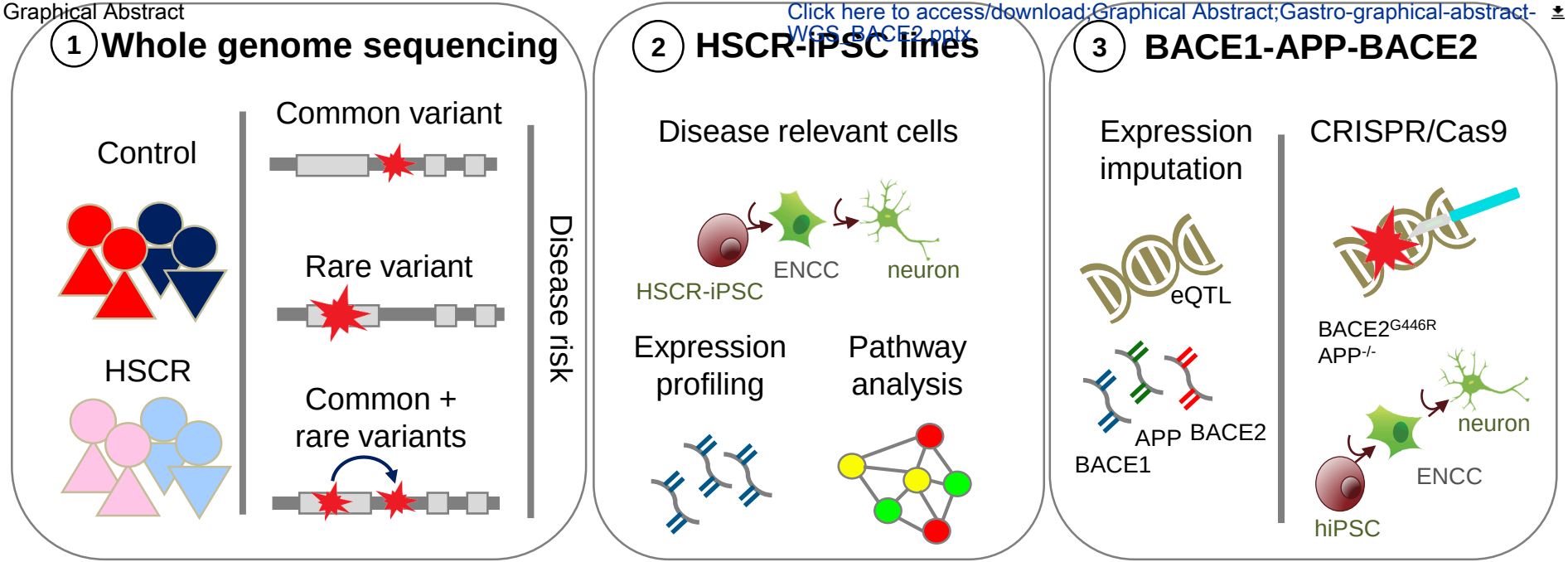
Graphical Abstract



Gastroenterology

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]Ø, Peng LI[1,a]Ø, Frank Pui-Ling LAI[1,6,a], Alexander Xi FU[7,c], Sin-Ting LAU[1,6,a], Man Ting SO[1,d], Kathy Nga-Chu LUI[1,c], Zhixin LI[1,6,c], Xuehan ZHUANG[1,c], Michelle YU[1,c], Xuelai LIU[8,e], Ngoc D NGO[9,e], Xiaoping MIAO[10,e], Xi ZHANG[11,c], Bin YI[11,h], Shaotao TANG[12,h], Xiaobing SUN[12,h], Furen ZHANG[13,h], Hong LIU[13,h], Qiji LIU[14,h], Ruizhong ZHANG[15,a], Hualong WANG[16,h], Liuming HUANG[17,h], Xiao DONG[18,h], Jinfa TOU[19,h], Kathy SE CHEAH[4,e],Wanling YANG[5,g], Zhenwei YUAN[20,f], Kevin Yuk-lap YIP[7,g], Pak-Chung SHAM[2,3,f], Paul Kwang-Hang TAM[1,f], Maria-Mercè GARCIA-BARCELO[1,g,*], Elly Sau-Wai NGAN[1, g,*]

[1]Department of Surgery, [2]Department of Psychiatry, [3]Centre for Genomic Sciences, [4]School of Biological Sciences, [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 250012, P.R. 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
[20]Department of Paediatric Surgery, Shengjing Hospital, China Medical University, Shenyang, China

ØAuthors have equal contribution

[a] Postdoctoral Fellow
[b] Research Assistant Professor
[c] Postgraduate Research Student
[d] Technician
[e] Non-clinical Professor

[f] Clinical Chair Professor
[g] Non-clinical Associate Professor
[h] Clinical Professor

Corresponding authors:
Dr. Elly SW Ngan
Associate Professor
Department of Surgery
University of Hong Kong, Pokfulam
Faculty of Medicine Building
21 Sassoon Road, Hong Kong, SAR, China.
Tel: (852) 2831-5410
Fax: (852) 3917-9621
Email: engan@hku.hk

Dr. Maria-Mercedes Garcia-Barcelo
Associate Professor
Department of Surgery
University of Hong Kong, Pokfulam
Faculty of Medicine Building
21 Sassoon Road, Hong Kong, SAR, China.
Tel: (852) 2831-5073
Fax: (852) 3917-9621
Email: mmgarcia@hku.hk

Short Title: Oligogenic aetiology of HSCR

The authors declare no conflicts of interest.

Author contributions: L.P., F.P.L.L. and S.T.L. performed functional studies. C.S.M.T., A.X.F. X.Z., K.Y.Y. and Z.L. performed genetics and bioinformatics analyses. MTS did the genetic sample preparation and Sanger sequencing. P.K.H.T. and MY provided clinical information together with XL, NGN, XM, XZ, BY, ST, XS, FZ, HL, QL, RZ, HW, LH, XD, JT, KC, WY and ZY who contributed with the clinical recruitment of patients and controls. E.S.W.N., P.C.S. and M.G. supervised the project and prepared the manuscript.

Grant Support:
This work was described in this paper was substantially supported by Theme-based Research scheme [grant number T12C-714/14-R.], UICP [UIM/299], General Research Fund [HKU17109215 to E.S.W.N., HKU17119514 to P.K.H. T.], Health Medical Research Fund [HMRF 02131866 and 01121516 to M.M.G.B, 0451966 to C.S.T., 03143236 to

E.S.W.N.] and Seed Fund for Basic Research of the University of Hong Kong [grant number 201606159005 to C.S.T].

Abbreviations: WGS: whole genome sequencing; HSCR: Hirschsprung disease; hiPSC: human induced pluripotent stem cell; APP: Amyloid Beta Precursor Protein; BACE2: Beta-Site APP-Cleaving Enzyme 2

ABSTRACT

Abstract:

Background & Aims: Hirschsprung disease, or congenital aganglionosis, is believed to be oligogenic—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 controls, and 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=7.4x10^{-7}). The patients had a significant excess of rare protein-altering variants in genes previously associated with Hirschsprung disease, and in the beta-secretase 2 gene (*BACE2*; P=2.9x10^{-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 APP 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.

Key words: genetics, Amyloid beta; enteric nervous system; CRISPR/Cas9

INTRODUCTION

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 three main types: 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, though 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, i.e., a larger effect in male, S-HSCR. On the contrary, the frequency of the *RET* coding mutations correlates positively with disease severity.

Our previous meta-analysis of genome-wide association studies (GWAS) 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 (i.e. L-HSCR and TCA). To identify novel HSCR genes and explore the oligogenic nature of the disease, we hereby 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 (hPSC)-based model, we further defined new biological pathways underlying the HSCR pathogenesis.

MATERIALS AND METHODS

Patients

The discovery cohort comprised 464 sporadic S-HSCR cases and 498 controls. Patients had been recruited at hospitals in China (n=341) and Hanoi, Vietnam (n=102). After quality control (detailed in Supplementary materials and methods), a total of 443 S-HSCR and 493 controls 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 institutional review board of the University of Hong Kong and the Hospital Authority ((HKU/HA HKW IRB) UW 13-225).

Whole genome sequencing and variant calling

All samples were whole genome sequenced (WGS) using Illumina HiSeq X Ten to a mean coverage of 30X. Sequence reads were then processed according to Genome Analysis Toolkit (GATK; version 3.4) Best Practices recommendations[5] (see Supplementary Information).

Variant annotation

Annotation was done using KGGseq for protein function against the RefGene, pathogenicity and population frequencies. We defined protein-truncating variants as those that lead to (i) gain of the stop codon, (ii) frameshift and (iii) 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. Lastly, 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 (URV) are defined as a singleton variant, that is, one that appeared only once in our whole dataset, not present in dbSNP138 or public databases (see supplementary information).

Known genes of ENS development and their interactome

Genes where mutations are reported to cause colonic aganglionosis in mutant mice according to the Mouse Genomics Informatics (MGI) were considered as known ENS genes. ENS interactome was defined by genes encoding proteins that show protein-protein interaction (PPI) with known ENS genes (see supplementary information).

Copy number variants (CNVs)

Overlapping CNVs for ENS genes were detected using 4 different yet complementary software to maximize the accuracy.

Gene-based and geneset-based burden test for rare variants

For the set of known ENS genes, we first assessed the enrichment of (i) damaging and (ii) all rare protein-altering variants collectively in cases compared to controls.

RET common and rare variants epistasis

To assess if 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) we stratified samples into 3 groups: carrying zero, one or at least two common HSCR-associated risk alleles. Samples were further subdivided into three subgroups (totaling 9=3x3

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 occurs in *cis* or *trans* with the enhancer variant (rs2435357), we performed read aware phasing as described in supplementary information.

Imputation of expression using PrediXcan

To impute the gene expression of *BACE2, BACE1* and *APP*, we considered two tissue models (each with >300 samples), (i) the neural (tibial nerve, 361 individuals) and (ii) whole blood (369 individuals) (see supplementary information)

Human induced pluripotent stem cells (hiPSC)

Two control (IMR90 and UE02302) and four HSCR (3 S-HSCR and 1 TCA) -iPSC lines were used to generate ENCCs and ENS neurons. IMR90 iPSC (clone#2) was purchased from Wicell Research Institute, UE02302 is 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 plate 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, CHIR99021 and then gradually switching to N2 medium and caudaulized with 1 μM retinoic acid as described previously[7]. Enteric NCCs (ENCCs) were then enriched using $p75^{NTR}$ and HNK1 antibodies.

FACS 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 ($3x10^4$) from the 10-day induction protocol were seeded as droplets on poly-ornithine/laminin/fibronectin (PO/LM/FN)-coated dishes in N2 medium containing 10 ng/ml FGF2 and 3μM CHIR99021. After 24 hours, N2 medium was replaced by the neuronal differentiation medium: N2 medium containing GDNF, ascorbic acid, brain-derived neurotrophic factor (BDNF), nerve growth factor (NGF), neurotrophin-3 (NT-3) and cyclic adenosine monophosphate (cAMP). 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 analyzing.

Immunofluorescence analysis

For immunofluorescence, the cells were fixed with 4% paraformaldehyde in PBS at room temperature for 30 min, blocked and incubated in primary antibody solutions and host-appropriate secondary antibody. Cells were then counterstained with mounting medium with DAPI (DAKO) and photographed using Carl Zeiss confocal microscope (LSM 800). Quantitative image analysis of differentiated neuronal cultures was done with ImageJ plugins. 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 ± SEM.

Cell culture, transfection and Immunoblotting

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% CO2, the culture medium was changed every other day. For transfection, around 1 million of cells were seeded onto 6-well plates (Nunc) 24 hours prior to transfection. GFP-tagged APP together with FLAG-tagged wild-type or mutant BACE2 were overexpressed in 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 (#89842, Thermo Fisher Scientific) according to the manufactor's protocol. 20µg of total protein from cell lysates was separated on 12% SDS-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 two-sided unpaired Student's t-test or 1-way ANOVA using GraphPad Prism 7 (GraphPad Software). The *P*-value is indicated by asterisks in the figures (*, $P < 0.05$; **, $P < 0.01$). Differences among group of $P< 0.05$ were considered statistically significant. All experiments were replicated at least three times and data are shown as means with standard error of mean (SEM) or standard derivation (SD).

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 controls in whom a total of 36.7 million autosomal variants (33.4 million single nucleotide variants (SNVs) and 3.3 million insertion-deletion (indels)) were called by the Genome Analysis Toolkit (GATK). On average, we identified 3.8 million SNVs and 0.5 million indels per individual. Most of these variants (77.5%) were novel or rare (MAF<1%).

We firstly performed genome-wide association analysis on common and low frequency biallelic variants (MAF>1%; n=7,224,040). The analysis revealed the strongest association of 328 variants with HSCR ($P<5x10^{-8}$), all of which mapped to the known disease susceptibility loci of *RET* and *NRG1* (Figure 1A, upper panel). Four new loci were

identified in this study with moderate association ($P<1x10^{-6}$) with the disease, which includes two intergenic (one between *LINC01518* and *LOC283028* in 10q11.21 and another between *SLC4A7* and *EOMES* in 3p24.1) and two intronic variants on phospholipase D1 (*PLD1*) and calsequestrin 2 (*CASQ2*) (Figure 1A, lower panel and Supplementary Table 3).

Increased burden of rare variants among S-HSCR

Considering the contribution of protein-truncating (stopgain, splicing or frameshift) ultra-rare variants (URVs) in L-HSCR[8], we assessed if 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 (genes having high probability of being intolerant to loss of function (Exome Aggregation Consortium (ExAC) pLI > 0.9); odds ratio (OR)=1.28; 95% confidence interval (CI):[1.06,1.53]; $P=8.6x10^{-3}$) but not in unconstrained genes nor for synonymous URVs (Figure 1B). Such over-representation was primarily detected for URVs that can elicit nonsense-mediated decay (NMD; $P=7.6x10^{-3}$ versus $P=0.57$ for those may escape from NMD). 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]; $P=0.021$). Consistently, the number of patients harboring at least one protein-truncating URVs was 9% higher than that of controls (42.7% S-HSCR versus 33.7% controls).

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 mutant mice as per the Mouse Genomics Informatics (MGI) 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=3.9x10^{-4} for damaging and P=7.8x10^{-5} for all protein-altering rare-variants; Table 1). In particular, all rare protein-altering variants in *RET* (CMC gene-based P=1.2x10^{-5}; Supplementary Table 4) and damaging rare variants in *EDNRB* (P=1.9x10^{-4}), *ERBB2* (P=7.7x10^{-3}) and *GFRA1* (P=0.019), were overrepresented in S-HSCR cases (false discovery rate (FDR)<0.1), thus vindicating the role of the three already known pathways involved in ENS development. Noteworthy, three large case-unique deletions (from 245kb to 16.7Mb) 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 three aforementioned significant ENS genes also increased risk, either synergistically or independently, to HSCR. We identified 13 S-HSCR cases with double-hits of rare variants among the known ENS genes (all rare protein-altering variants for *RET* while rare damaging variants for other ENS genes) and none in controls (Supplementary Table 6). Half of them are female (n=7) and nearly all have rare variants affecting multiple known ENS pathways. While 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 three major HSCR genes, i.e. *RET*, *EDNRB* and *ERBB2*. Among the 87 interactome genes tested, two genes encoding the ERBB2-interacting partners, *ITGB4* (P=1.04x10^{-3}) and *PTK2* (also interacts with RET and ITGB4; P=1.39x10^{-3}), were also significantly enriched with damaging rare variants (FDR<0.1; Figure 1 C and D; Supplementary Table 7).

The RET-sensitized genetic background in HSCR

Unlike *EDNRB* and *ERBB2*, many of 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 two *RET* risk alleles, further increasing their risk to S-HSCR at least two-fold on top of the effect of the common high risk allele (OR_{dosage}=2.34; 95%CI:[1.01-5.44;] P_{dosage}=0.049). On the other hand, near 58% of patients carrying *RET* damaging variants have a single common risk allele occurring mostly *in trans* compared to <17% for other patients carrying no or benign rare variants (Figure 1E). For these patients with a double-hit of common and rare variants, the damaging *RET* variants conferred a near 5-fold increase in the risk of S-HSCR (OR_{dosage}=4.82; 95%CI: [1.82,12.77]; P_{dosage}=$1.6x10^{-3}$). 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 human induced pluripotent stem cell (hiPSC)-based disease model[7] to elucidate the underlying biological pathways associated with the *RET* variants. A total of six hiPSC lines were used for comparison, including two control hiPSC lines derived from two healthy individuals (one female and one male) carrying non-risk *RET* alleles (CC) in rs2435357; two S-HSCR patients carrying two common high risk alleles (TT) in rs2435357; one S-HSCR patient harboring a rare damaging variant in *RET*, encoding T317P and a risk *RET* allele (CT in rs2435357); and one TCA-HSCR patient with an in-frame deletion rare variant in *RET* (G731del), representing the three major sub-classes of *RET* variants identified in the HSCR patients (Figure 2A). All the hiPSC lines were derived from the fibroblast 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], enteric neural crest cells (ENCCs) carrying the exact genetic makeup of the patients were obtained (Figure 2B). 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 two neural crest markers ($p75^{NTR}$ and HNK-1) (Figure 2C), but all the HSCR-ENCCs were less competent to make neurons and migrate than that of the control (Figure 2D). In particular, the differentiation defect of ENCCs derived from the TCA-HSCR (HSCR#6) iPSC were found relatively more severe than that of S-HSCR ENCCs (HSCR#3, HSCR#5 & HSCR#20) as monitored based on the formation of neuron-like processes and the expression of 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. In order to identify the most relevant pathways attributed by the genetic lesions with the relative 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 & UE02303); S-HSCR$^{common+rare}$ with a common high-risk allele and a damaging rare variant in *RET* (HSCR#20) and S-HSCR^{common} with two common *RET* risk-alleles (HSCR#3 & HSCR#5). A TCA-HSCR with an in-frame deletion rare variant in *RET* (HSCR#6) was also included for comparison. Cells derived from two 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 (DEG) in S-HSCR$^{common+rare}$ and S-HSCR^{common} groups, respectively, in comparison with the control group (log2 fold change >= 1.5; adjusted p <0.05), where 167 genes were dysregulated in both groups (Figure 2E and Supplementary Table 8). Interestingly, when we included the TCA-HSCR line for comparison, we found that DEGs exhibit extensive interactions within each group and among the three 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 interaction and cAMP signaling pathways) and the metabolic pathway GO terms were enriched in all groups. Additional DEGs belonging to phospholipase D signaling pathway GO terms (mediating cell growth/survival and differentiation) were identified in S-HSCR$^{common+rare}$ group, while GO terms related to cell migration (e.g. Rap1 signaling pathway and actin assembly molecules) were enriched in S-HSCR^{common} group (Figure 2F). These data suggest that the RET-GDNF signaling is the main signaling pathway implicated in HSCR pathogenesis, and dysregulation of distinctive biological pathways in addition to the common pathways are

likely involved in these three subgroups of HSCR patients, and that 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=3.8x10^{-4}) was considered to be intolerant to missense changes (ExAC missense *z*-score>1.96) among the top HSCR-associated genes (Figure 3A). To validate this finding, we Sanger sequenced an independent set of 534 ethnically-matched controls. The combined analysis of WGS and Sanger sequencing confidently established the association between *BACE2* and HSCR (P=2.9x10^{-6}) which surpassed multiple testing association of 11,898 genes (Supplementary Table 9).

BACE2 is a homolog of *BACE1*, and encodes a transmembrane aspartyl protease, β-secretase 2. A total of nine rare protein alternating variants, all validated by Sanger sequencing, were detected in this gene, affecting 18 HSCR patients and 3 controls. Eight variants were found only in cases. Six out of these eight 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 cases and controls (Figure 3B).

BACE2 has a board substrate specificity similar to BACE1[10-12]. Amyloid beta precursor protein (APP) is highly expressed in nerve tissue and it contains recognition sites for BACE1 and BACE2 (Figure 3C)[13, 14]. BACE1 and γ-secretase cleave APP sequentially and produce amyloid beta (Aβ) and C60, while 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 level of these genes in the individuals who carry *BACE2* rare variants and non-carriers were imputed based on the GTEx neural tissues (tibial nerve) or whole blood models using PrediXcan[16]. The relative expression level of these genes in the individuals carrying *BACE2* variant(s) was compared to their mean expression level in the control group (493 individuals). All these three genes were predicted to be expressed in relatively lower level in the three non-HSCR carriers. On the other hand, among the 18 S-HSCR patients, the majority of them (1) exhibited higher imputed expression level in at least one of these genes; (2) carried additional mutation(s) in known HSCR gene(s) and/or (3) harbored one of the four damaging variants in *BACE2* (Figure 3D). 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 GFP-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 3E, the four 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 two rare variants are residing adjacent to the transmembrane domain of BACE2 and that may interfere 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 utilized hiPSC to generate ENCCs (Figure 4A) 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 two isogenic hiPSC lines: one carried a *BACE2* null mutation (*BACE2*$^{-/-}$) (Figure 4B) and the other harbored a *BACE2* damaging rare variant (*BACE2*^{G446R}) (Figure 4C) using the CRISPR/Cas9-mediated genome editing technology (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 4D) and *BACE2*^{G446R} (Figure 4E) hiPSC lines. *BACE2*$^{-/-}$ and *BACE2*^{G446R} hiPSC lines could efficiently generate ENCCs with comparable yield as seen in the control hiPSC line, as measured by flow cytometry using antibodies against HNK-1 and p75^{NTR} (Supplementary Figure 4A). The majority of ENCCs derived from these hiPSC lines were co-expressing the two key ENCC markers, RET and SOX10 (Supplementary Figure 4), suggesting 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). It is concordant with a previous study showing 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.

eQTL prediction suggests that the relative expression level 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 three key developmental windows of the ENS. Western blot analysis showed that BACE1 is expressed in both hiPSC-derived ENCCs and ENS-neurons, but not

in hiPSCs. Low level of BACE2, on the other hand, was detected in hiPSCs, and its expression level were elevated overtime when hiPSC were differentiating into ENCCs and then to ENS-neurons, with the highest expression in the ENS-neurons. APP expression level was comparable in the hiPSCs and ENCCs, but it was remarkably reduced in the ENS neurons, inversely correlated to the BACE2 level (Supplementary Figure 6). Co-expression of these three proteins in enteric ganglion of human colon was consistently observed based on the immunohistochemistry data available in *The Human Protein Atlas* (Supplementary Figure 7), supporting the causal relationship of these molecules in the ENS development. Therefore, we further examined how the HSCR associated rare variants in *BACE2* may affect the endogenous APP processing in the ENS neurons. We directed the control, *BACE2*$^{-/-}$ or *BACE2*^{G446R} ENCCs to the neuronal lineage by culturing them in the neuronal differentiation medium for 30 days (Figure 4F). Immunocytochemistry analyses showed that both the control and *BACE2* mutant ENCCs can give rise to neurons efficiently, expressing the pan-neuronal marker, beta-III-tubulin (TUJ1). By day 30 of the neuronal differentiation, *BACE2*$^{-/-}$ and *BACE2*^{G446R} ENS neurons underwent obvious morphological changes with the enlarged cell bodies and granules. Subsequent immunocytochemistry using antibody against cleaved-caspase3 further revealed that many of *BACE2*$^{-/-}$ and *BACE2*^{G446R} ENS neurons are undergoing apoptosis and expressing high level of cleaved-caspase 3 (Figure 4G). Importantly, accumulation of Aβ oligomers was observed in the tightly arranged *BACE2*$^{-/-}$ and *BACE2*^{G446R} ENS neurons, such intensively stained granules were not found in the control cells carrying the wild-type *BACE2* (Figure 4H). *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 using Western blotting (Figure 4I). Genome-wide gene expression

profiles of the control and mutant ENS neurons were obtained by performing bulk RNA sequencing. We identified 2604 differentially expressed genes (DEG) in *BACE2*^{G446R} ENS neurons representing 1044 upregulated and 1,560 downregulated genes, while 1353 and 1875 genes were up- and down-regulated in *BACE2*$^{-/-}$ ENS neurons in comparison with the control (hiPSC-ENS neuron) (log2 fold change >= 1.5; adjusted p <0.05) (Supplementary Table 10 and Supplementary Figure 8). Gene ontology (GO) enrichment analysis showed enrichment of apoptosis GO terms in both *BACE2*$^{-/-}$ and *BACE2*^{G446R} ENS neurons (Figure 4J), further supporting that BACE2 deficiency may induce apoptosis of the ENS neurons.

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

DISCUSSION

Our WGS analysis of 443 S-HSCR patients and 493 ethnically-matched controls provides a glimpse into the genetic architecture of HSCR, in particular that of 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 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 utilizing a population-based association approach, not only did we vindicate the roles of *RET* and *EDNRB* as the two major HSCR genes, but we also firmly established *ERBB-NRG* as a core HSCR pathway with significant contribution of both common (*NRG1*) and rare (*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 (ECM) through interaction with ECM proteins, e.g. laminin and collagen. It has been shown that focal adhesion kinase (FAK) encoded by *PTK2* can be activated by β4-integrin. In addition, FAK 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 co-ordination 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 be proved 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 of 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 later, 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 suggested 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.

Importantly, our population-based rare variant association study identified *BACE2* as a novel HSCR gene. Unlike BACE1, the physiological 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 the Aβ production and prevents the accumulation of amyloid oligomers in order to protect neurons from undergoing apoptosis. More importantly, our data also indicated that the BACE1-APP-BACE2 pathway would be another key pathway underlying the HSCR pathogenesis. Similar to other HSCR pathways, double hits interrupting more than one gene within this pathway are likely required to cause the disease. It may not be limited to these three core genes (*BACE1*, *BACE2* and *APP*), but also other genes implicated in the Aβ production or processing. For instance, PLD1 is involved in the 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 demonstrated that the effect of rare variants among various ENS genes, BACE1-APP-BACE2 pathway genes, particularly, some of these rare variants are modified

by the genetic background sensitized by the common regulatory variants such as *RET*. No matter whether a hit is from protein-altering rare or common variant regulating gene expression, two or multiple hits are required to influence disease risk. Such epistasis highlights the importance of the underlying genetic background while assessing disease risk for complex disorders. Instead of testing the burden of rare variants and association of common variants separately, integration of statistical methods jointly considering both while 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, or *EDNRB* damaging mutations, or with 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. As 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.

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;46:568-80.
3. Emison ES, Garcia-Barcelo M, Grice EA, et al. Differential contributions of rare and common, coding and noncoding Ret mutations to multifactorial Hirschsprung disease liability. Am J Hum Genet 2010;87:60-74.
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;43:491-8.
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 e8.
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;25:543-58.
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;97:9712-7.
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;20:1369-76.
16. **Gamazon ER, Wheeler HE, Shah KP**, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091-8.
17. Fattahi F, Steinbeck JA, Kriks S, et al. Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. Nature 2016;531:105-9.
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;286:17292-302.
19. **Dominguez D, Tournoy J**, Hartmann D, et al. Phenotypic and biochemical analyses of BACE1- and BACE2-deficient mice. J Biol Chem 2005;280:30797-806.

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;103:1941-6.
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;103:1936-40.

FIGURE LEGENDS

Figure 1. Common and rare variants association analysis on WGS data. (A) Manhanttan plot of genome-wide association analysis results for variants with MAF>1%. Upper panel: SNPs within loci (±1Mb of top SNPs) passing genome-wide significance (horizontal dash line; P<5x10-8) are labelled in yellow and the top SNP of each region are labelled separately in orange diamond. Lower panel: Magnified plot for moderately associated variants. SNPs showing marginal association (5x10-8≤P<1x10-6) are denoted by dark grey circles. **(B)** Association between protein-truncating and synonymous ultra-rare variants (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 ExAC database. Counts of alternative allele in cases (top panel) and controls (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.

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-seq. **(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 h). The bar charts show mean± SEM from 3-7 independent experiments. "**" and "*" indicate significant different from the healthy control (P<0.001 and P<0.05, respectively). **(E)** Transcriptional profiles of $p75^{NTR+}HNK\text{-}1^{+}$ ENCCs derived from the control and HSCR-hiPSC lines were obtained by RNA sequencing. The Venn diagram shows differentially expressed genes (DEGs) commonly and uniquely found in S-$HSCR^{common+rare}$, S-$HSCR^{common}$ and TCA-HSCR groups in comparison of the control ENCCs. **(F)** GeneMANIA shows genes (network nodes) in the association networks created using the DEGs. Genes belonging to different GO terms and the involved biological processes are indicated.

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 cases (top) and controls (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 eQTL- predicted expression levels of *BACE1*, *BACE2* and *APP* in individuals carrying *BACE2* rare variant (3 control 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 293FT human cell line. The full-length and cleaved APP (GFP-tagged) were detected. C81 fragment represents BACE2-cleaved APP as marked by "*". The enzymatic cleavage ability of BACE2 were 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 three independent experiments).

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 Ctrl, *BACE2*$^{-/-}$ or *BACE2*^{G446R} hiPSC at day 30 with **(G)** TUJ1 or cleaved Caspase3 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.

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 day30 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.

ACKNOWLEDGEMENT

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

Table 1

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	**Cases**	**Controls**	**OR (95%CI)[a]**	***P*[b]**	**Cases**	**Controls**	**OR (95%CI)**	***P***
RET	19	11	1.96 (0.92,4.17)	0.075	63	29	2.65 (1.67,4.20)	**$1.2x10^{-5}$**
EDNRB	13	0	NA	**$1.9x10^{-4}$**	16	4	4.58 (1.52,13.81)	**$5.5x10^{-3}$**
ERBB2	12	3	4.55 (1.27,16.22)	**$7.7x10^{-3}$**	21	19	1.24 (0.66,2.34)	0.446
GFRA1	6	1	6.76 (0.81,65.33)	**0.019**	7	4	1.96 (0.57,6.75)	0.158
AEBP2	3	7	0.47 (0.12,1.84)	0.254	3	8	0.41 (0.11,1.57)	0.134
IHH	3	6	0.55 (0.14,2.23)	0.415	3	6	0.55 (0.14,2.23)	0.415
ITGB1	4	6	0.74 (0.21,2.64)	0.767	4	6	0.74 (0.21,2.64)	0.767
GDNF	6	6	1.11 (0.36,3.48)	0.769	7	6	1.30 (0.43,3.91)	0.769
SOX10	3	2	-	-	3	4	0.83 (0.19,3.74)	0.930
ZEB2	5	4	1.40 (0.37,5.23)	0.555	6	7	0.95 (0.32,2.86)	0.950

[a] OR refers to odds ratio compared by 2x2 table without adjustment by principal components (PCs)
[b] Gene-based *P*-valued for CMC burden test (score test) with 3 PCs as covariates

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*-value
	Damaging[b]	Benign			
≥2	+	-	5 (1.1)	17.57 (1.91, 161.41)	**0.01**
	-	+	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)	0.08
	-	-	64 (14.4)	1.21 (0.72, 2.03)	0.48
0	+	-	3 (0.7)	4.50 (0.84, 24.19)	0.08
	-	+	1 (0.2)	0.99 (0.10, 9.39)	0.99
	-	-	25 (5.6)	1	-

[a] ≥2: More than 2 risk alleles

[b] Rare variants are classified into damaging and benign according to *in silico* prediction

Figure 1

Click here to access/download;Figure;Figure 1.tif

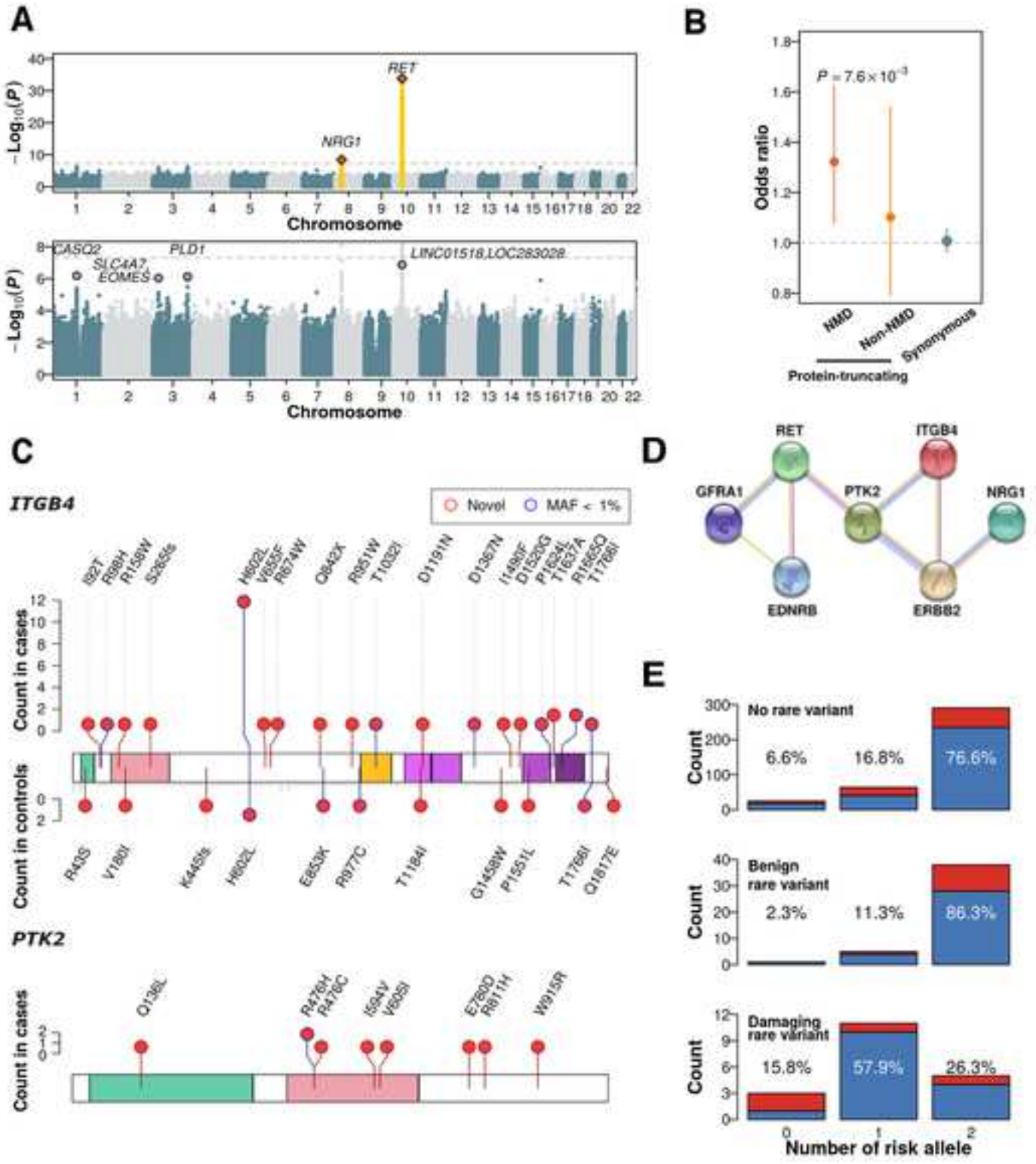
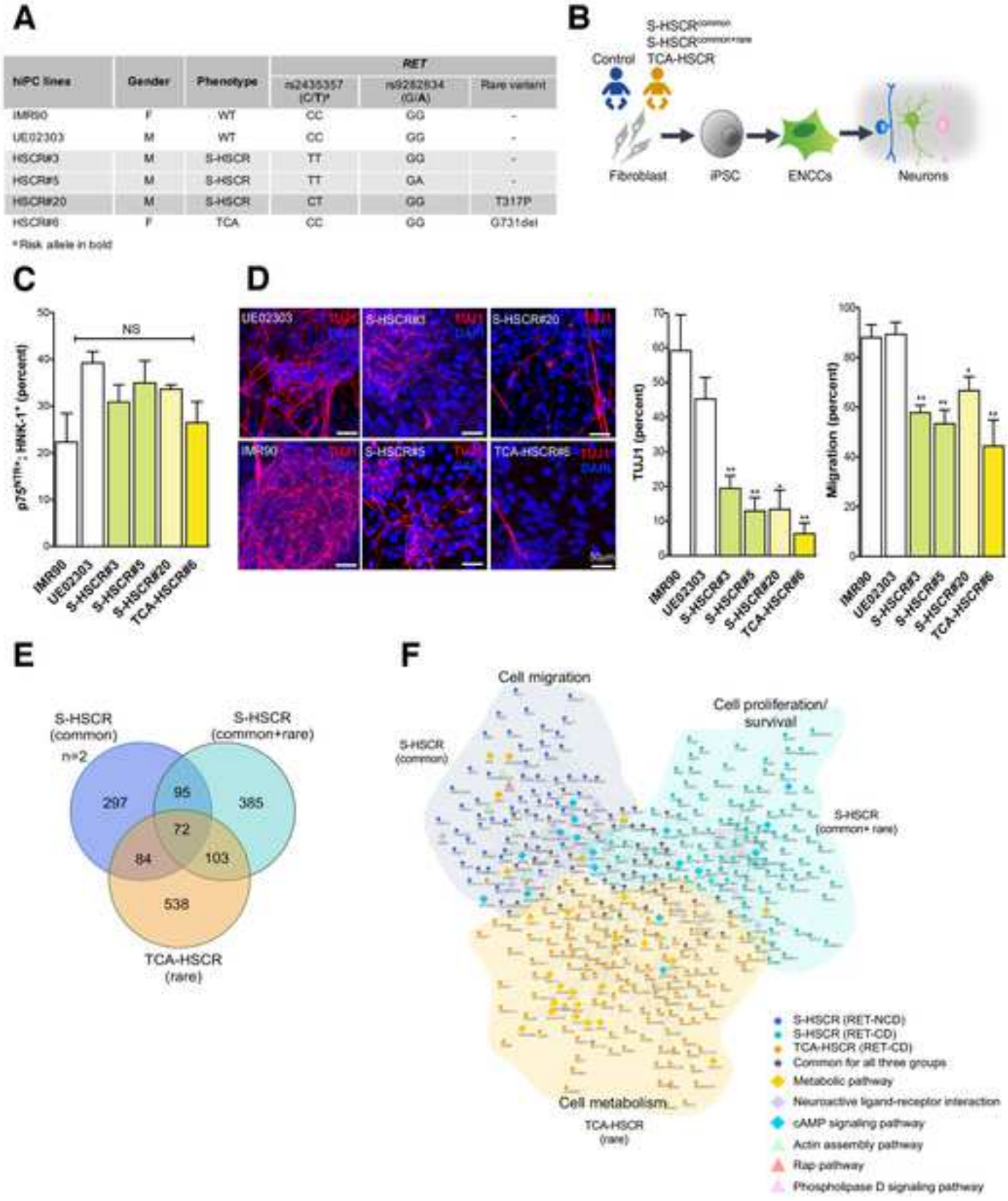


Figure 1. Common and rare variants association analysis on WGS data. (A) Manhanttan plot of genome-wide association analysis results for variants with MAF>1%. Upper panel: SNPs within loci (±1Mb of top SNPs) passing genome-wide significance (horizontal dash line; $P<5x10-8$) are labelled in yellow and the top SNP of each region are labelled separately in orange diamond. Lower panel: Magnified plot for moderately associated variants. SNPs showing marginal association ($5x10-8<P<1x10-6$) are denoted by dark grey circles. **(B)** Association between protein-truncating and synonymous ultra-rare variants (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 ExAC database. Counts of alternative allele in cases (top panel) and controls (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. Illoc bar: male; Red bar: female.

Figure 2

Click here to access/download;Figure;Figure 2.tif



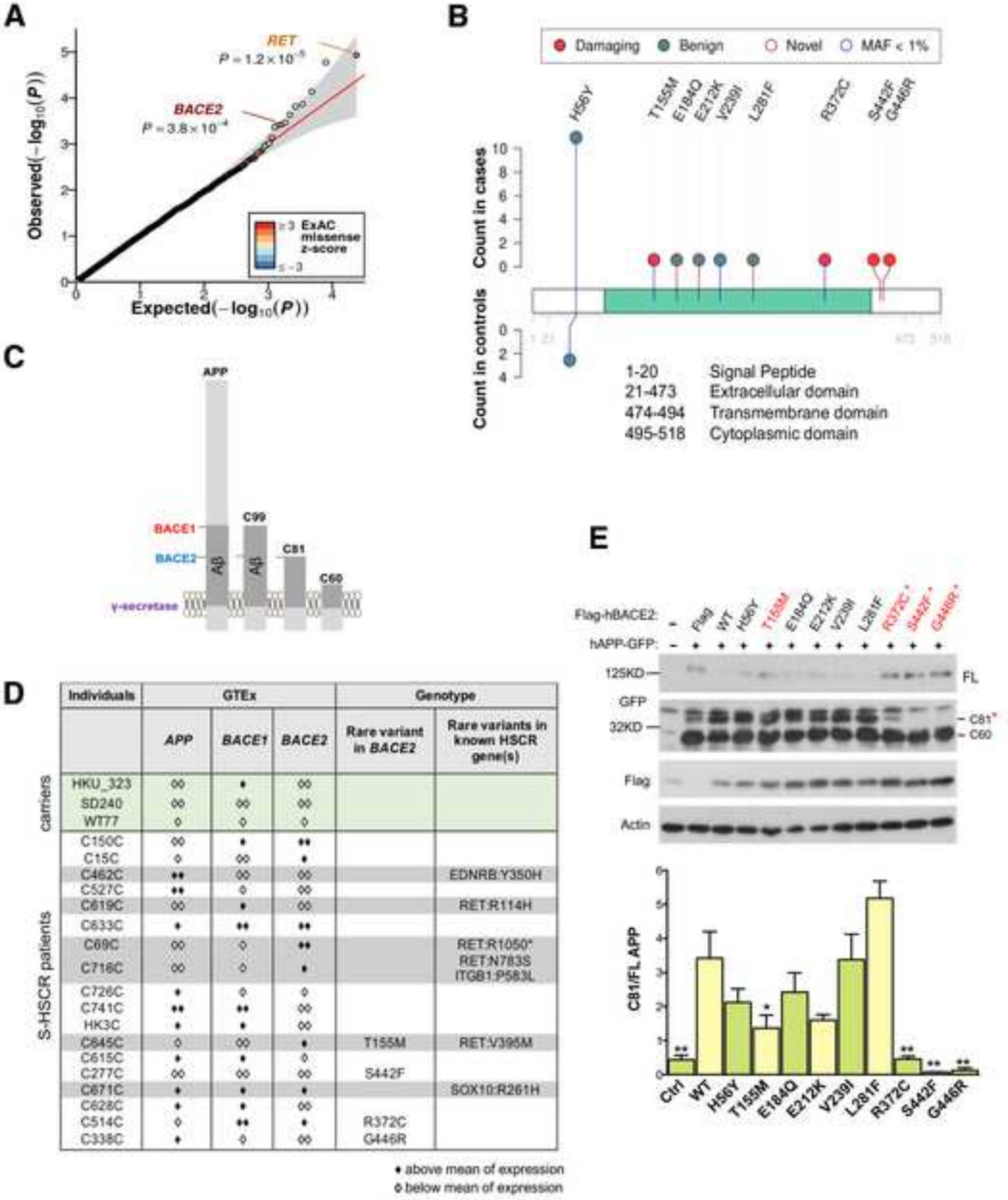
hiPC lines	Gender	Phenotype	RET		
			rs2435357 (C/T)*	rs9282834 (G/A)	Rare variant
IMR90	F	WT	CC	GG	-
UE02303	M	WT	CC	GG	-
HSCR#3	M	S-HSCR	TT	GG	-
HSCR#5	M	S-HSCR	TT	GA	-
HSCR#20	M	S-HSCR	CT	GG	T317P
HSCR#6	F	TCA	CC	GG	G731del

*Risk allele in bold

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-seq. **(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 h). The bar charts show mean± SEM from 3-7 independent experiments. "**" and "*" indicate significant different from the healthy control ($P<0.001$ and $P<0.05$, respectively). **(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 differentially expressed genes (DEGs) commonly and uniquely found in S-HSCR^{common}, S-HSCR$^{common+rare}$ and TCA-HSCR groups in comparison of the control ENCCs. **(F)** GeneMANIA shows genes (network nodes) in the association networks created using the DEGs. Genes belonging to different GO terms and the involved biological processes are indicated.

Figure 3

Click here to access/download;Figure;Figure 3.tif



	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

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 cases (top) and controls (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 eQTL-predicted expression levels of *BACE1*, *BACE2* and *APP* in individuals carrying *BACE2* rare variant (3 control 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 293FT human cell line. The full-length and cleaved APP (GFP-tagged) were detected. C81 fragment represents BACE2-cleaved APP as marked by "*". The enzymatic cleavage ability of BACE2 were 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 three independent experiments).

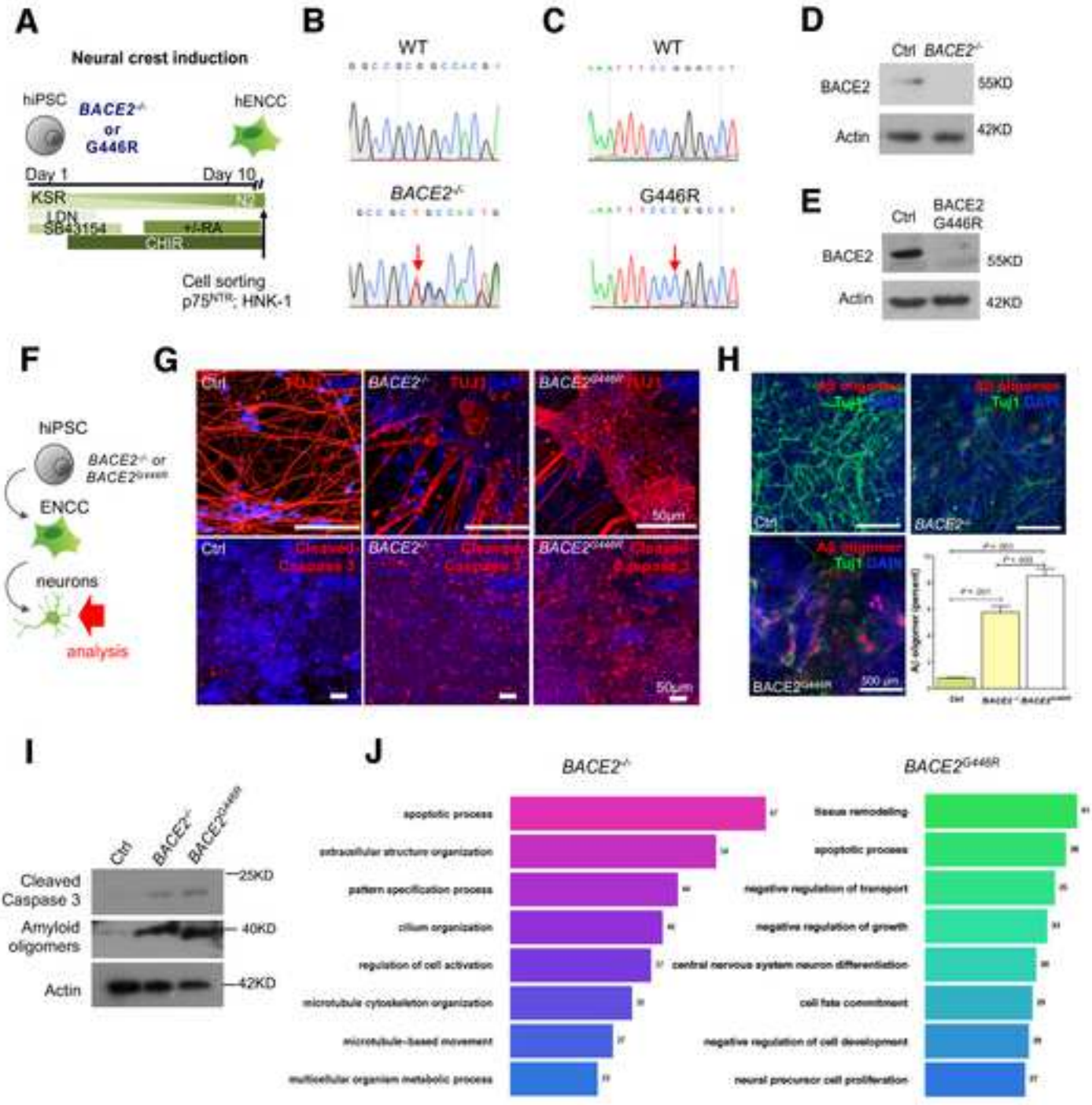


Figure 4. *BACE2* deficiency promotes Aβ accumulation and induces apoptosis in hiPSC-derived ENS neurons. (A) Stepwise differentiation protocol for the generation of ENCCs from hiPSC. Homozygous deletion and *BACE2*^{G446R} mutant hiPSC lines were generated using CRISPR/Cas9. Sanger sequencing confirms **(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 Ctrl, *BACE2*$^{-/-}$ or *BACE2*^{G446R} hiPSC at day 30 with **(G)** TUJ1 or cleaved Caspase3 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.

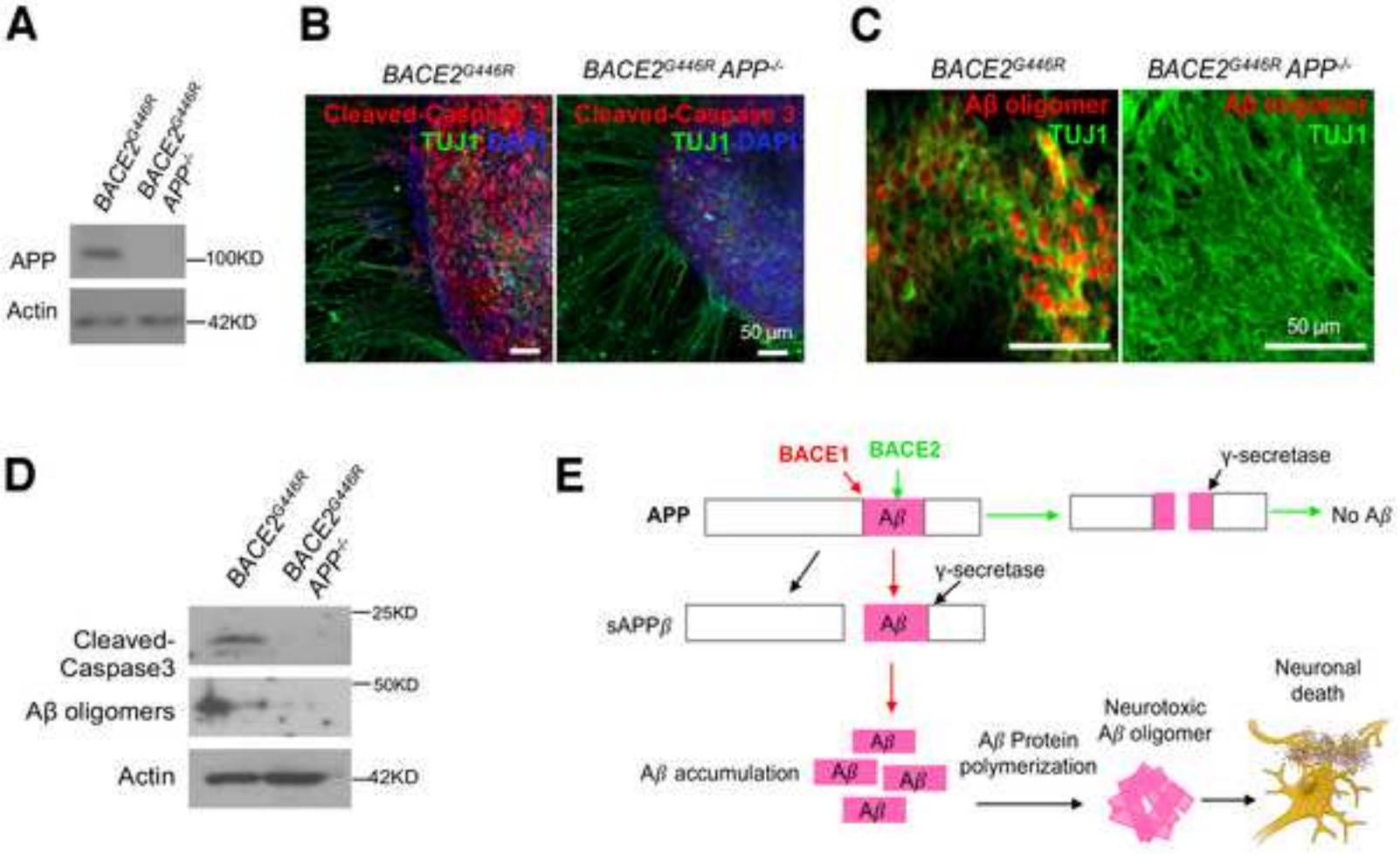


Figure 5. Deletion of *APP* improves the survival of *BACE2*^{G446R} hiPSC-derived ENS neurons. (A) Western blot analysis of protein lysates from *BACE2*^{G446R} and *BACE2*^{G446R}/*APP*$^{-/-}$ hiPSC-derived ENCCs confirms the complete knockout of *APP*. Actin was used as a loading control. Immunocytochemistry of ENS neurons derived from *BACE2*^{G446R} and *BACE2*^{G446R}/*APP*$^{-/-}$ hiPSC at day30 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.

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. In this study, we decoded the oligogenic nature of this complex disease using whole genome sequencing (WGS) and human pluripotent stem cell (hPSC)-based disease model.

NEW FINDINGS

- Four new HSCR susceptibility loci including *PLD1, CASQ2* and nine novel rare protein-altering mutations in *BACE2*
- Epistatic effects of common and rare variants across various disease loci provide a sensitized background that confers risk to disease
- Some common and distinctive biological pathways are associated with different *RET*-sensitized backgrounds
- A protective role of BACE2 in enteric neurons and the implications of BACE1-APP-BACE2 pathway in HSCR pathogenesis.

LIMITATIONS;

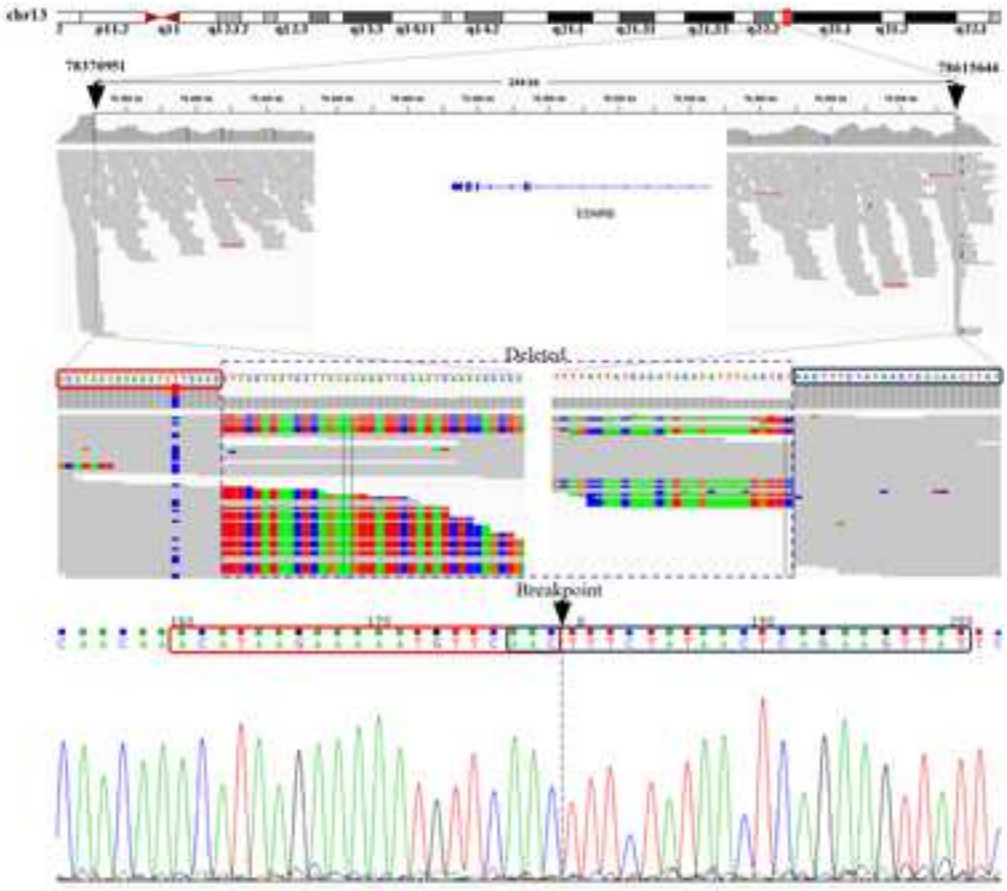
- Limited number of cases (443) and control (493) were included in the WGS analysis.
- Only *in vitro* hPSC-based model was used for functional studies.

IMPACT

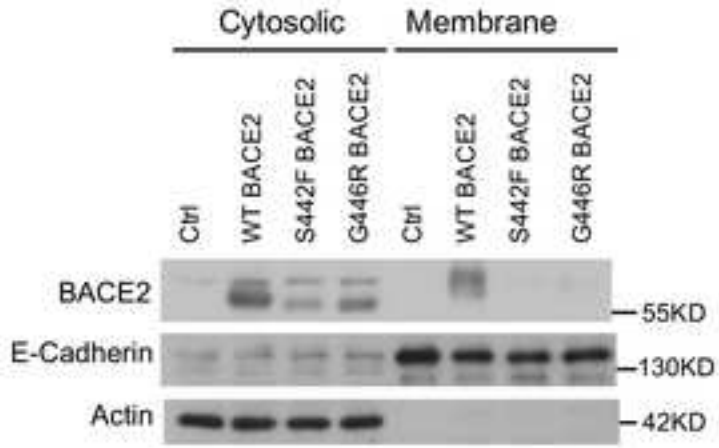
The integration of jointly analyzed rare and common variants with the use of hiPSC-based model represents a powerful approach for studying oligogenic diseases.

Lay Summary

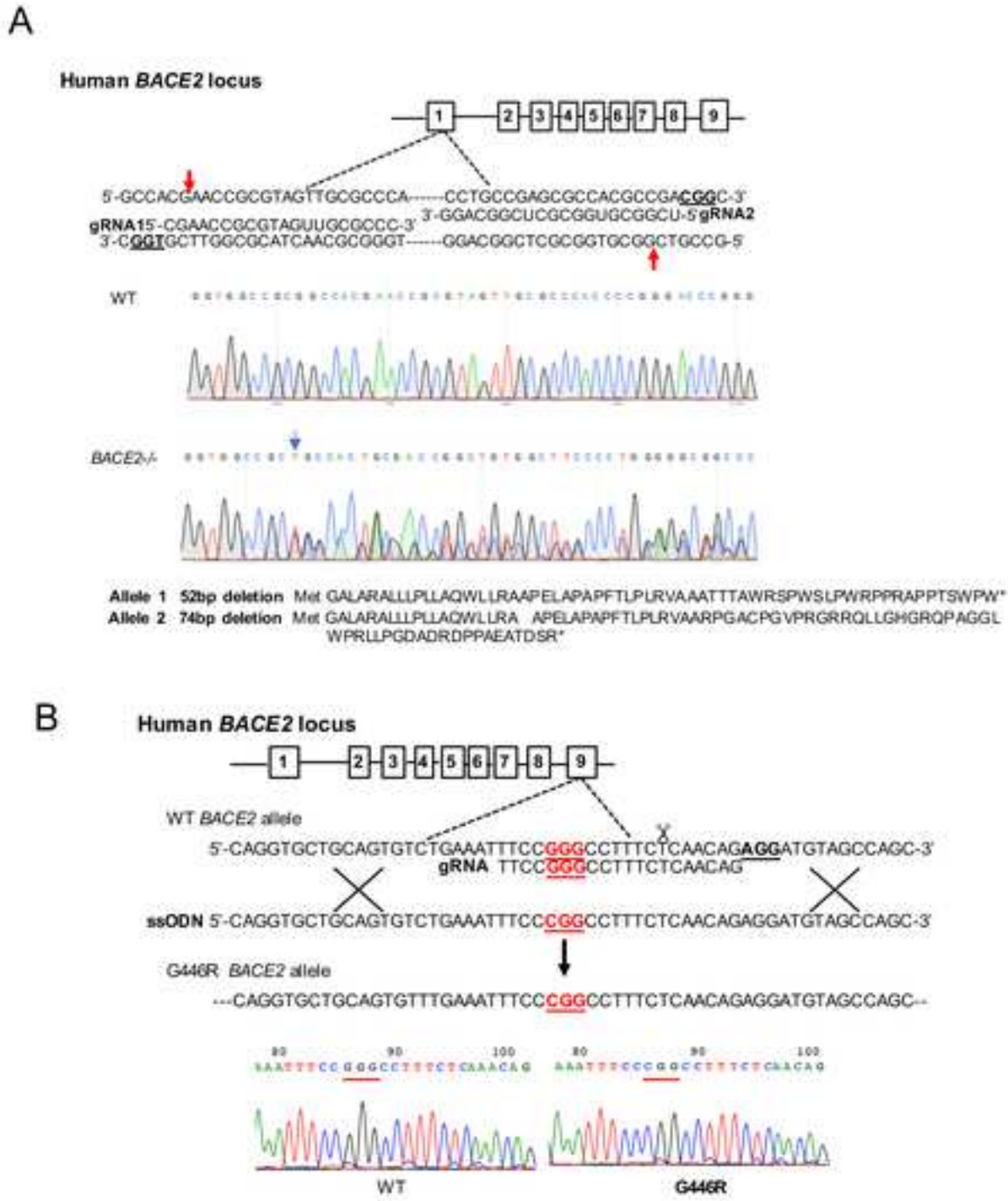
We have established an experimental paradigm by jointly analyzing rare and common genetic variants and integrating genetics with the human iPSC-based model to decode the oligogenic aetiology of HSCR.



Supplementary Figure 1. Example of *EDNRB* deletion detected by whole genome sequencing. Upper panel shows IGV plot, with soft clips, of a 245kb deletion detected by a combination of read depth, read pair and split read methods. Lower panel illustrates the validation of breakpoints by Sanger sequencing.

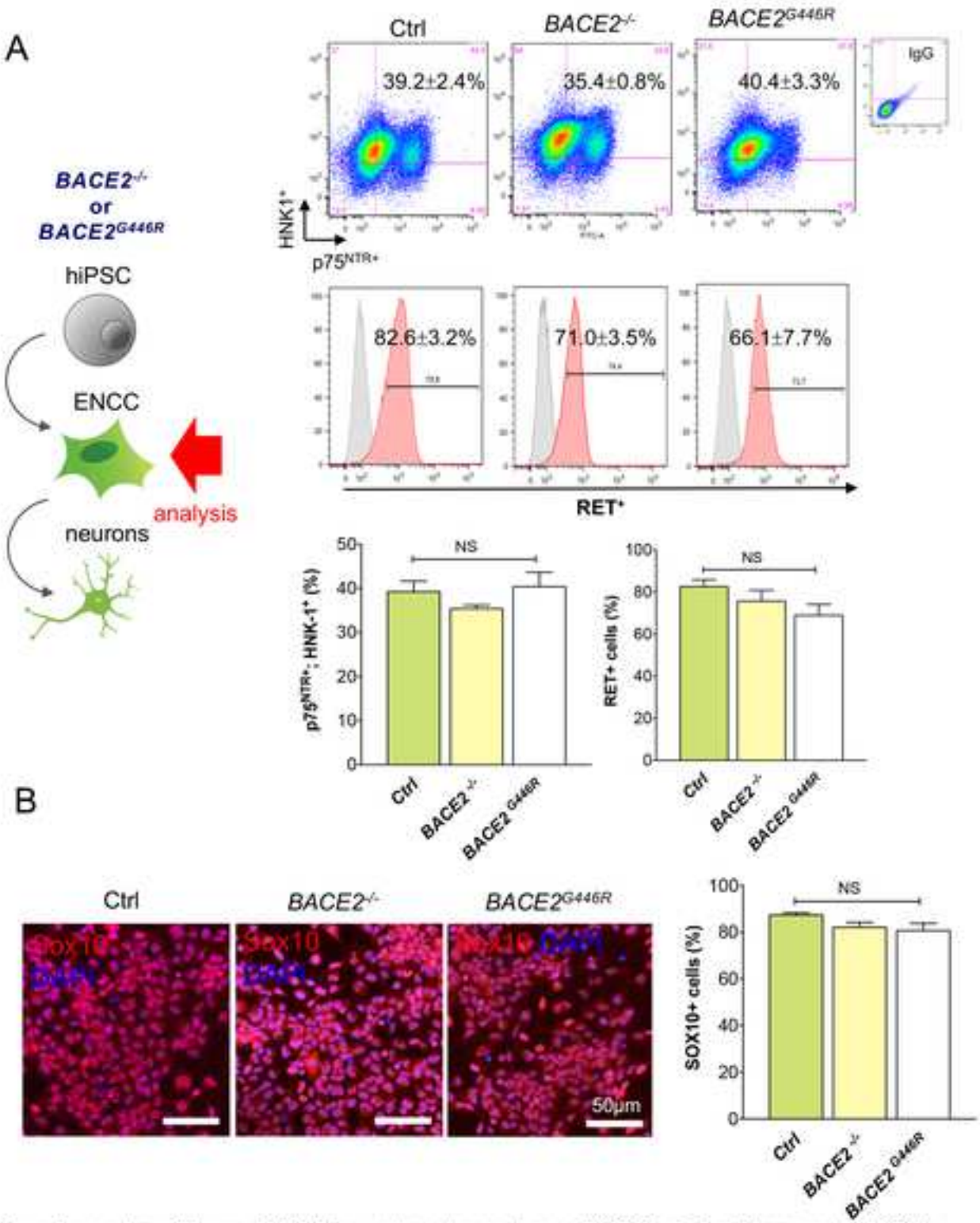


Supplementary Figure 2. Western blot analyses of proteins from membrane and cytosolic fractions of 293FT cells transfected with Flag (Ctrl), wild-type (WT) *BACE2*, S442F or G446R *BACE2* plasmids. Actin and E-cadherin were used as the loading controls for cytosolic membrane fraction, respectively.



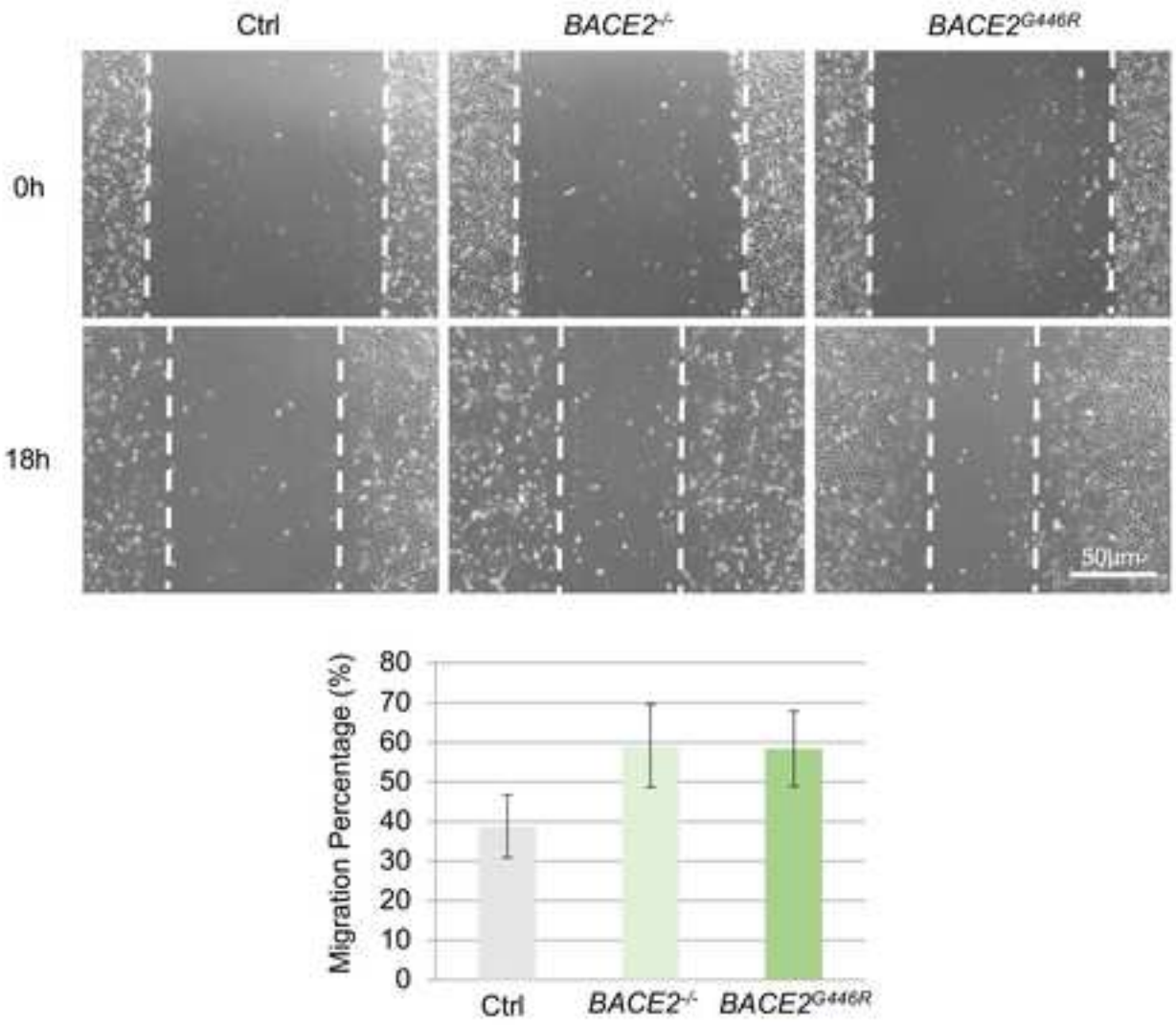
Supplementary Figure 3. (A) Schematic of CRIPSR targeting exon 1 of *BACE2* gene to generate isogenic human *BACE2* knockout from a control hPSC line. A pair of gRNAs, their targeting site and the putative cut site of Cas9D10A are indicated. Sanger sequencing shows that 52bp- and 72bp- deletions are introduced to allele1 and allele 2 of *BACE2* gene, resulting in premature termination of BACE2. Blue arrow marks the mutation site. **(B)** Targeting strategy for generation of homozygous $BACE2^{G446R}$ mutant hiPSC. gRNA targets exon 9 of *BACE2* and the putative cut site of Cas9 is indicated and $BACE2^{G446R}$ (1336G>C, shown in red) mutation was introduced into the genome through homology direct repair using a single-stranded oligo DNA nucleotide (ssODN) donor template. Sanger sequencing confirms the G to C switch at the target site.

Supplementary Figure 3



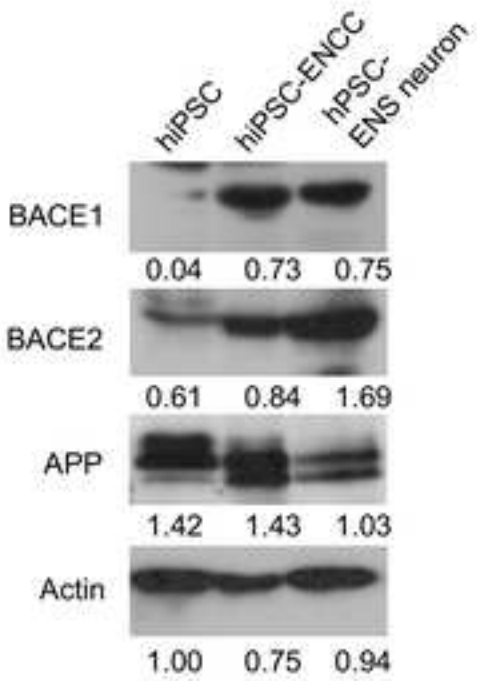
Supplementary Figure 4. (A) Flow cytometry analyses: ENCCs derived from control (Ctrl) and two BACE2 mutant (*BACE2*$^{-/-}$ and *BACE2*^{G446R}) hiPSC lines were marked by HNK1 and p75^{NTR} antibodies. RET expressing ENCCs were measured. Percentages of HNK1 and p75^{NTR} double positive and RET positive cells are shown in the bar charts. Bars represent means± SEM, n=3. **(B)** Immunocytochemistry analysis: SOX10 is expressed in FACS enriched ENCCs. Bar chart shows the percentages of SOX10+ cells in each group. Bars represent means± SEM, n=3. NS indicated no significant difference.

Supplementary Figure 4

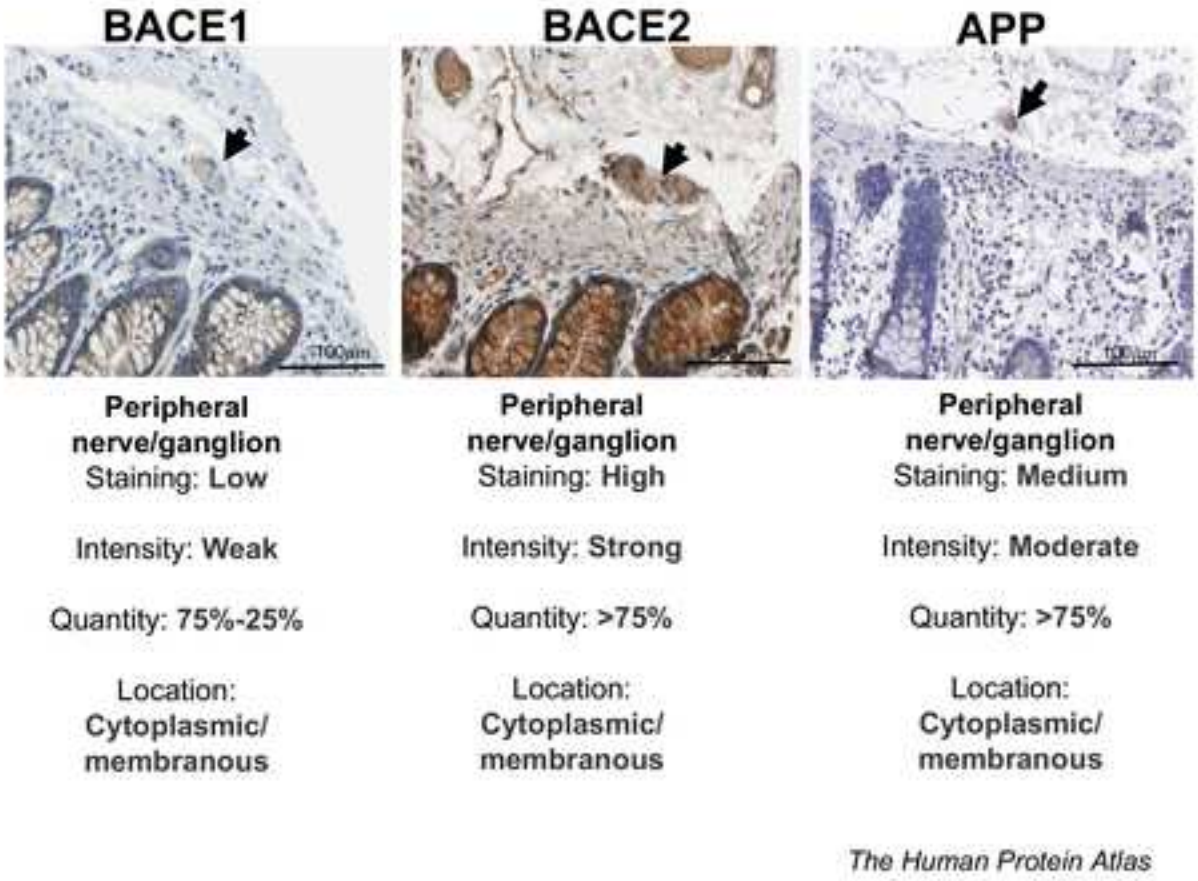


Supplementary Figure 5. Scratch assays of ENCCs derived from control (Ctrl), *BACE2*$^{-/-}$ and *BACE2*^{G446R} iPSC. The bar chart shows wound closure measured as the percentage of scar width over time (18 h). Data are shown as mean values ± SD. n = 2.

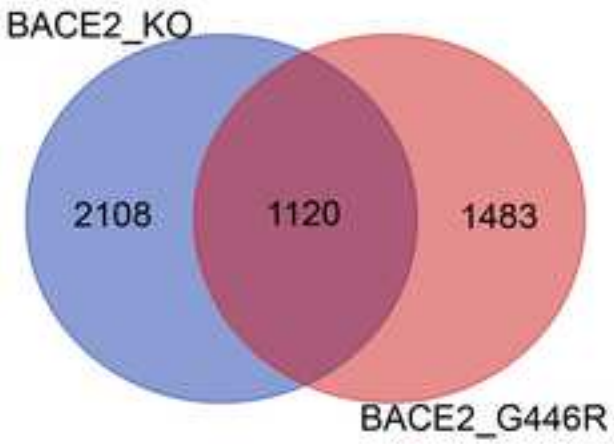
Supplementary Figure 5



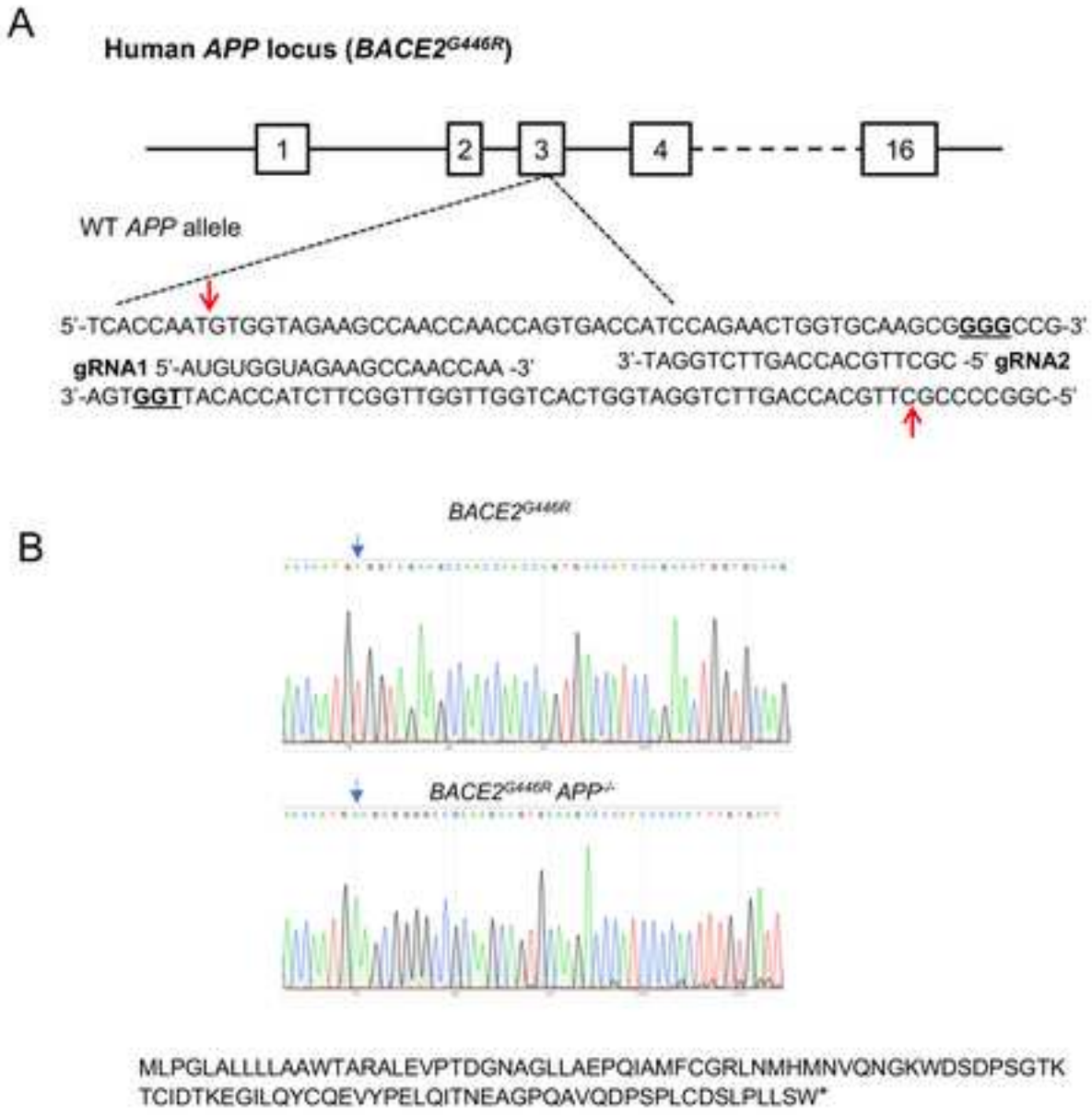
Supplementary Figure 6. Western blot analyses of proteins from control hiPSC, hiPSC-derived-ENCC and ENS neuron at day 30 of differentiation using BACE1, BACE2 and APP antibodies. Actin was used as a loading control. Numbers indicate the relative expression levels.



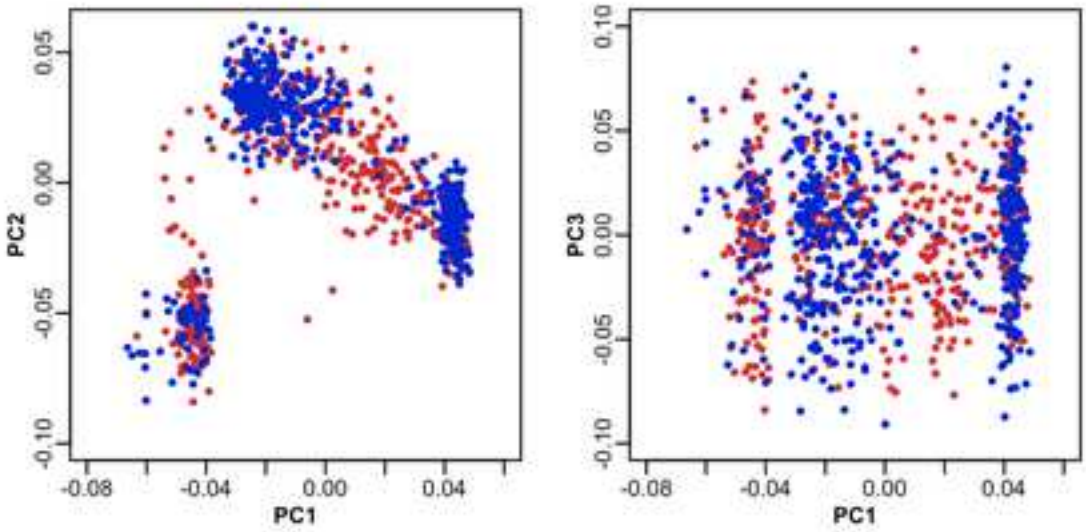
Supplementary Figure 7. Immunohistochemistry data show expression of BACE1, BACE2 and APP in enteric ganglion of human colon. Data was obtained from the Human Protein Atlas database.



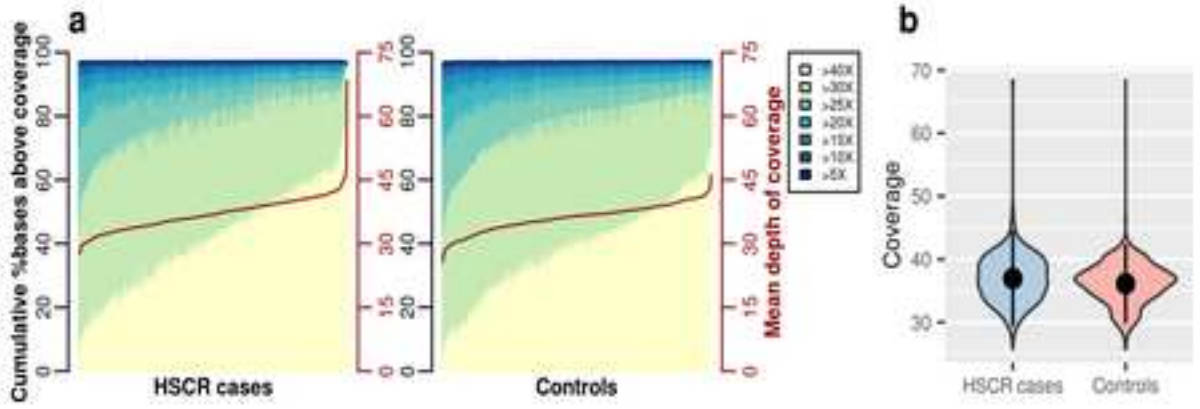
Supplementary Figure 8. Vann diagram shows the number of common and distinctive genes dysregulated in *BACE2*$^{-/-}$ and *BACE2*^{G446R} ENCC.



Supplementary Figure 9. (A) Schematic of CRIPSR targeting exon 3 of *APP* gene to generate *BACE2*^{G446R}/*APP*$^{-/-}$ hiPSC line from *BACE2*^{G446R} hPSC line. A pair of gRNAs, their targeting sites and the putative cut sites of Cas9D10A are indicated. **(B)** *BACE2*^{G446R}*APP*$^{-/-}$ carrying a 41bp-deletion in both alleles of *APP* are confirmed by Sanger sequencing. Blue arrows mark the mutation site. The corresponding amino acid sequences encoded by the mutant *APP* are shown, the deletion mutation introduces a premature stop code in APP.



Supplementary Figure 10. Principal component analysis of the WGS samples. First three principal components (PC) of cases (red) and controls (blue) are plotted.



Supplementary Figure 11. Distribution of (a) breadth of coverage and (b) mean read depth for S-HSCR cases and controls. (a) The cumulative percentages of bases above 5-40X are shown by corresponding coloured bars (left y-axis). Mean coverage of each samples (left: cases and right: controls) is denoted by the red line (right y-axis). (b) Violin plot of the mean read depth for S-HSCR cases and controls.

Supplementary Table 1: List of oligos used in this study.

Target gene	Oligo sequences	Annealing Temp (°C)	Product Size (bp)	Experiment used
BACE2 Exon 1	gRNA #1: 5'-GGGCGCAACTACGCGGTTCGTGG-3' gRNA #2: 5'-CCTGCCGAGCGCCACGCCGACGG-3'	-	-	Gene targeting
BACE2 Exon 9	gRNA: 5'-TTCCGGGCCTTTCTCAACAG-3'	-	-	Base editing
APP Exon 3	gRNA #1: 5'-TTGGTTGGCTTCTACCACATTGG-3' gRNA #2: 5'-ATCCAGAACTGGTGCAAGCGGGG-3'	-	-	Gene targeting
BACE2 Exon 1	Forward: 5'- GCTGCTGCTGCCTCTGCTGG-3' Reverse: 5'- CATGGCCAAGAAGTTGGCGG-3'	72	220	PCR sequencing
APP Exon 3	Forward: 5'- GCTTGTTAGATGCTTGTAAATG-3' Reverse: 5'- GGCTCACCTAAGCAGCGGTA-3'	58	208	PCR sequencing
BACE2 H56Y	Forward: 5'-CCCTGCCGAGCGCTATGCCGACGGCTTGG-3' Reverse: 5'-CCAAGCCGTCGGCATAGCGCTCGGCAGGG-3'	58	-	Mutagenesis
BACE2 T155M	Forward: 5'-CCAACGAAGCCCATCCAGCTTCCTTGTGTGTAC-3' Reverse: 5'-GTACACACAAGGAAGCTGGATGGGCTTCGTTGG-3'	58	-	Mutagenesis
BACE2 E184Q	Forward: 5'-TCTTGTCAACATTGCCACTATTTTTGAATCACAGAATTTCTTTTTGCC-3' Reverse: 5'-GGCAAAAAGAAATTCTGTGATTCAAAAATAGTGGCAATGTTGACAAGA-3'	58	-	Mutagenesis
BACE2 E212K	Forward: 5'-AGCCATCAAGTTCTCTGAAGACCTTCTTCGACTCC-3' Reverse: 5'- GGAGTCGAAGAAGGTCTTCAGAGAACTTGATGGCT-3'	58	-	Mutagenesis
BACE2 V239I	Forward: 5'-AGCCGGCTTGCCCATTGCTGGATCTGG-3' Reverse: 5'-CCAGATCCAGCAATGGGCAAGCCGGCT-3'	58	-	Mutagenesis
BACE2 L281F	Forward: 5'-CCAGATAGAAATTCTGAAATTCGAAATTGGAGGCCAAAGC-3' Reverse: 5'-GCTTTGGCCTCCAATTTCGAATTTCAGAATTTCTATCTGG-3'	58	-	Mutagenesis
BACE2 R372C	Forward: 5'-GGCAGGATTGTGATACAGAATGACCTGCTGGAG-3' Reverse: 5'-CTCCAGCAGGTCATTCTGTATCACAATCCTGCC-3'	58	-	Mutagenesis
BACE2 S442F	Forward: 5'-AATTGCAGGTGCTGCAGTGTTTGAAATTTCCGGG-3 Reverse: 5'-CCCGGAAATTTCAAACACTGCAGCACCTGCAATT-3'	58	-	Mutagenesis
BACE2 G446R	Forward: 5'-CTCTGTTGAGAAAGGCCGGGAAATTTCAGACACTG-3' Reverse: 5'-CAGTGTCTGAAATTTCCCGGCCTTTCTCAACAGAG-3'	58	-	Mutagenesis
BACE2 G446R	5'-CAGGTGCTGCAGTGTCTGAAATTTCCCGGCCTTTCTCAACAGAGGATGTAGCCAGC-3'	-	-	Base editing DNA template

Supplementary Table 2: List of antibodies used in this study.

Target Protein	Description	Company (Catalog number.)	Working conc	RRID number	Experiment used
RET	APC-conjugated anti-human RET	Neuromics (FC15017)	0.5µl/10^6cells	AB_1622004	Flow cytometry
HNK-1	APC-conjugated anti-human CD57 (HNK1)	BD Pharmingen (560845)	0.1µl/10^6cells	AB_10563760	FACS/flow cytometry
$p75^{NTR}$	FITC-conjugated anti-human CD271 ($p75^{NTR}$)	Miltenyi Biotec (130-091-917)	1µl/10^6cells	AB_871651	FACS/flow cytometry
Mouse IgG1 κ	APC-conjugated mouse IgG1 κ isotype control	BD Pharmingen (554681)	1µl/10^6cells	AB_398576	FACS/flow cytometry
Mouse IgG/IgM	FITC-conjugated goat anti-mouse IgG/IgM	BD Pharmingen (555988)	1µl/10^6cells	AB_396275	FACS/flow cytometry
Mouse IgM	APC-conjugated rat anti-mouse IgM	BD Pharmingen (550676)	1µl/10^6cells	AB_398464	FACS/flow cytometry
SOX10	Mouse anti-SOX10	R & D system (MAB2864)	1:500	AB_2195180	IF
TUJ1	Mouse monoclonal anti-TUJ1	Convance (PRB-435P)	1:500	AB_291637	IF
Cleaved-caspase-3	Rabbit monoclonal anti-cleaved caspase-3 (Asp175) (5A1E)	Cell signaling (#9664)	1:100	AB_2070042	IF
Amyloid oligomers	Rabbit polyclonal anti-Amyloid oligomers	Abcam (ab126892)	1:100	AB_11128526	IF
BACE2	Rabbit polyclonal anti-BACE2	Thermo Fisher Scientific (PA1-754)	1:1000	AB_326006	WB
Flag	Rabbit polyclonal anti-Flag	Sigma (A7425)	1:2000	AB_439687	WB
GFP	Goat polyclonal anti-GFP	Rockland (600-101-215)	1:2000	AB_218182	WB
Actin	Mouse monoclonal anti-Actin	Millipore (MAB1501)	1:5000	AB_2223041	WB
E-cadherin	Goat polyclonal anti-E-cadherin	R & D system AF648	1:2000	AB_355504	WB
Cleaved-Caspas3	Rabbit monoclonal anti-cleaved caspase-3 (Asp175) (5A1E)	Cell signaling (#9664)	1:1000	AB_2070042	WB
Amyloid oligomers	Rabbit polyclonal anti-Amyloid oligomers	Abcam (ab126892)	1:1000	AB_11128526	WB
APP	Rabbit polyclonal anti-Amyloid Precursor Protein, C-Terminal	Sigma (A8717)	1:1000	AB_258409	WB
BACE1	Rabbit polyclonal anti-BACE1	Thermo Fisher Scientific (PA1-757)	1:1000	AB_325863	WB
Mouse IgG	HRP conjugated anti-mouse antibody	DAKO (P0447)	1:2500	AB_2617137	WB
Rabbit IgG	HRP conjugated anti-rabbit antibody	DAKO (P0448)	1:2500	AB_2617138	WB
Goat IgG	HRP conjugated anti-goat antibody	DAKO (P0449)	1:2500	AB_2617143	WB

Supplementary Table 3. Common variants showing moderate-to-high association with HSCR with $P<1x10^{-6}$.

Chr	Position	Ref/Alt[a]	rsID	Genomic Feature	Genes	Alt allele frequency[a]		OR	95%CI		P^b
						S-HSCR	Controls				
10	43580224	C/A	rs2506008	Intronic	*RET*	0.175	0.506	0.24	0.19,	0.30	**$1.5x10^{-34}$**
8	32401501	T/G	rs7005606	Upstream	*NRG1*	0.348	0.216	1.92	1.55,	2.38	**$3.4x10^{-9}$**
10	43240428	A/T	rs1414027	Intergenic	*LINC01518,LOC283028*	0.210	0.325	0.55	0.44,	0.69	$1.4x10^{-7}$
1	116302923	C/T	rs9428225	Intronic	*CASQ2*	0.472	0.356	1.62	1.34,	1.96	$6.6x10^{-7}$
3	171511355	G/A	rs12632766	Intronic	*PLD1*	0.196	0.293	0.57	0.45,	0.71	$7.4x10^{-7}$
3	27601508	C/A	rs9851320	Intergenic	*SLC4A7,EOMES*	0.248	0.164	1.84	1.44,	2.34	$9.2x10^{-7}$

[a] Ref/Alt denotes the reference and alternatives alleles respectively

[b] Variants passing genome-wide significance level ($P<5x10^{-8}$) are highlighted in bold

Supplementary Table 4 Rare variants in the HSCR major gene, *RET*.

Chr	Position	Ref/Alt	rsID	Genomic Feature[a]	Nucleotide change	Protein change	Exon	Count		Max freq[b]	Damaging[c]
								S-HSCR	Controls		
10	43595990	G/A	.	missense	c.157G>A	p.V53I	2	1	0	$7.8x10^{-4}$	N
10	43596026	A/G	.	missense	c.193A>G	p.S65G	2	1	0	N	N
10	43596033	G/A	rs192489011	missense	c.200G>A	p.R67H	2	12	13	0.026	N
10	43596128	C/T	.	missense	c.295C>T	p.R99W	2	1	0	N	N
10	43597792	C/T	.	missense	c.340C>T	p.R114C	3	1	0	$8.1x10^{-6}$	N
10	43597793	G/A	rs76397662	missense	c.341G>A	p.R114H	3	30	11	$6.0x10^{-3}$	N
10	43597849	C/T	.	missense	c.397C>T	p.R133C	3	1	0	N	Y
10	43597916	C/T	.	missense	c.464C>T	p.P155L	3	1	0	N	Y
10	43597976	G/C	.	missense	c.524G>C	p.R175P	3	1	0	N	N
10	43600606	A/G	.	missense	c.832A>G	p.T278A	4	4	0	$1.0x10^{-3}$	N
10	43600607	C/A	rs35118262	missense	c.833C>A	p.T278N	4	5	16	0.019	N
10	43601830	G/A	rs34682185	missense	c.874G>A	p.V292M	5	11	9	0.017	Y
10	43601905	A/C	.	missense	c.949A>C	p.T317P	5	1	0	N	Y
10	43601945	G/A	rs80236571	missense	c.989G>A	p.R330Q	5	1	0	N	Y
10	43601972	C/T	.	missense	c.1016C>T	p.S339L	5	1	0	$8.1x10^{-6}$	N
10	43601986	G/A	.	missense	c.1030G>A	p.G344S	5	1	0	$8.1x10^{-6}$	N
10	43601990	G/T	.	missense	c.1034G>T	p.S345I	5	1	0	N	N
10	43601995	G/A	.	missense	c.1039G>A	p.V347M	5	1	0	N	N
10	43604592	T/A	.	missense	c.1177T>A	p.F393I	6	0	1	N	Y
10	43604598	G/A	.	missense	c.1183G>A	p.V395M	6	1	0	N	N
10	43606678	C/A	.	missense	c.1287C>A	p.N429K	7	0	1	N	N
10	43606776	C/A	.	**stopgain**	c.1385C>A	p.S462*	7	1	0	N	Y
10	43606832	C/G	.	missense	c.1441C>G	p.L481V	7	0	1	$3.3x10^{-5}$	N
10	43606850	G/A	.	missense	c.1459G>A	p.A487T	7	1	0	N	Y
10	43606856	G/A	rs9282834	missense	c.1465G>A	p.D489N	7	27	17	0.019	N

Supplementary Table 4 (continued). Rare variants in the HSCR major gene, *RET*.

Chr	Position	Ref/Alt	rsID	Genomic Feature[a]	Nucleotide change	Protein change	Exon	Count		Max freq[b]	Damaging
								S-HSCR	Controls		
10	43607621	G/A	rs75873440	missense	c.1597G>A	p.G533S	8	2	2	$1.0x10^{-3}$	Y
10	43607642	A/G	.	missense	c.1618A>G	p.R540G	8	0	2	$1.0x10^{-3}$	Y
10	43608354	G/A	rs140464432	missense	c.1702G>A	p.G568S	9	1	1	$1.2x10^{-4}$	Y
10	43609001	CAG/C--	.	**frameshift**	c.C1760del-AG	p.R587fs	splicing-exon10	1	0	N	Y
10	43609027	G/C	.	missense	c.1783G>C	p.E595Q	10	1	0	N	Y
10	43609042	C/T	.	missense	c.1798C>T	p.R600W	10	0	1	$8.1x10^{-6}$	NA
10	43609070	G/C	rs77939446	missense	c.1826G>C	p.C609S	10	1	0	N	Y
10	43609939	G/A	.	missense	c.1891G>A	p.D631N	11	0	1	$5.7x10^{-5}$	Y
10	43609942	G/A	.	missense	c.1894G>A	p.E632K	11	1	0	$2.0x10^{-4}$	Y
10	43612141	G/C	rs34288963	missense	c.2246G>C	p.R749T	12	0	3	$1.0x10^{-3}$	N
10	43613884	A/G	.	missense	c.2348A>G	p.N783S	13	3	0	N	Y
10	43615074	G/A	rs200127630	missense	c.2488G>A	p.G830R	14	0	1	$1.0x10^{-3}$	Y
10	43615084	G/A	.	missense	c.2498G>A	p.R833H	14	0	1	$2.4x10^{-5}$	Y
10	43615652	G/A	.	**splicing**	c.2730+1G>A		exon16GTdonor	1	0	N	Y
10	43620335	C/T	rs17158558	missense	c.2944C>T	p.R982C	18	15	15	0.031	Y
10	43620336	G/A	rs368550200	missense	c.2945G>A	p.R982H	18	0	1	$8.1x10^{-5}$	Y
10	43620345	T/G	.	missense	c.2954T>G	p.L985R	18	0	1	N	Y
10	43622131	C/T	.	**stopgain**	c.3148C>T	p.R1050*	19	1	0	N	Y
10	43622176	AT/A-	.	**Frameshift**[d]	c.A3194del-T	p.I1065fs	19	1	0	N	Y
10	43623643	C/A	.	missense	c.3271C>A	p.P1091T	20	0	1	N	N

[a] Genomic feature with reference to NM_020975; Loss-of-function ultra-rare variants are highlighted in bold

[b] Maximum frequency across public databases (ExAC, 1000 Genomes Project, ESP); N: absence in any database

[c] *In silico* prediction of damaging effect by KGGseq; Y: damaging; N: benign; NA : not annotated

[d] Genomic feature with reference to NM_020630

Supplementary Table 5. Large deletions overlapping *EDNRB*

Sample ID	Sex	Start	End	Length (kb)	Het:Hom ratio	Software called	Genes overlapped
HK136C	M	62422064	79108403	16686.3	8:217236	Delly,Lumpy,Seeksv	*BORA,CLN5,COMMD6,DACH1,DIS3,**EDNRB**,FBXL3,IRG1, KCTD12,KLF12,KLF5,KLHL1,LMO7,LMO7DN,MYCBP2,MZT1, PCDH9,PIBF1,SCEL,SLAIN1,TBC1D4,UCHL3*
C745C	M	72564111	83200496	10636.4	34:129529	Delly,Lumpy,Seeksv	*BORA,CLN5,COMMD6,DIS3,**EDNRB**,FBXL3,IRG1,KCTD12, KLF12,KLF5,LMO7,LMO7DN,MYCBP2,MZT1,NDFIP2,PIBF1, POU4F1,RBM26,RNF219,SCEL,SLAIN1,SPRY2,TBC1D4,UCHL3*
C531C	F	78370951	78615644	244.7	0:3210	Delly,Lumpy,Seeksv, CNVnator	***EDNRB***

Supplementary Table 6. Rare variant profile for individuals with double hit of known ENS genes

		Benign	Damaging						
Sample ID	**Sex**	***RET***	***RET***	***GDNF***	***GFRA1***	***EDNRB***	***EDN3***	***ERBB2***	**Other ENS genes**
C48C	M	-	+	-	-	-	-	+	
HK4C	M	-	+	-	-	-	-	-	*ZEB2*
C716C	M	-	+	-	-	-	-	-	*ITGB1*
C692C	F	+	-	-	-	-	-	+	
C718C	F	+	-	-	-	-	-	+	
HK19C	F	+	-	-	-	-	-	-	*IHH*
C626C	M	+	-	+	-	-	-	-	*SOX10*
HK114C	M	-	-	-	+	+	-	-	
HK8C	F	-	-	-	-	+	-	+	
C531C	F	-	-	-	-	+	-	-	*ITGB1*
VH88C	M	-	-	-	-	-	+	+	
HK134C	F	-	-	+	+	-	-	-	
C749C	F	-	-	+	-	-	+	-	

Supplementary Table 7. ERBB2-interacting partners with significant burden (FDR<0.1) of rare damaging variants

Gene	Protein-protein interacting partners		Count		OR (95% CI)	*P*
	Known ENS genes	Other ERBB2-interacting partners	Cases	Controls		
ITGB4	ERBB2	PTK2	31	12	3.02 (1.53, 5.95)	1.04×10^{-3}
PTK2	ERBB2 & RET	ITGB4, ERBB3, ITGB1 & DCC	9	0	NA	1.39×10^{-3}

Supplementary Table 8. Differentially expressed genes (DEGs) uniquely or commonly found in S-HSCR$^{common+rare}$, S-HSCR^{common} and TCA-HSCR-iPSC-derived ENCCs in comparison of the control ENCCs. Fold change (Log2FC) and *p*-values are shown.

Please refer to the excel file attached.

Supplementary Table 9. Rare missense *BACE2* variants detected in the discovery WGS and follow-up Sanger sequencing.

Stage	Chr	Position	Ref/Alt	rsID	Nucleotide change[a]	Protein change	Exon	Count		Max freq[b]	Damaging[c]
								S-HSCR	Controls		
Discovery & follow-up	21	42540356	C/T	rs200758268	c.166C>T	p.H56Y	1	11	4	$8.1x10^{-3}$	N
Discovery	21	42609502	C/T	.	c.464C>T	p.T155M	3	1	0	$4.1x10^{-5}$	Y
Discovery	21	42609588	G/C	.	c.550G>C	p.E184Q	3	1	0	N	N
Discovery	21	42613761	G/A	.	c.634G>A	p.E212K	4	1	0	N	N
Discovery	21	42613842	G/A	.	c.715G>A	p.V239I	4	1	0	$2.0x10^{-4}$	N
Discovery	21	42615398	G/C	.	c.843G>C	p.L281F	5	1	0	N	N
Discovery	21	42622808	C/T	rs139556900	c.1114C>T	p.R372C	7	1	0	$1.0x10^{-3}$	Y
Discovery	21	42647319	C/T	.	c.1325C>T	p.S442F	9	1	0	N	Y
Discovery	21	42647330	G/A	.	c.1336G>A	p.G446R	9	1	0	N	Y
Follow-up	21	42647478	G/A	rs756173057	c.1484G>A	p.R495Q	9	0	1	$8.2x10^{-6}$	N
Follow-up	21	42647496	G/A	rs749800434	c.1502G>A	p.R501H	9	0	1	$8.3x10^{-5}$	N

[a] Nucleotide change is based on BACE2 isoform with accession of NM_012105

[b] Maximum frequency across public databases (ExAC, 1000 Genomes Project, ESP); N: absence in any database

[c] *In silico* prediction of damaging effect by KGGseq; Y: damaging; N: benign

Supplementary Table 10. Differentially expressed genes (DEGs) found in *BACE2*$^{-/-}$ and *BACE2*^{G446R} iPSC-derived ENS neurons in comparison of the control. Log2 Fold changes and *p*-values are shown.

Please refer to the excel file attached.

Supplementary Table 11. Demographics and clinical characteristics of whole-genome sequenced samples

Sex	S-HSCR Cases (N=443)		Controls (N=493)	
	Chinese (N=341)	Vietnamese (N=102)	Chinese (N=442)	Vietnamese (N=51)
Males	267 {30}(9)	75 {1}(1)	363	36
Females	74 {11}(3)	27 {2}(2)	79	15

Numbers in { }, additional anomalies; numbers in ( ), Down syndrome.

Supplementary Table 12: Index variants used to predict *APP* expression based on GTEx whole blood model by PrediXcan and their linkage disequilibrium (LD) proxies (r^2=1).

Chr	Index variant	Proxy variant	Position	Ref/Alt[a]	1000 Genomes allele frequency[a]				*APP* Enhancer[b]	*APP* Promoter	Motif[c]
					AFR	AMR	ASN	EUR			
21	rs2830053		27471711	C/T	0.02	0.14	0.12	0.10	Neural		HAND1,PAX2
21	rs56034894		27482766	A/C	0.08	0.28	0.17	0.12			FOXD3,FOXF1,FOXJ2, FOXJ3, FUBP1,SRY
21	rs2830065		27493526	G/T	0.02	0.14	0.12	0.10	Neural+blood		FOXP3
21		rs4816276	27477155	C/G	0.03	0.14	0.12	0.10			IRF1,IRF2,IRF3,IRF4, IRF8,MLXIPL,NR2E3 PRDM1,STAT2
21		rs2830056	27478349	G/A	0.03	0.14	0.12	0.10	Neural		SMAD1,TP53,TP63, XBP1
21		rs3827166	27481317	T/G	0.03	0.14	0.12	0.10			
21		rs45495091	27485768	G/A	0.03	0.14	0.12	0.10	Neural		
21		rs73168372	27486006	T/C	0.03	0.14	0.12	0.10	Neural		ELF3,ETV7
21		rs2830061	27486262	T/A	0.02	0.14	0.12	0.10	Neural		TAL1
21		rs10482971	27493547	C/T	0.03	0.14	0.12	0.10	Neural+blood		POU6F1
21		rs2830067	27495183	G/A	0.03	0.14	0.12	0.10	Neural		RXRG
21	rs2830596		28355042	G/A	0.53	0.30	0.24	0.50			RFX3
21	rs2830050		27464270	A/G	0.79	0.80	0.94	0.74	Neural		
21	rs2830081		27507649	A/C	0.50	0.37	0.20	0.32	Neural		
21		rs13046704	27506969	G/C	0.50	0.37	0.19	0.32	Neural		FOXC2
21	rs4172		28025461	C/A	0.40	0.20	0.23	0.18			
21	rs45473297		27541906	G/A	0.33	0.29	0.10	0.31		Neural+blood	KLF8,MAZ

[a] Alternative allele frequency in 1000 Genomes Phase 1 data; AFR: African; AMR; Ad mixed American; ASN: East Asian and EUR: European

[b] Enhancer and promoter predictions are based on ChromHMM-based chromatin segmentation from ROADMAP. Neural: brain tissues and neuronal progenitors. Blood: T and B cells.

[c] Motif of transcription factors strongly affected by the variant

Supplementary Table 13: Index variants used to predict *BACE1* expression based on GTEx whole blood model by PrediXcan.

Chr	Index variant	Position	Ref/Alt	1000 Genomes allele frequency				*BACE1* Enhancer	Motif
				AFR	AMR	ASN	EUR		
11	rs58283940	116238556	G/A	0.05	0.06	0.25	0.09		
11	rs480101	116314401	C/T	0.31	0.22	0.29	0.18		NR2E3
11	rs11215985	116398763	G/A	0.69	0.58	0.41	0.56		
11	rs7936161	116411448	C/T	0.36	0.06	0.04	0.02	Neural	REST
11	rs7483863	116652490	A/G	0.99	0.89	0.76	0.91		GLI1,NFKB1,REST,ZFX
11	rs11216153	116705099	G/T	0.11	0.20	0.26	0.15		NFKB2
11	rs144037495	117035815	C/T	0.01	0.05	0.17	0.01		TCF4
11	rs1056136	117120997	T/G	0.35	0.18	0.27	0.20		ARID3A,FOXC1,FOXO4,FOXQ1
11	rs595297	117154090	G/T	0.87	0.86	0.89	0.83		FOXC2,FOXD3,FOXF1,FOXJ2,FOXJ3,FUBP1,SRY
11	rs525493	117182707	T/G	0.45	0.49	0.76	0.42		HIVEP1
11	rs3017608	117189215	G/A	0.5	0.53	0.77	0.44		TEAD1
11	rs560564	117195627	A/C	0.47	0.47	0.73	0.37		FOXA3,FOXD3,FOXF1,FOXJ2,FOXJ3
11	rs73014325	117196434	C/A	0.05	0.03	0.01	0.05		SPIB
11	rs549289	117358219	A/C	0.41	0.33	0.41	0.21		
11	rs12419495	117383815	T/C	0.09	0.07	0.1	0.03	Neural	
11	rs77944184	117388656	G/A	0.08	0.05	0.1	0.03		
11	rs498689	117462999	A/G	0.55	0.39	0.38	0.29	Neural	
11	rs694005	117464172	C/T	0.40	0.38	0.37	0.29		
11	rs10892178	117685342	C/T	0.25	0.38	0.24	0.41	Neural+blood	
11	rs553709	117726473	C/T	0.49	0.84	0.97	0.88	Neural	HEY2,HNF4G,ZBTB4
11	rs12788624	118022106	C/T	0.1	0.12	0.15	0.12		

[a] Alternative allele frequency in 1000 Genomes Phase 1 data; AFR: African; AMR; Ad mixed American; ASN: East Asian and EUR: European
[b] Enhancer and promoter predictions are based on chromatin segmentation from ROADMAP. Neural: brain tissues and neuronal progenitors. Blood: T and B cells.
[c] Motif of transcription factors strongly affected by the variants

METHODS

Patients

The discovery cohort comprised 464 S-HSCR cases and 498 controls analyzed by whole genome sequencing (WGS). All patients analyzed are sporadic, 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 analysed for rare coding variants in *RET*[1] and 98 S-HSCR cases were genotyped by Affymetrix 500K SNP array for genome-wide association analysis[2,3]. To avoid cofounding effects due to population stratification, controls were ascertained from the same or nearby cities to match with the subpopulations of the patients. Samples failing heterozygosity, gender 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 and 493 controls remained. PCA of the resulting genetic data indicated that cases and controls were well-matched for ethnic origin (Supplementary Fig. 10). Demographic information of these samples is summarized in Supplementary Table 11. The follow-up cohort included 534 ethnically-matched controls 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 ((HKU/HA HKW IRB) UW 13-225).

Whole genome sequencing and variant calling

All samples were whole genome sequenced using Illumina HiSeq X Ten at Macrogen (Korea) to a mean coverage of 30X. Raw sequence reads were first aligned to human reference genome (hg19) using Burrows–Wheeler Aligner (BWA-MEM)[4]. After alignment, we achieved an average mapped read depth of 26-64X (median of 36.6X)

and the breadth of coverage was highly comparable between cases and controls (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 (VQSR). 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, i.e., those with genotype quality less than 20 (GQ<20) or covered by less than 8 reads (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<1x10^{-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 (Ti/Tv) ratio for novel SNVs is 2.10 across the whole genome and 3.00 while considering only the exome. A genotype concordance of 99.96% between duplicates was achieved, which demonstrated 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 (e.g., PolyPhen2, SIFT, and CADD) as well as population frequencies (e.g., 1000 Genomes Project phase 3, Exome Aggregation Consortium (ExAC v0.2) and NHLBI Exome Sequencing Project (ESP) databases). We defined protein-truncating variants as variants that lead to (i) gain of the stop codon, (ii) frameshift and (iii) 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. Lastly, 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, Exome Aggregation Consortium (ExAC v0.2) and NHLBI Exome Sequencing Project databases.

Association between ultra-rare variants (URVs) and HSCR

URV was defined as a singleton variant, that is one that appeared only once in our whole dataset, not present in dbSNP138, and was not seen in public databases (1000 Genomes Project phase 3, Exome Aggregation Consortium (ExAC v0.2) and NHLBI Exome Sequencing Project databases). We adopted the definition of constrained gene as described previously[7], i.e., loss-of-function-intolerant (ExAC pLI≥0.9; n=3488). Protein-truncating URVs are further divided into two types, one that can elicit nonsense-mediated decay (NMD) and another that escapes NMD (non-NMD). Non-NMD URVs are protein-truncating variants with 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 up the corresponding number of singleton URVs in genes expressed in ENCC. 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 adjusted for the first three principal components (PCs). To avoid confounding due to technical artefacts, number of synonymous URVs was added as a covariate while 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 (MGI) were considered as known ENS genes. ENS interactome was defined by genes encoding proteins that show protein-protein interaction (PPI) 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, including 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 50bp. Only those CNVs (i) larger than 50bp, (ii) called by at least 3 software and (iii) were 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 identified as CNVs with <50% reciprocal overlap against all CNVs found in the WGS controls as well as those documented in the Database of Genomic Variants (DGV)[13] and in the population controls of DECIPHER[14] within the known ENS loci.

Gene-based and geneset-based burden test for rare variants

For the set of known ENS genes, we first assessed the enrichment of (i) damaging and (ii) all rare protein-altering variants collectively in cases compared to controls. Gene-based and geneset-based association tests adjusted for the first three principal

components from the aforementioned PCA were carried out using combined multivariate and collapsing (CMC) test[15] and SKAT-O respectively by rvtests[16]. Genes and genesets passing false discovery rate (FDR) <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 below: *Neural crest induction*), we performed gene-based association test for all rare protein-altering variants, regardless of the *in silico* deleterious prediction. Genes with association $P<4.0\text{x}10^{-6}$, 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 if 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 three groups carrying zero, one or at least two common HSCR-associated risk alleles. Samples were further subdivided into three subgroups (totalling 9=3x3 combinations), according to the presence of mutations and their predicted pathogenicity (damaging and benign). Next, we computed the odds ratio for individuals with each combination relative to the baselines, i.e. not carrying any common risk allele nor rare variants. For dosage-specific odds ratio (OR_{dosage}), groups of samples carrying no mutation under the same common risk allele dosage were used as 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 (PIRs) spanning at least 2 heterozyguous sites were obtained from the BAM files and were used to phase the common and rare variants in the genotype data in VCF file.

Imputation of expression using PrediXcan

To impute the gene expression of *BACE2, BACE1* and *APP*, we considered two tissue models (each with >300 samples), (i) the neural (tibial nerve, 361 individuals) and (ii) whole blood (369 individuals), built from expression quantitative loci (eQTLs) of the GTEx database (v6p release) using PrediXcan[18]. As no linear model on tibial nerve (FDR<5%) was available in PredictDB for *BACE1* and *APP*, we examined if whole blood tissue is a good surrogate for the unmeasured ENCC. 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. While many of selected eQTLs overlapped with neural (brain and neuronal progenitors) or blood samples enhancers or promoters in ROADMAP and were predicted to strongly affect motifs of transcription factors important in ENS development, e.g. 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 as it provides a better fit than whole blood tissue model (P=2.6x10^{-14} for tibial nerve versus P=3.4x10^{-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 controls.

Human induced pluripotent stem cells (hiPSC)

Two control human induced pluripotent stem cell lines (IMR90 and UE02302-hiPSC) were used to generate ENCCs and ENS neurons. IMR90 iPSC (clone#2) was purchased from Wicell Research Institute, UE02302 is 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 plate 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 (#48138 and #44720 respectively). The guide RNAs (gRNA) 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 #41824) was linearized with AflII and the gene-specific gRNA targeting sequence was incorporated into gRNA plasmid using Gibson assembly (New England Biolab, Cat no. E2611L) according to manufacturer's protocol. All the gRNA oligo sequences are listed in supplementary Table 1.

An expression plasmid containing human *BACE2* full length cDNA (NM_012105.3)

was purchased from Sino Biological Inc. (HG10783-NF). Human APP-GFP plasmid was purchased from Addgene (#69924). HSCR associated mutations were introduced into *BACE2* expression construct using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) 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 the exons 1 of *BACE2* gene or exons 3 of *APP* gene locus were created according the gRNA cloning protocol as described above. Four million of hiPSC were transfected with gRNA constructs and a GFP-fused Cas9^{D10A} nickase expression plasmid by electroporation using Nuclofector transfection kit (Lonza, VPH-5022). After transfection, the 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 plate by FACS to get single cell. Single colony was formed around 7-14 days and then the single colony was passaged twice using ReLeSR^{TM} (StemCell Technologies, 05872) according to the manufacturer's protocol. Subsequently, the genomic DNA was isolated and the targeted region of *BACE2* or *APP* gene was PCR amplified using primers listed in Supplementary Table 1. The PCR products were directly sequenced after treated with exonuclease I and shrimp alkaline phosphatase. Colonies carrying bi-allelic nonsense mutations were expanded and subjected to functional studies.

To generate *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 (ssODN) were cotransfected into the control hiPSC (UE020302)[22]. Mutant clones were selected and validated using Sanger sequencing as described above.

Neural crest induction

Control or mutant hiPSCs were seeded on matrigel-coated plate (10^5 cells cm^{-2}) in iPS cell medium containing 10 ng ml^{-1} FGF2 (PeproTech, 100-18B) and 10 μM Y-27632, at this stage, the cells were marked as day 0. Differentiation was initiated by replacing iPS cell medium with KSR medium, containing knockout 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 nM, Stemgent) and SB431542 (10 μM, Tocris). Following the differentiation, the KSR medium was then gradually changed to N2 medium at day 4 by increasing N2 from 25% to 75% from day 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 the enteric neural crest cell 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 μM CHIR99021 (from day 2 to day 10, Tocris Bioscience, 4423), and 1 μM 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% (w/v) Saponin solution, then washed and blocked in PBS with 2% FBS. The cells are then stained with antibodies for 30-45 minutes on ice. Approximately 10^6 cells were stained and labeled cells were detected using a FACSAriaIII (Becton Dickinson Immunocytometry Systems, San Jose, CA, USA). Isotype-matched antibodies were used as controls. FlowJo version 8.2 (Tree Star, Inc.) 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 fluorescence activated cell sorting (FACS) (BD FACSAria III Cell Sorter). The HNK-1 and p75^{NTR} double positive cells were gated and sorted using the four-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**

Around 40 thousand FACS-enriched ENCCs were seeded as droplets on poly-ornithine/laminin/fibronectin (PO/LM/FN)-coated 24 well plate in N2 medium containing 10ng/ml FGF2, 3 μM CHIR99021 and 10 μM Y-27632. After 24 hours, N2

medium was replaced by neuronal differentiation medium: N2 medium containing BDNF (10ng ml^{-1} PeproTech, 450-01), GDNF (10 ng ml^{-1}, Peprotech, 450-10) and ascorbic acid (200 μM, Sigma, A4034-100G), NT-3 (10ng ml^{-1}, PeproTech, 450-03), NGF (10ng ml^{-1}, PeproTech, 450-01) and cAMP (1 μM, Sigma, D0260). 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 analyzing.

Migration assay

FACS-enriched ENCCs were plated on human fibronectin coated 12-well culture plates (30,000 cells cm^{-2}). After 24 hours, cells were treated with mitomycin (10μg ml^{-1}) to stop the 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 min, followed by blocking with 1% bovine serum albumin (BSA) (Thermo 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 FITC or Texas-Red secondary antibody (Molecular Probes, Invitrogen) (Supplementary Table 2) for 1 h at room temperature. Cells were then

counterstained with mounting medium with DAPI (DAKO) to detect nuclei. Cells were photographed using Carl Zeiss confocal microscope (LSM 800). Quantitative image analysis of differentiated neuronal cultures was done with ImageJ plugins. 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 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 ± SEM.

Cell culture, transfection and Immunoblotting

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% CO2, the culture medium was changed every other day. For transfection, around 1 million of cells were seeded onto 6-well plates (Nunc) 24 hours prior to transfection. GFP-tagged APP together with FLAG-tagged wild-type or mutant BACE2 were overexpressed in 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 50mM Tris-HCl, pH7.5, 100mM NaCl, 1% Triton X-100, 0.1mM EDTA, 0.5mM MgCl2, 10% glycerol, protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche). After incubation on ice for 15 minutes and the total proteins were collected by centrifuge for 10 minutes at 12000rpm, 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 (#89842, Thermo Fisher Scientific) according to the manufactor's protocol. 20μg of total protein from cell lysates was separated on 12% SDS-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 two-sided unpaired Student's t-test or one-way ANOVA using GraphPad Prism 7 (GraphPad Software). The *P*-value is indicated by asterisks in the figures (*, $P < 0.05$; **, $P < 0.01$). Differences among group of $P < 0.05$ were considered statistically significant. All experiments were replicated at least three times and data are shown as means with standard error of mean (SEM) or standard derivation (SD).

References

1. So, M.T. *et al.* RET mutational spectrum in Hirschsprung disease: evaluation of 601 Chinese patients. *PLoS One* **6**, e28986 (2011).
2. Garcia-Barcelo, M.M. *et al.* Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc Natl Acad Sci U S A* **106**, 2694-9 (2009).
3. Tang, C.S. *et al.* Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease. *Hum Mol Genet* **25**, 5265-5275 (2016).
4. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
5. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
6. Li, M. *et al.* Robust and rapid algorithms facilitate large-scale whole genome sequencing downstream analysis in an integrative framework. *Nucleic Acids Res* (2017).
7. Samocha, K.E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944-50 (2014).
8. Abyzov, A., Urban, A.E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* **21**, 974-84 (2011).
9. Liang, Y. *et al.* Seeksv: an accurate tool for somatic structural variation and virus integration detection. *Bioinformatics* **33**, 184-191 (2017).
10. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333-i339 (2012).
11. Layer, R.M., Chiang, C., Quinlan, A.R. & Hall, I.M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* **15**, R84 (2014).
12. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 1-34 (2014).
13. MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L. & Scherer, S.W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res* **42**, D986-92 (2014).
14. Firth, H.V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* **84**, 524-33 (2009).
15. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* **83**, 311-21 (2008).
16. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-6 (2016).
17. Delaneau, O., Howie, B., Cox, A.J., Zagury, J.F. & Marchini, J. Haplotype estimation using sequencing reads. *Am J Hum Genet* **93**, 687-96 (2013).
18. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).
19. Coetzee, S.G., Coetzee, G.A. & Hazelett, D.J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847-9 (2015).
20. Xue, Y. *et al.* Generating a non-integrating human induced pluripotent stem cell bank from urine-derived cells. *PLoS One* **8**, e70573 (2013).

21. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6 (2013).
22. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125-9 (2016).
23. Lee, G., Chambers, S.M., Tomishima, M.J. & Studer, L. Derivation of neural crest cells from human pluripotent stem cells. *Nat Protoc* **5**, 688-701 (2010).
24. Fattahi, F. *et al.* Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. *Nature* **531**, 105-9 (2016).
25. Lee, G. *et al.* Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nat Biotechnol* **25**, 1468-75 (2007).
26. Zeltner, N., Lafaille, F.G., Fattahi, F. & Studer, L. Feeder-free derivation of neural crest progenitor cells from human pluripotent stem cells. *J Vis Exp* (2014).