Multi-ancestry genome-wide association meta-analysis identifies novel associations and informs genetic risk prediction for Hirschsprung disease

Yuanxin Zhong,[a] Man-Ting So,[b] Zuyi Ma,[b] Detao Zhang,[b] Yanbing Wang,[b] Zewei Xiong,[a] João Fadista,[c] You-Qiang Song,[d] Kathryn Song-Eng Cheah,[d] Maria M. Alves,[e] Salud Borrego,[f,g] Isabella Ceccherini,[h] Mikko P. Pakarinen,[i,j] Bjarke Feenstra,[c,k] Vincent Chi-hang Lui,[b,l] Maria-Merce Garcia-Barcelo,[b] Pak Chung Sham,[a,n,*] Paul Kwong-Hang Tam,[b,m,n,**] and Clara Sze-Man Tang[b,l,n,***]

[a]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region of China
[b]Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region of China
[c]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark
[d]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region of China
[e]Department of Clinical Genetics, Erasmus University Medical Centre, Sophia Children's Hospital, Rotterdam, the Netherlands
[f]Department of Genetics, Reproduction and Fetal Medicine, Institute of Biomedicine of Seville (IBIS), University Hospital Virgen del Rocio/CSIC/University of Seville, Seville, Spain
[g]Centre for Biomedical Network Research on Rare Diseases (CIBERER), Seville, Spain
[h]IRCCS Istituto Giannina Gaslini, Genoa, Italy
[i]Section of Pediatric Surgery, Helsinki University Hospital and University of Helsinki, Finland
[j]Pediatric Liver and Gut Research Group, University of Helsinki, Finland
[k]Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[l]Dr Li Dak-Sum Research Centre, The University of Hong Kong - Karolinska Institutet Collaboration in Regenerative Medicine, Hong Kong Special Administrative Region of China
[m]Faculty of Medicine, Macau University of Science and Technology, Macao, China

eBioMedicine
2025;115: 105680
Published Online xxx
https://doi.org/10.1016/j.ebiom.2025.105680

Summary

Background Hirschsprung disease (HSCR) is a rare, congenital disease characterized by the absence of enteric ganglia in the hindgut. Common genetic variation contributes substantially to the heritability of the disease yet only three HSCR-associated loci were identified from genome-wide association studies (GWAS) thus far.

Methods We performed the largest multi-ancestry meta-analysis of GWAS to date, totalling 1250 HSCR cases and 7140 controls. Prioritized candidate genes were further characterized using single-cell transcriptomic data of developing human and mouse gut for their roles in development of enteric nervous system (ENS). Functional characterisation using human cells and zebrafish models was performed. Global and ancestry-matched polygenic risk score (PRS) models were derived and evaluated for predicting risk of HSCR.

Findings We identified four HSCR-susceptibility loci, with three loci (*JAG1*, *HAND2* and *ZNF25*) reaching genome-wide significance and one putative locus (*UNC5C*) prioritized by functional relevance. Spatiotemporal analysis revealed hotspots of gene dysregulation during ENS development. Functional analyses further demonstrated that knockdown of the candidate genes impaired cell migration and zebrafish knockouts displayed abnormal ENS development. We also demonstrated comparable performance for a PRS model derived from multi-ancestry meta-analysis to those of ancestry-matched PRS models, supporting its potential clinical application in risk prediction of HSCR across populations.

Interpretation Overall, the meta-analysis implicated novel genes, pathways and spatiotemporal developmental hotspots in the genetic aetiology of HSCR. Development of a PRS universally applicable irrespective of ancestries may leverage its clinical utility in risk prediction.

Funding The full list of funding bodies can be found in the Acknowledgements section.

*Corresponding author.
**Corresponding author. Faculty of Medicine, Macau University of Science and Technology, Macao, China.
***Corresponding author. Division of Paediatric Surgery, Department of Surgery, The University of Hong Kong, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong.
E-mail addresses: pcsham@hku.hk (P.C. Sham), pkhtam@must.edu.mo (P.K.-H. Tam), claratang@hku.hk (C.S.-M. Tang).
[n]These authors jointly supervised the study.

Keywords: Hirschsprung disease; Multi-ancestry meta-analysis; Genome-wide association study; Polygenic risk score

Research in context

Evidence before this study

Hirschsprung disease (HSCR) stands as a model for rare, complex congenital disorders with substantial genetic contribution from both common and rare variants. Thus far, seven GWAS were performed, identifying three HSCR-associated loci and implicating novel pathological pathways. Still, a large proportion of disease heritability remains unexplained.

Added value of this study

We performed the largest multi-ancestry GWAS meta-analysis of HSCR to date, identifying four genomic loci, *JAG1*, *HAND2*, *ZNF25*, and *UNC5C* associated with HSCR. Single-cell spatiotemporal analysis of candidate genes together with the functional validation in human cells and zebrafish models supported their pathological relevance in HSCR. Based on the findings of the multi-ancestry meta-analysis, an 11 SNPs PRS model was developed and was shown to have good predictive performance for risk of HSCR irrespective of ancestry.

Implications of all the available evidence

The multi-ancestry meta-analysis implicated etiological roles of dysregulation of HAND2 and ZNF25, NOTCH signalling, netrin/UNC5C ligand-receptor coupling for further functional characterization. The potential development of a universally applicable PRS irrespective of ancestries would be of clinical relevance for disease risk prediction in the future and in advancing precision medicine for HSCR.

Introduction

Hirschsprung disease (HSCR), also known as congenital intestinal aganglionosis, is a rare congenital neurocristopathy associated with the absence of enteric ganglia in the myenteric and submucosal plexuses along a variable length of the hindgut.[1] Short-segment HSCR (S-HSCR), with aganglionosis limited to the sigmoid region, is the most common subtype of HSCR, accounting for approximately 80% of cases. Other more severe but less common subtypes include long-segment HSCR (L-HSCR; ~15%) and total colonic aganglionosis (TCA; ~5%), in which the aganglionic segment extends proximal to the sigmoid colon or beyond the ileocaecal valve respectively. The estimated incidence of HSCR is 1 in 5000 live births globally but the rate varies significantly by ethnic groups, with a twofold increase in incidence among Asians compared to Europeans (28 and 15 cases per 100,000 live births among Asians and Europeans, respectively).[2,3]

HSCR is characterized by high heritability and non-Mendelian inheritance in families.[4] To date, rare damaging coding variants in more than 15 genes have been linked to HSCR[5]; however, these rare variants altogether only account for a limited portion of the overall heritability of the disease. In addition to rare variants, common regulatory variants in *RET, NRG1,* and SEMA3C/D were also identified as being associated with HSCR from multiple ancestry-specific genome-wide association studies (GWAS). In particular, a recent genetic model comprising four common variants, including three variants in *RET* and one variant in SEMAC/D loci, was shown to have a comparable, if not greater, contribution to risk of HSCR when compared to the combined effect of all rare damaging variants.[6] Despite the failure of the previous multi-ancestry meta-analysis of GWAS to uncover any novel disease-susceptibility loci for HSCR, it highlighted ancestry-specific associations and suggested that a larger meta-analysis can potentially uncover the hidden heritability.[7]

To fully exploit the heritability captured by SNP arrays, we perform the largest multi-ancestry meta-analysis to date. The meta-analysis includes all seven published GWAS on HSCR, effectively doubling the sample size to 1250 HSCR cases and 7140 controls. Through our analysis and subsequent functional characterization and validation, we identify four novel HSCR-associated loci and four secondary association signals in *RET*. We also examine the spatiotemporal expression of the HSCR-associated genes in mouse and human foetal enteric nervous system (ENS), providing new insights into disease-relevant cell types and vulnerable developmental stages in the developing gut. Additionally, we explore the ancestry-specific features of the polygenic architecture of HSCR using polygenic risk score (PRS) analyses and derive a global PRS model which may have clinical significance in disease risk prediction regardless of the population. Overall, our findings contribute to a better understanding of the genetic basis of HSCR and provide potential implications for clinical applications in the future.

Methods

Study subjects and quality controls

Raw genotyping data and/or summary statistics of GWAS were collected from the seven research centres for the multi-ancestry meta-analysis, which included 1250 HSCR cases and 7140 controls of European and

East Asian ancestries. For the ancestry-specific meta-analyses, the European-specific GWAS included 610 HSCR cases and 5658 controls whereas the Asian-specific GWAS included 640 HSCR cases and 1482 controls. With this sample size, we estimated that we have 80% power to detect association for a common variant with an odds ratio of 1.34 and a risk allele frequency of 0.63 at genome-wide significance threshold ($p = 5 \times 10^{-8}$). An overview of each study, including the genotyping array, imputation reference panel, variant quality control, number of samples and the demographics, was provided in Supplementary Table S1.

Meta-analysis

Before performing the meta-analysis, we computed the inflation factor, lambda (λ), for each of the seven datasets to ensure the data quality (Supplementary Figure S1). Due to the minor inflation (λ = 1.17) observed in the Chinese GWAS dataset, we corrected the association test statistics of this dataset by genomic control adjustment. For each SNP, we multiplied the standard error of the effect (BETAs) by the inflation factor and recalculated the corresponding association *p*-value. After the adjustment, the inflation factor of this cohort was successfully reduced to 1.002, indicating the absence of inflation. Next, we used METAL[8] to perform fixed-effect meta-analyses using the inverse-variance method. We retained only those variants that were present in at least two GWAS in the European- and Asian-specific meta-analyses and three GWAS for the multi-ancestry meta-analysis. Within each meta-analysed set, we further confirmed the absence of inflation of the meta-analysis test statistics using LD score regression (LDSC).[9] Heterogeneity of effect was assessed using Cochran's Q statistics and I^2 for the significant loci.

Heritability estimation

We estimated the heritability explained by GWAS for both European and Asian ancestries using LDSC with default settings. The estimation was based on the summary statistics for SNPs with MAF >0.01 from the respective ancestry-specific meta-analyses and the pre-computed LD scores derived from the 1000 Genomes European and East Asian data. Estimates of heritability were converted from observed scale to liability scale using a population prevalence of 0.02% for European and 0.028% for Asian, respectively.

Defining independently associated loci

We used the FUMA (Functional Mapping and Annotation) software[10] to map the significant and putative loci of variants with associations surpassing genome-wide significance threshold of $p < 5 \times 10^{-8}$ or false discovery rate (FDR) <0.001 respectively with the default setting. Here, we adopted a more rigorous suggestive threshold FDR <0.001, as opposed to FDR <0.05 suggested in other studies.[11,12] This stringent approach could increase the likelihood of detecting more true positive signals while minimizing false discoveries for functional follow-up. Briefly, independent association signals were first obtained based on the LD between markers ($r^2 < 0.8$). Independent signals with a distance less than 250 kilobases apart were then merged to form an independently associated locus. Finally, within each independent locus, the SNP with the lowest *p*-value was defined as the lead SNP.

Gene set and pathway enrichment analyses

MAGMA gene-set analysis implemented in FUMA was performed for gene set and pathway enrichment analyses. Testing was done on the curated gene sets and Gene Ontology (GO) terms from the Molecular Signatures Database (MsigDB).[13,14] We used Bonferroni correction to adjust for multiple testing. Accounting for the number of terms (pathway/gene-set) examined, genome-wide significance threshold for set-based analysis was determined at $P_{bon} = 3.2 \times 10^{-6}$ (=0.05/15,485).

Estimation of age of the European-specific variants

DMLE + version 2.3 was used to estimate the age of the two European-specific HSCR-associated variants using a Bayesian approach. The method uses Markov Chain Monte Carlo algorithm to estimate the mutation age based on the observed LD between a disease variant and linked markers of unrelated normal individuals and affected patients. Demographic parameters of the European population were applied, with an estimated incidence of 0.02%, and a population growth rate was set at 0.025.

Prioritization of candidate genes

We obtained the list of all potential candidate genes mapped with positional, expression quantitative trait loci (eQTL), and chromatin interaction evidence as implemented in FUMA functional gene mapping analysis.[10] For each putative locus at FDR <0.001, we further used ToppGene[15] to prioritize the mapped candidate genes using all available functional annotations. We defined the training set as the HSCR candidate genes summarized in our previously published review.[5] To determine the significance cut-off threshold of prioritization by chance, we tested a background set of 18,199 protein-coding genes and determined that a prioritization *p*-value of less than 1.2×10^{-4} could effectively control the type I error rate at a level below 0.05. Among the three putative loci associated at FDR <0.001, only *UNC5C* at 4q22.3 can be successfully prioritized with prioritization *p* below this cut-off threshold. To ensure the robustness of the prioritization, we further applied a stepwise strategy to move the prioritized gene from the testing to the training set and repeated the prioritization analysis until all the remaining testing genes failed to reach the significant cut-off threshold. After the

iterations, no candidate genes other than *UNC5C* can be prioritized, suggesting that *UNC5C* is the most likely HSCR gene in the locus.

Fine-mapping

Assuming one causal variant per locus, we used PAINTOR3[16] to perform cross-population fine-mapping to compute the posterior probabilities of causality (PP) and to construct the 95% credible sets for the seven HSCR-associated loci. Except the *SEMA3D* and *ZNF25* loci showing European-specific association, we performed fine-mapping using the summary statistics from the combined multi-ancestry meta-analysis and tailored the trans-ancestry LD matrices (hereinafter referred to as EUR + EAS) of the pairwise correlations between SNPs using the European (EUR) and East Asian (EAS) samples from 1000 Genomes Phase 3 reference panel. For *SEMA3D* and *ZNF25*, we computed the LD matrix using the EUR reference panel only and fine-mapping was performed on the summary statistics from the European-specific meta-analysis. All variants within 100 kb (±50 kb on each side) of the lead variant were included in the fine-mapping.

Defining significant independently associated variants

To evaluate whether there is secondary association signal for each HSCR-associated locus, we performed association analyses while conditioning on the lead variant of each ancestry using GCTA-COJO. Stepwise conditional analysis was carried out to determine all the independently associated variants in the *RET* locus. Briefly, if there were more than one variant reaching genome-wide significance, we added the variant with the lowest *p*-value to the former regression model as covariate and re-performed the conditional analyses until no secondary signal reached genome-wide significance.

Characterization of HSCR-associated genes using single cell transcriptomes of human and mouse foetal gut

To better understand the functional impact and the underlying molecular mechanisms of the four newly identified HSCR-associated genes in ENS development, we explored the spatiotemporal expression pattern of these genes in single cell transcriptomic profiles of (i) human foetal gut from Gut Cell Altas,[17] (ii) mouse developing ENS from two published datasets,[18,19] and (iii) mouse small intestine at E14.5. Gut Cell Altas includes 62,849 cells isolated from 6 to 11 weeks post–conception developing human gut, interrogating expression of 26,609 GENCODE genes, including non-coding RNA genes, in 21 intestinal cell types. We first determined if the known and novel HSCR-associated genes are differentially expressed during human ENS development in each cell type or time point compared to the remaining data using Scanpy.[20] Differentially expressed genes (DEG) were defined as having log2 fold change larger than 1 (log2 FC >1) and q-value <0.01 based on Wilcoxon rank-sum test. Hypergeometric test was then used to determine the cell type(s) or time point(s) with significant DEG enrichment of the HSCR-associated genes. For the characterization of gene expression in the ENS lineage, we used the single cell transcriptome data of ENS cells isolated from Wnt1-Cre; R26-EYFP mouse embryos at E13.5[19] and Wnt1-Cre; R26R-Tomato mouse embryos at E15.5 and E18.5.[18] Expression matrices were downloaded from SRR15465838 of Sequence Read Archive (SRA) and GSE149524 of the Gene Expression Omnibus (GEO) database respectively and processed using Seurat V4[21] according to the methods and quality control parameters stated in the original studies.

Polygenic risk score (PRS) modelling

To better understand the cumulative effects of common variants, we derived PRS models using PRSice2.[22] We used the raw genotyping data of Korean and International HSCR Genetic Consortium (IHGC) cohorts as testing datasets for Asian and European populations respectively (Supplementary Table S1). To secure independence between the training and testing cohorts, we excluded the testing dataset and re-ran the meta-analyses following the same steps mentioned above to obtain the independent GWAS summary statistics for PRS construction. For the clumping and thresholding (C + T) method, we trained the PRS models with a series of *p*-value thresholds (Pt = 1×10^{-8}, 5×10^{-8}, 1×10^{-7}, 1×10^{-6}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1) and used the ancestry-matched reference panels of the 1000 Genomes Phase 3 data (i.e., EUR, EAS, and EUR + EAS for the European, Asian, and multi-ancestry combined meta-analyses respectively) to perform LD clumping. A multiple testing correction based on permutation implemented in PRSice2 was used, and an empirical $p < 0.001$ was regarded as the significance cut-off. The prediction accuracy/performance was evaluated with Lee's R2 and Nagelkerke's pseudo R2.

In an attempt to derive a predictive model that is simple, transparent and globally applicable, we trained a set of PRS models including only the 11 variants which represented the most likely causal variants for the seven HSCR-associated loci together with the four independently associated variants from the conditional analysis of *RET*. Four 11-SNP PRS models were compared, namely (i) combined, (ii) hybrid, (iii) matched and (iv) unmatched models, with effects sizes derived from different meta-analyses. For the combined, matched and unmatched models, we extracted the effect sizes from the multi-ancestry, target ancestry-matched (e.g., Asian-specific meta-analysis as training data for Asian target samples), and target ancestry-unmatched training data (e.g., Asian-specific meta-analysis as training data for

European target samples), respectively. To train the hybrid model, we used the target ancestry-matched training data for variants showing significant heterogeneity in genetic effect and multi-ancestry training data for those variants with homogeneous genetic effect. Additionally, effect sizes of the four secondary association signals from joint conditional analysis were used in the hybrid model. The *p*-value threshold was set at 1 in PRSice2 to force the inclusion of all 11 variants. For comparison, we trained a PRS model with the 4 common variants proposed in Tilghman et al. (2019).[6] In addition, to evaluate the prediction capability of the PRS model on disease subtypes, we further stratified the Korean cases into 47 L-HSCR/TCA and 75 S-HSCR as the target testing data and assessed the performance of the PRS model separately. Area under the ROC Curve (AUC) analysis was also performed to evaluate the prediction capability of the models. DeLong test was performed to compare for the significant difference in the predictive ability of the models based on AUCs. The odds of HSCR for individuals at each decile of the PRS was compared with the mid 40–50% decile (i.e. the 3rd quantile) to obtain odds ratio estimates with 95% confidence intervals.

Immunofluorescence staining

Intestine specimens from operation of the non-HSCR paediatric patients was fixed in 4% (w/v) PFA/PBS paraformaldehyde (Sigma #P6148) for 24 h at 4 °C, dehydrated in graded series of alcohol, cleared in xylene before being embedded in paraffin. Tissue sections of 6 μm in thickness were prepared and mounted onto microscope slide. Sections were dewaxed in xylene, hydrated in a graded series of alcohol and finally in distilled water. Antigen retrieval were performed by incubation in 10 mM sodium citrate buffer (pH 6.0) at 95 °C for 10 min. After blocking in PBS-T (PBS with 0.1% Triton) supplemented with 5% donkey serum (Bio-rad #C06SB) for 1 h at room temperature, sections were incubated with primary antibodies diluted in PBS-T overnight at 4 °C. The sections were washed three times with PBS before incubation with fluorochrome-conjugated secondary antibodies diluted in PBS-T for 1 h at room temperature. After washing with PBS, sections were incubated with DAPI (Thermofisher #62248) 1:1500 in PBS for 10 min at room temperature. After washing with PBS, the slides were mounted with Prolong gold antifade reagent (Thermofisher #P36930). Sections were imaged with upright LSM900 Confocal Microscope (Imaging and Flow Cytometry Core, Centre for PanorOmic Sciences, LKS Faculty of Medicine, HKU). The primary and secondary antibodies used in the study were listed in Supplementary Table S2.

Transwell migration assay

Human neuroblastoma cell line HTB11 (ATCC: SK-N-SH) was used for the transwell migration assay to evaluate the impact of the candidate genes on cell migration. The cells were cultured in DMEM medium (Gibco) supplemented with 10% foetal bovine serum (Gibco) and 100 U/mL of penicillin, 100 μg/mL streptomycin (Gibco) at 37 °C in 5% CO2. The validation information for HTB11 is in the reagent validation file (see Supplemental Data). Small interfering RNAs (siRNAs) targeting *HAND2, UNC5C* and *ZNF25* transcripts, and a scrambled siRNA were purchased from Origene. A total of 6×10^5 cells were seeded into each well of the 6-well plates and incubated for 24 h. Cells were then transfected with either gene-specific or scrambled siRNA using Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen) according to the manufacturer's instructions. After the incubation with siRNAs for 48 h, cells were re-seeded to 24-well plates with Transwell inserts containing 8-μm pore size PET membranes (Corning Inc.). A total of 10^5 cells/well were plated into the upper chamber in serum-free medium and complete medium was added to the lower chamber. After overnight incubation, cells on the upper side of the membrane were gently wiped off with cotton swabs. Cells on the bottom surface of the membrane were fixed in 4% paraformaldehyde for 15 min and subsequently stained with 0.5% crystal violet for 15 min. Five randomly selected fields were image captured with a microscope (Nikon Eclipse E600). Number of migrated cells were counted in triplicates and t-test was used to test for the mean difference.

Zebrafish

The transgenic zebrafish *Tg(−8.3phox2bb:Kaede)* line, which expresses strong Kaede fluorescent signal to serve as a marker for enteric neurons, was used to functionally evaluate the role of the HSCR-associated genes in the ENS development. The study was conducted in compliance with the ARRIVE guidelines. Zebrafish were kept in Zebrafish Core Facility (HKU). Maintenance of zebrafish and culture of embryos were carried out as described previously.[23] Zebrafish were cultured at 28.5 °C and embryos were obtained by natural spawning and staged by days post-fertilization (dpf).

Gene knockout by CRISPR/Cas9 in zebrafish model

The three zebrafish orthologs of the HSCR-associated genes, *jag1a*, *hand2*, and *unc5c*, were selected for functional characterization. The design and synthesis of short guide (sgRNA) were carried out as described.[23] Briefly, we purchased predesigned or custom designed sgRNA for each orthologue using the Alt-R CRISPR-Cas9 system at IDT (Supplementary Table S3). Designs with high predicted efficiency and low predicted off target effect were chosen. Zebrafish embryos were randomized for the co-injection of the sgRNA (150 pg/nl) and Cas9 protein (667 pg/nl) to generate indels at the target gene at 1-cell stage. The injected larvae and uninjected control were cultured to 5dpf for phenotype

checking and imaging. Larva was collected separately for DNA extraction, T7 endonuclease I assay and Sanger sequencing to confirm the presence of indel mutation (Supplementary Figure S2). Enteric neuron counting was performed based on the Kaede fluorescent signals in the distal most 300 μm of intestine for the crispants (*hand2:* n = 80; *jag1a:* n = 66 and *unc5c:* n = 72) and wild type control (n = 63) zebrafish. While the average number of neurons counted from the intestine of a single wild type zebrafish embryo is around 85 with a standard deviation of 17. Assuming a reduction of ~20% of neurons (number of neurons ~68) for a gene knockout, we estimated that we have >90% of power to detect the association at $p < 0.01$ for >30 zebrafish crispant embryos. Kolmogorov–Smirnov test for normality was performed to ensure the count of enteric neurons following normal distribution and t-test was used to test for the mean difference in number of enteric neurons between conditions.

Ethics

Written informed consent was obtained from all human participants and all study protocols were approved by the institutional review boards of each centre.[6,24–29] The Chinese and HKUTRS GWAS were approved by the institutional review board (IRB) of the University of Hong Kong and the Hospital Authority (UW_13–225). Korean GWAS was approved by the IRB of each hospital (IRB No. SMC_2010-02-028-003 of Samsung Medical Centre; 2010-0395 of Asan Medical Centre; 4-2010-0436 of Severance Children's Hospital; 1006-129-322 of Seoul National University Children's Hospital). Danish GWAS was approved by the Danish Scientific Research Ethics Committee of the Capital Region (Copenhagen) and the Danish Data Protection Agency. The Finnish GWAS was approved by the Ethics committee of the Hospital for Children and Adolescents, University of Helsinki, by the Pirkanmaa Hospital District and by the Ethics Committee at Karolinska Institutet, Stockholm. The US GWAS was approved by the Institutional Review Board of Johns Hopkins University School of Medicine.

All animal procedures were approved by the Committee of the Use of Laboratory and Research Animals (CULATR 4871-18) at the University of Hong Kong. The study was conducted in compliance with the ARRIVE guidelines.

Role of funders

The funders of the study had no role in the study design, data collection, data analysis, interpretation, or writing of this report.

Results

Overview of the multi-ancestry meta-analysis of GWAS

To identify novel HSCR-associated loci, we first performed a multi-ancestry meta-analysis of seven published GWAS on HSCR using a fixed-effect model, totalling 1250 HSCR cases and 7140 controls (Supplementary Table S1). European- and Asian-specific meta-analyses were carried out to evaluate if there is heterogeneity in genetic effects between populations. Using LDSC, we estimated that the proportions of variance explained by the current study were 27.9% ($p = 0.059$, linear regression) and 24.7% ($p = 0.049$) for Europeans and Asians respectively. The consistency in the ancestry-specific estimates implied a comparable overall contribution of common variants across populations.

Among the 8.6 million variants tested, we identified six genome-wide significant loci at $p < 5 \times 10^{-8}$ and three putative loci at a FDR of 0.001 (Fig. 1a; Supplementary Table S4). Three of the GWAS-significant loci were known to be associated with HSCR, including *RET*, *NRG1* and *SEMA3D* (Supplementary Figures S3–S8). As reported previously, heterogeneity in effect were detected for these known loci.[7] The *SEMA3D* locus showed a strong European-specific effect (Fig. 1b), with the lead variant being monomorphic in Asians. In contrast, the associations of *RET* and *NRG1* were detected in both populations albeit with significantly different effect sizes ($p < 0.01$ by Cochran's Q test). In Europeans, the genetic effect of the *RET* locus was generally larger, while the effect of the *NRG1* locus was smaller compared to the Asian population (Table 1). Gene set enrichment analysis by MAGMA highlighted significant enrichment for HSCR-relevant GO terms of "neural crest cell migration involved in autonomic nervous system development" ($P_{bon} = 0.014$ after Bonferroni correction) and "transmembrane receptor protein tyrosine kinase activator activity" ($P_{bon} = 0.044$).

Novel HSCR-associated loci at JAG1, ZNF25, HAND2, and UNC5C

We next mapped the candidate genes of the three novel GWAS-significant loci based on positional, eQTL, and chromatin interaction mapping using FUMA. The strongest novel association was detected for rs6032951 at 20p12.2 ($p = 2.8 \times 10^{-16}$; inverse-variance weighted fixed-effect meta-analysis OR [95% CI] = 1.6 [1.4–1.8]; Fig. 1c; Supplementary Figures S9 and S10), which mapped to *JAG1* encoding a NOTCH1 ligand. The other two GWAS-significant loci were mapped to *ZNF25* (rs79812746: $p = 1.3 \times 10^{-11}$; OR [95% CI] = 6.8 [3.9–11.8]) and *HAND2* (rs62337918: $p = 1.0 \times 10^{-10}$; OR [95% CI] = 1.5 [1.3–1.7]) based on eQTL mapping (Fig. 1d and e; Supplementary Figures S11–S14). Of note, like *SEMA3D*, association of *ZNF25* is European-specific and contributed mainly by low frequency variants (MAF~0.017) in Europeans. Through evolution analysis, we estimated that these European-specific variants could have emerged approximately 414 generations ago, supporting the presence of a founder effect within the European population (Supplementary Figure S15).

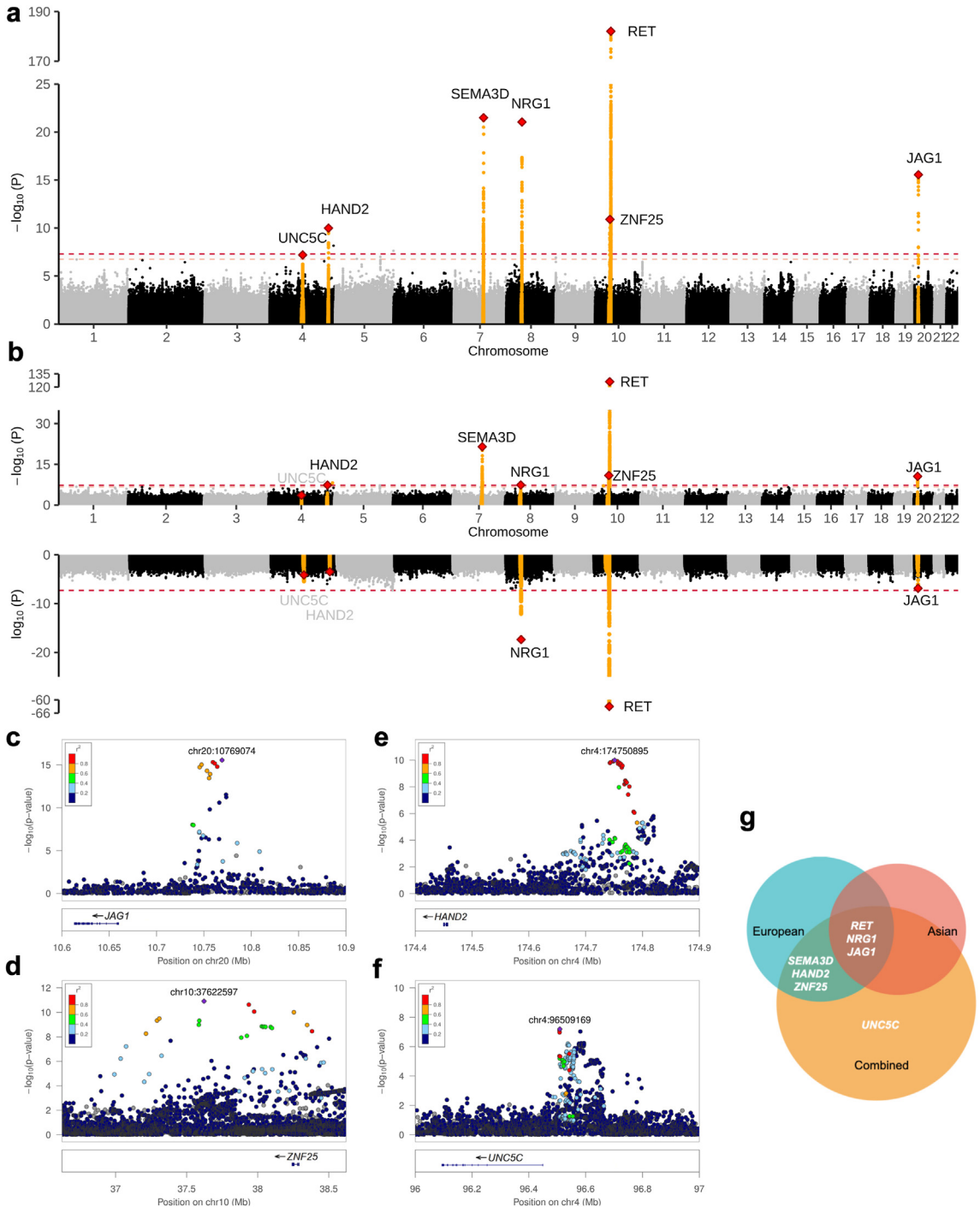


Fig. 1: **Association results of multi-ancestry and ancestry-specific meta-analyses**. Lead variant in each locus is marked as a red diamond. (a) Manhattan plot of the multi-ancestry meta-analysis. The red line represents genome-wide significant threshold of $p = 5 \times 10^{-8}$ and the orange horizontal line represents FDR = 0.001. Candidate genes with lead SNP surpassing genome-wide significance and the prioritized putative gene are shown. (b) Manhattan plots of European-specific (upper panel) and Asian-specific meta-analyses (lower panel). (c–f) Regional LocusZoom plots for *JAG1*, *ZNF25*, *HAND2*, and *UNC5C*, respectively for the multi-ancestry meta-analysis. (g) Venn diagram showing the overlapping candidate genes from the three meta-analyses of GWAS.

Variant	Position	Alleles[a]	Multi-ancestry meta-analysis		European-specific meta-analysis			Asian-specific meta-analysis			Gene[c]
			p	OR (95% CI)	EAF[b]	*p*	OR (95% CI)	EAF	*p*	OR (95% CI)	
GWAS-significant loci											
rs2506008	10:43,580,224	C/A	8.3×10^{-183}	4.3 (3.9, 4.7)	0.30	3.2×10^{-125}	5.3 (4.6, 6.1)	0.58	9.5×10^{-64}	3.4 (2.9, 3.9)	*RET*
rs117617821	7:84,322,189	C/T	3.2×10^{-22}	4.2 (3.1, 5.6)	0.052	3.2×10^{-22}	4.2 (3.1, 5.6)	0	–	–	*SEMA3D*
rs7005606	8:32,401,501	G/T	9.0×10^{-22}	1.7 (1.5, 1.8)	0.49	8.1×10^{-8}	1.4 (1.3, 1.6)	0.24	4.5×10^{-18}	2.1 (1.8, 2.5)	*NRG1*
rs6032951	20:10,769,074	T/G	2.8×10^{-16}	1.6 (1.4, 1.8)	0.25	2.6×10^{-11}	1.7 (1.5, 2.0)	0.34	9.1×10^{-7}	1.5 (1.3, 1.7)	***JAG1***
rs79812746	10:37,622,597	A/T	1.2×10^{-11}	6.8 (3.9, 11.8)	0.017	1.3×10^{-11}	6.8 (3.9, 11.8)	0	–	–	***ZNF25***
rs62337918	4:174,750,895	G/C	1.0×10^{-10}	1.5 (1.3, 1.6)	0.25	6.8×10^{-8}	1.5 (1.3, 1.7)	0.19	3.3×10^{-4}	1.4 (1.2, 1.7)	***HAND2***
Prioritized putative loci at FDR <0.001											
rs6532569	4:96,509,169	T/C	6.3×10^{-8}	1.3 (1.2, 1.5)	0.76	2.4×10^{-4}	1.3 (1.1, 1.6)	0.49	7.3×10^{-5}	1.3 (1.2, 1.5)	***UNC5C***

[a]Alleles: effective/non-effective allele. [b]EAF: allele frequency for the effective allele. [c]Novel loci are highlighted in bold.

Table 1: **Association results of the top associated loci from the multi-ancestry and ancestry-specific meta-analyses of genome-wide association studies.**

For the three putative loci associated at FDR <0.001, we performed a candidate gene prioritization using ToppGene with respect to their similarity in functional annotations to reported known HSCR genes. Based on the biological and functional evidence, we prioritized one candidate gene, *UNC5C*, at 4q22.3 encoding a receptor that binds to the laminin-related secretory protein, netrin-1 (NTN1) (rs6532569: $p = 6.3 \times 10^{-8}$; OR [95% CI] = 1.3 [1.2–1.5]) (Fig. 1f; Supplementary Figures S16 and S17). For these three novel HSCR-associated loci besides *ZNF25*, no heterogeneity in genetic effect was detected across populations (Cochran's Q test $p > 0.1$).

Secondary association signal analysis at RET

Complex genetic locus may harbour multiple causal variants that are collectively detected as a strong association signal in an independently associated locus. To investigate the presence of secondary association signals, we conducted conditional analysis using GCTA-COJO within each of the three known and the four newly identified HSCR-associated loci. Notably, in line with the observed heterogeneity in effect for the lead SNPs, significant difference in the secondary association signals were observed across populations specifically for the *RET* locus. In *RET*, three independently associated variants (rs1774171, rs115530198, and rs10899830) were detected in Europeans, whereas only one variant (rs140184225) was shown to have independent effect in the Asian population (Table 2). Two of these variants, rs115530198 and rs140184225 discovered in Europeans and Asians respectively, are ancestry-specific and show very low allele frequencies (MAF <0.002) in another population. The other two independently associated variants also demonstrated much larger effect among Europeans than Asians. Apart from the *RET* locus, no secondary association signal was detected in the other loci. Altogether, these findings demonstrated that genetic heterogeneity substantially underlies HSCR and such heterogeneity must be considered when modelling genetic risk of HSCR to improve the generalizability for prediction.

Fine-mapping

Fine-mapping of the truly causal variants within the disease-susceptibility loci can often help delineate the underlying molecular pathomechanisms. Taking advantage of our diverse multi-ancestry GWAS cohorts, we further performed cross-population fine-mapping analysis using PAINTOR3. By leveraging the variability in the local LD patterns across populations, we successfully fine-mapped these HSCR-associated loci, resulting in the 95% credible sets containing a median of eight causal variants (Supplementary Table S5). In

Ancestry	Variant	Position	Alleles[a]	EAF[b]		Ancestry-specific meta-analysis		Ancestry-specific joint conditional analysis[c]	
				European	Asian	*p*	OR (95% CI)	*p*	OR (95% CI)
European	rs1774171	10:43,434,211	C/T	0.5944	0.7202	5.6×10^{-32}	2.2 (2.0, 2.6)	1.1×10^{-19}	1.9 (1.6, 2.2)
	rs115530198	10:43,587,326	T/C	0.0139	0	4.0×10^{-8}	4.1 (2.4, 6.8)	1.5×10^{-10}	5.2 (3.2, 8.7)
	rs10899830	10:44,069,863	A/G	0.1093	0.0605	3.4×10^{-12}	2.0 (1.7, 2.5)	9.9×10^{-9}	1.8 (1.5, 2.3)
Asian	rs140184225	10:43,759,346	T/C	0	0.0278	1.4×10^{-5}	2.3 (1.6, 3.3)	5.5×10^{-9}	3.1 (2.1, 4.4)

[a]Alleles: effective/non-effective allele. [b]EAF: allele frequency for the effective allele. [c]Joint conditional analysis result of the independently associated variants identified from stepwise analysis together with the lead SNP.

Table 2: **Secondary association signals in *RET* from ancestry-specific conditional analyses.**

particular, we narrowed down the causal variant of *NRG1* to single-marker resolution, with the lead SNP having posterior probability of causality (PP) > 0.99. In addition, the lead SNPs of *SEMA3D*, *JAG1*, and *ZNF25* also showed high posterior probabilities (PP > 0.49) and were likely to be responsible for the association signal in the corresponding loci. The full list of all likely causal SNPs within the 95% credible set was included in Supplementary Table S5.

Functional annotation at single-cell level revealed spatiotemporal hotspots for HSCR-associated genes and provided mechanistic insights

Among the four newly identified HSCR-associated genes, only *HAND2* was previously reported to be linked to ENS development in animal models. *HAND2* encodes a transcription factor shown to promote neurogenesis of bipotent neural crest cells.[30] Conditional knockout of *Hand2* in the neural crest cells resulted in reduced numbers of enteric neurons, which prolonged gastrointestinal transit times and impaired colonic motility.[31,32] Immunofluorescence staining of these four HSCR-associated genes in intestine of non-HSCR controls revealed that all four genes were expressed in enteric neurons of the myenteric and submucosal plexuses (Fig. 2). Like HAND2, the three novel HSCR candidate genes are not neuron-specific and appeared to be broadly expressed in various intestinal cell types. Specifically, JAG1 and UNC5C exhibited strong expression in the intestinal stroma, which is consistent with their expression patterns reported in the single cell transcriptomic data from the Gut Cell Altas (Supplementary Figure S18). Additionally, we also confirmed that JAG1 is expressed in GFAP^{+} enteric glial cells (Supplementary Figures S18 and S19a). Such patterns of expression were consistently observed in developing mouse small intestine at E14.5 (Supplementary Figures S19b, S20a and S21).

To explore the molecular function of the HSCR-associated genes in human gut development, we analysed the Gut Cell Atlas characterizing the transcriptome of the 6–11 weeks post–conception developing human gut.[17] First, we examined the spatiotemporal expression of all seven HSCR-associated genes with respect to the 21 cell types and nine developmental time points of the developing human gut. After correcting for multiple testing, significant polygenic enrichment of these genes was found primarily in cells of the ENS lineage, including enteric neurons ($p = 1.7 \times 10^{-9}$, hypergeometric test) and neural crest cells ($p = 0.0099$). Besides recapitulating the known cell type–disease association, a mesenchymal subset, known

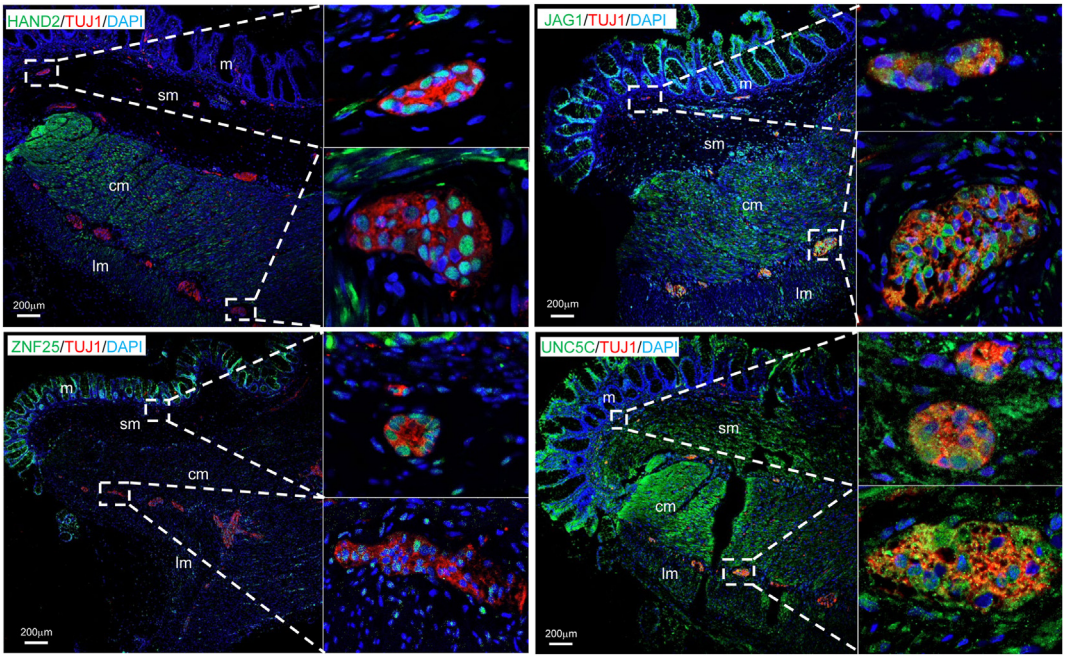


Fig. 2: **Immunofluorescence staining of the four newly identified HSCR-associated genes showing colocalization of expression in human ENS.** Representative immunofluorescence confocal images of HAND2, JAG1, ZNF25 and UNC5C (green) in the intestine tissues of non-HSCR controls, showing colocalization with TUJ1 (red) in enteric neurons in the ganglion cells of the submucosal and myenteric plexuses. DAPI nuclear staining is also shown (blue). The images at right panel represent the zoomed-in views of selected ganglion cells in submucosal (upper right) and myenteric (bottom right) plexuses as indicated by the dotted boxes. m: mucosa; sm: submucosa; cm: circular muscle layer; lm: longitudinal muscle layer. m, mucosa; sm, submucosa; cm, circular muscle; lm, longitudinal muscle.

as FRZB fibroblast cells, was also enriched with HSCR-associated genes identified from the meta-analysis ($p = 8.4 \times 10^{-5}$) (Fig. 3a). Temporally, we identified two developmental hotspots, F6.1 ($p = 0.0041$) and F8.4 ($p = 1.7 \times 10^{-5}$), where the HSCR-associated genes showed overexpression (Fig. 3b). Interestingly, the FRZB fibroblast cells, characterized by high expression of *FRZB* and *ZEB2*, were exclusively detected in foetal

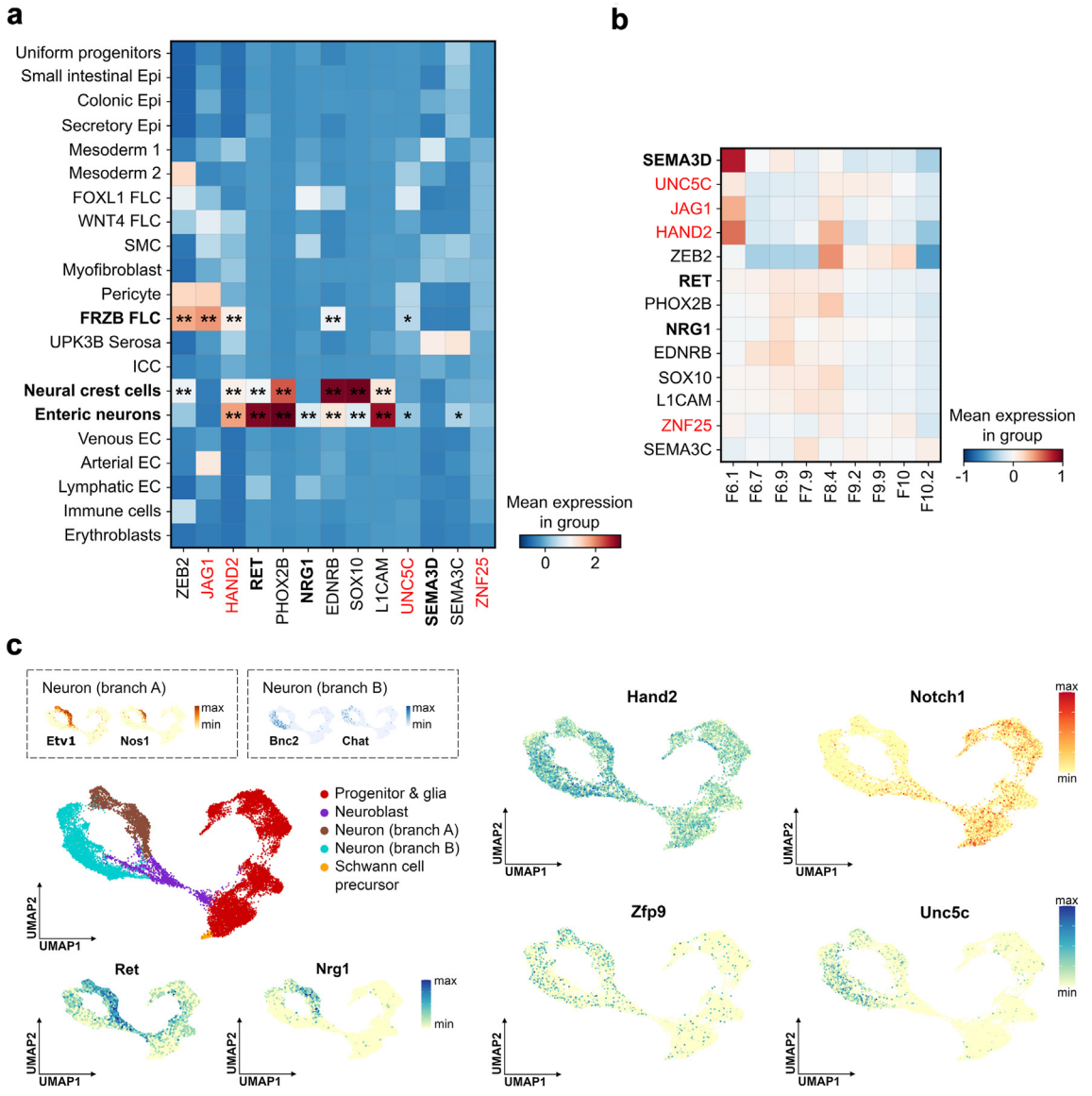


Fig. 3: **Spatiotemporal expression pattern of known HSCR genes and newly identified HSCR-associated genes in the developing intestine of human and mice foetuses**. (a) Spatial expression of the HSCR-associated genes. The three cell types (neural crest cells, enteric neurons, and FRZB fibroblast cells) in human developing intestine enriched with HSCR-associated genes are highlighted in bold on the y-axis. Known and novel HSCR-associated genes identified from the current meta-analysis are highlighted in bold and red respectively on the x-axis. Differentially expressed genes (log2 fold change>1 and adjusted *p*-value <0.01, Wilcoxon rank-sum test) are marked by * ($p < 0.01$) and ** ($p < 0.001$) in the corresponding cell type. Epi, epithelial cells; SMC, smooth muscle cells; FLC, fibroblast cells; ICC, interstitial cells of Cajal; EC, endothelial cells; (b) Temporal expression of the HSCR-associated genes. Candidate risk genes demonstrated signals of overexpression primarily at two developmental (F6.1 and F8.4) time points. (c) Expression patterns for the HSCR-associated genes in the developing ENS of mice using single cell transcriptome integrated from data at E13.5, E15.5 and E18.5. *Zfp9* represents the mouse orthologue to human *ZNF25*. *Jag1* is mainly expressed in the mesenchyme and thus the expression pattern of its receptor, *Notch1*, is shown. Branch A and B neurons were defined analogously to human Gut Cell Altas based on the cell-type specific markers of *Etv1* and *Bnc2* respectively. UMAP plots stratified by three developmental time points are included in Supplementary Figure S22.

colon at F6.1, whereby the ENS cells were predominately migrating/proliferating NCC regionally.[33] Three of the newly identified HSCR-associated genes, *JAG1*, *HAND2* and *UNC5C*, were also differentially expressed in this cell type at F6.1. In particular, *JAG1* was mainly expressed in the foetal gut mesenchyme and showed the strongest upregulation in these FRZB fibroblast cells. Synergistically, upregulation of the receptor of *JAG1*, *NOTCH1*, was observed in both NCC (log2 fold change (FC) = 2.44; $P_{adjust} = 9.6 \times 10^{-299}$) and enteric neurons (log2 FC = 1.22; $P_{adjust} = 4.7 \times 10^{-26}$) at F6.1 and F8.4 as well; however, whether and how the interaction of these ligands and receptors between the NCC, enteric neurons, and FRZB fibroblast cells affects the enteric NCC proliferation/migration or neurogenesis remains largely unclear.

To further delineate the role of these novel HSCR-associated genes in enteric neurogenesis and neuron differentiation, we explored the differentiation trajectory of the enteric NCCs in developing mouse ENS.[18,19] All the generic cell states of the developing ENS, from progenitor, neuroblast, to the two major neuronal subtypes, were represented in various proportions in the two published datasets of three time points (E13.5, E15.5 and E18.5; Supplementary Figure S22). By integrating these datasets, we analysed the dynamic expression pattern of the novel HSCR genes. Like the known ENS transcription factor *Phox2b*, *Hand2* was ubiquitously expressed along the differentiation trajectory, from cycling enteric NCCs to differentiated enteric neurons of both branches (Fig. 3c; Supplementary Figure S22). Additionally, while the expression of *Jag1* was low in mouse ENS, its receptor *Notch1* was highly expressed in the neural progenitors but gradually downregulated after branching. This suggested that JAG1/NOTCH1 signalling may be important in maintaining the undifferentiated progenitor state of the enteric NCCs. In contrast, we observed that *Zfp9*, the mouse orthologue of *ZNF25*, exhibited increased expression during neurogenesis and this enriched expression persisted in both branches of differentiated neurons. Intriguingly, the expression of *Unc5c* was mainly limited to the enteric neurons. Unlike *Ret* and *Nrg1* which were strongly expressed in nitric oxide synthase positive (*Nos1*+ Branch A neuron) inhibitory neurons, *Unc5c* was predominately expressed in excitatory neurons that co-express choline acetyltransferase (*Chat*) (Branch B neuron; Fig. 3c and Supplementary Figure S22). The distinct pattern of differentiation suggested that NTN/UNC5C signalling may affect ENS development antagonistically or independently from the RET signalling pathway. Meanwhile, it also highlighted that defective differentiation into functional excitatory motor neurons could underly the genetic aetiology of HSCR.

Inhibition of the novel HSCR-associated genes disrupted cell migration *in vitro*

Given the potential significance of these novel HSCR-associated genes in ENS development, we sought to evaluate their roles in regulating cell migration *in vitro*. As *JAG1* appeared to be predominantly expressed in the mesenchyme during human gut development and we consistently did not observe colocalization of expression of Jag1 in $p75^{NTR+}$ enteric NCCs in developing mouse intestine (Supplementary Figure S20b), we focused on assessing the effect of inhibiting the expression of the other three HSCR-associated genes, *HAND2*, *ZNF25*, and *UNC5C*, on cell migration. We knocked down these three genes using siRNA in SK-N-SH cells. Transwell migration assay showed delayed migration with decreased expression of *HAND2* ($p = 0.020$, t-test) and *ZNF25* ($p = 0.020$) (Fig. 4). Consisted with the distinct expression pattern of *UNC5C* in differentiated neurons following neuronal commitment, we did not observe significant difference in cell migration capacity upon downregulating *UNC5C* ($p = 0.35$). Furthermore, immunofluorescence staining further confirmed the absence of Unc5c expression in migrating enteric NCCs (Supplementary Figure S20c). Altogether our results suggested that genetic dysregulation of *HAND2*, *ZNF25*, and *JAG1* could impair enteric NCC migration, potentially leading to abnormalities in ENS development and increased HSCR susceptibility.

Knockout of novel HSCR-associated genes resulted in abnormal ENS development in zebrafish

Zebrafish has been a valuable model organism for studying the molecular function of HSCR candidate genes in ENS development.[34] Previous studies using zebrafish models demonstrated that *ret* homozygous mutants displayed HSCR-like phenotype, characterized by loss of enteric neurons from distal to proximal intestine, whereas knockdown of *nrg1*, *sema3c*, and *sema3d* led to reduction of enteric neurons.[24,35] We thus assessed the effect of depletion of these newly identified HSCR-associated genes on ENS development using the *Tg(−8.3phox2bb:Kaede)* transgenic zebrafish. This reporter line allows fluorescent Kaede labelling of the neural crest-derived enteric neuron precursors and differentiated enteric neurons for screening HSCR-like ENS phenotypes. As there is no zebrafish orthologue for *ZNF25*, CRISPR/Cas9-mediated knockout was performed only for the remaining three candidate genes. Although all larvae of *hand2*, *jag1a*, and *unc5c* crispants displayed complete innervation of the gut up to the anal opening, significant reduction in neuronal density were observed (all $p < 2.2 \times 10^{-16}$; Fig. 5a and b). On average, the numbers of enteric neurons were reduced by half in these crispant lines. Notably, the crispant larvae with greater numbers of neurons (above the upper quantile; Fig. 5a, left panel) exhibited ENS phenotypes resembling those seen below the lower quantile of the controls (Fig. 5a, top right panel). The result provided additional evidence supporting the genotype–phenotype association, whereby the dysregulation of these genes may contribute to an increased risk of developing HSCR.

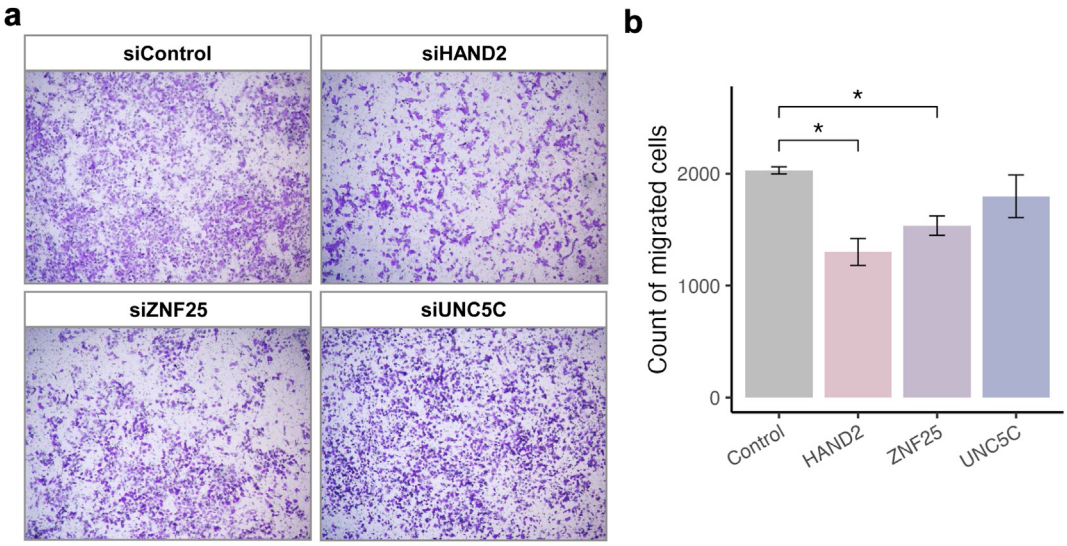


***Fig. 4:* Transwell migration assay showing migration defects with downregulation of the HSCR-associated genes.** (a) Representative images and quantification of the migrated cells are shown. (b) Number of migrated cells was counted in triplicates and is presented as mean ± standard error. Symbols denote the statistical significance of t-test (*$p < 0.05$, **$p < 0.01$).

PRS prediction models

PRS summarizes the combined effect of many genetic variants derived from GWAS and can provide individualized polygenic risk prediction for complex diseases. Despite having an overall similar contribution of common variants, our meta-analysis demonstrated an evident genetic heterogeneity for HSCR across populations. Such heterogeneity may reduce the performance and transferability of PRS in predicting genetic risk of HSCR. With the aim of better understanding the polygenic component of HSCR and to develop a global PRS applicable for genetic risk prediction across populations, we derived and compared PRS using 2 methods: (i) clumping and thresholding (C + T) and (ii) 11 SNPs models aggregating the genetic effect of the 11 independently associated variants across seven HSCR-associated loci (see Methods).

The C + T method is a traditional method commonly used for selecting independent variants in PRS construction. For the PRS derived using this method, the use of ancestry-matched training data gave optimal prediction in both Asian and European populations (Fig. 6a and b; Supplementary Table S6). However, in Europeans, PRS derived from the multi-ancestry combined meta-analysis performed equally well as those derived from the ancestry-matched meta-analysis data. Also, in Europeans, the performance of C + T PRS models using ancestry-matched training data was optimal at a lenient *p*-value threshold of 1×10^{-6}, suggesting the existence of additional HSCR-associated variants not reaching genome-wide significance.

Though the ancestry-specific PRS models using the C + T method achieved good prediction performance, the use of different variants (7 SNPs for Asians and 66 SNPs for Europeans) and different effect sizes across populations limited the generalizability of these PRS models to other ethnic groups or admixed populations. To address this limitation, we thus considered a simplified 11-SNP model comprising only the 11 SNPs independently associated with HSCR and tested the use of different effect sizes trained from the ancestry-matched/unmatched and the multi-ancestry meta-analyses (see Methods). Alternatively, we also considered a 11-SNP hybrid model using effect sizes derived from the ancestry-matched meta-analyses for variants with heterogeneous genetic effect across populations or otherwise from the multi-ancestry combined meta-analysis. Overall, using ancestry-matched training data, the 11-SNP models achieved prediction performance that are highly comparable with the PRS models produced by C + T method (Supplementary Table S7). Intriguingly, the two models (combined and hybrid) trained on the multi-ancestry meta-analysis demonstrated the highest predictive power (Supplementary Tables S7 and S8). Though the hybrid model marginally outperformed the combined model on both Asians and Europeans, the 11-SNP combined model can provide similar predictive accuracy when the underlying ancestry of the target samples is uncertain. Regardless of the inclusion of *RET* variants showing secondary association, we observed a significantly higher PRS for HSCR cases across groups stratified by the genotype at the lead SNP (rs2506008) of *RET*, which demonstrates the importance of the non-*RET* loci in predicting genetic risk of HSCR (Supplementary Table S9). Compared with the previously proposed 4-SNP combined model in

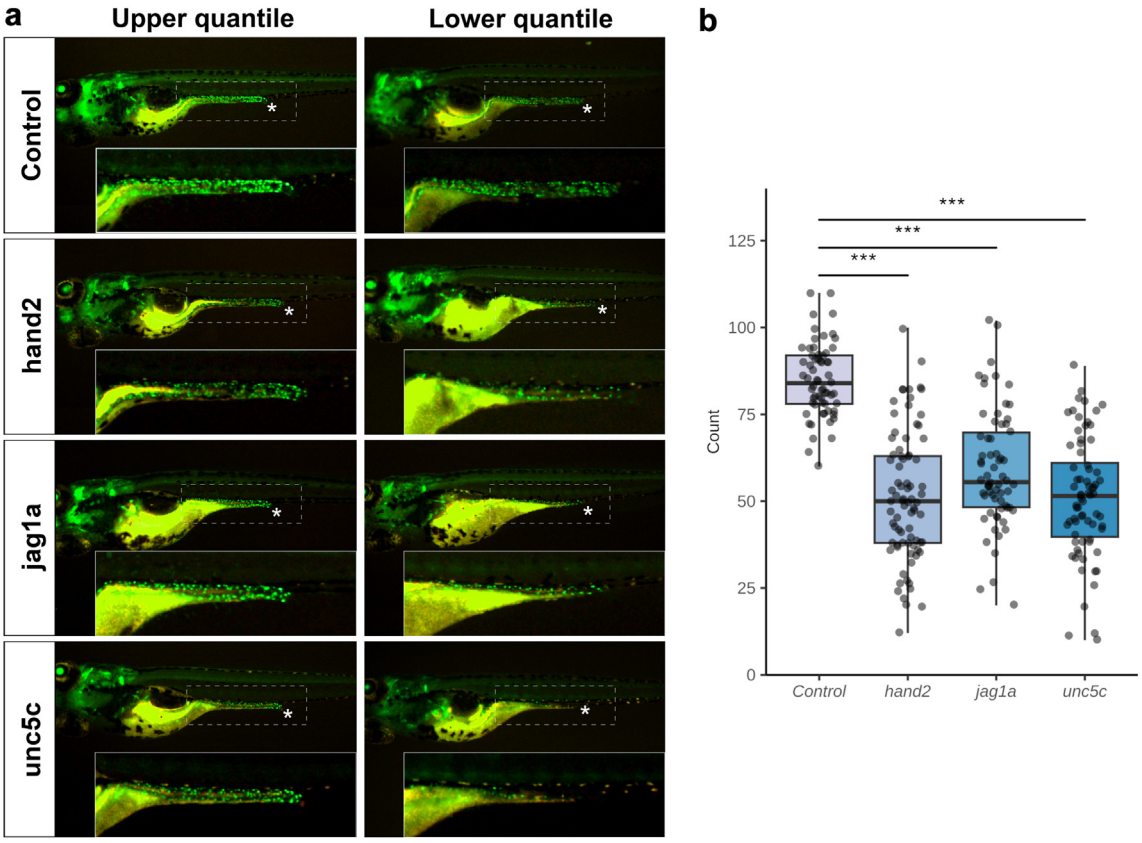
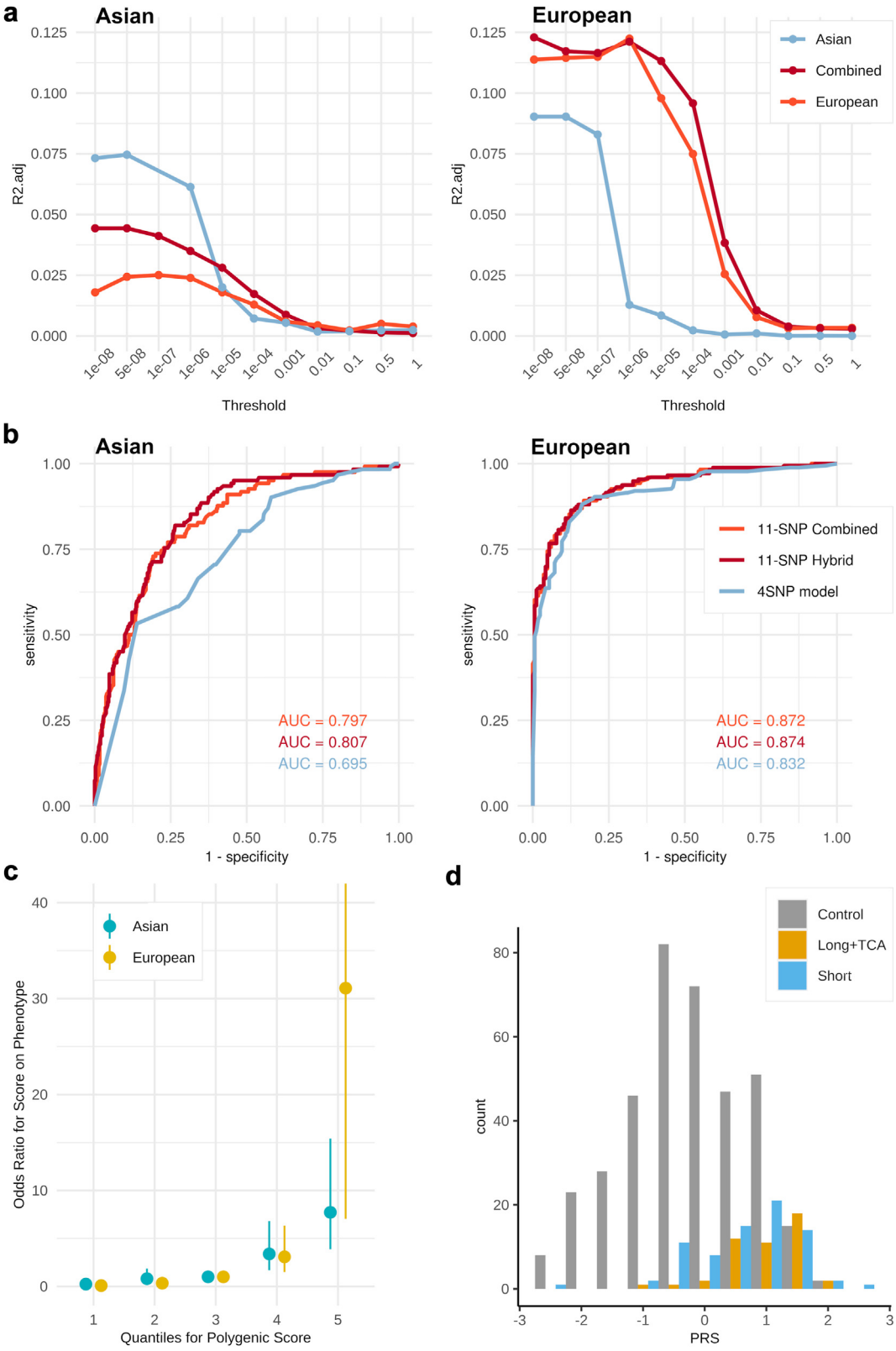


Fig. 5: **CRISPR/Cas9 knockout of *hand2*, *jag1a*, and *unc5c* impaired ENS development and resulted in decreased enteric neuronal number.** (a) Representative images of the wild type zebrafish (control), the *hand2*, *jag1a* and *unc5c* crispant zebrafish with number of neurons above the upper (left) and below the lower quantile (right) are shown. The image at the bottom right panel represents the zoomed-in view (indicated by the dotted box) visualizing the Kaede-expressing enteric neurons (green dots) in the intestine. Asterisk (*) indicates end of the gut tube (anal pore). (b) Significant reduction in the number of enteric neurons in *hand2* (n = 80), *jag1a* (n = 66) and *unc5c* (n = 72) crispants compared to wild type control fish (n = 63). *** T-test $p < 0.001$.

Tilghman et al. (2019),[6] the 11-SNP combined PRS model significantly improved risk prediction of HSCR in Asian population (AUC from 0.70 to 0.81; r^2 from 0.032 to 0.103; DeLong test $p < 1 \times 10^{-5}$; Supplementary Table S7) and modestly improved the prediction in European population (AUC from 0.83 to 0.87; r^2 from 0.12 to 0.17; $p < 0.01$) (Fig. 6b). Based on this combined model, the odds ratio of HSCR for the top decile of PRS (relative to the 3rd decile) was much higher than the odds ratios of other deciles, being 31.1 (95% CI 7.0–137.2) for the European and 7.7 (95% CI 3.9–15.4) for the Asian test samples (Fig. 6c). A similar pattern of association between risk of HSCR and deciles of PRS was also observed for the best-performing C + T models, which further demonstrated the robustness and the generalizability of the 11-SNP combined PRS (Supplementary Figure S23). Lastly, we observed a similar distribution of PRS across HSCR subtypes, indicating that this 11-SNP combined PRS model was not only applicable to predict genetic risk of the most common form of S-HSCR but also equally powerful to predict the risk of the rarer, severe subtype of L-HSCR and TCA (Fig. 6d).

Discussion

Here, we combined the findings of all available GWAS on HSCR to present the largest multi-ancestry meta-analysis on this disease. In addition to confirming previous associations, we identified four novel HSCR-susceptibility loci, including *JAG1*, *ZNF25*, *HAND2* and *UNC5C*, and four secondary association signals in *RET*. By comparing the results between the two ancestry-specific meta-analyses, we detected genetic heterogeneity in effects specifically in *RET* and *NRG1* between populations. Meanwhile, we highlighted the European-specific association of *ZNF25* in addition to the known HSCR-associated *SEMA3D* locus. To gain mechanistic insights into the genetic predisposition, we performed functional annotation of the HSCR-


a
Asian
European
R2.adj
Threshold
Asian
Combined
European
b
Asian
European
sensitivity
1 - specificity
AUC = 0.797
AUC = 0.807
AUC = 0.695
AUC = 0.872
AUC = 0.874
AUC = 0.832
11-SNP Combined
11-SNP Hybrid
4SNP model
c
Odds Ratio for Score on Phenotype
Quantiles for Polygenic Score
Asian
European
d
count
PRS
Control
Long+TCA
Short

associated genes using single cell transcriptomic data of human and mouse developing gut, which suggested spatiotemporal hotspots and novel biological pathways underlying the defective ENS development in patients with HSCR. Functional assays demonstrated that dysregulation of expression of these novel HSCR-associated genes impaired cell migration and resulted in abnormal ENS development in zebrafish model. Lastly, to assess the clinical utility of our findings, we developed a global PRS model aggregating the genetic effects of the most likely causal variants strongly associated with HSCR as well as the secondary association signals from the meta-analysis. The global 11-SNP PRS model demonstrated a substantially improved predictive ability on disease risk, particularly on East Asian populations, suggesting the potential of future clinical implementation of PRS in improving genetic counselling.

Major HSCR genes, like *RET, PHOX2B* and *SOX10,* were primarily expressed in cells of the ENS lineage. Upon entering the foregut, the spatiotemporal expression of these ENS genes together with their interacting partners modulate the migration/proliferation/differentiation behaviour of the enteric NCC through sensor-effector coupling. The identification of a subtype of fibroblast cells (FRZB fibroblast cells) uniquely present during colonization of the colon which are enriched with HSCR-associated genes suggested an underestimated importance of these fibroblast cells in supporting ENS development. Functional characterization of interaction between these fibroblast cells, the ENS precursors and enteric neurons in model organisms may provide clues into the underlying molecular mechanisms dysregulating ENS development in patients with HSCR. Of particular interest, these FRZB fibroblast cells strongly expressed one of the novel HSCR genes, *JAG1*. JAG1 expresses mainly in the mesenchyme and is a crucial member of the NOTCH signalling pathway implicated in the regulation of proliferation and differentiation of the enteric NCCs. In addition, reduced expression of *NOTCH1* was detected in the aganglionic segments of HSCR bowel.[36] Our previous pathway-based analysis uncovered that disruption of the epistasis between *PTCH1* of the Hedgehog signalling and *DLL3* of the NOTCH signalling may increase risk of HSCR and suggested a novel disease mechanism of Hedgehog/Notch-induced premature gliogenesis.[37] In both mouse and human developing ENS, *NOTCH1/2* is highly expressed in NCC while some NOTCH ligands, like *DLL1* and *DLL3*, are markers of neuroblasts governing neurogenesis; therefore, the early expression of *JAG1* in the developing intestinal stromal niche may help co-ordinate stem cell maintenance or binary neurogenic branching via interaction with other NOTCH signalling partners. Further study to elucidate the underlying molecular mechanism of disease pathogenesis in this locus is warranted.

In fact, in adult rat ENS, *Notch1* is preferentially expressed in the cholinergic neurons but not in the *Nos1*$^{+}$ inhibitory neurons.[38] Coincidentally, our novel HSCR gene, *UNC5C*, was also preferentially expressed in excitatory neurons lacking *NOS1*. Like other netrin receptors, *UNC5C* mediates chemoattraction or chemorepulsion upon NTN1 binding to guide cell migration as well as axon guidance during nervous system development.[39] Based on the published microarray-based expression profiles of the mouse intestine at E11.5 and E15.5,[40] although *UNC5C* was not directly implicated in ENS development, other ligand-receptor couples of the netrin signalling pathway, e.g., *Ntn4-Unc5d*, and *Ntng1-Lrrc4c*, were implicated in the orchestration of neuronal migration and synapse formation of the enteric networks in the developing mice ENS. The distinct expression pattern of *Unc5c* in the developing ENS contrasts with those of *Ret* and *Nrg1,* suggesting that NTN/UNC5C signalling may affect ENS development antagonistically or independently from RET signalling pathway and may have a complementary role in neurogenesis and maturation of the excitatory neurons. It also raised the possibility that genetic variants dysregulating the neuronal specification of excitatory neurons may also underlie disease pathogenesis.

Both HAND2 and ZNF25, the other two newly identified HSCR-associated genes, function as transcription factors. Consistent with the widespread expression of *Hand2* in developing mouse ENS, *HAND2* was found to be necessary for the specification of neurons that produce the vasoactive intestinal peptide (VIP), as well as their nitrergic and calretinin-expressing equivalents.[41] Knockdown of *Hand2* in enteric crest-derived cells was shown to prevent neuronal development. Conditional inactivation of *Hand2* in migrating NCC-derived cells resulted in absence of terminally differentiated enteric neurons.[42] The current study provided compelling evidence supporting the association of *HAND2* with HSCR from a common-variant perspective. Unlike *HAND2,* the precise molecular function of

Fig. 6: **Risk stratification of HSCR by global and ancestry-specific polygenic risk score. (a)** Proportion of risk variance in Asian (left) and European (right) test samples explained by PRS derived from Asian, European or multi-ancestry training samples using C + T approach across different *p*-value thresholds, showing superior performance in European than in Asian test samples. $R2_{adj}$ refers to Lee's R^2. (b) Receiver operating characteristic (ROC) curves of Asian (left) and European (right) test samples to evaluate the performance of the 4SNP model and the 11-SNP combined and hybrid PRS models derived from the multi-ancestry meta-analysis. Area under curve (AUC) of the models are shown. (c) Estimated odds ratios with 95% confidence intervals of HSCR in Asian and European test samples by PRS decile, relative to the 3rd decile for the 11-SNP combined PRS model. (d) Distribution of PRS of the 11-SNP combined PRS model for S-HSCR, L-HSCR and TCA samples relative to the controls.

ZNF25 remains relatively unexplored. Given the elevated expression of *Zfp9* during and after neurogenesis and the findings from our functional characterization assay, *ZNF25* may function similarly as *HAND2* and contribute to the development of HSCR by influencing cell migration.

While our current study is primarily focused on identifying HSCR-associated genes, exploring the population differences in genetic architecture as well as the derivation of polygenic risk score to predict the genetic risk of HSCR, the functional characterisation conducted thus far may not be adequate to conclusively confirm the proposed molecular pathomechanisms. Future research could involve more in-depth functional assays, for example, establishing transgenic murine models for lineage tracing or generating genome-edited intestinal organoids from induced pluripotent stem cells, to provide a more comprehensive understanding of the underlying pathological mechanisms linked with these genes.

Although it is well acknowledged that HSCR is genetically heterogeneous, the collective contribution of common variants to disease risk has been underappreciated thus far. The previous genetic model advocated by Tilghman et al. (2019) re-emphasized the importance of common variants in genetic risk prediction such that the majority of patients with HSCR of European ancestry have genetic burden contributed largely by common variants in *RET* and *SEMA3D*. However, this 4-SNP PRS model suffered from non-generalization across populations because of the heterogeneous effect of the *RET* variants and the non-inclusion of *NRG1* which has a larger impact in Asians. PRS models trained on Asian samples also failed to capture those founder variants like *SEMA3D* that showed European-specific associations. In the current study, we derived a 11-SNP combined PRS model aggregating the effects of 11 HSCR-associated variants and secondary association signals based on the multi-ancestry meta-analysis summary statistics. Making use of multi-ancestry data, this global PRS model improves the generalizability across populations, showing equally good predictive power in both European and Asian populations. Alternatively, when the ancestry of the testing samples is unambiguous, the hybrid model which utilizes a combination of ancestry-matched and multi-ancestry training samples can also be considered to further enhance the prediction performance. Intriguingly, our study also revealed a similar level of phenotypic variance explained at a *p*-value cut-off at 1×10^{-6} for both populations using the ancestry-matched C + T methods. The similar proportion of variance explained by twice as many variants at this nominal *p*-value compared to the genome-wide significance threshold implied the presence of additional causal variants with moderate associations yet to be discovered. Altogether, these findings underscore the importance of expanding the size and diversity of the HSCR cohorts in future GWAS to fully understand the genetic basis of HSCR.

In conclusion, our multi-ancestry meta-analysis implicated novel HSCR-associated loci and suggested etiological roles of dysregulation of *ZNF25*, *HAND2*, NOTCH signalling, netrin/UNC5C coupling in disease development. The global PRS model, when integrated additively with genetic models on rare coding and copy number variants, will lead to an improved genetic risk prediction for recurrence within family and aid genetic counselling. It brings us one step closer to the clinical implementation of genomic findings to improve clinical management of HSCR in the future.

Contributors

Y.Z. and C.S.T. verified the data and wrote the article. Y.Z. and D.Z. did the statistical analysis and data interpretation. M.T.S and Z.M. performed the functional characterization assays. Y.W., M.T.S, Z.X., V.C.L. and M.M.G.B. aided the data analysis and interpretation. J.F, S.B., I.C., M.P.P., B.F. and P.K.H.T. recruited the subjects, collected the samples, and provided the genotype data or summary statistics of individual GWAS. Y.S and K.S.C. provided the single cell transcriptomic data. P.K.H.T., C.S.T. and P.C.S. jointly supervised the research and conceived the experiments. P.C.S. and P.K.H.T. contributed to critical revision of the article. All authors have read and approved the final version of the manuscript.

Data sharing statement

The data of this study are available within the article and in the Supplementary Information, or from the authors upon request. The summary statistics and data that support these findings are publicly available and were accessed from several repositories. Finnish GWAS: https://www.danishnationalbiobank.com/gwas/hirschsprung; single cell expression data on mouse foetal gut was available in SRA (https://www.ncbi.nlm.nih.gov/sra/?term=SRR15465838) and GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149524); single cell expression data on human fetal intestine was available in the Gut Cell Altas (https://www.gutcellatlas.org/). Other GWAS results can be found in the original publication as denoted in Supplementary Table S1.

Declaration of interests

João Fadista is currently employed by Novo Nordisk, but all work was completed while at Statens Serum Institut. Other authors declare no competing interests.

Acknowledgements

We are grateful to the numerous patients, their families and referring physicians that have participated in these studies in our laboratories, and the numerous members of our laboratories for their valuable contributions over many years. We thank members of the International HSCR Genetics Consortium (IHGC) for the provision of data for meta-analysis and advices during manuscript drafting.

This work was supported by Theme-based Research scheme [grant number T12C-714/14-R to P.T.], the Science and Technology Development Fund (FDCT) [grant number 0097/2022/A2 to P.T.], REACH research grant, Health and Medical Research Fund (HMRF) [grant numbers 06171636 and 08193446 to C.S.T., 08191496 to Y.S., and PR-HKU-1 to P.T.], Instituto de Salud Carlos III, Spanish Ministry of Economy and Competitiveness, Spain [PI19-01550 and PI22-01428 to S.B.], The Autonomous Government of Andalusia [PEER-0470-2019 to S.B.], and PAIDI2020 of University of Seville [P20_00887 to S.B.]. The Danish HSCR GWAS was funded by the Danish Medical Research Council (DFF 4004-00512) and was conducted using the Danish National Biobank resource established with the support of major grants from the Novo Nordisk Foundation, the Danish Medical Research

Council and the Lundbeck Foundation. B.F. received support from the Oak Foundation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.105680.

References

1 Amiel J, Sproat-Emison E, Garcia-Barcelo M, et al. Hirschsprung disease, associated syndromes and genetics: a review. *J Med Genet*. 2008;45:1–14. https://doi.org/10.1136/jmg.2007.053959.
2 Chakravarti A, McCallion A, Lyonnet S. *Scrivers' online metabolic & molecular bases of inherited disease. Multisystem inborn errors dev. Hirschsprung McGraw Hill Educ*. 2006.
3 Tam PKH, Garcia-Barceló M. Genetic basis of Hirschsprung's disease. *Pediatr Surg Int*. 2009;25:543–558. https://doi.org/10.1007/s00383-009-2402-2.
4 Badner JA, Sieber WK, Garver KL, Chakravarti A. A genetic study of Hirschsprung disease. *Am J Hum Genet*. 1990;46:568–580.
5 Karim A, Tang CS-M, Tam PK-H. The emerging genetic landscape of hirschsprung disease and its potential clinical applications. *Front Pediatr*. 2021;9. https://doi.org/10.3389/fped.2021.638093.
6 Tilghman JM, Ling AY, Turner TN, et al. Molecular genetic anatomy and risk profile of Hirschsprung's disease. *N Engl J Med*. 2019;380:1421–1432. https://doi.org/10.1056/NEJMoa1706594.
7 Tang CS-M, Gui H, Kapoor A, et al. Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease. *Hum Mol Genet*. 2016;25:5265–5275. https://doi.org/10.1093/hmg/ddw333.
8 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. https://doi.org/10.1093/bioinformatics/btq340.
9 Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47:1236–1241. https://doi.org/10.1038/ng.3406.
10 Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1826. https://doi.org/10.1038/s41467-017-01261-5.
11 Chen Z, Boehnke M, Wen X, Mukherjee B. Revisiting the genome-wide significance threshold for common variant GWAS. *G3 (Bethesda)*. 2021;11. https://doi.org/10.1093/g3journal/jkaa056.
12 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003;100:9440–9445. https://doi.org/10.1073/pnas.1530509100.
13 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–425. https://doi.org/10.1016/j.cels.2015.12.004.
14 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739–1740. https://doi.org/10.1093/bioinformatics/btr260.
15 Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37:W305–W311. https://doi.org/10.1093/nar/gkp427.
16 Kichaev G, Pasaniuc B. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am J Hum Genet*. 2015;97:260–271. https://doi.org/10.1016/j.ajhg.2015.06.007.
17 Elmentaite R, Kumasaka N, Roberts K, et al. Cells of the human intestinal tract mapped across space and time. *Nature*. 2021;597:250–255. https://doi.org/10.1038/s41586-021-03852-1.
18 Morarach K, Mikhailova A, Knoflach V, et al. Diversification of molecularly defined myenteric neuron classes revealed by single-cell RNA sequencing. *Nat Neurosci*. 2021;24:34–46. https://doi.org/10.1038/s41593-020-00736-x.
19 Lai FP, Li Z, Zhou T, et al. Ciliary protein Kif7 regulates Gli and Ezh2 for initiating the neuronal differentiation of enteric neural crest cells during development. *Sci Adv*. 2021;7:eabf7472. https://doi.org/10.1126/sciadv.abf7472.
20 Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19:15. https://doi.org/10.1186/s13059-017-1382-0.
21 Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–3587.e3529. https://doi.org/10.1016/j.cell.2021.04.048.
22 Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience*. 2019;8. https://doi.org/10.1093/gigascience/giz082.
23 Gui H, Schriemer D, Cheng WW, et al. Whole exome sequencing coupled with unbiased functional analysis reveals new Hirschsprung disease genes. *Genome Biol*. 2017;18:48. https://doi.org/10.1186/s13059-017-1174-6.
24 Jiang Q, Arnold S, Heanue T, et al. Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet*. 2015;96:581–596. https://doi.org/10.1016/j.ajhg.2015.02.014.
25 Virtanen VB, Salo PP, Cao J, et al. Noncoding RET variants explain the strong association with Hirschsprung disease in patients without rare coding sequence variant. *Eur J Med Genet*. 2019;62:229–234. https://doi.org/10.1016/j.ejmg.2018.07.019.
26 Fadista J, Lund M, Skotte L, et al. Genome-wide association study of Hirschsprung disease detects a novel low-frequency variant at the RET locus. *Eur J Hum Genet*. 2018;26:561–569. https://doi.org/10.1038/s41431-017-0053-7.
27 Kim JH, Cheong HS, Sul JH, et al. A genome-wide association study identifies potential susceptibility loci for Hirschsprung disease. *PLoS One*. 2014;9:e110292. https://doi.org/10.1371/journal.pone.0110292.
28 Garcia-Barcelo MM, Tang CSM, Ngan ESW, et al. Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc Natl Acad Sci U S A*. 2009;106:2694–2699. https://doi.org/10.1073/pnas.0809630105.
29 Tang CS, et al. Identification of genes associated with Hirschsprung disease, based on whole-genome sequence analysis, and potential effects on enteric nervous system development. *Gastroenterology*. 2018;155:1908–1922.e1905. https://doi.org/10.1053/j.gastro.2018.09.012.
30 Lasrado R, Boesmans W, Kleinjung J, et al. Lineage-dependent spatial and functional organization of the mammalian enteric nervous system. *Science*. 2017;356:722–726. https://doi.org/10.1126/science.aam7511.
31 D'Autréaux F, Margolis KG, Roberts J, et al. Expression level of Hand 2 affects specification of enteric neurons and gastrointestinal function in mice. *Gastroenterology*. 2011;141:587.e571–587.e576. https://doi.org/10.1053/j.gastro.2011.04.059.
32 Hendershot TJ, Liu H, Sarkar AA, et al. Expression of Hand2 is sufficient for neurogenesis and cell type-specific gene expression in the enteric nervous system. *Dev Dyn*. 2007;236:93–105. https://doi.org/10.1002/dvdy.20989.
33 Elmentaite R, Ross ADB, Roberts K, et al. Single-Cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. *Dev Cell*. 2020;55:771–783.e775. https://doi.org/10.1016/j.devcel.2020.11.010.
34 Heanue TA, Boesmans W, Bell DM, Kawakami K, Vanden Berghe P, Pachnis V. A novel zebrafish ret heterozygous model of hirschsprung disease identifies a functional role for mapk10 as a modifier of enteric nervous system phenotype severity. *PLoS Genet*. 2016;12:e1006439. https://doi.org/10.1371/journal.pgen.1006439.
35 Pu J, Tang S, Tong Q, et al. Neuregulin 1 is involved in enteric nervous system development in zebrafish. *J Pediatr Surg*. 2017;52:1182–1187. https://doi.org/10.1016/j.jpedsurg.2017.01.005.
36 Jia H, Zhang K, Chen Q, Gao H, Wang W. Downregulation of Notch-1/Jagged-2 in human colon tissues from Hirschsprung disease patients. *Int J Colorectal Dis*. 2012;27:37–41. https://doi.org/10.1007/s00384-011-1295-4.
37 Ngan ES-W, Garcia-Barceló MM, Yip BHK, et al. Hedgehog/Notch-induced premature gliogenesis represents a new disease mechanism for Hirschsprung disease in mice and humans. *J Clin Invest*. 2011;121:3467–3478. https://doi.org/10.1172/JCI43737.
38 Sander GR, Brookes SJ, Powell BC. Expression of Notch1 and Jagged2 in the enteric nervous system. *J Histochem Cytochem*. 2003;51:969–972. https://doi.org/10.1177/002215540305100712.
39 Ko SY, Dass CR, Nurgali K. Netrin-1 in the developing enteric nervous system and colorectal cancer. *Trends Mol Med*. 2012;18:544–554. https://doi.org/10.1016/j.molmed.2012.07.001.
40 Memic F, Knoflach V, Morarach K, et al. Transcription and signaling regulators in developing neuronal subtypes of mouse and human enteric nervous system. *Gastroenterology*. 2018;154:624–636. https://doi.org/10.1053/j.gastro.2017.10.005.

41 Lei J, Howard MJ. Targeted deletion of Hand2 in enteric neural precursor cells affects its functions in neurogenesis, neurotransmitter specification and gangliogenesis, causing functional aganglionosis. *Development*. 2011;138:4789–4800. https://doi.org/10.1242/dev.060053.

42 D'Autréaux F, Morikawa Y, Cserjesi P, Gershon MD. Hand2 is necessary for terminal differentiation of enteric neurons from crest-derived precursors but not for their migration into the gut or for formation of glia. *Development*. 2007;134:2237–2249. https://doi.org/10.1242/dev.003814.