-100G), NT-3 (10 ng/mL; PeproTech, 450-03), NGF (10 ng/mL, PeproTech, 450-01), and cyclic adenosine monophosphate (1 μmol/L; Sigma, D0260). Cells were cultured in the neuronal differentiation medium for up to 30 days, and the culture medium was changed every 2 days. ENS neurons at differentiation day 30 were fixed for immunocytochemistry analyzes or harvested using Accutase for RNA sequencing and Western blot analyzing.

Migration Assay

FACS-enriched ENCCs were plated on human fibronectin–coated 12-well culture plates (30,000 cells/cm). After 24 hours, cells were treated with mitomycin (10 μg/mL) to stop cell proliferation. A wound was created in the center of each well by scratching with a pipette tip. Cells were allowed to migrate for 18 hours. The images of the initial wound and final wound were captured immediately and at 18 hours after scratching. The migration distance was obtained by comparing the width of the initial wound created and wound closure during 18 hours.

Immunofluorescence Analysis

For immunofluorescence, the cells were fixed with 4% paraformaldehyde in PBS at room temperature for 30 minutes, followed by blocking with 1% bovine serum albumin (Thermo Fisher Scientific, 23209) with or without 0.1% Triton X-100 (Sigma, T8787) in PBS buffer. Cells were incubated in primary antibody solutions overnight at 4°C and host-appropriate fluorescein isothiocyanate or Texas Red secondary antibody (Molecular Probes, Invitrogen) (Supplementary Table 2) for 1 hour at room temperature. Cells were then counterstained with mounting medium with DAPI (DAKO) to detect nuclei. Cells were photographed using a Carl Zeiss confocal microscope (LSM 800). Quantitative image analysis of differentiated neuronal cultures was done with ImageJ plugins (National Institutes of Health). In brief, intensity thresholds were set, blinded to sample identity, to selectively identify as positive cells, which displayed unambiguous signal intensity above local background. These parameters were used on all samples and were only minimally adjusted for different staining batches as necessary. A minimum of 4,000 cells were analyzed per sample. Percentages of neuronal cells were measured over the total number of cells (DAPI), and the values reported in bar charts represent the mean ± standard error of the mean.

Cell Culture, Transfection, and Immunoblotting

The 293FT cell line was used to analyze the biological impacts of *BACE2* variants in APP processing and BACE2 membrane localization. 293FT cells were cultured in DMEM medium, supplemented with 10% FBS and 1% penicillin/streptomycin, at 37°C in 5% CO_2, and the culture medium was changed every other day. For transfection, approximately 1 million cells were seeded onto 6-well plates (Nunc) 24 hours before transfection. GFP-tagged APP and FLAG-tagged wild-type or mutant BACE2 were overexpressed in the 293FT cell line by transfection using FuGene HD Transfection Reagent (Promega) according to the transfection protocol. Two days after transfection, the cells were collected and lysed using protein lysis buffer containing 50 mmol/L Tris-HCl pH 7.5, 100 mmol/L NaCl, 1% Triton X-100, 0.1 mmol/L EDTA, 0.5 mmol/L $MgCl_2$, 10% glycerol, protease inhibitor cocktail (Roche), and phosphatase inhibitor cocktail (Roche). After incubation on ice for 15 minutes, the total proteins were collected by centrifuge for 10 minutes at 12,000 rpm at 4°C. For the membrane and cytosolic protein fractionation, 293FT cells overexpressing FLAG-tagged wild-type BACE2, S442F, or G446R BACE2 were collected 48 hours after transfection. Membrane and cytosolic proteins were extracted using the Mem-PER plus membrane protein extraction kit (no. 89842, Thermo Fisher Scientific) according to the manufacturer's protocol. Next, 20 μg of total protein from cell lysates was separated on 12% sodium dodecyl sulfate–polyacrylamide gels and blotted with the corresponding primary antibodies. A list of primary antibodies and working dilutions is provided in Supplementary Table 12. The same membranes were stripped and hybridized with anti-β-actin monoclonal antibody (Millipore, MAB 1501) as a protein-loading control. All blots were incubated with secondary horseradish peroxidase–conjugated anti-mouse or anti-rabbit or anti-goat antibody (1:2500, DAKO).

Statistical Analysis

Statistical significance was determined by the 2-sided unpaired Student *t* test or 1-way analysis of variance using GraphPad Prism 7 (GraphPad Software). The *P* value is indicated by asterisks in the figures. Differences among groups of $P < .05$ were considered statistically significant. All experiments were replicated at least 3 times, and data are shown as means with SEM or standard derivation SD.

References

1. So MT, Leon TY, Cheng G, et al. RET mutational spectrum in Hirschsprung disease: evaluation of 601 Chinese patients. PLoS One 2011;6:e28986.
2. 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;106:2694–2699.
3. Tang CS, Gui H, Kapoor A, et al. Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease. Hum Mol Genet 2016;25:5265–5275.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754–1760.
5. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011; 43:491–498.
6. Li M, Li J, Li MJ, et al. Robust and rapid algorithms facilitate large-scale whole genome sequencing down-stream analysis in an integrative framework. Nucleic Acids Res 2017;45(9):e75.
7. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet 2014;46:944–950.
8. Abyzov A, Urban AE, Snyder M, et al. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 2011;21:974–984.
9. Liang Y, Qiu K, Liao B, et al. Seeksv: an accurate tool for somatic structural variation and virus integration detection. Bioinformatics 2017;33:184–191.
10. Rausch T, Zichner T, Schlattl A, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28:i333–i339.
11. Layer RM, Chiang C, Quinlan AR, et al. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 2014;15:R84.
12. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 2014;47:11 12 1–34.
13. MacDonald JR, Ziman R, Yuen RK, et al. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 2014; 42:D986–D992.
14. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 2009;84:524–533.
15. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008;83:311–321.
16. Zhan X, Hu Y, Li B, et al. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 2016;32:1423–1426.
17. Delaneau O, Howie B, Cox AJ, et al. Haplotype estimation using sequencing reads. Am J Hum Genet 2013; 93:687–696.
18. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091–1098.
19. Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics 2015; 31:3847–3849.
20. Xue Y, Cai X, Wang L, et al. Generating a non-integrating human induced pluripotent stem cell bank from urine-derived cells. PLoS One 2013;8:e70573.
21. Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. Science 2013;339:823–826.
22. Paquet D, Kwart D, Chen A, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature 2016;533:125–129.
23. Lee G, Chambers SM, Tomishima MJ, et al. Derivation of neural crest cells from human pluripotent stem cells. Nat Protoc 2010;5:688–701.
24. Fattahi F, Steinbeck JA, Kriks S, et al. Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. Nature 2016;531:105–109.
25. Lee G, Kim H, Elkabetz Y, et al. Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. Nat Biotechnol 2007; 25:1468–1475.
26. Zeltner N, Lafaille FG, Fattahi F, et al. Feeder-free derivation of neural crest progenitor cells from human pluripotent stem cells. J Vis Exp 2014.