<table>
<thead>
<tr>
<th>Title</th>
<th>Fine mapping of the NRG1 hirschsprung's disease locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Tang, CSM; Tang, WK; So, MT; Miao, XP; Leung, BMC; Yip, BHK; Leon, TYY; Ngan, ESW; Lui, VCH; Chen, Y; Chan, IHY; Chung, HY; Liu, XL; Wu, XZ; Wong, KKY; Sham, PC; Cherny, SS; Tam, PKH; GarciaBarceló, MM</td>
</tr>
<tr>
<td>Citation</td>
<td>Plos One, 2011, v. 6 n. 1</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2011</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/125323">http://hdl.handle.net/10722/125323</a></td>
</tr>
<tr>
<td>Rights</td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Fine Mapping of the NRG1 Hirschsprung’s Disease Locus

Clara Sze-Man Tang1,2,3, Wai-Kiu Tang2,9, Man-Ting So2,9, Xiao-Ping Miao6, Brian Man-Chun Leung2, Benjamin Hon-Kei Yip1,2, Thomas Yuk-Yu Leon2, Elly Sau-Wai Ngan2,4, Vincent Chi-Hang Lui2,4, Yan Chen2, Ivy Hau-Yee Chan2, Patrick Ho-Yu Chung2, Xue-Lai Liu2, Xuan-Zhao Wu5, Kenneth Kak-Yuen Wong2, Pak-Chung Sham1,3,4, Stacey S. Cherny1*, Paul Kwong-Hang Tam2,4, Maria-Mercè García-Barceló2,4*

1 Department of Psychiatry, University of Hong Kong, Hong Kong, China, 2 Department of Surgery, University of Hong Kong, Hong Kong, China, 3 Genome Research Centre, University of Hong Kong, Hong Kong, China, 4 Centre for Reproduction, Development and Growth, University of Hong Kong, Hong Kong, China, 5 Department of Surgery, Guiyang Medical College Affiliated Hospital, Guiyang, China, 6 Department of Epidemiology and Biostatistics, Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, China

Abstract

The primary pathology of Hirschsprung’s disease (HSCR, colon aganglionosis) is the absence of ganglia in variable lengths of the hindgut, resulting in functional obstruction. HSCR is attributed to a failure of migration of the enteric ganglion precursors along the developing gut. RET is a key regulator of the development of the enteric nervous system (ENS) and the major HSCR-causing gene. Yet the reduced penetrance of RET DNA HSCR-associated variants together with the phenotypic variability suggest the involvement of additional genes in the disease. Through a genome-wide association study, we uncovered a ~350 kb HSCR-associated region encompassing part of the neuregulin-1 gene (NRG1). To identify the causal NRG1 variants contributing to HSCR, we genotyped 243 SNPs variants on 343 ethnic Chinese HSCR patients and 359 controls. Genotype analysis coupled with imputation narrowed down the HSCR-associated region to 21 kb, with four of the most associated SNPs (rs10088313, rs10094655, rs4624987, and rs3884552) mapping to the NRG1 promoter. We investigated whether there was correlation between the genotype at the rs10088313 locus and the amount of NRG1 expression levels between diseased and control individuals bearing the same rs10088313 risk genotype. This indicates that the effects of NRG1 common variants are likely to depend on other alleles or epigenetic factors present in the patients and would account for the variability in the genetic predisposition to HSCR.

Introduction

Hirschsprung’s disease (HSCR, aganglionic megacolon) is a congenital disorder of the enteric nervous system (ENS) characterised by the absence of enteric ganglia along a variable length of the intestine. There is significant ethnic variation in the incidence of the disease, and it is most often found among Asians (2.8 per 10,000 live births)[1,2]. Non-familial HSCR has a complex pattern of inheritance and manifests with low, sex-dependent penetrance and variability in the length of the aganglionic segment, according to which patients are classified into short segment (S-HSCR; 80%), long segment (L-HSCR; 15%), and total colonic aganglionosis (TCA; 5%). The male:female ratio is 4:1 among S-HSCR patients and 1:1 among L-HSCR patients. The recurrence risk to sibs of S-HSCR probands ranges between 1.5% and 3.3%, while risk to sibs of L-HSCR probands varies from 2.9% to 17.6%[1].

The RET gene, encoding a tyrosine-kinase receptor, is the major HSCR causing gene[3,4] and its expression is crucial for the development of the enteric ganglia. Mutations in the coding sequence (CDS) of RET account for up to 50% of the familial cases and between 15%–20% of the sporadic cases[5]. Other HSCR genes identified so far mainly code for protein members of interrelated signalling pathways involved in the development of enteric ganglia: RET, endothelin receptor B (EDNRB), and the transcriptional regulator SOX10. Yet, mutations in genes other than RET account for only 7% of the cases[6,7,8,9,10,11,12,13]. Despite the importance of RET, additional genes (acting either in conjunction with or independently from RET) are necessary to explain not only the disease incidence but also its complex pattern of inheritance.

Through a genome-wide association study (GWAS) on Chinese individuals we identified the association of a 350 kb genomic region encompassing intron 1 of the NRG1 gene with Hirsch-
sprung’s disease[14]. Within the \textit{NRG1} region, the strongest overall associations were found for two SNPs located in intron 1 of the neuregulin1 gene (\textit{NRG1}) on 8p12, with rs16679552 and rs7835688 (underlined throughout the text; supplementary Figure S1) yielding odds ratios of 1.60 [CI95%:(1.40,2.00), \( p = 1.80 \times 10^{-8} \)] and 1.90 [CI95%:(1.59,2.47), \( p = 1.12 \times 10^{-9} \)], respectively, for the heterozygous risk genotypes under an additive model. \textit{NRG1} plays an important role in ENS development and maintenance[15,16,17,18].

As these intron 1 \textit{NRG1} HSCR-associated SNPs are not predicted to functionally affect the gene, we hypothesized that these loci are in linkage disequilibrium (LD) with a functional variant(s) not covered by the 500K Affymetrix chips used in our initial GWAS. To find those functional variant(s), we increased the genotype density within the region by genotyping 243 SNPs in 343 HSCR Chinese HSCR patients and 380 Chinese controls. Genotype imputation was used to further increase the SNPs density. As the \textit{NRG1} HSCR-associated region encompasses regulatory regions, we hypothesized that \textit{NRG1} SNPs could affect HSCR susceptibility by altering \textit{NRG1} expression. Thus, the difference between \textit{NRG1} expression levels in gut tissues of affected and non-affected individuals was tested, along with the relationship of expression with \textit{NRG1} genotype.

**Results and Discussion**

Fine mapping implicates the \textit{NRG1} promoter in HSCR susceptibility

Fine mapping of \textit{NRG1} was carried out by genotyping additional SNPs within the region delimited by the upstream and downstream recombination hot spots. We found 9 SNPs (“typed” in Table 1) more significantly associated with HSCR than rs7835688 (initially discovered in the GWAS, \( p = 5.92 \times 10^{-4} \); OR = 1.73 and 95% CI = 1.28–2.32 in this set) and one SNP with a \( p \)-value lower than that of the also GWAS identified rs16879552 (\( p = 8.99 \times 10^{-5} \); OR = 1.63 and 95% CI = 1.29–2.07; also in this set).

The SNP displaying the strongest association [rs10088313; in bold face in Table 1] mapped to the promoter region of most \textit{NRG1} isoforms, (except GGF2, associated with Schizophrenia[19]), and had the strongest LD (\( r^{2} = 0.84 \)) with rs16879552. Other top SNPs were also in moderate to high LD with the previously GWAS-implicated SNPs (rs7835688 and rs16879552) (Figure 1, upper panel).

We next tried to refine our findings by imputing un-typed SNP allelic dosage using MACH. Imputation (based on HapMap Phase II haplotypes) nominated three additional \textit{NRG1}-associated SNPs, rs10094655, rs4624987 and rs3884552, with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar allelic dosage using MACH. Imputation (based on HapMap Phase II haplotypes) nominated three additional \textit{NRG1} SNPs, rs10094655, rs4624987 and rs3884552, with highly similar Phase II. Imputation (based on HapMap Phase II haplotypes) nominated three additional HSCR-associated SNPs, rs10094655, rs4624987 and rs3884552, with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-associated SNPs with highly similar Phase II haplotypes) nominated three additional HSCR-ass
Significant interaction between genotype and case-control status ($F = 5.54, p = 0.0064$) was found, indicating that the relationship between genotype and expression level differs between cases and controls. When both groups (cases and controls) were stratified between genotype and expression level differs between cases and in bold), post hoc group (Figure 2). In this group, the highest differences among the three genotypes only within the control shows higher expression than GG in cases. These data were observed in individuals with the GG (risk) genotype and the lowest in individuals with the TT genotype. Furthermore, GG in individuals with the GG (risk) genotype and the TT risk allele and the NRG1 rs2435357 T risk allele is associated with reduction of RET expression levels differ between patients and controls for a given population prevents us from obtaining a large enough representation of all RET genotypes.

We have recently demonstrated that the RET rs2435357 T risk allele disrupts a SOX10 binding site that compromises RET transactivation[23] SOX10 is also known to regulate the expression of NRG1 receptor ErbB3 in neural crest cell receptors[15] Though the link between RET and NRG1 signalling in the development of the ENS is still to be elucidated our data suggests that NRG1 is player in the signaling network implicated in ENS development and maintenance and that may genetically and biologically interact with members of the RET signalling pathway during the ENS development. Biological interaction between RET and NRG1 signaling has been reported and linked to the survival and maintenance of the peripheral nervous system, where injury-induced expression of the RET ligand GDNF by non-myelinating Schwann cells is ErbB dependent[24].

We suspect that our expression data may be reflecting the result of the genetic NRG1-RET interaction previously described[14]. Somehow comparable findings have been reported in a study on NRG1 expression in post-mortem human brain samples of schizophrenic patients and control individuals where NRG1 expression levels differ between patients and controls for a given NRG1 schizophrenia susceptibility SNP genotype[25]. Collectively,

### Table 1. Fine mapping association results of NRG1 SNPs using logistic regression on MACH-imputed allelic dosage

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position</th>
<th>MAP $^{a,b}$</th>
<th>Allele $^c$</th>
<th>Type</th>
<th>Before imputation</th>
<th>Association values after imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16879425</td>
<td>32426748</td>
<td>0.44</td>
<td>A G</td>
<td>typed</td>
<td>2.29E-04</td>
<td>1.57 (1.23, 1.99) 2.21E-04</td>
</tr>
<tr>
<td>rs10954845</td>
<td>32439384</td>
<td>0.35</td>
<td>A G</td>
<td>typed</td>
<td>2.57E-04</td>
<td>1.59 (1.24, 2.05) 3.29E-04</td>
</tr>
<tr>
<td>rs4422736</td>
<td>32490062</td>
<td>0.45</td>
<td>C T</td>
<td>typed</td>
<td>4.00E-04</td>
<td>1.54 (1.20, 1.96) 4.00E-04</td>
</tr>
<tr>
<td>rs10113578</td>
<td>32503670</td>
<td>0.51</td>
<td>G A</td>
<td>imputed</td>
<td>NA</td>
<td>1.53 (1.22, 1.94) 3.18E-04</td>
</tr>
<tr>
<td>rs10088313</td>
<td>32509603</td>
<td>0.52</td>
<td>G T</td>
<td>typed</td>
<td>6.71E-05</td>
<td>1.60 (1.27, 2.02) 6.71E-05</td>
</tr>
<tr>
<td>rs10107065</td>
<td>32510100</td>
<td>0.52</td>
<td>A G</td>
<td>imputed</td>
<td>NA</td>
<td>1.60 (1.27, 2.01) 7.65E-05</td>
</tr>
<tr>
<td>rs10113593</td>
<td>32510900</td>
<td>0.51</td>
<td>T C</td>
<td>typed</td>
<td>2.18E-04</td>
<td>1.55 (1.23, 1.96) 2.04E-04</td>
</tr>
<tr>
<td>rs10094655</td>
<td>32513689</td>
<td>0.52</td>
<td>T A</td>
<td>imputed</td>
<td>NA</td>
<td>1.63 (1.29, 2.05) 4.27E-05</td>
</tr>
<tr>
<td>rs4624987</td>
<td>32516246</td>
<td>0.52</td>
<td>G A</td>
<td>imputed</td>
<td>NA</td>
<td>1.62 (1.29, 2.05) 4.69E-05</td>
</tr>
<tr>
<td>rs3884552</td>
<td>32519399</td>
<td>0.53</td>
<td>C T</td>
<td>imputed</td>
<td>NA</td>
<td>1.65 (1.30, 2.09) 4.13E-05</td>
</tr>
<tr>
<td>rs7826312</td>
<td>32519657</td>
<td>0.24</td>
<td>C T</td>
<td>typed</td>
<td>3.22E-04</td>
<td>1.77 (1.30, 2.41) 3.22E-04</td>
</tr>
<tr>
<td>rs3802159</td>
<td>32524243</td>
<td>0.51</td>
<td>G C</td>
<td>typed</td>
<td>1.25E-04</td>
<td>NA NA NA</td>
</tr>
<tr>
<td>rs7834206</td>
<td>32525690</td>
<td>0.27</td>
<td>A C</td>
<td>typed</td>
<td>2.27E-04</td>
<td>NA NA NA</td>
</tr>
<tr>
<td>rs16879552</td>
<td>32530758</td>
<td>0.51</td>
<td>C T</td>
<td>typed</td>
<td>8.99E-05</td>
<td>1.63 (1.29, 2.07) 4.19E-05</td>
</tr>
<tr>
<td>rs7835668</td>
<td>32531041</td>
<td>0.26</td>
<td>C G</td>
<td>typed</td>
<td>5.92E-04</td>
<td>1.73 (1.28, 2.32) 3.00E-04</td>
</tr>
<tr>
<td>rs16879576</td>
<td>32560777</td>
<td>0.51</td>
<td>C A</td>
<td>imputed</td>
<td>NA</td>
<td>1.57 (1.24, 1.99) 1.87E-04</td>
</tr>
<tr>
<td>rs12680129</td>
<td>32562687</td>
<td>0.51</td>
<td>A G</td>
<td>typed</td>
<td>1.86E-04</td>
<td>1.57 (1.24, 1.98) 1.83E-04</td>
</tr>
</tbody>
</table>

Only SNPs with $P$-value lower than either of the 2 previously implicated SNPs (rs16879552 and rs7835668) are shown; $^a$: minor allele frequency; underlined: SNPs found associated in the previously reported GWAS. $^b$: frequencies reported for imputed alleles (except for rs3802159 and rs7834206). $^c$: minor and major alleles in patients and controls combined; in bold genotype SNP with the lowest $p$ association value.

doi:10.1371/journal.pone.0016181.t001
these findings warrant further study on NRG1 regulation and its implication in diseases.

This expression study has limitations, some of which cannot be overcome. Firstly, the gut of newborn patients is in an advanced developmental stage (although not fully mature). Thus our analysis does not mimic the expression of NRG1 during the early developmental stages of the human gut, when expression patterns of other genes may be different. For obvious reasons, this cannot be surmounted. Secondly, also for obvious reasons, it is not possible to obtain gut tissue samples from controls, requiring us to use samples from individuals who underwent gut biopsy for other motives than HSCR. Thirdly, the limited sample size is hampering the expression study. Again, it is difficult to obtain a balanced representation of all NRG1 SNP genotypes and a desirable number of gut tissues from affected and non-affected individuals.

We conclude that NRG1 regulatory SNPs may conferring an increased risk of HSCR by interfering with the normal NRG1 expression in human developing gut and their effect is likely to depend on the genetic background of the individual, most likely DNA alterations on trans-acting regulatory proteins expressed in the developing gut in a time-specific manner or in other main.

Figure 1. Fine mapping association results of NRG1 SNPs. Association results shown before (upper panel) and after imputation (lower panel). Diamonds and circles indicate genotyped and imputed SNPs respectively. Colour gradient (red r² = 1 to white r² = 0) marks the LD of the SNPs with rs10088313, except green for rs7835688, dark green for rs16879552 and blue for itself; grey indicates no information on LD. The 21 kb region we narrowed down to is highlighted in pink. The fine-scale recombination rate across the region is represented by the light-blue line. Green lines symbolize a schematic representation of the NRG1 isoforms overlapping the associated region. The bottom green line represents the GGF2 isoform associated with schizophrenia that expands 900 kb upstream the transcription start site for the rest of NRG1 isoforms.

doi:10.1371/journal.pone.0016181.g001

Figure 2. NRG1 expression in human gut tissues. Quantitative RT-PCR analysis of log-transformed expression ratio of NRG1 to 18S in the colon tissues from the normal portions of HSCR patients and non-HSCR patients stratified according to the rs10088313 (G/T; G risk allele). Bars represent the standard error (SE). * p < 0.05; ** p < 0.005.
doi:10.1371/journal.pone.0016181.g002

HSCR genes such as RET. Why and how altered NRG1 expression may contribute to HSCR is yet to be learned.

Materials and Methods

Ethics Statement

At recruitment, informed consent was obtained from each subject. This study was approved by the institutional review board of the University of Hong Kong (UW 03-227 T/227).

Subjects

A total 343 ethnic Chinese patients diagnosed with sporadic HSCR were included for genotyping. Of those, 258 HSCR had been included in our previous GWAS[14] and 177 in the fine mapping of the 9q31 HSCR susceptibility locus[26]. The characteristics of the patients are summarized in supplementary Table S1.

Control individuals were obtained from the blood bank of the Hong Kong Red Cross. We included a total of 359 ethnic Chinese subjects without a diagnosis of HSCR.

Gut tissues were collected from 40 HSCR patients (subset of the 343 individuals genotyped) and 21 non-HSCR patients who had undergone colon surgery for reasons other than HSCR. For the 21 non-HSCR patients (12 affected with imperforate anus; 7 with necrotizing enterocolitis and 2 with mesenteric cysts), tissues were obtained from at least 2 cm away from the margin of the diseased bowel.

The HSCR diagnosis was histologically confirmed with either biopsy or surgical resection material for absence of enteric plexuses.

SNP selection. Following the identification of the NRG1 rs16879552 and rs7835688 HSCR-associated SNPs and the delimitation of the 350 kb HSCR-associated region (chromosome 8:32.235-32.575 Mb; hg18) as originally described, we proceeded with fine-mapping.

The 350 kb region is flanked by two major recombination hotspots (A: 32.235–32.245 Mb and C:32.565–32.575 Mb). A close-up look at the recombination rate within the region (downloaded from HapMap Phase II database) revealed an additional hotspot (B: 32.395–32.405 Mb) that disrupts slightly the LD of the main haplotype block (Supplementary Figure S1). LD was found to be stronger in the region flanked by hotspots B and C. This B–C region spans from the 5’ UTR to exon 2 of the NRG1 isoforms (except for isoform GGF-2 - Gial Growth Factor; NM_013962.2) and includes mostly non-coding sequence DNA containing conserved noncoding elements (CNEs) which are likely
to control gene regulation, chromosome structure, and other key functions.

Annotation and sequencing of the DNA polymorphisms in the 350 kb region were downloaded from the NCBI dbSNP database (dbSNP129). Among the 2675 polymorphisms downloaded, we excluded 320 multiallelic variants and those known to be monomorphic SNPs in Asians. SNPs with MAF as low as 1% in HapMap were not excluded as we attempted to fine map the casual variant affecting the disease, which could be quite rare in the general population.

In order to minimize the genotyping cost, we used WCLUS-TAG[27], an in-house developed tagging program, to identify tag SNPs among the HapMap genotyped SNPs ($r^2$>0.90).

For selection of SNPs without population frequencies available, we applied a tiered approach based on their functional significance: i) we force-included all nonsynonymous SNPs, since presumably they have higher impact on protein structure and function; and ii) we selected SNPs according to functional scores. For this, we relied on two databases, Ensembl and UCSC. The criteria for selection included accessibility to chromatin, CpG islands associated with promoter and the degree of conservation among species (Mu1iz 28-way). For each SNP, we also checked if the allelic variation introduces potential change in predicted transcription factor binding sites (TFBS). To this end, we used P-MATCH[29]. We assigned a higher score to SNPs whose variation disrupts an existing TFBS or creates a new one.

A total of 243 SNPs spanning about 350 kb of chromosome 8 (from 32,236 Mb to 32,575 Mb) were selected for genotyping.

**SNP genotyping.** The 243 SNPs selected were genotyped in 343 HSCR cases and in 359 controls using Sequenom technology as previously described[29]. After removing 12 cases and 8 controls with call rate <90%, 331 HSCR cases and 351 controls remained for association analysis. Standard quality control criteria for SNPs were employed, leaving a total of 207 SNPs with call rate >95%, MAF>1% and not violating Hardy Weinberg equilibrium ($p$>0.001).

**Real-time assay for gene expression.** Resected colon tissues were collected from 40 HSCR patients and 21 non-HSCR patients. No tissues were available from the rest of the patients. All resected tissues were immediately placed in liquid nitrogen and then stored at –80°C before analysis. Full-thickness tissues from ganglionic portions of bowel of each HSCR patient and colonos from non-HSCR patients were used for RNA extraction by Trizol Reagent (Life Technologies, Rockville, MD) and converted to cDNA using an oligo (dT)15 primer and Superscript III (Invitrogen, Carlsbad, CA). The cDNA products equivalent to 10 ng of total RNA were used for quantitative real-time PCR which was performed by ready-to-use TaqMan gene expression assays from Applied Biosystems. The assay for NRG1 was Hs00247620_m1, which targets all but type 3 NRG1 SMDF (this isoform plays a major role in myelination). Real-time qPCR was performed in triplicate (96-well plates) on an ABI 7900 (Applied Biosystems) machine using standard thermal cycling conditions (10 min at 95°C, 40 cycles for 15 s at 95°C, 1 min at 60°C). A standard curve was constructed for each PCR run with 10-fold serial dilutions containing 10, 10, 1, 0.1 and 0.01 ng/mL of cDNA from the neuroblastoma cell line HTB11. The amount of target gene per sample was interpolated according to the standard curves. All analyses were performed in a blinded fashion with the laboratory operators unaware of genotyping data.

**Statistical analyses.** Association was assessed by means of R and PLINK[30] using logistic regression under an additive model and sample origin (Northern vs. Southern Chinese) was included as a covariate to correct for population stratification.

To further evaluate the association of untyped markers tagged by the genotyped SNPs, imputation was carried out by MACH[31] using HapMap Phase II CHB haplotypes as reference. As our ultimate goal here was to pinpoint the causal variant of NRG1, increasing marker density outweighed the use of the less dense yet more accurate Phase III panel.

We examined the effects of case-control status and NRG1 genotype on log-transformed NRG1 expression in gut via two-way ANOVA and post-hoc test of all pair-wise differences. These statistical analyses were done using R. All statistical tests were twosided, and $p$<0.05 was considered significant.

**Supporting Information**

**Table S1** Characteristics of the Chinese HSCR patients included in the NRG1 genotyping. (DOCX)

**Figure S1** Schematic representation of the 350 kb NRG1 HSCR-associated region. On the top panel, close-up of the chromosome 8 association peak obtained in the GWAS. Middle panel, recombination rates throughout the region (red vertical lines). NRG1 isoforms are represented by grey lines (boxes represent exons). Bottom line, Haploview representation of the LD in the region ($D^*$). (TIF)

**Figure S2** Regional map of the 5 HSCR-associated SNPs. rs10088313, rs10094655, rs4624987, rs3884552 and rs16879552 depicted in top green panel, from left to right. Conservation information was given by PhastCon score for Multiz 28-way alignment for vertebrates (hg18). (TIF)

**Author Contributions**

Conceived and designed the experiments: MMGB CSMT SST PCS PKHT. Performed the experiments: WKT MTS XPM BMCL TYYL. Wrote the paper: CSMT SSC MMGB.

**References**


2. Torfs CP An epidemiological study of Hirschsprung disease in a multiracial California population; 1998; Evian, France.


