



POPULATION STUDY ARTICLE

Common variation of the *NSD1* gene is associated with susceptibility to Hirschsprung's disease in Chinese Han population

Xian-Xian Yu^{1,2,3}, Xun Chu^{2,3}, Wen-Jie Wu^{1,2,3}, Zhi-Liang Wei^{2,3}, Huan-Lei Song^{2,3}, Mei-Rong Bai^{2,3}, Yan-Jiao Lu^{2,3}, Bei-Lin Gu^{2,3}, Yi-Ming Gong^{1,2,3} and Wei Cai^{1,2,3}

BACKGROUND: Hirschsprung's disease (HSCR) is the most common congenital cause of intestinal obstruction in children. Sotos syndrome (SoS) is an overgrowth disorder with constipation and sometimes accompanied by HSCR. *NSD1* gene mutation is the main cause of SoS. We aimed to investigate association of *NSD1* common single nucleotide polymorphisms (SNPs) with HSCR susceptibility in Chinese Han population.

METHOD: We genotyped 15 SNPs encompassing *NSD1* gene region in 420 HSCR patients and 1665 controls on Fluidigm EP1 platform. Association analysis was performed between cases and controls.

RESULT: Rs244709 was the most associated SNP with HSCR susceptibility of the sample set ($P_{\text{Allelic}} = 9.69 \times 10^{-5}$, OR = 1.37, 95% CI: 1.17–1.61). Gender stratification analysis revealed that *NSD1* SNPs were associated with HSCR in males, but not in females. The nonsynonymous coding SNP rs28932178 in *NSD1* exon 5 represented the most significant signal in males ($P_{\text{Allelic}} = 6.43 \times 10^{-5}$, OR = 1.42, 95% CI: 1.20–1.69). The associated SNPs were expression quantitative trait loci (eQTLs) of nearby genes in multiple tissues. *NSD1* expression levels were higher in aganglionic colon tissues than ganglionic tissues ($P = 3.00 \times 10^{-6}$).

CONCLUSION: *NSD1* variation conferred risk to HSCR in males, indicating SoS and HSCR may share common genetic factors.

Pediatric Research (2021) 89:694–700; <https://doi.org/10.1038/s41390-020-0933-3>

IMPACT:

- This is the first study to reveal that *NSD1* variation conferred risk to Hirschsprung's disease susceptibility in males of Chinese Han population, indicating Sotos syndrome and Hirschsprung's disease may share some common genetic background.
- This study indicates more attention should be paid to the symptom of constipation in patients with Sotos syndrome.
- Our results raise questions about the role of *NSD1* in the development of enteric nervous system and the pathogenesis of Hirschsprung's disease.

INTRODUCTION

Hirschsprung's disease (HSCR) is a heterogeneous disorder characterized by the absence of ganglion cells along the intestine resulting in functional bowel obstruction. The incidence of HSCR is estimated to be approximately 1 in 5000 live births in Europeans and 1 in 2800 in Asians.¹ This neurodevelopmental birth defect has sex difference in incidence with the male:female ratio of 4:1.^{2,3} Clinically, according to the length of the aganglionic segment, HSCR can be mainly classified as short-segment HSCR (S-HSCR), long-segment HSCR (L-HSCR), and total colonic aganglionosis (TCA), which count for approximately 80%, 15% and 5% of all patients, respectively.¹ The underlying mechanism of HSCR pathogenesis is the failure of normal development of the enteric nervous system (ENS) in the progress of neural crest cell migration, proliferation, differentiation and survival.⁴ HSCR commonly occurs as an isolated phenotype, which was considered as a multifactorial

disease. Common variants in several genes were found be involved in the pathogenesis of HSCR in recent studies, such as *RET*, *NRG1* and *SEMA3C-D*.^{2,3}

Sotos syndrome (SoS) is a congenital overgrowth disorder with an incidence of approximately 1 in 14,000 live births.⁵ It is characterized by a distinctive facial appearance, learning disability and overgrowth. SoS is believed to be an autosomal dominant disorder, and *NSD1* mutations are detected in approximately 80–90% of SoS patients.⁵ *NSD1* encodes the nuclear receptor binding SET domain protein 1, which contains multiple functional domains and functions as a bifunctional transcriptional regulator that can activate or repress transcription in response to ligand binding. Sio et al.⁶ reported a case of SoS patient associated with HSCR, and identified a novel *NSD1* mutation responsible for this clinical syndrome. Considering that constipation is commonly seen in SoS and HSCR could be accompanied by SoS,^{7,8} it is

¹Department of Pediatric Surgery, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China; ²Shanghai Key Laboratory of Pediatric Gastroenterology and Nutrition, Shanghai, China and ³Shanghai Institute of Pediatric Research, Shanghai, China
Correspondence: Xun Chu (chuxun@xinhua.com.cn) or Wei Cai (caiw204@sjtu.edu.cn)
These authors contributed equally: Xian-Xian Yu, Xun Chu

Received: 11 December 2019 Revised: 1 April 2020 Accepted: 2 April 2020
Published online: 7 May 2020

reasonable to speculate that HSCR and SoS might share common genes. In this study, we therefore aimed to investigate association of common *NSD1* variation with risk to HSCR in Chinese Han population.

MATERIALS AND METHODS

Subjects

A total of 420 sporadic HSCR patients were recruited from 2008 to 2018 who have received treatments in Xinhua Hospital affiliated to Shanghai Jiao Tong University School of Medicine. All the cases were diagnosed with HSCR by examination of biopsy specimen for the absence of the enteric ganglia. Patients were composed of 323 S-HSCR, 58 L-HSCR, and 39 TCA. Among 420 patients, 322 were males and 98 were females (the male:female ratio of 3.29:1). A group of 1665 healthy, unrelated subjects visiting Xinhua Hospital for routine health check were randomly selected as controls including 1281 males and 384 females (the male:female ratio of 3.34:1). This study was approved by the institution review board of Xinhua Hospital. Written informed consent was obtained from all participants or their parents after the procedure had been fully explained. All data were recorded anonymously, but the data were destroyed if the participants asked to withdraw their file.

SNP selecting and genotyping

We investigated single nucleotide polymorphisms (SNPs) located in the genic region and 20 kb upstream and downstream flanking regions of *NSD1*. TagSNPs were selected using the Genome Variation Server (<http://gvs.gs.washington.edu/GVS150/>) with MAF (minor allele frequency) ≥ 0.01 and $r^2 \geq 0.8$ based on the HapMap HCB (Han Chinese in Beijing) data. Among a cluster of interchangeable tag SNPs, SNPs with predicted deleterious effect were preferred including nonsynonymous SNPs, expression quantitative trait loci (eQTL) SNPs and SNPs occurring within transcription factor binding sites or in a regulatory region. To avoid missing information caused by failure in assay designing and genotyping, nonsynonymous SNPs and 5'UTR SNPs in high LD were all selected. Together 15 SNPs were selected, namely rs244730, rs244731, rs71601339, rs244709, rs202177679, rs3733875, rs28932178, rs79600412, rs6886255, rs35848863, rs34165241, rs78247455, rs4631, rs3088050 and rs10051822.

The selected 15 SNPs were genotyped using Fluidigm 96.96 Dynamic Array IFCs on Fluidigm EP1 platform. Of the 15 SNPs genotyped, three SNPs (rs4631, rs3088050 and rs10051822) with call rate < 98% was removed for further association analysis. The data completion rate of the remaining 12 SNPs was 99.9%. The average interval between the 12 SNPs was 16.3 kb.

Functional annotation

Functional consequence of coding variants was predicted using PolyPhen 2, SIFT and PROVEAN. We visited GTEx Analysis Release v.8 (<http://www.gtexportal.org/>) to examine the eQTL for 12 selected SNPs of *NSD1*.⁹ We also investigated the functional consequences of the associated SNP by checking HaploRegv4.1 database. The gene expression profiles of associated genes were also acquired from GTEx database. We explored protein–protein interaction using STRING database (<http://string-db.org/>).¹⁰ *NSD1* and its neighbor genes *MXD3* and *RGS14* were looked up in STRING database to explore the interaction with HSCR underlying pathway genes (*RET*, *EDNRB*, *NRG1*, *SEMA3*, *SOX10* and *PHOX2B*).

Quantitative real-time PCR analysis

Aganglionic colon tissues and adjacent ganglionic tissues were collected from 28 HSCR patients who underwent surgery in Xinhua Hospital affiliated to Shanghai Jiao Tong University School of Medicine. Total RNA of aganglionic colon tissues and ganglionic tissues was isolated using RNeasy Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. cDNAs

were synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific™, MA USA). Quantitative real-time PCR (RT-PCR) was performed with SYBR Green Kit (TaKaRa, Otsu, Japan) on a QuantStudio Dx Real-Time PCR (RT-PCR) Instrument (Applied Biosystems, Foster City, CA, USA). RT-PCR of each sample was performed in triplicate with *GAPDH* as an internal control. Primers are listed in Supplementary Table S1. The expression levels between aganglionic colon tissues and ganglionic tissues were compared using a paired Student's *t* test.

Statistical analysis

Genetic analysis was performed using PLINK 1.09. The genotype distribution of each SNP was tested for Hardy–Weinberg equilibrium in both case and control population.

Four genetic models, including the allelic, additive, dominant and recessive models, together with a genotypic association test (2df test) were used to analyze the association for each SNP. Both allele and genotype frequencies were assessed by χ^2 -test between the cases and the controls. Conditional logistic analysis was performed to find additional markers with independent effect by adding the top associated markers as covariates in logistic regression. The linkage disequilibrium (LD) structure of 12 SNPs was examined by program Haploview 4.2. A study-level *P* value < 0.0042 (0.05/12) was considered statistically significant.

RESULTS

Associations of *NSD1* SNPs with HSCR

The genotype distribution of all 12 SNPs genotyped conformed to the Hardy–Weinberg equilibrium ($P > 0.05$). The allelic and genotypic distribution between cases and controls are shown in Tables 1 and 2, respectively. After multiple testing, we found that six SNPs reached the study-level significance.

Of the 12 SNPs genotyped, rs244709 is the most associated SNP located in intron 2 of *NSD1*. The minor allele T of rs244709 was associated with HSCR risk ($P_{\text{Allelic}} = 9.69 \times 10^{-5}$, OR = 1.37, 95% CI: 1.17–1.61; Table 1). The frequency of rs244709 allele T was 0.49 in cases and 0.42 in controls. The second association signal was two nonsynonymous SNPs in the fifth exon of *NSD1*, namely rs3733875 and rs28932178. These two SNPs were in perfect LD ($r^2 = 1$) and were in high LD with rs244709 ($r^2 = 0.83$; Fig. 1). The major allele T of rs28932178 was associated with HSCR risk ($P_{\text{Allelic}} = 7.68 \times 10^{-4}$, OR = 1.30, 95% CI: 1.11–1.51; Table 1).

HSCR is more common in males than in females; we therefore performed association analysis in males and females, respectively. None of the 12 SNPs showed association with HSCR in females ($P > 0.05$; Table 3). However, a significant increase in risk was observed when comparing male group with the combined samples. The nonsynonymous coding SNP rs28932178 represented the most significant signal ($P_{\text{Allelic}} = 6.43 \times 10^{-5}$, OR = 1.42, 95% CI: 1.20–1.69; Table 3). rs71601339 located 5' upstream of *NSD1* was in high LD with the above two missense SNPs ($r^2 = 0.98$) and also showed significant association with HSCR risk ($P_{\text{Allelic}} = 1.63 \times 10^{-4}$, OR = 1.43, 95% CI: 1.19–1.73, Table 3). Since six SNPs showed association with disease risk in males, we performed conditional logistic analysis to identify independent association signal. When conditioning on the most associated SNPs rs28932178, no other SNPs show significant association ($P > 0.05$).

Genotype–phenotype association study

HSCR is classified as S-HSCR, L-HSCR and TCA. S-HSCR is the most common form, which is characterized by multifactorial inheritance.^{3,11} L-HSCR and TCA is often characterized by autosomal dominant inheritance. However, rs2435357 risk allele at *RET*, the major HSCR gene, was observed to be more common in S-HSCR/L-HSCR when compared with TCA HSCR.¹² Consequently, we investigated whether all six significant SNPs were restricted to specific markers of subgroups by comparing frequencies of risk

Table 1. Association analysis of 12 SNPs in 420 HSCR patients and 1665 controls.

SNP	BP	Nucleotide amino acid	Feature	Risk allele	RAF (%)		Additive <i>P</i> value	Allelic <i>P</i> value	OR (95%CI)
					Case	Control			
rs244730	176,539,212	—	5'upstream	G	98.0	96.9	1.14×10^{-1}	1.08×10^{-1}	1.53 (0.91–2.57)
rs244731	176,539,679	—	5'upstream	G	98.0	96.9	1.12×10^{-1}	1.08×10^{-1}	1.53 (0.91–2.57)
rs71601339	176,544,496	—	5'upstream	A	53.5	46.9	1.61×10^{-4}	1.33×10^{-3}	1.31 (1.11–1.54)
rs244709	176,587,111	—	Intron 2	T	49.5	41.6	1.19×10^{-4}	9.69×10^{-5}	1.37 (1.17–1.61)
rs202177679	176,608,428	—	Intron 2	C	53.6	47.1	9.82×10^{-4}	8.58×10^{-4}	1.29 (1.11–1.51)
rs3733875	176,637,240	V614L	Exon 5	C	53.3	46.9	8.86×10^{-4}	8.15×10^{-4}	1.30 (1.11–1.51)
rs28932178	176,637,576	S726P	Exon 5	T	53.3	46.8	8.34×10^{-4}	7.68×10^{-4}	1.30 (1.11–1.51)
rs79600412	176,713,634	—	Intron 20	A	99.6	99.3	2.12×10^{-2}	2.14×10^{-1}	2.11 (0.63–6.99)
rs6886255	176,718,361	—	Intron 21	G	51.2	45.0	1.41×10^{-3}	1.24×10^{-3}	1.28 (1.10–1.49)
rs35848863	176,721,119	M2250I	Exon 23	A	2.1	1.9	5.90×10^{-1}	5.90×10^{-1}	1.16 (0.68–1.97)
rs34165241	176,721,151	M2261T	Exon 23	G	2.1	1.9	5.94×10^{-1}	5.93×10^{-1}	1.16 (0.68–1.96)
rs78247455	176,722,005	A2546T	Exon 23	A	2.1	1.8	5.54×10^{-1}	5.53×10^{-1}	1.17 (0.69–2.00)

SNP single nucleotide polymorphism, BP base pair, RAF risk allele frequency, OR odds ratio, CI confidence interval.

alleles within cases classified by segment length of aganglionosis. Firstly, we grouped S-HSCR/L-HSCR together and compared them with TCA HSCR cases. Secondly, we grouped L-HSCR/TCA together and compared them with S-HSCR cases. Our results found none of the six associated SNPs showed significant genotype–phenotype associations as shown in Table 4.

Functional annotation of associated SNPs

SNP rs28932178 is a nucleotide transition (T > C) in exon 5 of *NSD1* that causes the amino acid substitution P726S. The functional prediction results of rs28932178 from PolyPhen 2, SIFT and PROVEAN were incongruent. PolyPhen 2 predicts a probably damaging variant with a score of 0.998, PROVEAN predicts a neutral variant with a score of 0.99 (cutoff = –2.5), SIFT predicts a damaging variant with a score of 0.013 (cutoff = 0.05). Unexpectedly, the risk allele T of rs28932178 was the ancestral one. The other nonsynonymous SNP rs3733875 in the fifth exon of *NSD1* was predicted to a neutral variant in three databases.

NSD1 was widely expressed and the highest level was found in cerebellar hemisphere and cerebellum from the Genotype-Tissue Expression (GTEx) portal (Supplementary Fig. S1). A higher expression level was also found in gastrointestinal tract tissues (Supplementary Fig. S1). The expression profiles of *NSD1* also indicated that it might play a critical role in the development of central nerve system and enteric nerve system.

None of the six associated SNPs was identified as eQTL of *NSD1*, while four of the six SNPs (rs71601339, rs3733875, rs28932178, rs6886255) were found to be correlated with *MXD3* expression in multiple gastrointestinal tissues (Table 5). rs244709 was associated with the expression of *RGS14* as well (Table 5). Both *MXD3* and *RGS14* are neighboring genes of *NSD1*. Protein–protein interaction analysis found no interaction between *NSD1*, *MXD3* and *RGS14* with HSCR underlying pathway genes (*RET*, *EDNRB*, *NRG1*, *SEMA3*, *SOX10*, *PHOX2B*).

Gene expression analysis

We further compared the RNA expression level of *NSD1*, *MXD3* and *RGS14* in aganglionic and ganglionic colon tissues from 28 HSCR patients. The expression level of *NSD1* was higher in aganglionic tissues than ganglionic tissues, with the *P* value of 3.00×10^{-6} (Fig. 2a), while the expression levels of *RGS14* (*P* = 0.42) and *MXD3* (*P* = 0.48) showed no difference (Fig. 2b, c), indicating *NSD1* may participate in the pathogenesis of HSCR.

DISCUSSION

SoS was first described by Juan Sotos in 1964 in five patients with prenatal and postnatal overgrowth, characteristic facial appearance, advanced bone age and developmental delay.¹³ The diagnosis of SoS is mainly based on the disease history and physical examination, which includes the cardinal features of typical facial dysmorphisms, increased velocity of growth and weight with advanced bone age, and some form of developmental delay such as in talking, walking, or other developmental milestone.⁶ Genetic analysis revealed that SoS was caused by mutation of *NSD1*.⁸ SoS is often accompanied by other symptoms, such as congenital heart disease, neurologic abnormalities and genitourinary problems.¹³ Notably, some SoS patients' complaint was gastrointestinal symptoms, especially constipation.^{7,14} Sio et al.⁶ recently reported an SoS patient with HSCR and identified two novel rare disease-causing mutations of *NSD1*, c.2465C4A (p. Ser822Tyr) and c.4347T4A (p.Cys1449). In this study, we found that six common SNPs in *NSD1* were associated with isolated HSCR, indicating SoS and HSCR may share overlapping genetic background. Stratification analysis revealed that *NSD1* SNPs were associated with HSCR in males, but not in females.

Out of the six associated markers, rs3733875 and rs28932178 are two missense variants in the exon 5 of *NSD1*. rs3733875 is a G > T transversion, changing a valine residue to a leucine residue; while rs28932178 is a T > C transversion, changing a serine residue to a proline residue. The amino acid substitutions caused by missense variants are more likely to have serious consequences for the function or structural stability of *NSD1*, especially when the interval between these two SNPs is only 336 bps in location.^{15,16} The remaining four associated SNPs (rs71601339, rs3733875, rs28932178 and rs6886255) are located in the untranslated region of *NSD1*. We searched all the four positive SNPs in HaploReg to further explore their functions.¹⁷ Interestingly, we found rs28932178 and rs6886255 affected a PAX-family binding site (Supplementary Table S2), while *PAX3* has been reported to play an important role in ENS development and in the pathogenesis of colonic aganglionosis.¹⁸ Therefore, these four noncoding SNPs might also have some functional effect involved in the pathogenesis of HSCR.

Considering the clinical manifestations, surgical options, and post-operation complications of L-HSCR and TCA patients differ significantly from S-HSCR patients, we performed an association analysis stratified by clinical subtype and did not find SNPs associated with specific subtype. In a gender-stratified analysis, we

Table 2. Comparisons of genotype distribution of 12 SNPs between HSCR patients and healthy controls.

SNP	Genotype	Genotype distribution <i>N</i> (%)		Genotypic <i>P</i> value	Dominant <i>P</i> value	Recessive <i>P</i> value
		Case	Control			
rs244730	G/G	402 (95.9)	1565 (94.0)	NA	NA	NA
	G/A	17 (4.1)	96 (5.8)			
	A/A	0	3 (0.2)			
rs244731	G/G	402 (95.9)	1565 (94.0)	NA	NA	NA
	G/A	17 (4.1)	96 (5.8)			
	A/A	0	3 (0.2)			
rs71601339	A/A	108 (29.4)	324 (22.7)	6.86×10^{-3}	1.06×10^{-2}	8.54×10^{-3}
	A/G	177 (48.2)	683 (48.4)			
rs244709	G/G	82 (22.3)	412 (28.9)	3.16×10^{-4}	1.05×10^{-4}	2.14×10^{-2}
	T/T	89 (23.5)	276 (18.3)			
	T/C	197 (52.0)	706 (47.4)			
rs202177679	C/C	93 (24.5)	529 (34.3)	3.72×10^{-3}	2.23×10^{-3}	1.94×10^{-2}
	C/A	117 (27.9)	381 (22.9)			
	C/A	213 (50.8)	804 (48.3)			
rs3733875	A/A	89 (21.2)	476 (28.6)	3.39×10^{-3}	2.14×10^{-3}	1.86×10^{-2}
	C/C	117 (27.9)	374 (22.5)			
	C/A	213 (50.8)	813 (48.8)			
rs28932178	A/A	89 (21.2)	478 (28.7)	3.23×10^{-3}	2.09×10^{-3}	1.75×10^{-2}
	T/T	117 (27.9)	373 (22.4)			
	T/C	213 (50.8)	813 (48.9)			
rs79600412	C/C	89 (21.2)	478 (28.7)	NA	NA	NA
	A/A	416 (99.3)	1640 (98.5)			
	A/C	3 (0.7)	25 (1.5)			
rs6886255	C/C	0	0	3.16×10^{-3}	9.58×10^{-4}	6.25×10^{-2}
	G/G	106 (25.3)	350 (21.1)			
	G/A	217 (51.8)	793 (47.8)			
rs35848863	A/A	96 (22.9)	517 (31.2)	NA	NA	NA
	A/A	0	1 (0.1)			
	A/G	18 (4.3)	60 (3.6)			
rs34165241	G/G	401 (95.7)	1604 (96.3)	NA	NA	NA
	G/G	0	1 (0.1)			
	G/A	18 (4.3)	60 (3.6)			
rs78247455	A/A	401 (95.7)	1602 (96.3)	NA	NA	NA
	A/A	0	1 (0.1)			
	A/G	18 (4.3)	59 (3.6)			
	G/G	401 (95.7)	1602 (96.4)			

found common SNPs in *NSD1* were specially associated with HSCR in males. The male-specific association might be explained by that *NSD1* transcription was regulated by androgen; however, experimental evidence is still needed. Another explanation might be that the small sample size of females (98 female patients vs. 384 female controls) had not enough power to detect the moderate association of common SNPs with this disease.

NSD1 contains 23 exons and encodes a histone H3 lysine 36 methyltransferase, which is important for multiple aspects of normal embryonic development.¹⁹ *NSD1* binds to various promoter elements to regulate gene transcriptions.²⁰ *NSD1* was recently recognized as an indicative factor of transcriptionally permissive chromatin state, which regulates development progress through the epigenetic mechanisms and plays a critical role in ENS development.^{21–24} It was also reported that *NSD1* mutations altered genome-wide DNA methylation in patients with SoS.²⁰ Our expression analysis showed that *NSD1* had a

higher expression level in aganglionic colon tissues than in ganglionic tissues. It is possible that *NSD1* regulates expression level of the major HSCR genes such as *RET*, *NRG1* and *SEMA3C-D*,^{2,3} and therefore contributes to the pathogenesis of HSCR. As *NSD1* SNPs showed male-specific association with HSCR, it could be hypothesized that the methyltransferase function of *NSD1* might be regulated by androgen. The underlying molecular basis need to be elucidated in future study.

In addition, *NSD1* protein interacts with retinoic acid receptors.²⁵ Retinoic acid is required for the efficient migration of enteric neural crest-derived cells (ENCDCs) by reducing the levels of phosphatase and tensin homolog (PTEN) protein, a critical negative regulator of ENCDC migration and proliferation.²⁶ Whether disease-associated variation influences expression of *NSD1* and thus alter RA signal was expected to be further explored. Furthermore, depletion of *NSD1* reduced the expression of bone morphogenetic protein gene 4 (*BMP4*).²⁷ *BMP4* is known

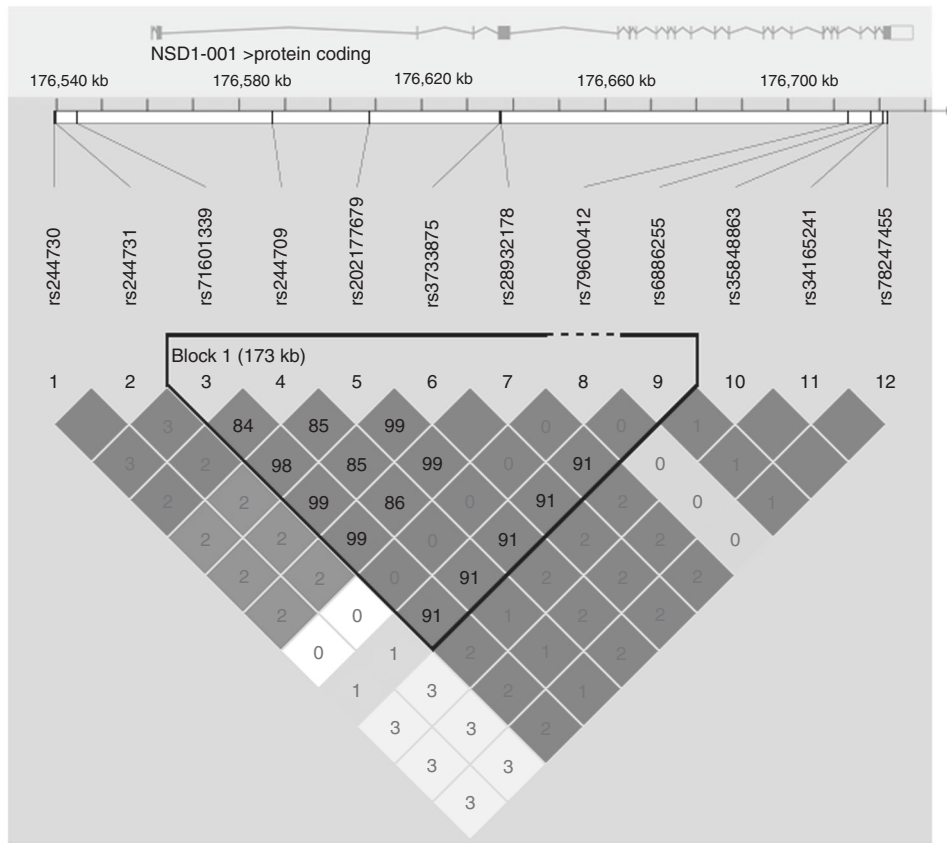


Fig. 1 Linkage disequilibrium (LD) map of 12 SNPs in *NSD1* region of Han Chinese population. The plots were constructed with the program Haploview²⁷, and r^2 values ($\times 100$) were depicted in the diamonds. r^2 values were estimated from the data of 420 HSCR patients and 1663 controls enrolled in the current study. r^2 values of 1 represent complete LD, r^2 values greater than 0.8 represent strong evidence of LD, r^2 values of 0.2–0.8 represent inconclusive LD, and r^2 less than 0.2 represent negligible evidence of LD. Blocks were determined using the default confidence intervals.

SNP	Feature	Risk allele	RAF in males		Allelic P_{Male}	OR (95%CI)	RAF in females (%)		Allelic P_{Female}	OR (95%CI)
			Case	Control			Case	Control		
rs244730	5'upstream	G	98.0	97.2	0.27	1.40 (0.77–2.54)	98.0	96.1	0.21	1.95 (0.68–5.61)
rs244731	5'upstream	G	98.0	97.2	0.27	1.40 (0.77–2.54)	98.0	96.1	0.21	1.95 (0.68–5.61)
rs71601339	5'upstream	A	55.2	46.3	1.63×10^{-4}	1.43 (1.19–1.73)	48.3	49.0	0.88	0.97 (0.70–1.36)
rs244709	Intron 2	T	50.3	41.6	1.44×10^{-4}	1.42 (1.19–1.71)	46.6	41.7	0.24	1.22 (0.88–1.70)
rs202177679	Intron 2	C	55.1	46.6	9.95×10^{-5}	1.41 (1.19–1.68)	48.5	49.1	0.88	0.98 (0.71–1.34)
rs3733875	Exon 5	C	55.0	46.2	6.96×10^{-5}	1.42 (1.20–1.69)	48.0	49.1	0.78	0.96 (0.70–1.31)
rs28932178	Exon 5	T	55.0	46.2	6.43×10^{-5}	1.42 (1.20–1.69)	48.0	49.1	0.78	0.96 (0.70–1.31)
rs79600412	Intron 20	A	99.5	99.4	0.72	1.25 (0.36–4.35)	100.0	98.7	0.11	\
rs6886255	Intron 21	G	52.8	44.4	1.46×10^{-4}	1.40 (1.18–1.66)	45.9	46.7	0.84	0.97 (0.71–1.33)
rs35848863	Exon 23	A	2.2	1.8	0.52	1.22 (0.67–2.23)	2.0	2.1	0.97	0.98 (0.32–2.96)
rs34165241	Exon 23	G	2.2	1.8	0.52	1.22 (0.67–2.23)	2.0	2.1	0.97	0.98 (0.32–2.96)
rs78247455	Exon 23	A	2.2	1.8	0.48	1.24 (0.68–2.28)	2.0	2.1	0.97	0.98 (0.32–2.96)

to play an important role in the formation of ENS by enhancing neuronal aggregation and controlling the clustering of ENDCs into definitive ganglia.²⁸ Since both *RA* and *BMP4* influence the migration of ENDCs, *NSD1* might be involved in related signaling pathway, and plays a role in the pathogenesis of HSCR.

Of note, eQTL data from GTEx database revealed the associated SNPs in *NSD1* region were eQTLs of the adjacent

genes *MXD3* and *RGS14*, but not associated with the *NSD1* expression level (Table 5). Since gene expression regulation is spatio-temporal progress, the eQTL data from adult tissues could not present the expression level in embryonic stages, when the ENS developed. Therefore, whether the associated SNPs influence the expression of *NSD1* in the stage of ENS development still needs further investigation. *MXD3* is a

Table 4. Allele distribution of six significantly associated SNPs grouped by segment length of aganglionosis.

SNP	Reference allele	Reference allele frequency (%)				P value	
		Controls	S-HSCR	S-HSCR /L-HSCR	L-HSCR /TCA	S/L-HSCR vs. TCA	S-HSCR vs. L-SCR/TCA
rs71601339	A	0.47	0.54	0.53	0.54	0.85	0.96
rs244709	T	0.42	0.49	0.49	0.51	0.56	0.61
rs202177679	C	0.47	0.53	0.54	0.55	0.96	0.74
rs3733875	C	0.47	0.53	0.53	0.55	0.93	0.68
rs28932178	T	0.47	0.53	0.53	0.55	0.93	0.68
rs6886255	G	0.45	0.51	0.51	0.53	0.80	0.55

Table 5. eQTL signal of the associated SNPs.

SNP	Gene	eQTL P value from different tissues ^a					
		Esophagus mucosa (n = 497)	Esophagus muscularis (n = 465)	Stomach (n = 324)	Small intestine (n = 174)	Sigmoid colon (n = 318)	Transverse colon (n = 368)
rs244709	<i>RGS14</i>	1.1×10^{-10}	6.9×10^{-13}	3.8×10^{-7}	—	7.1×10^{-8}	7.2×10^{-8}
rs71601339	<i>MXD3</i>	1.6×10^{-10}	1.6×10^{-16}	1.5×10^{-9}	2.9×10^{-7}	2.4×10^{-9}	9.0×10^{-13}
rs3733875	<i>MXD3</i>	7.5×10^{-12}	2.3×10^{-17}	2.5×10^{-9}	—	2.7×10^{-9}	6.9×10^{-13}
rs28932178	<i>MXD3</i>	3.4×10^{-9}	3.0×10^{-14}	4.7×10^{-7}	—	8.5×10^{-8}	1.3×10^{-9}
rs6886255	<i>MXD3</i>	2.3×10^{-10}	7.3×10^{-17}	3.7×10^{-8}	—	1.0×10^{-12}	1.3×10^{-10}
rs202177679	—	—	—	—	—	—	—

^aData were extracted from GTEx portal (GTEx Analysis Release v.8, <http://www.gtexportal.org/>).

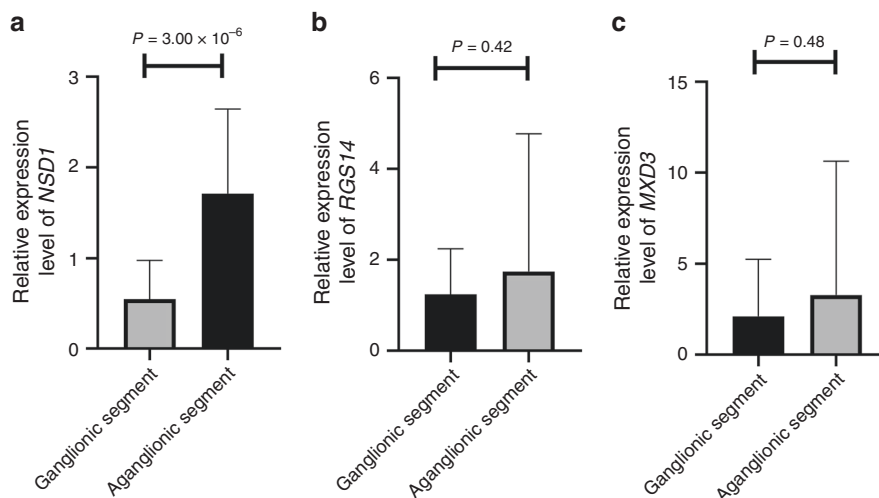


Fig. 2 The relative expression levels of *NSD1*, *MXD3* and *RGS14* in aganglionic and ganglionic colon tissues from HSCR patients. **a** The expression level of *NSD1* was higher in aganglionic segment colon than ganglionic segment colon ($P = 3.00 \times 10^{-6}$; $n = 28$). **b, c** The expression levels of *RGS14* and *MXD3* showed no difference in two segments ($n = 28$).

transcription factor involving in the MYC/MAX/MXD transcriptional network. Knocking down of *MXD3* led to a reduction in cell numbers, and overexpression of *MXD3* promoted cell proliferation in mouse cerebellar.²⁹ Interestingly, it was demonstrated that acute activation of *MXD3* resulted in a transient increase in cell proliferation, while persistent activation of *MXD3* resulted in an overall decrease in cell numbers, indicating *MXD3* had a complicated role in the regulation of cell proliferation.³⁰ *RGS14* is a multifunctional scaffolding protein that integrates G protein and MAPK signaling pathway to regulate synaptic plasticity.³¹ Since these three neighboring genes all have roles in regulating the neural proliferation, it is also possible that they may work as complex to influence the development of ENS.

Our results revealed *NSD1* as a susceptibility locus to HSCR in males and indicated new pathway underlying pathogenesis of HSCR. Further replication studies in other independent sample set were needed to validate this finding and functional experiments were also required to investigate the mechanisms of *NSD1* influence on the development of ENS and the pathogenesis of HSCR.

CONCLUSION

Common variations of *NSD1* conferred risk to HSCR susceptibility in Chinese Han population, indicating SoS and HSCR may share partial overlapping genetic background.

ACKNOWLEDGEMENTS

We thank all the participants of this research. This work was supported by National Natural Science Foundation of China (31671317, 31471190 and Key Program 81630039), Foundation of Shanghai Municipal Health Commission (201840014) and Shanghai Key Laboratory of Pediatric Gastroenterology and Nutrition (17DZ2272000).

AUTHOR CONTRIBUTIONS

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work; W.C. and X.C. designed the research and revised the article; X.-X.Y. drafted the article; X.C. and X.-X.Y. analyzed the data; X.-X.Y., W.-J.W., H.-L.S., Z.-L.W., M.-R.B., Y.J.L., and Y.-M.G. performed the experiment and collected the data.

ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41390-020-0933-3>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Consent statement: In the present study, written informed consent was obtained from all participants or their parents.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Haricharan, R. N. & Georgeson, K. E. Hirschsprung disease. *Semin. Pediatr. Surg.* **17**, 266–275 (2008).
2. Garcia-Barcelo, M.-M. et al. Genome-wide association study identifies *NRG1* as a susceptibility locus for Hirschsprung's disease. *Proc. Natl Acad. Sci. USA* **106**, 2694–2699 (2009).
3. Jiang, Q. 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.* **96**, 581–596 (2015).
4. Burkhardt, D. D., Graham, J. M., Short, S. S. & Frykman, P. K. Advances in Hirschsprung disease genetics and treatment strategies: an update for the primary care pediatrician. *Clin. Pediatr.* **53**, 71–81 (2014).
5. Lane, C., Milne, E. & Freeth, M. Cognition and behaviour in Sotos syndrome: a systematic review. *PLoS ONE* **11**, e0149189 (2016).
6. Sio, C. A. et al. Sotos syndrome associated with Hirschsprung's disease: a new case and exome-sequencing analysis. *Pediatr. Res.* **82**, 87–92 (2017).
7. Faravelli, F. *NSD1* mutations in Sotos syndrome. *Am. J. Med. Genet. Part C: Semin. Med. Genet.* **137C**, 24–31 (2005).
8. Kurotaki, N. et al. Haploinsufficiency of *NSD1* causes Sotos syndrome. *Nat. Genet.* **30**, 365–366 (2002).
9. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multi-tissue gene regulation in humans. *Science* **348**, 648–660 (2015).

10. Szklarczyk, D. et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
11. Tam, P. K. Hirschsprung's disease: a bridge for science and surgery. *J. Pediatr. Surg.* **51**, 18–22 (2016).
12. Emison, E. S. et al. Differential contributions of rare and common, coding and noncoding *Ret* mutations to multifactorial Hirschsprung disease liability. *Am. J. Hum. Genet.* **87**, 60–74 (2010).
13. Leventopoulos, G. et al. A clinical study of Sotos Syndrome patients with review of the literature. *Pediatr. Neurol.* **40**, 357–64 (2009).
14. Baujat, G. & Cormier-Daire, V. Sotos syndrome. *Orphanet J. Rare Dis.* **2**, 36 (2007).
15. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95–108 (2005).
16. Orr, N. & Chanock, S. Common genetic variation and human disease. *Adv. Genet.* **62**, 1–32 (2008).
17. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
18. Barlow, A. J., Dixon, J., Dixon, M. & Trainor, P. A. *Tcof1* acts as a modifier of *Pax3* during enteric nervous system development and in the pathogenesis of colonic aganglionosis. *Hum. Mol. Genet.* **22**, 1206–17 (2013).
19. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–68 (2010).
20. Choufani, S. et al. *NSD1* mutations generate a genome-wide DNA methylation signature. *Nat. Commun.* **6**, 10207 (2015).
21. Hu, N., Strobl-Mazzulla, P. H. & Bronner, M. E. Epigenetic regulation in neural crest development. *Dev. Biol.* **396**, 159–68 (2014).
22. Torroglosa, A. et al. Epigenetics in ENS development and Hirschsprung disease. *Dev. Biol.* **417**, 209–216 (2016).
23. Strobl-Mazzulla, P. H., Sauka-Spengler, T. & Bronner-Fraser, M. Histone demethylase *Jmjd2A* regulates neural crest specification. *Dev. Cell* **19**, 460–468 (2010).
24. Tan, H., Wu, S., Wang, J. & Zhao, Z. K. The *JMJD2* members of histone demethylase revisited. *Mol. Biol. Rep.* **35**, 551–556 (2008).
25. Huang, N. et al. Two distinct nuclear receptor interaction domains in *NSD1*, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J.* **17**, 3398–3412 (1998).
26. Sato, Y. & Heuckeroth, R. O. Retinoic acid regulates murine enteric nervous system precursor proliferation, enhances neuronal precursor differentiation, and reduces neurite growth in vitro. *Dev. Biol.* **320**, 185–98 (2008).
27. Lucio-Eterovic, A. K. et al. Role for the nuclear receptor-binding SET domain protein 1 (*NSD1*) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc. Natl Acad. Sci. USA* **107**, 16952–16957 (2010).
28. Goldstein, A. M., Brewer, K. C., Doyle, A. M., Nagy, N. & Roberts, D. J. BMP signaling is necessary for neural crest cell migration and ganglion formation in the enteric nervous system. *Mech. Dev.* **122**, 821–33 (2005).
29. Ngo, T. et al. Alternative splicing of *MXD3* and its regulation of *MXD3* levels in glioblastoma. *Front. Mol. Biosci.* **6**, 5 (2019).
30. Ngo, T., Barisone, G. A., Lam, K. S. & Diaz, E. *MXD3* regulation of DAOY cell proliferation dictated by time course of activation. *BMC Cell Biol.* **15**, 30 (2014).
31. Evans, P. R., Lee, S. E., Smith, Y. & Hepler, J. R. Postnatal developmental expression of regulator of G protein signaling 14 (*RGS14*) in the mouse brain. *J. Comp. Neurol.* **522**, 186–203 (2014).