Genetics of Hirschsprung disease
Schriemer, Duco

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
DE NOVO MUTATIONS IN HIRSCHSPRUNG PATIENTS LINK CENTRAL NERVOUS SYSTEM GENES TO THE DEVELOPMENT OF THE ENTERIC NERVOUS SYSTEM

Hongsheng Gui1,2*, Duco Schriemer3*, William W.C. Cheng1,4*, Rajendra K. Chauhan4, Guillermo Antiñolo5,6, Courtney Berrios7, Marta Bleda6,8, Alice S. Brooks4, Rutger W.W. Brouwer9, Alan J. Burns4,10, Stacey S. Cherny2, Joaquin Dopazo5,6, Bart J.L. Eggen3, Paola Griseri11, Binta Jalloh12, Thuy-Linh Le13,14, Vincent C.H. Lui1, Berta Luzón-Toro5,6, Ivana Materia11, Elly S.W. Ngan1, Anna Pelet13,14, Macarena Ruiz-Ferrer5,6, Pak C. Sham2, Iain T. Shepherd12, Man-Ting So1, Yunia Sribudiani4,15, Clara S.M. Tang1, Mirjam C.G.N. van den Hout9, Wilfred F.J. van Ijcken9, Joke B.G.M. Verheij16, Jeanne Amiel13,14, Salud Borrego5,6, Isabella Ceccherini11, Aravinda Chakravarti7, Stanislas Lyonnet13,14, Paul K.H. Tam1, Maria-Mercè Garcia-Barceló1# & Robert M.W. Hofstra4,10#

1 Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, SAR, China
2 Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, SAR, China
3 Department of Neuroscience, section Medical Physiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
4 Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands
5 Department of Genetics, Reproduction and Fetal Medicine, Institute of Biomedicine of Seville (IBIS), University Hospital Virgen del Rocío/CSIC/University of Seville, Seville, Spain
6 Centre for Biomedical Network Research on Rare Diseases (CIBERER), Seville, Spain
7 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, USA
8 Department of Medicine, School of Clinical Medicine, University of Cambridge, Addenbrooke’s Hospital, Cambridge, United Kingdom
9 Erasmus Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands.
10 Stem Cells and Regenerative Medicine, Birth Defects Research Centre, UCL Institute of Child Health, London, UK
11 UOC Genetica Medica, Istituto Gaslini, Genova, Italy
12 Department of Biology, Emory University, Atlanta, USA
13 Département de Génétique, Faculté de Médecine, Université Paris Descartes, Paris, France
14 INSERM U-781, AP-HP Hôpital Necker-Enfants Malades, Paris, France
15 Department of Biochemistry and Molecular Biology, Faculty of Medicine, Universitas Padjadjaran, Bandung, Indonesia
16 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

*Equally contributing authors; #Equally contributing corresponding authors.

Submitted
ABSTRACT

Hirschsprung disease (HSCR), the most common form of congenital bowel obstruction, results from a failure of enteric nervous system (ENS) progenitors to migrate, proliferate, differentiate or survive to and within the gastrointestinal tract, resulting in aganglionosis in the distal colon. The HSCR genes identified to date are known to be involved in ENS development. Therefore, the search for genes solving the missing heritability in HSCR has focused on ENS-related pathways. A de novo mutation (DNM) screening in 24 HSCR patients revealed 20 DNMs in 20 genes besides 8 DNMs in the known HSCR gene RET. Knockdown of genes carrying missense and loss of function DNMs identified 4 genes indispensable for ENS development in zebrafish. Moreover, these 4 genes, which are expressed in the gut or ENS progenitors, are also involved in central nervous system (CNS) development. These newly identified HSCR genes indicate that CNS-associated genes also play a major role in ENS development.

**Keywords:** De novo mutations, Hirschsprung disease, neural crest, ENS, CNS
INTRODUCTION

Hirschsprung disease (HSCR) is the most common form of congenital obstruction of the bowel, with an incidence of ~1 per 5,000 live births. However, the incidence varies significantly between ethnic groups with the highest incidence reported in the Asian population, with 2.8 per 10,000 live births\(^1\). HSCR results from a failure of the neural crest cells, that give rise to the enteric nervous system (ENS), to migrate, proliferate, differentiate or survive in the bowel wall, resulting in aganglionosis of the distal part of the gastrointestinal tract. This results in clinically severe and sometimes life-threatening bowel obstruction. As HSCR is a highly heritable disorder, genetic variation (mutations) in the genomes of these patients must largely explain disease development. The mode of inheritance of HSCR can be recessive mostly in syndromic cases, or dominant with incomplete penetrance in non-syndromic HSCR families, to oligogenic/polygenic in sporadic cases\(^3\). So far >15 HSCR susceptibility genes have been found as are 6 linkage regions\(^1\) and three associated loci\(^2,4\). The genes identified belong to a limited number of pathways, which have been shown to be relevant to the development of the ENS, of which the RET pathway and the endothelin pathway are the most important ones. However, the identified genes and variants in these genes explain no more than 25% of the overall genetic risk\(^2,4\). Thus, the vast majority of cases cannot yet be explained by the identified HSCR-associated variants. These findings indicate that the majority of the disease risk must be due to as yet unidentified rare or common variants in the known HSCR genes or, more likely, variants in yet unknown genes, acting alone or in combination.

Exome sequencing followed by selection of genes that can be functionally linked to the pathways already known to be involved in the disease is the current approach in the field of human genetics. Variants in genes totally unlinked to the known genes or pathways are largely neglected. This study aimed to determine the contribution of rare exonic, non-synonymous de novo mutations (DNMs) to HSCR without any a priori selection. Therefore, not only did we perform ‘standard’ exome sequencing analyses, followed by burden tests and in silico prediction, but we also carried out an unbiased in vivo analysis of the mutated genes in a zebrafish model.
METHODS

Study samples

Trios
A total of 24 trios (affected child and unaffected parents) without family history of HSCR recruited in 5 different centers were included for Whole Exome Sequencing (WES). The patients were all non-syndromic. Five trios were of Chinese origin whereas 19 were of Caucasian ancestry. We prioritized the most/more severe and rarer HSCR cases for this study, namely female patients with long segment or total colonic aganglionosis. Sixteen out of the 24 patients had previously tested negative for RET damaging variants by traditional technologies. Characteristics of the patients are presented in Supplementary Table 1. Informed consent was obtained from all participants.

Case-control
WES data from 28 additional sporadic HSCR patients without sub-phenotype limitation (singletons) and 212 controls were used to check gene recurrence and assess the gene burden for rare variants (Supplementary Table 1).

Data generation

Whole exome sequencing
DNA samples were sequenced in four centers. The exome-capture kit and sequence platforms used per center are detailed in Supplementary Table 2. Appropriate mapping tools (Burrows-Wheeler aligner–BWA- for Illumina data and Bfast for Solid data) were used to align sequence reads to the human reference genome (build 19)\(^5\). Sequence quality was re-evaluated using the FastQC toolbox, Picard’s metric summary and the GATK Depth-of-Coverage module. After initial quality control (QC) all eligible sequences were pre-processed for local indel realignment, PCR duplicate removal and base quality recalibration\(^6\).

Genome-wide SNP array
To determine copy number variants (CNVs) and regions of homozygosity, DNA was hybridized to the HumanCyto SNP12 BeadChip (Illumina, San Diego, CA, USA) according to standard protocols.
**Variant calling and prioritization**

Aligned reads from all sequenced samples were pre-processed according to standard guidelines. Variant calling was done independently for Illumina reads or Solid reads using the Genome Analysis Toolkit (GATK) unified Genotyper 2.0. To avoid mismatched regions across different capture kits, calling was performed on whole genome wide without limiting on any capture array. Special setting (allow potentially miscoded quality scores) was used to make color-spaced solid reads compatible to the program (Broad institute). Raw variants (including single nucleotide variants and short insertions/deletions) with individual genotypes and their affiliated quality scores were stored in a standard VCF format after calling. Quality assessment (QA) and QC were then adopted on a few set of variants (raw variants, exonic variants, rare variants) to generate a confident variant set for downstream prioritization (Supplementary Note).

Clean variant set at exonic regions was produced after variant-level and genotype-level quality control. Rare coding sequence variants were then prioritized by filtering out those variants with minor allele frequency >0.01 in any of these public databases (dbSNP137, 1000 Human Genome project and NHLBI Exome Sequencing project). An automatic pipeline integrating GATK, KGGSeq, Annovar and Plink was used to generate final set of qualified variants (Supplementary Figure 1).

**Identification of DNM**

**WES DNM detection**

Rare, exonic variants present in the probands but absent in both parents were considered DNM. To select putative DNM (or de novo variations) the following criteria were used: 1) minimal coverage of 5 in patients and parents; 2) a minimal genotype quality score of 10 for both patients and parents; 3) at least 10% of the reads showed the alternative allele in patients; and 4) not more than 10% of the reads showed the alternative allele in parents. Subsequently all remaining DNM variants were manually inspected using the Integrated Genome Viewer (IGV) and classified into 5 different confidence ranks according to their base-calling quality and strand bias. The first two ranks of DNM candidates were selected for validation by Sanger sequencing; while the other three classes of candidates were re-evaluated by a model trained from variants submitted for Sanger sequencing (Supplementary Note).
RET gene inspection
To guarantee that no de novo mutations had been missed in the major HSCR gene, the depth of coverage of each of the 21 exons of RET was manually inspected for each patient. All exons with a coverage <10 were Sanger sequenced. Mutation Detector software (Thermo Fisher Scientific) was used to identify rare coding sequencing mutations from raw Sanger sequences; any mutation found in trio proband was further checked in his/her parents. Besides rare mutations, bi-allelic genotypes for the common risk single nucleotide polymorphisms (IVS1+9494, rs2435357T) were extracted from local databases or newly genotyped.

Copy number variation detection
The Nexus® software program (Biodiscovery, El Segundo, CA, USA) was used to normalize and analyse the SNP array data as mentioned above. Loss is defined as the loss of a minimum of 5 probes in a 150kb region, with a minimum Log R ratio – 0.2. Gain is defined as the gain of a minimum of 7 probes in 200kb region, with minimum Log R ratio 0.15. The minimum length of regions of homozygosity analysed was 2Mb. The identified CNVs were reviewed for pathogenicity using the genome browser UCSC (http://genome.ucsc.edu), the DGV database (http://dgv.tcag.ca/dgv/app/home), the Decipher database (https://decipher.sanger.ac.uk/) and our in-house local reference data base that consists of 250 healthy controls and 250 individuals of the general population.

Statistical tests
De novo mutation rate
All proven DNMs were classified into loss-of-function (nonsense Single Nucleotide Variants (SNVs), frame-shift indels and splicing sites), missense SNVs, in-frame indels and synonymous SNVs. The counts of DNM per trio were fitted to Poisson distribution with lamda as observed mean. De novo mutation rates were calculated for these DNM subtypes and compared to 677 published healthy trios and neurodevelopmental disease trios using a binomial test. Given per-gene mutation rate in Samocha et al. paper, statistical over-representation of mutations in all 24 genes were calculated using Fisher's exact test.

Gene-wide burden analysis
Genes with DNM were further scrutinized for the presence of inherited rare damaging variants in the trios as well as in HSCR singletons for whom WES data
were available. A detailed analytical protocol was shared before running association in each centre. Briefly, genotypes of rare damaging variants (as previously defined) in genes carrying ≥1 de novo mutation were extracted from raw sequencing reads. CMC test in Rvtest package was used to collapse multiple variants into the same gene (boundary defined using hg19 refgene) and compare overall burden between cases and local matched controls. *P*-values were estimated by asymptotic chi-square distribution. Gene-wise *p*-value, burden direction and variant count per gene were exported. Ultimately sample-size weighted Z-score method was used to conduct meta-analysis on gene-wise summary statistics from three centres using the same protocol.

**Bioinformatics analysis**

**Variant-level implication**

The impact of each DNM to its carrying gene was predicted using several of bioinformatics tools or databases. The conservation of missense SNVs was predicted using GERP and PhyloP across 29 different species. The deleteriousness of missense or nonsense SNVs were determined by a logit model incorporating 5 prediction programs (Polyphen2, Sift, MutationTaster, PhyloP and Likelihood ratio). Human Splicing finder was used to predict whether DNMs causing synonymous change or locating at splicing sites (exon +/- 2bp) created or disrupted splice sites. To further implicate the possible role of synonymous DNMs on transcription, RNAmute was used to predicted the RNA substructure change due to corresponding site mutation. Finally, ClinVar and PubMed were searched for the same or similar mutations in the same gene that present in healthy controls or other disease patients.

**Gene-level implication**

The evidence of gene-level implication was collected from two aspects. On one side, those 24 genes carrying DNMs were searched against databases (ATGU's Server) for other disease patients or healthy samples. On the other side, ENS candidate genes/gene-sets (Supplementary Table 8; Supplementary Note) were linked to newly identified genes using pathway or PPI network information. Disease Association Protein-Protein Link Evaluator (DAPPLE) was used to test whether the genes carrying DNM in our study are functionally connected to each other. The significance of observed pathway enrichment and network connectivity was evaluated empirically using randomly selected genes, genes having the same
genomic size as the identified DNM genes. InWeb and Ingenuity Pathway Analysis were used to detect direct and indirect protein interactions between ENS-related genes and genes with DNMs.

**Gene expression in ENS**

In order to test the involvement of the newly identified genes in enteric nervous system development, in house expression data was shared from other in-parallel projects in Hong Kong, Rotterdam centre. The first expression dataset was from RNA sequencing on an iPSC-induced enteric neural crest cell (ENCC) for a HSCR patient; the second and third expression dataset was from microarray chips on embryonic mouse gut and ENCC.

**Zebrafish**

*Tg(-8.3bphox2b:Kaede)* transgenic zebrafish (*Danio rerio*) embryos were obtained from natural spawning. Maintenance of zebrafish and culture of embryos were carried out as described previously. Embryos were staged by days post-fertilization (dpf) at 28.5°C.

*Gene knockdown by antisense morpholino*

Antisense morpholinos (MO) (Gene Tools LLC) targeting the zebrafish orthologues of the candidate genes, by blocking either translation or splicing, were microinjected to 1 to 4-cell stage *Tg(-8.3bphox2b:Kaede)* transgenic zebrafish embryos as previously described. For candidate genes that are duplicated in the zebrafish genome, morpholinos targeting all paralogs were co-injected. Standard control morpholino and 5-nucleotide mismatch control morpholino for *ckap2l, dennd3a, dennd3b, ncl1, nup98* and *tbata* were used as negative control. Embryos were raised to 5 dpf, analysed and imaged under a stereo fluorescence microscope (Leica MZ16FA and DFC300FX). An HSCR-like phenotype was defined as the absence of enteric neurons in the distal intestine in 5 dpf embryos. Sequences and dosages of all morpholinos used are listed in Supplementary table 9.

*Expression analysis*

To confirm the target gene were successfully knockdown, total RNA were extracted from 1 dpf embryos (n=50) injected with the splice blocking morpholino using RNA Bee (Amsbio) and cDNA were reverse transcribed using iScript cDNA Synthesis Kit (Bio-rad). qPCR was performed using KAPA Sybr® Fast qPCR Kit
**RESULTS**

**Identification of de novo mutations**

We performed whole-exome sequencing (WES) on 24 trios composed of a sporadic non-syndromic HSCR patient and the unaffected parents (72 individuals; Supplementary Table 1) and focused on *de novo* variants. Sporadic female cases with a long segment (LS) HSCR were overrepresented as the load of *de novo* rare coding variant is presumed to be the highest in this group. The depth coverage of the targeted sequences ranged from 18X to 74X (average 46X), and the targeted exome covered by at least 10 sequence reads ranged from 65% to 98% (average 88%). Sequencing metrics after standard analytical pipeline (Supplementary Figure 1) were in normal ranges (Supplementary Note; see Supplementary Table 2 and Supplementary Figure 2 for detail).

All *de novo* variations were carefully selected, validated and/or statistically predicted (Methods and Supplementary Note; see prediction result in Supplementary Table 3). After Sanger sequencing validation, a total of 28 DNMs in 14 patients were identified (Table 1). The overall DNM rate per individual was 1.2 per exome per generation (Poisson distribution with $\lambda=1.2$; Kolmogorov-Smirnov test, $p=0.893$; Supplementary Figure 3) which is in accordance with the expected mutation rate in the general population. Several studies have shown that the DNM rates are similar between patients and healthy controls, but found that patients have a significantly higher fraction of loss of function (LOF) DNMs$^{8,9}$. Indeed, in our
Table 1. *De novo* mutations in Hirschsprung disease probands

<table>
<thead>
<tr>
<th>Trio</th>
<th>Pheno-type</th>
<th>Gene</th>
<th>De novo mutation</th>
<th>Type</th>
<th>Prediction deleteriousness*</th>
<th>MAF (dbSNP137/ESP6500)%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L, F</td>
<td>RET</td>
<td>3splicing9+1</td>
<td>splicing</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c.474C&gt;T: p.L158L</td>
<td>synonymous</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>2</td>
<td>L, F</td>
<td>RET</td>
<td>c.2511_2519delCCCTGG ACC:p.S837fs</td>
<td>frameshift</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>COL6A3 c.3327C&gt;T: p.H1109H</td>
<td>synonymous</td>
<td>-</td>
<td>0.00042 (rs114845780) / N</td>
</tr>
<tr>
<td>3</td>
<td>L, F</td>
<td>RET</td>
<td>c.1818_1819insGGCAC: p.Y606fs</td>
<td>frameshift</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>4</td>
<td>L, F</td>
<td>DAB2IP</td>
<td>c.2339C&gt;T: p.T780M*</td>
<td>missense</td>
<td>No</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ISG20L</td>
<td>c.961G&gt;A: p.G321R</td>
<td>missense</td>
<td>Yes</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MED26</td>
<td>c.675C&gt;T: p.A225A</td>
<td>synonymous</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NCLN</td>
<td>c.496C&gt;T: p.Q166X*</td>
<td>nonsense</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NUP98</td>
<td>c.5207A&gt;G: p.N1736S</td>
<td>missense</td>
<td>Yes</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vezf1</td>
<td>c.584C&gt;T: p.S195F</td>
<td>missense</td>
<td>Yes</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZNF57</td>
<td>c.570C&gt;T: p.A190D</td>
<td>synonymous</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>5</td>
<td>L, F</td>
<td>RET</td>
<td>c.1761delG: p.G588fs</td>
<td>frameshift</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCUBE3</td>
<td>c.1493A&gt;T: p.N498I</td>
<td>missense</td>
<td>No</td>
<td>N / N</td>
</tr>
<tr>
<td>6</td>
<td>L, M</td>
<td>AFF3</td>
<td>c.1975G&gt;C: p.V659L</td>
<td>missense</td>
<td>No</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLEKHG5</td>
<td>c.2628G&gt;T: p.T876T</td>
<td>synonymous</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>7</td>
<td>L, M</td>
<td>KDM4A</td>
<td>c.26A&gt;G: p.N9S</td>
<td>missense</td>
<td>No</td>
<td>N / N</td>
</tr>
<tr>
<td>8</td>
<td>L, M</td>
<td>MAP4</td>
<td>c.3351C&gt;T: p.W1117G</td>
<td>synonymous</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>9</td>
<td>L, M</td>
<td>RET</td>
<td>c.1858T&gt;C: p.G620R</td>
<td>missense</td>
<td>Yes</td>
<td>0 (rs7316810) / N</td>
</tr>
<tr>
<td>10</td>
<td>TCA, M</td>
<td>CKAP2L</td>
<td>c.555_556delAA: p.E186fs</td>
<td>frameshift</td>
<td>-</td>
<td>N / 0.00002</td>
</tr>
<tr>
<td>11</td>
<td>L, F</td>
<td>RET</td>
<td>c.4097T&gt;G: p.C1377G</td>
<td>missense</td>
<td>Yes</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMCN1</td>
<td>c.10326G&gt;A: p.A3456T</td>
<td>missense</td>
<td>No</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TUBG1</td>
<td>c.699T&gt;C: p.S233S</td>
<td>synonymous</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>12</td>
<td>L, F</td>
<td>CCR2</td>
<td>c.848T&gt;A: p.L283Q</td>
<td>missense</td>
<td>Yes</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DENND3</td>
<td>c.1921delT: p.K640fs</td>
<td>frameshift</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>13</td>
<td>L, F</td>
<td>RET</td>
<td>c.1710C&gt;T: p.C570X</td>
<td>nonsense</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td>14</td>
<td>L, F</td>
<td>RET</td>
<td>c.526_528delGCA: p.R175del</td>
<td>non-frameshift</td>
<td>-</td>
<td>N / N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TBATA</td>
<td>c.157C&gt;T: p.R53C</td>
<td>missense</td>
<td>No</td>
<td>N / N</td>
</tr>
</tbody>
</table>

F: Female; M: Male; L: Long-segment HSCR; TCA: Total Colonic Aganglionosis; *: Disease-causal prediction by KGGSeq[^57], a software that uses a weighted logistic regression to combine multiple prediction scores; #mosaic mutation; Dark grey: *de novo* RET mutations; Light grey: genes giving a HSCR-like phenotype in zebrafish; %: minor allele frequency in dbSNP137 or ESP database, with ‘N’ standing for no data available.
HSCR patient cohort, the rate of loss of function DNMs (LOF; N=8, including nonsense, frameshift and splice site changes) is significantly higher than that of healthy trios \((p=0.011)\) or unaffected siblings of neuropsychiatric patients \((p=0.001)\) from multiple published studies\(^8,^{10-12,22}\) (Supplementary Table 4). The 28 DNMs were localised in 21 genes. 8 DNMs were found in \(RET\), the major HSCR gene\(^23\). Among the DNMs in \(RET\) was the Cys620Arg variant, known to cause both HSCR and Multiple Endocrine Neoplasia type 2A\(^24\). In this study, the observed rate for \(RET\) DNMs (0.33 per trio) was significantly higher (binomial test, \(p<2*10^{-16}\)) than that modelled for \(RET\) DNMs in the general population (0.000133 per trio) according to Samocha et al.\(^14\).

One of the patients analysed carried a total of 7 DNMs, two of which (in \(NCLN\) and \(DAB2IP\)) were mosaic mutations (Supplementary Figure 4). This finding is in line with a recent report stating that 6.5% of all DNMs are in fact mosaic and occur post-zygotic\(^25\). Within the 24 patients we looked for inherited rare damaging variants in the 21 genes that carried DNMs (Supplementary Note; Methods). Inherited damaging mutations were found in \(RET\), \(HMCN1\), \(PLEKHG5\), \(MAP4\), \(SCUBE3\), and \(KDM4A\) (Supplementary Table 5). Neither \textit{de novo} nor inherited copy number variants (CNVs) were detected in any of the trios.

### Mutation profile of HSCR patients

In general, disease-associated common variants confer a liability to disease to the individuals of the general population. These common variants, in combination with environmental and/or rare variants finally result in manifestation of the disease. Thus, since both rare and common variants jointly contribute to HSCR we carefully examined the genetic profile of our patients to assess the genetic background on which the DNMs reside. Each patient was investigated for the presence or absence of the common HSCR-associated \(RET\) allele (IVS1+9494, rs2435357T)\(^26-29\) as well as for the presence of rare variants (inherited from unaffected parents) in a set of 116 pre-selected genes known to be involved in ENS development (Supplementary Notes; Supplementary Tables 3 and 6).

The mutation profile for all patients is shown in Supplementary Table 5. We observe that 29% of the patients with \(\geq 1\) DNM and 60% of the patients without any DNM carry the common \(RET\) risk genotype TT (rs2435357T). Moreover, patients with DNM carry on average 1.4 inherited rare damaging variants in ENS genes, compared to an average of 2.4 in patients without any DNM. Notably, six out of the 14 patients carried DNMs without co-occurrence of
a *RET* coding sequence mutation. Although the differences are not statistically significant, these observations suggest that the new genes identified may, independently of the genetic background, play a role in the pathology of the disorder, and prompted us to further investigate those genes using *in silico* and *in vivo* approaches.

**Determining pathogenicity of the DNMs *in silico***

The recurrence of a mutation or the identification of a recurrently mutated gene in an independent group of patients or unrelated controls can provide corroborating evidence of pathogenicity or neutrality\(^{30}\). Therefore, all the genes in which we identified DNMs were checked against public databases (ATGU's Gene-Mutation-Constraint Server) for DNM recurrence. Only one missense DNM (different from that identified in this study) in *MAP4* was found in a patient with autism spectrum disorder (ASD). A few genes (*SCUBE3, RBM25 and TUBG1*; Table 2) were identified evolutionary constrained genes in which functional variants are more likely to be deleterious\(^{14}\).

To establish whether genes with DNMs carry significantly more rare variants in HSCR patients than in controls, we used the WES data from the 20 eligible HSCR trio-probands, 28 additional HSCR patients and 212 control individuals to calculate the variation burden per gene (Methods). Nine of the twenty-one genes (*RET, KDM4A, HMCN1, MAP4, NUP98, AFF3, COL6A3, CCR2, and CKAP2L*) were found recurrently mutated in multiple HSCR patients with different rare damaging mutation sites (Supplementary Table 7). Meta-analysis of our gene burden tests showed that *RET* and *CKAP2L* were enriched for rare damaging variants in the HSCR patients (nominal \(p<0.05\); Table 2 and Supplementary Table 7). However, cross-checking of these 21 genes in another in-parallel HSCR exome study (190 cases and 740 controls) revealed only *RET* was significantly overrepresented with deleterious variants (\(p < 0.001\); manuscript in preparation, A. Chakravarti).

The possible impact of DNMs on gene function was explored using bioinformatic prediction tools (Methods). Besides the 8 LOF mutations, 6 out of twelve missense mutations were consistently predicted deleterious (Table 1). As for the seven synonymous DNMs, we found no *in silico* evidence indicating that those changes interfered with splicing and/or significantly changed the RNA structure (Supplementary Table 8).
We next checked whether the genes with DNMs are functionally related to each other and/or to the signalling networks known to govern ENS development. ISG20L2 and MAP4 showed more indirect interactions with other genes carrying DNMs than expected by chance ($p=0.0063$ and $p=0.0167$ respectively) as predicted by DAPPLE, though no direct in silico interactions were found among those 21 genes. A list of 116 known ENS related genes (Supplementary Table 6) was used to study the functional link between genes with DNMs (other than RET) and the ENS. Only a single interaction was identified in the InWeb protein interaction catalogue (COL6A3 interacts with ITGB1). Using Ingenuity Pathway Analysis, we identified additional direct and indirect relationships with ENS-related genes for MAP4, COL6A3, RBM25 and TUBG1 (Supplementary Figure 5). All genes carrying DNMs were either expressed in human iPSC-derived enteric neuron precursors or in primary murine enteric neuron precursors (Table 2).

Table 2. Genes carrying de novo mutations

<table>
<thead>
<tr>
<th>Gene</th>
<th>Number of amino acids</th>
<th>Co-occurrence with RET DNM</th>
<th>Burden test meta-analyses (p-value)</th>
<th>Zebrafish ENS phenotype</th>
<th>Gut expression (human; mouse; zebrafish)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLEKHG5</td>
<td>1062</td>
<td>No</td>
<td>0.3997</td>
<td>NT</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>KDM4A</td>
<td>1064</td>
<td>No</td>
<td>0.1190</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>ISG20L2</td>
<td>353</td>
<td>No</td>
<td>0.4949</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>HMCN1</td>
<td>5635</td>
<td>Yes</td>
<td>0.9789</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>AFF3</td>
<td>1226</td>
<td>No</td>
<td>0.4745</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>CKAP2L</td>
<td>745</td>
<td>No</td>
<td>0.0178</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>COL6A3</td>
<td>3177</td>
<td>Yes</td>
<td>0.6398</td>
<td>NT</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>CCR2</td>
<td>374</td>
<td>No</td>
<td>0.4745</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>MAP4</td>
<td>1152</td>
<td>No</td>
<td>0.4851</td>
<td>NT</td>
<td>Yes; No; -</td>
</tr>
<tr>
<td>SCUBE3*</td>
<td>993</td>
<td>Yes</td>
<td>0.7133</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>DENND3</td>
<td>1198</td>
<td>No</td>
<td>0.5977</td>
<td>Yes</td>
<td>Yes; Yes; Yes</td>
</tr>
<tr>
<td>DAB2IP</td>
<td>1189</td>
<td>No</td>
<td>0.9819</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>RET</td>
<td>1114</td>
<td>-</td>
<td>0.0078</td>
<td>Yes</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>TBATA</td>
<td>351</td>
<td>Yes</td>
<td>0.8028</td>
<td>Yes</td>
<td>No; Yes; Yes</td>
</tr>
<tr>
<td>NUP98</td>
<td>1817</td>
<td>No</td>
<td>0.7243</td>
<td>Yes</td>
<td>Yes; Yes; Yes</td>
</tr>
<tr>
<td>RBM25*</td>
<td>843</td>
<td>Yes</td>
<td>0.0846</td>
<td>NT</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>TUBG1*</td>
<td>451</td>
<td>Yes</td>
<td>1.0000</td>
<td>NT</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>VEZF1</td>
<td>521</td>
<td>No</td>
<td>0.6717</td>
<td>No</td>
<td>Yes; Yes; -</td>
</tr>
<tr>
<td>ZNF57</td>
<td>555</td>
<td>No</td>
<td>0.3808</td>
<td>NT</td>
<td>Yes; No; -</td>
</tr>
<tr>
<td>NCLN</td>
<td>563</td>
<td>No</td>
<td>1.0000</td>
<td>Yes</td>
<td>Yes; Yes; Yes</td>
</tr>
<tr>
<td>MED26</td>
<td>600</td>
<td>No</td>
<td>1.0000</td>
<td>NT</td>
<td>Yes; Yes; -</td>
</tr>
</tbody>
</table>

*genes evolutionary constrained as per Samocha et al. 2014; NT: not tested (gene carries synonymous mutation and/or has no orthologue in zebrafish); *data from in-house hiPSC-derived neural crest, mouse expression data, and RT-PCR in zebrafish (test only for 4 novel genes).
Determination of pathogenicity of the DNMs in vivo

As no proof of functional effects for any of the synonymous DNMs was found, we further focused on the 13 genes (other than RET) that have a LOF or missense mutation. Because none of these 13 genes were obvious candidates for HSCR, we used the zebrafish model system to further investigate the function of these genes in ENS development. Previous studies have shown that morpholino-mediated knockdown of orthologues of known HSCR genes results in an HSCR-like phenotype in zebrafish\(^4,31-35\). Except CCR2, all 13 genes with nonsynonymous DNMs have zebrafish orthologues. Splice-blocking morpholinos (SBMOs) were designed to knock down the orthologues for these 12 genes (Methods). The SBMOs were injected into Tg(-8.3bphox2b:Kaede) transgenic zebrafish\(^19\) embryos that express the fluorescent protein Kaede in enteric neuron precursors and differentiated enteric neurons. Initially, knockdown of 5 orthologues (ckap2l, dennd3a and dennd3b, ncl1, nup98 and tbata) resulted in a HSCR-like phenotype as enteric neurons were absent in the distal intestine of 5 dpf embryos, while embryos injected with 5-nucleotide mismatch control morpholinos had normal ENS development with enteric neurons present along the entire length of intestine. We then coinjected the SBMOs with p53 morpholinos to verify the phenotype did not result from non-specific, p53-induced apoptosis. Co-injection of p53 morpholino with dennd3a and dennd3b, ncl1, nup98 or tbata SBMOs resulted in the same phenotype (Figure 1), indicating the phenotype was not caused by non-specific apoptosis. On the contrary, the phenotype could not be reproduced in ckap2l SBMO and p53 morpholino co-injection (Figure 1). To further demonstrate the absence of enteric neuron was specific to the knockdown of the orthologues, we repeated the experiment by injecting translation-blocking morpholinos (TBMOs) against dennd3a, dennd3b, ncl1, nup98 and tbata and the phenotype was reproduced (data not shown). Therefore we concluded that knockdown of the DENND3, NCLN, NUP98 and TBATA orthologues disrupted ENS development and caused a HSCR-like phenotype in vivo.

To confirm the SBMOs knockdown effect, qPCR was performed to compare the expression of the target genes between SBMO-injected and control morpholino-injected embryos. Expression of dennd3a, dennd3b, nup98 and tbata was markedly reduced in the SBMO-injected embryos (Supplementary Figure 6). Intriguingly, there was no significant reduction in ncl1 expression in the ncl1 SBMO injected embryos.
Therefore we further investigated it by performing RT-PCR on individual embryos and found that there was a large variation in *ncl1* expression between embryos injected with the SBMO, with some of them showing a clear reduction in *ncl1* transcript level (Supplementary Figure 7). Of the zebrafish orthologues that did not show a specific HSCR-like phenotype after SBMOs injection, all demonstrated significant reductions in expressions except for *aff3*, *scube3* and *vezf1a* (Supplementary Figure 6).

In addition we performed RT-PCR and whole mount in situ hybridization (WISH) experiments to determine if the gene expression patterns of the zebrafish orthologues were consistent with a predicted role in ENS development. Temporal analysis using RT-PCR revealed that zebrafish orthologues of *DENND3*, *NCLN* and *NUP98* were maternally and zygotically expressed from 0-120hpf while the *TBATA* orthologue is only zygotically expressed from 24-120hpf (Supplementary Fig 8).

**Figure 1. Pathogenicity analysis in vivo by morpholino gene knockdown in zebrafish.** Knockdown of *nc11*, *dennd3*, *nup98* and *tbata* resulted in HSCR-like phenotype that kaede-expressing enteric neurons were absent in the distal intestine at 5 dpf and the results were reproduced in the presence of p53 morpholino. Aganglionosis observed in *ckap2l* knockdown was caused by non-specific apoptosis as the result was not reproducible in p53 morpholino co-injection. Number of embryos with phenotype out of total number of embryos observed is shown. Dotted lines outline the intestines. Asterisks indicate the positions of anus. Arrows indicate the position where the aganglionic region begins.
WISH analysis showed that the orthologues for all 4 genes were expressed in distinct spatial locations specifically in the intestine and the anterior CNS from 24-96hpf (Figure 2).

DISCUSSION

Over the last years a large number of papers have been published on de novo mutation screening in human diseases. This has resulted in the identification of many new disease-associated genes. Genes are considered as true disease causing when at least two unlinked patients are found with a mutation in the same gene. This works well for diseases that are relatively homogeneous or for which many patients can be investigated. For the more heterogeneous rare diseases for which only small cohorts are available this poses a problem. Often possible disease causing genes are found in a single patient. How to decide whether this finding is of importance? Expression of the gene in the relevant tissues can be considered as additional evidence, as is networks analysis. However, making strong statements for private disease genes is, and will be, extremely difficult. It also results in a bias towards genes in the known disease causing gene networks. Genes not fitting the current knowledge are often discarded as uninteresting. In the current study we wanted to take this all one step further.

Therefore, we decided that the best way to obtain sound evidence for involvement of new candidate genes in HSCR should come from functional analysis. We opted for an in vivo approach using the zebrafish model system. We knocked down the expression of zebrafish orthologues of 12 of the 13 genes in which loss of function or missense DNMs were identified in a transgenic reporter zebrafish line (Tg(-8.3bphox2b:Kaede)). The orthologues of 9 of the 12 genes were successfully knockdown by morpholinos, and from which we discovered that 4 genes when functionally perturbed resulted in loss of neurons in the distal gut, as in the HSCR patients. It is noteworthy that the SBMOs targeting 3 of the orthologues (aff3, scube3 and vezf1a) did not knock down the target transcripts as expected, which highlighted the limitation of morpholinos and might lead to false-negative results\textsuperscript{36}. To bypass this limitation, other loss-of-function approaches should be considered to further study these genes, such as CRISPR/Cas9 knockout\textsuperscript{37}. Finding 4 genes that when knocked down in zebrafish give a hindgut phenotype resembling the human patients in which the DNMs were found, clearly
DE NOVO MUTATIONS IN HSCR PATIENTS LINK CNS GENES TO THE DEVELOPMENT OF THE ENS

demonstrates that genes that never would have been followed up, based on the usual gene selection criteria, should not be ignored.

Using the bioinformatics prediction and statistics, we would have focused on RET and CKAP2L only as they were significantly enriched for rare variants in the HSCR patients (nominal \( p<0.05 \); Table 2).

We wondered whether any or all of these 4 genes can be linked to the ENS or whether they play relevant roles in neuronal development or neural crest derived cell types in general. In fact by studying these genes in more depth we noticed that all 4, despite lack of obvious connection to the known ENS pathways, are involved in the development of the CNS or the neural crest, making these not as random as they might first appear.

DENN/MADD Domain Containing 3 (DENND3) is a guanine nucleotide exchange factor (GEF) that is involved in intracellular trafficking by activation of the small GTPase RAB12\(^{38}\). In zebrafish, Rab12 and other Rab GTPases are highly expressed by pre-migratory neural crest cells and their expression is dysregulated in Ovo1 morphant zebrafish that display altered migration of neural crest cells\(^{39}\). Independently of RAB12, DENND3 also regulates Akt activity, which is involved in the proliferation and survival of enteric neural crest cells\(^{38,40}\).

![Figure 2. Temporal and spatial expression patterns of zebrafish orthologues.](image)

Whole mount in situ hybridized embryos hybridized with antisense riboprobes for dennd3a, dennd3b, ncl1, nup98 and tbata at the indicated developmental stages. All columns show lateral views. Anterior CNS expression is apparent at all stages for all probes while intestinal expression for all probes is apparent from 48hpf onwards.
Nicalin (NCLN) is a key component of a protein complex that antagonizes Nodal signalling\(^{41}\). In vertebrates, Nodal signalling is involved in induction of the mesoderm and endoderm\(^{42}\). In contrast, inhibition of Nodal signalling is required for the specification of human embryonic stem cells into neuroectoderm, including the neural crest\(^{43,44}\). The antagonizing function of Nicalin on Nodal signalling is therefore consistent with the neural crest specification that is required for ENS development.

The *NUP98* gene encodes a precursor protein that is autoproteolytically cleaved to produce two proteins: NUP98 from the N-terminus and NUP96 from the C-terminus\(^{45,46}\). A missense DNM was identified in the last exon of the *NUP98* gene and therefore affects the NUP96 protein. As in humans, zebrafish Nup96 is produced by cleavage of the Nup98 precursor protein. Since morpholino’s act on mRNA level, both *nup98* and *nup96* were targeted in our zebrafish experiments. It is therefore unclear whether the observed aganglionosis is caused by loss of Nup98 or Nup96. NUP96 is one of approximately 30 proteins in the nuclear pore complex (NPC)\(^{47}\) and its expression level regulates the rate of proliferation\(^{48}\). Two other members of the NPC (*Nup133* and *Nup210*) are involved in neural differentiation in mice\(^{49,50}\). Moreover, NUP96 interacts with NUP98 and NUP98 is involved in the transcriptional regulation of the HSCR genes *SEMA3A*, *DSCAM*, *NRG1* and the *NRG1* receptor *ERBB4* in human neural progenitor cells\(^{51}\). Therefore, it is likely that loss of either NUP protein (NUP96 or NUP98) could contribute to HSCR development.

The mouse orthologue of Thymus, Brain And Testes Associated (*TBATA*) is called Spatial and is highly expressed during differentiation of several tissues\(^{52}\). These include the cerebellum, hippocampus and Purkinje cells in the brain, where *TBATA*/Spatial is expressed in early differentiating neurons\(^{53}\). In mouse hippocampal neurons, *TBATA*/Spatial is required for neurite outgrowth and dendrite patterning\(^{54}\).

The 4 newly identified candidate genes for HSCR all seem to play a role in neuronal development and could potentially be involved in HSCR (Figure 3). This also suggests a clear link between CNS and ENS development. This is not surprising as a number of studies have described the strong correlation between Down syndrome and syndromic HSCR and several known HSCR genes (e.g. *KBP*, *SOX10*, *NRG1*, *IKBKAP*, *ZEB2*, *PHOX2B*) have been reported to be involved in both CNS and ENS pathologies\(^{2,55–57}\). In humans, *SOX10* mutations cause myelin deficiencies and sensory neuropathies as well as the neurological variant of
Waardenburg-Shah syndrome which includes HSCR in the phenotypic spectrum. Likewise, NRG1 is associated with schizophrenia and Nrg1 mutations in mice cause peripheral sensory neuropathies\(^6^6\). IKBKAP mutations are associated with the Riley-Day syndrome or familial dysautonomia (FD)\(^5^8,^5^9\). Notably, some patients with FD also suffer from gastrointestinal dysfunction shortly after birth and interestingly, the co-occurrence of both FD and HSCR has been reported\(^6^0\). In addition, knockdown of ikbkap in zebrafish also generates a HSCR-like phenotype\(^3^5\). Further, KBP mutations are associated with Goldberg-Shprintzen syndrome\(^6^1\) (MIM 609460), a rare autosomal recessive inherited syndrome, where patients present with HSCR, microcephaly polymicrogyria and moderate mental retardation.

Besides the fact that several HSCR/neuromuscular genes are known to be associated with CNS defects, the opposite is also described. Many neurological and

**Figure 3. Newly identified genes in ENS development.** All symbols represent proteins coded by Hirschsprung known genes or novel genes identified in this study. The effect of gene NUP98 is shown by protein NUP96. The interaction effects between different proteins are illustrated by four different lines representing binding, secreted/express, phosphorylation and activation. ENCC, enteric neural crest cell.
psychiatric disorders are associated with constipation, and sometimes defects in the ENS are reported\textsuperscript{62}. For instance, it has recently been described that mutations in \textit{CDH8} result in a specific subtype of autism in combination with gastrointestinal problems. A \textit{cdh8}\textsuperscript{−/−} zebrafish recapitulates the human phenotype, including increased head size (expansion of the forebrain/midbrain), an impairment of gastrointestinal motility and a reduction in post-mitotic enteric neurons\textsuperscript{63}. Besides, a search of CNS and autism in Phenolyzer\textsuperscript{64} returned two genes (\textit{APP} and \textit{MECP2}) that have been implicated in ENS development\textsuperscript{65,66}.

Thus, given all of the above, and the fact that HSCR occurs with neurological disorders more often than would be expected by chance, it is not surprising that dysfunction of these newly identified neurological related genes results in dysregulation of the neural crest-derived cells that form the ENS, and hence in HSCR. These data are further corroborated by the expression patterns we observed for the orthologues of these 4 genes in zebrafish embryos (Figure 2), with all 4 having clear expression in both the brain and the gut.

Finding a niche for these genes in ENS development will help to open new avenues of research which, eventually, will enhance our knowledge about ENS development and HSCR disease mechanisms. Until now, we believed that the number of cellular processes involved in the development of HSCR was limited. Clearly this idea needs to be revisited as the novel genes we identified are not directly linked to any of the currently known HSCR gene networks. In spite of the plethora of databases and prediction tools available, very little is known about the intricate ways in which genes interact in the development of the ENS, or the function of many genes.

**URLS**

Genome analysis toolkit (GATK) (https://www.broadinstitute.org/gatk/);
ANNOVAR (http://annovar.openbioinformatics.org/en/latest/);
PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/);
KGGSeq (http://statgenpro.psychiatry.hku.hk/limx/kggseq/);
ATGU’s Server (http://atgu.mgh.harvard.edu/webtools/gene-lookup/);
DAPPLE (http://www.broadinstitute.org/mpg/dapple/dappleTMP.php);
ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/)
ACKNOWLEDGMENTS
The authors would like to thank the patients and families involved in this study. DS, AB, RC, BE, YS, RH are supported by research grants from ZonMW (TOP-subsidie 40-00812-98-10042 to BJLE/RMWH) and the Maag Lever Darm stichting (WO09-62 to RMWH); AC is supported by NIH grant R37 HD28088; HG, WC, VL, EN, PS, MS, CT, PT, MMG-B are supported by the Health and Medical Research Fund (HMRF 01121326 to VCHL; HMRF 02131866 to MMGB; and 01121476 to ESWN); General Research Fund (HKU 777612M to PKT); HKU seed funding for basic research (201110159001 to PKT) and small project funding (201309176158 to CSMT). The work described was also partially supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, Project No. [T12C-714/14R] to PKT; CH, PG are supported by grants from the Italian Ministry of Health through “Cinqu per mille” and Ricerca Corrente to the Gaslini Institute; GA, MB, BL-T, MR-F, SB are supported by the Spanish Ministry of Economy and Competitiveness (Institute of Health Carlos III (ISCIII), PI13/01560 and CDTI, FEDER-Innterconecta EXP00052887/ITC-20111037), and the Regional Ministry of Innovation, Science and Enterprise of the Autonomous Government of Andalusia (CTS-7447); NINDS (5R21NS082546) awarded to ITS.

AUTHOR CONTRIBUTIONS
H.G. and D.S. performed the exome sequencing analyses and wrote the manuscript. W.C. together with A.J.B., R.C., V.L., B.J. and I.T.S. performed the zebrafish experiments and prepared the figures for the manuscript. Y.S. and C.S.T. conducted the CNV analyses. Sanger sequencing validation was performed by P.G., I.M., A.P., M.T.S., M.R.F., B.L-T. and D.S. Statistical support was provided by H.G., M.B., R.W.W.B., T.L., S.C., P.S. and A.C. Expression data was obtained and analyzed by Y.S. and E.S.W.N. Bioinformatics support was provided by S.C., P.S., M.v.d.H., W.v.IJ. and J.B.G.M.V. A.S.B., C.B., P.T., J.A., S.L., R.H., B.E., M.M.G.B., G.A., S.B. and I.C were involved in patient recruitment and clinical aspect of the study. P.T., J.A., S.L., R.H., B.E., M.M.G.B., S.B., I.C. and A.C. conceived and design the project. All authors contributed to writing and editing.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.
REFERENCES

DE NOVO MUTATIONS IN HSCR PATIENTS LINK CNS GENES TO THE DEVELOPMENT OF THE ENS


SUPPLEMENTARY NOTES

Quality assessment and control for exome variants
Concrete criterions in quality assessment (QA) include: total number of variants; dbSNP137 coverage; Transition/Transversion (Ti/Tv) ratio; genotype concordance rate and cross-sample identical-by-decent (IBD) relatedness. Two complementary steps were applied in quality control (QC), including variant-level filtering (hard filtration or variant quality recalibration (VQSR)) and genotype-level filtering. In detail, we annotated GATK-called variants as low quality SNPs (“QD <2.0” or “MQ <40.0” or “FS >60.0” or “HaplotypeScore >13.0” or “MQRankSum < -12.5” or “ReadPosRankSum < -8.0” in their ‘info’ field) and low quality Indels (“QD <2.0” or “ReadPosRankSum < -20.0” or “InbreedingCoeff < -0.8” or “FS >200.0 in ‘info’ field); in addition, VQSR differentiated a few relatively low quality SNVs (labeled as “TruthSensitivityTranche99.90to100.00” after Gaussian mixture modeling at true sensitivity 99%) from other passed SNVs. On the other hand, individual genotypes were evaluated by quality parameters in the field of genotyping, mainly reflecting the likelihood of three possible genotypes (reference homozygous, heterozygous and alternative homozygous). A heterozygous genotype was kept only if it was supported by >4 total reads, and the ratio for alternative allele is above 0.25. Comparatively, a reference or alternative homozygous genotype was accepted if it was supported by > 4 total reads, and ratio for reference or alternative allele is above 0.95.

Supplementary Table 2 shows the details of quality statistics for samples from different sequencing centers at variant level. The total count of SNVs (20~30K) or Indels (1~2K), Transition/Transversion (Ti/Tv) ratio (above 3.0), dbSNP137 coverage (above 95%) and GWAS genotype concordance (>99%) are all in normal range. No trio violated relatedness checking; meanwhile, no batch effects or close relatedness (pi-hat coefficient > 0.125 as first cousin or above) were found among the HSCR patients from different centers (Supplementary Figure 2). All these quality metrics or statistics showed data quality at exonic regions that were comparatively good for trios from different platforms or resources, and justified our unbiased searching of de novo mutations in the following stages.

Mutation validation and prediction
Each DNM candidate was manually inspected using the Integrative Genomic Viewer (IGV) and they were categorized into five different groups: probably true
positive, possibly true positive, unclear, possibly false positive and probably false positive. Two lists of putative DNM candidates were generated for confirmation by Sanger sequencing. The first list contains 74 variants with high confidence ranking (probably true positive and possibly true positive). Raw data were then re-evaluated to generate 48 candidates with relatively low-confidence (unclear), especially for those trios without any confirmed DNM in the first round. Rare (minor allele frequency < 0.01 in public databases) predicted damaging variants in genes carrying confirmed de novo mutations were extracted from exome calls and submitted for Sanger validation. The allele origin was determined by checking the mutation site in both parents. Phasing of DNM and inherited variants in the same gene was also performed by Sanger sequencing. Rare damaging inherited variants located in 116 ENS candidate genes were extracted from exome reads using the same pipeline (Supplementary Figure 1); and the transmission patterns of these variants were determined by referring to parental and maternal genotypes at the same site.

Stepwise logistic regression was used to select effective predictors of the de novo status in a trio and for the presence or absence of a mutation in a given individual. The performance of these prediction models was evaluated using 10-fold cross validation by the software WEKA. For model fitting to DNM status in the trios, genotype quality (represented by normalized phred likelihood score for the second most likely genotype) in the child and alternative allelic ratio in the parents were prioritized. The Area Under the Receiver Operating Characteristic Curve (AUC) was 0.959 (Supplementary Table 3) which suggests that the model predicts the DNM status accurately. This model was then adopted to test all other unvalidated de novo candidates (falling under the "unclear", "possibly false positive" or "probably false positive" categories), which all turned out to be negatives. For model fitting to the presence or absence of a variant in the patients, genotype quality and alternative allelic ratio in each individual were retained. The AUC was 0.824 (Supplementary Table 3). This second model was then used to help predict the presence of rare variants in the DNM genes or ENS genes. Only those variants predicted as positive candidates were shown (Supplementary Table 5).

**Generation of ENS candidate genes**
Candidate genes were selected by a literature review on Hirschsprung disease research, which included both genetic and functional studies. Most of them were also covered in Jiang et al.\(^2\) and Gui et al.\(^3\), which previously summarized possible
genes related to HSCR or involved in ENS development. The genes were categorized into 4 major types, genes selected based on: genetic linkage, genetic association, microarray expression, and animal models. In total 116 genes were selected that fit more than 1 category (Supplementary Table 6). A few of these genes fall into the same pathways previously implicated in neural crest cell migration, proliferation and differentiation. Three pathways (RET signaling pathway, EDNRB signaling pathway and KBP signaling pathway) were key partners involved in ENS development.

**SUPPLEMENTARY TABLES**

**Supplementary Table 1. Information of sample included in the study.**

<table>
<thead>
<tr>
<th>HSCR patients (N=52)</th>
<th>Trios (N=24)</th>
<th>Singletons (N=28)</th>
<th>Controls (N=212)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short (N=1)</td>
<td>Long/TCA (N=23)</td>
<td>Short (N=15)</td>
</tr>
<tr>
<td>Males</td>
<td>0</td>
<td>7 (4)</td>
<td>13</td>
</tr>
<tr>
<td>Females</td>
<td>1 (0)</td>
<td>16 (10)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>117</td>
</tr>
</tbody>
</table>

Trios used to detect de novo mutations in coding sequences. Case/control samples were exome sequenced by the same protocol in each cohort, and used to calculate gene-level burden for all genes carrying a de novo mutation. ( ) : number of patients with validated DNM. TCA: total colonic aganglionosis.

**Supplementary Table 2. Quality metrics for sequencing reads and variants from different cohorts.**

<table>
<thead>
<tr>
<th>Centre</th>
<th># of Trios</th>
<th>Capture array</th>
<th>Target region</th>
<th>Sequencer</th>
<th>Mean coverage</th>
<th>&gt;10X SNVs/indels per patient</th>
<th>Ti/Tv ratio</th>
<th>Concordance rate</th>
<th>dbSNP v137 coverage</th>
<th>RV per patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>HK</td>
<td>5</td>
<td>Illumina Truseq</td>
<td>62.3 M</td>
<td>Illumina GAII</td>
<td>27.9 X</td>
<td>74%</td>
<td>10475 / 234</td>
<td>3.52</td>
<td>NA</td>
<td>99.17%</td>
</tr>
<tr>
<td>NL</td>
<td>10</td>
<td>Agilent SS V4</td>
<td>51.4 M</td>
<td>Illumina HiSeq2000</td>
<td>53.8 X</td>
<td>95%</td>
<td>13603 / 342</td>
<td>3.34</td>
<td>99.10%</td>
<td>99.29%</td>
</tr>
<tr>
<td>FR</td>
<td>5</td>
<td>Agilent SS V4</td>
<td>51.4 M</td>
<td>Illumina HiSeq2000</td>
<td>51.8 X</td>
<td>92%</td>
<td>12432 / 287</td>
<td>3.42</td>
<td>NA</td>
<td>99.51%</td>
</tr>
<tr>
<td>SP</td>
<td>4</td>
<td>NimbleGen V2</td>
<td>36.5 M</td>
<td>ABI Solid4</td>
<td>47.4 X</td>
<td>82%</td>
<td>10502 / 530</td>
<td>3.59</td>
<td>NA</td>
<td>95.50%</td>
</tr>
</tbody>
</table>

This shows comparable read depth and % of targeted exonic bases on the intersected exonic regions (~30Mb) for different cohorts; in addition, variant-level metrics are also comparable at exonic regions (Ti/Tv ratio, SNP/Indel counts, dbSNP137 coverage). #: SNVs passing variant quality recalibration filtering were counted; #: only SNVs in exonic regions were used to estimate Ti/Tv ratio; %: concordance between GWAS array and exome data, NA data not available; $: RV, rare variants with minor allele frequency < 0.01 in dbsnp137, 1000 genome 2012 and ESP 65000 databases; SS: Sure Select.
## Supplementary Table 3. Statistical models for mutation prediction.

<table>
<thead>
<tr>
<th>Model</th>
<th>Classifier</th>
<th>Confusion matrix</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F-measure</th>
<th>AUC (10-fold CV)¹#</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNM status in trios</td>
<td>2ndPL patient + FA parents</td>
<td>93</td>
<td>3</td>
<td>0.692</td>
<td>0.969</td>
<td>0.857</td>
<td>0.766</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variant presence/absence in patients</td>
<td>2ndPL patient + FA patient</td>
<td>68</td>
<td>10</td>
<td>0.703</td>
<td>0.872</td>
<td>0.839</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two models were trained by stepwise logistic regression on sequencing quality metrics and then used to predict the *de novo* mutation status in a trio or the variant presence/absence status in exome individuals. Training data was from true or false variants validated by Sanger sequencing, as shown in confusion matrix. *: 2ndPL_patient means "second minimum phred-scaled likelihood (PL) score" in the trio proband; FA_parents means maximum "fractions of reads (FA) supporting each reported alternative allele" from two parents. 2ndPL_patient, FA_patient means PL or FA value for given patient. #: Area under curve (AUC) calculated from 10-fold cross-validation. Confusion matrix, F-measure and AUC were acquired from WEKA output.

## Supplementary Table 4. Comparison of *de novo* mutation rates

<table>
<thead>
<tr>
<th>Mutation type</th>
<th>HSCR-trios (N=24)</th>
<th>Healthy-trios* (N=54)</th>
<th>p-value</th>
<th>Unaffected siblings (N=677)²</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count (rate)</td>
<td>Count (rate)</td>
<td>A vs. B</td>
<td>Count (rate)</td>
<td>A vs. C</td>
</tr>
<tr>
<td>All DNM</td>
<td>28 (1.17)</td>
<td>44 (0.81)</td>
<td>0.159</td>
<td>547 (0.81)</td>
<td>0.065</td>
</tr>
<tr>
<td>LOF DNM</td>
<td>8 (0.33)</td>
<td>4 (0.07)</td>
<td>0.011⁴</td>
<td>54 (0.08)</td>
<td>0.001⁵#</td>
</tr>
<tr>
<td>Non-RET LOF DNM</td>
<td>3 (0.13)</td>
<td>4 (0.07)</td>
<td>0.447</td>
<td>54 (0.08)</td>
<td>0.447</td>
</tr>
<tr>
<td>Synonymous DNM</td>
<td>7 (0.29)</td>
<td>12 (0.22)</td>
<td>0.62</td>
<td>143 (0.21)</td>
<td>0.365</td>
</tr>
</tbody>
</table>

DNM mutation rate by different categories (All, LOF only, non-RET LOF, synonymous) were compared between HSCR trios included in this study and those published healthy trios or unaffected siblings to neurodevelopmental diseases. *: Data from Rauch (2012) and Xu (2012); &: data from Iossifov (2012), O’Roak (2012), Sanders (2012) and Gulsuner (2013); #: nominally significant at 0.05; ##: significant after Bonferroni correction.
### Supplementary Table 5. Joint distribution of common and rare variants for each trio proband.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>RET rs2435357: T/C</th>
<th>De novo mutations</th>
<th>Inherited mutations in genes in which de novo mutations were found</th>
<th>Inherited mutations in 116 ENS/HSCR candidate genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>L, F</td>
<td>CC</td>
<td>RET: 3splicing9+1 (splicing site); RBM25: L158L (synonymous); RET: S837fs (frameshift)</td>
<td>RET: L56M (missense) (P)</td>
<td>SMO (P), KIAA1279 (M)</td>
</tr>
<tr>
<td>L, F</td>
<td>CC</td>
<td></td>
<td></td>
<td>DCC (P)</td>
</tr>
<tr>
<td>L, F</td>
<td>TC</td>
<td>RET: 6006fs (frameshift);</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>CT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>CC</td>
<td>SCUBE3: N498I (missense); RET: G588fs (frameshift)</td>
<td>PLEKHG5: E800fs (frameshift) (U)</td>
<td>NOTCH3 (M)</td>
</tr>
<tr>
<td>L, M</td>
<td>TT</td>
<td>PLEK9G5: T876T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L, M</td>
<td>TT</td>
<td>KDM4A: N9S (missense); MAP4: A882G (missense) (M)</td>
<td>ECE1 (P), JAG1 (P)</td>
<td></td>
</tr>
<tr>
<td>L, M</td>
<td>CT</td>
<td>MAP4: G1117G (synonymous)</td>
<td></td>
<td>PCDHA1 (P), DCC (M), NOTCH3 (P)</td>
</tr>
<tr>
<td>L, F</td>
<td>TT</td>
<td>RET: C620R (missense)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCA, M</td>
<td>TT</td>
<td>CKAP2L: E186fs (frameshift)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>C/T</td>
<td>HMCM1: A3456T (missense); RET: C137G (missense); TUBG1: S2338S (synonymous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>C/T</td>
<td>CCR2: L283Q (missense); DNN2D: K460fs (frameshift)</td>
<td>IKBKAP (P), JAG1 (M)</td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>C/T</td>
<td>RET: G570* (stopgain)</td>
<td></td>
<td>ECE1 (P)</td>
</tr>
<tr>
<td>L, F</td>
<td>C/T</td>
<td>RET: R175del (non-frameshift); TBATA: R53C (missense)</td>
<td></td>
<td>NOTCH1 (P), PFKL (P)</td>
</tr>
<tr>
<td>L, F</td>
<td>CC</td>
<td>HMCM1: P1269T (missense) (M)</td>
<td>IKBKAP (P), EDN1 (P), JAG1 (M)</td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>CC</td>
<td>HMCM1: N2461S (missense) (M)</td>
<td>PHACTR4 (P), GLI3 (M), SHH (M), HMX3 (M), NAV2 (M), PRPH (P), PSPN (P)</td>
<td></td>
</tr>
<tr>
<td>L, M</td>
<td>TT</td>
<td></td>
<td>IHH (P), PFKL (P), JAG1 (M)</td>
<td></td>
</tr>
<tr>
<td>S, F</td>
<td>TT</td>
<td>JAG1 (P)</td>
<td>ELAVL4 (P), SERPINI1 (U), PTCH1 (U), IKBKAP (P)</td>
<td></td>
</tr>
<tr>
<td>TCA, M</td>
<td>TT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>TT</td>
<td>SCUBE3: R907C (missense) (P)</td>
<td>NRG1 (M), IFNGR2 (P)</td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>TT</td>
<td></td>
<td>TAGLIN3 (P)</td>
<td></td>
</tr>
<tr>
<td>L, F</td>
<td>TC</td>
<td>DAB2I: P338T (missense) (P); KDM4A: Y989M (M)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Common risk SNP (RET rs2435357), DNMs, inherited damaging variants in genes carrying DNMs, rare damaging variants in ENS candidate genes were tabulated for each HSCR patient. DNMs and inherited variants in DNMs were confirmed by Sanger sequencing. Rare damaging variants in ENS candidate genes were all predicted as true according to training model 2 (see Supplementary Table 3). #: L: Long segment aganglionosis; S: Short segment aganglionosis; TCA: Total colonic aganglionosis; F: Female; M: Male. #: rs2435357, T is risk allele and minor allele. %: genes functionally validated in bold; $: parent of origin for mutation in candidate genes; P for paternal (P); M for Maternal, U for Unsure; & 116 ENS-related HSCR candidate genes (as listed in Supplementary Table 6).
### Supplementary Table 6. Characteristics of 116 ENS-related HSCR candidate genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene name</th>
<th>Chromosome</th>
<th>Evidence</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALDH1A2</td>
<td>aldehyde dehydrogenase 1 family, member A2</td>
<td>15q22.1</td>
<td>Mouse (Absence EN)</td>
<td>5</td>
</tr>
<tr>
<td>ARHGEF3</td>
<td>Rho guanine nucleotide exchange factor (GEF) 3</td>
<td>3p14.3</td>
<td>Expression</td>
<td>6,7</td>
</tr>
<tr>
<td>ARTN</td>
<td>artemin</td>
<td>1p34.1</td>
<td>Mouse (Abnormal ENS morphology)</td>
<td>8</td>
</tr>
<tr>
<td>ASCL1</td>
<td>achaete-scute complex homolog 1 (Drosophila)</td>
<td>12q23.2</td>
<td>Mouse (Absence EN)/Expression</td>
<td>7–9</td>
</tr>
<tr>
<td>CADM1</td>
<td>cell adhesion molecule 1</td>
<td>11q23.2</td>
<td>Expression</td>
<td>7,12</td>
</tr>
<tr>
<td>CARTPT</td>
<td>CART prepropeptide</td>
<td>5q13.2</td>
<td>Expression</td>
<td>7</td>
</tr>
<tr>
<td>CBR1</td>
<td>carbonyl reductase 1</td>
<td>21q22.13</td>
<td>Expression</td>
<td>13</td>
</tr>
<tr>
<td>CDH2</td>
<td>cadherin 2, type 1, N-cadherin (neuronal)</td>
<td>1q21.2.2</td>
<td>Expression</td>
<td>9</td>
</tr>
<tr>
<td>CRMP1</td>
<td>collapsin response mediator protein 1</td>
<td>1q16.1</td>
<td>Expression</td>
<td>7,15</td>
</tr>
<tr>
<td>CSTB</td>
<td>cystatin B (stefin B)</td>
<td>21q22.3</td>
<td>Expression</td>
<td>16</td>
</tr>
<tr>
<td>CTTNL1</td>
<td>catterin (cadherin-associated protein), alpha-like 1</td>
<td>9q31.3</td>
<td>Expression</td>
<td>7</td>
</tr>
<tr>
<td>DCC</td>
<td>deleted in colorectal carcinoma</td>
<td>18p21.2</td>
<td>Mouse (Absence submucosal ganglia)</td>
<td>17</td>
</tr>
<tr>
<td>DCX</td>
<td>doublecortin</td>
<td>Xq22.3-q23</td>
<td>Expression</td>
<td>7</td>
</tr>
<tr>
<td>DLL1</td>
<td>delta-like 1 (Drosophila)</td>
<td>6q27</td>
<td>Not described</td>
<td>10,11</td>
</tr>
<tr>
<td>DLL3</td>
<td>delta-like 3 (Drosophila)</td>
<td>19q13.2</td>
<td>Not described</td>
<td>10,11</td>
</tr>
<tr>
<td>Dlx1</td>
<td>distal-less homeobox 1</td>
<td>2q32</td>
<td>Expression</td>
<td>7,16,1</td>
</tr>
<tr>
<td>DPYSL3</td>
<td>dihydropyrimidinase-like 3</td>
<td>5q22</td>
<td>Expression</td>
<td>7,20</td>
</tr>
<tr>
<td>EBF3</td>
<td>early B-cell factor 3</td>
<td>10q26.3</td>
<td>Expression</td>
<td>7,21</td>
</tr>
<tr>
<td>ECE1</td>
<td>endothelin converting enzyme 1</td>
<td>1p36</td>
<td>Human (Linkage/Mouse (Absence EN))</td>
<td>17,22</td>
</tr>
<tr>
<td>EDN3</td>
<td>endothelin 3</td>
<td>20q13</td>
<td>Human (Linkage/Mouse (Absence EN))</td>
<td>17,22,</td>
</tr>
<tr>
<td>EDNRB</td>
<td>endothelin receptor type B</td>
<td>13q22</td>
<td>Human (Linkage/CNV/Mouse (Absence EN))</td>
<td>26,27</td>
</tr>
<tr>
<td>ELAVL2</td>
<td>ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)</td>
<td>9p21</td>
<td>Expression</td>
<td>7,28</td>
</tr>
<tr>
<td>ELAVL4</td>
<td>ELAV (embryonic lethal, abnormal vision, Drosophila)-like 4 (Hu antigen D)</td>
<td>1p34</td>
<td>Expression</td>
<td>7,29</td>
</tr>
<tr>
<td>ERBB2</td>
<td>oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)</td>
<td>17q12</td>
<td>Mouse (Abnormal ENS morphology)</td>
<td>30,31</td>
</tr>
<tr>
<td>ERBB3</td>
<td>v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, avian</td>
<td>12q13.2</td>
<td>Mouse (Abnormal ENS morphology)</td>
<td>30,31</td>
</tr>
<tr>
<td>ERBB4</td>
<td>v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)</td>
<td>2q33.3-q34</td>
<td>Human (CNV)</td>
<td>32</td>
</tr>
<tr>
<td>ETV1</td>
<td>ets variant 1</td>
<td>7p21.3</td>
<td>Expression</td>
<td>7,33</td>
</tr>
<tr>
<td>FGF13</td>
<td>fibroblast growth factor 13</td>
<td>Xq26.3</td>
<td>Expression</td>
<td>7,34</td>
</tr>
<tr>
<td>GAP43</td>
<td>growth associated protein 43</td>
<td>3q13.1-q13.2</td>
<td>Expression</td>
<td>7,35</td>
</tr>
<tr>
<td>GDNF</td>
<td>glial cell derived neurotrophic factor</td>
<td>5p13</td>
<td>Human (Linkage/Mouse (Absence EN)/Expression</td>
<td>17,22,</td>
</tr>
<tr>
<td>GFRA1</td>
<td>GDNF family receptor alpha 1</td>
<td>10q25</td>
<td>Human (1 patient)/Mouse (Absence EN)/Expression</td>
<td>7,38,</td>
</tr>
<tr>
<td>GFRA2</td>
<td>similar to GDNF family receptor alpha 2; GDNF family receptor alpha 2</td>
<td>8p21.3</td>
<td>Mouse (Abnormal ENS morphology)</td>
<td>8</td>
</tr>
<tr>
<td>GFRA3</td>
<td>GDNF family receptor alpha 3</td>
<td>5q11.2</td>
<td>Mouse (Abnormal sympathetic system)</td>
<td>8</td>
</tr>
<tr>
<td>GFRA4</td>
<td>GDNF family receptor alpha 4</td>
<td>20p13</td>
<td>Not described</td>
<td>8</td>
</tr>
<tr>
<td>GLI1</td>
<td>GLI family zinc finger 1</td>
<td>12q13.3</td>
<td>Mouse (Abnormal intestinal morphology)</td>
<td>41,42</td>
</tr>
<tr>
<td>GLI2</td>
<td>GLI family zinc finger 2</td>
<td>2q14</td>
<td>Mouse (Abnormal intestinal morphology)</td>
<td>41,42</td>
</tr>
<tr>
<td>GLI3</td>
<td>GLI family zinc finger 3</td>
<td>7p14</td>
<td>Mouse (Abnormal intestinal morphology)</td>
<td>41,42</td>
</tr>
<tr>
<td>GNG2</td>
<td>guanine nucleotide binding protein (G protein), gamma 2</td>
<td>14q21</td>
<td>Expression</td>
<td>7</td>
</tr>
</tbody>
</table>
**De novo mutations in HSCR patients link CNS genes to the development of the ENS**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Chromosome</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNG3</td>
<td>Guanine nucleotide binding protein (G protein), gamma 3</td>
<td>11p11</td>
<td>7,43</td>
</tr>
<tr>
<td>GRB10</td>
<td>Growth factor receptor-bound protein 10</td>
<td>7p12.2</td>
<td>38</td>
</tr>
<tr>
<td>HES1</td>
<td>Hairy and enhancer of split 1, (Drosophila)</td>
<td>3p29</td>
<td>41,42</td>
</tr>
<tr>
<td>HMP19</td>
<td>HMP19 protein</td>
<td>5q35.2</td>
<td>7</td>
</tr>
<tr>
<td>HMX3</td>
<td>H6 family homeobox 3</td>
<td>10q26.13</td>
<td>7,44</td>
</tr>
<tr>
<td>HOXB5</td>
<td>Homeobox B5</td>
<td>17q21.3</td>
<td>7,45</td>
</tr>
<tr>
<td>HOXD4</td>
<td>Homeobox D4</td>
<td>2q31.1</td>
<td>7,46</td>
</tr>
<tr>
<td>IFNGR2</td>
<td>Interferon gamma receptor 2 (interferon gamma transducer 1)</td>
<td>21q22.11</td>
<td>47</td>
</tr>
<tr>
<td>IHH</td>
<td>Indian hedgehog homolog (Drosophila), inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein</td>
<td>2q35</td>
<td>41,42</td>
</tr>
<tr>
<td>IL10RB</td>
<td>Interleukin 10 receptor, beta integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)</td>
<td>21q22.11</td>
<td>49</td>
</tr>
<tr>
<td>ITGB1</td>
<td>Integrin beta 3 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)</td>
<td>10p11.22</td>
<td>17</td>
</tr>
<tr>
<td>JAG1</td>
<td>Jagged 1 (Alagille syndrome)</td>
<td>20p12.1</td>
<td>10,11</td>
</tr>
<tr>
<td>JAG2</td>
<td>Jagged 2</td>
<td>14q32.33</td>
<td>10,11</td>
</tr>
<tr>
<td>KIAA1279</td>
<td>KIAA1279</td>
<td>10q21</td>
<td>50</td>
</tr>
<tr>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>9q31</td>
<td>41,42</td>
</tr>
<tr>
<td>L1CAM</td>
<td>L1 cell adhesion molecule</td>
<td>Xq28</td>
<td>48</td>
</tr>
<tr>
<td>MAB21L1</td>
<td>Mab-21-like 1 (C. elegans)</td>
<td>13q13</td>
<td>7,54</td>
</tr>
<tr>
<td>MAPK10</td>
<td>Mitogen-activated protein kinase 10</td>
<td>4q22.1-q23</td>
<td>7,55</td>
</tr>
<tr>
<td>MAPT</td>
<td>Microtubule-associated protein tau myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11</td>
<td>17q21.1</td>
<td>7</td>
</tr>
<tr>
<td>MLLT11</td>
<td>Mllt1 (C. elegans)</td>
<td>1q21</td>
<td>7,56</td>
</tr>
<tr>
<td>NAV2</td>
<td>Neuron navigator 2</td>
<td>11p15.1</td>
<td>57</td>
</tr>
<tr>
<td>NKX2-1</td>
<td>NK2 homeobox 1</td>
<td>14q13</td>
<td>17,22</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>Notch homolog 1, translocation-associated (Drosophila)</td>
<td>9q34</td>
<td>10,11</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>Notch homolog 2 (Drosophila)</td>
<td>6q27</td>
<td>10,11</td>
</tr>
<tr>
<td>NOTCH3</td>
<td>Notch homolog 3 (Drosophila)</td>
<td>19p13.12</td>
<td>10,11</td>
</tr>
<tr>
<td>NRG1</td>
<td>Neuregulin 1</td>
<td>8p12</td>
<td>58</td>
</tr>
<tr>
<td>NRG3</td>
<td>Neuregulin 3</td>
<td>10q22-q23</td>
<td>32,59</td>
</tr>
<tr>
<td>NRP1</td>
<td>Neuropilin 1</td>
<td>10p11.22</td>
<td>60,61</td>
</tr>
<tr>
<td>NRTN</td>
<td>Neurturin</td>
<td>19p13</td>
<td>62</td>
</tr>
<tr>
<td>NTF3</td>
<td>3'-nucleotidase</td>
<td>1p36.11</td>
<td>17</td>
</tr>
<tr>
<td>NTRK3</td>
<td>Neurotrophic tyrosine kinase, receptor, type 3</td>
<td>15q25</td>
<td>17</td>
</tr>
<tr>
<td>PAX3</td>
<td>Paired box 3</td>
<td>2q36</td>
<td>8,57</td>
</tr>
<tr>
<td>PCDHA1</td>
<td>Protocadherin alpha 1; protocadherin alpha 4</td>
<td>5q31</td>
<td>7,63</td>
</tr>
<tr>
<td>PFKL</td>
<td>Phosphofructokinase, liver</td>
<td>21q22.3</td>
<td>64</td>
</tr>
<tr>
<td>PHACTR4</td>
<td>Phosphatase and actin regulator 4</td>
<td>1p35.3</td>
<td>65,66</td>
</tr>
<tr>
<td>PHOX2A</td>
<td>Paired-like homeobox 2a</td>
<td>11q13.2</td>
<td>7,67</td>
</tr>
<tr>
<td>PHOX2B</td>
<td>Paired-like homeobox 2b</td>
<td>4p13</td>
<td>7,68, 69</td>
</tr>
<tr>
<td>PLXNA1</td>
<td>Plexin A1</td>
<td>3q21.3</td>
<td>60,61</td>
</tr>
<tr>
<td>PLXNB1</td>
<td>Plexin B1</td>
<td>3p21</td>
<td>70,71</td>
</tr>
</tbody>
</table>
### Chapter 2

#### Supplementary Information

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Chromosome</th>
<th>Mouse</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>POFUT1</td>
<td>protein O-fucosyltransferase 1</td>
<td>20q11.21</td>
<td>Mouse (Absence EN)</td>
<td>10,11</td>
</tr>
<tr>
<td>PROK1</td>
<td>prokineticin 1</td>
<td>1p13</td>
<td>Not described</td>
<td>72,73</td>
</tr>
<tr>
<td>PROK2</td>
<td>prokineticin 2</td>
<td>3p13</td>
<td>Not described</td>
<td>72,73</td>
</tr>
<tr>
<td>PROKR1</td>
<td>prokineticin receptor 1</td>
<td>2p14</td>
<td>Not described</td>
<td>72,73</td>
</tr>
<tr>
<td>PROKR2</td>
<td>prokineticin receptor 2</td>
<td>20p12</td>
<td>Not described</td>
<td>72,73</td>
</tr>
<tr>
<td>PRPH</td>
<td>peripherin</td>
<td>12q12-q13</td>
<td>Expression</td>
<td>7,74</td>
</tr>
<tr>
<td>PSPN</td>
<td>perserin</td>
<td>19p13.3</td>
<td>Not described</td>
<td>8</td>
</tr>
<tr>
<td>PTC1</td>
<td>patched homolog 1 (Drosophila)</td>
<td>9q22.32</td>
<td>Mouse (Abnormal intestinal morphology)</td>
<td>41,42</td>
</tr>
<tr>
<td>RET</td>
<td>ret proto-oncogene</td>
<td>10q11</td>
<td>(Linkage/CNV/Exome)/Mouse (Absence EN)</td>
<td>7,57,75,76</td>
</tr>
<tr>
<td>SALL4</td>
<td>sal-like 4 (Drosophila)</td>
<td>20q13.2</td>
<td>Mouse (Absence EN)</td>
<td>17</td>
</tr>
<tr>
<td>SCG3</td>
<td>secretogranin III</td>
<td>15q26</td>
<td>Not described</td>
<td>8</td>
</tr>
<tr>
<td>SEMA3A</td>
<td>sema domain, immunoglobulin domain (ig), short basic domain, secreted, (semaphorin) 3A</td>
<td>7p12</td>
<td>(Expression/Human)</td>
<td>57,77</td>
</tr>
<tr>
<td>SEMA3C/D</td>
<td>sema domain, immunoglobulin domain (ig), short basic domain, secreted, (semaphorin) 3D</td>
<td>7p12</td>
<td>Human (GWAS/Others)</td>
<td>60,61</td>
</tr>
<tr>
<td>SERPINI1</td>
<td>serpin peptidase inhibitor, clade I (neuroserpin), member 1</td>
<td>3q26.1</td>
<td>Expression</td>
<td>7,78,79</td>
</tr>
<tr>
<td>SHH</td>
<td>sonic hedgehog homolog (Drosophila)</td>
<td>7q36</td>
<td>Mouse (Ectopic enteric ganglia formation)</td>
<td>41,42</td>
</tr>
<tr>
<td>SMO</td>
<td>smoothened homolog (Drosophila)</td>
<td>7q32</td>
<td>Mouse (Abnormal neural crest cell migration)</td>
<td>41,42</td>
</tr>
<tr>
<td>SQD1</td>
<td>superoxide dismutase 1, soluble</td>
<td>21q22.11</td>
<td>Expression</td>
<td>80,81</td>
</tr>
<tr>
<td>SON</td>
<td>SON DNA binding protein</td>
<td>21q22.11</td>
<td>Expression</td>
<td>82</td>
</tr>
<tr>
<td>SOX10</td>
<td>SRY (sex determining region Y)-box 10</td>
<td>22q13</td>
<td>(Expression/Human)</td>
<td>7,83</td>
</tr>
<tr>
<td>SOX2</td>
<td>SRY (sex determining region Y)-box 2</td>
<td>3q26.3-q27</td>
<td>Expression</td>
<td>7,84</td>
</tr>
<tr>
<td>SPRY2</td>
<td>sprouty homolog 2 (Drosophila)</td>
<td>13q31.1</td>
<td>Mouse (Increased EN)</td>
<td>17</td>
</tr>
<tr>
<td>STMN2</td>
<td>stathmin-like 2</td>
<td>8q21.13</td>
<td>Expression</td>
<td>7,85</td>
</tr>
<tr>
<td>STMN3</td>
<td>stathmin-like 3</td>
<td>20q13.3</td>
<td>Expression</td>
<td>7</td>
</tr>
<tr>
<td>SUFU</td>
<td>suppressor of fused homolog (Drosophila)</td>
<td>10q24.32</td>
<td>Mouse (Abnormal neural tube morphology)</td>
<td>41,42</td>
</tr>
<tr>
<td>SYT11</td>
<td>synaptotagmin XI</td>
<td>1q21.2</td>
<td>Expression</td>
<td>7,86</td>
</tr>
<tr>
<td>TAGLN3</td>
<td>transgelin 3</td>
<td>3q13.2</td>
<td>Expression</td>
<td>7</td>
</tr>
<tr>
<td>TBX3</td>
<td>T-box 3</td>
<td>12q24.1</td>
<td>Expression</td>
<td>87,88</td>
</tr>
<tr>
<td>TFF3</td>
<td>trefoil factor 3 (intestinal)</td>
<td>21q22.3</td>
<td>Expression</td>
<td>7,89,90</td>
</tr>
<tr>
<td>TGBF2</td>
<td>transforming growth factor, beta 2</td>
<td>1q41</td>
<td>Expression</td>
<td>7,91</td>
</tr>
<tr>
<td>TMEFF2</td>
<td>transmembrane protein with EGF-like and two follistatin-like domains 2</td>
<td>2q32.3</td>
<td>Expression</td>
<td>7,92</td>
</tr>
<tr>
<td>TREX1</td>
<td>three prime repair exonuclease 1</td>
<td>3p21.31</td>
<td>Human (Linkage)</td>
<td>70,71</td>
</tr>
<tr>
<td>TTC3</td>
<td>tetratricopeptide repeat domain 3; tetratricopeptide repeat domain 3-like tubulin, beta 3; melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)</td>
<td>21q22.2</td>
<td>Expression</td>
<td>93</td>
</tr>
<tr>
<td>TUBB3</td>
<td>ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)</td>
<td>16q24.3</td>
<td>Expression</td>
<td>7,94</td>
</tr>
<tr>
<td>VIP</td>
<td>vasoactive intestinal peptide</td>
<td>4p14</td>
<td>Expression</td>
<td>7,95</td>
</tr>
<tr>
<td>VIP</td>
<td>vasoactive intestinal peptide</td>
<td>6q25</td>
<td>Expression</td>
<td>7,96</td>
</tr>
<tr>
<td>ZEB2</td>
<td>zinc finger E-box binding homeobox 2</td>
<td>2q22</td>
<td>Human (Linkage/MW syndrome/CNV/Mouse (Abnormal NCC migration)</td>
<td>97–99</td>
</tr>
<tr>
<td>ZIC2</td>
<td>Zic family member 2 (odd-paired homolog, Drosophila)</td>
<td>13q32</td>
<td>Mouse</td>
<td>100,101</td>
</tr>
</tbody>
</table>

Notes: updated gene symbols used for ZFHX1B (replaced by ZEB2), TRKC (replaced by NTRK3) and RALDH2 (replaced by ALDH1A2); EN: enteric neurons; NCC: neural crest cells.
**DE NOVO MUTATIONS IN HSCR PATIENTS LINK CNS GENES TO THE DEVELOPMENT OF THE ENS**

**Supplementary Table 7. Gene recurrence and burden test.**

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>HK (14/73)</th>
<th>Spain (15/100)*</th>
<th>Rotterdam (19/39)</th>
<th>Meta-analysis (48/212)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>Direction*</td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td>AFF3</td>
<td>1.0000</td>
<td>-1</td>
<td>0.6973</td>
<td>0.0392</td>
</tr>
<tr>
<td>CCR2</td>
<td>1.0000</td>
<td>-1</td>
<td>0.6973</td>
<td>0.0392</td>
</tr>
<tr>
<td>CKAP2L</td>
<td>0.0216</td>
<td>1</td>
<td>0.1175</td>
<td>1.0000</td>
</tr>
<tr>
<td>COL6A3</td>
<td>0.5883</td>
<td>-1</td>
<td>0.5430</td>
<td>0.5970</td>
</tr>
<tr>
<td>DAB2IP</td>
<td>0.4402</td>
<td>-1</td>
<td>0.6973</td>
<td>0.1484</td>
</tr>
<tr>
<td>DENND3</td>
<td>0.3699</td>
<td>-1</td>
<td>0.6973</td>
<td>0.5970</td>
</tr>
<tr>
<td>HMCN1</td>
<td>0.4308</td>
<td>-1</td>
<td>0.3662</td>
<td>0.8013</td>
</tr>
<tr>
<td>ISG20L2</td>
<td>1.0000</td>
<td>-1</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>KDM4A</td>
<td>0.0216</td>
<td>1</td>
<td>0.4967</td>
<td>0.1484</td>
</tr>
<tr>
<td>MAP4</td>
<td>0.6596</td>
<td>-1</td>
<td>0.2903</td>
<td>0.5970</td>
</tr>
<tr>
<td>MED26</td>
<td>1.0000</td>
<td>-1</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>NCLN</td>
<td>1.0000</td>
<td>-1</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>NUP98</td>
<td>0.5309</td>
<td>-1</td>
<td>0.6973</td>
<td>0.0392</td>
</tr>
<tr>
<td>PLEKHG5</td>
<td>1.0000</td>
<td>-1</td>
<td>0.1999</td>
<td>0.9826</td>
</tr>
<tr>
<td>RBM25</td>
<td>1.0000</td>
<td>-1</td>
<td>0.0095</td>
<td>1.0000</td>
</tr>
<tr>
<td>RET</td>
<td>0.1867</td>
<td>1</td>
<td>0.6367</td>
<td>0.0008</td>
</tr>
<tr>
<td>SCUBE3</td>
<td>1.0000</td>
<td>-1</td>
<td>0.5806</td>
<td>1.0000</td>
</tr>
<tr>
<td>TBA7A</td>
<td>1.0000</td>
<td>-1</td>
<td>1.0000</td>
<td>0.5970</td>
</tr>
<tr>
<td>TUBG1</td>
<td>1.0000</td>
<td>-1</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>VEZF1</td>
<td>1.0000</td>
<td>-1</td>
<td>0.6973</td>
<td>0.1484</td>
</tr>
<tr>
<td>ZNF57</td>
<td>0.0216</td>
<td>1</td>
<td>0.4967</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Genes with DNMs were checked for the presence of rare damaging mutations in additional HSCR patients. The burden of rare, damaging mutations in HSCR patients was compared to that of a local population-matched controls; in addition, gene-wise burden test p-values from three cohorts (HK, Spain and Rotterdam) were combined using meta-analysis. Number of cases and controls are given in parentheses. *: Direction 1 means rare damaging variants enriched in cases, -1 means rare variants enriched in controls; #: 4 HSCR patients in discovery trios were not included due to mismatched platform with control data. Nominal P-values from meta-analyses < 0.05 are given in bold (CKAP2L and RET).
### Supplementary Table 8. Bioinformatics prediction of the functional impact of DNMs.

<table>
<thead>
<tr>
<th>Gene and mutation</th>
<th>RNA structure change</th>
<th>Human splicing finder</th>
<th>Conservation (PhyloP)</th>
<th>Gene-level relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>RET:splicing9+1</td>
<td>0.7866</td>
<td>splice donor</td>
<td>7.88</td>
<td>Major HSCR gene</td>
</tr>
<tr>
<td>RBM25:L158L</td>
<td>0.8224</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RET:S837fs</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COL6A3:H1109H</td>
<td>0.4898</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RET:Y606fs</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAB21P:1132Y</td>
<td>0.8539</td>
<td></td>
<td>1.76</td>
<td></td>
</tr>
<tr>
<td>ISG20L2:G321R</td>
<td>0.495</td>
<td></td>
<td>7.59</td>
<td></td>
</tr>
<tr>
<td>MED22:A225A</td>
<td>0.9717</td>
<td></td>
<td></td>
<td>Pathway sharing with NOTCH genes</td>
</tr>
<tr>
<td>NCLN:Q166*</td>
<td>0.5467</td>
<td></td>
<td>7.38</td>
<td>Nodal signaling, involves in CNS development</td>
</tr>
<tr>
<td>NUP98:N1662S</td>
<td>0.5235</td>
<td></td>
<td>5.95</td>
<td>Regulation of known HSCR genes, involves in CNS development</td>
</tr>
<tr>
<td>VEZF1:S195F</td>
<td>0.0565</td>
<td></td>
<td>9.86</td>
<td></td>
</tr>
<tr>
<td>ZNF57:D190D</td>
<td>0.5217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RET:G588fs</td>
<td>NA</td>
<td></td>
<td></td>
<td>Major HSCR gene</td>
</tr>
<tr>
<td>SCUBE3:N498I</td>
<td>0.5473</td>
<td></td>
<td>1.96</td>
<td>Hedgehog signaling</td>
</tr>
<tr>
<td>KDM4A:N9S</td>
<td>0.5813</td>
<td></td>
<td>3.54</td>
<td>Neural crest specification in chicken</td>
</tr>
<tr>
<td>PLEKHG5:T876T</td>
<td>0.4096</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFF3:V659L</td>
<td>0.0272</td>
<td></td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>MAP4:G1117G</td>
<td>0.996</td>
<td></td>
<td>1.40</td>
<td>Interacts with CDH2, ERBB2, MAPT and DCX</td>
</tr>
<tr>
<td>RET:C620R</td>
<td>0.4286</td>
<td></td>
<td>4.80</td>
<td>Major HSCR gene</td>
</tr>
<tr>
<td>CKAP2LE186fs</td>
<td>NA</td>
<td></td>
<td></td>
<td>Involves in CNS development</td>
</tr>
<tr>
<td>RET:C137G</td>
<td>0.6841</td>
<td></td>
<td>3.73</td>
<td>Major HSCR gene</td>
</tr>
<tr>
<td>HMCN1:A3456T</td>
<td>0.6906</td>
<td></td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>TUBG1:S233S</td>
<td>0.5802</td>
<td></td>
<td></td>
<td>Interacts with SOX2</td>
</tr>
<tr>
<td>CCR2:L283O</td>
<td>0.0659</td>
<td></td>
<td>5.87</td>
<td>3p21; interacts with GLI2; shares pathway with EDN3</td>
</tr>
<tr>
<td>DENND3K640fs</td>
<td>NA</td>
<td></td>
<td></td>
<td>Involves in CNS development</td>
</tr>
<tr>
<td>RET:C570*</td>
<td>0.1453</td>
<td></td>
<td>-0.02</td>
<td>Major HSCR gene</td>
</tr>
<tr>
<td>RET:R175del</td>
<td>NA</td>
<td></td>
<td></td>
<td>Major HSCR gene</td>
</tr>
<tr>
<td>TBAT:A53C</td>
<td>0.5526</td>
<td></td>
<td>1.71</td>
<td>Involves in CNS development</td>
</tr>
</tbody>
</table>

Bioinformatics prediction tools, databases and literature were used to predict functional impact of DNMs and the genes carrying DNMs. *: significant changes (< 0.2) are in bold; #: only potential splice sites (donor or acceptor) are shown; %: PhyloP score > 2 means conservative; $: evidence collected from PubMed literature and bioinformatics databases (STRING, MsigDB pathways).
**Supplementary Table 9. Sequence and dosage of antisense morpholino.**

### a. Splice-blocking morpholino

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Human ortholog</th>
<th>Sequence</th>
<th>Dosage (ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aff3</td>
<td>AFF3</td>
<td>AAATGTCTTTCCCCCCTCACCTTTC</td>
<td>6</td>
</tr>
<tr>
<td>ckap2l</td>
<td>CKAP2L</td>
<td>TGAAGTAAACTCAAGTCTTTCTTC</td>
<td>6</td>
</tr>
<tr>
<td>dab2ipa</td>
<td>DAB2IP*</td>
<td>AGGTCAACGACTCACCTCGAGCA</td>
<td>6</td>
</tr>
<tr>
<td>dab2ipb</td>
<td>DAB2IP*</td>
<td>GCTTTCCACTACCTCTCCAGC</td>
<td>6</td>
</tr>
<tr>
<td>dennd3a</td>
<td>DENND3*</td>
<td>CATCTTTACCTGTGCAAGAAGTTA</td>
<td>6</td>
</tr>
<tr>
<td>dennd3b</td>
<td>DENND3*</td>
<td>CCATTCAATTGGTTACCTGGA</td>
<td>6</td>
</tr>
<tr>
<td>hmcn1</td>
<td>HMCN1</td>
<td>GCACAAAAGATTTCCCCCTACCTGA</td>
<td>6</td>
</tr>
<tr>
<td>isg20l2</td>
<td>ISG20L2</td>
<td>CTACTGATGCTATTTTACCTTCT</td>
<td>6</td>
</tr>
<tr>
<td>kdm4aa</td>
<td>KDM4A*</td>
<td>GACACAGAATGACAGTACAGGGA</td>
<td>6</td>
</tr>
<tr>
<td>kdm4ab</td>
<td>KDM4A*</td>
<td>AGTTGAACAGAACATACTTGCT</td>
<td>6</td>
</tr>
<tr>
<td>ncl1</td>
<td>NCLN</td>
<td>GAACCTGCAATGGATGTGTTTAT</td>
<td>6</td>
</tr>
<tr>
<td>nup98</td>
<td>NUP98</td>
<td>GTATGAACAGCATTTACCAGTGT</td>
<td>1</td>
</tr>
<tr>
<td>scube3</td>
<td>SCUBE3</td>
<td>ACCTAGATGAAGGACTCACTTGC</td>
<td>6</td>
</tr>
<tr>
<td>tbata</td>
<td>TBATA</td>
<td>GATAGGCCAATACCTGTATCCC</td>
<td>4</td>
</tr>
<tr>
<td>vezf1a</td>
<td>VEZF1*</td>
<td>AGCCAAATCGCACTAGCCTACCTT</td>
<td>6</td>
</tr>
<tr>
<td>vezf1b</td>
<td>VEZF1*</td>
<td>ATCCAAAATGCCAACCACCTAGA</td>
<td>6</td>
</tr>
</tbody>
</table>

### b. Translation-blocking morpholino

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Human ortholog</th>
<th>Sequence</th>
<th>Dosage (ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ckap2l</td>
<td>CKAP2L</td>
<td>GTCTTCACTCTCGATCCTTCT</td>
<td>6</td>
</tr>
<tr>
<td>dennd3a</td>
<td>DENND3*</td>
<td>GACCGTCTGCATTGAAAAATGCA</td>
<td>8</td>
</tr>
<tr>
<td>dennd3b</td>
<td>DENND3*</td>
<td>GACCGTCCTGGAAATCAGCAGC</td>
<td>8</td>
</tr>
<tr>
<td>ncl1</td>
<td>NCLN</td>
<td>ACCTCAGCGCTGTCAACACATGC</td>
<td>0.8</td>
</tr>
<tr>
<td>nup98</td>
<td>NUP98</td>
<td>GTTGAAACATCTTGCACCTGCTAG</td>
<td>12</td>
</tr>
<tr>
<td>tbata</td>
<td>TBATA</td>
<td>AGCACCTGCACAAACAAATACAG</td>
<td>6</td>
</tr>
</tbody>
</table>

### c. Control morpholino

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Human ortholog</th>
<th>Sequence$^\text{a}$</th>
<th>Dosage (ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ckap2l</td>
<td>CKAP2L</td>
<td>TGIATCAAAGTCACATGTTTCT</td>
<td>6</td>
</tr>
<tr>
<td>dennd3a</td>
<td>DENND3*</td>
<td>CAaACGTATCAGAATCAATTA</td>
<td>6</td>
</tr>
<tr>
<td>dennd3b</td>
<td>DENND3*</td>
<td>CCaATGATTTCGCTGGGA</td>
<td>6</td>
</tr>
<tr>
<td>ncl1</td>
<td>NCLN</td>
<td>GAACATCGAATTCTGGTGTTTA</td>
<td>6</td>
</tr>
<tr>
<td>nup98</td>
<td>NUP98</td>
<td>GTTACGACATTAAACCGTTTCT</td>
<td>1</td>
</tr>
<tr>
<td>tbata</td>
<td>TBATA</td>
<td>GAaACAGGCCAATCTGACCTG</td>
<td>4</td>
</tr>
<tr>
<td>p53</td>
<td>P53</td>
<td>GCACACATGTGGCAGAATTTG</td>
<td>2</td>
</tr>
<tr>
<td>HBB$^\text{b}$</td>
<td></td>
<td>CCTCTGCTCTAGGTATTAATTG</td>
<td>12</td>
</tr>
</tbody>
</table>

$^*: DENND3, DAB2IP, KDM4A and VEZF1 are duplicated in zebrafish genome. #: There was no suitable target site in tbata for translation-blocking morpholino. A second non-overlapping splice-blocking morpholino was used instead. %: Small letters indicate the mismatch nucleotides to the corresponding splice-blocking morpholino. $: morpholino against human beta-globin as an universal negative control.
### Supplementary Table 10. qPCR/RT-PCR primers.

<table>
<thead>
<tr>
<th>Target transcript</th>
<th>Forward / Reverse</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>af3</td>
<td>Forward</td>
<td>AAAGCAGCAGTCACCAGTTC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CATCTGTCACATCGCAATGC</td>
</tr>
<tr>
<td>ckap2l</td>
<td>Forward</td>
<td>TGGATCAAGCCACACAGAAG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTTCGCACAGCAAGCACAATGC</td>
</tr>
<tr>
<td>dab2ipa</td>
<td>Forward</td>
<td>TGGGACAGGATTTCTGCTTC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GCAAGACAGTCACATACATTC</td>
</tr>
<tr>
<td>dab2ipb</td>
<td>Forward</td>
<td>GCACACCTGCCACATCGCAATGG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACGGATCCATCGCAATCC</td>
</tr>
<tr>
<td>dennd3a</td>
<td>Forward</td>
<td>TGCTTGAGTGTCAAGCAGA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ATAACGCTGGGAGGGTGGCA</td>
</tr>
<tr>
<td>dennd3b</td>
<td>Forward</td>
<td>GCACGCTCTGTGATGTCCT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTTCGCCAGCTATGAGGCA</td>
</tr>
<tr>
<td>hmcn1</td>
<td>Forward</td>
<td>GAAGAAATAGCCGCTGCAAG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACGAGGTGAACTTTGAGGG</td>
</tr>
<tr>
<td>isg20l2</td>
<td>Forward</td>
<td>ACTGACGTGGAGTGGAATGC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GAGATCGATGCCACATGG</td>
</tr>
<tr>
<td>kdm4aa</td>
<td>Forward</td>
<td>GGGATGTCAGAGGACACATG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TGCTCTGGGAGGGCACAAGA</td>
</tr>
<tr>
<td>kdm4ab</td>
<td>Forward</td>
<td>TGAAAGGTATGCTGGTGGA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACGCTCCATAGATGGGAGG</td>
</tr>
<tr>
<td>ncl1</td>
<td>Forward</td>
<td>CTGGTTCTGTCGGTGGGGA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ATCACACAGCGACACTG</td>
</tr>
<tr>
<td>nup98</td>
<td>Forward</td>
<td>GAACCTGGGATCTGGTGTCT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CCACGATACGTTCTGCAAT</td>
</tr>
<tr>
<td>scube3</td>
<td>Forward</td>
<td>TCTCCGTCTGCTGGAACACC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACTCCACATGCTGCTGGTAG</td>
</tr>
<tr>
<td>tbata</td>
<td>Forward</td>
<td>CTGGAGCTGGCGCTGGAGGA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTGTGTGTGTGTCGTAGC</td>
</tr>
<tr>
<td>vezf1a</td>
<td>Forward</td>
<td>GATGGAGGTGTCACAAACCC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GCAGGCCGTATCTGACATT</td>
</tr>
<tr>
<td>vezf1b</td>
<td>Forward</td>
<td>GCAAGAGCCTATACGCTGCT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TGGGATTAAAAGCTGCTGGTC</td>
</tr>
<tr>
<td>elfa</td>
<td>Forward</td>
<td>CTTCTCAGGGCTGCTGTCG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CGCGTACGTTACCCCT</td>
</tr>
<tr>
<td>actb</td>
<td>Forward</td>
<td>TACAAATGAGCTTCGGTGTC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GTTCCCATCTCTGCTACAA</td>
</tr>
</tbody>
</table>
**DE NOVO MUTATIONS IN HSCR PATIENTS LINK CNS GENES TO THE DEVELOPMENT OF THE ENS**

**SUPPLEMENTARY FIGURES**

Supplementary Figure 1. Analytical pipeline for exome sequence filtration and prioritization. 1: GATK; 2: KGGSeq; 3: PLINK; 4: ANNOVAR. KGGSeq integrates different kinds of knowledge resources from (epi)genetic databases, pathways databases and protein-protein interaction networks to annotate the genes that harbor any post-QC variants as well as to predict the potential pathogenicity of their variants. For deleteriousness prediction, KGGSeq integrates 5 prediction programs (Polyphen2, Sift, MutationTaster, PhyloP and Likelihood ratio) which are weighted by logistic regression. Annovar is mainly used to double-check the final remaining variant for annotation, and provides supplementary features from Database of genomic variation (DGV) and clinical variation database (ClinVAR).
Supplementary Figure 2. Relatedness plotting of HSCR exome sequences. Around 17K common SNPs (minor allele frequency > 0.01 in 1000Genomes European populations) were used to calculate identical by descent (IBD) and identical by state (IBS) proportion. Each cell shows pi_hat statistics \( \hat{p} \) (IBD proportion, calculated from \( P(\text{IBD}=2)+0.5*P(\text{IBD}=1) \); http://pngu.mgh.harvard.edu/~purcell/plink/ibdibs.shtml) between two patients. No pairwise pi_hat coefficients are above 0.125 (the first cousin relationship); the light blue cells represent 0.07~0.11 for samples mainly from HK population, which is expected to be different from other European patients.
**DE NOVO MUTATIONS IN HSCR PATIENTS LINK CNS GENES TO THE DEVELOPMENT OF THE ENS**

Supplementary Figure 3. Distribution of *de novo* mutations per trio. A) Number of DNMs (separated by mutation type) in each trio, categorized into three different types (Loss of function, synonymous and others). B) Distribution of observed counts of DNMs per trio and expected counts per trio calculated from Poisson distribution (lambda at 1.2).
Supplementary Figure 4. Sanger confirmation of mosaic DNMs in DAB2IP and NCLN. Two out of 28 de novo mutations (in DAB2IP and NCLN) were confirmed as mosaic mutations by Sanger sequencing (forward and reverse Sequencing direction). A) Peak for the DAB2IP heterozygous mosaic mutation. B) Peak for the NCLN heterozygous mosaic mutation.
Supplementary Figure 5. Connection of DNM genes and ENS genes at pathway/network level. Ingenuity Pathway Analysis (IPA) was used to link 116 ENS candidate genes (left, Supplementary Table 8) with the 20 newly found genes harboring de novo mutations (right). Solid and dotted lines represent direct and indirect interactions, respectively.
Supplementary Figure 6. qPCR confirmation of gene knockdown by SBMO. Relative expression of the candidate genes between SBMO-injected (grey bar) and control morpholino-injected embryos (black bar) by qPCR.
Supplementary Figure 7. RT-PCR confirmation of ncl1 SBMO knockdown. ncl1 expressions in six 1dpf embryos injected with ncl1 SBMO were compared to control MO injected embryos. Arrow indicated the expected amplicon. L: ladder; C1: control MO-injected embryo; C2: RT negative control.

Supplementary Figure 8. RT-PCR for expression of 4 candidate genes in zebrafish. Temporal expression pattern of zebrafish orthologue genes. RT-PCR for dennd3a, dennd3b, ncl1, nup98 and tbata was performed on RNA isolated from wild type embryos at 0, 24, 48, 72, 96 and 120 hpf.
REFERENCES


DE NOVO MUTATIONS IN HSCR PATIENTS LINK CNS GENES TO THE DEVELOPMENT OF THE ENS


72. Ngan, E. S. W. et al. Prokineticin-1 (Prok-1) works coordinately with glial cell line-derived neurotrophic factor (GDNF) to mediate proliferation and differentiation of enteric neural crest.


