Genetic and Functional Studies of Hirschsprung Disease

Rajendra Kumar Chauhan

The work described in this thesis was financially supported by grants from ZonMW (TOP-subsidie 40-00812-98-10042) and the Maag lever Darm stichting (W009-62).

All research described in the PhD thesis was conducted at the Department of Clinical genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. Part of this work was conducted on behalf of the International Hirschsprung Disease Consortium.

The production costs of the thesis were supported by:

Erasmus University Rotterdam, The Netherlands

Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands

ISBN: 978-90-8219-688-7

Cover: Rajendra Kumar Chauhan, Chaitali Paul and Tom de Vries-Lentsch

Chapter illustrations: Chaitali Paul

Layout: Chaitali Paul

Printing: PrintSupport4U

Genetic and Functional Studies of Hirschsprung Disease

Genetische en Functionele Studies voor ziekte van Hirschsprung

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the

rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision by the Doctorate Board.

The public defense shall be held on

Tuesday 1 November 2016 at 13:30 hrs

by

Rajendra Kumar Chauhan

born in Sarkaghat, Mandi, India

Erasmus University Rotterdam

Doctoral Committee:

Promoter:	Prof.dr. R.M.W. Hofstra
Assessment committee:	Prof.dr. D. Tibboel
	Prof.dr. R. Willemsen
	Dr. I. Shepherd
Co-promoters:	Dr. B.J.L. Eggen
	Dr. A.J. Burns

Table of Contents

List of Abbreviations

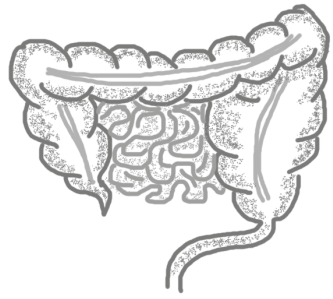
ATP5O	ATP Synthase, H+ Transporting, Mitochondrial F1 Complex, O Subunit
5HT	5-hydroxytryptamin
ARNT2	Aryl-hydracarbon receptor translocator 2
ARTN	Artemin
ASD	Autism spectrum disorder
AVPR2	Arginine Vasopressin Receptor2
BWA	Burrows-Wheeler aligner
CCHS	Congenital Central Hypoventilation syndrome
CGRP	Calcitonin gene-related peptide
Chr	Chromosome
CNS	Central nervous system
CNVs	Copy number variants
DAPPLE	Disease Association Protein-Protein Link Evaluator
DHH	Desert hedgehog
DNA	Deoxyribonucleic acid
DNM	De novo mutation
DNMT3B	DNA (Cytosine-5-)-Methyltransferase 3 Beta
DPF	Days post fertilization
DS	Down syndrome
DSCAM	Down syndrome cell adhesion molecule
ECE1	Endothelin converting enzyme 1
EDC	Erasmus Dierexperimenteel Centrum
EDN3	Endothelin 3
EDNRB	Endothelin receptor type B
EGF	Epidermal growth factor
EMT	Epithelial to mesenchymal transformation
ENCC	Enteric neural crest cells
ENS	Enteric nervous system
ERK	Extracellular signal–regulated kinase
FD	Familial dysautonomia
GATK	Genome Analysis Toolkit
GDNF	Glial cell line-derived neurotrophic factor
GFRα1	GDNF family receptor alpha-1
GI	Gastrointestinal
GLI	GLI Family Zinc Finger
GOSHS	Goldberg-Shprintzen syndrome
GPI	Glycosylphosphatidylinositol
GWAS	Genome wide association studies
hESCs	Human embryonic stem cells

Hh	Hedgehog
HNK1	Human natural killer-1
HOXB5	Homeobox B5
Hsa21	Human chromosome 21
HSCR	Hirschsprung disease
ICC	Interstitial cell of cajal
IGV	Integreted Genome Viewer
IHH	Indian hedgehog
IKBKAP	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex associated protein
IPA	Ingenuity pathway analysis
iPS	Induced pluripotent stem (iPS)
JAK/STAT	Janus kinase/signal transducers and activators of transcription
JNK	Jun amino-terminal kinase
KBP	Kinesin binding protein
KIF26A	Kinesin Family Member 26 A
L1CAM	L1 Cell Adhesion Molecule
L-HSCR	Long segment Hirschsprung disease
LOD	Logarithm of the odds
LOF	Loss of function
LRBA	LPS Responsive Beige-Like Anchor Protein
MEN2A	Multiple endocrine neoplasia of type 2A
MEN2B	Multiple endocrine neoplasia of type 2B
MO	Morpholino
mRNA	Messenger ribonucleic acid
MTC	Medullary thyroid carcinoma
NAV2	Neuron Navigator 2
NCC	Neural crest cells
NCLN	Nicalin
NCSC	Neural crest stem cell
NGS	Next Generation Sequencing
NKA	Neurkinin A
NLBs	Neurosphere like bodies
NO	Nitric oxide
NPC	Neural progenitor cell
NRG1	Neuregulin 1
NRTN	Neuturin
NUP98	Nucleoporin 98kDa
NXF (NPAS4)	Neuronal PAS domain protein 4
OSCP	Oligomycin Sensitivity Conferral Protein
PACAP	Pituitary adenylate cyclase activating peptide

PCR	Polymerase chain reaction
PHOX2B	Paired-like homeobox 2b
PI3K/AKT	Phosphatidylinositol 3-kinase/AKT
PKC	Protein kinase C
PNS	Peripheral nervous system
PSPN	Persephin
PTCH1	Patched 1
QA	Quality assesment
RAIR	Recto-anal inhibitory reflex
RAS/MAPK	Ras/mitogen activated protein kinase
RET	Rearranged during transfection
SBMOs	Splice-blocking morpholinos
SDM	Site directed mutagenesis
SEMA	Semaphorin
SHH	Sonic hedgehog
S-HSCR	Short segment Hirschsprung disease
SMO	Smoothened
SNP	Single nucleotide polymorphism
SNVs	Single nucleotide variants
SOX10	SRY (Sex determining region Y)-box 10
TBATA	Thymus, Brain And Testes Associated
TBMOs	Translation-blocking morpholinos
TCA	Total colonic aganglionosis
TF	Transcription factor
TIA	Total intestinal aganglionosis
TSS	Transcriptional Start Site
TTF1	Thyroid transcription factor 1
UTR	Untranslated region
VCF	Variant call format
VIP	Vasoactive intestinal polypeptide
WES	Whole Exome Sequencing
WGS	Whole genome sequencing
WISH	Whole mount in situ hybridization
WS4	Waardenburg syndrome, type IV
WT	Wild type
ZEB2	Zinc finger E-box-binding homeobox2
ZIRC	Zebrafish International Resource Centrum

CHAPTER 1

General Introduction and Scope of the Thesis



ABSTRACT

Complex (genetic) diseases are caused by many genetic, epigenetic and environmental factors that in concert result in a disease phenotype. Identifying the contribution of an individual gene(s), epigenetic aberrations or environmental factors is extremely challenging. It is this understanding of complex diseases that has become the major topic in the field of human genetics. Hirschsprung disease (HSCR) is one such complex genetic disorder. It is the most common forms of congenital obstruction of the bowel, and results from a failure of the neural crest-derived progenitor cells of the enteric nervous system (ENS) to migrate, proliferate, differentiate or survive in the gut wall during early embryonic development. The phenotype of this defect(s) is a variable length of aganglionosis in the distal part of the bowel. Since the 1990s, a multitude of genetic studies based on linkage analysis, homozygosity mapping, and genome wide association studies (GWAS) resulted in the identification of many susceptibility loci and genes involved in this complex disease. In the last decade, extraordinary progress has been made in genome sequencing technologies, collectively referred to as Next Generation Sequencing (NGS). This has greatly enhanced our knowledge and understanding of the role of novel genes and genetic variability in the pathogenesis of diseases. Combining NGS-based strategies with traditional linkage or expression studies has resulted in the identification of new HSCR genes. Moreover, the use of *in vitro* and *in vivo* assays to establish genotype-phenotype associations has further enhanced our understanding of the different mechanisms associated with ENS development in general and HSCR in particular. In this thesis, we aim to better understand and unravel the complexity of HSCR genetics using genomics approach and in vivo modelling of HSCR in zebrafish model.

THE ENTERIC NERVOUS SYSTEM (ENS)

The gastrointestinal (GI) tract is an internal organ which requires the coordinated activity of its neuromuscular components for the mixing and propulsion of food, for breakdown of complex foods during digestion, and for secretion, absorption and excretion. The functions of the GI tract are governed by the ENS, an extensive network of neurons and glial cells that form a meshwork of interconnected ganglia along the entire bowel[1]. These comprise the outer myenteric (Auerbach's) plexus which resides between the circular and longitudinal smooth muscle layers, and the inner submucosal (Meissner's) plexus.

The myenteric plexus provides motor innervation to the muscle layers and secretomotor innervation to the mucosa[1].

Embryonic origin of ENS

The entire ENS is derived from the neural crest (NC). The NC is a transient population of cells that emerges/detaches from the dorsal neural tube during early embryonic development. This happens after undergoing epithelial to mesenchymal transformation (EMT) and these neural crest cells (NCC) start migration to various locations throughout the embryo. NCC are multipotent and differentiate into a wide range of cell types during vertebrate development including elements of the craniofacial skeleton, peripheral neurons, glia, melanocytes and connective, endocrine and adipose tissues. The neural crest is the distinguishing feature of vertebrates and its regulation is highly conserved among humans and many other vertebrate species. The neural crest origin of the ENS was first demonstrated by using avian embryos to show that upon ablation of the vagal NC region, the enteric ganglia failed to form in the gut[2]. Subsequently, a number of classical transplantation experiments using chick-quail chimeras were fundamental in understanding the fate of NCC and established the vagal neural crest (adjacent to somite 1–7) as the major source of ENS precursors[3]. Vagal NCC colonizes the gut by rostro-caudal migration along the entire length of the gut and a further contribution to the distal ENS arises from sacral NCC (posterior to somite 28). It was shown in mouse and chick that these cells colonize the most distal part of the GI tract by migration in an opposing caudo-rostral direction (Figure 1)[4-7]. It has also been reported that the anterior NCCs contribute to the foregut ENS[8]. NCCs enter mouse foregut at embryonic day E 9.5, and colonization of the mouse gut is complete by E14.5. In humans, NCCs enter the foregut by week 4 and at week 7 colonization of the entire gut is complete[7,9,10].

Enteric NCC (ENCC) derived cells also undergo a secondary inward radial migration in mice after initial colonization to form mucosal ganglia[11]. Contrary to the inwards migration, there is also an outwards migration from the submucosa in the case of avian[4]. It has also been shown that during early development of mice (E10.5-11.5), the vagal neural crest derived ENCC also migrate in a trans-mesenteric direction perpendicular to rostro-caudal migration[12]. In contrast to these complex migrations of ENCC, the zebrafish ENS completely derives from vagal NC and there is no evidence to support a sacral NCC contribution to the ENS[13]. In humans, as enteric NCCs migrate along the gut, they proliferate and differentiate into different neuronal subtypes and glial cells to form interconnected ganglia (Figure 1). Defects in the development of NCCs result in myriad of neurocristopathies. One of the most common diseases affecting the ENS

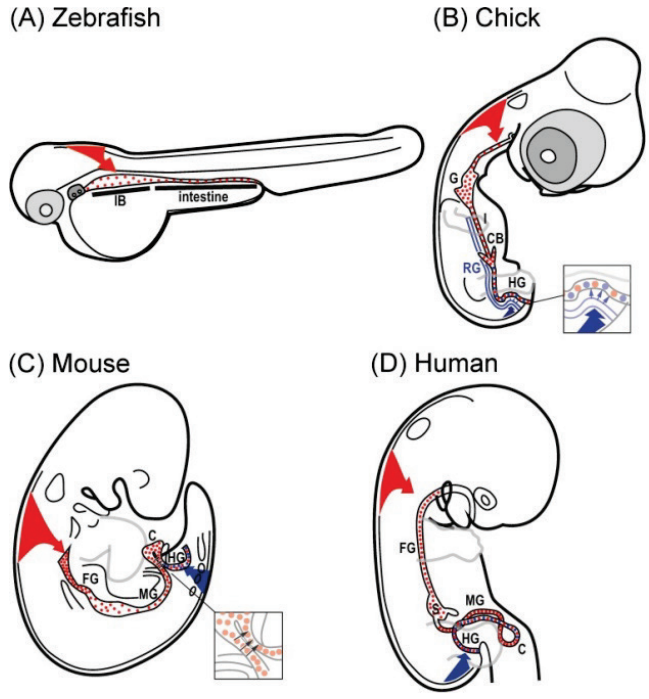


Figure 1. Representation of embryonic origin of the ENS in diverse vertebrate models during early development. Embryonic origins of the ENS in diverse vertebrate models. (A) The ENS of zebrafish derives from vagal NCC (red arrow) that enters the rostral gut tube at approximately 36 hours post-fertilization (hpf). Cells migrate caudally and progressively colonize the intestinal bulb (IB) and intestine. The gut is fully colonized by these vagal neural crest-derived ENS progenitors (red dots) by 66 hpf. (B) In the chick, the ENS is formed primarily from vagal NCC at the level of somite 1–7 (red arrow) that enter the foregut (FG) at approximately embryonic day (E) 3–3.5 and migrate caudally to progressively colonize the gizzard (G) (mechanical stomach), intestine (I), cecal buds (CB) and hindgut, a process that is complete by E7.5 (red dots). Sacral NCC, arising caudal to somite 28 (blue arrow), also contribute to the ENS, first forming the extramural nerve of Remak (RG) (blue), and then migrating into the hindgut (inset, blue arrows) to colonize primarily the distal hindgut (blue dots). (C) The mouse ENS is formed principally from vagal NCC from the level of somite 1–7 (red arrow), which enter the foregut at approximately E9, and migrate caudally to colonize the foregut (FG), midgut (MG), cecum (C), and hindgut (HG) (red dots). In addition to rostrocaudal migration, trans-mesenteric migration of vagal NCC from the midgut to the hindgut also occurs (inset, arrows). Colonization of the length of the gut is complete by E14. An additional ENS contribution arises from NCC that migrate from the sacral region (blue arrow). These cells initially form pelvic ganglia adjacent to the hindgut, then migrate into the gut and primarily occupy the hindgut and caudal midgut (blue dots). (D) In the human, the ENS derives from vagal NCC (red arrow) that enter the foregut (FG) at 4 weeks of gestation and migrate along the gut to fully colonize the foregut, stomach (S), midgut (MG), cecum (C), and hindgut (HG) by week 7 (red dots). It is inferred, from mouse data, that sacral NCC also contribute to the hindgut ENS (blue hatched arrow), however no experimental evidence is yet available to confirm this. (Adapted from[14]).

is HSCR, which is attributed to the failure of NCCs to migrate, differentiate, proliferate or survive and thereby form a functional ENS. HSCR research is

concentrated on gaining a better understanding of the underlying pathophysiology of this enteric neuropathy, including the genetics.

HIRSCHSPRUNG DISEASE

HSCR, also known as congenital megacolon, is the most common causes of neonatal intestinal obstruction, if left untreated then it is life threatening (Figure 2). The first description of HSCR comes from ancient writings (between 1200 BC and 600 BC) of Ayurvedic Sushruta Samhita. It has described "Baddha Gudodaram," a condition analogous to modern day HSCR[15]. The name Hirschsprung originates from Harald Hirschsprung, a Danish physician who in 1888 described 2 boys, aged 8 and 11 months, respectively, who died due to severe constipation[16]. The first clinical description of what we now call HSCR is ascribed to a Dutch anatomist "Frederick Ruysch" in 1691, who reported a 5 year old girl with abdominal pain "*enormis intestine cono dilatatio*"[17]. The relation between congenital megacolon and bowel aganglionosis was not understood until 1948 and until then HSCR remained a fatal disease (Figure 2). Pioneer studies by Swenson and Bill along with others recognized the histopathological features of HSCR by using full thickness rectal biopsies and correlated it with the absence of enteric ganglia in the intestinal segment below the dilated part of colon and these findings became significant for diagnosis and surgical intervention[18-22]. These studies allowed development of a simple and reliable diagnostic confirmation for HSCR using histochemical staining for acetylcholinesterase (AchE)[23]. These findings led to the identification of the underlying cause of the severe constipation seen in HSCR; an absence of enteric neurons in the myentric (Auerbach's) plexus and the submucosal (Meissner's) plexus in a length of the gut.

Clinical features and diagnosis

HSCR is clinically characterized by failure to pass meconium (the first stool) within the first 48 hours after birth, severe constipation, bilious vomiting, abdominal distention and recurrent neonatal enterocolitis[24]. Physical examination of children with HSCR show an enlarged abdominal circumference with numerous fecal masses (Figure 2)[25].

HSCR is usually diagnosed with radiographic studies combined with barium enema, anorectal manometry and rectal biopsies. Imaging studies incorporating non-invasive radiograph tests can be obtained first for children and infants suspected of having HSCR with history of pain. Abdominal X-ray showing a distended small bowel and proximal colon and an empty rectum can give an

early indication of HSCR. A water soluble (contrast) enema typically shows a narrow distal segment, a funnel-shaped dilatation, characteristically localized at the level of the transition zone, and a marked dilatation of the proximal colon and poor emptying after 24 hours. Anorectal manometry measures the contractility in the anus and rectum. Usually, anal and rectal muscles are tightened to hold on bowel movement and they relax in order to pass feces. Absence of the recto-anal inhibitory reflex (RAIR), when the rectum is distended, can be helpful in diagnosing HSCR[26], although there is a perception that it is unnecessary in most cases[27]. The gold standard for a definitive diagnosis of HSCR relies on histopathology of a full thickness rectal biopsy. In this biopsy the pathologist searches for the absence of ganglia at the plexuses (myenteric and submucosal) of gut wall. Pathological evaluation of rectal biopsies is based on enzymatic histochemistry using frozen sections to establish the presence of ganglia and the analysis of AchE positive nerve fibers. Another approach is based on paraffin sections stained with hematoxylin and eosin (H&E), and sometimes complemented with AchE histochemistry[28,29].

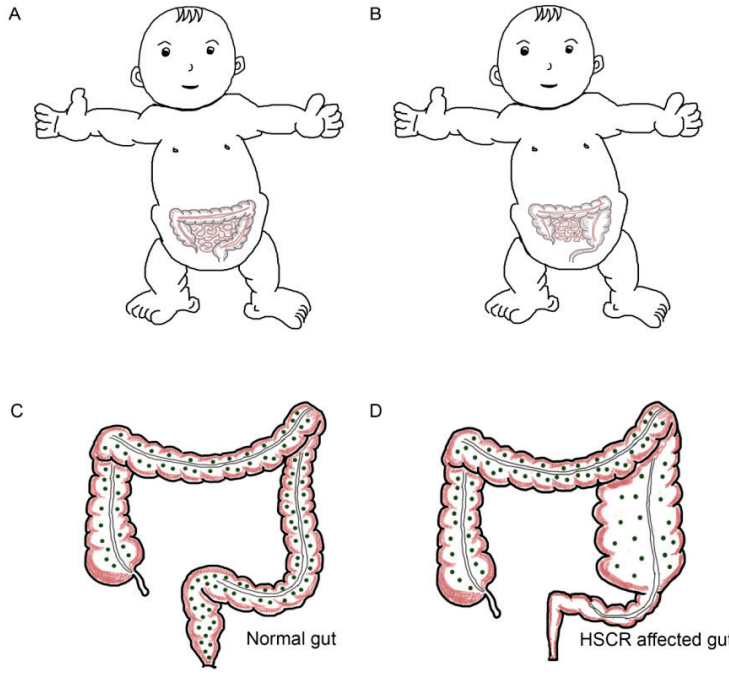


Figure 2. Hirschsprung disease (A) Cartoon of a child with a normal colon. (B) Cartoon of child with HSCR (intestinal megacolon). (C) Enteric ganglia are represented by green dots in the large intestine and in a normal gut the large intestine is fully colonized. (D) Aganglionosis of distal colon can be seen in HSCR affected colon.

Classification

Based on the variability in the length of affected region, HSCR is further classified as short-segment HSCR (S-HSCR), long segment HSCR (L-HSCR) or total colonic aganglionosis (TCA). S-HSCR affects around 80% of patients and the aganglionosis does not extend beyond the sigmoid colon. In the case of L-HSCR, (which affects 15-20% of patients) the aganglionosis is also observed proximal to the sigmoid colon. In TCA (around 5% of cases), the aganglionosis affects the entire large intestine and may also affect the small intestine. In very rare cases, the whole bowel is affected and known as total intestinal aganglionosis (TIA).

Incidence and prevalence

The prevalence of HSCR is estimated to be ~1 in 5000 newborns, however this varies between ethnic groups (1.0/5000 for Hispanics, 1.5/5000 for Caucasians, 2.1/5000 for African-Americans and 2.8/5000 for Asians) and males are affected more than the females (4:1) in all ethnic populations[30-33]. The male preponderance persists and decreases with the length of agangionosis. It varies from 4:1 in S-HSCR to 1:1-1:2 in L-HSCR and reverses to 0.8:1 in TCA[34-36]. There is no evidence of X-linked loci in HSCR and the exact cause behind sex bias remains largely unexplained.

Association with other anomalies

HSCR occurs as an isolated trait in 70% of the cases. Consequently, in 30% of the cases HSCR is associated with other congenital anomalies, which includes in addition to other GI malformations, cleft palate, cardiac malformations, craniofacial anomalies and polydactyly[37]. These anomalies can occur by chance or can be part of a (un)known syndrome, such as Waardenburg-Shah syndrome, type IV (WS4), Congenital Central Hypoventilation syndrome (CCHS), Mowat-Wilson syndrome, cartilage-hair hypoplasia syndrome, Goldberg-Shprintzen syndrome (GOSHS) and Smith-Lemli-Opitz syndrome[37].

The most common syndrome of which HSCR is part of is Down syndrome (DS). Therefore, it is not surprising that trisomy 21 is the most common chromosomal abnormality seen in HSCR. The overall incidence of DS ranges from 2-10% in all HSCR cases[37-39]. In total chromosomal abnormalities are identified in ~12% of all the HSCR cases.

Treatment and future therapies

Current treatment for HSCR consists of surgical resection (pull-through) of the aganglionic segment of the intestine and reconnection of the proximal bowel to the anus[40]. Alternatively, individuals with extensive intestinal aganglionosis may

require intestinal transplantation[41]. Over the past decades the surgical management of HSCR has continued to evolve with the adoption of minimally invasive approaches. However, even after surgical treatment for HSCR, it is associated with long term morbidities such as constipation, fecal soiling and enteritis[42].

In recent years there has been much interest in the use of stem cell therapy for HSCR. Here the idea is that ENS progenitor cells/stem cells could be transplanted into the aganglionic region of the bowel to reconstruct the absent ENS[43]. In support of this approach, cultures of multipotent ENS progenitors that generate neurospheres or neurosphere like bodies (NLBs) containing ENCCs have been shown to differentiate to form mature neurons and glial cells from embryonic and postnatal mouse gut[44,45]. Human enteric neurospheres have also been isolated and grown successfully from full thickness and mucosal biopsies obtained from fetal, postnatal and adult bowel[46-50]. Upon transplantation, these ENCCs can colonize recipient bowel *in vitro* and have the capability of migration, proliferation and neuroglial differentiation[44-47]. Moreover, recent *in vivo* transplantation studies have demonstrated functional integration of ENCCs with the endogenous ENS[51,52]. ENS progenitors have also been isolated from the aganglionic gut of HSCR patients and it was found that p75-positive neural crest derived cells present in the thickened nerve trunk gave rise to neurons in culture[53].

Simultaneous studies have focused on generation of neural crest derived peripheral neurons from mouse and human embryonic stem cells (hESCs) using *in vitro* differentiation protocols[54,55]. It has been shown that neural crest stem cells (NCSC) can be derived from *in vitro* differentiated hESCs expressing neural crest markers such as p75, HNK1 (human natural killer-1) and SOX10 (SRY (Sex determining region Y)-box 10), which give rise to multiple neural crest lineages[56,57]. Transplantation of hESCs derived NCSCs into the recipient chick embryo and adult mouse demonstrated survival, migration and differentiation *in vivo*[57]. Similarly, human induced pluripotent stem (iPS) cells can be differentiated *in vitro* to NCSCs that can be further differentiated into neurons and glial cells[58]. More recently, the *in vitro* differentiation of hESCs and human pluripotent stem cells into ENS progenitors and their further differentiation into functional enteric neurons have been established[59]. Transplantation of *in vitro* derived ENS precursors displayed targeted migration in the chick embryo, colonization of mouse gut, and reversal of disease-associated mortality in an HSCR mouse model[59]. These studies have also led to the identification of pepstetin A as a novel candidate therapeutic target that reverses the impaired migratory potential of NCC[59]. Despite these advancements and development of new treatment strategies

based on the stem cell therapy of the affected ENS, there are numerous concerns, such as optimal source, delivery method and safety concerns among others and these issues are being approached by many scientists to reach up to a consensus and address them for future clinical trials[60].

HSCR GENETICS

Segregation studies in HSCR have demonstrated that the recurrence risk to siblings varies from 1.5-33% depending on the sex of the affected person and extent of aganglionosis[31,33]. HSCR is considered an inherited disease based on the fact that there are familial cases (~ 5%), and an overall elevated risk of 4% (relative risk as high as 200) of the disease in siblings the presence of chromosomal abnormalities and occurrence of naturally occurring animal models[37].

HSCR susceptibility loci by linkage analysis

The first linkage studies on large multigenerational HSCR families identified 10q11.2 as the major locus for HSCR[61,62]. These studies led to the earliest identification of *RET* (Rearranged during transfection) gene mutations in HSCR probands[63,64].

Bolk *et al.* conducted a linkage analysis study on 12 multiplex HSCR cases and found that all but one family showed linkage to the *RET* locus. Only half of the families carried a *RET* coding variant. Intriguingly, the families that did not have a *RET* coding variant showed linkage to 9q31. It was hypothesized that the gene in 9q31 might be a modifier for development of HSCR[65].

In another study by Gabriel *et al.* a genome-wide scan was conducted in small HSCR families with S-HSCR. Sib pair analysis identified susceptibility loci at 3p21 and 19q12 in addition to the *RET* locus[66]. A fifth locus at 13q21 was identified by Puffenberger *et al.* after performing linkage analysis in 43 Mennonite trios belonging to the same kindred. Within the linkage region, *EDNRB* (Endothelin Receptor type B) was pinpointed as the causative HSCR gene. The linkage study also revealed the presence of genetic modifier of HSCR on 21q22[67]. A sixth locus was found by studying a large multi-generational Dutch family with isolated HSCR. It resulted in the identification of a HSCR susceptible locus at 4q31.3-q32.3[68].

HSCR susceptibility loci by GWAS

In HSCR, three GWAS have been performed mainly on sporadic HSCR cases from different ethnicities to identify additional HSCR genetic loci that could contribute to the disease risk[69-71]. GWAS conducted in a Chinese population found association with *NRG1* (Neurogelin1)[69]. GWAS performed by the International HSCR Consortium on HSCR patients of European ancestry found an association downstream from the protein *SEMA3D* (Semaphorin3D) and upstream from *SEMA3A* (Semaphorin 3A) and mutational screening of these genes identified several coding variants in these genes[72]. Another GWAS performed on Thai population also found an association with *RET* and *NRG1*[71].

Table 1. HSCR-associated genes and loci

Gene	Locus	Phenotype	Frequency of the coding mutations	Inheritance
RET	10q11.2	Non-syndromic HSCR/ MEN2A	50% familial, 15-35% sporadic	Dominant, incomplete penetrance
GDNF	5p13.1	Non-syndromic HSCR	Rare	Non-Mendelian
GFRA1	10q25.3	Non-syndromic HSCR	1 case reported	Dominant
NRTN	19p13.3	Non-syndromic HSCR	Very rare	Non-Mendelian
PSPN	19p13.3	Non-syndromic HSCR	Very rare	Non-Mendelian
EDNRB	13q22.3	Non-syndromic HSCR, WS	3-7%	Dominant (*de novo* in 80%)
EDN3	20q13.32	Non-syndromic HSCR, WS	<5%	Dominant, incomplete penetrance Recessive
ECE1	1p36.12	HSCR, craniofacial and cardiac defects	1 case reported	Dominant
NRG1	8p12	Non-syndromic HSCR	6%	Dominant, incomplete penetrance
NRG3	10q23.1	Non-syndromic HSCR	Rare	Dominant, incomplete penetrance
SEMA3C	7q21.11	Non-syndromic HSCR	Rare	Non-Mendelian
SEMA3D	7q21.11	Non-syndromic HSCR	Rare	Non-Mendelian
SOX10	22q13.1	Non-syndromic HSCR, WS	>5%	Dominant (*de novo* in 75%)
PHOX2B	4p13	CCHS	<5%	Dominant (*de novo* in 90%)
ZFHX1B / ZEB2	2q22.3	Mowat-Wilson syndrome	<5%	Dominant (de novo in 100%)
TCF4	18q21.2	Pitt-Hopkins syndrome	1 case reported	Dominant
NKX2-1 / TTF1	14q13.3	Non-syndromic HSCR	1 case reported	Dominant
L1CAM	Xq28	X-linked hydrocephalus and HSCR	Rare	X-linked dominant
DSCAM	21q22.2	Non-syndromic HSCR, DS	Association of common variants	Non-Mendelian
KBP / KIAA1279	10q22.1	Goldberg-Schprintzen syndrome	Rare	Recessive
DNMT3B	20q11.21	Non-syndromic HSCR	Rare	Dominant, incomplete penetrance
PTCH1	9q22.32	Non-syndromic HSCR	Association of common variants	Non-Mendelian
DLL3	6q27	Non-syndromic HSCR	Association of common variants	Non-Mendelian
IKBKAP	9q31	Non-syndromic HSCR, dysautonomia	Association of common variants	Dominant, incomplete penetrance
Unknown	3p21	Non-syndromic HSCR		Non-Mendelian
Unknown	16q23	WS		Non-Mendelian
Unknown	4q31-q32	Non-syndromic HSCR		Dominant, incomplete penetrance
Unknown	19q12	Non-syndromic HSCR		Non-Mendelian
Unknown	21q22	WS		Recessive

HSCR: Hirschsprung disease, MEN2A: multiple endocrine neoplasia type 2, WS: Waardenburg-Shah syndrome, CCHS: Congenital central hypoventilation syndrome, DS: Down syndrome

Genes associated with HSCR

To date mutations in 16 genes (Table 1, Figure 3) have been identified that can cause, or contribute to the development of HSCR[70,73].

RET gene mutations

The proto-oncogene *RET*, is the predominant gene associated with HSCR. *RET* is considered to be the major HSCR gene, as 50% of the familial cases and 15-35% of the sporadic cases have a mutation in the *RET* coding region or in the regions involved in mRNA (messenger ribonucleic acid) splicing[74]. *RET* mutations in HSCR generally result in loss of function of the encoded protein supporting a haplo-insufficiency effect in disease pathogenesis[75].

A variety of *RET* mutations have been identified, including microdeletions, insertions, frameshifts, splice variants, nonsense and missense mutations. They can be found along the entire length of gene[74,76,77]. The inactivating mutations of *RET* affect its function due to various molecular mechanisms such as, protein misfolding, failure of protein transport to the cell surface and suppression of its biological activity[75,78-80]. The penetrance of *RET* mutations is incomplete in familial HSCR and it is higher in males (72%) than in females (51%) supporting the existence of one of more modifier genes to develop the disease[76]. It should be noted that *RET* mutations are also implied in other pathologies: multiple endocrine neoplasia of type 2A (MEN2A) and 2B (MEN2B) and familial and sporadic forms of medullary thyroid carcinoma (MTC) and papillary thyroid cancer[81]. In all these diseases, however the mutations are affecting specific codons and result in an activation of the mutated protein. Moreover, HSCR can be found in association with MEN2A and familial MTC with mutation in *RET* gene[82,83].

Non-coding mutations and HSCR

As mentioned earlier, *RET* coding mutations have been identified in 50% of the familial cases of HSCR and failure to identify coding region mutations in some of the *RET*-linked families suggested that mutations in *RET* regulatory regions might contribute significantly to the disease[65,66]. This idea was further corroborated in studies by different groups on sporadic HSCR cases that consisted Caucasians and Asian population with and without coding mutation in *RET*. In all these studies a common disease associated *RET* haplotype was identified[84-88]. This haplotype spans approximately 27 Kb and it includes 4Kb of 5' UTR (untranslated region), exon1, intron1 and exon2. Common variants in the *RET* promoter (rs10900296 and rs10900297) upstream of the *RET* transcription start site were identified[85]. These SNPs (single nucleotide polymorphisms) were shown to reduce the

binding affinity of the transcription factor TTF1 (Thyroid transcription factor 1) and thereby the variant was believed to result in a decrease of RET transcription[89]. However, reduction in RET expression was later found to be cell line dependent[90]. Comparative genomics approaches identified a multispecies conserved enhancer region in intron1 of RET and within this enhancer region two SNPs (rs2435357 and rs2506004) are present. For these SNPs it was shown that both could negatively influence the enhancer activity leading to lower RET expression, independently[91,92]. It was also shown that SOX10 binding was disrupted by rs2435357, and rs2506004 served as binding site for NXF-ARNT2 and SIM2-ARNT2 transcription factor heterodimers, respectively[92,93]. Taken together, these findings strongly support that both rare, coding mutations and common non-coding variants in *RET* contribute to HSCR development. RET is a transmembrane tyrosine kinase receptor expressed at highest levels during early embryogenesis in the developing excretory system, in all lineages of the PNS, and in motor and catecholaminergic neurons of the central nervous system (CNS)[94]. Alternative splicing generates three *RET* isoforms containing 51 (RET51), 43 (RET43) and 9 (RET9) amino acids in the carboxyl (C)-terminal tail[95]. RET has a large extracellular domain, a transmembrane region and an intracellular kinase domain[96]. It is a signaling receptor for four ligands, GDNF (glial cell line-derived neurotrophic factor), NRTN (neuturin), ARTN (artemin) and PSPN (persephin) (Figure 3)[97].

These ligands activate RET by binding to the GPI (glycosyl phosphatidyl inosated)-linked GDNF family of receptors (GFRα1-4) respectively. Upon binding of the ligand-co-receptor complex, RET dimerization and autophosphorylation of the tyrosine residues occur in the intracellular domain. These tyrosine residues act as docking sites for adapter and signaling proteins to stimulate multiple downstream pathways[98,99]. These pathways include JAK-STAT, RAS-MAPK, PI3-AKT, ERK, PKC and JNK, that can promote cell growth, proliferation, survival or differentiation[100].

The RET-GDNF-GFRα signaling pathway

Through *Ret*, GDNF/GFRα1 signaling stimulates ENS progenitor proliferation, directional migration, survival and differentiation along the developing gut[101-107]. Deletion of *Ret* in mouse leads to complete intestinal aganglionosis[108]. *ret* knockdown in zebrafish embryos also display absence of enteric neurons in the gut[13].

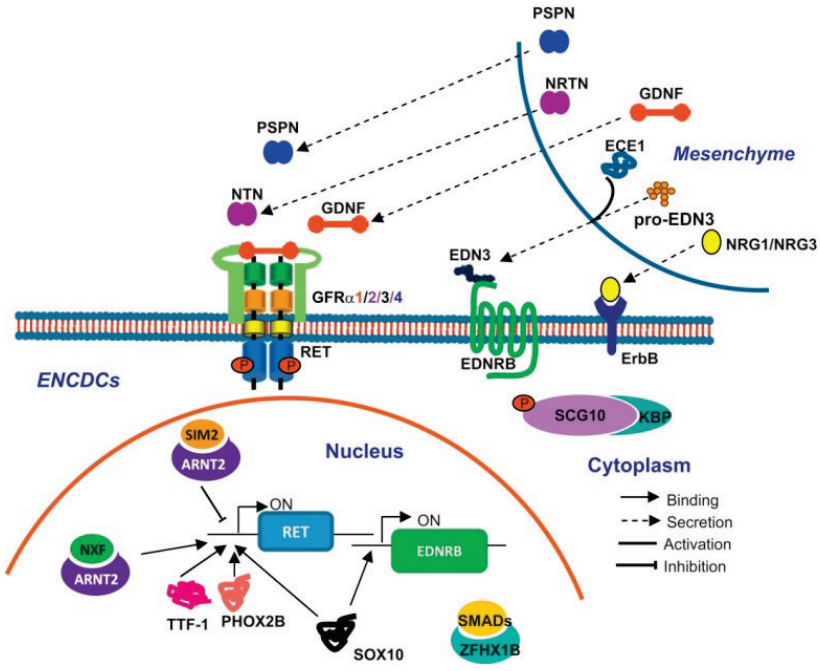


Figure 3.
Proteins encoded by the identified HSCR susceptibility genes and possible interaction between the different protein identified (Adapted from[73]).

GDNF acts as a chemo-attractant for the RET positive vagal ENCCs and it is highly expressed in the stomach ahead of the migrating ENCCs wave-front and its expression is elevated in the caecum, while ENCCs migrate towards the distal part of the gut[107,109]. Heterozygous *GDNF* mutations have also been reported in sporadic HSCR patients[74,110,111]. Recently, it has also been found that a kinesin, KIF26A acts a negative regulator of GDNF-Ret signaling in ENS development[112].

NRTN and GFRα2 are other members of RET ligand-co-receptor complex that have been implicated in ENS development. The ENS is formed in adult *NRTN* and *GFRα2*-null mice, but myenteric neurons display decreased soma size and fewer excitatory nerve fibres in the myenteric plexus[113-115]. In rare cases of HSCR, patients have been identified carrying mutations in *NRTN*[116].

The EDNRB/EDN3 Signaling Pathway

A second signaling pathway involved with HSCR and in ENS development is the Endothelin Receptor type B (EDNRB) pathway. EDNRB is a G-protein coupled receptor expressed in NC derivatives and Endothelin 3 (EDN3) mediates its activation. EDN3-EDNRB signaling is required for the development of melanocytes and enteric neurons[117]. HSCR patients have been identified carrying

mutations in EDNRB, EDN3 and the endothelin converting enzyme ECE1 (which converts an inactive precursor form of EDN3 into an active form)[118-120]. Mutations in these genes are present in approximately 5% of HSCR patients[121]. Inherited mutations in these genes are generally seen in the context of Waardenburg-Shah syndrome, type IV (WS4), a disorder that includes pigmentation defects, sensorineural deafness, dysmorphic facial features and aganglionic megacolon in humans[122]. Avian studies have established that EDNRB transcripts are present in NCC , before as well as during their emigration from the neural tube at all levels of the neuraxis and EDN3 dramatically enhances the proliferation of NCC[123]. The mutant mice carrying mutations in *Ednrb*, *Edn3* and *Ece1* also exhibit aganglionosis and pigmentation defects[124-126]. EDNRB/EDN3 signaling is involved with regulation of ENCCs migration as EDNRB is expressed by migrating ENCCs, whereas *Edn3* is expressed in the midgut and hindgut mesoderm and highly expressed in caecum and proximal colon[127,128]. Other studies have also suggested a role of EDNRB/EDN3 signaling in ENS development as activation of EDNRB by EDN3 induces enteric NCCs to proliferate, maintain their precursor state and prevent premature differentiation[102,129,130].

Neuregulin signaling

A GWAS conducted in a Chinese population found association with *NRG1* (Neuregulin1). The neuregulins (NRGs) are cell-cell signaling proteins that are ligands for receptor tyrosine kinases of the ErbB family. *NRG1* is believed to be involved in ENS development and it interacts with major HSCR gene *RET*[69].

Fine mapping of *NRG1* locus by SNP genotyping resulted in identification of four highly associated SNPs on *NRG1* promoter[131]. Later on, implication of *NRG1* in HSCR was demonstrated through the identification of pathogenic coding mutations using different functional approaches[132]. Involvement of *NRG1* variants in the etiology of HSCR was further confirmed in Thai HSCR population and in Caucasian HSCR patients as well[133,134]. Moreover, copy number variants (CNVs) in a paralog of *NRG1* gene, *NRG3* were found corroborating the importance of neuregulin signaling in HSCR[135]. Exome sequencing studies in Chinese family have also identified *NRG3* as a susceptible gene for HSCR[136].

The protein encoded by *NRG1* is a membrane glycoprotein that plays a critical role in the growth and development of multiple organ systems. A variety of different isoforms are derived from the neuregulin gene and the *NRG1* isoforms exert their effects through a heterodimeric complex consisting of members of the EGF (epidermal growth factor) receptor tyrosine kinases ErbB3/ErbB4 in the heart and ErbB2/ErbB3 in NCCs[137,138]. *NRG1* is suggested to promote the survival and maintenance of the ENS and it is expressed in both neurons and glial cells of

enteric ganglia, nerve fibers and it acts as a neurotrophic factor for the ENS[139]. Conditional knockout of *Erbb2* in mice display loss of enteric ganglia and distention of colon indicating the relevance of Nrg-ErbB signaling in murine ENS development[140].

Hedgehog Signaling

Mammals have three Hedgehog homologues, Desert (*DHH*), Indian (*IHH*), and Sonic (*SHH*). The hedgehog pathway is involved in the development of the ENS and hedgehog proteins have important function as morphogens. Activation of Hh signal requires two transmembrane proteins including a receptor Patched (ptch) and signal activator Smoothened (Smo). The downstream signal activation of Hh pathway is mainly mediated by a family of zinc-fingers containing the transcriptional factors, the Gli proteins (Gli1, Gli2 and Gli3) and they act as activators and/or repressors[141].

It has been shown that localized expression of hedgehog proteins in the epithelium is essential for concentric patterning of the bowel wall [142]. Sonic hedgehog promotes the proliferation and inhibits the differentiation of crest-derived cells from embryonic mice *in vitro*[143]. Mice lacking either the Indian Hh- or Shh-secreted proteins display partial intestinal aganglionosis, accompanied by megacolon or ectopic ganglia formation[142]. Pathed-1 (Ptch1), the receptor of Hedgehog ligands is expressed by enteric neural progenitors[143]. Deletion of Ptch1 in ENCCs results in elevated expression of Gli1 and inhibition of neurogenesis[144]. In zebrafish, sonic hedgehog is required for the migration of neural crest from the hindbrain into the anterior gut[145]. Recently, it has also been shown that some HSCR patients that lacked *RET* coding mutations have mutations in *GLI* genes[146].

Sema3C/D signaling

Semaphorins are transmembrane, secreted, or GPI-linked proteins known to be involved with neuronal migration, proliferation, survival, or axonal guidance[147]. These proteins are grouped into different classes and many subgroups. In the developing colon and cecum, semaphorin 3A (*Sema3A*) is expressed by the inner mesenchyme, while the coreceptor for Sema3A, neuropilin-1, is expressed by all enteric neural crest derived cells[148]. Earlier studies had also postulated a role of semaphorin signaling in ENS development[149,150]. Further descriptive studies of the associated region led to the functional validation of the semaphorin genes in the aetiology of HSCR. It was found that *Sema3a*, *Sema3c* and *Sema3d* were expressed in the mouse ENS and knockdown of *sema3* in zebrafish display reduction in the migration of ENS precursors[70].

Transcriptional regulation of ENS development and other genes

Peripheral autonomic neurons and ENS development is regulated by a network of transcriptional factors[151]. Mutations in transcriptional factors have majorly been implicated in the genetic etiology of syndromic forms of HSCR, such as WS4, CCHS and Mowat-Wilson syndrome.

SOX10

Transcriptional regulator Sox10 is expressed in the vagal NCC, ENCCs and its mutations disrupts neural crest development in a HSCR mouse model called 'Dom'[126,152]. The identification of a mutation in the Dom mouse led in finding of *SOX10* mutations in patients with Waardenburg-Hirschsprung disease and established its role in the development of ENS[153]. Cell death is increased in undifferentiated, postmigratory NCC lacking *Sox10*[154]. One of the earliest zebrafish mutants of the ENS, cls (colourless) mutant, was identified in a genetic screen for pigmentation defects. *cls* serves as a Waardenburg-Shah syndrome model since fish lack pigment cells, and have reduced enteric neurons as well as additional NCC defects[155]. Subsequently, it was shown that the *cls* locus mapped to the *sox10* gene which is known to be required for neural crest development[156].

PHOX2B

HSCR is also associated with congenital central hypoventilation syndrome (CCHS), a disorder characterized by an idiopathic failure of the automatic control of breathing and primarily caused due to mutations in *PHOX2B* (Paired-like homeobox 2b) gene[157,158]. The homeobox protein Phox2b is essential for the development of autonomic neural crest derivatives and it is expressed by migrating ENCCs, and mice lacking Phox2b display aganglionosis due to failure of ENCCs to colonize the gut[159]. Similarly, *Phox2b* function is conserved in zebrafish and upon its knockdown, a HSCR-like phenotype is observed[160]. Like *SOX10*, *PHOX2B* is also required for Ret expression in enteric NCCs[159].

ZEB2/ZFHX1B or *SIP1*

ZEB2 (Zinc finger E-box-binding homeobox2) is a transcription factor involved in neural specification and in epithelial-mesenchymal transition (EMT) during early neural crest development. Mowat-Wilson syndrome, characterized by mental retardation, facial abnormalities, epilepsy along with HSCR is caused by mutations in *ZEB2*[161,162]. *Zeb2*$^{-/-}$ mice show a complete lack of vagal neural crest and die around E9.5[163]. Targeted ablation of *Zeb2* in neural crest of mice displays craniofacial, heart, pigment, PNS deformities and aganglionosis of entire colon extending up to the small intestine[162].

L1CAM

L1CAM (L1 Cell Adhesion Molecule) encodes a neuronal cell adhesion molecule and it is the only known X-linked gene associated with HSCR, but only a small subset (3%) of patients with mutations in *L1CAM* display HSCR and it is thought to be an X-linked HSCR modifier gene[164,165].

IKBKAP

Fine mapping of the RET dependent modifier in the 9q31 Hirschsprung's disease locus pointed towards *IKBKAP* (inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex associated protein) as the most likely candidate gene[166]. Depletion of *Ikbkap* in zebrafish leads to a HSCR disease - like phenotype[167].

HOXB5

Disruption of the transcription factor Hoxb5 (Homeobox B5) results in Ret haploinsufficiency and failure of ENCCs migration in the distal colon of mouse gut[168]. Perturbation of Hoxb5 in mice displays Sox9 downregulation, NCC apoptosis, hypoplastic sympathetic and dorsal root ganglia, hypopigmentation and ENS defects[169].

DSCAM

DSCAM (Down syndrome cell adhesion molecule) is a cell adhesion molecule and using SNP association analysis, it was postulated that excessive of DSCAM may explain HSCR associated Down syndrome, but this association has not been yet confirmed experimentally in a model system[170].

KBP

Goldberg-Shprintzen syndrome (GOSHS) is a syndromic form of HSCR. It is characterized by polymicrogyria, mental retardation, microcephaly, facial dysmorphisms and in most cases, by HSCR[171,172]. Homozygosity mapping in a consanguineous family identified homozygous nonsense mutations in *KIAA1279* (now called *KBP*)[171]. KBP (Kinesin binding protein) interacts with microtubule associated proteins and it is required for neuronal differentiation and neurite outgrowth[173]. *KIAA1279* mutations have not been yet identified in isolated HSCR patients.

DNMT3B

DNMT3B (DNA (Cytosine-5-)-Methyltransferase 3 Beta) encodes for one of the *de novo* methyltransferases and it is essential in establishing CpG methylation patterns and it is proposed to play a role in ENS development and HSCR. Recently, it has been described that ENCCs isolated from HSCR patients display lower level of *DNMT3B* expression as compared to the control individuals and damaging mutations were found in a HSCR patient cohort[174].

Genetic interaction between RET and EDNRB signaling

Initially, there was no any connection between RET and EDNRB signaling and these pathways were believed to work independently. Now it has been demonstrated that interaction between the RET and EDNRB signaling pathways does exist and control ENS development throughout the intestine[127]. Moreover, genome-wide association studies conducted on a Mennonite population, where the incidence of HSCR was ten-fold higher (1/500) than in the normal situation, showed a statistically significant co-transmission of *EDNRB* and *RET* alleles in affected individuals[175]. Activation of EDNRB specifically enhances the effect of RET signaling on the proliferation of uncommitted ENS progenitors[127]. EDNRB has also been shown to modify the migratory response of NCC to GDNF[176]. Finally, it has been shown that partial loss of Ednrb in Sox10 heterozygous mice impairs colonization of the gut by enteric crest cells[177].

Non-genetic factors and HSCR

Few studies have focused on the hypothesis of involvement of non-genetic factors in the ENS development. It has been shown that adequate vitamin A levels are required during early gestation for proper development of ENS[178]. It was also found that mycophenolate, an inhibitor of *de novo* guanine nucleotide biosynthesis impaired ENS development in mice and zebrafish[179]. In yet another study by the same group, a common used drug ibuprofen was also shown to disturb bowel colonization by ENS precursors in zebrafish, mouse and chick[180]. These studies provide the earliest evidences of involvement of non-genetic factors influencing the ENS development and probably contribute to HSCR in some cases.

SCOPE OF THE THESIS

HSCR is a heterogeneous, complex genetic disease involving mutations (in combination) in several genes. The differential contribution of rare and common coding and non-coding variants vary in accordance with gender and length of aganglionosis. Major mutations in genes involved in HSCR are also linked to ENS development, and defects in ENCCs migration, proliferation, differentiation and survival display HSCR like phenotype in animal models. Exome sequencing of families with unidentified mutations in known HSCR genes have led in identification of new HSCR genes. Mutations in regulatory elements encompassing the RET locus are associated with HSCR. Detailed epigenome profiling of different cell types and tissues by the Human epigenome atlas project have now provided desired datasets for interrogating the role of epigenetic marks in HSCR. Development of new transgenic models in model organisms such as, zebrafish has also facilitated the validation studies of genomic findings. All these studies identified a large number of genes and loci however they explain only part of the total genetic risk for HSCR. In this thesis we aim at finding novel genes, mutation within these genes and ENS specific regulatory regions that could explain part of the missing heritability.

An overview of the ENS development, HSCR disease diagnosis, treatment and pathophysiology of HSCR is described in **chapter 1**.

In **chapter 2**, we describe the role of *de novo* mutations in long-segment HSCR. Since genes carrying *de novo* mutations were not linked to ENS development based on bioinformatics prediction, we tested the functional contribution of these genes to ENS development in a zebrafish model.

In a previous linkage study performed by Brooks *et al.* on a multigeneration Dutch family with HSCR, the 4q31-32 region was identified as a new HSCR susceptibility locus. In **chapter 3,** we tried to unravel the genetics within this family focusing not only on genes within the linkage region, but also on non-linked candidate HSCR genes.

Most of the HSCR research had been mainly focused to identify coding variants in the genes associated with HSCR, although coding mutations had not been able to explain more than 25 % of all the cases. **Chapter 4** describes our strategy to identify novel regulatory elements involved with ENS development and potentially with HSCR

As the incidence of HSCR is over a hundred times higher in Down syndrome patients than in the general population, it is hypothesized that the trisomy of one or more genes on chromosome 21 contribute to the development

of HSCR. To find this gene(s), we overexpressed highly conserved chromosome 21 mRNAs in a transgenic zebrafish. This work is described in **chapter 5**.

Pinpointing the functional relevance of newly identified genetic variants in HSCR is rather difficult. In **chapter 6**, we review the use of the zebrafish model in HSCR research and screening ENS for functional validation of newly identified disease variants.

Finally, **in chapter 7** we summarize and discuss the work presented in this thesis and we discuss work that might be done in future to further understand HSCR genetics and disease development.

REFERENCES

1. Furness, J.B. & Wiley, I. The enteric nervous system. (Blackwell Pub., Malden, Mass., 2006).
2. Yntema, C.L. & Hammond, W.S. The origin of intrinsic ganglia of trunk viscera from vagal neural crest in the chick embryo. *J Comp Neurol* **101**, 515-41 (1954).
3. Le Douarin, N.M. & Teillet, M.A. The migration of neural crest cells to the wall of the digestive tract in avian embryo. *J Embryol Exp Morphol* **30**, 31-48 (1973).
4. Burns, A.J. & Douarin, N.M. The sacral neural crest contributes neurons and glia to the post-umbilical gut: spatiotemporal analysis of the development of the enteric nervous system. *Development* **125**, 4335-47 (1998).
5. Wang, X., Chan, A.K., Sham, M.H., Burns, A.J. & Chan, W.Y. Analysis of the sacral neural crest cell contribution to the hindgut enteric nervous system in the mouse embryo. *Gastroenterology* **141**, 992-1002 e1-6 (2011).
6. Gershon, M.D., Chalazonitis, A. & Rothman, T.P. From neural crest to bowel: development of the enteric nervous system. *J Neurobiol* **24**, 199-214 (1993).
7. Wallace, A.S. & Burns, A.J. Development of the enteric nervous system, smooth muscle and interstitial cells of Cajal in the human gastrointestinal tract. *Cell Tissue Res* **319**, 367-82 (2005).
8. Durbec, P.L., Larsson-Blomberg, L.B., Schuchardt, A., Costantini, F. & Pachnis, V. Common origin and developmental dependence on c-ret of subsets of enteric and sympathetic neuroblasts. *Development* **122**, 349-58 (1996).
9. Young, H.M. *et al.* A single rostrocaudal colonization of the rodent intestine by enteric neuron precursors is revealed by the expression of Phox2b, Ret, and p75 and by explants grown under the kidney capsule or in organ culture. *Dev Biol* **202**, 67-84 (1998).
10. Kapur, R.P., Yost, C. & Palmiter, R.D. A transgenic model for studying development of the enteric nervous system in normal and aganglionic mice. *Development* **116**, 167-75 (1992).
11. Jiang, Y., Liu, M.T. & Gershon, M.D. Netrins and DCC in the guidance of migrating neural crest-derived cells in the developing bowel and pancreas. *Dev Biol* **258**, 364-84 (2003).
12. Nishiyama, C. *et al.* Trans-mesenteric neural crest cells are the principal source of the colonic enteric nervous system. *Nat Neurosci* **15**, 1211-8 (2012).
13. Shepherd, I.T., Pietsch, J., Elworthy, S., Kelsh, R.N. & Raible, D.W. Roles for GFRalpha1 receptors in zebrafish enteric nervous system development. *Development* **131**, 241-9 (2004).
14. Heanue, T.A., Shepherd, I.T. & Burns, A.J. Enteric nervous system development in avian and zebrafish models. *Dev Biol* (2016).
15. Raveenthiran, V. Knowledge of ancient Hindu surgeons on Hirschsprung disease: evidence from Sushruta Samhita of circa 1200-600 BC. *J Pediatr Surg* **46**, 2204-8 (2011).
16. Hirschsprung, H. Struhltragheit Neugeborener in folge von Dilatation and Hypertrophie des colons. **Jahr- buch Kinderheilkunde 27**, 1-7 (1888).

17. Leenders, E. & Sieber, W.K. Congenital megacolon observation by Frederick Ruysch--1691. *J Pediatr Surg* **5**, 1-3 (1970).
18. Swenson, O. & Bill, A.H., Jr. Resection of rectum and rectosigmoid with preservation of the sphincter for benign spastic lesions producing megacolon; an experimental study. *Surgery* **24**, 212-20 (1948).
19. Swenson, O., Rheinlander, H.F. & Diamond, I. Hirschsprung's disease; a new concept of the etiology; operative results in 34 patients. *N Engl J Med* **241**, 551-6 (1949).
20. Bodian, M., Stephens, F.D. & Ward, B.C. Hirschsprung's disease and idiopathic megacolon. *Lancet* **1**, 6-11 (1949).
21. Zuelzer, W.W. & Wilson, J.L. Functional intestinal obstruction on a congenital neurogenic basis in infancy. *Am J Dis Child* **75**, 40-64 (1948).
22. Whitehouse, F.R. & Kernohan, J.W. Myenteric plexus in congenital megacolon; study of 11 cases. *Arch Intern Med (Chic)* **82**, 75-111 (1948).
23. Meier-Ruge, W. *et al.* Acetylcholinesterase activity in suction biopsies of the rectum in the diagnosis of Hirschsprung's disease. *J Pediatr Surg* **7**, 11-7 (1972).
24. Puri, P. & Montedonico, S. Hirschsprung's Disease: Clinical Features. in *Hirschsprung's Disease and Allied Disorders* (eds. Holschneider, M.A. & Puri, P.) 107-113 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).
25. Sergi, C. Hirschsprung's disease: Historical notes and pathological diagnosis on the occasion of the 100(th) anniversary of Dr. Harald Hirschsprung's death. *World J Clin Pediatr* **4**, 120-5 (2015).
26. Emir, H. *et al.* Anorectal manometry during the neonatal period: its specificity in the diagnosis of Hirschsprung's disease. *Eur J Pediatr Surg* **9**, 101-3 (1999).
27. Vult von Steyern, K., Wingren, P., Wiklund, M., Stenstrom, P. & Arnbjornsson, E. Visualisation of the rectoanal inhibitory reflex with a modified contrast enema in children with suspected Hirschsprung disease. *Pediatr Radiol* **43**, 950-7 (2013).
28. Knowles, C.H. *et al.* Gastrointestinal neuromuscular pathology: guidelines for histological techniques and reporting on behalf of the Gastro 2009 International Working Group. *Acta Neuropathol* **118**, 271-301 (2009).
29. Kapur, R.P. Practical pathology and genetics of Hirschsprung's disease. *Semin Pediatr Surg* **18**, 212-23 (2009).
30. Torfs, C.P. An epidemiological study of Hirschsprung disease in a multiracial California population. (1998).
31. Bodian, M. & Carter, O.O. A family study of Hirschsprung's disease. *Annals of Human Genetics* **26**, 261–277 (1963).
32. Passarge, E. The Genetics of Hirschsprung's Disease. *New England Journal of Medicine* **276**, 138-143 (1967).
33. Badner, J.A., Sieber, W.K., Garver, K.L. & Chakravarti, A. A genetic study of Hirschsprung disease. *Am J Hum Genet* **46**, 568-80 (1990).
34. Orr, J.D. & Scobie, W.G. Presentation and incidence of Hirschsprung's disease. *Br Med J (Clin Res Ed)* **287**, 1671 (1983).
35. Ikeda, K. & Goto, S. Diagnosis and treatment of Hirschsprung's disease in Japan. An analysis of 1628 patients. *Ann Surg* **199**, 400-5 (1984).
36. Kleinhaus, S., Boley, S.J., Sheran, M. & Sieber, W.K. Hirschsprung's disease -- a survey of the members of the Surgical Section of the American Academy of Pediatrics. *J Pediatr Surg* **14**, 588-97 (1979).
37. Amiel, J. *et al.* Hirschsprung disease, associated syndromes and genetics: a review. *J Med Genet* **45**, 1-14 (2008).
38. Moore, S.W. & Johnson, A.G. Hirschsprung's disease: genetic and functional associations of Down's and Waardenburg syndromes. *Semin Pediatr Surg* **7**, 156-61 (1998).
39. Friedmacher, F. & Puri, P. Hirschsprung's disease associated with Down syndrome: a meta-analysis of incidence, functional outcomes and mortality. *Pediatr Surg Int* **29**, 937-46 (2013).
40. Swenson, O. Early history of the therapy of Hirschsprung's disease: facts and personal observations over 50 years. *J Pediatr Surg* **31**, 1003-8 (1996).
41. Bond, G.J. & Reyes, J.D. Intestinal transplantation for total/near-total aganglionosis and intestinal pseudo-obstruction. *Semin Pediatr Surg* **13**, 286-92 (2004).

42. Gershon, M.D. Transplanting the enteric nervous system: a step closer to treatment for aganglionosis. *Gut* **56**, 459-61 (2007).
43. Hotta, R. & Thapar, N. Advances in enteric neurobiology: how close are we to clinical use? *J Pediatr Gastroenterol Nutr* **53 Suppl 2**, S43-5 (2011).
44. Natarajan, D., Grigoriou, M., Marcos-Gutierrez, C.V., Atkins, C. & Pachnis, V. Multipotential progenitors of the mammalian enteric nervous system capable of colonising aganglionic bowel in organ culture. *Development* **126**, 157-68 (1999).
45. Bondurand, N., Natarajan, D., Thapar, N., Atkins, C. & Pachnis, V. Neuron and glia generating progenitors of the mammalian enteric nervous system isolated from foetal and postnatal gut cultures. *Development* **130**, 6387-400 (2003).
46. Almond, S., Lindley, R.M., Kenny, S.E., Connell, M.G. & Edgar, D.H. Characterisation and transplantation of enteric nervous system progenitor cells. *Gut* **56**, 489-96 (2007).
47. Lindley, R.M. *et al.* Human and mouse enteric nervous system neurosphere transplants regulate the function of aganglionic embryonic distal colon. *Gastroenterology* **135**, 205-216.e6 (2008).
48. Metzger, M., Caldwell, C., Barlow, A.J., Burns, A.J. & Thapar, N. Enteric nervous system stem cells derived from human gut mucosa for the treatment of aganglionic gut disorders. *Gastroenterology* **136**, 2214-25.e1-3 (2009).
49. Hetz, S. *et al.* In vivo transplantation of neurosphere-like bodies derived from the human postnatal and adult enteric nervous system: a pilot study. *PLoS One* **9**, e93605 (2014).
50. Rauch, U., Hansgen, A., Hagl, C., Holland-Cunz, S. & Schafer, K.H. Isolation and cultivation of neuronal precursor cells from the developing human enteric nervous system as a tool for cell therapy in dysganglionosis. *Int J Colorectal Dis* **21**, 554-9 (2006).
51. Cooper, J.E. *et al.* In Vivo Transplantation of Enteric Neural Crest Cells into Mouse Gut; Engraftment, Functional Integration and Long-Term Safety. *PLoS One* **11**, e0147989 (2016).
52. Hotta, R. *et al.* Isogenic enteric neural progenitor cells can replace missing neurons and glia in mice with Hirschsprung disease. *Neurogastroenterol Motil* **28**, 498-512 (2016).
53. Wilkinson, D.J., Bethell, G.S., Shukla, R., Kenny, S.E. & Edgar, D.H. Isolation of Enteric Nervous System Progenitor Cells from the Aganglionic Gut of Patients with Hirschsprung's Disease. *PLoS One* **10**, e0125724 (2015).
54. Mizuseki, K. *et al.* Generation of neural crest-derived peripheral neurons and floor plate cells from mouse and primate embryonic stem cells. *Proc Natl Acad Sci U S A* **100**, 5828-33 (2003).
55. Pomp, O., Brokhman, I., Ben-Dor, I., Reubinoff, B. & Goldstein, R.S. Generation of peripheral sensory and sympathetic neurons and neural crest cells from human embryonic stem cells. *Stem Cells* **23**, 923-30 (2005).
56. Jiang, X. *et al.* Isolation and characterization of neural crest stem cells derived from in vitro-differentiated human embryonic stem cells. *Stem Cells Dev* **18**, 1059-70 (2009).
57. Lee, G. *et al.* Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nat Biotechnol* **25**, 1468-75 (2007).
58. Chambers, S.M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol* **27**, 275-80 (2009).
59. Fattahi, F. *et al.* Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. *Nature* **531**, 105-9 (2016).
60. Burns, A.J. *et al.* White paper on guidelines concerning enteric nervous system stem cell therapy for enteric neuropathies. *Dev Biol* (2016).
61. Lyonnet, S. *et al.* A gene for Hirschsprung disease maps to the proximal long arm of chromosome 10. *Nat Genet* **4**, 346-50 (1993).
62. Angrist, M. *et al.* A gene for Hirschsprung disease (megacolon) in the pericentromeric region of human chromosome 10. *Nat Genet* **4**, 351-6 (1993).
63. Edery, P. *et al.* Mutations of the RET proto-oncogene in Hirschsprung's disease. *Nature* **367**, 378-80 (1994).
64. Romeo, G. *et al.* Point mutations affecting the tyrosine kinase domain of the RET proto-oncogene in Hirschsprung's disease. *Nature* **367**, 377-8 (1994).
65. Bolk, S. *et al.* A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. *Proc Natl Acad Sci U S A* **97**, 268-73 (2000).

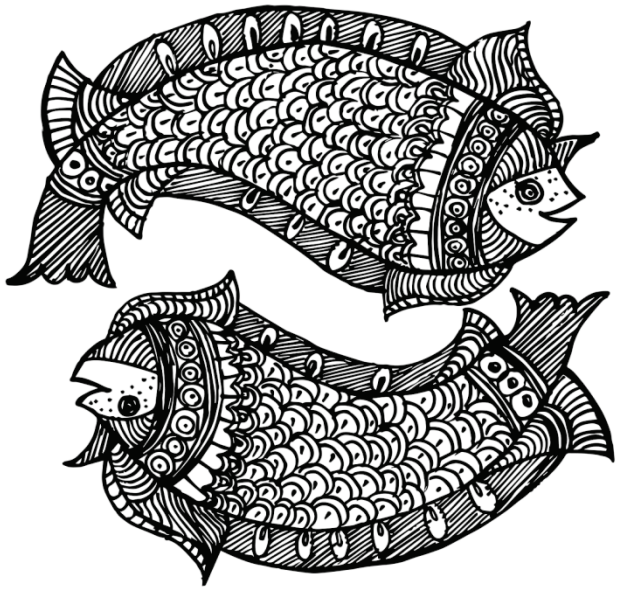
66. Gabriel, S.B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* **31**, 89-93 (2002).
67. Puffenberger, E.G. *et al.* Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum Mol Genet* **3**, 1217-25 (1994).
68. Brooks, A.S. *et al.* A novel susceptibility locus for Hirschsprung's disease maps to 4q31.3-q32.3. *J Med Genet* **43**, e35 (2006).
69. Garcia-Barcelo, M.M. *et al.* Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc Natl Acad Sci U S A* **106**, 2694-9 (2009).
70. 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**, 581-96 (2015).
71. Kim, J.H. *et al.* A genome-wide association study identifies potential susceptibility loci for Hirschsprung disease. *PLoS One* **9**, e110292 (2014).
72. Luzon-Toro, B. *et al.* Mutational spectrum of semaphorin 3A and semaphorin 3D genes in Spanish Hirschsprung patients. *PLoS One* **8**, e54800 (2013).
73. Alves, M.M. *et al.* Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev Biol* **382**, 320-9 (2013).
74. Hofstra, R.M. *et al.* RET and GDNF gene scanning in Hirschsprung patients using two dual denaturing gel systems. *Hum Mutat* **15**, 418-29 (2000).
75. Pasini, B. *et al.* Loss of function effect of RET mutations causing Hirschsprung disease. *Nat Genet* **10**, 35-40 (1995).
76. Attie, T. *et al.* Diversity of RET proto-oncogene mutations in familial and sporadic Hirschsprung disease. *Hum Mol Genet* **4**, 1381-6 (1995).
77. Angrist, M. *et al.* Mutation analysis of the RET receptor tyrosine kinase in Hirschsprung disease. *Hum Mol Genet* **4**, 821-30 (1995).
78. Carlomagno, F. *et al.* Molecular heterogeneity of RET loss of function in Hirschsprung's disease. *Embo J* **15**, 2717-25 (1996).
79. Pelet, A. *et al.* Various mechanisms cause RET-mediated signaling defects in Hirschsprung's disease. *J Clin Invest* **101**, 1415-23 (1998).
80. Iwashita, T., Murakami, H., Asai, N. & Takahashi, M. Mechanism of ret dysfunction by Hirschsprung mutations affecting its extracellular domain. *Hum Mol Genet* **5**, 1577-80 (1996).
81. de Groot, J.W., Links, T.P., Plukker, J.T., Lips, C.J. & Hofstra, R.M. RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. *Endocr Rev* **27**, 535-60 (2006).
82. Takahashi, M. *et al.* Co-segregation of MEN2 and Hirschsprung's disease: the same mutation of RET with both gain and loss-of-function? *Hum Mutat* **13**, 331-6 (1999).
83. Peretz, H. *et al.* Cys 618 Arg mutation in the RET proto-oncogene associated with familial medullary thyroid carcinoma and maternally transmitted Hirschsprung's disease suggesting a role for imprinting. *Hum Mutat* **10**, 155-9 (1997).
84. Burzynski, G.M. *et al.* Localizing a putative mutation as the major contributor to the development of sporadic Hirschsprung disease to the RET genomic sequence between the promoter region and exon 2. *Eur J Hum Genet* **12**, 604-12 (2004).
85. Fitze, G. *et al.* Functional haplotypes of the RET proto-oncogene promoter are associated with Hirschsprung disease (HSCR). *Hum Mol Genet* **12**, 3207-14 (2003).
86. Pelet, A. *et al.* Homozygosity for a frequent and weakly penetrant predisposing allele at the RET locus in sporadic Hirschsprung disease. *J Med Genet* **42**, e18 (2005).
87. Sancandi, M. *et al.* Single nucleotide polymorphic alleles in the 5' region of the RET proto-oncogene define a risk haplotype in Hirschsprung's disease. *J Med Genet* **40**, 714-8 (2003).
88. Borrego, S. *et al.* A founding locus within the RET proto-oncogene may account for a large proportion of apparently sporadic Hirschsprung disease and a subset of cases of sporadic medullary thyroid carcinoma. *Am J Hum Genet* **72**, 88-100 (2003).
89. Garcia-Barcelo, M. *et al.* TTF-1 and RET promoter SNPs: regulation of RET transcription in Hirschsprung's disease. *Hum Mol Genet* **14**, 191-204 (2005).
90. Griseri, P. *et al.* A common haplotype at the 5' end of the RET proto-oncogene, overrepresented in Hirschsprung patients, is associated with reduced gene expression. *Hum Mutat* **25**, 189-95 (2005).

91. Emison, E.S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857-63 (2005).
92. Sribudiani, Y. *et al.* Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology* **140**, 572-582 e2 (2011).
93. 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).
94. Marcos, C. & Pachnis, V. The effect of the ret- mutation on the normal development of the central and parasympathetic nervous systems. *Int J Dev Biol* **Suppl 1**, 137S-138S (1996).
95. Myers, S.M., Eng, C., Ponder, B.A. & Mulligan, L.M. Characterization of RET proto-oncogene 3' splicing variants and polyadenylation sites: a novel C-terminus for RET. *Oncogene* **11**, 2039-45 (1995).
96. Takahashi, M. *et al.* Cloning and expression of the ret proto-oncogene encoding a tyrosine kinase with two potential transmembrane domains. *Oncogene* **3**, 571-8 (1988).
97. Baloh, R.H., Enomoto, H., Johnson, E.M., Jr. & Milbrandt, J. The GDNF family ligands and receptors - implications for neural development. *Curr Opin Neurobiol* **10**, 103-10 (2000).
98. Arighi, E., Borrello, M.G. & Sariola, H. RET tyrosine kinase signaling in development and cancer. *Cytokine Growth Factor Rev* **16**, 441-67 (2005).
99. Ibanez, C.F. Structure and physiology of the RET receptor tyrosine kinase. *Cold Spring Harb Perspect Biol* **5**(2013).
100. Mulligan, L.M. RET revisited: expanding the oncogenic portfolio. *Nat Rev Cancer* **14**, 173-86 (2014).
101. Gianino, S., Grider, J.R., Cresswell, J., Enomoto, H. & Heuckeroth, R.O. GDNF availability determines enteric neuron number by controlling precursor proliferation. *Development* **130**, 2187-98 (2003).
102. Hearn, C.J., Murphy, M. & Newgreen, D. GDNF and ET-3 differentially modulate the numbers of avian enteric neural crest cells and enteric neurons in vitro. *Dev Biol* **197**, 93-105 (1998).
103. Chalazonitis, A., Rothman, T.P., Chen, J. & Gershon, M.D. Age-dependent differences in the effects of GDNF and NT-3 on the development of neurons and glia from neural crest-derived precursors immunoselected from the fetal rat gut: expression of GFRalpha-1 in vitro and in vivo. *Dev Biol* **204**, 385-406 (1998).
104. Taraviras, S. *et al.* Signalling by the RET receptor tyrosine kinase and its role in the development of the mammalian enteric nervous system. *Development* **126**, 2785-97 (1999).
105. Wang, H. *et al.* The timing and location of glial cell line-derived neurotrophic factor expression determine enteric nervous system structure and function. *J Neurosci* **30**, 1523-38 (2010).
106. Young, H.M. *et al.* GDNF is a chemoattractant for enteric neural cells. *Dev Biol* **229**, 503-16 (2001).
107. Natarajan, D., Marcos-Gutierrez, C., Pachnis, V. & de Graaff, E. Requirement of signalling by receptor tyrosine kinase RET for the directed migration of enteric nervous system progenitor cells during mammalian embryogenesis. *Development* **129**, 5151-60 (2002).
108. Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F. & Pachnis, V. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. *Nature* **367**, 380-3 (1994).
109. Young, H.M. & Newgreen, D. Enteric neural crest-derived cells: origin, identification, migration, and differentiation. *Anat Rec* **262**, 1-15 (2001).
110. Angrist, M., Bolk, S., Halushka, M., Lapchak, P.A. & Chakravarti, A. Germline mutations in glial cell line-derived neurotrophic factor (GDNF) and RET in a Hirschsprung disease patient. *Nat Genet* **14**, 341-4 (1996).
111. Ivanchuk, S.M., Myers, S.M., Eng, C. & Mulligan, L.M. De novo mutation of GDNF, ligand for the RET/GDNFR-alpha receptor complex, in Hirschsprung disease. *Hum Mol Genet* **5**, 2023-6 (1996).
112. Zhou, R., Niwa, S., Homma, N., Takei, Y. & Hirokawa, N. KIF26A is an unconventional kinesin and regulates GDNF-Ret signaling in enteric neuronal development. *Cell* **139**, 802-13 (2009).
113. Heuckeroth, R.O. *et al.* Gene targeting reveals a critical role for neurturin in the development and maintenance of enteric, sensory, and parasympathetic neurons. *Neuron* **22**, 253-63 (1999).

114. Rossi, J. *et al.* Alimentary tract innervation deficits and dysfunction in mice lacking GDNF family receptor alpha2. *J Clin Invest* **112**, 707-16 (2003).
115. Rossi, J. *et al.* Retarded growth and deficits in the enteric and parasympathetic nervous system in mice lacking GFR alpha2, a functional neurturin receptor. *Neuron* **22**, 243-52 (1999).
116. Doray, B. *et al.* Mutation of the RET ligand, neurturin, supports multigenic inheritance in Hirschsprung disease. *Hum Mol Genet* **7**, 1449-52 (1998).
117. Shin, M.K., Levorse, J.M., Ingram, R.S. & Tilghman, S.M. The temporal requirement for endothelin receptor-B signalling during neural crest development. *Nature* **402**, 496-501 (1999).
118. Puffenberger, E.G. *et al.* A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* **79**, 1257-66 (1994).
119. Hofstra, R.M. *et al.* A homozygous mutation in the endothelin-3 gene associated with a combined Waardenburg type 2 and Hirschsprung phenotype (Shah-Waardenburg syndrome). *Nat Genet* **12**, 445-7 (1996).
120. Hofstra, R.M. *et al.* A loss-of-function mutation in the endothelin-converting enzyme 1 (ECE-1) associated with Hirschsprung disease, cardiac defects, and autonomic dysfunction. *Am J Hum Genet* **64**, 304-8 (1999).
121. Amiel, J. & Lyonnet, S. Hirschsprung disease, associated syndromes, and genetics: a review. *J Med Genet* **38**, 729-39 (2001).
122. Lake, J.I. & Heuckeroth, R.O. Enteric nervous system development: migration, differentiation, and disease. *Am J Physiol Gastrointest Liver Physiol* **305**, G1-24 (2013).
123. Nataf, V., Lecoin, L., Eichmann, A. & Le Douarin, N.M. Endothelin-B receptor is expressed by neural crest cells in the avian embryo. *Proc Natl Acad Sci U S A* **93**, 9645-50 (1996).
124. Baynash, A.G. *et al.* Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. *Cell* **79**, 1277-85 (1994).
125. Hosoda, K. *et al.* Targeted and natural (piebald-lethal) mutations of endothelin-B receptor gene produce megacolon associated with spotted coat color in mice. *Cell* **79**, 1267-76 (1994).
126. Yanagisawa, H. *et al.* Dual genetic pathways of endothelin-mediated intercellular signaling revealed by targeted disruption of endothelin converting enzyme-1 gene. *Development* **125**, 825-36 (1998).
127. Barlow, A., de Graaff, E. & Pachnis, V. Enteric nervous system progenitors are coordinately controlled by the G protein-coupled receptor EDNRB and the receptor tyrosine kinase RET. *Neuron* **40**, 905-16 (2003).
128. Leibl, M.A. *et al.* Expression of endothelin 3 by mesenchymal cells of embryonic mouse caecum. *Gut* **44**, 246-52 (1999).
129. Wu, J.J., Chen, J.X., Rothman, T.P. & Gershon, M.D. Inhibition of in vitro enteric neuronal development by endothelin-3: mediation by endothelin B receptors. *Development* **126**, 1161-73 (1999).
130. Nagy, N. & Goldstein, A.M. Endothelin-3 regulates neural crest cell proliferation and differentiation in the hindgut enteric nervous system. *Dev Biol* **293**, 203-17 (2006).
131. Tang, C.S. *et al.* Fine mapping of the NRG1 Hirschsprung's disease locus. *PLoS One* **6**, e16181 (2011).
132. Tang, C.S. *et al.* Mutations in the NRG1 gene are associated with Hirschsprung disease. *Hum Genet* **131**, 67-76 (2012).
133. Phusantisampan, T. *et al.* Association of genetic polymorphisms in the RET-protooncogene and NRG1 with Hirschsprung disease in Thai patients. *J Hum Genet* **57**, 286-93 (2012).
134. Luzon-Toro, B. *et al.* Comprehensive analysis of NRG1 common and rare variants in Hirschsprung patients. *PLoS One* **7**, e36524 (2012).
135. Tang, C.S. *et al.* Genome-wide copy number analysis uncovers a new HSCR gene: NRG3. *PLoS Genet* **8**, e1002687 (2012).
136. Yang, J. *et al.* Exome sequencing identified NRG3 as a novel susceptible gene of Hirschsprung's disease in a Chinese population. *Mol Neurobiol* **47**, 957-66 (2013).
137. Britsch, S. *et al.* The ErbB2 and ErbB3 receptors and their ligand, neuregulin-1, are essential for development of the sympathetic nervous system. *Genes Dev* **12**, 1825-36 (1998).

138. Garratt, A.N., Voiculescu, O., Topilko, P., Charnay, P. & Birchmeier, C. A dual role of erbB2 in myelination and in expansion of the schwann cell precursor pool. *J Cell Biol* **148**, 1035-46 (2000).
139. Barrenschee, M. *et al.* Expression and function of Neuregulin 1 and its signaling system ERBB2/3 in the enteric nervous system. *Front Cell Neurosci* **9**, 360 (2015).
140. Crone, S.A., Negro, A., Trumpp, A., Giovannini, M. & Lee, K.F. Colonic epithelial expression of ErbB2 is required for postnatal maintenance of the enteric nervous system. *Neuron* **37**, 29-40 (2003).
141. Hui, C.C., Slusarski, D., Platt, K.A., Holmgren, R. & Joyner, A.L. Expression of three mouse homologs of the Drosophila segment polarity gene cubitus interruptus, Gli, Gli-2, and Gli-3, in ectoderm- and mesoderm-derived tissues suggests multiple roles during postimplantation development. *Dev Biol* **162**, 402-13 (1994).
142. Ramalho-Santos, M., Melton, D.A. & McMahon, A.P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763-72 (2000).
143. Fu, M., Lui, V.C., Sham, M.H., Pachnis, V. & Tam, P.K. Sonic hedgehog regulates the proliferation, differentiation, and migration of enteric neural crest cells in gut. *J Cell Biol* **166**, 673-84 (2004).
144. Ngan, E.S. *et al.* Hedgehog/Notch-induced premature gliogenesis represents a new disease mechanism for Hirschsprung disease in mice and humans. *J Clin Invest* **121**, 3467-78 (2011).
145. Reichenbach, B. *et al.* Endoderm-derived Sonic hedgehog and mesoderm Hand2 expression are required for enteric nervous system development in zebrafish. *Dev Biol* **318**, 52-64 (2008).
146. Liu, J.A. *et al.* Identification of GLI Mutations in Patients With Hirschsprung Disease That Disrupt Enteric Nervous System Development in Mice. *Gastroenterology* **149**, 1837-1848 e5 (2015).
147. Kruger, R.P., Aurandt, J. & Guan, K.L. Semaphorins command cells to move. *Nat Rev Mol Cell Biol* **6**, 789-800 (2005).
148. Anderson, R.B. *et al.* Effects of different regions of the developing gut on the migration of enteric neural crest-derived cells: a role for Sema3A, but not Sema3F. *Dev Biol* **305**, 287-99 (2007).
149. Heanue, T.A. & Pachnis, V. Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. *Proc Natl Acad Sci U S A* **103**, 6919-24 (2006).
150. Shepherd, I.T. & Raper, J.A. Collapsin-1/semaphorin D is a repellent for chick ganglion of Remak axons. *Dev Biol* **212**, 42-53 (1999).
151. Howard, M.J. Mechanisms and perspectives on differentiation of autonomic neurons. *Dev Biol* **277**, 271-86 (2005).
152. Herbarth, B. *et al.* Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease. *Proc Natl Acad Sci U S A* **95**, 5161-5 (1998).
153. Pingault, V. *et al.* SOX10 mutations in patients with Waardenburg-Hirschsprung disease. *Nat Genet* **18**, 171-3 (1998).
154. Paratore, C., Goerich, D.E., Suter, U., Wegner, M. & Sommer, L. Survival and glial fate acquisition of neural crest cells are regulated by an interplay between the transcription factor Sox10 and extrinsic combinatorial signaling. *Development* **128**, 3949-61 (2001).
155. Kelsh, R.N. & Eisen, J.S. The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. *Development* **127**, 515-25 (2000).
156. Dutton, K.A. *et al.* Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development* **128**, 4113-25 (2001).
157. Amiel, J. *et al.* Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. *Nat Genet* **33**, 459-61 (2003).
158. Sasaki, A. *et al.* Molecular analysis of congenital central hypoventilation syndrome. *Hum Genet* **114**, 22-6 (2003).
159. Pattyn, A., Morin, X., Cremer, H., Goridis, C. & Brunet, J.F. The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. *Nature* **399**, 366-70 (1999).

160. Elworthy, S., Pinto, J.P., Pettifer, A., Cancela, M.L. & Kelsh, R.N. Phox2b function in the enteric nervous system is conserved in zebrafish and is sox10-dependent. *Mech Dev* **122**, 659-69 (2005).
161. Mowat, D.R. *et al.* Hirschsprung disease, microcephaly, mental retardation, and characteristic facial features: delineation of a new syndrome and identification of a locus at chromosome 2q22-q23. *J Med Genet* **35**, 617-23 (1998).
162. Wakamatsu, N. *et al.* Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. *Nat Genet* **27**, 369-70 (2001).
163. Van de Putte, T. *et al.* Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. *Am J Hum Genet* **72**, 465-70 (2003).
164. Griseri, P. *et al.* Complex pathogenesis of Hirschsprung's disease in a patient with hydrocephalus, vesico-ureteral reflux and a balanced translocation t(3;17)(p12;q11). *Eur J Hum Genet* **17**, 483-90 (2009).
165. Fernandez, R.M. *et al.* Association of X-linked hydrocephalus and Hirschsprung disease: report of a new patient with a mutation in the L1CAM gene. *Am J Med Genet A* **158A**, 816-20 (2012).
166. Tang, C.S. *et al.* Fine mapping of the 9q31 Hirschsprung's disease locus. *Hum Genet* **127**, 675-83 (2010).
167. Cheng, W.W. *et al.* Depletion of the IKBKAP ortholog in zebrafish leads to hirschsprung disease-like phenotype. *World J Gastroenterol* **21**, 2040-6 (2015).
168. Lui, V.C. *et al.* Perturbation of hoxb5 signaling in vagal neural crests down-regulates ret leading to intestinal hypoganglionosis in mice. *Gastroenterology* **134**, 1104-15 (2008).
169. Kam, M.K. *et al.* Perturbation of Hoxb5 signaling in vagal and trunk neural crest cells causes apoptosis and neurocristopathies in mice. *Cell Death Differ* **21**, 278-89 (2014).
170. Jannot, A.S. *et al.* Chromosome 21 scan in Down syndrome reveals DSCAM as a predisposing locus in Hirschsprung disease. *PLoS One* **8**, e62519 (2013).
171. Brooks, A.S. *et al.* Homozygous nonsense mutations in KIAA1279 are associated with malformations of the central and enteric nervous systems. *Am J Hum Genet* **77**, 120-6 (2005).
172. Goldberg, R.B. & Shprintzen, R.J. Hirschsprung megacolon and cleft palate in two sibs. *J Craniofac Genet Dev Biol* **1**, 185-9 (1981).
173. Alves, M.M. *et al.* KBP interacts with SCG10, linking Goldberg-Shprintzen syndrome to microtubule dynamics and neuronal differentiation. *Hum Mol Genet* **19**, 3642-51 (2010).
174. Torroglosa, A. *et al.* Involvement of DNMT3B in the pathogenesis of Hirschsprung disease and its possible role as a regulator of neurogenesis in the human enteric nervous system. *Genet Med* **16**, 703-10 (2014).
175. Carrasquillo, M.M. *et al.* Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. *Nat Genet* **32**, 237-44 (2002).
176. Kruger, G.M. *et al.* Temporally distinct requirements for endothelin receptor B in the generation and migration of gut neural crest stem cells. *Neuron* **40**, 917-29 (2003).
177. Stanchina, L. *et al.* Interactions between Sox10, Edn3 and Ednrb during enteric nervous system and melanocyte development. *Dev Biol* **295**, 232-49 (2006).
178. Fu, M. *et al.* Vitamin A facilitates enteric nervous system precursor migration by reducing Pten accumulation. *Development* **137**, 631-40 (2010).
179. Lake, J.I., Tusheva, O.A., Graham, B.L. & Heuckeroth, R.O. Hirschsprung-like disease is exacerbated by reduced de novo GMP synthesis. *J Clin Invest* **123**, 4875-87 (2013).
180. Schill, E.M. *et al.* Ibuprofen slows migration and inhibits bowel colonization by enteric nervous system precursors in zebrafish, chick and mouse. *Dev Biol* **409**, 473-88 (2016).



CHAPTER 2

De novo mutations in Hirschsprung patients link Central Nervous System genes to the development of the Enteric Nervous System

Hongsheng Gui[1,2*], Duco Schriemer[3*], William W.C. Cheng[1,4*], Rajendra K. Chauhan[4], Guillermo Antiňolo[5,6], Courtney Berrios[7], Marta Bleda[6,8], Alice S. Brooks[4], Rutger W.W. Brouwer[9], Alan J. Burns[4,10], Stacey S. Cherny[2], Joaquin Dopazo[5,6], Bart J.L. Eggen[3], Paola Griseri[11], Binta Jalloh[12], Thuy-Linh Le[13,14], Vincent C.H. Lui[1], Berta Luzón-Toro[5,6], Ivana Matera[11], Elly S.W. Ngan[1], Anna Pelet[13,14], Macarena Ruiz-Ferrer[5,6], Pak C. Sham[2], Iain T. Shepherd[12], Man-Ting So[1], Yunia Sribudiani[4,15], Clara S.M. Tang[1], Mirjam C.G.N. van den Hout[9], Wilfred F.J. van IJcken[9], Joke B.G.M. Verheij[16], Jeanne Amiel[13,14], Salud Borrego[5,6], Isabella Ceccherini[11], Aravinda Chakravarti[7], Stanislas Lyonnet[13,14], Paul K.H. Tam[1], Maria-Mercè Garcia-Barceló[1#] & Robert M.W. Hofstra[4,10#]

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

*Equally contributing authors; #Equally contributing corresponding authors.

Submitted

ABSTRACT

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

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

INTRODUCTION

Hirschsprung disease (HSCR) is the most common form of congenital obstruction of the bowel, with an incidence of ~1 per 5000 live births. However, the incidence varies significantly between ethnic groups with the highest incidence reported in the Asian population, with 2.8 per 10,000 live births[1,2]. HSCR results from a failure of the neural crest cells, that give rise to the enteric nervous system (ENS), to migrate, proliferate, differentiate or survive in the bowel wall, resulting in aganglionosis of the distal part of the gastrointestinal tract. This results in clinically severe and sometimes life-threatening bowel obstruction. As HSCR is a highly heritable disorder, genetic variation (mutations) in the genomes of these patients must largely explain disease development. The mode of inheritance of HSCR can be recessive mostly in syndromic cases, or dominant with incomplete penetrance in non-syndromic HSCR families, to oligogenic/polygenic in sporadic cases[3]. So far >15 HSCR susceptibility genes have been found as are 6 linkage regions[1] and three associated loci[2,4]. The genes identified belong to a limited number of pathways, which have been shown to be relevant to the development of the ENS, of which the RET pathway and the endothelin pathway are the most important ones. However, the identified genes and variants in these genes explain no more than 25% of the overall genetic risk[2,4]. Thus, the vast majority of cases cannot yet be explained by the identified HSCR-associated variants. These findings indicate that the majority of the disease risk must be due to as yet unidentified rare or common variants in the known HSCR genes or, more likely, variants in yet unknown genes, acting alone or in combination.

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

METHODS

Study samples

Trios

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

Case-control

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

Data generation

Whole exome sequencing

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

Genome-wide SNP array

To determine copy number variants (CNVs) and regions of homozygosity, DNA was hybridized to the HumanCyto SNP12 BeadChip (Illumina, San Diego, CA, USA) according to standard protocols.

Variant calling and prioritization

Aligned reads from all sequenced samples were pre-processed according to standard guidelines[6]. Variant calling was done independently for Illumina reads

or Solid reads using the Genome Analysis Toolkit (GATK) unified Genotyper 2.0[7]. To avoid mismatched regions across different capture kits, calling was performed on whole genome wide without limiting on any capture array. Special setting (allow potentially miscoded quality scores) was used to make color-spaced solid reads compatible to the program (Broad institute). Raw variants (including single nucleotide variants and short insertions/deletions) with individual genotypes and their affiliated quality scores were stored in a standard VCF format after calling. Quality assessment (QA) and QC were then adopted on a few set of variants (raw variants, exonic variants, rare variants) to generate a confident variant set for downstream prioritization (Supplementary Note).

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

Identification of DNM

WES DNM detection

Rare, exonic variants present in the probands but absent in both parents were considered DNM. To select putative DNM (or *de novo* variations) the following criteria were used: 1) minimal coverage of 5 in patients and parents; 2) a minimal genotype quality score of 10 for both patients and parents; 3) at least 10% of the reads showed the alternative allele in patients; and 4) not more than 10% of the reads showed the alternative allele in parents. Subsequently all remaining DNM variants were manually inspected using the Integrated Genome Viewer (IGV) and classified into 5 different confidence ranks according to their base-calling quality and strand bias. The first two ranks of DNM candidates were selected for validation by Sanger sequencing; while the other three classes of candidates were re-evaluated by a model trained from variants submitted for Sanger sequencing (Supplementary Note).

RET gene inspection

To guarantee that no *de novo* mutations had been missed in the major HSCR gene, the depth of coverage of each of the 21 exons of *RET* was manually inspected for each patient. All exons with a coverage <10 were Sanger sequenced. Mutation Detector software (Thermo Fisher Scientific) was used to identify rare coding sequencing mutations from raw Sanger sequences; any mutation found in trio

proband was further checked in his/her parents. Besides rare mutations, bi-allelic genotypes for the common risk single nucleotide polymorphisms (IVS1+9494, rs2435357T) were extracted from local databases or newly genotyped.

Copy number variation detection

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

Statistical tests

De novo mutation rate

All proven DNMs were classified into loss-of-function (nonsense Single Nucleotide Variants (SNVs), frame-shift indels and splicing sites), missense SNVs, in-frame indels and synonymous SNVs. The counts of DNM per trio were fitted to Poisson distribution with lamda as observed mean. *De novo* mutation rates were calculated for these DNM subtypes and compared to 677 published healthy trios and neurodevelopmental disease trios using a binomial test[8–13]. Given per-gene mutation rate in Samocha *et al*. paper[14], statistical over-representation of mutations in all 24 genes were calculated using Fisher's exact test.

Gene-wide burden analysis

Genes with DNM were further scrutinized for the presence of inherited rare damaging variants in the trios as well as in HSCR singletons for whom WES data were available. A detailed analytical protocol was shared before running association in each centre. Briefly, genotypes of rare damaging variants (as previously defined) in genes carrying ≥1 *de novo* mutation were extracted from raw sequencing reads. CMC test in Rvtest package was used to collapse multiple variants into the same gene (boundary defined using hg19 refgene) and compare overall burden between cases and local matched controls[15]. *P-values* were estimated by asymptotic chi-square distribution. Gene-wise *p*-value, burden direction and variant count per gene were exported. Ultimately sample-size

weighted Z-score method was used to conduct meta-analysis on gene-wise summary statistics from three centres using the same protocol.

Bioinformatics analysis

Variant-level implication

The impact of each DNM to its carrying gene was predicted using several of bioinformatics tools or databases. The conservation of missense SNVs was predicted using GERP and PhyloP across 29 different species. The deleteriousness of missense or nonsense SNVs were determined by a logit model incorporating 5 prediction programs (Polyphen2, Sift, MutationTaster, PhyloP and Likelihood ratio)[16]. Human Splicing finder was used to predict whether DNMs causing synonymous change or locating at splicing sites (exon +/- 2bp) created or disrupted splice sites[17]. To further implicate the possible role of synonymous DNMs on transcription, RNAmute was used to predicted the RNA substructure change due to corresponding site mutation[18]. Finally, ClinVar and PubMed were searched for the same or similar mutations in the same gene that present in healthy controls or other disease patients.

Gene-level implication

The evidence of gene-level implication was collected from two aspects. On one side, those 24 genes carrying DNMs were searched against databases (ATGU's Server) for other disease patients or healthy samples[14]. On the other side, ENS candidate genes/gene-sets (Supplementary Table 8; Supplementary Note) were linked to newly identified genes using pathway or PPI network information. Disease Association Protein-Protein Link Evaluator (DAPPLE) was used to test whether the genes carrying DNM in our study are functionally connected to each other. The significance of observed pathway enrichment and network connectivity was evaluated empirically using randomly selected genes, genes having the same genomic size as the identified DNM genes. InWeb and Ingenuity Pathway Analysis were used to detect direct and indirect protein interactions between ENS-related genes and genes with DNMs.

Gene expression in ENS

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

Zebrafish

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

Gene knockdown by antisense morpholino

Antisense morpholinos (MO) (Gene Tools LLC) targeting the zebrafish orthologues of the candidate genes, by blocking either translation or splicing, were microinjected to 1 to 4-cell stage *Tg(-8.3bphox2b:Kaede)* transgenic zebrafish embryos as previously described[19]. For candidate genes that are duplicated in the zebrafish genome, morpholinos targeting all paralogs were co-injected. Standard control morpholino and 5-nucleotide mismatch control morpholino for *ckap2l, dennd3a, dennd3b, ncl1, nup98* and *tbata* were used as negative control. Embryos were raised to 5 dpf, analysed and imaged under a stereo fluorescence microscope (Leica MZ16FA and DFC300FX). An HSCR-like phenotype was defined as the absence of enteric neurons in the distal intestine in 5 dpf embryos. Sequences and dosages of all morpholinos used are listed in Supplementary table 9.

Expression analysis

To confirm the target gene were successfully knockdown, total RNA were extracted from 1 dpf embryos (n=50) injected with the splice blocking morpholino using RNA Bee (Amsbio) and cDNA were reverse transcribed using iScript cDNA Synthesis Kit (Bio-rad). qPCR were performed using KAPA Sybr® Fast qPCR Kit (KAPA Biosystems; see Supplemantary Table 10 for primer detail) and the expression of the target gene was normalized by the mean expression of two housekeeping genes (*elfa* and *actb*). Relative expression of the target gene in the splice blocking morpholino-injected embryos to the control morpholino-injected embryos was determined by Livak method[20].

To determine the temporal expression of the zebrafish orthologues, RT-PCR was performed at various time points with primers used to amplify up a segment of the open reading frame of each gene. To determine the spatial expression patterns of *dennd3a*, *dennd3b*, *ncl1*, *nup98* and *tbata*, antisense Digoxigenin-labeled probes for both genes were generated and whole-mount *in situ* hybridization was performed as described by Thisse *et al.*[21].

RESULTS

Identification of *de novo* mutations

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

All *de novo* variations were carefully selected, validated and/or statistically predicted (Methods and Supplementary Note; see prediction result in Supplementary

Table 1. *De novo* mutations in Hirschsprung disease probands

Trio	Pheno-type	Gene	*De novo* mutation	Type	Prediction delete-riousness*	MAF (dbSNP137/ ESP6500)%
1	L, F	*RET*	3splicing9+1	splicing	-	N / N
		RBM25	c.474C>T: p.L158L	synonymous	-	N / N
2	L, F	*RET*	c.2511_2519delCCCTGGACC:p.S837fs	frameshift	-	N / N
		COL6A3	c.3327C>T: p.H1109H	synonymous	-	0.00042 (rs114845780) / N
3	L, F	*RET*	c.1818_1819insGGCAC: p.Y606fs	frameshift	-	N / N
4	L, F	*DAB2IP*	c.2339C>T:p.T780M#	missense	No	N / N
		ISG20L2	c.961G>A:p.G321R	missense	Yes	N / N
		MED26	c.675C>T:p.A225A	synonymous	-	N / N
		NCLN	c.496C>T:p.Q166X#	nonsense	-	N / N
		NUP98	c.5207A>G:p.N1736S	missense	Yes	N / N
		VEZF1	c.584C>T:p.S195F	missense	Yes	N / N
		ZNF57	c.570C>T:p.D190D	synonymous	-	N / N
5	L, F	*RET*	c.1761delG :p.G588fs	frameshift	-	N / N
		SCUBE3	c.1493A>T:p.N498I	missense	No	N / N
6	L, M	*AFF3*	c.1975G>C:p.V659L	missense	No	N / N
		PLEKHG5	c.2628G>T:p.T876T	synonymous	-	N / N
7	L, M	*KDM4A*	c.26A>G:p.N9S	missense	No	N / N

8	L, M	*MAP4*	c.3351C>T:p.G1117G	synonymous	-	N / N
9	L, F	*RET*	c.1858T>C:p.C620R	missense	Yes	0 (rs77316810) / N
10	TCA, M	*CKAP2L*	c.555_556delAA: p.E186fs	frameshift	-	N / 0.00002
11	L, F	*RET*	c.409T>G:p.C137G	missense	Yes	N / N
		HMCN1	c.10366G>A:p.A3456T	missense	No	N / N
		TUBG1	c.699T>C:p.S233S	synonymous	-	N / N
12	L, F	*CCR2*	c.848T>A:p.L283Q	missense	Yes	N / N
		DENND3	c.1921delT:p.K640fs	frameshift	-	N / N
13	L, F	*RET*	c.1710C>A:p.C570X	nonsense	-	N / N
14	L, F	*RET*	c.526_528delGCA: p.R175del	non-frameshift	-	N / N
		TBATA	c.157C>T:p.R53C	missense	No	N / N

F: Female; M: Male; L: Long-segment HSCR; TCA: Total Colonic Aganglionosis; *: Disease-causal prediction by KGGSeq[57], a software that uses a weighted logistic regression to combine multiple prediction scores; #mosaic mutation; Dark grey: *de novo RET* mutations; Light grey: genes giving a HSCR-like phenotype in zebrafish; %: minor allele frequency in dbSNP137 or ESP database, with 'N' standing for no data available.

Table 3). After Sanger sequencing validation, a total of 28 DNMs in 14 patients were identified (Table 1). The overall DNM rate per individual was 1.2 per exome per generation (Poisson distribution with λ=1.2; Kolmogorov-Sminov test, p=0.893; Supplementary Figure 3) which is in accordance with the expected mutation rate in the general population. Several studies have shown that the DNM rates are similar between patients and healthy controls, but found that patients have a significantly higher fraction of loss of function (LOF) DNMs[8,9]. Indeed, in our HSCR patient cohort, the rate of loss of function DNMs (LOF; N=8, including nonsense, frameshift and splice site changes) is significantly higher than that of healthy trios (p=0.011) or unaffected siblings of neuropsychiatric patients (p=0.001) from multiple published studies[8,10–12,22] (Supplementary Table 4). The 28 DNMs were localised in 21 genes. 8 DNMs were found in *RET*, the major HSCR gene[23]. Among the DNMs in *RET* was the Cys620Arg variant, known to cause both HSCR and Multiple Endocrine Neoplasia type 2A[24]. In this study, the observed rate for *RET* DNMs (0.33 per trio) was significantly higher (binomial test, $p<2*10^{-16}$) than that modelled for *RET* DNMs in the general population (0.000133 per trio) according to Samocha *et al.*[14].

One of the patients analysed carried a total of 7 DNMs, two of which (in *NCLN* and *DAB2IP*) were mosaic mutations (Supplementary Figure 4). This finding is in line with a recent report stating that 6.5% of all DNMs are in fact mosaic and occur post-zygotic[25]. Within the 24 patients we looked for inherited rare damaging variants in the 21 genes that carried DNMs (Supplementary Note, Methods). Inherited damaging mutations were found in *RET*, *HMCN1, PLEKHG5, MAP4,*

SCUBE3, and *KDM4A* (Supplementary Table 5). Neither *de novo* nor inherited copy number variants (CNVs) were detected in any of the trios.

Mutation profile of HSCR patients

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

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

Determining pathogenicity of the DNMs *in silico*

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

To establish whether genes with DNMs carry significantly more rare variants in HSCR patients than in controls, we used the WES data from the 20 eligible HSCR trio-probands, 28 additional HSCR patients and 212 control individuals to calculate the variation burden per gene (Methods). Nine of the twenty-one genes

(*RET, KDM4A, HMCN1, MAP4, NUP98, AFF3, COL6A3, CCR2, and CKAP2L)* were found recurrently mutated in multiple HSCR patients with different rare damaging mutation sites (Supplementary Table 7). Meta-analysis of our gene burden tests showed that *RET* and *CKAP2L* were enriched for rare damaging variants in the HSCR patients (nominal $p<0.05$; Table 2 and Supplementary Table 7). However, cross-checking of these 21 genes in another in-parallel HSCR exome study (190 cases and 740 controls) revealed only *RET* was significantly overrepresented with deleterious variants ($p < 0.001$; manuscript in preparation**,** A. Chakravarti).

The possible impact of DNMs on gene function was explored using bioinformatic prediction tools (Methods). Besides the 8 LOF mutations, 6 out of twelve missense mutations were consistently predicted deleterious (Table 1). As for the seven synonymous DNMs, we found no *in silico* evidence indicating that those changes interfered with splicing and/or significantly changed the RNA structure (Supplementary Table 8).

Table 2. Genes carrying *de novo* mutations

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

*genes evolutionary constrained as per Samocha *et al.* 2014; NT: not tested (gene carries synonymous mutation and/or has no ortholog in zebrafish); [&]data from in-house hIPSC-derived neural crest, mouse expression data, and RT-PCR in zebrafish (test only for 4 novel genes).

We next checked whether the genes with DNMs are functionally related to each other and/or to the signalling networks known to govern ENS development. *ISG20L2* and *MAP4* showed more indirect interactions with other genes carrying DNMs than expected by chance (p=0.0063 and p=0.0167 respectively) as predicted by DAPPLE, though no direct *in silico* interactions were found among those 21 genes. A list of 116 known ENS related genes (Supplementary Table 6) was used to study the functional link between genes with DNMs (other than *RET*) and the ENS. Only a single interaction was identified in the InWeb protein interaction catalogue (*COL6A3* interacts with *ITGB1*). Using Ingenuity Pathway Analysis, we identified additional direct and indirect relationships with ENS-related genes for *MAP4*, *COL6A3*, *RBM25* and *TUBG1* (Supplementary Figure 5). All genes carrying DNMs were either expressed in human iPSC-derived enteric neuron precursors or in primary murine enteric neuron precursors (Table 2).

Determining pathogenicity of the DNMs *in vivo*

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

of the orthologues, we repeated the experiment by injecting translation-blocking morpholinos (TBMOs) against *dennd3a, dennd3b, ncl1, nup98* and *tbata* and the phenotype was reproduced (data not shown). Therefore we concluded that knockdown of the *DENND3*, *NCLN*, *NUP98* and *TBATA* orthologues disrupted ENS development and caused a HSCR-like phenotype *in vivo*.

To confirm the SBMOs knockdown effect, qPCR was performed to compare the expressions of the target genes between SBMO-injected and control morpholino-injected embryos. Expression of *dennd3a*, *dennd3b*, *nup98* and *tbata* was markedly reduced in the SBMO-injected embryos (Supplementary Figure 6). Intriguingly, there was no significant reduction in *ncl1* expression in the *ncl1* SBMO injected embryos.

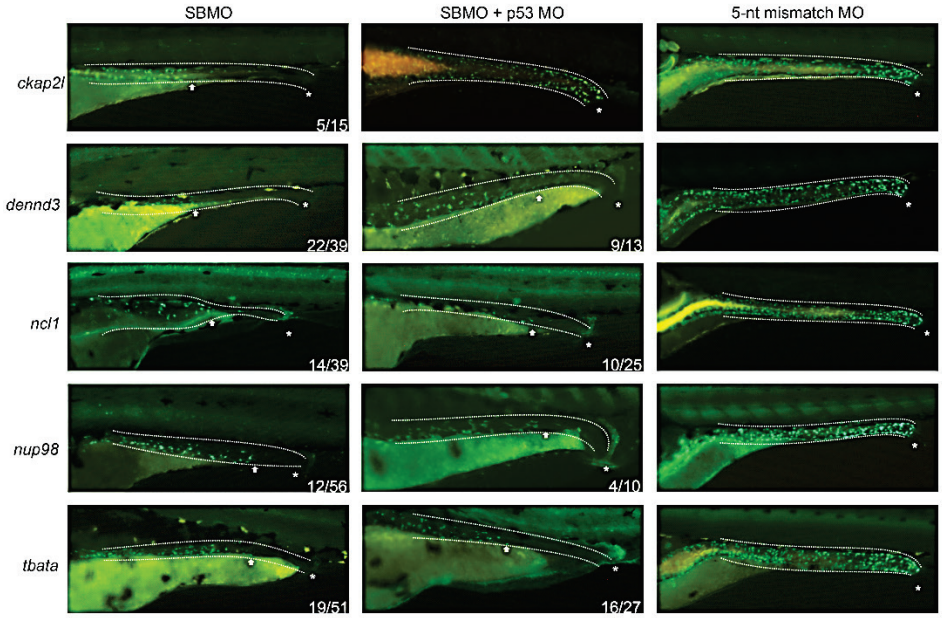


Figure 1. Pathogenicity analysis *in vivo* by morpholino gene knockdown in zebrafish. Knockdown of *ncl1, dennd3, nup98* and *tbata* resulted in HSCR-like phenotype that kaede-expressing enteric neurons were absent in the distal intestine at 5 dpf and the results were reproduced in the presence of p53 morpholino. Aganglionosis observed in *ckap2l* knockdown was caused by non-specific apoptosis as the result was not reproducible in p53 morpholino co-injection. Number of embryos with phenotype out of total number of embryos observed is shown. Dotted lines outline the intestines. Asterisks indicate the positions of anus. Arrows indicate the position where the aganglionic region begins.

Therefore we further investigated it by performing RT-PCR on individual embryos and found that there was a large variation in *ncl1* expression between

embryos injected with the SBMO, with some of them showing a clear reduction in *ncl1* transcript level (Supplementary Figure 7). Of the zebrafish orthologues that did not show a specific HSCR-like phenotype after SBMOs injection, all demonstrated significant reductions in expressions except for *aff3*, *scube3* and *vezf1a* (Supplementary Figure 6).

In addition we performed RT-PCR and whole mount in situ hybridization (WISH) experiments to determine if the gene expression patterns of the zebrafish orthologues were consistent with a predicted role in ENS development. Temporal analysis using RT-PCR revealed that zebrafish orthologues of *DENND3, NCLN* and *NUP98* were maternally and zygotically expressed from 0-120hpf while the *TBATA* orthologue is only zygotically expressed from 24-120hpf (Supplementary Fig 8). WISH analysis showed that the orthologues for all 4 genes were expressed in distinct spatial locations specifically in the intestine and the anterior CNS from 24-96hpf (Figure 2).

DISCUSSION

Over the last years a large number of papers have been published on *de novo* mutation screening in human diseases. This has resulted in the identification of many new disease associated genes. Genes are considered as true disease causing when at least 2 unlinked patients are found with a mutation in the same gene. This works well for diseases that are relatively homogeneous or for which many patients can be investigated. For the more heterogeneous rare diseases for which only small cohorts are available this poses a problem. Often possible disease causing genes are found in a single patient. How to decide whether this finding is of importance? Expression of the gene in the relevant tissues can be considered as additional evidence, as is networks analysis. However, making strong statements for private disease genes is, and will be, extremely difficult. It also results in a bias towards genes in the known disease causing gene networks. Genes not fitting the current knowledge are often discarded as uninteresting. In the current study we wanted to take this all one step further.

Therefore, we decided that the best way to obtain sound evidence for involvement of new candidate genes in HSCR should come from functional analysis. We opted for an *in vivo* approach using the zebrafish model system. We knocked down the expression of zebrafish orthologues of 12 of the 13 genes in which loss of function or missense DNMs were identified in a transgenic reporter zebrafish line (*Tg(-8.3bphox2b:Kaede)*). The orthologues of 9 of the 12 genes were successfully knockdown by morpholinos, and from which we discovered that 4

genes when functionally perturbed resulted in loss of neurons in the distal gut, as in the HSCR patients. It is noteworthy that the SBMOs targeting 3 of the orthologues (*aff3*, *scube3* and *vezf1a*) did not knockdown the target transcripts as expected, which highlighted the limitation of morpholinos and might lead to false-negative results[36]. To bypass this limitation, other loss-of-function approaches should be considered to further study these genes, such as CRISPR/Cas9 knockout[37].Finding 4 genes that when knocked-down in zebrafish give a hindgut phenotype resembling the human patients in which the DNMs were found, clearly demonstrates that genes that never would have been followed up, based on the usual gene selection criteria, should not be ignored.

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

We wondered whether any or all of these 4 genes can be linked to the ENS or whether they play relevant roles in neuronal development or neural crest derived cell types in general. In fact by studying these genes in more depth

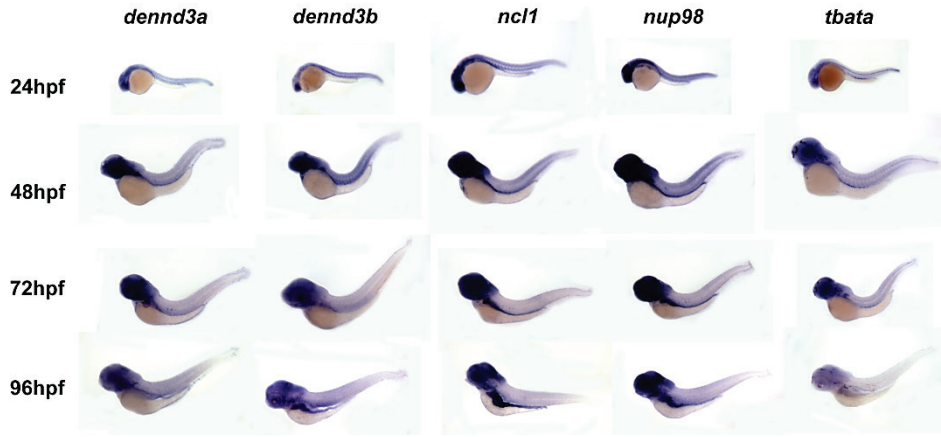


Figure 2. Temporal and spatial expression patterns of zebrafish orthologues. Whole mount *in situ* hybridized embryos hybridized with antisense riboprobes for *dennd3a, dennd3b, ncl1, nup98 and tbata* at the indicated developmental stages. All columns show lateral views. Anterior CNS expression is apparent at all stages for all probes while intestinal expression for all probes is apparent from 48hpf onwards.

we noticed that all 4, despite lack of obvious connection to the known ENS pathways, are involved in the development of the CNS or the neural crest, making these not as random as they might first appear.

DENN/MADD Domain Containing 3 (*DENND3*) is a guanine nucleotide exchange factor (GEF) that is involved in intracellular trafficking by activation of

the small GTPase RAB12[38]. In zebrafish, Rab12 and other Rab GTPases are highly expressed by pre-migratory neural crest cells and their expression is dysregulated in Ovo1 morphant zebrafish that display altered migration of neural crest cells[39]. Independently of *RAB12, DENND3* also regulates Akt activity, which is involved in the proliferation and survival of enteric neural crest cells[38,40].

Nicalin (*NCLN*) is a key component of a protein complex that antagonizes Nodal signalling[41]. In vertebrates, Nodal signalling is involved in induction of the mesoderm and endoderm[42]. In contrast, inhibition of Nodal signalling is required for the specification of human embryonic stem cells into neuroectoderm, including the neural crest[43,44]. The antagonizing function of Nicalin on Nodal signalling is therefore consistent with the neural crest specification that is required for ENS development.

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

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

The 4 newly identified candidate genes for HSCR all seem to play a role in neuronal development and could potentially be involved in HSCR (Figure 3). This also suggests a clear link between CNS and ENS development. This is not surprising as a number of studies have described the strong correlation between Down syndrome and syndromic HSCR and several known HSCR genes (e.g. *KBP*, *SOX10, NRG1, IKBKAP, ZEB2, PHOX2B*) have been reported to be involved in both

CNS and ENS pathologies[2,55–57]. In humans, *SOX10* mutations cause myelin deficiencies and sensory neuropathies as well as the neurological variant of Waardenburg-Shah syndrome which includes HSCR in the phenotypic spectrum.

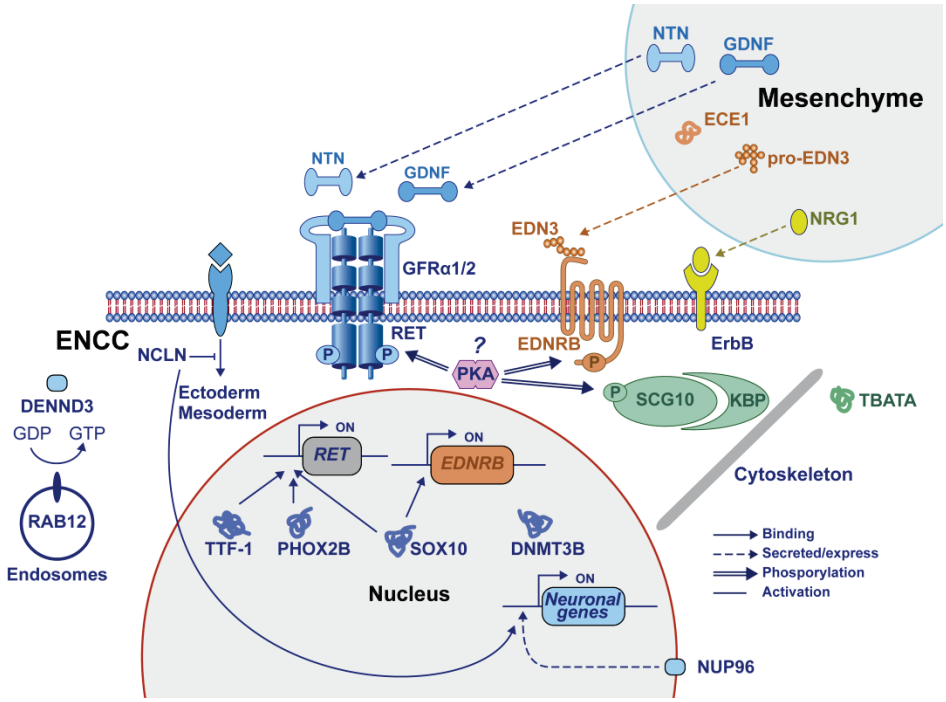


Figure 3. **Newly identified genes in ENS development.** All symbols represent proteins coded by Hirschsprung known genes or novel genes identified in this study. The effect of gene *NUP98* is shown by protein NUP96. The interaction effects between different proteins are illustrated by four different lines representing binding, secreted/express, phosphorylation and activation. ENCC, enteric neural crest cell.

Likewise, *NRG1* is associated with schizophrenia and *Nrg1* mutations in mice cause peripheral sensory neuropathies[46]. *IKBKAP* mutations are associated with the

Riley-Day syndrome or familial dysautonomia (FD)[58,59]. Notably, some patients with FD also suffer from gastrointestinal dysfunction shortly after birth and interestingly, the co-occurrence of both FD and HSCR has been reported[60]. In addition, knockdown of *ikbkap* in zebrafish also generates a HSCR-like phenotype[35]. Further, *KBP* mutations are associated with Goldberg-Shprintzen syndrome[61] (MIM 609460), a rare autosomal recessive inherited syndrome,

where patients present with HSCR, microcephaly polymicrogyria and moderate mental retardation.

Besides the fact that several HSCR/neuromuscular genes are known to be associated with CNS defects, the opposite is also described. Many neurological and psychiatric disorders are associated with constipation, and sometimes defects in the ENS are reported[62]. For instance, it has recently been described that mutations in *CDH8* result in a specific subtype of autism in combination with gastrointestinal problems. A *cdh8*$^{-/-}$ zebrafish recapitulates the human phenotype, including increased head size (expansion of the forebrain/midbrain), an impairment of gastrointestinal motility and a reduction in post-mitotic enteric neurons[63]. Besides, a search of CNS and autism in Phenolyzer[64] returned two genes (*APP* and *MECP2*) that have been implicated in ENS development[65,66].

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

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

URLS

Genome analysis toolkit (GATK) (https://www.broadinstitute.org/gatk/);
ANNOVAR (http://annovar.openbioinformatics.org/en/latest/);
PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/);
KGGSeq (http://statgenpro.psychiatry.hku.hk/limx/kggseq/);
ATGU's Server (http://atgu.mgh.harvard.edu/webtools/gene-lookup/);
DAPPLE (http://www.broadinstitute.org/mpg/dapple/dappleTMP.php);
ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/)

ACKNOWLEDGMENTS

The authors would like to thank the patients and families involved in this study. DS, AB, RC, BE, YS, RH are supported by research grants from ZonMW (TOP-subsidie 40-00812-98-10042 to BJLE/RMWH) and the Maag Lever Darm stichting (WO09-62 to RMWH); AC is supported by NIH grant R37 HD28088; HG, WC, VL, EN, PS, MS, CT, PT, MMG-B are supported Health and Medical Research Fund (HMRF 01121326 to VCHL; HMRF 02131866 to MMGB; and 01121476 to ESWN); General Research Fund (HKU 777612M to PKT); HKU seed funding for basic research (201110159001 to PKT) and small project funding (201309176158 to CSMT). The work described was also partially supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, Project No. [T12C-714/14R] to PKT; CH, PG are supported by grants from the Italian Ministry of Health through "Cinque per mille" and Ricerca Corrente to the Gaslini Institute; GA, MB, BL-T, MR-F, SB are supported by the Spanish Ministry of Economy and Competitiveness (Institute of Health Carlos III (ISCIII), PI13/01560 and CDTI, FEDER-Innterconecta EXP00052887/ITC-20111037), and the Regional Ministry of Innovation, Science and Enterprise of the Autonomous Government of Andalusia (CTS-7447); NINDS (5R21NS082546) awarded to ITS.

AUTHOR CONTRIBUTIONS

H.G. and D.S. performed the exome sequencing analyses and wrote the manuscript. W.C. together with A.J.B., R.C., V.L., B. J. and I.T.S. performed the zebrafish experiments and prepared the figures for the manuscript. Y.S. and C.S.T. conducted the CNV analyses.

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Amiel, J. *et al.* Hirschsprung disease, associated syndromes and genetics: a review. *J. Med. Genet.* **45,** 1–14 (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. U. S. A.* **106,** 2694–9 (2009).
3. Alves, M. M. *et al.* Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev. Biol.* **382,** 320–9 (2013).
4. 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,** 581–96 (2015).
5. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
6. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43,** 491–498 (2011).
7. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).
8. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485,** 237–241 (2012).
9. Iossifov, I. *et al.* De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* **74,** 285–299 (2012).
10. Xu, B. *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat. Genet.* **43,** 864–8 (2011).
11. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485,** 246–250 (2012).
12. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet* **380,** 1674–1682 (2012).
13. Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154,** 518–529 (2013).
14. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46,** 944–950 (2014).
15. Li, B. & Leal, S. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83,** 311–321 (2008).
16. Li, M. X. *et al.* Predicting Mendelian Disease-Causing Non-Synonymous Single Nucleotide Variants in Exome Sequencing Studies. *PLoS Genet.* **9,** (2013).
17. Desmet, F. O. *et al.* Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37,** (2009).
18. Churkin, A. & Barash, D. RNAmute: RNA secondary structure mutation analysis tool. *BMC Bioinformatics* **7,** 221 (2006).
19. Harrison, C., Wabbersen, T. & Shepherd, I. T. In vivo visualization of the development of the enteric nervous system using a Tg(-8.3bphox2b:Kaede) transgenic zebrafish. *Genesis* **52,** 985–90 (2014).
20. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25,** 402–8 (2001).
21. Thisse, C., Thisse, B., Schilling, T. F. & Postlethwait, J. H. Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development* **119,** 1203–15 (1993).
22. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *November* **13,** 216–221 (2014).
23. Heanue, T. a & Pachnis, V. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat. Rev. Neurosci.* **8,** 466–79 (2007).
24. Romeo, G. *et al.* Association of multiple endocrine neoplasia type 2 and Hirschsprung disease. in *Journal of Internal Medicine* **243,** 515–520 (1998).
25. Acuna-Hidalgo, R. *et al.* Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *Am. J. Hum. Genet.* **97,** 67–74 (2015).
26. Emison, E. S. *et al.* A common sex-dependent mutation in a RET enhancer underlies

Hirschsprung disease risk. *Nature* **434,** 857–63 (2005).
27. 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).
28. Burzynski, G. M. *et al.* Identifying candidate Hirschsprung disease-associated RET variants. *Am. J. Hum. Genet.* **76,** 850–858 (2005).
29. Sribudiani, Y. *et al.* Variants in RET associated with hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology* **140,** 572–582 (2011).
30. Do, R., Kathiresan, S. & Abecasis, G. R. Exome sequencing and complex disease: Practical aspects of rare variant association studies. *Hum. Mol. Genet.* **21,** (2012).
31. Shepherd, I. T., Pietsch, J., Elworthy, S., Kelsh, R. N. & Raible, D. W. Roles for GFR alpha 1 receptors in zebrafish enteric nervous system development. *Development* **131,** 241–249 (2004).
32. Shepherd, I. T., Beattie, C. E. & Raible, D. W. Functional analysis of zebrafish GDNF. *Dev. Biol.* **231,** 420–435 (2001).
33. Elworthy, S., Pinto, J. P., Pettifer, A., Cancela, M. L. & Kelsh, R. N. Phox2b function in the enteric nervous system is conserved in zebrafish and is sox10-dependent. *Mech. Dev.* **122,** 659–669 (2005).
34. Dutton, K., Dutton, J. R., Pauliny, A. & Kelsh, R. N. A morpholino phenocopy of the colourless mutant. *Genesis* **30,** 188–9 (2001).
35. Cheng, W. W.-C. *et al.* Depletion of the IKBKAP ortholog in zebrafish leads to hirschsprung disease-like phenotype. *World J. Gastroenterol.* **21,** 2040–6 (2015).
36. Bedell, V. M., Westcot, S. E. & Ekker, S. C. Lessons from morpholino-based screening in zebrafish. *Brief. Funct. Genomics* **10,** 181–188 (2011).
37. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–55 (2014).
38. Matsui, T., Noguchi, K. & Fukuda, M. Dennd3 Functions as a Guanine Nucleotide Exchange Factor for Small GTPase Rab12 in Mouse Embryonic Fibroblasts. *J. Biol. Chem.* **289,** 13986–95 (2014).
39. Piloto, S. & Schilling, T. F. Ovo1 links Wnt signaling with N-cadherin localization during neural crest migration. *Development* **137,** 1981–1990 (2010).
40. Srinivasan, S., Anitha, M., Mwangi, S. & Heuckeroth, R. O. Enteric neuroblasts require the phosphatidylinositol 3-kinase/Akt/Forkhead pathway for GDNF-stimulated survival. *Mol. Cell. Neurosci.* **29,** 107–19 (2005).
41. Haffner, C. *et al.* Nicalin and its binding partner Nomo are novel Nodal signaling antagonists. *EMBO J.* **23,** 3041–3050 (2004).
42. Schier, A. F. Nodal signaling in vertebrate development. *Annu. Rev. Cell Dev. Biol.* **19,** 589–621 (2003).
43. Smith, J. R. *et al.* Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev. Biol.* **313,** 107–117 (2008).
44. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27,** 275–280 (2009).
45. Fontoura, B. M. A., Blobel, G. & Matunis, M. J. A conserved biogenesis pathway for nucleoporins: Proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. *J. Cell Biol.* **144,** 1097–1112 (1999).
46. Rosenblum, J. S. & Blobel, G. Autoproteolysis in nucleoporin biogenesis. *Proc Natl Acad Sci U S A* **96,** 11370–11375 (1999).
47. Tran, E. J. & Wente, S. R. Dynamic Nuclear Pore Complexes: Life on the Edge. *Cell* **125,** 1041–1053 (2006).
48. Chakraborty, P. *et al.* Nucleoporin Levels Regulate Cell Cycle Progression and Phase-Specific Gene Expression. *Dev. Cell* **15,** 657–667 (2008).
49. Lupu, F., Alves, A., Anderson, K., Doye, V. & Lacy, E. Nuclear Pore Composition Regulates Neural Stem/Progenitor Cell Differentiation in the Mouse Embryo. *Dev. Cell* **14,** 831–842 (2008).
50. D'Angelo, M. A., Gomez-Cavazos, J. S., Mei, A., Lackner, D. H. & Hetzer, M. W. A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation. *Dev. Cell* **22,** 446–458 (2012).
51. Liang, Y., Franks, T. M., Marchetto, M. C., Gage, F. H. & Hetzer, M. W. Dynamic Association of

NUP98 with the Human Genome. *PLoS Genet.* **9,** (2013).
52. Irla, M. *et al.* Genomic organization and the tissue distribution of alternatively spliced isoforms of the mouse Spatial gene. *BMC Genomics* **5,** 41 (2004).
53. Irla, M. *et al.* Neuronal distribution of Spatial in the developing cerebellum and hippocampus and its somatodendritic association with the kinesin motor KIF17. *Exp. Cell Res.* **313,** 4107–4119 (2007).
54. Yammine, M., Saade, M., Chauvet, S. & Nguyen, C. Spatial gene's (Tbata) implication in neurite outgrowth and dendrite patterning in hippocampal neurons. *Mol. Cell. Neurosci.* **59,** 1–9 (2014).
55. Tang, C. S. *et al.* Fine mapping of the 9q31 Hirschsprung's disease locus. *Hum. Genet.* **127,** 675–683 (2010).
56. Pingault, V. *et al.* Peripheral neuropathy with chronic intestinal pseudo-obstruction and deafness: A developmental 'neural crest syndrome' related to a SOX10 mutation. *Ann. Neurol.* **48,** 671–676 (2000).
57. Harrison, P. J. & Law, A. J. Neuregulin 1 and Schizophrenia: Genetics, Gene Expression, and Neurobiology. *Biological Psychiatry* **60,** 132–140 (2006).
58. Anderson, S. L. *et al.* Familial dysautonomia is caused by mutations of the IKAP gene. *Am. J. Hum. Genet.* **68,** 753–8 (2001).
59. Slaugenhaupt, S. A. *et al.* Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am. J. Hum. Genet.* **68,** 598–605 (2001).
60. Azizi, E., Berlowitz, I., Vinograd, I., Reif, R. & Mundel, G. Congenital megacolon associated with familial dysautonomia. *Eur. J. Pediatr.* **142,** 68–9 (1984).
61. Brooks, a S. *et al.* A consanguineous family with Hirschsprung disease, microcephaly, and mental retardation (Goldberg-Shprintzen syndrome). *J. Med. Genet.* **36,** 485–9 (1999).
62. Winge, K., Rasmussen, D. & Werdelin, L. Constipation in neurological diseases. *J. Neurol. Neurosurg. Psychiatry* **74,** 13–19 (2003).
63. Bernier, R. *et al.* Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158,** 263–276 (2014).
64. Yang, H., Robinson, P. N. & Wang, K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat. Methods* 1–6 (2015). doi:10.1038/nmeth.3484
65. Van Ginneken, C., Schäfer, K. H., Van Dam, D., Huygelen, V. & De Deyn, P. P. Morphological changes in the enteric nervous system of aging and APP23 transgenic mice. *Brain Res.* **1378,** 43–53 (2011).
66. Wahba, G. *et al.* MeCP2 in the enteric nervous system. *Neurogastroenterol. Motil.* **27,** 1156–61 (2015).

SUPPLEMENTARY NOTES

Quality assessment and control for exome variants

Concrete criterions in quality assessment (QA) include: total number of variants; dbSNP137 coverage; Transition/Transversion (Ti/Tv) ratio; genotype concordance rate and cross-sample identical-by-decent (IBD) relatedness[1]. Two complementary steps were applied in quality control (QC), including variant-level filtering (hard filtration or variant quality recalibration (VQSR)) and genotype-level filtering. In detail, we annotated GATK-called variants as low quality SNPs ("QD <2.0" or "MQ <40.0" or "FS >60.0" or "HaplotypeScore >13.0" or "MQRankSum <-12.5" or "ReadPosRankSum <-8.0" in their 'info' field) and low quality Indels ("QD <2.0" or "ReadPosRankSum <-20.0" or "InbreedingCoeff <-0.8" or "FS >200.0 in 'info' field); in addition, VQSR differentiated a few relatively low quality SNVs (labeled as "TruthSensitivityTranche99.90to100.00" after Gaussian mixture modeling at true sensitivity 99%) from other passed SNVs. On the other hand, individual genotypes were evaluated by quality parameters in the field of genotyping, mainly reflecting the likelihood of three possible genotypes (reference homozygous, heterozygous and alternative homozygous). A heterozygous genotype was kept only if it was supported by >4 total reads, and the ratio for alternative allele is above 0.25. Comparatively, a reference or alternative homozygous genotype was accepted if it was supported by > 4 total reads, and ratio for reference or alternative allele is above 0.95.

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

Mutation validation and prediction

Each DNM candidate was manually inspected using the Integrative Genomic Viewer (IGV) and they were categorized into five different groups: probably true positive, possibly true positive, unclear, possibly false positive and probably false positive. Two lists of putative DNM candidates were generated for confirmation by

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

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

Generation of ENS candidate genes

Candidate genes were selected by a literature review on Hirschsprung disease research, which included both genetic and functional studies. Most of them were also covered in Jiang et al.[2] and Gui et al.[3], which previously summarized possible genes related to HSCR or involved in ENS development. The genes were categorized into 4 major types, genes selected based on: genetic linkage, genetic association, microarray expression, and animal models. In total 116 genes were selected that fit more than 1 category (Supplementary Table 6). A few of these

genes fall into the same pathways previously implicated in neural crest cell migration, proliferation and differentiation. Three pathways (*RET* signaling pathway, *EDNRB* signaling pathway and *KBP* signaling pathway) were key partners involved in ENS development[4].

SUPPLEMENTARY TABLES

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

HSCR patients (N=52)	**Trios** (N=24)		**Singletons** (N=28)		**Controls** (N=212)
	Short	Long/TCA	Short	Long/TCA	
	(N=1)	(N=23)	(N=15)	(N=13)	
Males	0	7 (4)	13	4	117
Females	1 (0)	16 (10)	2	9	95

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

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

Centre	# of Trios	Capture array	Target region	Sequencer	Mean coverage	>10X	SNVs/indels per patient[1]	Ti/Tv[2]	Concordance rate[3]	dbSNP v137 coverage	RV per patient[4]
HK	5	Illumina Truseq	62.3 M	Illumina GAII	27.9 X	74%	10475 / 234	3.52	NA	99.17%	228
NL	10	Agilent SS V4	51.4 M	Illumina HiSeq2000	53.8 X	95%	13603 / 342	3.34	99.10%	99.29%	340
FR	5	Agilent SS V4	51.4 M	Illumina HiSeq2000	51.8 X	92%	12432 / 287	3.42	NA	99.51%	234
SP	4	NimbleGen V2	36.5 M	ABISolid4	47.4 X	82%	10502 / 530	3.59	NA	95.50%	713

This shows comparable read depth and % of targeted exonic bases on the intersected exonic regions (~ 30Mb) for different cohorts; in addition, variant-level metrics are also comparable at exonic regions (Ti/Tv ratio, SNP/Indel counts, dbSNP137 coverage). 1: SNVs passing variant quality recalibration filtering were counted; 2: only SNVs in exonic regions were used to estimate Ti/Tv ratio; 3: concordance between GWAS array and exome data, NA data not available; 4: RV, rare variants with minor allele frequency < 0.01 in dbsnp137, 1000 genome 2012 and ESP 6500 databases; SS: Sure Select.

Supplementary Table 3. Statistical models for mutation prediction

Model	Classifier[1]	Confusion matrix		Sensitivity	Specificity	Precision	F-Measure	AUC (10-fold CV)[2]
DNM status in trios	2ndPL patient + FA parents	93	3	0.692	0.969	0.857	0.766	0.959
		8	18					
Variant presence/ absence in patients	2ndPL patient + FA patient	68	10	0.703	0.872	0.839	0.765	0.824
		22	52					

Two models were trained by stepwise logistic regression on sequencing quality metrics and then used to predict the *de novo* mutation status in a trio or the variant presence/absence status in exome individuals. Training data was from true or false variants validated by Sanger sequencing, as shown in confusion matrix. 1: 2ndPL_patient means "second minimum phred-scaled likelihood (PL) score" in the trio proband; FA_parents means maximum "fractions of reads (FA) supporting each reported alternative allele" from two parents. 2ndPL_patient, FA_patient means PL or FA value for given patient. 2: Area under curve (AUC) calculated from 10-fold cross-validation. Confusion matrix, F-measure and AUC were acquired from WEKA output.

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

	A	B		C	
Mutation type	**HSCR-trios (N=24)**	**Healthy-trios# (N=54)**	***p*-value**	**Unaffected siblings (N=677)&**	***p*-value**
	Count (rate)	**Count (rate)**	**A *vs.* B**	**Count (rate)**	**A *vs.* C**
All DNMs	28 (1.17)$	44 (0.81)	0.159	547 (0.81)	0.065
LOF DNMs	8 (0.33)	4 (0.07)	0.011*	54 (0.08)	0.001**
Non-*RET* LOF DNMs	3 (0.13)	4 (0.07)	0.447	54 (0.08)	0.447
Synonymous DNMs	7 (0.29)	12 (0.22)	0.62	143 (0.21)	0.365

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

Supplementary Table 5. Joint distribution of common and rare variants for each trio proband.

Pheno-type[1]	*RET* rs2435357: T/C	*De novo* mutations[3]	Inherited mutations in genes in which *de novo* mutations were found[3]	Inherited mutations in 116 ENS/HSCR candidate genes[5]
L, F	CC	***RET*: 3splicing9+1 (splicing site)**; *RBM25*: L158L (synonymous)	***RET*: L56M (missense) (P)**	*SMO* (P), *KIAA1279* (M)
L, F	CC	*COL6A3*: H1109H (synonymous); ***RET*:S837fs (frameshift)**		*DCC* (P)
L, F	TC	***RET*:Y606fs (frameshift)**;		*SON* (M)
L, F	CT	*DAB2IP*: H1132Y (missense); ***NUP98*:N1662S (missense)**; *VEZF1*:S195F (missense); *ZNF57*:D190D (synonymous); *ISG20L2*:G321R (missense); *MED26*:A225A (synonymous); ***NCLN*:Q166* (stopgain)**	***NUP98*:I1609T (missense) (M)**	*IKBKAP* (P), *SOX10* (M)
L, F	CC	*SCUBE3*: N498I (missense); ***RET*: G588fs (frameshift)**	*PLEKHG5:* E800fs (frameshift) (U)	*NOTCH3* (M)
L, M	TT	*PLEKHG5:*T876T (synonymous); *AFF3*:V659L(missense)		
L, M	TT	*KDM4A*: N9S (missense)	*MAP4*: A882G (missense) (M)	*ECE1* (P), *JAG1* (P)
L, M	CT	*MAP4*:G1117G (synonymous)		*PCDHA1* (P), *DCC* (M), *NOTCH3* (P)
L, F	TT	***RET*:C620R (missense)**		
TCA, M	TT	***CKAP2L*:E186fs (frameshift)**		*CBR1* (M)
L, F	C/T	*HMCN1*:A3456T (missense); ***RET*:C137G (missense)**; *TUBG1*:S233S (synonymous)		
L, F	C/T	*CCR2*:L283Q (missense); ***DENND3*:K640fs (frameshift)**		*IKBKAP* (P), *JAG1* (M)
L, F	C/T	***RET*:C570* (stopgain)**		*ECE1* (P)
L, F	C/T	***RET*:R175del (non-frameshift); *TBATA*:R53C (missense)**		*NOTCH1* (P), *PFKL* (P)
L, F	CC		*HMCN1*: P1269T (missense) (P)	*IKBKAP* (P), *EDNRB* (P), *JAG1* (M)
L, F	CC		*HMCN1*: N2461S (missense) (M)	*PHACTR4* (P), *GLI3* (M), *SHH* (M), *HMX3* (M), *NAV2* (M), *PRPH* (P), *PSPN* (P)
L, M	TT			*IHH* (P), *PFKL* (P, M)

S, F	TT		*JAG1* (P)
TCA, M	TT		*ELAVL4* (P), *SERPINI1* (U), *PTCH1* (U), *IKBKAP* (M)
TCA, M	TT		*JAG1* (M)
L, F	TT		*PLXNB1* (P)
L, F	TT	*SCUBE3*:R907C (missense) (P)	*NRG1* (M), *IFNGR2* (P)
L, F	TC		*TAGLN3* (P)
L, F	TC	*DAB2IP*:A338T (missense) (P); *KDM4A*: V988M (M)	*SON* (P)

Common risk SNP (RET rs2435357), DNMs, inherited damaging variants in genes carrying DNMs, rare damaging variants in ENS candidate genes were tabulated for each HSCR patient. DNMs and inherited variants in DNM genes were confirmed by Sanger sequencing. Rare damaging variants in ENS candidate genes were all predicted as true according to training model 2 (see Supplementary Table 3). 1: L: Long segment aganglionosis; S: Short segment aganglionosis; TCA: Total colonic aganglionosis; F: Female; M: Male. 2: rs2435357, T is risk allele, reference allele, and minor allele. 3: genes functionally validated in bold; 4: parent of origin for mutation in candidate genes, P for paternal (P); M for Maternal, U for Unsure; 5: 116 ENS-related HSCR candidate genes (as listed in Supplementary Table 6).

Supplementary Table 6. Characteristics of 116 ENS-related HSCR candidate genes.

Gene	Gene name	Chromosome	Evidence	Ref
ALDH1A2	aldehyde dehydrogenase 1 family, member A2	15q22.1	Mouse (Absence EN)	5
ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3	3p14.3	Expression	6,7
ARTN	artemin	1p34.1	Mouse (Abnormal ENS morphology)	8
ASCL1	achaete-scute complex homolog 1 (Drosophila)	12q23.2	Mouse (Absence EN)/Expression	7,9–11
CADM1	cell adhesion molecule 1	11q23.2	Expression	7,12
CARTPT	CART prepropeptide	5q13.2	Expression	7
CBR1	carbonyl reductase 1	21q22.13	Expression	13
CDH2	cadherin 2, type 1, N-cadherin (neuronal)	18q11.2	Expression	7,14
CRMP1	collapsin response mediator protein 1	4p16.1	Expression	7,15
CSTB	cystatin B (stefin B)	21q22.3	Expression	16
CTNNAL1	catenin (cadherin-associated protein), alpha-like 1	9q31.3	Expression	7
DCC	deleted in colorectal carcinoma	18q21.2	Mouse (Absence submucosal ganglia)	17
DCX	doublecortin	Xq22.3-q23	Expression	7
DLL1	delta-like 1 (Drosophila)	6q27	Not described	10,11
DLL3	delta-like 3 (Drosophila)	19q13.2	Not described	10,11
DLX1	distal-less homeobox 1	2q32	Expression	7,18,19
DPYSL3	dihydropyrimidinase-like 3	5q32	Expression	7,20
EBF3	early B-cell factor 3	10q26.3	Expression	7,21
ECE1	endothelin converting enzyme 1	1p36	Human (Linkage)/Mouse (Absence EN)	17,22,23
EDN3	endothelin 3	20q13	Human (Linkage)/Mouse (Absence EN)	17,22,24,25
EDNRB	endothelin receptor type B	13q22	Human (Linkage/CNV)/Mouse (Absence EN)	17,22,26,27
ELAVL2	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)	9p21	Expression	7,28
ELAVL4	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 4 (Hu antigen D)	1p34	Expression	7,29
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	17q12	Mouse (Abnormal ENS morphology)	30,31
ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	12q13.2	Mouse (Abnormal ENS morphology)	30,31
ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	2q33.3-q34	Human (CNV)	32

ETV1	ets variant 1	7p21.3	Expression	7,33
FGF13	fibroblast growth factor 13	Xq26.3	Expression	7,34
GAP43	growth associated protein 43	3q13.1-q13.2	Expression	7,35
GDNF	glial cell derived neurotrophic factor	5p13	Human (Linkage)/Mouse (Absence EN)/Expression	17,22,36–39
GFRA1	GDNF family receptor alpha 1	10q25	Human (1 patient)/Mouse (Absence EN)/Expression	7,8,38,40
GFRA2	similar to GDNF family receptor alpha 2; GDNF family receptor alpha 2	8p21.3	Mouse (Abnormal ENS morphology)	8
GFRA3	GDNF family receptor alpha 3	5q11.2	Mouse (Abnormal sympathetic system)	8
GFRA4	GDNF family receptor alpha 4	20p13	Not described	8
GLI1	GLI family zinc finger 1	12q13.3	Mouse (Abnormal intestinal morphology)	41,42
GLI2	GLI family zinc finger 2	2q14	Mouse (Abnormal intestinal morphology)	41,42
GLI3	GLI family zinc finger 3	7p14	Mouse (Abnormal intestinal morphology)	41,42
GNG2	guanine nucleotide binding protein (G protein), gamma 2	14q21	Expression	7
GNG3	guanine nucleotide binding protein (G protein), gamma 3	11p11	Expression	7,43
GRB10	growth factor receptor-bound protein 10	7p12.2	Human (Linkage)	38
HES1	hairy and enhancer of split 1, (Drosophila)	3q29	Mouse (Abnormal intestinal morphology)	41,42
HLX	H2.0-like homeobox	1q41	Mouse (Hypoaganglionosis)	17
HMP19	HMP19 protein	5q35.2	Expression	7
HMX3	H6 family homeobox 3	10q26.13	Expression	7,44
HOXB5	homeobox B5	17q21.3	Expression	7,45
HOXD4	homeobox D4	2q31.1	Expression	7,46
IFNGR2	interferon gamma receptor 2 (interferon gamma transducer 1)	21q22.11	Expression	47
IHH	Indian hedgehog homolog (Drosophila)	2q35	Mouse (Absence EN)	41,42
IKBKAP	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	9q31.3	Human (Co-Expression)	48
IL10RB	interleukin 10 receptor, beta	21q22.11	Expression	49
ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	10p11.22	Mouse (Absence EN)	17
JAG1	jagged 1 (Alagille syndrome)	20p12.1	Not described	10,11
JAG2	jagged 2	14q32.33	Not described	10,11
KIAA1279	KIAA1279	10q21	Human (Linkage/GSM syndrome)	50
KLF4	Kruppel-like factor 4 (gut)	9q31	Mouse (Abnormal intestinal morphology)	41,42
L1CAM	L1 cell adhesion molecule	Xq28	Human (Hydrocephalus)/Mouse (Delayed NCC differentiation)/Expression	7,51–53
MAB21L1	mab-21-like 1 (C. elegans)	13q13	Expression	7,54
MAPK10	mitogen-activated protein kinase 10	4q22.1-q23	Expression	7,55
MAPT	microtubule-associated protein tau	17q21.1	Expression	7
MLLT11	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11	1q21	Expression	7,56
NAV2	neuron navigator 2	11p15.1	Human (Exome)	57
NKX2-1	NK2 homeobox 1	14q13	Human (Case report)	17,22
NOTCH1	Notch homolog 1, translocation-associated (Drosophila)	9q34	Not described	10,11
NOTCH2	Notch homolog 2 (Drosophila)	6q27	Not described	10,11
NOTCH3	Notch homolog 3 (Drosophila)	19p13.12	Not described	10,11
NRG1	neuregulin 1	8p12	Human (GWAS)/Mouse (Abnormal NCC migration)	58

NRG3	neuregulin 3	10q22-q23	Human (CNV/Exome)	32,59
NRP1	neuropilin 1	10p11.22	Not described	60,61
NRTN	neurturin	19p13	Human (Linkage)/Mouse (Abnormal ENS)	62
NTF3	3'-nucleotidase	1p36.11	Mouse (Reduced enteric ganglia)	17
NTRK3	neurotrophic tyrosine kinase, receptor, type 3	15q25	Mouse (Reduced enteric ganglia)	17
PAX3	paired box 3	2q36	Human (Exome)/Mouse (Absence EN)	8,57
PCDHA1	protocadherin alpha 1; protocadherin alpha 4	5q31	Expression	7,63
PFKL	phosphofructokinase, liver	21q22.3	Expression	64
PHACTR4	phosphatase and actin regulator 4	1p35.3	Mouse	65,66
PHOX2A	paired-like homeobox 2a	11q13.2	Expression	7,67
PHOX2B	paired-like homeobox 2b	4p13	Human (Haddad syndrome)/Mouse (Abnormal ENS)/Expression	7,68, 69
PLXNA1	plexin A1	3q21.3	Not described	60,61
PLXNB1	plexin B1	3p21	Human (Linkage)	70,71
POFUT1	protein O-fucosyltransferase 1	20q11.21	Mouse (Absence EN)	10,11
PROK1	prokineticin 1	1p13	Not described	72,73
PROK2	prokineticin 2	3p13	Not described	72,73
PROKR1	prokineticin receptor 1	2p14	Not described	72,73
PROKR2	prokineticin receptor 2	20p12	Not described	72,73
PRPH	peripherin	12q12-q13	Expression	7,74
PSPN	persephin	19p13.3	Not described	8
PTCH1	patched homolog 1 (Drosophila)	9q22.32	Mouse (Abnormal intestinal morphology)	41,42
RET	ret proto-oncogene	10q11	Human (Linkage/CNV/Exome)/Mouse (Absence EN)	7,57, 75,76
SALL4	sal-like 4 (Drosophila)	20q13.2	Mouse (Absence EN)	17
SCG3	secretogranin III	15q21	Expression	7
SEMA3A	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A	7p12	Human (Association)/Mouse (Abnormal ENS morphology)	57,77
SEMA3C/D	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D	7p12	Human (GWAS)/Others (Not described)	60,61
SERPINI1	serpin peptidase inhibitor, clade I (neuroserpin), member 1	3q26.1	Expression	7,78, 79
SHH	sonic hedgehog homolog (Drosophila)	7q36	Mouse (Ectopic enteric ganglia formation)	41,42
SMO	smoothened homolog (Drosophila)	7q32	Mouse (Abnormal neural crest cell migration)	41,42
SOD1	superoxide dismutase 1, soluble	21q22.11	Expression	80,81
SON	SON DNA binding protein	21q22.11	Expression	82
SOX10	SRY (sex determining region Y)-box 10	22q13	Human (Linkage/WS4)/Mouse (Absence EN)	7,83
SOX2	SRY (sex determining region Y)-box 2	3q26.3-q27	Expression	7,84
SPRY2	sprouty homolog 2 (Drosophila)	13q31.1	Mouse (Increased EN)	17
STMN2	stathmin-like 2	8q21.13	Expression	7,85
STMN3	stathmin-like 3	20q13.3	Expression	7
SUFU	suppressor of fused homolog (Drosophila)	10q24.32	Mouse (Abnormal neural tube morphology)	41,42
SYT11	synaptotagmin XI	1q21.2	Expression	7,86
TAGLN3	transgelin 3	3q13.2	Expression	7
TBX3	T-box 3	12q24.1	Expression	87,88
TFF3	trefoil factor 3 (intestinal)	21q22.3	Expression	7,89, 90
TGFB2	transforming growth factor, beta 2	1q41	Expression	7,91
TMEFF2	transmembrane protein with EGF-like	2q32.3	Expression	7,92

	and two follistatin-like domains 2			
TREX1	three prime repair exonuclease 1	3p21.31	Human (Linkage)	70,71
TTC3	tetratricopeptide repeat domain 3; tetratricopeptide repeat domain 3-like	21q22.2	Expression	93
TUBB3	tubulin, beta 3; melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	16q24.3	Expression	7,94
UCHL1	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	4p14	Expression	7,95
VIP	vasoactive intestinal peptide	6q25	Expression	7,96
ZEB2	zinc finger E-box binding homeobox 2	2q22	Human (Linkage/MW syndrome/CNV)/Mouse (Abnormal NCC migration)	97–99
ZIC2	Zic family member 2 (odd-paired homolog, Drosophila)	13q32	Mouse	100, 101

Notes: updated gene symbols used for ZFHX1B (replaced by ZEB2), TRKC (replaced by NTRK3) and RALDH2 (replaced by ALDH1A2); EN: enteric neurons; NCC: neural crest cells.

Supplementary Table 7. Gene recurrence and burden test.

Gene symbol	HK (14/73)		Spain (15/100)%		Rotterdam (19/39)		Meta-analysis (48/212)	
	p-value	Direction*	*p*-value	Direction*	*p*-value	direction	*p*-value	Direction*
AFF3	1.0000	-1	0.6973	-1	0.0392	1	0.4745	1
CCR2	1.0000	-1	0.6973	-1	0.0392	1	0.4745	1
CKAP2L	0.0216	1	0.1175	1	1.0000	-1	**0.0178**	1
COL6A3	0.5883	-1	0.5430	-1	0.5970	1	0.6398	-1
DAB2IP	0.4402	-1	0.6973	-1	0.1484	1	0.9819	-1
DENND3	0.3699	-1	0.6973	-1	0.5970	1	0.5977	-1
HMCN1	0.4308	-1	0.3662	1	0.8013	-1	0.9789	1
ISG20L2	1.0000	-1	1.0000	-1	0.1484	1	0.4949	1
KDM4A	0.0216	1	0.4967	-1	0.1484	1	0.1190	1
MAP4	0.6596	-1	0.2903	1	0.5970	1	0.4851	1
MED26	1.0000	-1	1.0000	-1	1.0000	-1	1.0000	-1
NCLN	1.0000	-1	1.0000	-1	0.1484	1	0.4949	-1
NUP98	0.5309	-1	0.6973	-1	0.0392	1	0.7243	1
PLEKHG5	1.0000	-1	0.1999	-1	0.9826	1	0.3997	-1
RBM25	1.0000	-1	0.0095	1	1.0000	-1	0.0846	1
RET	0.1867	1	0.6367	1	0.0008	1	**0.0078**	1
SCUBE3	1.0000	-1	0.5806	-1	1.0000	-1	0.7133	-1
TBATA	1.0000	-1	1.0000	-1	0.5970	1	0.8028	1
TUBG1	1.0000	-1	1.0000	-1	1.0000	-1	1.0000	-1
VEZF1	1.0000	-1	0.6973	-1	0.1484	1	0.6717	1
ZNF57	0.0216	1	0.4967	-1	1.0000	-1	0.3808	1

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

Supplementary Table 8. Bioinformatics prediction of the functional impact of DNMs.

Gene and mutation	RNA structure change[1]	Human splicing finder[2]	Conservation (PhyloP)[3]	Gene-level relevance[4]
RET:splicing9+1	0.7866	splice donor	7.88	Major HSCR gene
RBM25:L158L	0.8224			Constrained gene[102], interacts with *PAX3*
RET:S837fs	NA			Major HSCR gene
COL6A3:H1109H	0.4898			Interacts with *ERBB2*, *ITGB1*, shares pathway with *NRTN, GDNF*
RET:Y606fs	NA			Major HSCR gene
DAB2IP:H1132Y	0.8539		1.76	
ISG20L2:G321R	0.495		7.59	
MED26:A225A	0.9717			Pathway sharing with *NOTCH* genes
NCLN:Q166*	0.5467		7.38	Nodal signaling, involves in CNS development
NUP98:N1662S	0.5235		5.95	Regulation of known HSCR genes, involves in CNS development
VEZF1:S195F	**0.0565**		9.86	
ZNF57:D190D	0.5217			
RET:G588fs	NA	splice-acceptor		Major HSCR gene
SCUBE3:N498I	0.5473		1.96	Hedgehog signaling
KDM4A:N9S	0.5813		3.54	Neural crest specification in chicken
PLEKHG5:T876T	0.4096			
AFF3:V659L	**0.0272**		1.84	
MAP4:G1117G	0.996		1.40	Interacts with *CDH2*, *ERBB2*, *MAPT* and *DCX*
RET:C620R	0.4286		4.80	Major HSCR gene
CKAP2L:E186fs	NA			Involves in CNS development
RET:C137G	0.6841		3.73	Major HSCR gene
HMCN1:A3456T	0.6906		0.60	
TUBG1:S233S	0.5802			Interacts with *SOX2*
CCR2:L283Q	**0.0659**		5.87	3p21; interacts with *GLI2*; shares pathway with *EDN3*
DENND3:K640fs	NA			Involves in CNS development
RET:C570*	**0.1453**		-0.02	Major HSCR gene
RET:R175del	NA			Major HSCR gene
TBATA:R53C	0.5526		1.71	Involves in CNS development

Bioinformatics prediction tools, databases and literature were used to predict functional impact of DNMs and the genes carrying DNMs. 1: significant changes (< 0.2) are in bold; 2: only potential splice sites (donor or acceptor) are shown; 3: Phylop score > 2 means conservative; 4: evidence collected from PubMed literature and bioinformatics databases (STRING, MsigDB pathways).

Supplementary Table 9. Sequence and dosage of antisense morpholino.

a. Splice-blocking morpholino

Target gene	Human ortholog	Sequence	Dosage (ng)
aff3	*AFF3*	AAATGTCTTTCCCCCCTCACCTTTC	6
ckap2l	*CKAP2L*	TGAAGTAAACTCACAGTCTTTCCTC	6
dab2ipa	*DAB2IP**	AGGTCAGCAGACTCACCTCGAAGCA	6
dab2ipb	*DAB2IP**	GCTTTCCACTAACACCTTACCCAGC	6
dennd3a	*DENND3**	CATCTTTACCCTGTGCGAAAAGTTA	6
dennd3b	*DENND3**	CCATTCAATTTTGTTTCACCTGGAA	6
hmcn1	*HMCN1*	GCACAAAGATTTCCCCTTACCCTGA	6
isg20l2	*ISG20L2*	CTACTGATGCTTATTTCATACCTCT	6
kdm4aa	*KDM4A**	GACACAAGCAATGACAGTACCAGGA	6
kdm4ab	*KDM4A**	AGTTGAACAGAACATACTTGTCGCT	6
ncl1	*NCLN*	GAACCTGCCAATGGATGTGGTTTAT	6
nup98	*NUP98*	GTATGGAGCAGCTAAACTTACGGTT	1
scube3	*SCUBE3*	ACTAGATGAAGGGACTCACTCTTGC	6
tbata	*TBATA*	GATAGAGCCCAATACTGTACCTCCC	4

vezf1a	*VEZF1**	AGCCAATCGCACTAGCCTTACCTTT	6
vezf1b	*VEZF1**	ATCCAAAATGCTAAACCCACCTAGA	6

b. Translation-blocking morpholino

Target gene	Human ortholog	Sequence	Dosage (ng)
ckap2l	*CKAP2L*	GTCTTCATCAGTCATCGTTTCCATC	6
dennd3a	*DENND3**	GACCACACGGCACATTATCAGCCAT	8
dennd3b	*DENND3**	GACCGTCTGCCATTGAAAATCAACA	8
ncl1	*NCLN*	ACCTCACCAGCCTCCTCGAACATGC	0.8
nup98	*NUP98*	GTTGAACATCTTGCACTGCTATAGA	12
tbata	*TBATA*	AGCACCTGCACAAACAAATCAGACT#	6

c. Control morpholino

Target gene	Human ortholog	Sequence^	Dosage (ng)
ckap2l	*CKAP2L*	TGtAcTAAAgTCACAcTgTTTCCTC	6
dennd3a	*DENND3**	CAaCaTTACgCTGTGCcAAAAcTTA	6
dennd3b	*DENND3**	CCAaTgAATTTTcTTTCACgTcGAA	6
ncl1	*NCLN*	GAACaTcCCAATGaATcTGaTTTAT	6
nup98	*NUP98*	GTtTcGAGCAcCTAAAgTTACcGTT	1
tbata	*TBATA*	GAaAcAGCCgAATACTcTAgCTCCC	4
p53	*P53*	GCGCCATTGCTTTGCAAGAATTG	2
HBB%		CCTCTTACCTCAGTTACAATTTATA	12

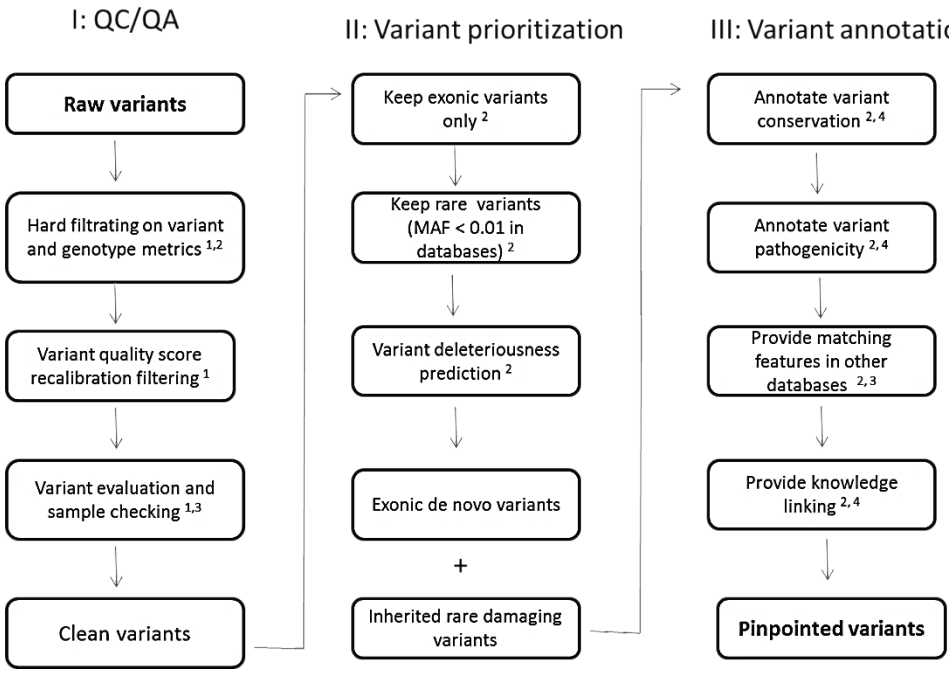
**DENND3*, *DAB2IP*, *KDM4A* and *VEZF1* are duplicated in zebrafish genome. #There was no suitable target site in *tbata* for translation-blocking morpholino. A second non-overlapping splice-blocking morpholino was used instead. ^Small letters indicate the mismatch nucleotides to the corresponding splice-blocking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

Supplementary Table 10. qPCR/RT-PCR primers.

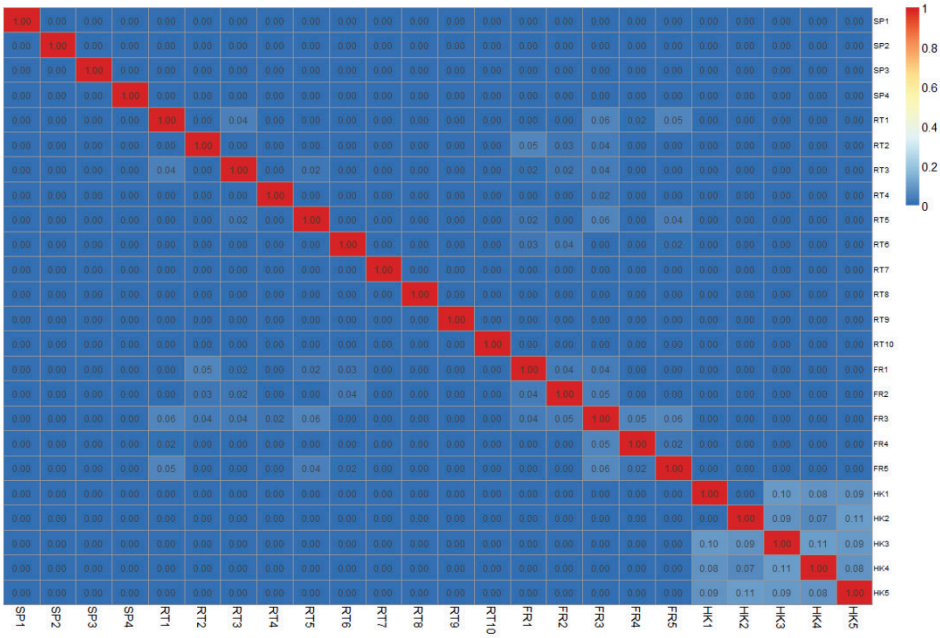
Target transcript	Forward / Reverse	Sequence (5' to 3')
aff3	Forward	AAAGCAGCAGTCAACGTTCC
	Reverse	CATCTGTCCAACTGCCAATG
ckap2l	Forward	TGAGATCCAACCACACCAAG
	Reverse	GTTCCACAGCGAAGACAATG
dab2ipa	Forward	TGGGACAGGATTTCTGCTTC
	Reverse	GCACAGCACGTCTCAAATTC
dab2ipb	Forward	GCACTAAAGCCATCGAGGAG
	Reverse	ACGGGTCCACTTCACAGTTC
dennd3a	Forward	TGCTTGGAGTGTCAAACGAG
	Reverse	ATAAACGGTGGAGCGTGAAC
dennd3b	Forward	GCAGCCTCTGATGATTGTCCT
	Reverse	GTTGGGACAGTATGGGCACA
hmcn1	Forward	GAAGAAATTGCCTCGACCAG
	Reverse	AGCAGGTGAACCTTTGAGGA
isg20l2	Forward	ACTCGCTGGAGTGGAATCAG
	Reverse	GGATAGCATGTCCCACAACC
kdm4aa	Forward	GGGATGTGGAAGAGCACATT
	Reverse	TGCTCTGGAGGCACAACATA
kdm4ab	Forward	TGAAAGAGTTCCGCAAAACC
	Reverse	CAGCTCCATAGATGGGAGGA
ncl1	Forward	CTGTTTCTGTCGGTCGGAAT
	Reverse	ATCACACAGCGACGACTCAG

nup98	Forward	GAACCTGGGGTTTGGATTCT
	Reverse	CCAGCATCACTTCCTCCAAT
scube3	Forward	TCTCCTGTCCTGGAAACACC
	Reverse	ACTCCACATTGGCTGGGTAG
tbata	Forward	CTGAAAGCTGGCGTGAGGAA
	Reverse	GTGTGTGTGTTGTCGTACGC
vezf1a	Forward	GATGGAGGTGTCCACAAACC
	Reverse	GCAGGCCGTTACTTGACATT
vezf1b	Forward	GCACAAGCCCTACATCTGCT
	Reverse	TGGCATTTAAAGGGTCGTTC
elfa	Forward	CTTCTCAGGCTGACTGTGC
	Reverse	CCGCTAGCATTACCCTCC
actb	Forward	TACAATGAGCTCCGTGTTGC
	Reverse	GTTCCCATCTCCTGCTCAAA

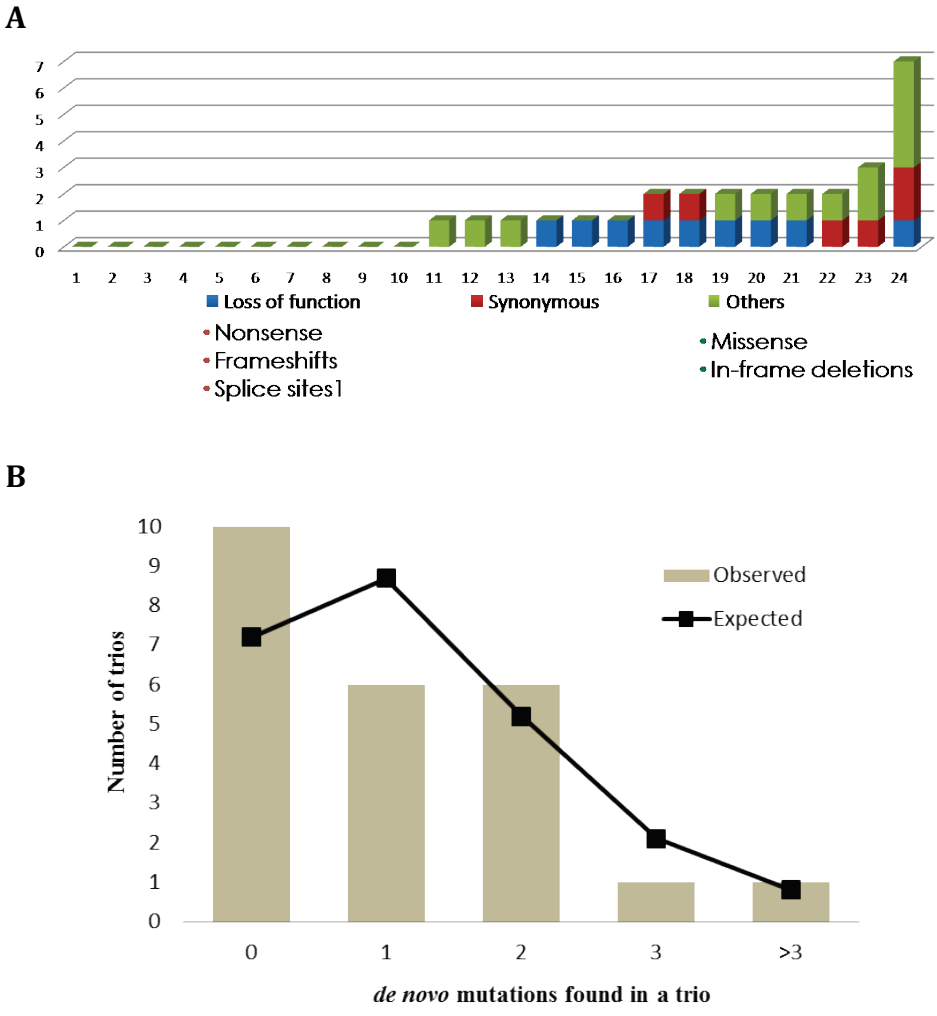
SUPPLEMENTARY FIGURES



Supplementary Figure 1. Analytical pipeline for exome sequence filtration and prioritization. 1: GATK; 2: KGGSeq; 3: PLINK; 4: ANNOVAR. KGGSeq integrates different kinds of knowledge resources from (epi)genetic databases, pathways databases and protein-protein interaction networks to annotate the genes that harbor any post-QC variants as well as to predict the potential pathogenicity of their variants. For deleteriousness prediction, KGGSeq integrates 5 prediction programs (Polyphen2, Sift, MutationTaster, PhyloP and Likelihood ratio) which are weighted by logistic regression[103]. Annovar is mainly used to double-check the final remaining variant for annotation, and provides supplementary features from Database of genomic variation (DGV) and clinical variation database (ClinVAR).

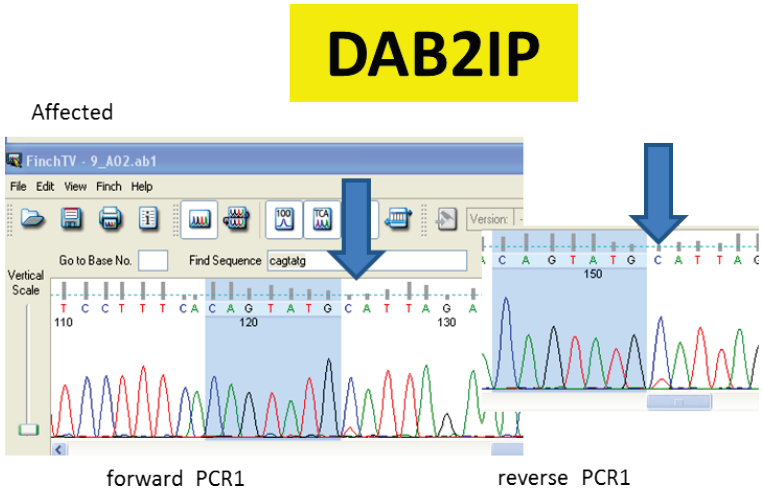


Supplementary Figure 2. Relatedness plotting of HSCR exome sequences. Around 17K common SNPs (minor allele frequency > 0.01 in 1000Genomes European populations) were used to calculate identical by descent (IBD) and identical by state (IBS) proportion. Each cell shows pi_hat statistics[1] (IBD proportion, calculated from P(IBD=2)+0.5*P(IBD=1); http://pngu.mgh.harvard.edu/~purcell/plink/ibdibs.shtml) between two patients. No pairwise pi_hat coefficients are above 0.125 (the first cousin relationship); the light blue cells represent 0.07~0.11 for samples mainly from HK population, which is expected to be different from other European patients.

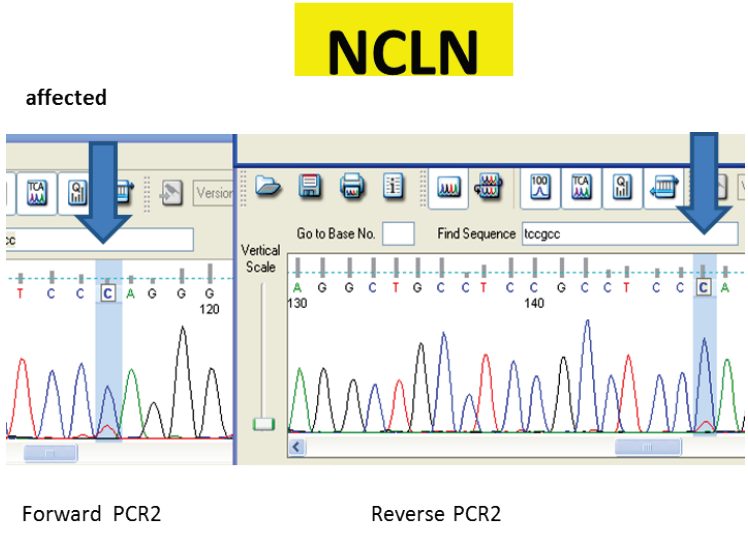


Supplementary Figure 3. Distribution of *de novo* mutations per trio. A) Number of DNMs (separated by mutation type) in each trio, categorized into three different types (Loss of function, synonymous and others). B) Distribution of observed counts of DNMs per trio and expected counts per trio calculated from Poisson distribution (lambda at 1.2).

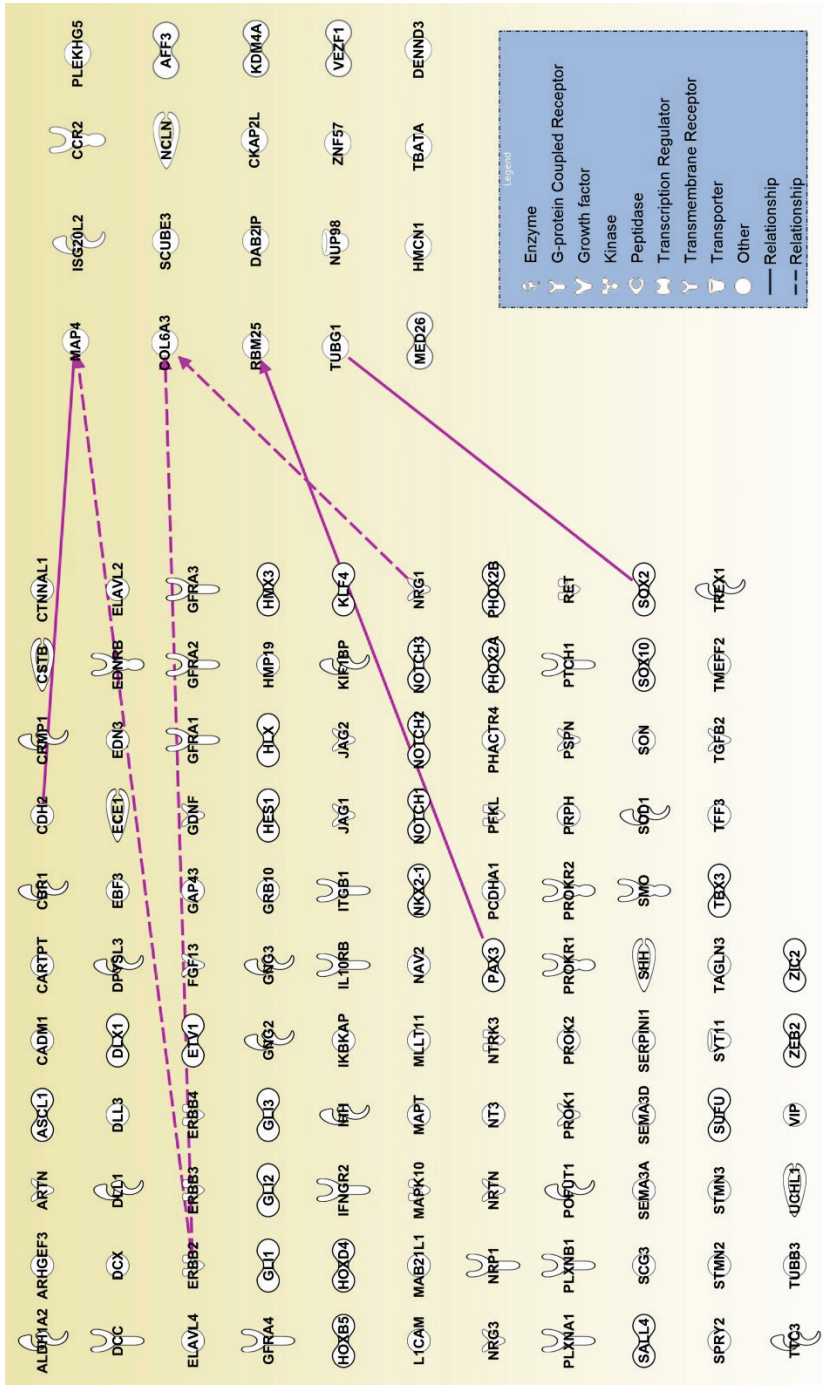
A



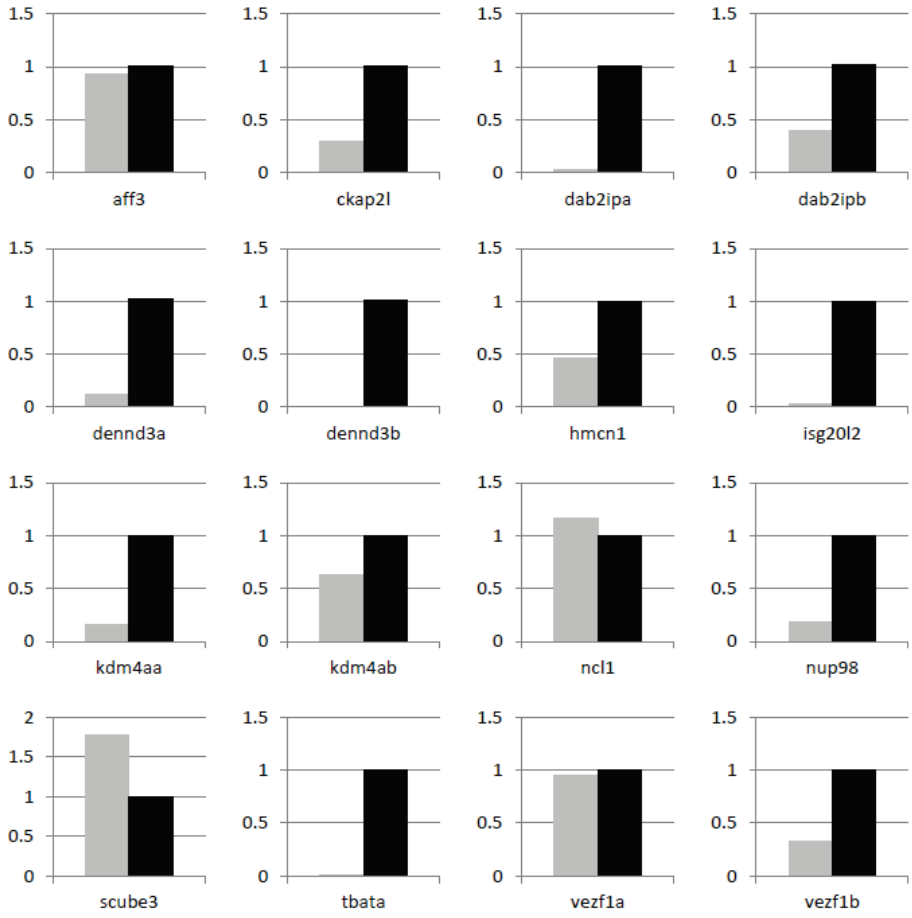
B



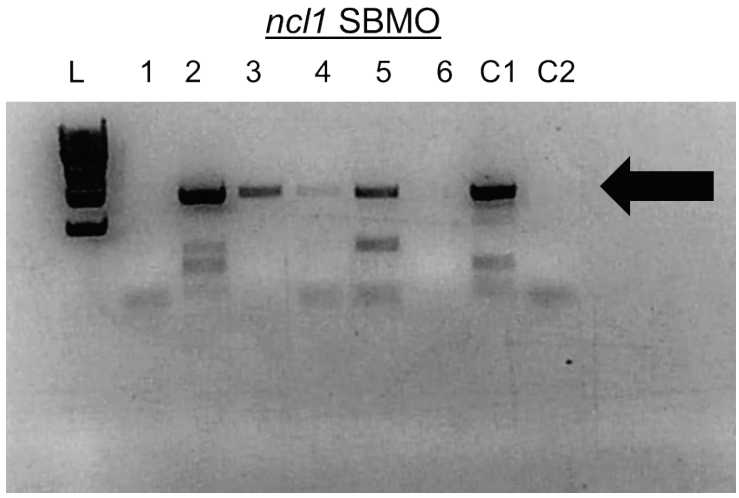
Supplementary Figure 4. Sanger confirmation of mosaic DNMs in *DAB2IP* and *NCLN*. Two out of 28 *de novo* mutations (in DAB2IP and *NCLN*) were confirmed as mosaic mutations by Sanger sequencing (forward and reverse Sequencing direction). A) Peak for the *DAB2IP* heterozygous mosaic mutation. B) peak for the *NCLN* heterozygous mosaic mutation.



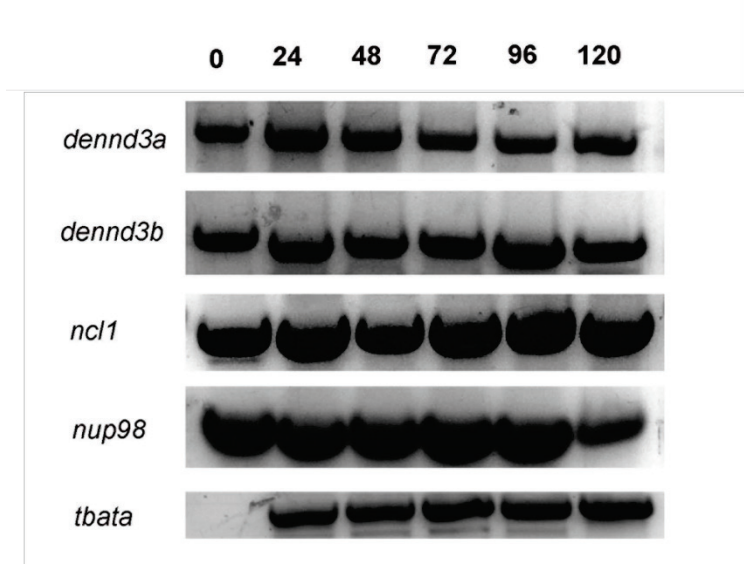
Supplementary Figure 5. Connection of DNM genes and ENS genes at pathway/network level. Ingenuity Pathway Analysis (IPA) was used to link 116 ENS candidate genes (left, Supplementary Table 8) with the 20 newly found genes harboring *de novo* mutations (right). Solid and dotted lines represent direct and indirect interactions, respectively.



Supplementary Figure 6. qPCR confirmation of gene knockdown by SBMO. Relative expression of the candidate genes between SBMO-injected (grey bar) and control morpholino-injected embryos (black bar) by qPCR.



Supplementary Figure 7. RT-PCR confirmation of *ncl1* SBMO knockdown. *ncl1* expressions in six 1dpf embryos injected with *ncl1* SBMO were compared to control MO injected embryos. Arrow indicated the expected amplicon. L: ladder; C1: control MO-injected embryo; C2: RT negative control.



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

REFERENCES

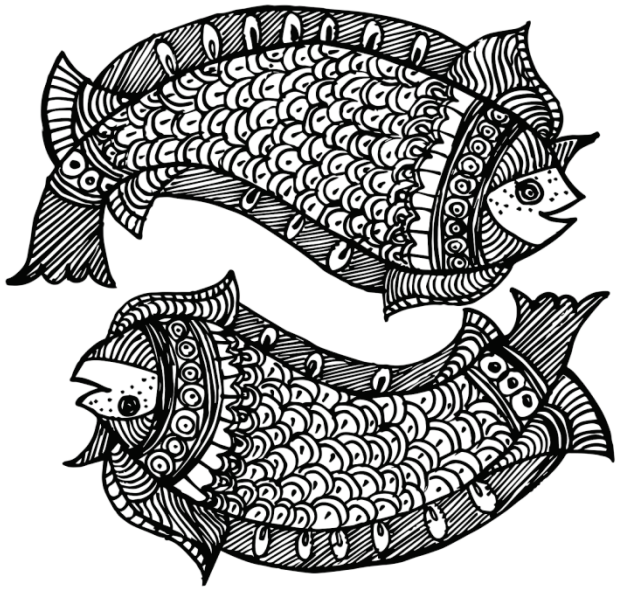
1. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
2. Jiang, Q., Ho, Y.-Y., Hao, L., Nichols Berrios, C. & Chakravarti, A. Copy number variants in candidate genes are genetic modifiers of Hirschsprung disease. *PLoS One* **6,** e21219 (2011).
3. Gui, H. *et al.* Targeted next-generation sequencing on hirschsprung disease: A pilot study exploits DNA pooling. *Ann. Hum. Genet.* **78,** 381–387 (2014).
4. Alves, M. M. *et al.* Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev. Biol.* **382,** 320–9 (2013).
5. Anderson, R. B., Newgreen, D. F. & Young, H. M. Neural crest and the development of the enteric nervous system. *Advances in Experimental Medicine and Biology* **589,** 181–196 (2006).
6. Thiesen, S., Kübart, S., Ropers, H. H. & Nothwang, H. G. Isolation of two novel human RhoGEFs, ARHGEF3 and ARHGEF4, in 3p13-21 and 2q22. *Biochem. Biophys. Res. Commun.* **273,** 364–9 (2000).
7. Heanue, T. A. & Pachnis, V. Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6919–24 (2006).
8. Pachnis, V., Durbec, P., Taraviras, S., Grigoriou, M. & Natarajan, D. Role Of the RET signal transduction pathway in development of the mammalian enteric nervous system. *Am. J. Physiol.* **275,** 321–327 (1998).
9. Huang, H. S. *et al.* Direct transcriptional induction of Gadd45gamma by Ascl1 during neuronal differentiation. *Mol. Cell. Neurosci.* **44,** 282–96 (2010).
10. Okamura, Y. & Saga, Y. Notch signaling is required for the maintenance of enteric neural crest progenitors. *Development* **135,** 3555–3565 (2008).
11. Morrison, S. J. *et al.* Transient Notch activation initiates an irreversible switch from neurogenesis to gliogenesis by neural crest stem cells. *Cell* **101,** 499–510 (2000).
12. Hagiyama, M., Ichiyanagi, N., Kimura, K. B., Murakami, Y. & Ito, A. Expression of a soluble isoform of cell adhesion molecule 1 in the brain and its involvement in directional neurite outgrowth. *Am. J. Pathol.* **174,** 2278–89 (2009).
13. Sultan, M. *et al.* Gene expression variation in Down's syndrome mice allows prioritization of candidate genes. *Genome Biol.* **8,** R91 (2007).
14. Hermiston, M. L. & Gordon, J. I. Inflammatory bowel disease and adenomas in mice expressing a dominant negative N-cadherin. *Science* **270,** 1203–1207 (1995).
15. Schmidt, E. F., Shim, S.-O. & Strittmatter, S. M. Release of MICAL autoinhibition by semaphorin-plexin signaling promotes interaction with collapsin response mediator protein. *J. Neurosci.* **28,** 2287–2297 (2008).
16. Brännvall, K. *et al.* Cystatin-B is expressed by neural stem cells and by differentiated neurons and astrocytes. *Biochem. Biophys. Res. Commun.* **308,** 369–374 (2003).
17. Heanue, T. a & Pachnis, V. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat. Rev. Neurosci.* **8,** 466–79 (2007).
18. Simeone, A. *et al.* Cloning and characterization of two members of the vertebrate Dlx gene family. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2250–4 (1994).
19. Cobos, I., Borello, U. & Rubenstein, J. L. R. Dlx Transcription Factors Promote Migration through Repression of Axon and Dendrite Growth. *Neuron* **54,** 873–888 (2007).
20. Hamajima, N. *et al.* A novel gene family defined by human dihydropyrimidinase and three related proteins with differential tissue distribution. *Gene* **180,** 157–163 (1996).
21. Zardo, G. *et al.* Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat. Genet.* **32,** 453–458 (2002).
22. Amiel, J. & Lyonnet, S. Hirschsprung disease, associated syndromes, and genetics: a review. *J. Med. Genet.* **38,** 729–739 (2001).
23. Hofstra, R. M. *et al.* A loss-of-function mutation in the endothelin-converting enzyme 1 (ECE-1) associated with Hirschsprung disease, cardiac defects, and autonomic dysfunction. *Am. J. Hum. Genet.* **64,** 304–8 (1999).
24. Bidaud, C. *et al.* Mutations of the endothelin-3 gene in isolated and syndromic forms of Hirschsprung disease. *Gastroenterol Clin Biol* **21,** 548–554 (1997).
25. Sánchez-Mejías, A., Fernández, R. M., López-Alonso, M., Antiñolo, G. & Borrego, S. Contribution of RET, NTRK3 and EDN3 to the expression of Hirschsprung disease in a multiplex family. *J.*

Med. Genet. **46,** 862–864 (2009).
26. Carrasquillo, M. M. *et al.* Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. *Nat. Genet.* **32,** 237–44 (2002).
27. Kenny, S. E. *et al.* Reduced endothelin-3 expression in sporadic Hirschsprung disease. *Br. J. Surg.* **87,** 580–585 (2000).
28. Han, J., Knops, J. F., Longshore, J. W. & King, P. H. Localization of human elav-like neuronal protein 1 (Hel-N1) on chromosome 9p21 by chromosome microdissection polymerase chain reaction and fluorescence in situ hybridization. *Genomics* **36,** 189–191 (1996).
29. Akamatsu, W. *et al.* The RNA-binding protein HuD regulates neuronal cell identity and maturation. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 4625–4630 (2005).
30. Paratore, C., Goerich, D. E., Suter, U., Wegner, M. & Sommer, L. Survival and glial fate acquisition of neural crest cells are regulated by an interplay between the transcription factor Sox10 and extrinsic combinatorial signaling. *Development* **128,** 3949–61 (2001).
31. Britsch, S. The neuregulin-I/ErbB signaling system in development and disease. *Adv. Anat. Embryol. Cell Biol.* **190,** 1–65 (2007).
32. Garcia-Barcelo, M.-M. *et al.* Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 2694–9 (2009).
33. Flames, N. & Hobert, O. Gene regulatory logic of dopamine neuron differentiation. *Nature* **458,** 885–889 (2009).
34. Smallwood, P. M. *et al.* Fibroblast growth factor (FGF) homologous factors: new members of the FGF family implicated in nervous system development. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 9850–7 (1996).
35. Strittmatter, S. M., Fankhauser, C., Huang, P. L., Mashimo, H. & Fishman, M. C. Neuronal pathfinding is abnormal in mice lacking the neuronal growth cone protein GAP-43. *Cell* **80,** 445–452 (1995).
36. Salomon, R. *et al.* Germline mutations of the RET ligand GDNF are not sufficient to cause Hirschsprung disease. *Nat. Genet.* **14,** 345–7 (1996).
37. Ivanchuk, S. M., Myers, S. M., Eng, C. & Mulligan, L. M. De novo mutation of GDNF, ligand for the RET/GDNFR-alpha receptor complex, in Hirschsprung disease. *Hum. Mol. Genet.* **5,** 2023–2026 (1996).
38. Angrist, M. *et al.* Genomic structure of the gene for the SH2 and pleckstrin homology domain-containing protein GRB10 and evaluation of its role in Hirschsprung disease. *Oncogene* **17,** 3065–3070 (1998).
39. Fernandez, R. M., Ruiz-Ferrer, M., Lopez-Alonso, M., Antiñolo, G. & Borrego, S. Polymorphisms in the genes encoding the 4 RET ligands, GDNF, NTN, ARTN, PSPN, and susceptibility to Hirschsprung disease. *J. Pediatr. Surg.* **43,** 2042–2047 (2008).
40. Myers, S. M. *et al.* Investigation of germline GFR alpha-1 mutations in Hirschsprung disease. *J.Med.Genet.* **36,** 217–220 (1999).
41. Fu, M. Sonic hedgehog regulates the proliferation, differentiation, and migration of enteric neural crest cells in gut. *J. Cell Biol.* **166,** 673–684 (2004).
42. Reichenbach, B. *et al.* Endoderm-derived Sonic hedgehog and mesoderm Hand2 expression are required for enteric nervous system development in zebrafish. *Dev. Biol.* **318,** 52–64 (2008).
43. Hurowitz, E. H. *et al.* Genomic characterization of the human heterotrimeric G protein alpha, beta, and gamma subunit genes. *DNA Res.* **7,** 111–120 (2000).
44. Bober, E., Baum, C., Braun, T. & Arnold, H. H. A novel NK-related mouse homeobox gene: expression in central and peripheral nervous structures during embryonic development. *Dev. Biol.* **162,** 288–303 (1994).
45. Fu, M., Lui, V. C. H., Sham, M. H., Cheung, A. N. Y. & Tam, P. K. H. HOXB5 expression is spatially and temporarily regulated in human embryonic gut during neural crest cell colonization and differentiation of enteric neuroblasts. *Dev. Dyn.* **228,** 1–10 (2003).
46. Mavilio, F. *et al.* Differential and stage-related expression in embryonic tissues of a new human homeobox gene. *Nature* **324,** 664–668 (1986).
47. Murai, M. *et al.* Interleukin 10 acts on regulatory T cells to maintain expression of the transcription factor Foxp3 and suppressive function in mice with colitis. *Nat. Immunol.* **10,** 1178–84 (2009).
48. Bolk, S. *et al.* A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. *Proc. Natl. Acad. Sci.*

U. S. A. **97,** 268–73 (2000).
49. Glocker, E. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* **361,** 2033–45 (2009).
50. Brooks, A. S. *et al.* Homozygous nonsense mutations in KIAA1279 are associated with malformations of the central and enteric nervous systems. *Am. J. Hum. Genet.* **77,** 120–6 (2005).
51. Griseri, P. *et al.* Complex pathogenesis of Hirschsprung's disease in a patient with hydrocephalus, vesico-ureteral reflux and a balanced translocation t(3;17)(p12;q11). *Eur. J. Hum. Genet.* **17,** 483–90 (2009).
52. Turner, K. N., Schachner, M. & Anderson, R. B. Cell adhesion molecule L1 affects the rate of differentiation of enteric neurons in the developing gut. *Dev. Dyn.* **238,** 708–15 (2009).
53. Kenwrick, S. Neural cell recognition molecule L1: relating biological complexity to human disease mutations. *Hum. Mol. Genet.* **9,** 879–886 (2000).
54. Mariani, M. *et al.* Two murine and human homologs of mab-21, a cell fate determination gene involved in Caenorhabditis elegans neural development. *Hum. Mol. Genet.* **8,** 2397–2406 (1999).
55. Gupta, S. *et al.* Selective interaction of JNK protein kinase isoforms with transcription factors. *EMBO J.* **15,** 2760–70 (1996).
56. Tse, W., Zhu, W., Chen, H. S. & Cohen, A. A novel gene, AF1q, fused to MLL in t(1;11) (q21;q23), is specifically expressed in leukemic and immature hematopoietic cells. *Blood* **85,** 650–6 (1995).
57. K. D. Nguyen, T. T. The burden of coding, non-coding and chromosomal mutations in Hirschsprung disease. *ASHG 2013 Abstr.* (2013).
58. Garcia-Barcelo, M.-M. *et al.* Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 2694–9 (2009).
59. Yang, J. *et al.* Exome sequencing identified NRG3 as a novel susceptible gene of Hirschsprung's disease in a Chinese population. *Mol. Neurobiol.* **47,** 957–966 (2013).
60. 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,** 581–96 (2015).
61. Kruger, R. P., Aurandt, J. & Guan, K.-L. Semaphorins command cells to move. *Nat. Rev. Mol. Cell Biol.* **6,** 789–800 (2005).
62. Doray, B. *et al.* Mutation of the RET ligand, neurturin, supports multigenic inheritance in Hirschsprung disease. *Hum. Mol. Genet.* **7,** 1449–1452 (1998).
63. Sugino, H. *et al.* Genomic organization of the family of CNR cadherin genes in mice and humans. *Genomics* **63,** 75–87 (2000).
64. Elson, A., Levanon, D., Weiss, Y. & Groner, Y. Overexpression of liver-type phosphofructokinase (PFKL) in transgenic-PFKL mice: implication for gene dosage in trisomy 21. *Biochem. J.* **299 (Pt 2,** 409–15 (1994).
65. Allen, P. B., Greenfield, A. T., Svenningsson, P., Haspeslagh, D. C. & Greengard, P. Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7187–92 (2004).
66. Kim, T. H., Goodman, J., Anderson, K. V. & Niswander, L. Phactr4 Regulates Neural Tube and Optic Fissure Closure by Controlling PP1-, Rb-, and E2F1-Regulated Cell-Cycle Progression. *Dev. Cell* **13,** 87–102 (2007).
67. Sasaki, A. *et al.* Molecular analysis of congenital central hypoventilation syndrome. *Hum. Genet.* **114,** 22–26 (2003).
68. Garcia-Barceló, M. *et al.* Association study of PHOX2B as a candidate gene for Hirschsprung's disease. *Gut* **52,** 563–7 (2003).
69. Ou-Yang, M. C. *et al.* Concomitant existence of total bowel aganglionosis and congenital central hypoventilation syndrome in a neonate with PHOX2B gene mutation. *J. Pediatr. Surg.* **42,** (2007).
70. Gabriel, S. B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat. Genet.* **31,** 89–93 (2002).
71. Garcia-Barceló, M.-M. *et al.* Mapping of a Hirschsprung's disease locus in 3p21. *Eur. J. Hum. Genet.* **16,** 833–40 (2008).
72. Ngan, E. S. W. *et al.* Prokineticin-1 (Prok-1) works coordinately with glial cell line-derived neurotrophic factor (GDNF) to mediate proliferation and differentiation of enteric neural crest

cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1783,** 467–478 (2008).
73. Ngan, E. S. W. *et al.* Prokineticin-1 modulates proliferation and differentiation of enteric neural crest cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1773,** 536–545 (2007).
74. Oblinger, M. M., Wong, J. & Parysek, L. M. Axotomy-induced changes in the expression of a type III neuronal intermediate filament gene. *J Neurosci* **9,** 3766–3775 (1989).
75. Edery, P. *et al.* Mutations of the RET proto-oncogene in Hirschsprung's disease. *Nature* **367,** 378–380 (1994).
76. Emison, E. S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434,** 857–63 (2005).
77. Behar, O., Golden, J. A., Mashimo, H., Schoen, F. J. & Fishman, M. C. Semaphorin III is needed for normal patterning and growth of nerves, bones and heart. *Nature* **383,** 525–8 (1996).
78. Stoeckli, E. T. *et al.* Identification of proteins secreted from axons of embryonic dorsal-root-ganglia neurons. *Eur. J. Biochem.* **180,** 249–258 (1989).
79. Lee, M. S. *et al.* Selection of neural differentiation-specific genes by comparing profiles of random differentiation. *Stem Cells* **24,** 1946–1955 (2006).
80. Moldrich, R. X. *et al.* Proliferation deficits and gene expression dysregulation in Down's syndrome (Ts1Cje) neural progenitor cells cultured from neurospheres. *J. Neurosci. Res.* **87,** 3143–3152 (2009).
81. Le Pecheur, M. *et al.* Oxidized SOD1 alters proteasome activities in vitro and in the cortex of SOD1 overexpressing mice. *FEBS Lett.* **579,** 3613–8 (2005).
82. Wynn, S. L. *et al.* Organization and conservation of the GART/SON/DONSON locus in mouse and human genomes. *Genomics* **68,** 57–62 (2000).
83. Southard-Smith, E. M., Kos, L. & Pavan, W. J. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat. Genet.* **18,** 60–64 (1998).
84. Bylund, M., Andersson, E., Novitch, B. G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* **6,** 1162–1168 (2003).
85. Bahn, S. *et al.* Neuronal target genes of the neuron-restrictive silencer factor in neurospheres derived from fetuses with Down's syndrome: a gene expression study. *Lancet (London, England)* **359,** 310–5 (2002).
86. Craxton, M. Genomic analysis of synaptotagmin genes. *Genomics* **77,** 43–9 (2001).
87. Han, J. *et al.* Tbx3 improves the germ-line competency of induced pluripotent stem cells. *Nature* **463,** 1096–100 (2010).
88. Paulsen, F. P. *et al.* Intestinal trefoil factor/TFF3 promotes re-epithelialization of corneal wounds. *J. Biol. Chem.* **283,** 13418–13427 (2008).
89. Mashimo, H., Wu, D. C., Podolsky, D. K. & Fishman, M. C. Impaired defense of intestinal mucosa in mice lacking intestinal trefoil factor. *Science* **274,** 262–265 (1996).
90. Böttner, M., Krieglstein, K. & Unsicker, K. The transforming growth factor-betas: structure, signaling, and roles in nervous system development and functions. *J. Neurochem.* **75,** 2227–40 (2000).
91. Krieglstein, K., Strelau, J., Schober, A., Sullivan, A. & Unsicker, K. TGF-beta and the regulation of neuron survival and death. *J. Physiol. Paris* **96,** 25–30 (2002).
92. Eib, D. W. *et al.* Expression of the follistatin/EGF-containing transmembrane protein M7365 (tomoregulin-1) during mouse development. *Mech. Dev.* **97,** 167–71 (2000).
93. Tsukahara, F., Hattori, M., Muraki, T. & Sakaki, Y. Identification and cloning of a novel cDNA belonging to tetratricopeptide repeat gene family from Down syndrome-critical region 21q22.2. *J Biochem* **120,** 820–827 (1996).
94. Mongroo, P. S. & Rustgi, A. K. The role of the miR-200 family in epithelial-mesenchymal transition. *Cancer Biology and Therapy* **10,** 219–222 (2010).
95. Sakurai, M. *et al.* Ubiquitin C-terminal hydrolase L1 regulates the morphology of neural progenitor cells and modulates their differentiation. *J. Cell Sci.* **119,** 162–171 (2006).
96. Larsson, L. T. Hirschsprung's disease--immunohistochemical findings. *Histol Histopathol* **9,** 615–629 (1994).
97. Wakamatsu, N. *et al.* Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. *Nat. Genet.* **27,** 369–70 (2001).
98. Amiel, J. *et al.* Large-scale deletions and SMADIP1 truncating mutations in syndromic Hirschsprung disease with involvement of midline structures. *Am. J. Hum. Genet.* **69,** 1370–7 (2001).
99. Van de Putte, T., Francis, A., Nelles, L., van Grunsven, L. A. & Huylebroeck, D. Neural crest-

specific removal of Zfhx1b in mouse leads to a wide range of neurocristopathies reminiscent of Mowat-Wilson syndrome. *Hum. Mol. Genet.* **16,** 1423–1436 (2007).
100. Nagai, T. *et al.* The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation. *Dev. Biol.* **182,** 299–313 (1997).
101. Sánchez-Camacho, C. & Bovolenta, P. Autonomous and non-autonomous Shh signalling mediate the in vivo growth and guidance of mouse retinal ganglion cell axons. *Development* **135,** 3531–41 (2008).
102. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46,** 944–950 (2014).
103. Li, M. X. *et al.* Predicting Mendelian Disease-Causing Non-Synonymous Single Nucleotide Variants in Exome Sequencing Studies. *PLoS Genet.* **9,** (2013).



CHAPTER 3

Oligogenic inheritance in Hirschsprung disease: implications of RET and Hedgehog signaling in ENS development

Y. Sribudiani[1*#], R.K.Chauhan[1*], L. Petrova[2], C. Harrison[2], T.D. Wabbersen[2], E. Brosens[1], B.M. de Graaf[1], G.Burzynski[2], R.W.W. Brouwer[3], W. F.J. van IJcken[3], A. de Klein[1], J. Osinga[4], B.J.L. Eggen[5], M.M. Alves[1], A.S. Brooks[1], I.T. Shepherd[2], R.M.W. Hofstra[1, 6]

[1] Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands
[2] Department of Biology, Emory University, Atlanta, USA
[3] Erasmus Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands
[4] Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[5] Department of Neuroscience, section Medical Physiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[6] Stem Cells and Regenerative Medicine, UCL Great Ormond Street Institute of Child Health, London, UK

*These authors contributed equally to this work
[#]Current address: Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Universitas Padjadjaran, Bandung, Indonesia

Corresponding author: Robert MW Hofstra, Department of Clinical Genetics, Erasmus University Medical Center, PO BOX 2040, 3000CA Rotterdam, The Netherlands. Tel: +31-10-7037643. E-mail: r.hofstra@erasmusmc.nl

Manuscript in preparation

ABSTRACT

Hirschsprung disease (HSCR) is a congenital disorder characterized by the absence of enteric ganglia in a variable length of the intestinal tract. A linkage study previously performed on a large Dutch multi-generational HSCR family revealed linkage to 4q31.3-q32.3. To determine the genetic cause of HSCR in this family, we performed exome sequencing and variant prioritization. We identified one main candidate in the linkage interval in exon 20 of *LRBA.* Although *in silico* prediction suggested an effect on mRNA splicing, but functional assays did not confirm this. A role as an enhancer mutation for *MAB21L2,* a gene embedded within an intron of *LRBA,* was also not confirmed. However, we show that *MAB21L2* is important for enteric neural crest cells (NCC) differentiation during enteric nervous system (ENS) development in a zebrafish model. The incomplete penetrance of the HSCR in the family suggests the involvement of other (rare) variants elsewhere in the genome. Therefore, we searched for variants present especially in known HSCR genes and genes associated with ENS development. We identified several candidates, among which a rare *RET* coding variant in one branch of the family and inherited *IHH*, *GLI3* variants along with a *de novo* mutation in *GDNF* (RET ligand) in the second branch of the family. Functional studies confirmed the pathogenic nature of the variants identified in *RET* and *IHH,* confirming the importance of RET and Hedgehog signaling for ENS development. This study demonstrates that rare variants in multiple genes, lead to the development of HSCR, further illustrating the complexity of HSCR genetics.

INTRODUCTION

Hirschsprung disease (HSCR) is a congenital disorder characterized by the absence of enteric ganglia in the submucosal and myenteric plexuses of the intestinal tract along a variable length of the distal gut. This aganglionosis leads to intestinal obstruction by dysregulated muscle relaxation, HSCR results from a failure of enteric neural crest cells (NCC) to migrate, differentiate, proliferate or survive and thereby colonize the gut to form a functional network of neurons and glia, called the enteric nervous system (ENS)[1].HSCR mainly occurs as a sporadic disorder and in most cases only a short segment of the terminal bowel is affected (termed short segment HSCR)[2]. Based on the familial occurrence, chromosomal abnormalities and on the presence of a many naturally occurring animal models with an aganglionic colon, HSCR is considered to be an inherited disease. The mode of inheritance can be dominant with reduced penetrance, mostly found in non-syndromic familial HSCR cases, whereas in families with syndromic HSCR, a recessive pattern of inheritance is often observed[3].

Numerous studies have been performed to find genes involved in HSCR development. To date mutations in 16 genes have been identified that can cause, or contribute to, the development of HSCR[4,5]. Of these genes, *RET* is considered to be the major HSCR gene, as 50% of the familial cases and 15-35% of the sporadic cases have a mutation in the *RET* coding region or affecting mRNA splicing regions[6]. Most of the other genes have been identified in rare (familial) syndromic HSCR cases. The majority of sporadic cases are suspected to be oligogenic or polygenic in nature and have not been resolved genetically. However, a low penetrant *RET* variant is often present in intron 1 in the majority of (sporadic) patients (OR=2 when present heterozygous and OR=20 when present homozygous)[7]. Combinations of distinct rare mutations resulting in the disease have yet hardly been reported. However, sib pair-analysis clearly points towards an oligogenic inheritance of HSCR, with the involvement of at least three loci[8] for which, besides *RET*, no other gene has been identified as yet.

In a previous linkage study, we identified a 12.2 Mb linkage interval on 4q31.3-q32.3 (chr4: 154,674,450-167,058,075 (Hg19) in a multi-generational, Dutch HSCR family with five affected members[9]. The pattern of inheritance in this family appeared to be autosomal dominant with reduced penetrance, indicating that a mutation in this region would be necessary but not sufficient to cause the disease. Therefore, it is most likely that in addition to the mutated gene on chromosome 4 (Chr4), variants in genes located elsewhere in the genome must also contribute to the development of the disease as well. In the linkage interval, 57 genes were present. Among these genes, *MAB21L2* was considered to be the

most promising candidate HSCR gene based on its expression in the central nervous system and neural crest in mid gestation embryogenesis in mice[10]. Moreover, *MAB21L2* is also linked (downstream target) to the TGF-β signaling pathway in which another known syndromic HSCR gene *ZEB2,* has been previously identified[11,12]. Mutational screening of *MAB21L2* was performed in this family, but no mutation was found in its coding region[9].

In an attempt to identify the causal genes and disease-associated variants leading to HSCR in this family, we have now exome sequenced two affected family members from different branches of the family (Fig.1: V-1 and V-4) and subsequently, we used whole exome sequencing for the trio containing patient V-4 to identify *de novo* mutations (Figure 1A: shown in enclosed dotted box). We determined the segregation patterns for the candidate variants identified and functionally studied the effects of several of them to reveal the complex genetics of HSCR.

MATERIALS AND METHODS

Patients and DNA samples

The multigenerational Dutch family included in this study is composed of five individuals diagnosed with HSCR and two diagnosed with functional constipation (Figure 1A). A detailed description of the phenotypes has been previously reported[9]. Genomic DNA was isolated from peripheral blood leucocytes using a standard protocol[13].

Exome sequencing; samples and variant prioritization

For exome sequencing, we initially selected two HSCR affected individuals (V-1 and V-4) from different branches of the family. In a later stage of the study we also selected a trio consisting of IV-4, IV-5 and V-4 (Figure 1A) Three micrograms of DNA from each of the individuals was sheared using acoustic technology (Covaris, Inc. Woburn, Massachusetts, USA). Target enrichment for V-1 and V-4 was performed with the SureSelect Human All Exon 50 Mb Targeted exome enrichment kit v4 and for the trio (IV-4, IV-5 and V-4) the Agilent Sureselect CRE capture kit (Agilent Technologies, Inc., Santa Clara, California). The captured fragments were sequenced (paired-end 101 bp read length) on the Illumina HiSeq2000 sequencer (Illumina, San Diego, USA). De-multiplexing, alignment to the human genome build 19 (Hg19) reference genome using the Burrows-Wheeler Aligner version 0.6.2[14] and curation of low quality reads were done as described by our in-house developed NARWHAL pipeline[15]. Chromosome sorted

BAM-files were generated with SAMtools version 0.1.12a[16]. Variant calling with the Bayesian genotyper incorporated in the genome analysis toolkit version 1.2.9[17] resulted in variant files of VCFv4 format, which were uploaded into Cartagenia Bench NGS version 3.04 (Cartagenia Inc, Boston, MA, USA) for variant filtering. Four types of analysis were performed. One was specifically focused on shared variants located inside the linkage interval, with another aimed at identifying shared rare variants outside the linkage interval. Other was aimed at identifying variants in known HSCR genes and also genes known to be involved in ENS development. The latter was done using a set of candidate HSCR genes, HSCR associated linkage intervals, genes identified via genome wide association studies (GWAS), and gene expression profiling studies and previously reported animal models. The fourth and final analysis consisted of *de novo* mutation identification in a trio by filtering out the parental variants.

Variants were selected with a read depth $\geq$ 6 for the analysis. We focused on nonsense, missense, InDels or synonymous variants located near exon-intron borders. (3 bases in the exon and 3 bases in the intron) Common variants (minor allele frequency ≥ 0.03) were excluded using an in-house cohort of 300 unaffected individuals, ESP6500; 1000 genomes; and GoNL (http://www.nlgenome.nl/) databases.

Validation of candidate variants and family screening

Candidate variants were validated by Sanger sequencing as previously described[18]. Segregation analysis was also performed in the family members for which DNA was available (II-2, III-2, IV-1 until IV-5, V-1, V-2, V-3 and V-4).

Vectors design and site direct mutagenesis (SDM)

Approximately 400 bp from the genomic region of LRBA, containing exon 20 and its flanking sequence, was amplified from control and patient DNA to get both the wild type (WT) and mutant (Mut) (NM_001199282.2:c.2444A>G) alleles, named LRBA-Enh-WT and LRBA-Enh-Mut, respectively. The PCR products obtained were inserted into the pCR™2.1-TOPO® vector, subsequently digested with XhoI and KpnI, and cloned into a pGL3-SV40 promoter (SV40-P), upstream of the luciferase gene (Promega, Madison, USA). The SV40-P and pGL3-SV40p-Luc-LRBA-Enh-WT/Mut (LRBA-WT/ LRBA-Mut) vectors were used for Luciferase reporter assays. The same LRBA products were also directly cloned into the exon trapping vector pSPL3 (Invitrogen), and named pSPL3-LRBA-WT/Mut. The pCMV-RET-WT vector, encoding the short isoform of human RET (RET9), was used to create the pCMV-RET-Mut (P399L) by site-directed mutagenesis, according to the manufacturer's instructions (Stratagene, La jolla, USA). The pCMV-IHH-FLAG-WT

vector was used to create the pCMV-IHH-FLAG-Mut (Q51K) by site-directed mutagenesis, according to the manufacturer's instructions (Primers details in Supplementary Table 1)[19]. All inserts were Sanger sequenced to confirm the presence of the WT and Mut variant, and the orientation of the inserted fragments.

Cloning of zebrafish *lrba, mab21l2* and *ihh* and whole mount *in situ* hybridization expression analysis

Zebrafish *lrba mab21l2* and *ihh* genes were amplified by RT-PCR using a One-Step RT-PCR Kit (Qiagen) from 48 hpf total mRNA using the following primers: *lrba* F-CTTTTGACCAAAGGAATGGGTTACG, R-TCCAAGCATGACTTCTGCTTTCC; *mab21l2* F- ATTCGCTCCCGCTTTCAG, R-TCGTCCCAGTCAGTCTCCC; *ihh*, F-GAATTTTACGCACGGACGAT R-CGTAATGCAGCGAATCTTCA. Amplified bands were gel purified and subcloned into TOPO TA PCRII vector (Thermo Fisher). Digoxigenin labeled antisense probes (Roche) were generated using SP6 polymerases (Roche) after linearizing the plasmid templates using *Not1* restriction enzymes (New England Biochemicals). Embryos were collected and processed for whole-mount in situ hybridization as previously described[20]**.** Digoxigenin-labeled probes were visualized with NBT/BCIP coloration reactions.

Zebrafish *mab21l2 lrba* and *ihh* Morphant Analysis

Two *lrba* splice blocking morpholinos (SBMOs) were designed, one to exon 13 (AGTTGGTTTAGTCTCTTACCGAGAC) and the other to exon 24 (ACTGCATACTAACCGAAGAAGAAGT). A previously described translation blocking morpholino (TBMO) for mab21l2 (ACTGTAGACCGGAGTTTCGCAGTAC) was obtained from Genetools LLC[21]. An *ihh* morpholino was designed to target the transcription start site as previously described[22]. The sequence of the *ihh* morpholino is: GGAGACGCATTCCACCGCAAGCG. The effectiveness of these *lrba* SBMOs were confirmed by RT-PCR. Morphants were generated by injecting 100μM morpholinos into one-cell embryos. Morphants and controls embryo were allowed to develop to 120hpf and were then fixed and antibody stained for ENS neurons using the HuC/D antibody (Invitrogen) as previously described[23]. Control embryos were injected with scrambled morpholinos and experimental and control embryos were also co-injected with a *p53* control morpholino (Gene Tools) designed to suppress apoptotic effects induced by morpholinos as previously described[24].

Cell culture and transfections

Human embryonic kidney (HEK293) cells, COS-7 cells and control fibroblasts were cultured in DMEM (GIBCO) containing 10% fetal bovine serum (GIBCO) and penicillin/streptomycin (GIBCO). The neuroblastoma cell line (Neuro-2a) (ATCC # CCL-131) was cultured according to the ATCC's protocol (LGC Standards, Middlesex, UK). All the cell lines were incubated at 37°C, and supplied with 5% of CO_2. Approximately 10^6 cells were cultured in 1 well of a 6-wells plate for 24 hours prior to transient transfections for the cell lines and approximately 300,000 control human fibroblast cells were plated prior to transfections. All the cells were transfected using 3µl genejuice transfection reagent (Novagen, 70967, Millipore) according to the manufacturer's instructions.

Exon trapping assay for functional analysis of a possible splice site mutation

The *LRBA* variant has been predicted to affect the mRNA splicing. Therefore, the effect of this variant was analysed using an exon trap assay as described earlier[25]. Shortly, COS-7 cells were transfected with 1 µg pSPL3-*LRBA*-WT/Mut. The transfected cells were incubated for 48 hours and total RNA was isolated from the transfected cells using the RNA easy Mini Kit (Qiagen) and cDNA was synthesized using the iScript™ cDNA Synthesis Kit (Bio-Rad) using equal amounts of RNA from all conditions. RT-PCR was performed for 30 cycles using SA2 and SD6 primers (Supplementary Table 2) that flanked the vector spanning internal exons. The RT-PCR products were run on a 1.5 % agarose gel and stained with gel red dye.

Luciferase assays to determine enhancer activity of the LRBA variant

To establish/determine whether the variant in *LRBA* exon 20 interferes with transcription regulation (of *MAB21L2*) we determined whether the mutation containing fragment has any enhancer activity and whether the mutation has any influence on this possible enhancer activity. Cells were transfected with 1 µg of SV40-P or LRBA-WT/Mut (sense) and co-transfected with 10 ng of internal control pRL-SV40-Renilla Luciferase (Promega, Madison, USA). The Luciferase activity was measured using the Dual-Luciferase-Reporter Assay (Promega), 48 hours after transfection on a glomax platform (Promega, Madison, USA). The ratio of Firefly Luciferase compared to Renilla Luciferase (normalized data) was used to determine the activity of each enhancer element LRBA-Enh-WT/Mutant (sense). The results were presented as a fold change of normalized data of each enhancer to the promoter only construct (SV40-P). As a negative control, we used SV40-E (without any promoter) and RET-WT-enhancer was used as a positive

control as published earlier[26]. The luciferase assays were performed in three independent, triplicate experiments (n=9).

GDNF activation and Western blot

pCMV-*RET*-WT and pCMV-*RET*-Mut were transiently co-transfected in combination with pCMV-GFRα1 and pNE-GFP into HEK293 cells. After 24 hours of culturing, cells were treated with 50ng/ml of GDNF (Prepotech EC,London, UK) for 15 minutes and they were lysed using a lysis buffer containing 150mM NaCl, 20mM Tris-HCl pH 7.4, 1% Triton X-100, protease inhibitors (Roche) and phosphatase inhibitors (Thermo Scientific, Waltman, MA, USA). Cell lysates were collected after centrifugation at 10000 rpm, at 4ºC for 10 minutes. Protein concentration was measured using the BCA kit (Pierce Biotechnology, Rockford, USA) according to the manufacturer's protocol. Forty micrograms of total protein was run on mini-Protean TGX 4-15 % (Bio-Rad) for separation and transferred onto a nitrocellulose membrane (GE). Five percent of skimmed milk in phosphate buffered saline (PBS) containing 0.1% Tween-20 (PBST) was used to block the nitrocellulose membrane. Western blot was performed using the primary antibodies and secondary antibodies described in Supplementary Table 3. The primary antibodies were applied overnight at 4ºC and respective secondary antibodies were applied for 1 hour at room temperature and the nitrocellulose membrane was washed 3X for 5 minutes with 1XPBST. The membrane was scanned with the Odyssey™ infrared Imager (Li-COR Biosciences).

Indian Hedgehog (IHH) WT/Mutant treatment and quantification of GLI1 expression

To determine whether the *IHH* variant identified resulted in a less functional IHH protein, we analysed a downstream target of IHH signaling, Gli1, after exposing fibroblasts with WT and mutant IHH proteins. HEK293 cells cultured in a 6 well plate were transiently transfected with pCMV-IHH-FLAG-WT/Mut. After 24 hours, the medium of transfected cells (conditioned medium) was collected and filtered using a 0.45 µm filter. Two - three hundred thousand control human fibroblast cells were cultured in a 6 well plate for 24 hours and the medium was replaced by 1 ml fresh complete medium and 500 µl of conditioned medium (containing secreted IHH-WT or IHH-Mut). Conditioned medium from non-transfected HEK293 cells was used as a negative control. Medium supplemented with 20 µM of Purmophamine (Calbiochem), an agonist of SMO and an intermediate in IHH-GLI signaling pathway, was used as a positive control for activation of Hedgehog (Hh) signaling. Five hundred microliters of conditioned medium was concentrated using an M-10 filter (Millipore) and used for Western

blot to determine the level of IHH-WT and IHH-Mut protein secretion into the medium. To quantify *GLI1* expression, quantitative real-time (qRT) Sybr Green PCR was performed using the 7300 Real time PCR platform system (Applied Biosystem). Approximately 20-25 ng of cDNA from fibroblasts treated with IHH-WT and IHH-Mut conditioned medium, and with Purmophamine was used as a template for qRT-PCR. *GLI1* expression was determined and *CLK2* housekeeping gene was used to normalize the *GLI1* expression (Primer details in Supplementary Table 4). qRT-PCR data were analysed using the method described by Livak[27], and presented as fold change after comparison with the normalized *GLI1* expression in fibroblasts treated with non-transfected HEK293 conditioned medium. The experiments were performed in three independent triplicate experiments (n = 9).

Statistical analysis

All results are expressed as the mean ± standard deviation (SD) or SEM. All data were analyzed using a 2-tailed Student's *t* test or the χ^2 test. $P < 0.05$ was considered statistically significant.

RESULTS

Exome sequencing

HSCR patient V-1, V-4, IV-4 and IV-5 (as shown in Figure 1 A) from previously identified family were exome sequenced. Target coverage statistics of the exome sequencing data of patient V-1, V-4, IV-4 and IV-5 are depicted in Supplementary Table 5.

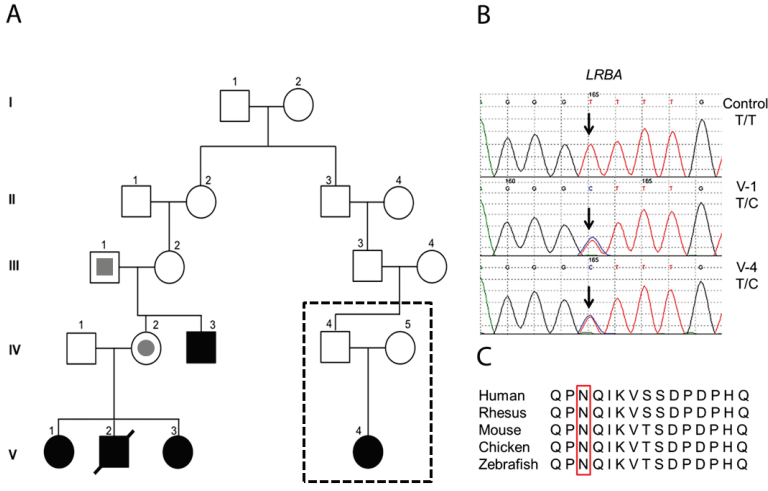


Figure 1. (A) Pedigree of a multigenerational Dutch family with HSCR. Patients affected with HSCR are represented as black symbols and two members with functional constipation are marked in grey. The exome sequenced trio for de novo mutation identification is enclosed in dotted box. (B) Chromatogram showing the *LRBA* variant in exon 20 of two affected individuals (V-1 and V-4) and in one control (C) Evolutionary conservation of the *LRBA* variant (p.N815) across vertebrates (enclosed in red box).

LRBA and *TMEM144* variants in the linkage interval

We first looked for variants present in the Chr4 linkage interval as reported earlier[9]. We found two variants in this linkage interval which were located in *LRBA* (NM_001199282.2:c.2444A>G) and *TMEM144* (NM_018342.4:c.715A>T) (Table 1). *TMEM144* is not expressed in mouse gut at E14.5 (data not shown). On the contrary we could detect expression of *LRBA* in the mouse gut and in ENS progenitors at E14.5 (data not shown).

Table 1. List of rare variants identified and shared between two HSCR patients V-1 and V-4

Gene	HGVS cDNA-level	Exon	ExAC MAF	GoNl MAF	dbSNP	db SNP buil d	≥ 2 EMC
SMPD4	NM_017951.4:c.1450T>G	15	0.01161	-	rs148027738	136	yes
NRP2	NM_201266.1:c.1000C>T	7	0.001574	0.003	rs114144673	134	no
ARMC9	NM_001271466.2:c.1645C>T	17	0.01037	0.021	rs148296188	134	yes
CCRL2	NM_003965.4:c.11A>G	2	0.02773	0.024	rs11574443	120	yes
MTTP	NM_000253.2:c.502G>A	6	0.01018	0.004	rs61750974	129	yes
PGRMC2	NM_006320.4:c.185G>A	1	0.000347 5	-	rs41298555	127	no
LRBA	**NM_001199282.2:c.2444A>G**	**20**	**0.00253 4**	**0.009**	**rs14066684 8**	**134**	**no**
TMEM144	**NM_018342.4:c.715A>T**	**10**	**0.00511 7**	**0.014**	**rs62335898**	**129**	**no**
ADAMTS2	NM_014244.4:c.1993G>A	13	0.007849	0.011	rs35372714	126	yes
FNDC1	NM_032532.2:c.4429A>G	14	-	-	-	-	yes
FNDC1	NM_032532.2:c.4436C>G	14	-	-	rs398066440	138	yes
PTPRD	NM_002839.3:c.2341A>G	26	0.02691	0.028	rs72694737	130	yes
ZNF518A	NM_001278524.1:c.1477G>C	7	0.002935	0.007	rs41291602	127	yes
ATM	NM_000051.3:c.3161C>G	22	0.01692	0.026	rs1800057	89	yes
KDELC2	NM_153705.4:c.431C>T	3	0.01606	0.025	rs74911261	131	yes
OR6T1	NM_001005187.1:c.107T>C	1	0.01039	0.010	rs140244798	134	yes
OR8D1	NM_001002917.1:c.304T>G	1	0.01701	0.019	rs2510433	100	yes
OR1F1	NM_012360.1:c.47G>A	1	0.002605	0.008	rs142486394	134	yes
CLUH	NM_015229.3:c.3547G>C	24	0.000164 3	-	rs201361018	137	no

PELP1	NM_014389.2:c.2696T>C	16	0.007134	0.009	rs199636910	137	yes
PELP1	NM_014389.2:c.2161A>G	16	0.001568	0.006	rs200062536	137	no
NADK	NM_001198994.1:c.1769_1771dupAGG	14	-		rs150880809	134	yes

Sanger sequencing validation of the identified *LRBA* variant (NM_001199282.2:c.2444A>G) shared by V-1 and V-4 is shown in Figure 1 (B). The identified LRBA variant is evolutionary conserved in other vertebrates as shown in Figure 1 (C). The *in silico* prediction of rare variants identified and shared by both the individual are depicted in Table 2.

Table 2. *In silico* prediction of rare variants shared in the HSCR patients V-1 and V-4

Gene	PHAST	GERP++ neutral rate	GERP++ RS	PhyloP	SiPhy	Mutation Taster	SIFT	PolyPhen2 HumVar	LRT	Mutation Assessor	FATHMM	BLOSUM62
SMPD4	-	4.24	4.24	1.538	11.314	DC	0.19	PoD	D	M		-1
NRP2	0.9	5.91	5.03	1.505	15.056	DC	0.02	ProD	D	M	-4.81	-3
ARMC9	-	-	-	-	-	-	-	-	-	-	-	-3
CCRL2	0	5.4	-0.048	0.042	4.516	P	0.07	B	N	L	0.15	-2
MTTP	1	5.9	5.07	1.519	13.015	DC	-	-	D	M	0.99	3
PGRMC2	-	3.81	1.87	0.927	7.764	DC	1	B	N	N	-1.13	-2
LRBA	1	5.66	4.47	0.96	12.981	DC	0.01	PoD	D	M	-0.12	1
TMEM144	1	5.35	5.35	2.027	14.329	DC	0.05	ProD	D	M	0.61	0
ADAMTS2	0.9	5.37	4.5	1.269	13.289	DC	0.06	B	N	L	-0.49	-2
FNDC1	-	5.32	1.53	0.022	6.352	DC	1	B	N	L	2.91	0
FNDC1	-	5.32	5.32	2.476	18.601	P	0.48	PoD	N	L	2.93	1
PTPRD	1	5.95	4.83	2.281	9.61	DC	-	B	D	N	0.28	0
ZNF518A	-	-	-	-	-	-	-	-	-	-	-	-2
ATM	1	5.63	5.63	2.798	20.05	DC	0	ProD	D	M	-0.56	-2
KDELC2	0.9	4.68	4.68	2.871	18.902	DC	0	ProD	D	M	1.57	-3
OR6T1	0	4.26	1.86	-0.202	8.753	P	0.21	B		N	4.02	-1
OR8D1	1	4.29	4.29	1.813	8.849	DC	0.01	PoD	U	M	7.56	-1
OR1F1	1	4.97	4.97	2.456	16.064	DC	0.01	ProD	D	H	5.95	-2
CLUH	1	5.07	3.02	1.248	6.899	DC	0.08	B	N	N	-1.58	-1
PELP1	0	4.42	0.488	0.187	0.625	P	0.64	B	N	L	0.88	0

PELP1	0.9	5.56	0.466	0.051	3.9	P	1	B	N	N	0.92	1
NADK	-	-	-	-	-	-	-	-	-	-	-	-

Build hg19, ProD= probably damaging, PoD= possibly damaging, D= deleterious, N = neutral, P = polymorphism, DC = disease causing, B = Benign, U = Unknown, N = Neutral, L = Low, M = medium, H = High

Moreover, we determined which exons were not totally covered within the linkage region and sequenced those (17 regions) by regular Sanger sequencing. We did not find any rare variant that could be linked to the disease phenotype (data not shown).

***RET* and *IHH* variants in V-1 and V-4 exomes.**

As we hypothesized that the variant in the linkage interval is not enough to cause the disease, we subsequently focussed on non-shared rare variants outside the linkage region present in any of the two individuals in genes of the HSCR gene target panel. Of these (Table 3) we prioritized variants based on their function and deleteriousness (Table 4) for further evaluation. We identified a rare *RET* variant (NM_020975.4:c.1196C>T) in patient V-1 and four rare variants in patient V-4. These were in the Indian hedgehog (*IHH*) gene (NM_002181.3:c.151C>A), Neuron Navigator 2 (*NAV2)* (NM_001244963.1:c.2569C>T), Arginine Vasopressin Receptor 2 (*AVPR2)* (NM_000054.4:c.1110_1112delATC) and GLI family zinc finger 3 (*GLI3)* (NM_000168.5:c.2119C>T). No allelic frequencies of these variants were found in ExAC.

Table 3. Exonic variants present in the 'HSCR/ENS gene panel' identified in the two HSCR patients (V-1 and V-4).

Individual	Gene	HGVS cDNA-level	Exon	Exac MAF	GoNL MAF	dbSNP	dbSNP build	≥ 2 EMC
V-4	*NOTCH 2*	NM_024408.3:c.7223T>A	34	0.001788	0.004	rs35586704	126	No
V-4	*TGFB2*	NM_001135599.2:c.272G>A	1	0.005275	0.006	rs10482721	119	No
V-4	*IHH*	NM_002181.3:c.151C>A	1	-	-			No
V-4	*GLI3*	NM_000168.5:c.2119C>T	14	0.0001977	0.001	rs121917716	133	No
V-4	*FKTN*	NM_001079802.1:c.1336A>G	11	0.01189	0.010	rs41313301	127	Yes
V-4	*NAV2*	NM_001244963.1:c.2569C>T	11	0.0001568	-	rs144875196	134	No
V-4	*BBS10*	NM_024685.3:c.424G>A	2	0.00837	0.019	rs142863601	134	Yes
V-4	*AMH*	NM_000479.3:c.1556C>T	5	0.002029	-	rs200031151	137	No
V-4	*AVPR2*	NM_000054.4:c.1110_1112delATC	3		-			No
V-1	*PCDHA 4*	NM_018907.3:c.920A>G	1	0.01291	0.027	rs145409201	134	Yes
V-1	*PCDHA 8*	NM_018911.2:c.337C>T	1	0.006259	0.014	rs146047089	134	Yes
V-1	*PCDHA #*	NM_018901.3:c.2414G>A	2	0.001467	0.004	rs150254638	134	Yes
V-1	*FKTN*	NM_001079802.1:c.166C>T	5	0.02353	0.027	rs41277797	127	No

V-1	*SVEP1*##	NM_153366.3:c.7244G>A	38	0.001953		rs192347509	135	No
V-1	*SVEP1*	NM_153366.3:c.1849A>G	9	0.01806	0.019	rs74597491	132	Yes
V-1	*NOTCH1*	NM_017617.3:c.6853G>A	34	0.02541	0.006	rs61751489	129	Yes
V-1	*RET*	NM_020975.4:c.1196C>T	6	-	-			No

Build hg19, #PCDHA 1-13 and AC1-2, ## not replicated with Sanger sequencing

De novo mutations in patient V-4

De novo mutations in patient V-4 were identified in Glial cell-derived neurotrophic factor (*GDNF*), suppression of tumorigenicity 18, zinc finger (*ST18*) and alstrom syndrome protein 1 (*ALMS1*), respectively. No allelic frequencies of these variants were found in ExAC. An inframe insertion of 3 bp was found in *ALMS1* (NM_015120.4:c.72_74dupGGA). The *de novo* variant of *ST18* (NM_014682.2:c.1009G>A) in the exon10 results is a missense mutation. In the *GDNF* gene, a heterozygous inframe insertion of 6bp just before the last codon of exon 3 was found.

Table 4. *In silico* prediction of the rare variants identified in the two HSCR patients (V-1 and V-4) present in the 'HSCR/ENS gene panel'.

Individual	Gene	HGVS	PHAST	GERP++ neutral rate	GERP++ RS	PhyloP	SiPhy	Mutation Taster	SIFT	PolyPhen2 HumVar	LRT	Mutation Assessor	FATHMM
V-4	*NOTCH2*	NM_024408.3:c.7223T>A	1	5.35	5.35	2.027	14.507	DC	0	PrD	U	L	-1.1
V-4	*TGFB2*	NM_001135599.2:c.272G>A	1	5.45	5.45	2.837	13.71	DC	0.07	PoD	D	L	-0.11
V-4	*IHH*	NM_002181.3:c.151C>A	-	4.22	4.22	2.18	12.671	DC	0	PrD	D	H	-6.03
V-4	*GLI3*	NM_000168.5:c.2119C>T	1	5.82	4.94	1.468	14.65	DCA	0.01	PrD	N	M	2.18
V-4	*FKTN*	NM_001079802.1:c.1336A>G	1	6.04	6.04	2.317	15.77	DC	0.02	PrD	D	M	0.87
V-4	*NAV2*	NM_001244963.1:c.2569C>T	0.9	5.02	0.954	0.298	14.955	DC		PrD	D	M	1.2
V-4	*BBS10*	NM_024685.3:c.424G>A	0.9	5.34	5.34	2.937	16.922	DC	0.05	PoD	D	L	-2.46
V-4	*AMH*	NM_000479.3:c.1556C>T	0	3.88	1.72	0.639	8.128	P	0.14	B	U	N	-1.81

V-4	*AVPR2*	NM_000054.4:c.1110_1112delATC	-	-	-	-	-	-	-	-	-	-	-
V-1	*PCDHA4*	NM_018907.3:c.920A>G	-	4.34	0	0.15	2.273	P	0.04	B	U	L	0.71
V-1	*PCDHA8*	NM_018911.2:c.337C>T	1	3.72	3.72	1.794	15.924	DC	0.02	PoD	U	M	0.58
V-1	*PCDHA #*	NM_018901.3:c.2414G>A	-	5.31	5.31	2.478	17.155	DC	0	PrD	N	M	2.17
V-1	*FKTN*	NM_001079802.1:c.166C>T	1	5.65	5.65	2.668	13.651	DC		PrD	D	L	-2.83
V-1	*SVEP1*	NM_153366.3:c.7244G>A	1	5.8	0.873	-0.091	5.678	DC	0.92	B	N	N	-0.05
V-1	*SVEP1*	NM_153366.3:c.1849A>G	1	4.85	4.85	1.822	14.413	DC	0.03	-	D	L	2.21
V-1	*NOTCH1*	NM_017617.3:c.6853G>A	0	5.55	2.74	0.405	8.604	P	0.4	B	N	N	-1.49
V-1	*RET*	NM_020975.4:c.1196C>T	-	5.13	4.22	1.151	10.524	DC	0	PoD	D	M	-3.02

Build hg19, #PCDHA 1-13 and AC1-2, PrD= probably damaging, PoD= possibly damaging, D= deleterious, N = neutral, P = polymorphism, DC = disease causing, , DCA = Disease causing automatic, B = Benign, U = Unknown, N = Neutral, L = Low, M = medium, H = High

Validation of candidate variants, segregation analysis and *MAB21L2* sequencing

We validated the candidate variants by Sanger sequencing in all the family members for whom the DNA was available (n=11). Details of the primer pair used for amplification and sequencing are provided in supplementary Table 6. Segregation analysis of the entire validated DNA variants is shown in Table 5. *MAB21L2* is embedded within intron 42 of LRBA gene (Figure 2).

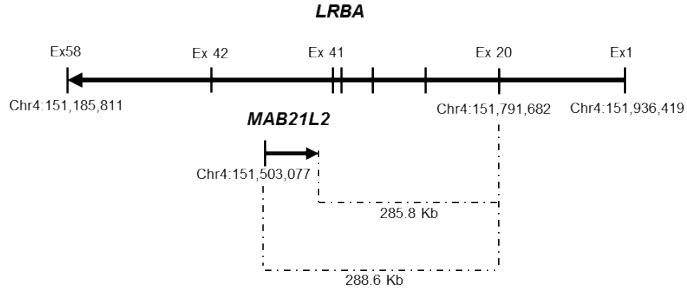


Figure 2. Schematic overview of the genomic region of *LRBA* and *MAB21L2* as a nested pair (located in intron 42 of *LRBA*) and their respective positions in the human genome (hg19). The variant in exon 20 of *LRBA* is located 288.6 Kb away from the start site of the *MAB21L2* gene.

Subsequently we sequenced 16 Kb upstream of *MAB21L2* assuming that a mutation in the regulatory region may also have an impact on the phenotype, but we did not find any rare variants that could be linked to the disease phenotype.

Table 5. Segregation analysis of candidate variants linked to HSCR identified by exome sequencing in the family members

Gene	II-2	III-2	IV-1	IV-2	IV-3	IV-4	IV-5	V-1	V-2	V-3	V-4
LRBA (c.2444A>G)	+	+	-	+	+	+	-	+	+	+	+
RET (c.1196C>T)	-	-	-	+	+	-	-	+	+	+	-
IHH (c.151C>A)	-	-	-	-	-	+	-	-	-	-	+
NAV2 (c.2569C>T)	N.A	N.A	-	-	N.A	+	-	-	N.A	N.A	+
AVPR2 (c.1110_1112delATC)	N.A	N.A	-	-	N.A	-	+	-	N.A	N.A	+
TMEM144 (c.715A>T)	+	+	-	+	+	+	+	+	+	+	+/+
GLI3 (c.2119C>T)	N.A	N.A	-	-	N.A	+	-	-	N.A	N.A	+
NRP2 (c.1000C>T)	+	+	-	+	-	+	-	+	-	-	+

+: Present, +/+ homozygous, Absent, N.A Not applied

The variant in *LRBA* does not affect splicing

The rare variant identified in exon 20 of *LRBA* (NM_001199282.2:c.2444A>G) has been predicted to affect mRNA splicing by one of the five splice site prediction programs included in the Alamut splicing prediction module, (http://www.interactive-biosoftware.com/alamut-visual/). The *in-vitro* splicing assay did not identify any splice defect caused by the identified *LRBA* variant, as similar sized bands of spliced product were observed in both the WT and mutant situations (Figure 3).

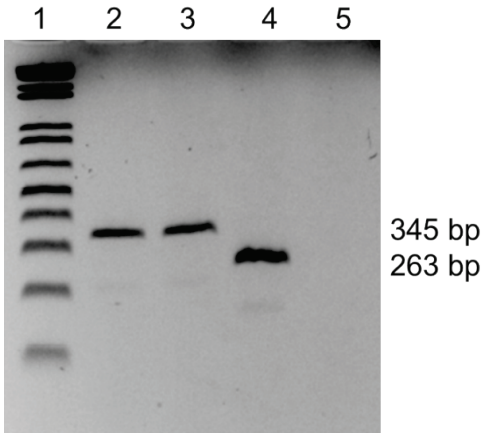


Figure 3. *In vitro* splice assay using the exon trapping method. Gel electrophoresis of cDNA-PCR products generated from wild type and mutant constructs (LRBA-WT/Mut) after transfection into COS-7 cells. Lane 1: DNA marker, 1Kb+ (Invitrogen); Lane 2: Splice product of exon 20 wild-type construct; Lane 3: Splice product of exon 20 mutant construct; Lane 4: Splice product of pSPL3; Lane 5: Un-transfected control. In the absence of an insert a splice product of 263 bp is produced. A splice product of 345 bp is produced in case of a construct containing wild type LRBA sequence or the mutant version. No difference was observed between the wild type and mutant sequence on the size of the splicing product using this exon trapping method.

The variant in *LRBA* does not disturb enhancer activity

The *MAB21L2* gene is located within the *LRBA* gene, specifically in intron 42 of *LRBA* (Figure 2). Interestingly, several enhancer elements for *MAB21L2* reside within the *LRBA* gene, and are known to drive expression of *MAB21L2* in a tissue specific manner[28]. We hypothesized that exon 20 of *LRBA* might be such an

enhancer for *MAB21L2*, and that the variant identified in this family enhances or diminishes the function of this prospective enhancer element.

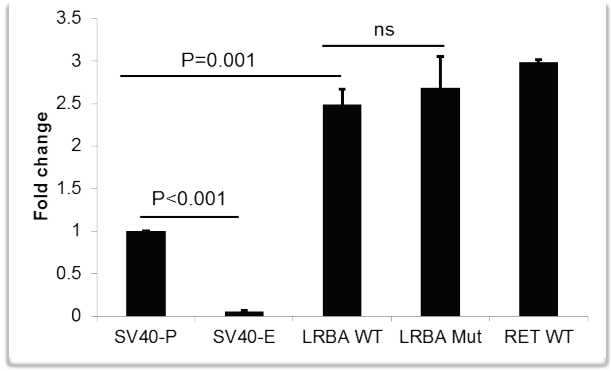


Figure 4. Enhancer activity effect of the LRBA exon 20 variant (c.2444A>G) and its flanking sequence. Exon 20 of LRBA proved to have enhancer activity when coupled to a SV40 promoter (SV40-P). However, no difference in luciferase activity could be detected between LRBA WT and LRBA Mut (c.2444A>G) constructs. SV40-E was used as a negative control. , A *RET* intronic enhancer element (RET WT) was used as a positive control for the luciferase enhancer activity.

To test this hypothesis we performed a series of luciferase assays using exon 20 of LRBA and its flanking regions containing the WT or the mutant sequence (c.2444A>G). Our results showed that exon 20 of *LRBA* with its flanking region could enhance the promoter activity of SV40, as hypothesized (Figure 4). However, no difference was detected when the *LRBA* variant (c.2444A>G) was introduced (Figure 4). As a positive control for the luciferase enhancer activity, we used a *RET* intronic enhancer (*RET*- WT) previously published[26].

Knockdown of *lrba* gene did not perturb ENS development in Zebrafish

To investigate the *in vivo* function of *lrba* in ENS development we utilized the zebrafish model system. A single zebrafish orthologue for *lrba* was identified in an Ensemble gene search and showed strong sequence similarity, as well as genome organization, to its human orthologue (82% sequence identity). Whole-mount in situ hybridization (WISH) studies revealed that zebrafish *lrba* expression has a comparatively restricted expression pattern. *lrba* is expressed along the yolk sack boundary and weakly in the hindbrain at 24hpf (Supplementary Figure 1). At 48hpf there is very weak expression in the hindbrain but no apparent expression elsewhere in the embryo (Supplementary Figure 1). The weak expression in the hindbrain continues from 72-96hpf and at 72 hpf expression appears in the intestinal bulb and continues at 96hpf (Supplemental Figure S1). To determine the functional significance of *lrba* in zebrafish ENS development genes we designed two different splice blocking

morpholinos (SBMOs). When examined at 120hpf *lrba* morphants had a shortened body axis and subtle gut morphogenesis defects but no significant reduction in enteric neuron number when compared to control embryos (Figure 5A, C).

Knockdown of *mab21l2* gene causes a reduction in the number of enteric neurons in Zebrafish

As *MAB21L2* is residing within the *LRBA* gene, and because we hypothesized that the *LRBA* mutation might have an influence on *MAB21L2* expression, we performed expression and knockdown experiments in zebrafish for this gene. As in the human the zebrafish *mab21l2* gene is located within the *lrba* gene.

Whole-mount in situ hybridization (WISH) studies revealed that *mab21l2* is strongly expressed in zebrafish embryos from 24-96hpf (Supplementary Figure 2). It has particularly strong expression in the hindbrain and cranial neural crest especially at 48hpf (Supplementary Figure 2). This cranial neural crest expression can be most clearly seen in the pharyngeal arches (Supplementary Figure 2). This pattern of expression is consistent with previously reported *mab21l2* expression[21]. Significantly though we observed that *mab21l2* is expressed in the gut mesoderm from 48hpf onward, which had not been previously reported (Supplementary Figure 2).

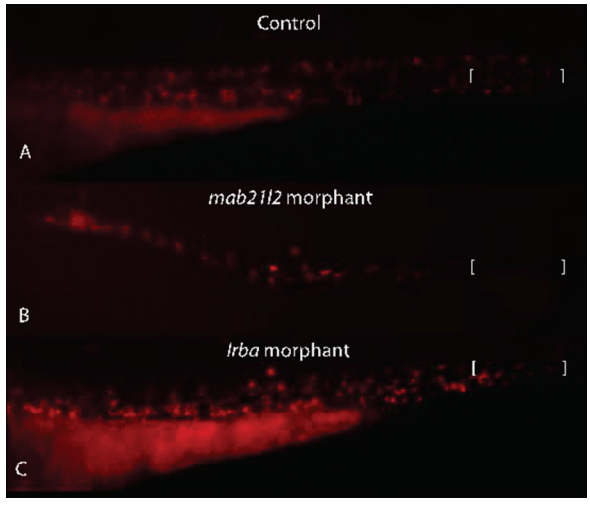


Figure 5. Enteric neurons in control, *mab21l2* morphant and *lrba* morphant embryos. HuC/Elavl3 antibody stain shows differentiated neurons in control and morphant embryos at 120 hpf. *mab21l2* morphants have reduced numbers ofenteric neurons and aganglionosis in the distal part of the intestines compared to control embryos (A-B) (brackets indicate aganglionic region in B and equivalent regions in A and C). Enteric neuron number and distribution along the gut in *lrba* morphants is similar to controls.

To investigate the *in vivo* function of this gene we used a morpholino knock down approach. Knock down of *mab21l2* causes defects in pharyngeal arches and intestinal smooth muscle development. These phenotypes were identical to those previously published[21]. Critically, *mab21l2* morphants also display a significant reduction in the number of enteric neurons (72% reduction in enteric neurons as compared to controls) (Figure 5B).

RET-P399L disturbs protein glycosylation and affects phosphorylation upon GDNF activation

RET is a transmembrane receptor tyrosine kinase required for normal development, maturation and maintenance of a limited number of cell types including the neural crest-drived precursor cells of the ENS. To determine the effect of the *RET* rare variant identified (c.1196C>T, p.P399L), cell lysates extracted from HEK293 cells transiently transfected with pCMV-RET-WT and pCMV-RET-Mut (P399L) expressing vectors, were analysed by Western-blot. In the presence of the RET-WT expressing vector, two bands were identified for RET protein as expected. The lower one (~150 kDa) corresponds to the un-glycosylated RET protein; while the upper band (~170 kDa) is the glycosylated (mature) RET protein. In the presence of the RET-Mut (RET-P399L) expressing vector, only the lower band was detected, suggesting that the variant identified disturbs protein glycosylation (Figure 6).

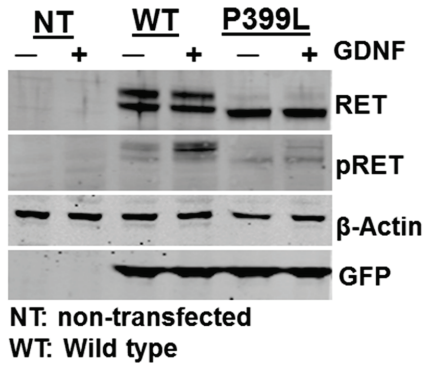


Figure 6. RET variant (c. 1196C>T, p.RET-P399L) disturbs protein glycosylation and phosphorylation. Western blot analysis of cell lysates isolated from HEK293 cells transiently transfected with pCMV-RET-WT and RET-P399L expressing vectors in the presence (+) or absence (-) of GDNF (50ng/ml). Anti-RET and anti-phospho RET (pRET) primary antibodies were used for detection of RET expression and activation levels. β-actin was used as a loading control and GFP as a transfecting control. M: Marker, NT: non-transfected, WT: wild type.

RET phosphorylation was also investigated upon GDNF stimulation, and we observed that in the presence of the RET-Mut (RET-P399L) expressing vector, RET phosphorylation was dramatically reduced (Figure 6). These result suggests that the variant identified is likely pathogenic.

IHH-Q51K disturbs/impairs the activation of Hedgehog signaling

IHH is a member of the Hedgehog signaling pathway. IHH binds to the patched (PTCH1) receptor, which together with Smoothened (SMO) activates the transcription of several target genes. In order to study the affect of this *IHH* variant identified (c.151C>A, p.Q51K) on Hh signaling, we transiently transfected HEK293 cells with plasmids containing *IHH*-WT-FLAG and *IHH*-Q51K-FLAG followed by collection of the culture medium (conditioned medium) and cell lysates after 24 hours. Western blot analysis was performed for the cell lysates and conditioned medium using an anti-FLAG antibody.

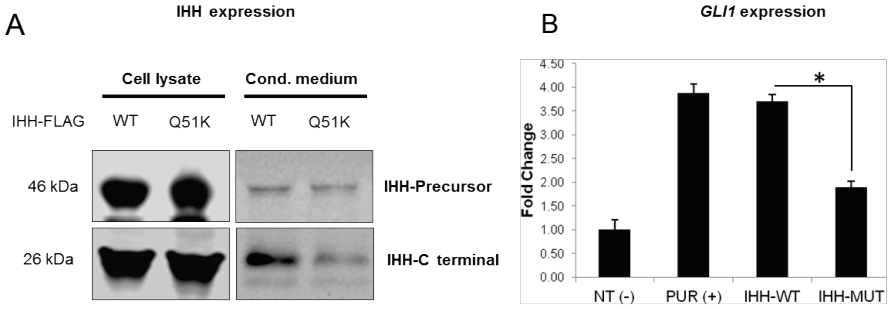


Figure 7. (A) Expression of IHH. The IHH-WT-FLAG, IHH-Q51K-FLAG precursor (~46 kDa) and C-terminal fraction (~26 kDa) are equally expressed in the cells and in secreted as detected by western blot analysis. **(B) GLI1 expression.** Quantification of *GLI1* expression by qPCR in fibroblasts grown in the presence of conditioned medium containing IHH-WT or IHH-Q51K secreted proteins. Purmorphamine (PUR+), an activator of Hh signaling was used as a positive control. Gli1 expression is lower for in cells that were stimulated with the mutant IHH, when compared to cells stimulated with WT IHH.

We found comparative expression levels for the precursor form of IHH-WT (~46 kDa) and IHH-Q51K in the cell lysate and in the conditioned medium from the transfected HEK293 cells (Figure 7A). To determine whether the *IHH* rare variant identified had an effect on overall Hh signaling, we determined *GLI1* (transcriptional target of Hh signaling) expression by qPCR in fibroblasts after growing them in conditioned medium containing the secreted form of either IHH-WT or IHH-Q51K. Our data shows a significantly lower *GLI1* expression for the fibroblast cells that were stimulated with conditioned medium containing the secreted form of mutant IHH when compared to the cells stimulated with WT IHH (Figure 7B).

***Ihh* is required for ENS development in zebrafish**

Although, a role for *shh* in zebrafish ENS development, have been previously shown, the role of *ihh* in ENS development in zebrafish has not been elucidated[29]. To address this we injected *Tg(-8.3phox2b:Kaede)* transgenic embryos with an *ihh*

morpholino, and imaged at 120hpf to determine if these embryos have an enteric neuronal phenotype (Figure 8). *ihh* morphants have a curved body, small eyes, no swim bladder as well as a 87.0% decrease in enteric neurons as compared to uninjected controls (Figure 8 B, D).

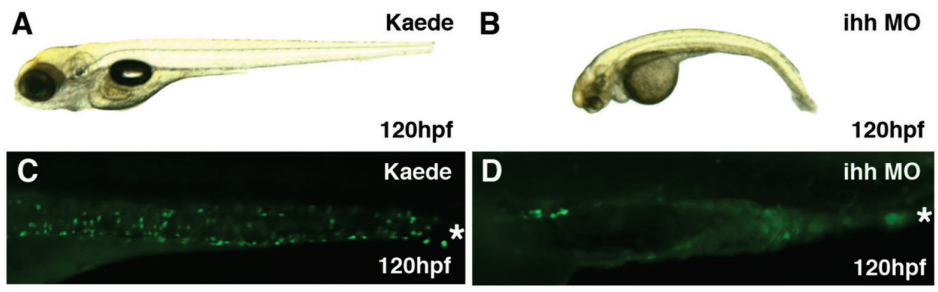


Figure 8. *ihh* morphant embryos exhibit a significant decrease in enteric neurons. (A, C) Kaede uninfected control embryos and (B,D) *ihh* MO injected embryos. (A, B) 120hpf *ihh* MO injected embryos has a curvature of the body, a smaller eye, craniofacial abnormalities and a loss of swim bladder. (C, D) Lateral views of 120hpf embryos stained with *anti-Elavl3* antibody show an 87.0% decrease in enteric neurons.

DISCUSSION

Variants identified within the linkage interval of chromosome 4

After performing exome sequencing on two distantly related, affected members of the large HSCR family, we identified two rare missense variants in the linkage interval: one in the exon 20 of the *LRBA* gene (NM_001199282.2:c.2444A>G) and the other one in *TMEM144* (NM_018342.4:c.715A>T).

TMEM144 is a protein-coding gene with unknown function. It has been reported to be over-expressed in brain, fetal myelinating oligodendrocyte cells and cerebral cortex[30]. However *Tmem144* is not expressed in mouse gut or ENS percursors at E14.5 (in-house data not shown). The Minor allele frequency of the *TMEM144* variant (NM_018342.4:c.715A>T) is 0.005117 in the Exome Aggregation Consortium (ExAC) database (Cambridge, MA, URL: http://exac.broadinstitute.org) although, it is relatively common in the Dutch population (GoNL 0.014). We therefore concluded that it is unlikely that *TMEM144* is involved in HSCR development.

The other variant present in the *LRBA* gene (NM_001199282.2:c.2444A>G) is also rare, with a minor allele frequency of 0.002534 in the ExAC and an allele frequency of 0.009 in GoNL. *LRBA* encodes for the LPS-Responsive vesicle trafficking, Beach and Anchor containing protein. It has been shown that *LRBA* is involved in cancer cell growth and it is hypothesized

that it is a positive regulator of cell survival and is anti-apoptotic[31]. Homozygous variants in *LRBA* have been implicated in common variable immunodeficiency with autoimmunity (CVID) [32] and inflammatory bowel disease (IBD)[33]. In this context it is intriguing that ENS abnormalities in IBD patients have also been described and reported[34,35]. The enteric neuroglial apoptosis in IBD points towards a defective ENS and since HSCR is an ENS defect it is possible that *LRBA* might also play a role in HSCR.

LRBA is a cytosolic protein expressed in almost all cell types, but highly expressed in immune cells[36]. The function of LRBA is regulating the endosomal trafficking, particularly endocytosis of ligand-activated receptors[37]. This gene belongs to the WDL-BEACH-WD (WBW) gene family and genes of this family share a conserved WBW multidomain structure at their C terminal[36]. WBW proteins appear to function as scaffolding proteins in vesicle trafficking and among the 8 members of this protein family, NBEA regulates neurotransmitter receptor trafficking to the synapses and it is known for its role in neuronal development and synaptic functions[38]. *NBEA* has already been implicated in autism spectrum disorders[39,40]. Since NBEA has 75% protein homology with LRBA, it is possible that a connection between LRBA and ENS development/HSCR exists. *In silico* analysis predicted that the *LRBA* variant could be a weak splice site variant. However, we were not able to confirm this prediction with our *in vitro* splicing assay, making it difficult to prove that the variant affects normal splicing (Figure 3).

Our results from zebrafish studies also provided no evidence to support a direct role for *lrba* in zebrafish ENS development. This result is striking as orthologs of nearly every other previously identified HSCR genes have been shown to have an evolutionarily conserved function in zebrafish ENS development[41]. As no evidence was found that the synonymous *LRBA* variant in exon 20 has any effect on the encoded LRBA protein via splicing nor did we find evidence from our zebrafish experiments that *LRBA* could be the HSCR gene we were looking for, we hypothesized that this variant could have an influence on the expression and regulation of *MAB21L2*, a gene present in intron 42 of *LRBA* and known to be involved in neural development (Figure 2). This idea was triggered by the study of Tsang et al, which showed that non-coding sequences scattered throughout *Lrba* give rise to tissue-specific expression of a reporter gene during mouse embryonic development[28]. Therefore, we hypothesized that the *LRBA* variant (c.2444A>G) might reside in an enhancer element that could influence the expression of *MAB21L2*. To prove this hypothesis we performed luciferase assays and were able to show that this DNA sequence (*LRBA*-WT) can enhance the expression of the luciferase gene. However, we could not detect any difference in

luciferase expression between the WT and the mutant variant (Figure 4), and therefore it could not be proved that this variant has an influence on *MAB21L2* expression.

Although our *in vitro* assays could not pinpoint *MAB21L2* as the disease associated gene on chromosome 4, we performed expression studies and morpholino induced knockdown studies in zebrafish for *MAB21L2*. mRNA expression analysis of the zebrafish ortholog of *mab21l2* showed that this gene's expression is consistent with playing a role in ENS and GI tract development. Furthermore our results confirm and extend the previously reported expression data for this gene in zebrafish[21]. Our *in vivo* functional analysis using morpholinos also confirm the previously reported *mab21l2* morphant phenotype but also reveals an essential role for this gene in zebrafish ENS development. This result is potentially significant and suggests that *MAB21L2* is indeed a potential HSCR gene. Further functional analysis in mammalian model systems will be required to confirm this.

Conclusion: a HSCR gene on chromosome 4

Based on bioinformatics data, published reports and *in vivo* data presented here, *MAB21L2* seems to be a perfect candidate for HSCR in the linkage interval. However, we have not found any genetic evidence for its involvement with HSCR in this family. We can also not totally exclude that another gene in the linkage region is involved or that the linkage found happened by chance. The parametric multipoint LOD score of 2.7 which was found in the family should be considered as 'suggestive linkage'. However, assuming that there is true linkage, we consider that *MAB21L2* is the best candidate gene in the linkage interval.

Rare variants in known HSCR genes

Hedgehog signaling

In one branch of this family, in the patient V-4, we detected not only the rare variant in *LRBA*, but also variants in the Indian hedgehog (*IHH*) gene (NM_002181.3:c.151C>A), Neuron Navigator 2 (*NAV2)* (NM_001244963.1:c.2569C>T), Arginine Vasopressin Receptor 2 (*AVPR2)* (NM_000054.4:c.1110_1112delATC) and GLI family zinc finger 3 (*GLI3)* (NM_000168.5:c.2119C>T). The *IHH, LRBA, NAV2 and GLI3* variant were inherited from the father (IV-4), while the deletion in *AVPR2* was inherited from the mother (IV-5).

NAV2 (Neuron Navigator 2) is a protein-coding gene, which plays a role in cellular growth and migration. It has been shown that *Nav2* is important for normal development of cranial neuronal fibres in mice during embryonic

development and regulation of blood pressure[42]. *AVPR2* gene encodes the vasopressin receptor, belonging to seven-transmembrane-domain G protein-coupled receptor (GPCR) superfamily, type 2. Mutations in *AVPR2* are associated with nephrogenic diabetes insipidus[43]. *GLI3* is a protein coding gene for one of the transcriptional factor and mediator of Hh signaling. It has a dual function, acting as an activator and repressor, and is known to play an important role during embryogenesis and limb development. Loss of function mutations in *GLI3* have been described in patients with Pallister Hall syndrome[44] (PHS, MIM 146510) and in Greig cephalopolysyndactyly syndrome[45] (GCPS, MIM 175700).

Recently, mutations in *GLI3* have also been identified in HSCR patients and it was also shown that disruption of Gli activity in mice interrupts with the ENS development[46]. The *GLI3* variant (c.2119C>T) present in patient V-4 is inherited from the father (IV-4). Remarkably, neither IV-4 nor V-4 has mesoaxial or postaxial polydactyly, bifid epiglottis, hypothalamic hamartoma, genitourinary, lung or skeletal anomalies seen in patients with PHS[9,44]. This could be due to the mild nature of the variant, however we also cannot exclude that the variant is non-causative.

IHH is part of the Hedgehog (Hh) signaling pathway which is known to play a diverse and important role in embryogenesis, including the development of the gastrointestinal tract. A Xenopus model for *ihh* indicates that this is a crucial gene for neural crest cell formation, maintenance and migration of neural crest cells[47]. Ramalho *et al*. showed that 50% of *Ihh* knock out mice develop a HSCR-like phenotype[48], confirming the importance of this gene for ENS development. However, Ihh depletion in mice is not fully penetrant, indicating that additional mutations in other gene are required for the ENS phenotype. Activation of Hh signaling activates GLI transcriptional factors, which are crucial in expression and regulation of many important ENS genes such as *SOX10*, a well-known transcription factor for ENS development found mutated in syndromic HSCR cases. All this evidence makes the *IHH* variant in exon 1 (NM_002181.3:c.151C>A) the perfect candidate for HSCR development. This is further corroborated by the fact that this variant is located in the active site (N-terminal) of the IHH protein and is predicted to be damaging, possibly by disturbing the secretion or stability of IHH, and hence disturbing the Hh signaling pathway. Our results confirmed the pathogenic nature of the *IHH* variant identified, showing that it disturbs Hh signaling via *GLI1* (Figure 7).

Our zebrafish *in vivo* results further support an evolutionarily conserved role for IHH in ENS development. Sonic hedgehog signaling has been previously shown to be required for normal ENS development in zebrafish[29]. Hedgehog pathway signaling is required in two phases of zebrafish ENS development, an

early migratory phase as well as during a proliferative phase. The precise role of *ihh* in ENS development has yet to be elucidated. In zebrafish *ihh* has been found to play a critical role in esophageal and swim bladder development and was shown to interact with *fgf10*[22]. The swim bladder arises as an outgrowth of foregut endoderm[49]. Further studies are required to determine if the zebrafish ENS defect is due to a failure of migration of the enteric neural crest precursors from the vagal neural crest region to the anterior end of the gut tube, similar to loss seen when either Hedgehog signaling *is* perturbed or whether IHH is required for proliferation of ENCCs once they are in the GI tract[29].

As patient V-4 inherited all but one variant from the unaffected father, along with mutation in LRBA, we reasoned that there might be a chance of strong *de novo* mutation. Therefore, we screened the patient and his parents via exome sequencing for such a mutation. We identified *de novo* mutations in 3 genes including one in *GDNF*, the gene encoding the ligand of RET.

RET/GDNF signaling

We identified a heterozygous rare variant in the coding region of *RET* (c.1196C>T, p.P399L) in patient V-1 a mutation previously missed by our DGGE screen[6]. Segregation analysis showed that her two affected siblings (V-2, V-3), the unaffected mother (IV-2) and the affected maternal uncle (IV-3) also have the same heterozygous *RET* variant, while the grandmother (III-2) does not. The mother (IV-2) and grandfather (III-1) are considered to be unaffected, but they were reported to suffer from severe constipation in childhood. Most likely, this *RET* variant was inherited from the grandfather (III-1), but since his DNA was unfortunately not available, we were unable to confirm our suspicions. Previous study on this family reported that all the three affected siblings (V-1,V-2,V-3) inherited a heterozygous common risk haplotype in *RET* from their father[9]. This common risk haplotype is marked by 14 SNPs scattered from 2 kb upstream of *RET* until the beginning of exon 2[7]. The heterozygous variant in the exon 20 of *LRBA* and common risk haplotype in *RET* might therefore, contribute to the development of HSCR and enhance the penetrance of the *RET* coding variant (c.1196C>T) to their offspring (V-1, V-2, V-3), possibly explaining why all of their children are affected. We performed functional analysis for this *RET* coding variant and could show that the identified *RET* variant disturbs RET glycosylation (Figure 6) Non-glycosylated RET proteins do not generally get transported to the plasma membrane, and as a consequence, cannot be activated by GDNF. We were able to confirm this effect with our in vitro assays (Figure 6), proving that the RET-P399L variant is pathogenic and results in RET dysfunction.

As mentioned in V-4 we identified a *de novo* mutation in GDNF, comprising a six base pair deletion. It is an in-frame deletion resulting in the loss of two amino acids. At this moment we cannot be certain whether this variant is truly pathogenic. However, the chance of finding in a *de novo* mutation in a well-known HSCR gene in a HSCR family is extremely low, making this mutation very suspicious.

CONCLUSIONS

This study used a combination of linkage analysis and whole exome sequencing to determine the genetic cause of HSCR in a multigenerational Dutch family. It perfectly shows the complexity of HSCR genetics. We have identified a number of possible causal variants and have demonstrated, for most of them, their contribution for ENS development. In addition, our studies have highlighted the role of Hh signaling for the development of HSCR in humans and reiterates the role of RET signaling. A complete understanding of the genetics of an inherited complex disease is a major challenge requiring substantial efforts, and it is the combination of genetics and functional studies that has given us these new insights.

ACKNOWLEDGEMENTS

The authors would like to thank the patients and families involved in this study. The pCMV-IHH-FLAG-WT vector was kindly provided by Dr.Gang Ma from Shanghai Jiaotong University. This study was supported by research grant from ZonMW (TOP-subsidie 40-00812-98-10042) and the Maag Lever Darm stichting. (WO09-62)

REFERENCES

1. Heanue, T.A. & Pachnis, V. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat Rev Neurosci* **8**, 466-79 (2007).
2. Brooks, A.S., Oostra, B.A. & Hofstra, R.M. Studying the genetics of Hirschsprung's disease: unraveling an oligogenic disorder. *Clin Genet* **67**, 6-14 (2005).
3. Badner, J.A., Sieber, W.K., Garver, K.L. & Chakravarti, A. A genetic study of Hirschsprung disease. *Am J Hum Genet* **46**, 568-80 (1990).
4. Alves, M.M. *et al.* Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev Biol* **382**, 320-9 (2013).
5. 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**, 581-96 (2015).
6. Hofstra, R.M. *et al.* RET and GDNF gene scanning in Hirschsprung patients using two dual denaturing gel systems. *Hum Mutat* **15**, 418-29 (2000).

7. Burzynski, G.M. *et al.* Identifying candidate Hirschsprung disease-associated RET variants. *Am J Hum Genet* **76**, 850-8 (2005).
8. Gabriel, S.B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* **31**, 89-93 (2002).
9. Brooks, A.S. *et al.* A novel susceptibility locus for Hirschsprung's disease maps to 4q31.3-q32.3. *J Med Genet* **43**, e35 (2006).
10. Mariani, M. *et al.* Two murine and human homologs of mab-21, a cell fate determination gene involved in Caenorhabditis elegans neural development. *Hum Mol Genet* **8**, 2397-406 (1999).
11. Wakamatsu, N. *et al.* Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. *Nat Genet* **27**, 369-70 (2001).
12. Verschueren, K. *et al.* SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem* **274**, 20489-98 (1999).
13. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215 (1988).
14. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).
15. Brouwer, R.W., van den Hout, M.C., Grosveld, F.G. & van Ijcken, W.F. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28**, 284-5 (2012).
16. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
17. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
18. Widowati, T. *et al.* RET and EDNRB mutation screening in patients with Hirschsprung disease: Functional studies and its implications for genetic counseling. *Eur J Hum Genet* (2015).
19. Ma, G. *et al.* Indian hedgehog mutations causing brachydactyly type A1 impair Hedgehog signal transduction at multiple levels. *Cell Res* **21**, 1343-57 (2011).
20. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* **3**, 59-69 (2008).
21. Kennedy, B.N. *et al.* Zebrafish rx3 and mab21l2 are required during eye morphogenesis. *Dev Biol* **270**, 336-49 (2004).
22. Korzh, S. *et al.* The interaction of epithelial Ihha and mesenchymal Fgf10 in zebrafish esophageal and swimbladder development. *Dev Biol* **359**, 262-76 (2011).
23. Shepherd, I.T., Pietsch, J., Elworthy, S., Kelsh, R.N. & Raible, D.W. Roles for GFRalpha1 receptors in zebrafish enteric nervous system development. *Development* **131**, 241-9 (2004).
24. Robu, M.E. *et al.* p53 activation by knockdown technologies. *PLoS Genet* **3**, e78 (2007).
25. Van Der Werf, C.S. *et al.* CLMP is required for intestinal development, and loss-of-function mutations cause congenital short-bowel syndrome. *Gastroenterology* **142**, 453-462 e3 (2012).
26. Sribudiani, Y. *et al.* Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology* **140**, 572-582 e2 (2011).
27. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).
28. Tsang, W.H., Shek, K.F., Lee, T.Y. & Chow, K.L. An evolutionarily conserved nested gene pair - Mab21 and Lrba/Nbea in metazoan. *Genomics* **94**, 177-87 (2009).
29. Reichenbach, B. *et al.* Endoderm-derived Sonic hedgehog and mesoderm Hand2 expression are required for enteric nervous system development in zebrafish. *Dev Biol* **318**, 52-64 (2008).
30. Carithers, L.J. & Moore, H.M. The Genotype-Tissue Expression (GTEx) Project. *Biopreserv Biobank* **13**, 307-8 (2015).
31. Wang, J.W. *et al.* Deregulated expression of LRBA facilitates cancer cell growth. *Oncogene* **23**, 4089-97 (2004).
32. Lopez-Herrera, G. *et al.* Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. *Am J Hum Genet* **90**, 986-1001 (2012).

33. Serwas, N.K. *et al.* Atypical manifestation of LRBA deficiency with predominant IBD-like phenotype. *Inflamm Bowel Dis* **21**, 40-7 (2015).
34. Villanacci, V. *et al.* Enteric nervous system abnormalities in inflammatory bowel diseases. *Neurogastroenterol Motil* **20**, 1009-16 (2008).
35. Bassotti, G. *et al.* Enteric neuroglial apoptosis in inflammatory bowel diseases. *J Crohns Colitis* **3**, 264-70 (2009).
36. Wang, J.W., Howson, J., Haller, E. & Kerr, W.G. Identification of a novel lipopolysaccharide-inducible gene with key features of both A kinase anchor proteins and chs1/beige proteins. *J Immunol* **166**, 4586-95 (2001).
37. Alangari, A. *et al.* LPS-responsive beige-like anchor (LRBA) gene mutation in a family with inflammatory bowel disease and combined immunodeficiency. *J Allergy Clin Immunol* **130**, 481-8 e2 (2012).
38. Nair, R. *et al.* Neurobeachin regulates neurotransmitter receptor trafficking to synapses. *J Cell Biol* **200**, 61-80 (2013).
39. Castermans, D. *et al.* The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. *J Med Genet* **40**, 352-6 (2003).
40. Volders, K., Nuytens, K. & Creemers, J.W. The autism candidate gene Neurobeachin encodes a scaffolding protein implicated in membrane trafficking and signaling. *Curr Mol Med* **11**, 204-17 (2011).
41. Heanue, T.A., Shepherd, I.T. & Burns, A.J. Enteric nervous system development in avian and zebrafish models. *Dev Biol* (2016).
42. McNeill, E.M., Roos, K.P., Moechars, D. & Clagett-Dame, M. Nav2 is necessary for cranial nerve development and blood pressure regulation. *Neural Dev* **5**, 6 (2010).
43. van den Ouweland, A.M. *et al.* Mutations in the vasopressin type 2 receptor gene (AVPR2) associated with nephrogenic diabetes insipidus. *Nat Genet* **2**, 99-102 (1992).
44. Biesecker, L.G. Pallister-Hall Syndrome. (1993).
45. Biesecker, L.G. The Greig cephalopolysyndactyly syndrome. *Orphanet J Rare Dis* **3**, 10 (2008).
46. Liu, J.A. *et al.* Identification of GLI Mutations in Patients with Hirschsprung Disease That Disrupt Enteric Nervous System Development in Mice. *Gastroenterology* (2015).
47. Aguero, T.H., Fernandez, J.P., Lopez, G.A., Tribulo, C. & Aybar, M.J. Indian hedgehog signaling is required for proper formation, maintenance and migration of Xenopus neural crest. *Dev Biol* **364**, 99-113 (2012).
48. Ramalho-Santos, M., Melton, D.A. & McMahon, A.P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763-72 (2000).
49. Winata, C.L. *et al.* Development of zebrafish swimbladder: The requirement of Hedgehog signaling in specification and organization of the three tissue layers. *Dev Biol* **331**, 222-36 (2009).

SUPPLEMENTARY FIGURES

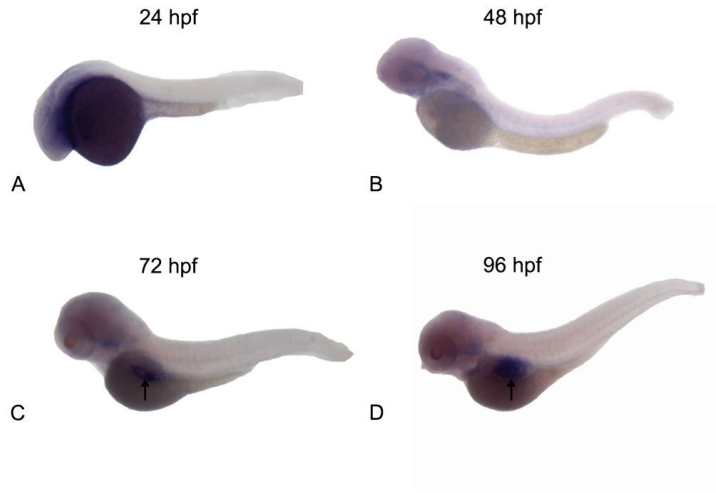


Figure 1. *lrba* expression pattern from 24-96 hpf. *In situ* hybridization showing that *lrba* has a very discrete expression pattern through all time points observed (Arrows indicate intestinal bulb expression). Expression is present along the yolk sack and weakly in the hindbrain at 24 hpf (A). Weak expression in the hindbrain continues throughout all times observed. Strong expression appears in the intestinal bulb from 72-96 hpf (C-D).

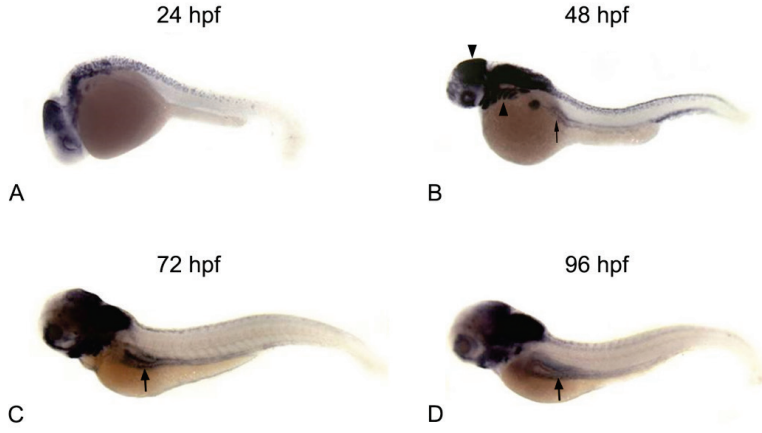


Figure 2. *mab21l2* expression pattern from 24-96 hpf. *In situ* hybridization showing that *mab21l2* has a very strong expression pattern through all time points observed (arrowhead indicates expression in hindbrain and pharyngeal arches and arrows indicate gut mesoderm expression). Expression is present in the hindbrain, cranial neural crest and pharyngeal arches (A-D). Strong expression appears in the gut mesoderm from 48 hpf onwards (B-D).

SUPPLEMENTARY TABLES

Supplementary Table 1. Primers used for site-directed mutagenesis

Gene/ Construct	Forward	Reverse
pCMV-RET-WT/Mut	5'-CGTGTCGGTGCTGC**T**GGTCAGCCTGCAC-3'	5'-GTGCAGGCTGACC**A**GCAGCACCGACACG-3'
pCMV-IHH-FLAG WT/Mut	5'-CGCTCGCCTACAAG**A**AGTTCAGCCCCAATG-3'	5 '-CATTGGGGCTGAACT**T**CTTGTAGGCGAGCG-3'

Supplementary Table 2. Primers used for *in vitro* splicing assay

SD6	5'-TCTGAGTCACCTGGACAACC-3'
SA2	5'-ATCTCAGTGGTATTTGTGAGC-3'

Supplementary Table 3. Antibodies used in Western blot

	Antibodies	Host	Dilution	Supplier
Primary	RET	Rabbit	1:1000	Santa Cruz Biotechnology
	p-RET	Rabbit	1:1000	Santa Cruz Biotechnology
	β-Actin	Mouse	1:500	Santa Cruz Biotechnology
	GFP	Rabbit	1:1000	Abcam
	Flag	Mouse	1:1000	Sigma
Secondary	IRDDye 800	Goat	1:10000	Licor
	IRDDye 680	Goat	1:10000	Licor

Supplementary Table 4. Primers used for qRT-PCR

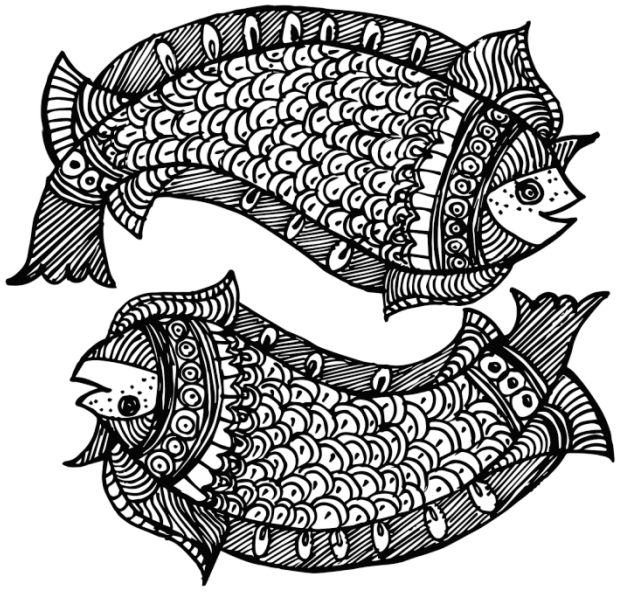
Gene	Forward	Reverse
GLI1	5'-TCCCCATGACTCTGCCCG-3'	5'- CCAGCATGTCCAGCTCAGA-3'
CLK2	5'-TCGTTAGCACCTTAGGAGAGG-3'	5'-TGATCTTCAGGGCAACTCG-3'

Supplementary Table 5. Summary statistics of the exome sequencing data

Exome data	V-1	V-4	IV-4	IV-5
Fraction of targets covered ≥ 10X (%)	97.6	97.2	97	96
Fraction of targets covered ≥ 20X (%)	92.1	90.5	92	89
Fraction of targets covered ≥ 30X (%)	84.1	81.1	85	80

Supplementary Table 6. Primers used for PCR amplification and Sanger sequencing

Gene	Forward	Reverse
LRBA	5'-CCACATAACTTAAGGTTGATTC -3'	5'-GATATAAGGAGATGTGGCTG-3'
RET	5'-CTGGCCAGCCCATCTTGG -3'	5'- CCGAGTCACCATATGCAGATTTACC-3'
IHH	5'-ATCAGCCCACCAGGAGACC -3'	5'-CATCAGCCCACCAGGAGACC-3'
AVPR2	5'-CCACCAGCCATCCTGAACC -3'	5'-CAGCTGGGGATGTGGAGACC-3'
NAV2	5'-CAGCCCTCGGCTCCAAGC -3'	5'-CTGGCCAAGCCTGGACTACC-3'
TMEM144	5'-GTGAGCCACTGCGCTCTGC-3'	5'-CACAGAGGATGGCTTTGTTTCC-3'
GLI3	5'-AGTGGCCAGCTCCATTCACC-3'	5'-GGTTACAGCGTCATTTTAGGACTGG-3'
NRP2	5'-GAATTGCAAACTGATACTAATTAC-3'	5'-CAAGGCCCTCTCTCCTGTAG-3'
GDNF	5'-TTTCAAACCCTAATGCACTTTTATTCC-3'	5'-TGACCTGGAAAAGGCCAAGG-3'
ALMS1	5'-GGCAAACATTTCCTGGGAACC-3'	5'-GGCTGGTGAGTGACAAAGTAGGG-3'
ST18	5'-GCCTAAGCTGGGCCACAACC-3'	5'-GGGGCCCGTAGTGAGAGTCG-3'



CHAPTER 4

Identification of predictive regulatory sequences for gut development

R.K. Chauhan[1], M. van der Sijde[2], C. Paul[3], J. Fu[2,4], R.M.W. Hofstra[1,5#] B.J.L. Eggen[3], A. Ijpma[6]

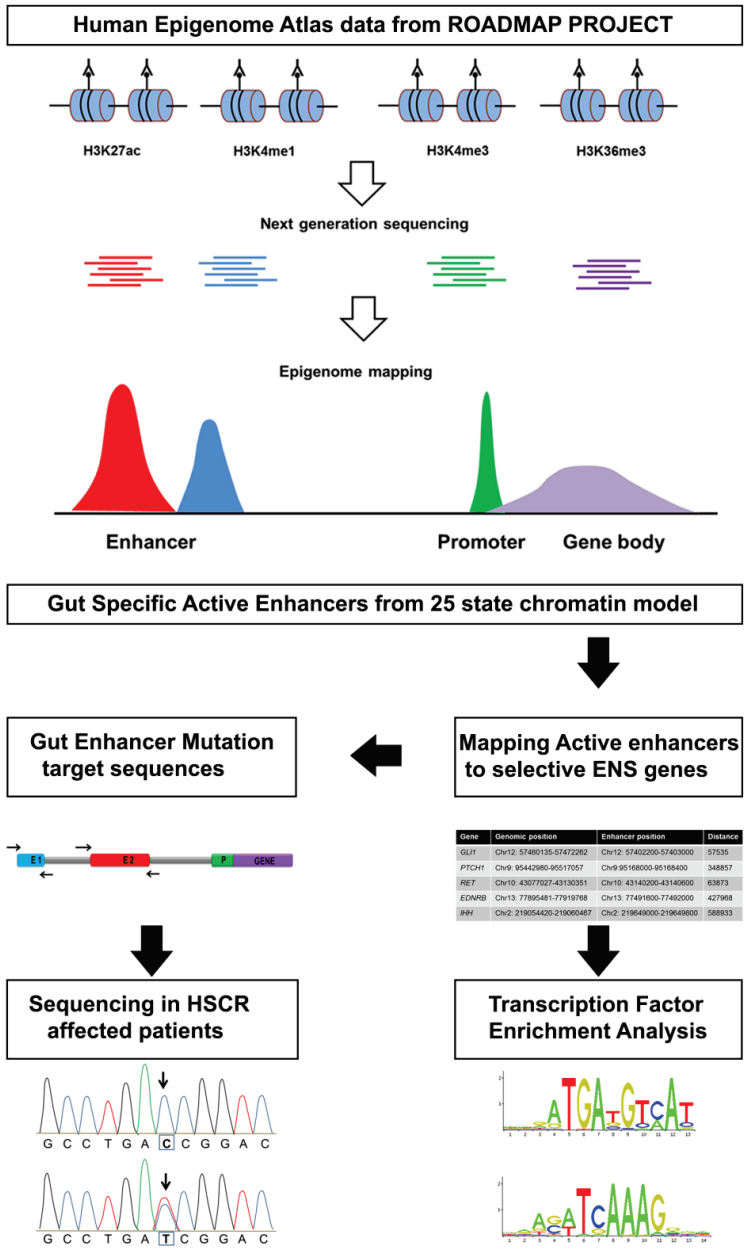
[1] Department of Clinical Genetics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands
[2] University of Groningen, University Medical Center Groningen, Department of Genetics, 9700 RB, Groningen, Netherlands
[3] Department of Neuroscience, section Medical Physiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[4] University of Groningen, University Medical Center Groningen, Department of Pediatrics, 9700 RB Groningen, Netherlands
[5] Stem Cells and Regenerative Medicine, UCL Great Ormond Street Institute of Child Health, London, UK
[6] Department of Bioinformatics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands

#Corresponding author: Robert MW Hofstra, Department of Clinical Genetics, Erasmus University Medical Center, PO BOX 2040, 3000CA Rotterdam, The Netherlands. Tel: +31-10-7037643. E-mail: r.hofstra@erasmusmc.nl

Manuscript in preparation

ABSTRACT

Hirschsprung disease is a congenital disorder characterized by a lack of enteric innervation of the distal gastrointestinal tract. Identification of genes causing HSCR mainly focused on the identification of coding variants. Collectively, the identified coding mutations explain approximately 25 % of the overall genetic risk on HSCR. The identification of pathogenic, non-coding DNA variants linked to complex diseases has been difficult, although many non-coding SNPs have been linked to human diseases. To identify novel non-coding DNA variants that alter regulatory elements involved with ENS development and hence potentially candidate variants for HSCR, we used existing epigenome atlas datasets. With a computational approach we catalogued all the active gene enhancers of sigmoid colon and fetal large intestine and intersected them with candidate HSCR gene involved with ENS development. These active enhancer regions can be further investigated for screening non-coding mutations and could be applicable in targeted sequencing approaches for understanding yet unexplained HSCR genetics.



Graphical Abstract: Schema of identifying gut specific active enhancers for ENS genes.

INTRODUCTION

Hirschsprung disease (HSCR) is one of the most common congenital disorders of the gastrointestinal (GI) tract, it caused by a partial absence of the enteric nervous system (ENS), called aganglionosis, in the most distal part of the GI tract. This aganglionosis results in functional obstruction of the bowel and in life threatening constipation. HSCR is an inherited disorder with a high recurrence risk for sibling. This recurrence risk depends on the length of the aganglionosis and the gender of the affected patient within the family. The recurrence risk ranges from 1 to 33%[1]. Moreover, familial occurrence is seen in approximately 10% of all cases and chromosomal abnormalities in 12 % of all cases. Finally, co-occurrence of additional malformations and syndromes and the finding of naturally occurring animal models all point to the involvement of inherited factors. Disease transmission in families can be either dominant or recessive, with incomplete penetrance and variable expressivity[1]. In sporadic HSCR the disease the inheritance is thought to be non-Mendelian (polygenic).

In the last two decades all kinds of genetic studies have been conducted ranging from linkage analysis in multigenerational (syndromic) families, sibpair analysis on smaller families, haplotype sharing studies in founder populations (for an overview see Brooks *et al*. 2005, Alves *et al*. 2013). These studies resulted in coding mutations in 16 genes that can cause, or contribute to the development of HSCR[2,3]. Besides these studies also association studies, including genome wide association studies (GWAS) on sporadic patients have been performed and these resulted in association with a locus with a 25 Kb region in 5'end of *RET*. Comparative genomic approaches identified a conserved enhancer region within intron1 of *RET* and within this enhancer region reside two strongly disease associated non-coding SNPs (rs2435357 and rs2506004) that negatively influence the enhancer activity leading to lower RET expression[4,5]. It pointed towards the involvement of non-coding variants in the development of HSCR.

This was not the first indication for the involvement of non-coding variants in HSCR. Bolk *et al*. conducted a linkage analysis study on 12 multiplex HSCR cases and found that 11 the families have linkage to the RET locus, but only half of them carry *RET* coding variants[6]. So, it was hypothesize that non-coding variants in *RET* should be present to explain the linkage to the *RET* locus. Genetic studies in HSCR altogether don't explain more than 25% of the overall genetic risk. The inability to understand and find genetic aberrations in HSCR, suggests that coding variants in not yet screened genes or variants in the unscreened regulatory genome should be searched for to unravel the genetic based of HSCR.

In the human genome, protein coding genes account for only 1.5% of the genome which increases to 2% if untranslated regions (UTRs) are also included[7]. Multiple studies outline the importance of the vast information hidden in the non-coding DNA, especially in the regulatory elements (such as enhancers) for gene regulation at the level of transcription and translation. Enhancers are genomic elements that regulate gene expression. Enhancers can be proximal or distal to the transcriptional start sites (TSS) of a gene and work independently of position and orientation[8,9]. They function as binding platforms for transcription factors and are characterized by specific epigenetic modifications[10]. Mutations in cis-regulatory genome have to date only been identified in a few known human diseases, for example in cancers, preaxial polydactyly, pancreatic agenesis, congenital heart disease and Parkinson's disease[11-15]. Historically, the search for mutations contributing to human diseases, as just describe for HSCR, had been limited to the coding part of genes and non-coding sequences remained poorly investigated.

The integrative analysis of the human epigenomic landscape for primary cells and tissues has revealed a genome-wide map of regulatory regions. Approximately 5% of each reference epigenome has enhancer and promoter signatures[16]. Imputation and annotation of epigenome marks to predict different chromatin states across various reference genomes has opened new applications that were previously not possible[17]. Disease-causing regulatory mutations at enhancer sequences are increasingly recognized, drawing attention to their importance in complex diseases, such as HSCR.

In this study, we shortlisted candidates HSCR genes, and by using the epigenetic information from the Roadmap project, we identified gut specific enhancers within the genomic regions in which these genes reside. Furthermore, we performed transcription factor (TF) enrichment analysis on gut enhancers to identify putative master TFs, which could regulate the expression of ENS genes. The identified gut enhancers can be screened in HSCR patients in conjunction with exome sequencing to interrogate the role of non-coding enhancer sequences in the etiology of HSCR.

METHODS

Selection of the ENS gene set

All the genes known to be associated with HSCR, genes involved in ENS development from previous expression studies in mouse and all known animal models were selected and combined to generate a list of ENS genes (data not shown). The entire gene coordinates, TSS (transcription start site) and strand

details were mapped using Biomart tool (http://www.ensembl.org/info/data/biomart).

Enhancer datasets of the ENS genes

We used publically available dataset consisting of imputed epigenome from roadmap epigenomics project that has predicted target marks of epigenetic regulation of different cells and tissue types based on the reference epigenome (http://egg2.wustl.edu/roadmap/web_portal/imputed.html)[17]. We downloaded imputed data consisting of 25 states and 12 marks for sigmoid colon (E106), fetal intestine large (E084) fetal thymus (E093) and fetal lung tissue (E088). H3K27ac data (EnhA1, EnhA2 and EnhAF), that marks active enhancers was then filtered out for further analysis and termed as active enhancer hereon. Hg38 assembly and annotation was used for all the analysis. Custom perl scripts were used to couple the ENS genes to their prospective active enhancers by scanning +/- 1MB distance from the TSS of the gene, generating enhancer-gene couplets.

Transcription factor identification and enrichment

One enhancer region bed-file was generated by combining the sigmoid colon and fetal large intestine active enhancer datasets. Similarly, one control region bed-file was generated by combining enhancer region files of fetal thymus and fetal lung. These bed-files were then converted to BAM files using the bed to BAM converter on the Galaxy platform[18]. The BAM files were uploaded into the ChipSeq workflow of Partek Genomics Suite 6.6 (Partek Incorporated, 624 Trade Center Boulevard, St. Louis, Missouri 63005, USA). We used the workflow that incorporates the JASPAR database[19] for transcription factor recognition sites, in order to identify significantly enriched TFs for the active enhancers of colon and fetal large Intestine tissues as compared to the control dataset (fetal thymus and fetal lung tissues enhancers). To identify enriched functions associated with the uniquely identified TFs for the colon and fetal large intestine enhancers, we analyzed these genes using the Ingenuity Pathway Analysis (IPA) tool (Qiagen Silicon Valley, 1700 Seaport Blvd, 3rd Floor, Redwood City, CA 94063).

RESULTS

Mapping and integration of gut enhancers

For our studies, we selected 115 ENS genes consisting of known HSCR genes, genes from associated HSCR loci and genes known to be involved in ENS development from previous expression studies and animal models. The majority

of these genes are enriched for biological functions related to neural crest, such as neural crest cell development, cell fate and neuronal differentiation (Figure 1).

To identify active enhancers of the selected ENS genes in the gut (sigmoid colon and fetal large intestine), we used a chromatin state model, based

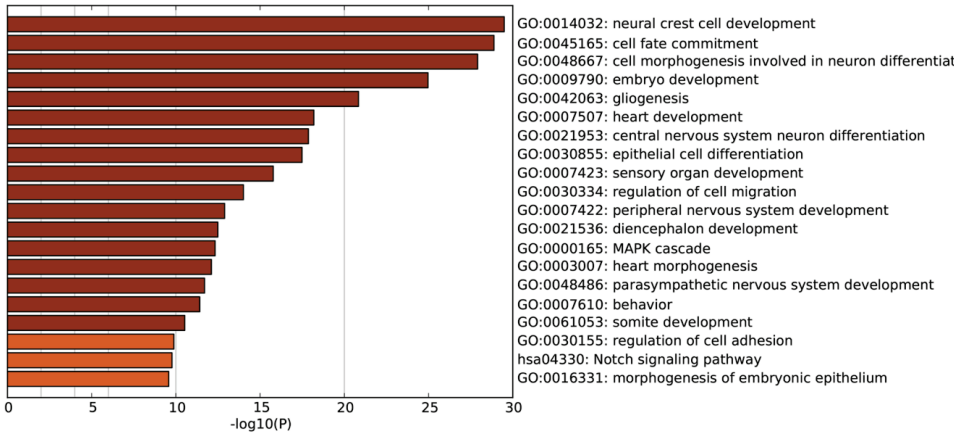


Figure 1. Gene Ontology of candidate ENS genes that were mapped to identify gut specific enhancers. The majority of genes are involved in neural crest development, cell fate determination and neuronal differentiation.

on the imputed epigenomic data for 12 specific marks (H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac, H4K20me1, H3K79me2, H3K36me3, H3K9me3, H3K27me3, H2A.Z, and DNase), and by doing so extracted all active enhancers (EnhA1, EnhA2 and EnhAF) that are marked by H3K27ac[17]. We used criteria of +/- 1MB distance from the TSS of the ENS genes and coupled all the active enhancers to ENS genes. We identified 7297 unique active enhancers for the sigmoid colon and 10127 enhancers for fetal large intestine (data not shown).

Transcription factor binding sites specific to active gut enhancers

We hypothesized that interrogation of DNA sequence motifs, that are most recurrent at gut enhancers compared to the control data set, would reveal enriched TFs driving ENS specific gene expression. Five TFs (Bcl6, JUN, RAP1, MSC and Ddit::Cebpa) motifs are significantly overrepresented ($p \leq 0.0000001$) in the gut datsets compared to the control datasets. When lowering the p value to ≤0.00001 the number increases to 20 TFs (Table 1). The majority of the TF binding sites represent enhancers associated with either the ATF-2 or the AP-1 pathways, respectively (Figure 2A). To identify the biological activities associated with these TFs, IPA was used, it was found that most of the TFs are either associated with cell proliferation or cell differentiation (Figure 2B). Some of them

were also involved with cell adhesion (JUNB, JUN and RAP1A) and neuronal mislocalization (HSF2 and ZNF423).

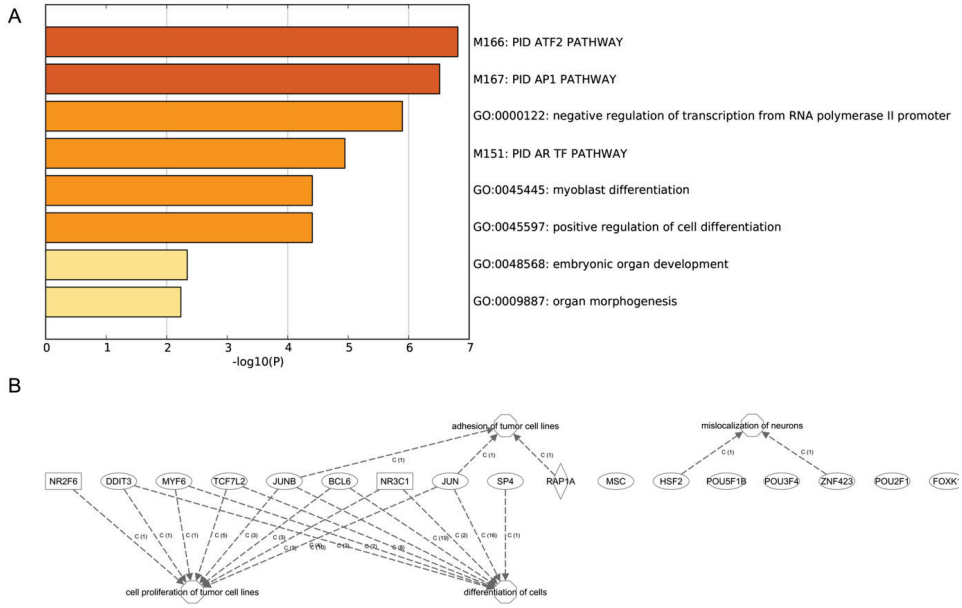


Figure 2. (A) Gene Ontology based functions of TFs enriched on gut enhancers as compared to the control dataset. (B) IPA pathway analysis of the enriched TFs, showing their association with cell proliferation, differentiation, adhesion and mislocalization of neurons.

Table 1. Transcription factors highly enriched on active enhancers compared to control dataset

S.No.	Transcription factor	Actual number of occurrences	p value
1	Bcl6	80	6.7342E-10
2	JUN	194	7.45125E-9
3	RAP1	275	1.30309E-9
4	MSC	150	1.6584E-8
5	Ddit3::Cebpa	323	1.37263E-13
6	pha-4	677	3.84999E-15
7	POU2F1	270	8.84969E-11
8	Nr2f6	9	1.60626E-10
9	POU3F4	383	2.72565E-9
10	Foxk1	55	6.97314E-9
11	NR3C1	75	2.65751E-8
12	POU5F1B	723	7.12749E-8
13	achi	3634	1.4113E-7
14	SP4	12	2.1961E-7
15	Znf423	35	2.38229E-7
16	TCF7L2	167	8.3263E-7
17	HSF2	38	1.17006E-6
18	dl	143	2.44125E-6
19	MYF6	171	3.67337E-6
20	JUNB	313	5.28966E-6

Upstream regulator analysis identifies key TFs

With the upstream regulator IPA tool, we predicted all the upstream regulators of the selected 116 ENS genes. The TF binding sites overrepresented in the gut specific enhancers were intersected with these predicted upstream regulators of the ENS genes to identify shared TFs. JUN and TCF7L2 were identified as significantly ($p<0.005$) enriched. The transcription factor JUN is associated with ERK/JNK pathway.

DISCUSSION

Using publically available Roadmap epigenomics project datasets and computational analysis of the data, we have catalogued thousands of predicted gut-specific enhancers that could potentially contribute to ENS gene expression. Here we present shortlisted regions of non-coding sequences that could be further investigated for identifying causal regulatory SNPs or novel variants associated with HSCR. The computational framework and the methods used can be also applied to other cell type/tissue and screens for active enhancers regulating gene expression.

By comparing transcription factor enrichment on the gut enhancers and control datasets (fetal thymus and fetal lung) for same set of ENS genes, we identified 20 TFs that are overrepresented on gut enhancers. During development TFs regulate gene expression by binding to enhancers and recruit coactivators and RNA polymerase II to target genes[20,21] One the highly significant TF JUN (c-Jun in combination with c-Fos, forms the AP-1 early response transcription factor) has been implicated in a large variety of biological processes including cell differentiation, proliferation, apoptosis oncogenic transformation, embryogenesis and organogenesis[22]. The JUN gene encodes protein c-Jun and the c-Jun N-Terminal Kinase (JNK) pathway is known to be involved in ENS Development (for review see[23,24]). Upstream regulator analysis of the set of ENS genes and an overlap with highly enriched TF on gut enhancer also resulted in identification of JUN along with another TF, TCF7L2. TCF7L2 is a central transcription factor in the canonical wingless-type MMTV integration site (WNT) signaling pathway, and genetic variants in TCF7L2 are associated with type 2 diabetes[25]. Early NCC migration is regulated by non-canonical Wnt signaling and dysregulation of non-canonical Wnt signaling inhibits NCC migration[26] Moreover TCF7L2 (previously known as TCF4) is also associated with Pitt-Hopkins syndrome (PHS) a rare syndromic form of HSCR[27].

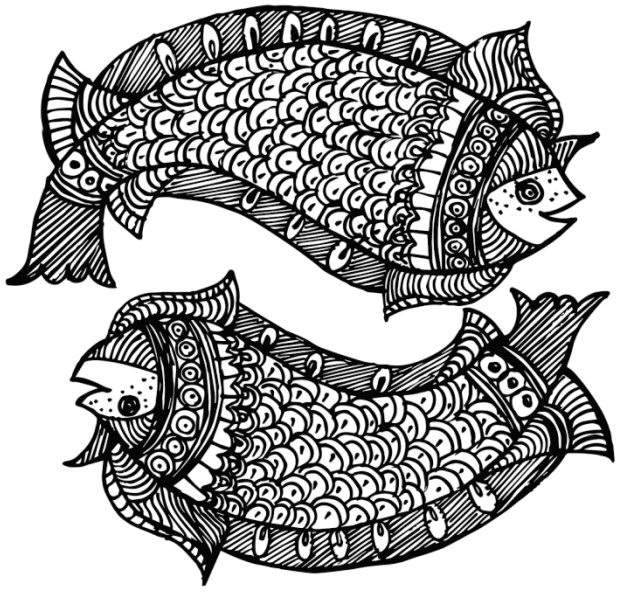
Genetic variants identified by GWAS usually explain only a small fraction of complex disease susceptibility with limited success in explaining genetic variance and its relation to phenotypic variability (giving rise to the concept of 'missing heritability')[28] Epigenomics facilitates interpretation of previously unsolved GWAS studies and many new tools and bioinformatics approaches have also been developed to understand functional relevance of non-coding mutations[29]. Our studies offer datasets for functional follow-up of GWAS loci, including fine mapping of GWAS signal(s), prioritization of putative disease causing SNPs by looking into enhancer mutations, which could potentially dysregulate ENS gene expression and contribute to HSCR development. It also offers genomic coordinates to screen non- coding modifier mutations or rare or private variants associated with HSCR.

ACKNOWLEDGEMENTS:
This study was supported by research grant from ZonMW (TOP-subsidie 40-00812-98-10042).

REFERENCES

1. Badner, J.A., Sieber, W.K., Garver, K.L. & Chakravarti, A. A genetic study of Hirschsprung disease. *Am J Hum Genet* **46**, 568-80 (1990).
2. Alves, M.M. *et al.* Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev Biol* **382**, 320-9 (2013).
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**, 581-96 (2015).
4. Emison, E.S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857-63 (2005).
5. Sribudiani, Y. *et al.* Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology* **140**, 572-582 e2 (2011).
6. Bolk, S. *et al.* A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. *Proc Natl Acad Sci U S A* **97**, 268-73 (2000).
7. Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M.B. Annotating non-coding regions of the genome. *Nat Rev Genet* **11**, 559-71 (2010).
8. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327-39 (2011).
9. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr Biol* **20**, R754-63 (2010).
10. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049-59 (2015).
11. Lettice, L.A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-35 (2003).
12. Weedon, M.N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* **46**, 61-4 (2014).

13. Postma, A.V., Bezzina, C.R. & Christoffels, V.M. Genetics of congenital heart disease: the contribution of the noncoding regulatory genome. *J Hum Genet* **61**, 13-9 (2016).
14. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* **533**, 95-9 (2016).
15. Weinhold, N., Jacobsen, A., Schultz, N., Sander, C. & Lee, W. Genome-wide analysis of noncoding regulatory mutations in cancer. **46**, 1160-5 (2014).
16. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
17. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* **33**, 364-76 (2015).
18. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
19. Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42**, D142-7 (2014).
20. Lelli, K.M., Slattery, M. & Mann, R.S. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet* **46**, 43-68 (2012).
21. Ong, C.T. & Corces, V.G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* **12**, 283-93 (2011).
22. Jochum, W., Passegue, E. & Wagner, E.F. AP-1 in mouse development and tumorigenesis. *Oncogene* **20**, 2401-12 (2001).
23. Ichihara, M., Murakumo, Y. & Takahashi, M. RET and neuroendocrine tumors. *Cancer Lett* **204**, 197-211 (2004).
24. Heanue, T.A. & Pachnis, V. Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. *Proc Natl Acad Sci U S A* **103**, 6919-24 (2006).
25. Hindy, G., Mollet, I.G., Rukh, G., Ericson, U. & Orho-Melander, M. Several type 2 diabetes-associated variants in genes annotated to WNT signaling interact with dietary fiber in relation to incidence of type 2 diabetes. *Genes Nutr* **11**, 6 (2016).
26. De Calisto, J., Araya, C., Marchant, L., Riaz, C.F. & Mayor, R. Essential role of non-canonical Wnt signalling in neural crest migration. *Development* **132**, 2587-97 (2005).
27. Amiel, J. *et al.* Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am J Hum Genet* **80**, 988-93 (2007).
28. Hu, J.X., Thomas, C.E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* (2016).
29. Tak, Y.G. & Farnham, P.J. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57 (2015).



CHAPTER 5

Overexpression of the chromosome 21 gene *ATP5O* results in fewer enteric neurons: the missing link between Down syndrome and Hirschsprung disease?

Rajendra K. Chauhan[1], Rizky Lasabuda[1], Duco Schriemer[2], William W.C. Cheng[1], Zakia Azmani[1], Bianca M. de Graaf[1], Alice S. Brooks[1], Sarah Edie[3], Roger H. Reeves[3], Bart J.L. Eggen[2], Alan J.Burns[1,5], Iain T. Shepherd[4], Robert M.W. Hofstra[1,5]

1 Department of Clinical Genetics, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands
2 Department of Neuroscience, section Medical Physiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
3 Johns Hopkins University School of Medicine, Department of Physiology and McKusick-Nathans Institute for Genetic Medicine, Baltimore, USA
4 Department of Biology, Emory University, Atlanta, USA
5 Stem Cells and Regenerative Medicine, UCL Great Ormond Street Institute of Child Health, London, UK

Manuscript in preparation

ABSTRACT

Hirschsprung disease (HSCR) is characterized by the absence of enteric ganglia in the distal region of the gastrointestinal tract, leading to severe intestinal obstruction. Around 12% of patients with HSCR have a chromosomal abnormality, the most of which have Down Syndrome (DS), trisomy 21. Moreover, individuals with DS have a >100 fold higher risk of developing HSCR than the general population. This suggests that overexpression of human chromosome 21 (Hsa21) genes contribute to the etiology of HSCR. To identify the gene(s) contributing to HSCR in DS, we overexpressed candidate genes in a reporter zebrafish, *Tg(-8.3bphox2b:Kaede)* where neural crest derived cells express the fluorescent kaede protein. We prioritized 21 genes and overexpressed them by microinjecting capped mRNAs in single-cell stage zebrafish embryos and scored them at 5 days post fertilization (dpf). We show that overexpression of *ATP5O (*ATP synthase, H^+ transporting, mitochondrial F1 complex, O subunit*)* leads to a disturbed enteric nervous system (ENS) with a reduced number of enteric neurons, strongly implicating *ATP5O* as a contributor to a HSCR phenotype. The *ATP5O* gene encodes a component of the F-type ATPase found in the mitochondrial matrix and participates in ATP synthesis coupled proton transport. *ATP5O* does not link to the known HSCR pathways, and although we show expression of the protein in the enteric ganglia, its involvement in disease development and ENS development is yet to be uncovered.

INTRODUCTION

Hirschsprung disease (HSCR, MIM #142623) is a complex congenital gut motility disorder resulting from a failure in the development of the enteric nervous system (ENS) of the gastrointestinal (GI) tract. It is characterized by the absence of enteric ganglia in a variable length of the distal gut. HSCR is recognized by a failure to pass meconium in the first 48 hr after birth, abdominal distention, vomiting, and neonatal enterocolitis. It leads to severe intestinal obstruction and life threatening constipation. The prevalence of HSCR is 1 in 5000 live births and there is an unexplained sex bias of four males to one female[1]. The lack of neurons in the distal part of the GI tract results from a failure of enteric neural crest cells (NCC) to migrate, differentiate, proliferate or survive and thereby colonize the gut and form a functional network of neurons and glia (reviewed by Sasselli *et al*., 2012[2]).

HSCR is considered as an inherited disease, based on the fact that there are familial cases (~5%), and the 200-fold increased risk of HSCR to siblings of patients[3]. Highly penetrant, coding mutations, in approximately 15 genes, have been identified to cause or contribute to HSCR (for review see[4]). The major gene in HSCR is *RET*, with a mutation prevalence of 50% in familial HSCR and 15% in sporadic HSCR[5,6]. However, cumulatively all the mutations in HSCR-associated genes explain only a small fraction of cases. In addition to the high penetrant coding mutations, common low-penetrance polymorphic variants at *RET*, in the region containing *SEMA3C/SEMA3D* and in *NRG1* are also associated with HSCR[7-10]. However, all together the heritability of the vast majority (~80%) of HSCR cases is still to be uncovered. Finding genetic factors that may explain the missing heritability could come from analysis of known HSCR linkage regions, syndromic HSCR cases, or from the chromosomal abnormalities often identified in HSCR patients.

HSCR is associated with chromosomal abnormalities in 12% of all cases. In this study we focused on the most common chromosomal abnormality found in HSCR, Trisomy 21. Trisomy 21, leading to Down Syndrome (DS), is the most frequent cause of learning difficulties with an incidence of 1 in 750 live births[11]. The incidence of DS among HSCR patients ranges from 2% to 10%[3]. Moreover, DS patients have >100 fold higher risk of developing HSCR than the general population[3]. This suggests that overexpression of one or more genes on chromosome 21 may have a substantial contribution to HSCR development in DS

associated HSCR cases. However, none of the established HSCR genes are localized on chromosome 21. Existing animal models for DS have not, as yet, been explored in detail for any ENS related defects and despite the vast knowledge available, this association still remains poorly understood.

Here we aimed to identify the gene(s) on chromosome 21 that could contribute to the HSCR phenotype. We injected mRNA of selected Hsa21 genes into a transgenic zebrafish reporter model, and found that elevated levels of one of the chromosome 21 genes, *ATP5O*, resulted in altered ENS development and a HSCR-like phenotype. Moreover, we show that *ATP5O* is expressed in the zebrafish gut and in the myenteric and submucosal ganglia of human postnatal colon sections.

METHODS

Prioritizing Hsa21 candidate HSCR Genes

In this study we first prioritized candidate genes based on genetic data and literature. The genetic data we used was: conservation of the genes between human and mouse[12,13]; expression of the genes in mouse enteric NCC (in-house RNA sequencing data); whether genes encode transcription factors and; presence of the genes in segmental duplicated regions of chromosome 21 in DS/HSCR patients[14]. In our literature search we took in consideration: previous studies on associations between DS and HSCR; Hsa21 genes that are involved in ENS and gut development; genes related to neuronal development; genes involved in neuronal signaling; known animal models of DS.

Hsa21 clone sets

To be able to microinject capped human mRNAs into 1-cell stage *Tg(-8.3bphox2b:Kaede)* zebrafish, the set of prioritized Hsa21 genes were sub-cloned in pCS2+ and were grown overnight followed by plasmid isolation and purification using the NucleoBond® Xtra plasmid purification system (Marchery-Nagel, Nagel, 2012). All the constructs were verified by DNA sequencing. A pSG5-hu*APP*-695 construct was used for the *APP* clone.

Zebrafish husbandry and strains

The *Tg(-8.3bphox2b:Kaede)* zebrafish line expresses the fluorescent Kaede protein in *phox2b* expressing cells, including those of the ENS[15]. *ret*$^{sa2684/+}$

zebrafish line was obtained directly from Zebrafish International Resource Center (ZIRC)[16]. Both the *Tg(-8.3bphox2b:Kaede)* and *ret*$^{sa2684/+}$ zebrafish lines were maintained by pairwise mating. A cross between *ret*$^{sa2684/+}$ and *Tg(-8.3bphox2b:Kaede)* was performed to generate *Tg(-8.3bphox2b:Kaede); ret*$^{sa2684/+}$ fish. Zebrafish were maintained at 28°C according to the standard zebrafish laboratory protocols[17]. Embryos were scored for ENS defects and abnormal phenotypes at 5 dpf as described below. The institutional review board for experimental animals of Erasmus MC, Rotterdam approved the use of zebrafish embryos for this study. All procedures and fish experiments were performed in accordance with Dutch animal welfare legislations and those of the Erasmus Dierexperimenteel Centrum (EDC).

In vitro transcription of mRNA and microinjections into zebrafish embryos

In order to generate capped mRNA for microinjections, the plasmids were linearized with an appropriate restriction enzyme. After digestion, the plasmid DNA was cleaned using a phenol chloroform extraction method followed by ethanol precipitation. The linearized plasmids were used for *in vitro* synthesis of capped mRNA using the mMEssage mMachine SP6 kit (Ambion Inc., AM1340). Total RNA was purified using the RNeasy mini kit (Qiagen,Inc., 74104) and loaded on a 2% agarose gel to assess RNA quality and integrity. Capped mRNA quantification was done using the Nanodrop8000 (Thermo). RNA samples were stored at -80°C. Capped mRNA was diluted in nuclease free water and microinjections were done in 1-cell stage zebrafish embryo to overexpress the genes, as described previously[18]. Different dosages of each mRNA (5pg, 10pg, 50pg, 100pg, 150pg, 200pg and 250pg) were injected to determine their effect on ENS development. The mRNA-injected animals were raised in E3 media until 5dpf at 28°C. Non-injected control (NIC) embryos served as positive controls for survival.

Imaging and neuronal counting in zebrafish

Zebrafish embryos injected with capped mRNA and NIC were scored using a Leica MZ16FA microscope for any visible phenotype under bright field and by using a GFP filter to image phox2b-positive enteric NCC. To analyze the zebrafish embryos, they were anesthetized using tricaine in E3 media and then mounted on 0.3% agarose gel for capturing the images. The digital images were made using a Leica MZ16FA microscope. Fluorescent imaging was made under the same

settings for each image. Images were processed with Leica LAS and Adobe Photoshop CS software. To count the number of enteric neurons, we used an in-house made algorithm with image analysis software from FIJI in a semi-automated way (Figure 1C, D).

Whole mount *in situ* hybridization in zebrafish

A fragment of 686bp of zebrafish *atp5o* cDNA was amplified using RT-PCR using primer pair 5'-TTTCATCCCAGACCAGTACG-3' (forward) and 5'-GGTATCCCTGATCAGCTTGG-3' (reverse). The amplified PCR product was ligated directly into the pCR®II-TOPO vector using the Dual Promoter TA cloning kit (Invitrogen). Positive clones were confirmed by DNA sequencing for the orientation and the correct sequence and used to generate antisense and sense

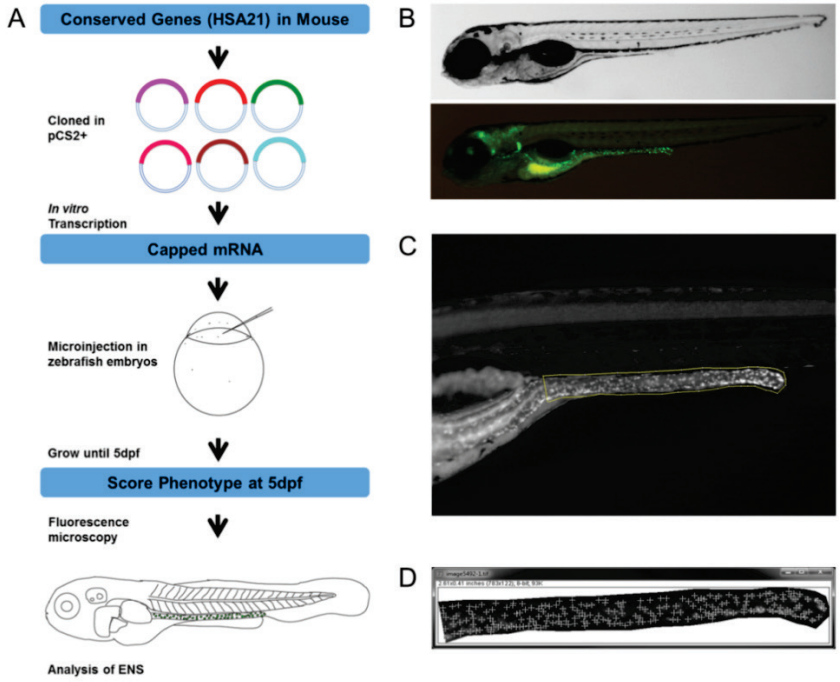


Figure 1. Schematic overview of experimental procedure. A) Schema for overexpression of prioritized candidate genes from Hsa21. B) *Tg(-8.3bphox2b:Kaede)* zebrafish line in bright field and under GFP filter, *phox2b* expressing neural crest cells are marked with fluorescent kaede protein. C) The enteric neurons of the intestinal region corresponding to 8 myotomes from the urogenital opening were selected as shown. D) Using FIJI software, the enteric neurons were counted as represented in the picture.

probes to detect *atp5o* mRNA expression. Whole-mount *in situ* hybridization was carried out as previously described[19]. We used the DIG RNA labelling kit (Roche) to generate digoxygenin-labeled riboprobes against *atp5o.* Stained embryos were mounted in 70% glycerol. The images were acquired using a Leica MZ16FA microscope.

Zebrafish genotyping

Tg(-8.3bphox2b:Kaede); ret$^{sa2684/+}$ embryos were grown until 5dpf for phenotyping. DNA was extracted from individual embryos and genotyping PCR was performed to distinguish mutants from wildtype using the gene-specific primers *ret*-wt-F1 (5'GATCTCGTTCGCCTGGC3'), *ret*-mut-F1 (5'GATCTCGTTCGCCTGGT3') and *ret*-wt- R1 (5'GGGGGCGTGTGACTAATTT3').

Immunohistochemistry on human colon material

Control postnatal human colon tissues were obtained from the Pathology Department repository of the Erasmus University Medical Center. Immunohistochemical (IHC) staining was performed using the Ventana Benchmark Ultra automated staining system (Ventana Medical System, Tuscon, AZ, USA). Briefly, after deparaffination the sectioned specimens for IHC detection of *ATP50* were processed for 60 min antigen retrieval using Cell Conditioning Solution (CC1, Ventana 950-124). After 30 minutes incubation with the primary antibody at 36˚C (ATP5O 1:200), detection with UltraView Universal DAB detection kit (Ventana 760-500) was performed after amplification with Ultraview amplification kit (Ventana 760-080). The sections were counterstained with hematoxylin II (Ventana 790-2208).

Epistasis between *ATP5O* and *ret* in zebrafish

1 ng of translation-blocking antisense morpholino against *ret*[20] and 50 pg of *ATP5O* capped mRNA were co-injected in 1 cell-stage *Tg(-8.3bphox2b:Kaede)* embryos. Embryos injected with either *ret* morpholino or *ATP5O* capped mRNA served as controls. At 5 dpf, embryos were imaged and enteric neurons present in the three myotome-length long, distal-most intestine were counted and compared.

Cell culture and transfections

The SK-N-SH Neuroblastoma cell line (ATCC # HTB-11) was cultured according to the ATCC's protocol (LGC Standards, Middlesex, UK) and incubated at 37°C, supplied with 5% of CO_2. Approximately 10^6 cells were cultured in 1 well of a 6-wells plate for 24 hr prior to transient transfections. Cells were transfected with 1µg of DNA construct containing *ATP5O* (pCS2+/*ATP5O*) or empty vector (pCS2+) and we used untransfected (UT) cells as a negative control. Transfections were done using 4µl GeneJuice transfection reagent (Novagen, 70967, Millipore) according to the manufacturer's instructions. Cells were starved in serum free media for 48 hr prior to harvesting and analysis.

Cell apoptosis and cell proliferation assay

Cell apoptosis was assessed by FACS analysis using PE Annexin V Apoptosis Detection Kit I (BD Pharmingen™) as per manufacturer's instructions. Cells were washed with PBS. Early apoptotic cells were identified as PE Annexin V-positive and 7AAD-negative, while cells positive for both, PE Annexin V and 7AAD were marked as apoptotic cells. For cell cycle staining assays, ethanol fixed cells were stained with propidium iodide (PI) for 30 min at room temperature. Stained cells were analyzed on a FACS flow cytometer (BD Biosciences, San Jose, CA) and for both assays data analysis was performed using FlowJo.

Statistical analysis

Results are presented as means ± standard deviation (SD). Data were analyzed by unpaired two-tailed t-test (comparisons of two groups) for the statistical significance.

RESULTS

Prioritization of candidate genes and generation of cDNA clones

To test which gene(s) on chromosome 21 contribute(s) to HSCR in DS patients, a selection of the most promising candidate genes was made. A total of 169 genes were initially assembled for screening. They consisted of 149 Hsa21 genes that are conserved between human and mouse and another 20 genes that are non-conserved, but are potentially interesting, and human specific[12,13]. From these 169 candidates we selected genes encoding transcription factors, genes involved

in neuronal development and genes reported as involved in DS with or without gut abnormalities (such as HSCR). Following these criteria, we generated a subset of 65 candidate genes and among them we further prioritized the genes based on their expression in E14.5 mouse enteric NCC (in-house RNA sequencing data), and based on functional evidence from studies in other model organisms. This pipeline resulted in a shortlist of 28 genes (Table 1) and we were able to synthesize 21 capped mRNAs (technical difficulties made us exclude 7 genes). A list of prioritized genes is presented in Table 1. A schematic of the experimental design is shown in Figure 1A.

Table 1. List of prioritized 28 genes in Hsa21 for overexpression.

No.	Gene Name	Gene Start (hg19) (bp)	Accession number	Conservation in zebrafish	Microinjection status
1	*APP*	27252861	NM_201414	Yes	Yes
2	*ATP5O*	35275757	NM_001697	Yes	Yes
3	*BACH1*	30566392	BC063307	Yes	Yes
4	*BRWD1*	40556102	NM_001007246	Yes	Yes
5	*BTG3*	18965971	NM_001130914	Yes	Yes
6	*CBR1*	37442239	NM_001757	Yes	Yes
7	*CHAF1B*	37757676	NM_005441	Yes	Yes
8	*CHODL*	19165801	NM_024944	Yes	Yes
9	*DSCAM*	41382926	AB384859	Yes	Yes
10	*DYRK1A*	38739236	BC156309	Yes	Yes
11	*HMGN1*	40714241	NM_004965	No	Yes
12	*PCBP3*	47063608	BC012061	Yes	Yes
13	*PDE9A*	44073746	NM_001001567	Yes	Yes
14	*PIGP*	38435146	NM_153681	Yes	Yes
15	*PKNOX1*	44394620	NM_004571	Yes	Yes
16	*RCAN1*	35885440	BC002864	Yes	Yes
17	*RUNX1*	36160098	BC069929	Yes	Yes
18	*SH3BGR*	40817781	NM_001001713	Yes	Yes
19	*SIM2*	38071433	NM_005069	Yes	Yes
20	*SOD1*	33031935	NM_000454	Yes	Yes
21	*SUMO3*	46191374	NM_006936	Yes	Yes
22	*DSCR3*	38591910	BC110655	Yes	No
23	*ETS2*	40177231	NM_005239	Yes	No
24	*HLCS*	38123493	NM_000411	Yes	No
25	*ITSN1*	35014706	BC116186	Yes	No
26	*TIAM1*	32361860	BC117196	Yes	No
27	*TTC3*	38445571	BC137345	Yes	No
28	*WRB*	40752170	NM_004627	Yes	No

Overview of the 28 prioritized genes for overexpression. mRNA was injected into the *Tg(-8.3bphox2b:Kaede)* zebrafish for the first 21 genes. The last 7 genes were omitted due to failed mRNA generation.

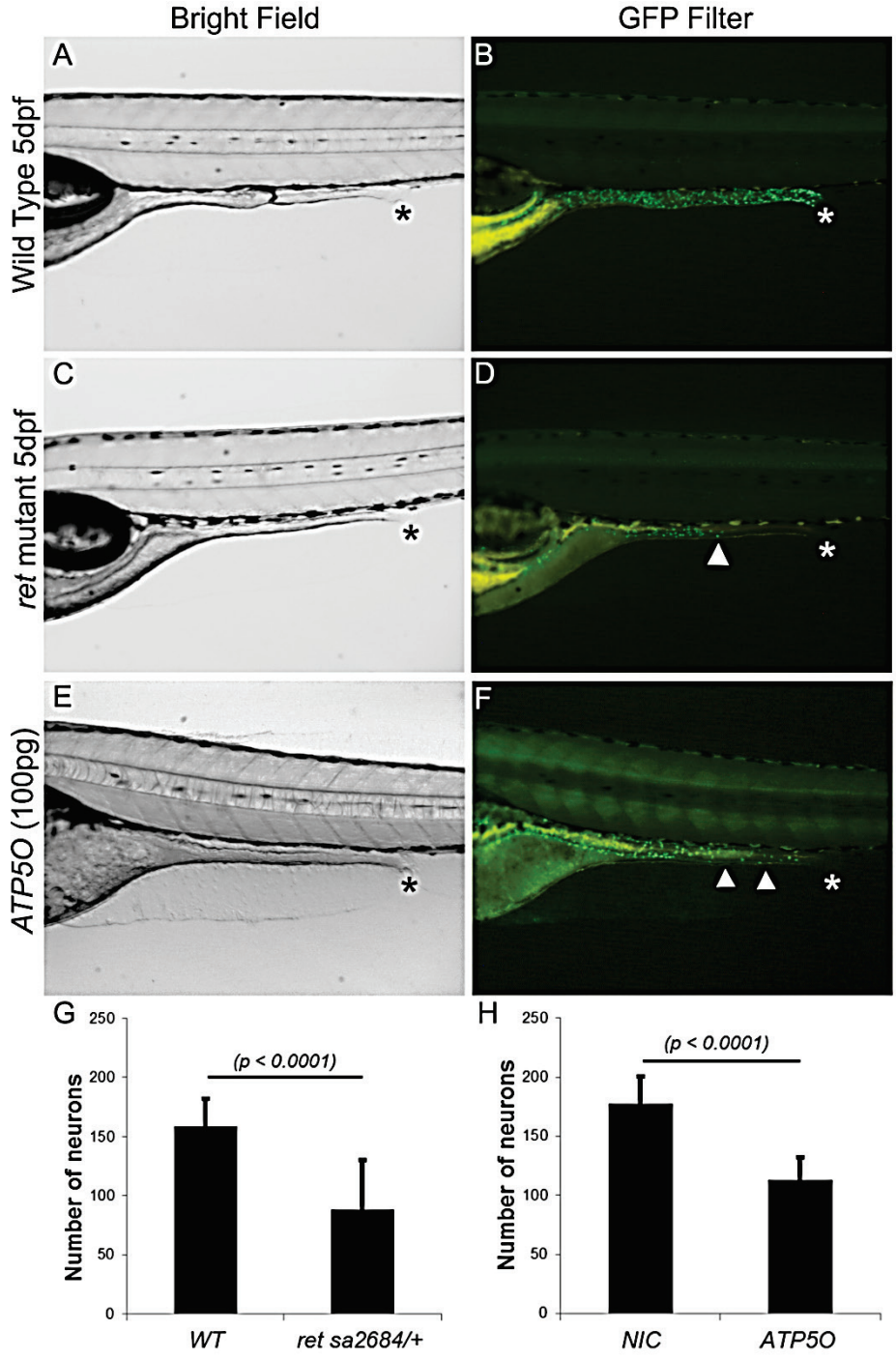
Tg(-8.3bphox2b:Kaede); ret$^{sa2684/+}$ mutants display ENS defect

In humans, loss of function mutations in the *RET* gene result in HSCR. In this study we used the *ret*$^{sa2684/+}$ zebrafish that was identified in an ENU mutagenesis project as a positive control for a HSCR-like phenotype[16]. The *ret*$^{sa2684/+}$ line was crossed with the *Tg(-8.3bphox2b:Kaede)* reporter zebrafish line and the number of enteric neurons was scored at 5dpf followed by genotyping. The *ret*$^{sa2684/+}$ mutant embryos contained significantly less enteric neurons in the gut, indicating an HSCR-like phenotype, when compared to control animals (Figure 2A-D). The quantification of enteric neurons, corresponding to 8 myotomes from the urogenital opening, demonstrated a significant reduction in number of enteric neurons in *ret*$^{sa2684/+}$ fish (88 ± 41) compared to WT fish (158 ± 23) ($p<0.0001$, Figure 2G). These data show that the *Tg(-8.3bphox2b:Kaede)* is a suitable animal model for HSCR-like aganglionosis.

Overexpression of selected candidate gene mRNA in a zebrafish model

Capped mRNAs of the 21 selected Hsa21 genes were injected into *Tg(-8.3bphox2b:Kaede)* zebrafish (Figure 1A,B), which were subsequently examined at 5 dpf (as described in the Methods section). The mRNA dosage was titrated in a range of 5pg to 250pg, to find the optimal dosage for each mRNA based on the lethality and phenotype observed. Injections of mRNAs resulted in normal ENS phenotypes for all mRNAs, except one. Only when overexpressing *ATP5O* (100pg), a reduction in the number of enteric neurons was observed along the entire intestine with normal gross morphology when compared to the non-injected controls (Figure 2E,F).

Figure 2. Reduced numbers of enteric neurons in the *ret*$^{sa2684/+}$ mutant fish and ATP5O mRNA injected fish. A,B) The control *Tg(-8.3bphox2b:kaede)* fish at 5 dpf in the bright field and under GFP filter showing fluorescently tagged *phox2b* expressing cells and the gut is completely colonized with enteric neurons until the urogenital opening. The asterisk indicates the urogenital opening. C,D) Enteric neurons along the gut of the *Tg(-8.3bphox2b:Kaede); ret*$^{sa2684/+}$. Heterozygous *ret* mutant displayed less neurons in the gut at 5dpf and discontinuity of colonization of the gut is indicated by arrowhead. E,F) The zebrafish injected with *ATP5O* mRNA at 100pg dosage show less enteric neurons in the gut. G) Quantification of enteric neurons in the intestine corresponding to 8 myotomes of 5dpf *ret* mutant compared to the control fish, marked significant reduction. H) Quantification of enteric neuronal count of non-injected controls compared to the *ATP5O* overexpressed embryos. A total of 37.5 % of embryos injected with 100pg of *ATP5O* displayed reduction in enteric neurons.


Bright Field
GFP Filter
A
B
C
D
E
F
Wild Type 5dpf
ret mutant 5dpf
ATP5O (100pg)
G
H
Number of neurons
(p < 0.0001)
WT
ret sa2684/+
NIC
ATP5O
0
50
100
150
200
250

The percentage of zebrafish displaying reduction in enteric neurons remained similar at higher dosage (150pg). Counting the enteric neurons within the gut, corresponding to 8 myotomes from the urogenital opening, the average count for the controls was 178 ± 23 neurons (Figure 2H). We classified a gut as hypo-neuronal when the fish contained 2 SD less enteric neurons compared to the average control zebrafish. For the fish injected with *ATP5O* mRNA we found that 37.5% (15/40) of the embryos displayed such a reduction in the number of enteric neurons in the gut. The enteric neuron count in the affected embryos displaying reduced enteric neurons was 113 ± 18 (Figure 2H), showing that elevated levels of *ATP5O* interfere with normal development of the ENS.

Expression of *atp5o* in zebrafish

Whole mount *in situ* hybridization (ISH) was used to determine the spatio-temporal expression pattern of *atp5o* between 1dpf and 5dpf of zebrafish development. RNA *in situ* hybridization revealed expression of *atp5o* in different organs at different developmental stages. At 1dpf, *atp5o* expression was seen ubiquitously (Figure 3B,C). At 2dpf the expression was restricted to the cerebellum, the otolith and the whole gut (Figure 3E,F). Between 3dpf to 5dpf *atp5o* was predominantly expressed in the intestine and cerebellum (Figure 3E,F,H,I,K,L,N,O). At 5dpf, high *atp5o* expression was observed in the proximal and mid intestine along with the caudal vein (Figure 3N). The sense probe did not show any staining at 1dpf – 5dpf stages (Figure 3A,D,G,J,M), confirming the specificity of the probe.

Expression of *ATP5O* in postnatal human colon

To assess whether *ATP5O* is also expressed in the human colon, immunohistochemistry was performed on postnatal colon from healthy individuals. *ATP5O* was specifically detected in the ganglia present in the submucosal (Figure 4A,B) and myenteric plexuses (Figure 4C,D). In addition, *ATP5O* was also detected in the colon epithelium. These results suggest that *ATP5O* may be important for ENS development in humans as well.

***In vitro* assays for cell apoptosis and cell cycle analysis**

To examine whether *ATP5O* overexpression affects early apoptosis or the cell cycle and thereby leads to less neurons in zebrafish gut, we used a human

neuroblastoma cell line (SK-N-SH) and assayed cell apoptosis and cell cycle. SK-N-SH cells expressing *ATP5O* were cultured in the absence of serum for 48 hr and tested for both alterations in apoptosis and cell cycle. Flow cytometry analysis didn't indicate any significant changes in the early and late apoptosis in cells expressing *ATP5O*, when compared to other conditions (Supplementary Figure 1A).

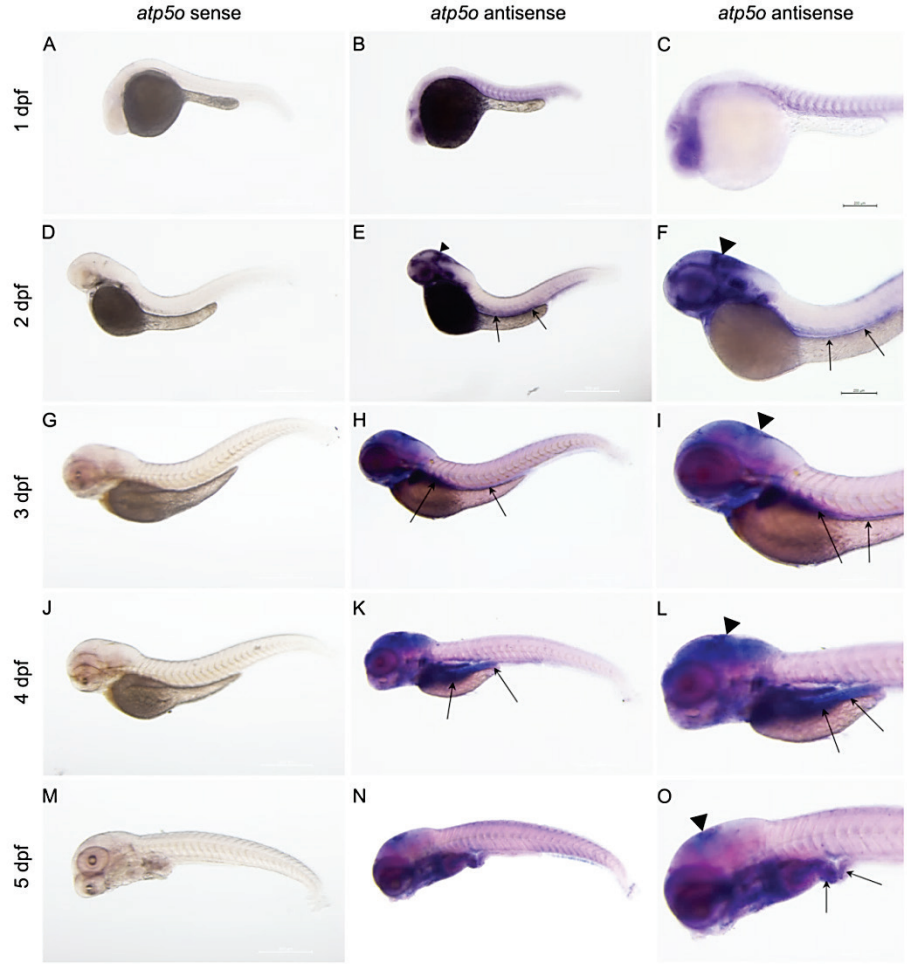


Figure 3. Spatio-temporal expression of *atp5o* in zebrafish. *atp5o* expression at indicated developmental stages ranging from 1–5dpf in zebrafish embryos (lateral view) detected using ISH. *atp5o* is expressed along the GI tract in all the stages. It is expressed ubiquitously at 1dpf (B,C) and the expression becomes restricted to cerebellum, otolith and whole gut by 2dpf (E,F). Arrowheads indicate expression in the brain and arrow marks indicate expression in the gastrointestinal tract (E, F, H, I, K, L, N, O). The sense probe shows no staining at 1–5dpf developmental stages as shown (A, D, G, J, and M).

In order to identify the impact of *ATP5O* overexpression on the cell cycle using flow cytometry, we observed a slight increase in the fraction of cells in the G1 phase as a result of *ATP5O* overexpression (Supplementary Figure 1B). Although the observed difference is not statistically significant, it could indicate that there can be some effect on the cell cycle arrest in *ATP5O*-overexpressing cells.

Epistasis between *ATP5O* and *ret* in zebrafish

To investigate whether *ATP5O* interacts with *RET* during the development of the ENS and in the pathogenesis of HSCR in DS, we knocked down *ret* and overexpressed *ATP5O* simultaneously by co-injecting *ret* translation-blocking morpholino (1 ng) and *ATP5O* capped mRNA (50 pg) and compared enteric neurons in distal intestine at 5 dpf to controls injected with either *ret* morpholino or *ATP5O* mRNA alone (Figure 5). The doses were chosen so that neither was sufficient to induce severe ENS defect by itself, and any synergistic effect between *ret* knockdown and *ATP5O* overexpression would readily be observed. The *ret* morpholino caused a mild decrease in enteric neuron number in the distal intestine compared to *ATP5O* mRNA control. However, co-injection of *ret* morpholino and *ATP5O* mRNA did not result in further significant reduction, suggesting limited or no synergistic effect between *ret* knockdown and *ATP5O* overexpression.

Other phenotypic effects of injection of *DSCAM* and *SIM2* mRNA

The use of this zebrafish model and its optical transparency allowed us to detect other gross developmental abnormalities upon overexpression of the prioritized genes. Injection of two candidate genes (*SIM2* and *DSCAM*) resulted in an abnormal phenotype. Overexpression of *SIM2* (100 pg) resulted in notochord defects in 66% of the injected embryos at 5dpf (Figure 6A,C) and 33% among them also displayed craniofacial abnormalities (Supplementary figure 2A,B). Overexpression of *DSCAM* (200 pg) resulted in deformed notochord and myotomes in 68% of the embryos at 5dpf (Figure 6B,D). The majority of these embryos also lacked the swim bladder. Microinjection of higher dosages (>200 pg for *DSCAM* and >100 pg for *SIM2*) of these mRNAs induced lethality.

DISCUSSION

This study reports a role for a chromosome 21 gene, *ATP5O*, in the development of the ENS in zebrafish using an mRNA overexpression screen, as overexpression of *ATP5O* results in reduced numbers of enteric neurons in the zebrafish gut. This phenotype is comparable to that of the *ret*$^{sa2684/+}$ zebrafish line that carries a mutation in *ret*, a known HSCR gene in humans. This makes us hypothesize that elevated levels of *ATP5O*, as likely in the case of DS, could contribute to the high prevalence of HSCR among DS patients.

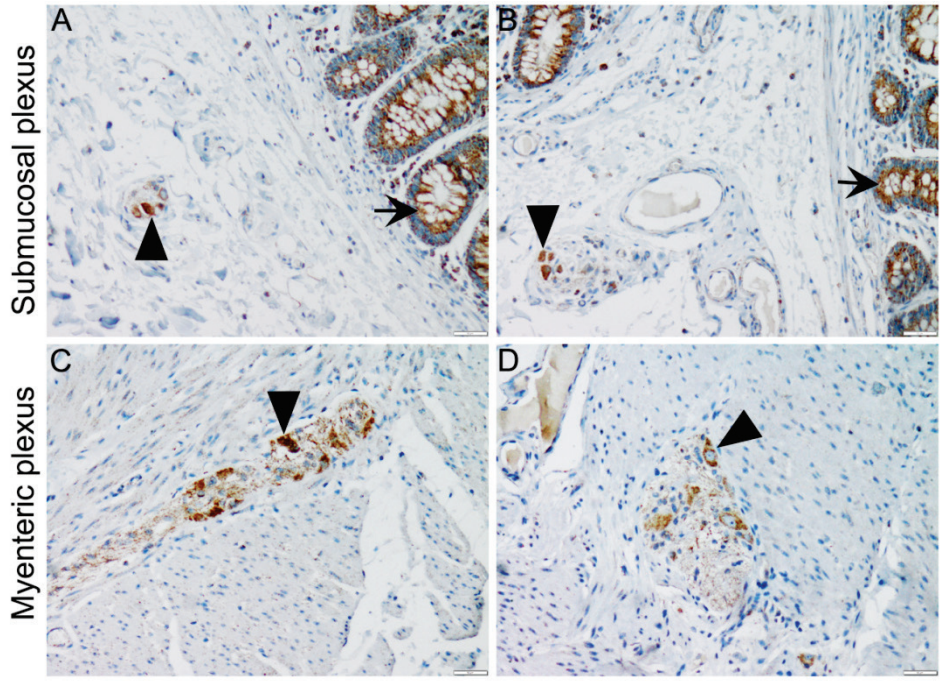


Figure 4. Expression of *ATP5O* in postnatal human colon. Expression of *ATP5O* detected by immunohistochemistry on paraffin embedded post-natal colon sections. Arrowheads indicate expression of *ATP5O* in submucosal plexus (A, B) and myenteric plexus (C, D). *ATP5O* is also expressed in the gut epithelia as shown by arrows (A, B).

The zebrafish as a model organism for human enteric neuropathies

The intestinal architecture and anatomy of zebrafish closely resembles that of mammals[21]. The zebrafish gut undergoes rapid development and by 5dpf the whole GI tract is functional[22]. In contrast to amniotes, the zebrafish gut is simpler,

it lacks submucosal layer and myenteric neurons are arranged as neuronal pairs or single neurons[21]. The zebrafish ENS is also derived from NCC, as in other vertebrates[23]. In zebrafish, NCC migrate as two parallel chains of cells to colonize the whole gut and differentiate into enteric neurons and glia[24]. Despite these differences, the organization of the ENS, which modulates functions such as motility, homeostasis and secretion, is comparable but less complex compared to mammals making it a good model for human GI diseases[25]. Previous studies have shown that perturbation of zebrafish orthologues of known human HSCR genes using morpholino mediated knockdown, but also some mutant zebrafish for genes not connected to HSCR, leads to loss of enteric neurons in zebrafish gut and recapitulates the human HSCR phenotype[10,20,24,26-29]. In particular *RET* is known to be the major player in HSCR and in ENS development[4,30]. For these reasons we included the *ret*$^{sa2684/+}$ mutant zebrafish line as positive control. Indeed when quantifying the number of neurons in the hindgut, the most distal part of zebrafish intestine, the region in which mostly the aganglionosis in HSCR patients is observed, the number of neurons in this mutant fish was reduced.

***ATP5O* overexpression results in reduced enteric neurons**

Microinjection of *ATP5O* mRNA resulted in reduced numbers of enteric neurons in the zebrafish gut comparable to what was found in case of the *Tg(-8.3bphox2b:Kaede); ret*$^{sa2684/+}$ zebrafish (Figure 2A-H). *ATP5O* was the only gene for which overexpression resulted in ENS defects. The fact that overexpression results in fewer enteric neurons might not be a real surprise as mouse *Atp5o* is highly expressed in mouse enteric NCC (in-house RNA sequencing data). *ATP5O* is also expressed in the ganglia of submucosal and myenteric plexuses as shown in our studies using control postnatal colon sections. Similarly, *atp5o* is also expressed in the zebrafish gut during early embryonic development and the ENS also forms during this period. Furthermore, meta-analysis of DS phenotypes in segmental trisomy's and its association with congenital gut abnormalities such as HSCR, duodenal stenosis and intestinal atresia suggested a critical GI region of <13 MB. This region also includes *ATP5O*[14]. Previous identity-by-descent (IBD) and association mapping in a large (inbred) Mennonite population also showed that *ATP5O* is within the IBD region associated with HSCR[31]. All these data suggest that *ATP5O* might well be responsible, or at least contribute to, the HSCR phenotype often seen in DS patients.

The role of ATP5O in HSCR

ATP5O is a mitochondrial gene, which encodes the ATP synthase H+ transporting, mitochondrial F1 complex, O subunit protein and is also known as Oligomycin Sensitivity Conferral Protein (OSCP). It is a component of ATP synthase (F(1)F(0) ATP synthase or Complex V) found in the mitochondrial matrix. ATP synthase is composed of an extramembranous catalytic core (F1) and a peripheral membrane proton channel (F0). The encoded protein appears to be part of the connector linking these two subunits and may be involved in transmission of conformational changes or proton conductance. It produces ATP from ADP via oxidative phosphorylation in the presence of a proton gradient across the mitochondrial membrane. Electron transport complexes of the respiratory chain generate this gradient[32,33]. The gene ontology (GO) annotation of *ATP5O* associates it with drug binding and transporter activity. It was hypothesized that overexpression of *ATP5O* could interfere with the normal subunit composition of

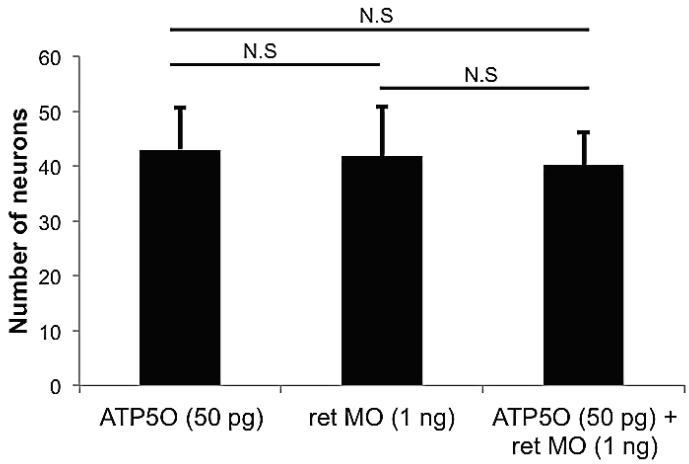


Figure 5. Epistasis between *ATP5O* and *ret*. Quantification of enteric neurons at 5dpf in the distal most intestine corresponding to 3 myotomes of zebrafish embryos for epistatic interaction between *ATP5O* and *ret*. Embryos were injected with *ATP5O* (50ng), *ret* MO (1ng) and a combination of both and the enteric neuronal count is plotted in the graph. There are no significant differences between *ATP5O* (50pg) vs 1 ng *ret* MO (p=0.6587), 1 ng *ret* MO vs *ATP5O* (50pg) + 1ng *ret* MO (p=0.5437) and *ATP5O* (50pg) vs *ATP5O* (50pg) + 1 ng *ret* MO (p=0.2146).

ATP synthase, resulting in an impairment of oxidative phosphorylation[34]. An imbalance of expression, as generated in our zebrafish, could potentially impair the subunit composition of ATP synthase, leading to oxidative phosphorylation disruption and eventual perturbed proliferation of these cells. We found a slight but not significant effect on cell cycle arrest (G1 phase) upon overexpression of

ATP5O in SK-N-SH cells. It has also been shown that disruption of oxidative phosphorylation can have a neurotoxic effect on neuronal progenitor cells[35], and overproduction of ATP synthase in *Escherichia coli* has already been implicated in cell division and growth[36]. During early ENS development, the enteric NCCs migrate, proliferate extensively and differentiate into neurons and glia. *ATP5O* overexpression could potentially affect enteric NCC proliferation and lead to fewer neurons in zebrafish gut, as observed in our experiments.

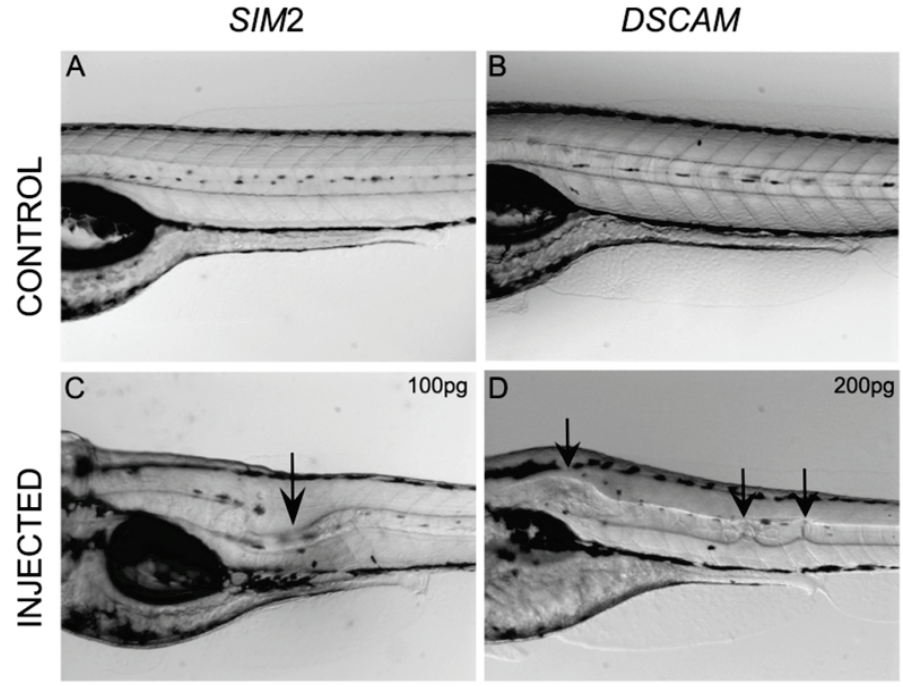


Figure 6. Notochord defects in *SIM2* and *DSCAM* mRNA injected zebrafish. Overexpression of *SIM2* and *DSCAM* lead to defects in the notochord, as represented by arrows in the bright field images of 5dpf embryos as compared to respective controls (A, B). The notochord is discontinuous and deformed in embryos in which *SIM2* and DSCAM are overexpressed (C, D).

ATP5O does not interact with *ret*

A previous study showed over-representation of the enhancer polymorphism *RET*+9.7 (rs2435357:C>T) in DS-HSCR[37]. The disease-associated allele was significantly different between individuals with DS alone, HSCR alone, and those with HSCR and DS, demonstrating an association and interaction between RET and chromosome 21 gene dosage. However, our zebrafish data did

not demonstrate any interaction between *ret* and *ATP5O* in ENS development, suggesting that they acted independently in separate pathways.

Additional phenotypes due to overexpression of Hsa21 genes

In this overexpression screen of 21 candidate genes from Hsa21, we also identified phenotypic defects other than that of the ENS for *DSCAM* and *SIM2*. Zebrafish *dscam* is highly expressed in the developing brain. It is thought to be involved in shaping the nervous system and early morphogenesis of the zebrafish embryo[38]. The expression pattern of *sim2* in zebrafish has been reported using whole mount ISH; it is expressed mainly in the diencephalon, the midbrain and the pharyngeal arches[39]. Overexpression of *DSCAM* and *SIM2* in zebrafish displayed defects mainly in notochord development and in the floor plate, exhibiting discontinuity with some twists and folds upon overexpression of these genes. The notochord is essential for proper vertebrate development by producing secreted factors that signal to the surrounding tissues. It is also important for specification of the ventral fates in the CNS and it plays an important role in patterning and in a proper structural integrity. The defects in notochord development are possibly due to uneven cell patterning or selective cell death or defects in signaling pathways required for normal notochord development (reviewed by Stemple, 2005[40]). Moreover, besides notochord defects we also observed craniofacial abnormalities on overexpression of *SIM2* in a subset of embryos at 5dpf along with notochord defects, (Supplementary Figure 2A, B) indicating its prospective contribution to the phenotype observed in DS affected individuals.

To our surprise, some candidates (such as *DSCAM*, *SIM2*, and *APP*) already associated with ENS phenotypes, based on previous genetic studies and murine models did not display any visible ENS phenotype following their overexpression in the zebrafish model. *DSCAM* has been highlighted as a predisposing locus to HSCR in patients with DS[14,31,41]. Our previous studies, using *in vitro* methods, have shown that overexpression of *SIM2* leads to a down regulation of the *RET* gene[8]. Similarly, a transgenic *APP* mouse model displayed reduction in myenteric neuronal density and delay in gut transit[42]. These are characteristic features of HSCR in humans, but we were not able to recapitulate similar phenotypes in the zebrafish model upon their overexpression. This could be due to the fact that the regulatory mechanism required for efficient translation of certain human RNAs was not equally efficient in zebrafish, or that the human

protein does not have the same effect as the zebrafish protein or alternatively the threshold dosage of RNA required resulting in a phenotype may not have been achieved in our study. Furthermore, we cannot rule out the presence of overexpressed mRNA in the embryos at critical stages of zebrafish ENS development or any unknown feedback mechanisms resulting in the net neutrality of overexpression. On the other hand, a phenotypic effect may simply require a combinatorial overexpression of more than one gene.

Within the list of 21 genes we did not include *COL6A4* although it was recently shown that overexpression of *Col6a4* in transgenic mice could lead to a HSCR-like phenotype[43]. The reason for not including it was the fact that we had not found any direct or indirect evidence for the involvement of *Col6a4* with ENS development nor did we see the gene being expressed in the mouse enteric NCC at E14.5 (in-house RNA sequencing data).

CONCLUSIONS

Although the association of DS with HSCR is well recognized, the causative link between them is not well understood. The majority of DS affected individual exhibit GI abnormalities[44], which might be related to abnormal ENS development. Here, we used a transgenic zebrafish line, whose ENS is marked with the fluorescent Kaede protein, to assay the functional effects of overexpression of Hsa21 candidate genes. We found that *ATP5O* affects ENS development in zebrafish. The use of a vertebrate model to find the missing link between DS and HSCR opens the door for larger screens and better understanding of this complex association.

ACKNOWLEDGEMENTS

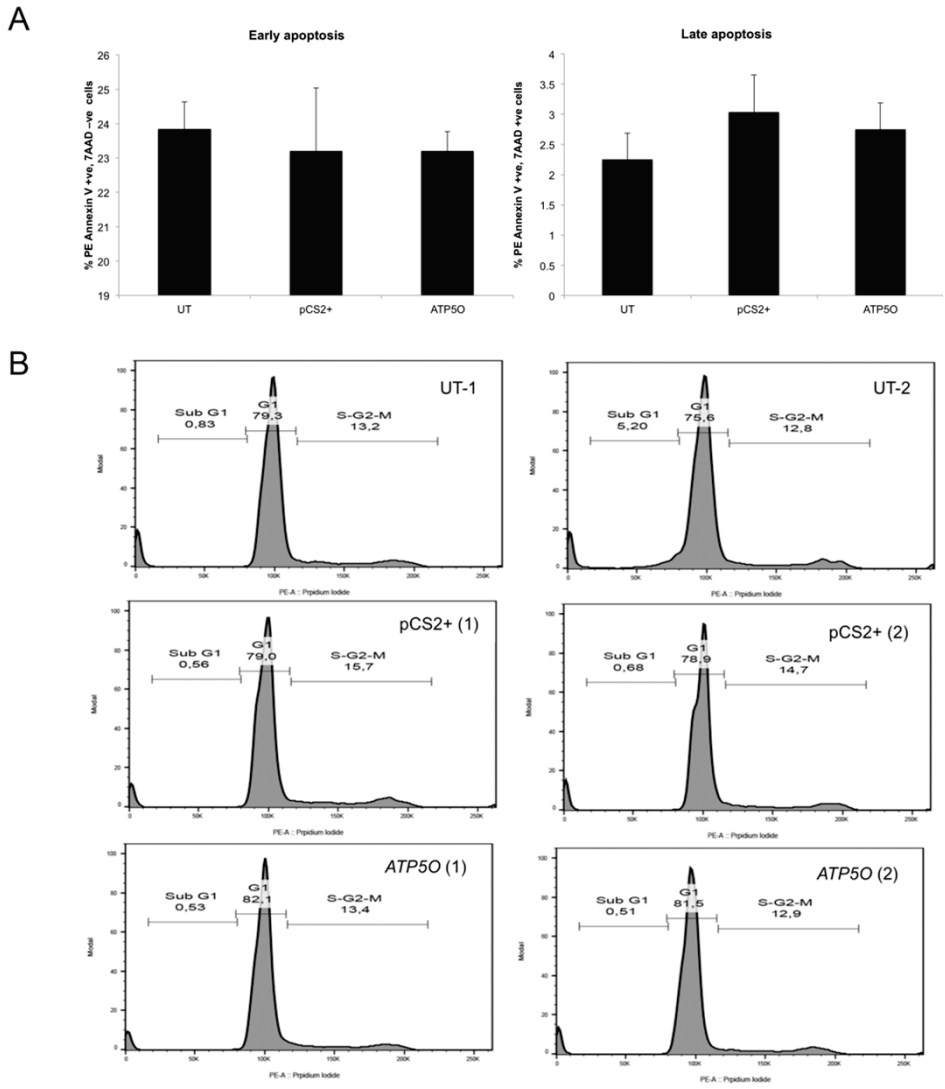
This study was supported by research grant from ZonMW (TOP-subsidie 40-00812-98-10042) and the Maag Lever Darm stichting (WO09-62). We thank Simar (Department of Immunology, Erasmus MC) for help with FACS. We also thank Prof. Bart de Strooper (Leuven, Belgium) for providing us with the *APP* plasmid; Herma van der Linde and Maria Alves for technical advice and helpful discussions.

REFERENCES

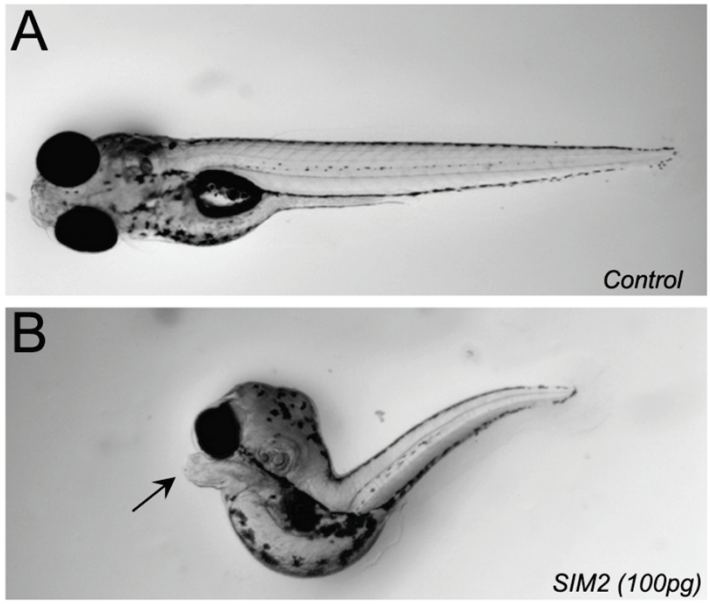
1. Badner, J.A., et al., *A genetic study of Hirschsprung disease.* Am J Hum Genet, 1990. **46**(3): p. 568-80.
2. Sasselli, V., V. Pachnis, and A.J. Burns, *The enteric nervous system.* Dev Biol, 2012. **366**(1): p. 64-73.
3. Amiel, J., et al., *Hirschsprung disease, associated syndromes and genetics: a review.* J Med Genet, 2008. **45**(1): p. 1-14.
4. Alves, M.M., et al., *Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model.* Dev Biol, 2013. **382**(1): p. 320-9.
5. Attie, T., et al., *Diversity of RET proto-oncogene mutations in familial and sporadic Hirschsprung disease.* Hum Mol Genet, 1995. **4**(8): p. 1381-6.
6. Hofstra, R.M., et al., *RET and GDNF gene scanning in Hirschsprung patients using two dual denaturing gel systems.* Hum Mutat, 2000. **15**(5): p. 418-29.
7. Emison, E.S., et al., *A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk.* Nature, 2005. **434**(7035): p. 857-63.
8. Sribudiani, Y., et al., *Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression.* Gastroenterology, 2011. **140**(2): p. 572-582 e2.
9. Garcia-Barcelo, M.M., et al., *Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease.* Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2694-9.
10. 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, 2015. **96**(4): p. 581-96.
11. Parker, S.E., et al., *Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006.* Birth Defects Res A Clin Mol Teratol, 2010. **88**(12): p. 1008-16.
12. Gardiner, K., et al., *Mouse models of Down syndrome: how useful can they be? Comparison of the gene content of human chromosome 21 with orthologous mouse genomic regions.* Gene, 2003. **318**: p. 137-47.
13. Sturgeon, X. and K.J. Gardiner, *Transcript catalogs of human chromosome 21 and orthologous chimpanzee and mouse regions.* Mamm Genome, 2011. **22**(5-6): p. 261-71.
14. Korbel, J.O., et al., *The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies.* Proc Natl Acad Sci U S A, 2009. **106**(29): p. 12031-6.
15. Harrison, C., T. Wabbersen, and I.T. Shepherd, *In vivo visualization of the development of the enteric nervous system using a Tg(-8.3bphox2b:Kaede) transgenic zebrafish.* Genesis, 2014. **52**(12): p. 985-90.
16. Busch-Nentwich, E., Kettleborough, R., Harvey, S., Collins, J., Ding, M., Dooley, C., Fenyes, F., Gibbons, R., Herd, C., Mehroke, S., Scahill, C., Sealy, I., Wali, N., White, R., and Stemple, D.L. , *Sanger Institute Zebrafish Mutation Project mutant, phenotype and image data submission*. 2012.
17. Westerfield, M., *The zebrafish book : a guide for the laboratory use of zebrafish (Danio rerio)*. 2007, [Eugene, OR]: M. Westerfield.
18. Hyatt, T.M. and S.C. Ekker, *Vectors and techniques for ectopic gene expression in zebrafish.* Methods Cell Biol, 1999. **59**: p. 117-26.
19. Thisse, C. and B. Thisse, *High-resolution in situ hybridization to whole-mount zebrafish embryos.* Nat Protoc, 2008. **3**(1): p. 59-69.
20. Heanue, T.A. and V. Pachnis, *Ret isoform function and marker gene expression in the enteric nervous system is conserved across diverse vertebrate species.* Mech Dev, 2008. **125**(8): p. 687-99.
21. Wallace, K.N., et al., *Intestinal growth and differentiation in zebrafish.* Mech Dev, 2005. **122**(2): p. 157-73.
22. Wallace, K.N. and M. Pack, *Unique and conserved aspects of gut development in zebrafish.* Dev Biol, 2003. **255**(1): p. 12-29.
23. Kelsh, R.N. and J.S. Eisen, *The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives.* Development, 2000. **127**(3): p. 515-25.
24. Shepherd, I.T., et al., *Roles for GFRalpha1 receptors in zebrafish enteric nervous system development.* Development, 2004. **131**(1): p. 241-9.

25. Shepherd, I. and J. Eisen, *Development of the zebrafish enteric nervous system.* Methods Cell Biol, 2011. **101**: p. 143-60.
26. Cheng, W.W., et al., *Depletion of the IKBKAP ortholog in zebrafish leads to hirschsprung disease-like phenotype.* World J Gastroenterol, 2015. **21**(7): p. 2040-6.
27. Elworthy, S., et al., *Phox2b function in the enteric nervous system is conserved in zebrafish and is sox10-dependent.* Mech Dev, 2005. **122**(5): p. 659-69.
28. Uyttebroek, L., et al., *The zebrafish mutant lessen: an experimental model for congenital enteric neuropathies.* Neurogastroenterol Motil, 2016. **28**(3): p. 345-57.
29. Simonson, L.W., et al., *Characterization of enteric neurons in wild-type and mutant zebrafish using semi-automated cell counting and co-expression analysis.* Zebrafish, 2013. **10**(2): p. 147-53.
30. 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, 2010. **87**(1): p. 60-74.
31. Puffenberger, E.G., et al., *Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22.* Hum Mol Genet, 1994. **3**(8): p. 1217-25.
32. Devenish, R.J., et al., *The oligomycin axis of mitochondrial ATP synthase: OSCP and the proton channel.* J Bioenerg Biomembr, 2000. **32**(5): p. 507-15.
33. Ronn, T., et al., *Genetic variation in ATP5O is associated with skeletal muscle ATP50 mRNA expression and glucose uptake in young twins.* PLoS One, 2009. **4**(3): p. e4793.
34. Chen, H., et al., *Cloning of the cDNA for the human ATP synthase OSCP subunit (ATP5O) by exon trapping and mapping to chromosome 21q22.1-q22.2.* Genomics, 1995. **28**(3): p. 470-6.
35. Lee, Y., et al., *Selective impairment on the proliferation of neural progenitor cells by oxidative phosphorylation disruption.* Neurosci Lett, 2013. **535**: p. 134-9.
36. von Meyenburg, K., B.B. Jorgensen, and B. van Deurs, *Physiological and morphological effects of overproduction of membrane-bound ATP synthase in Escherichia coli K-12.* Embo J, 1984. **3**(8): p. 1791-7.
37. Arnold, S., et al., *Interaction between a chromosome 10 RET enhancer and chromosome 21 in the Down syndrome-Hirschsprung disease association.* Hum Mutat, 2009. **30**(5): p. 771-5.
38. Yimlamai, D., et al., *The zebrafish down syndrome cell adhesion molecule is involved in cell movement during embryogenesis.* Dev Biol, 2005. **279**(1): p. 44-57.
39. Thisse B, T.C., *Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submission* (http://zfinorg). 2004.
40. Stemple, D.L., *Structure and function of the notochord: an essential organ for chordate development.* Development, 2005. **132**(11): p. 2503-12.
41. Jannot, A.S., et al., *Chromosome 21 scan in Down syndrome reveals DSCAM as a predisposing locus in Hirschsprung disease.* PLoS One, 2013. **8**(5): p. e62519.
42. Semar, S., et al., *Changes of the enteric nervous system in amyloid-beta protein precursor transgenic mice correlate with disease progression.* J Alzheimers Dis, 2013. **36**(1): p. 7-20.
43. Soret, R., et al., *A collagen VI-dependent pathogenic mechanism for Hirschsprung's disease.* J Clin Invest, 2015. **125**(12): p. 4483-96.
44. Buchin, P.J., J.S. Levy, and J.N. Schullinger, *Down's syndrome and the gastrointestinal tract.* J Clin Gastroenterol, 1986. **8**(2): p. 111-4.

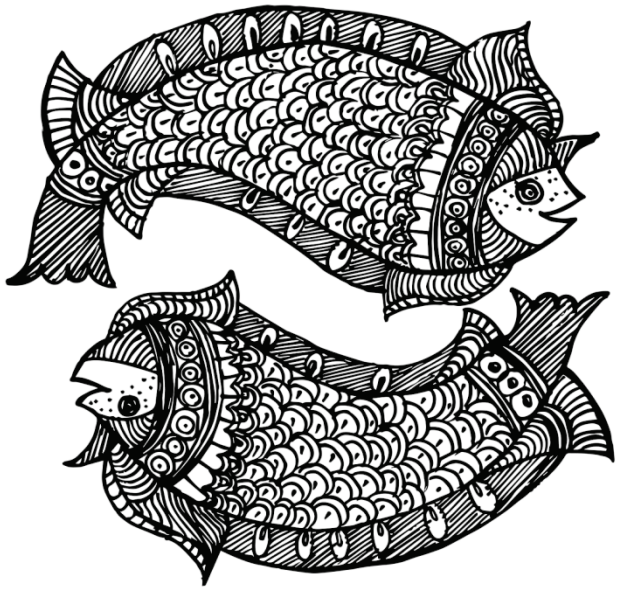
SUPPLEMENTARY INFORMATION



Supplementary Figure 1. *In vitro* apoptosis and cell cycle assays. SK-N-SH neuroblastoma cells were transected with construct containing *ATP5O* and the cells were starved for 48 hours and assayed for apoptosis and cell cycle assays by FACS analysis. Data are represented as mean ± SD for two independent experiments. A) Cells that were PE Annexin V-positive and 7AAD-negative were classified as early apoptotic, while cells positive for both PE Annexin V and 7AAD were marked as apoptotic. There were no differences between *ATP5O*-transfected cells and controls. B) Cell cycle analysis using propidium iodide (PI) DNA staining indicated no major differences between the cell phases, but there was a slight increase in G1 phase for *ATP5O*-transfected cells as compared to the controls.



Supplementary Figure 2. Craniofacial abnormalities by *SIM2* overexpression. A) Control embryo at 5dpf. (B) *SIM2* (100pg) overexpressing embryo at 5dpf displaying craniofacial abnormality as shown by arrows in 33% of them along with the notochord phenotype. The lower jaw appears dislodged compared to the control.



CHAPTER 6

Functional analysis of Hirschsprung disease genetic data using the zebrafish model

Rajendra K. Chauhan[1], Bart J. L. Eggen[2], Alan J. Burns[1,3], Iain T. Shepherd[4], Robert M.W. Hofstra[1,3], William. W. Cheng[1]

[1] Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.
[2] Department of Neuroscience, Section Medical Physiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
[3] Stem Cells and Regenerative Medicine, UCL Great Ormond Street Institute of Child Health, London, UK
[4] Department of Biology, Emory University, USA.

Corresponding author:
Robert MW Hofstra, Department of Clinical Genetics, Erasmus University Medical Center, PO BOX 2040, 3000CA Rotterdam, The Netherlands. Tel: +31-10-7037643. E-mail: r.hofstra@erasmusmc.nl

Manuscript in preparation

ABSTRACT

Zebrafish has emerged as a prominent vertebrate model for studying development processes and modeling human diseases. Next generation sequencing technology has been able to explain the genetics of many diseases, but many variants could not be linked to disease. Here we describe the utility of zebrafish (*Danio rerio*) for validating the newly identified genes involved with ENS development and HSCR. We then discuss different available methods to study ENS development and perform functional genetics in zebrafish by traditional methods and emphasize on newly developed genome editing techniques to query the gene activity. Finally, we discuss current methods and assays for phenotypic analysis in zebrafish ENS.

INTRODUCTION

The enteric nervous system (ENS) is one of the most complex subdivisions of the peripheral nervous system (PNS) composed of an intermeshed network of neurons and glial cells[1]. The enteric neurons and glial cells are organized to form enteric ganglia along entire length of gastrointestinal (GI) tract and ENS is the intrinsic nervous system of the GI tract. Enteric ganglia are arranged along the GI tract in two concentric plexi, the outer myenteric (Auerbach's) plexus between the circular and longitudinal smooth muscle layer wall and second is submucosal (Meissner's) plexus. The GI tract is essential for transporting, absorbing, digesting, and excreting food and waste, but also for protecting the host from ingested pathogens, allergens, and toxins. A complete ENS is essential to regulate gut motility and perform all these tasks of GI tract[2].

In vertebrates, ENS is entirely derived from the multipotent stem cells of neural crest. Specifically, neural crest cells (NCC) migrate extensively from the hindbrain, the vagal region of the neural tube, into and along the entire length of the GI tract[3]. A second contribution to the ENS arises from sacral neural crest cells. In both mouse and chick it has been shown that these cells colonize the distal part of the GI tract[4-7]. In zebrafish there is no evidence to support a sacral neural crest cell contribution to the ENS[8]. The NCC differentiates into various cell types during vertebrate development. These include bones; tendons; neurons; glia; melanocytes and connective, endocrine and adipose tissues.

Abnormality in the development of neural crest results in myriad of neurocristopathies. The enteric NCC must differentiate into different neuronal subtypes and glial cells to form a proper ENS. One of the most common diseases affecting the ENS is Hirschsprung's disease (HSCR)[9]. It is attributed to the failure of neural crest cells to migrate, differentiate, proliferate or survive and thereby form a functional ENS network. HSCR research is concentrated on gaining a better understanding of the underlying pathophysiology of enteric neuropathies, including the genetics.

HIRSCHSPRUNG DISEASE

HSCR is one of the most common causes of life threatening intestinal obstruction in neonates. The prevalence of HSCR is 1 in 5000 newborns and it affects more males than females (4:1)[10]. It is characterized by the absence of enteric ganglia in the distal part of the gastrointestinal (GI) tract due to an incomplete colonization of intestine by enteric NCC. The length of the aganglionic segment among HSCR patients differs. In the vast majority of cases (80%) only the most distal part of

the GI tract is affected and is termed short segment HSCR (S-HSCR) when only the rectum and sigmoid colon is aganglionic. In a smaller group (approximately 20% of cases), the aganglionosis extends proximal to include the sigmoid colon and is termed long segment HSCR (L-HSCR). In rare cases the aganglionosis can extend to the whole colon (total colonic aganglionosis - TCA), or the entire bowel (total intestinal aganglionosis)[11].

While in most cases patients only have HSCR (non-syndromic), in approximately one third of children HSCR is accompanied with other congenital anomalies. However, in only a minority of these children is a syndrome diagnosis established. Syndrome diagnosed in HSCR patients includes those caused by chromosomal abnormalities (like trisomy 21) and monogenic inherited disorders. Some of the best-known syndromes for which HSCR is part of the diagnosis are Waardenburg-shah type 4, congenital central hyperventilation syndrome, Goldberg-sphrintzen syndrome and Down syndrome[11].

HSCR genetics

HSCR is considered an inherited disease as familial cases occur in ~5% of all cases, there is an elevated risk for sibs (ranging from 1 to 44%), chromosomal abnormalities are found, and HSCR can be part of a syndrome. Although the disease can be inherited, the majority (80%) of cases occur sporadically. The recurrent risk for a second child with HSCR within a family ranges from 1% to 33% depending on the gender and the length of the aganglionic gut of the affected child[10]. Chromosomal abnormalities are seen in 12% of HSCR cases[11]. The mode of inheritance for the familial cases, in particular those in families where HSCR is the only disease entity seen, is autosomal dominant (with reduced penetrance). However, in families with syndromic HCSR the mode of inheritance is mostly autosomal recessive. The sporadic cases are believed to have a more complex mode of inheritance, with involvement of multiple genetic and non-genetic factors.

To date 16 genes have been reported as disease contributing. Most mutations are in genes belonging to the RET or Endothelin signaling pathways. The major gene by far is RET. Mutations are found in half of the familial cases and in around 15% of sporadic cases. Altogether, mutations in these 16 genes explain no more than 20-25% of all cases. The RET and Endothelin pathways are known to be involved in ENS development and are well conserved between human and mice[12,13].

However, not only do these mutations not explain most of the total disease risk for the entire patient population, they also only explain part of the

disease risk for individual patients. Genetic interactions between genes, for instance between *RET* and the Endothelin Receptor B gene, also play a major role in disease development[14-18]. This is further corroborated by the observed phenotypic variability.

Identifying new genes for HSCR

To find the missing heritability in HSCR next generation sequencing (NGS) techniques, such as whole exome sequencing (WES) and whole genome sequencing (WGS), are being used[19191919]. The list of candidate genes associated with HSCR has increased in the last decade and the quest is still ongoing[20-26].

Pinpointing the functional relevance of newly identified genetic variants in HSCR and ENS development has been challenging. Nevertheless, genetic studies *in vivo* and *in vitro* have been extremely helpful. In particular the zebrafish has proven to be powerful tools with many genes such as *ret*, *phox2b, sox10, gfrα1* and *gdnf* have been shown to be involved in ENS development[8,27-32].

ZEBRAFISH MODEL

Despite significant advances in the post genomic era, the elucidation of numerous pathogenic variants identified from NGS-based approaches needs a robust screening method to assign genes as disease causing. Given that HSCR is considered to be a disorder of ENS development that arises during early embryonic stages, it is necessary to study the processes involved *in vivo* rather than exclusively using cell culture techniques that present a poor indication of ENS development. Modeling of human disease phenotypes using murine models has been eminent due to many conserved developmental pathways and available genetic tools (such as conditional knockouts). Despite these advantages, using a murine model is time consuming, expensive, labor intensive and not well suited for large scale high-throughput genetic screens.

Zebrafish (*Danio rerio*) emerged as an attractive vertebrate model for the geneticists almost 3 decades ago[33]. Zebrafish belong to the group of lower vertebrate organisms with a relatively simple developmental plan. Compared to higher vertebrate models, the zebrafish is better suited for genetic analysis and *in vivo* studies because the embryonic development is *ex utero* and embryos can be grown in a culture dish. The embryos are virtually transparent allowing visualization of internal organs in a non-invasive way through early development. The zebrafish has high fecundity and one breeding pair can produce around 100-150 embryos. It is much closer to human than yeast, worms and flies and its

genome has been sequenced, well annotated and around 71.4% of human genes have at least one zebrafish orthologue[34]. These features have made zebrafish a tractable vertebrate model for genetics and developmental biology studies. In fact many human defects of heart, pigmentation, kidney, and retina, and diseases such as cancer, neurological disorders and many other congenital disorders have already been modeled using zebrafish and are reviewed elsewhere[35,36].

Zebrafish gut and ENS

The intestinal architecture and anatomy of the zebrafish closely resembles that of mammals. [37] The zebrafish gut undergoes rapid development and by 5 days post fertilization (dpf) the entire GI tract is functional. [38] In contrast to amniotes, the zebrafish gut is simpler: it lacks a submucosal layer and myenteric neurons are arranged as neuronal pairs or single neurons. [37] Like other vertebrates, the zebrafish ENS is also derived from the neural crest. [31] In zebrafish, neural crest cells (NCC) migrate as two parallel chains of cells to colonize the whole gut and differentiate into enteric neurons and glia[8]. The zebrafish ENS is also comparable to the mouse ENS based on the gene expression and function studies [13]. By 4dpf, regular anterograde and retrograde contractions are already being generated in the intestine and can be easily visualized [39]. The enteric innervation is well developed by 4dpf and already functional by 5dpf, when they start feeding (Figure 1).

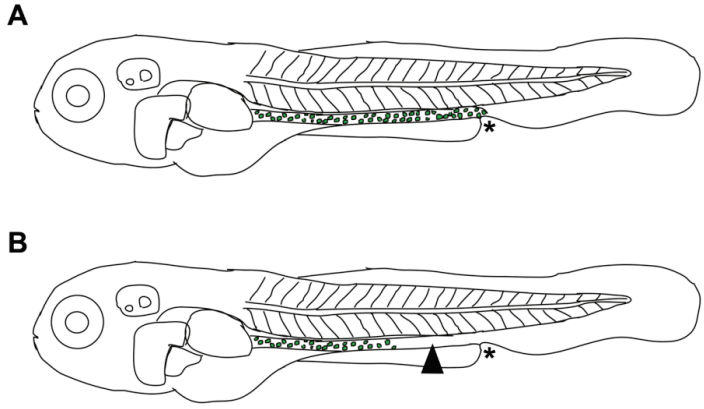


Figure 1. Zebrafish cartoon model for studying HSCR. (A) The enteric neurons are shown in dots along the zebrafish intestine. In normal zebrafish neural crest cells migrate until the anus (shown by star mark) and differentiate into enteric neurons demonstrating normal colonization of the intestine. (B) In HSCR model of zebrafish, the intestine doesn't get fully colonized (aganglionosis) and distal part of intestine is affected (as shown by arrow head). Comparing an affected (HSCR) zebrafish with unaffected zebrafish can be done by screening for enteric neurons.

Major neurotransmitters, such as serotonin 5-hydroxytryptamin (5HT), neurkinin A (NKA), vasoactive intestinal polypeptide (VIP), pituitary adenylate cyclase activating peptide (PACAP), nitric oxide (NO) and calcitonin gene-related peptide (CGRP) are also present in the zebrafish gut [40,41]. This combination of above mentioned features and presence of a functional gut containing an ENS derived from NCC make zebrafish as a good vertebrate model for studying gastrointestinal diseases including those affecting ENS development. (reviewed elsewhere[12,42]). Such studies have enabled the identification of new genes and characterization of known ENS development genes to better understand underlying molecular mechanisms in more detail[8,31,32,43]. These attributes qualify its utility for modeling HSCR and ENS development.

METHODS FOR STUDYING ZEBRAFISH ENS

Zebrafish reporter lines for ENS

Detection of enteric NCC and enteric neurons in zebrafish is the vital first step to determine if the ENS develops normally or abnormally (presence of HSCR-like phenotype). This is conventionally done by whole mount *in situ* hybridization or immunohistochemistry using markers such as *sox10*, *phox2b*, *ret* and *elavl3* (huc). However, lengthy protocols, poor resolution at tissue and cellular levels, and lack of antibodies against zebrafish proteins hinder the efficacy of these methods. Transgenic zebrafish reporter lines, mostly making use of promoter and regulatory element of the *phox2b* or *sox10* genes to drive expression of fluorescent proteins, provide an alternative to conventional cell labeling methods (Table 1). For example, the *Tg(-4.8phox2b:kaede)* line, with green fluorescent protein kaede expressed in migrating enteric NCC and differentiated enteric neurons, allows the rapid detection of mature ENS cells from 4 dpf onwards. The photoconvertible nature of the kaede protein also makes live cell tracing and imaging possible[44]. All in all, ENS specific reporter zebrafish lines greatly enhance the phenotypic analysis and their use should be encouraged.

Zebrafish mutant models for HSCR

Large-scale forward genetic screens in zebrafish have led to the identification of new genes and pathways for vertebrate development[45,46]. One of the earliest zebrafish mutants of the ENS, *cls* (colourless) mutant, was identified in a genetic screen for pigmentation defects. *cls* serves as a Waardenburg-shah syndrome model since fish lack pigment cells, and have reduce enteric neurons as well as additional NCC defects [31]. Subsequently, it was shown that the *cls* locus mapped to

the *sox10* gene which is known to be required for neural crest development[30]. Forward genetic screens have identified a number of other zebrafish mutants mimicking the HSCR phenotype. Specifically, two studies have mainly focused on

Table 1. Transgenic zebrafish reporter lines for studying neural crest derivatives.

Reporter	line	construct	Reference
egfp	w37tg	*phox2b:egfp*	(Boer et al., 2015)
	ba2tg	*sox10:egfp*	(Uribe and Bronner, 2015)
	bu1tg	*spon1b:egfp*	(Akle et al., 2012)
	jh102tg	*sox10:egfp**	(Seiler et al., 2010)
	jh105tg	*sox10:egfp**	(Seiler et al., 2010)
	jh109tg	*sox10:egfp**	(Seiler et al., 2010)
	knu3tg	*HuC/D:egfp*	(Olden et al., 2008)
			(Reichenbach et al., 2008)
			(Park et al., 2000)
	zf15tg	*foxd3:egfp*	(Alves et al., 2010)
DsRed	zf148tg	*NBT:DsRed*	(Davuluri et al., 2010)
			(Seiler et al., 2010)
			(Peri and Nusslein-Volhard, 2008)
kaede	em2tg	*phox2b:kaede*	(Harrison et al., 2014)
mCherry	c264Tg; gmc607Et	*N.D*	(Kok et al., 2012)
mrfp	N.D	*sox10:mrfp*	(Uribe and Bronner, 2015)

N.D: Not defined

genes essential for ENS development and normal gut function by examining the distribution of enteric neurons[47,48]. Pietsch *et al*. isolated 6 mutations and described the *lessen* mutant which has a reduction in enteric neurons in the distal intestine. *lessen* was subsequently mapped to the *med24* gene[42,45]. Recently, it was shown that *lessen* displayed delayed onset of motility and disturbed interstitial cell of cajal (ICC) in addition to the ENS defects[46]. Similar genetic screens performed by Kulhman et al. isolated 13 mutations affecting zebrafish ENS with 4 mutants displaying ENS-specific defects and the other 9 displaying pleiotropic effects along with ENS defects. Of note among the ENS mutants identified by Kulhman *et al*. was *gutwrencher,* which had fewer enteric neurons and less coordinated waves of contraction along the gut in mutants at 5.5dpf[47].

ENS FUNCTIONAL GENETICS

Morpholino-mediated gene knockdown

Antisense morpholinos (MO) are synthetic oligonucleotides that can be used to knockdown target genes by blocking either protein translation[48] or splicing[49] and have been widely used in developmental biology research. MO knockdown in zebrafish for genes known to be involved in HSCR have reproduced the absence/loss of enteric neurons observed in HSCR patients and murine

models[8,29,32,50]. The same approach has been used in reverse genetics to study the function of newly identified candidate genes identified through genetic analysis of HSCR patients[21,51]. Recently our group evaluated the function of 20 genes, in which *de novo* mutations were discovered in HSCR patients, by MO knockdown in zebrafish and confirmed 4 of them were important for ENS development (manuscript submitted). Despite its proven usefulness in research, there are still potential problems with the use of MO[52], notably off-target effects where any phenotype observed could be caused by the unintended knockdown of an irrelevant gene. In addition, there is an ongoing debate on how well MO knockdown and mutant phenotypes in zebrafish are correlated[53,54]. Therefore it is recommended that while deciding to use MO to interrogate gene function in ENS development in zebrafish, the experiments have to be carefully designed with all necessary controls included to enable one to distinguish between target gene-specific and off-target effects. Ideally, the MO knockdown phenotype should be verified by an independent experiment, such as mutant or knockout by targeted genome editing.

Targeted genome editing

Targeted genome editing is a relatively new technique which can be used to induce random insertion or deletion mutation (indel) at target coding sequences and thereby knock out genes as truncated proteins that are very often non-functional. The latest generation of targeted genome editing utilizes the Crispr/Cas9 (clustered regularly interspaced palindromic repeats/CRISPR-associated) system[55-58] and has already been successfully adopted to knockout target genes in zebrafish[59]. The protocol has since been modified to induce biallelic indel in F0 progeny[60]. The first report using Crispr/Cas9 to study the ENS came when Bernier et al.[61] knocked out the autism associated gene *chd8* in zebrafish and reproduced the phenotype of reduced number of enteric neurons as observed in *chd8* MO knockdown. Due to the high knockout efficiency by Crispr/Cas9, it is feasible to analyze the phenotype in F0 and hence replace MO with Crispr/Cas9 as a loss-of-function approach to study gene function. Based on this idea, rapid, high-throughput screening methods using Crispr/Cas9 in zebrafish with low off-target effect have been developed[62,63]. These methods will be useful for the HSCR research field, as more candidate genes are identified from NGS of patient genome a robust and reliable method is needed for their functional analysis. As a proof of principle we injected gRNAs targeting *ret* or *sox10* together with Cas9 protein to 1 cell-stage *Tg(phox2b:kaede)* embryos and in both scenarios observed loss of enteric neurons at 5 dpf (Figure 2). This data demonstrates the

potential of Crispr/Cas9-mediated knockout as a primary tool for the functional study of HSCR genes.

In addition to generating random indel, Crispr/Cas9 can be used to knock in specific SNP to target regions in human and mouse cells[55,58]. The protocol was adopted to correct the nonsense mutation in *slca45a2*$^{b4/b4}$ mutant zebrafish and rescue the pigmentation defect[64]. The possibility of targeted knock-in SNP in zebrafish is valuable for analyzing variants identified from HSCR patients, especially in the case of missense mutations when loss-of-function approach by MO knockdown and Crispr/Cas9 knockout might not accurately reflect the functional consequences of the mutations.

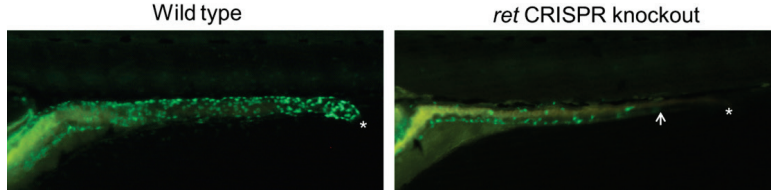


Figure 2. Crispr/Cas9 mediated knowndown in *Tg(phox2b:kaede)* zebrafish display aganglionosis of the distal intestine. The observed phenotype is similar to the phenotype seen on morpholino knockdown and that of *ret* mutant zebrafish.

mRNA overexpression

Most variants identified in HSCR patients are predicted to have loss-of-function effects and therefore gene knockdown or knockout in zebrafish will be the obvious choice to interrogate functional consequences. In certain scenarios, such as in Down syndrome-associated HSCR cases, a different approach is required. Down syndrome is the chromosomal anomaly most frequently associated with HSCR (2 to 10% of total HSCR cases)[11]. It has been hypothesized that over-expression of gene(s) on chromosome 21 leads to a higher risk for developing HSCR. To prove this hypothesis a gain-of-function, over-expression model is required and this can be achieved by injecting *in vitro* transcribed mRNA of interest to zebrafish embryos[65]. Using this method our group investigated a selection of genes from human chromosome 21 and the effect of their overexpression in ENS development. We discovered that overexpression of *ATP5O*, encoding a sub-unit of ATP synthase complex, resulted in fewer enteric neurons, a phenotype that could explain the missing heritability of Down syndrome-associated HSCR. The same approach will also be useful to test the recently proposed mechanism that excessive deposition of extracellular matrix molecules along migration pathways within the gut delays enteric NCC migration and hence could be a potential cause of HSCR[66]. Although, one of the key issue

with mRNA overexpression is the stability of the overexpressed mRNA and for how long does it lasts in the embryo.

Chemical Genetics

Although genetic defects undoubtedly contribute to HSCR, to date known genetic variants only account for less than half of the cases. Furthermore, incomplete penetrance and difference in expressivity remain unexplained. These facts have led to the hypothesis that non-genetic factors might also be involved in disease development. To identify drugs that could disrupt ENS development, Lake et al.[67] used zebrafish embryos to screen a library of 1508 compounds. One drug, mycophenolate, was singled out as its administration to zebrafish led to incomplete gut colonization by enteric NCC and impaired ENS development. The same research group later published another report which used zebrafish embryos to screen common medicines frequently taken by women during early pregnancy and discovered ibuprofen caused HSCR-like absence of enteric neurons[68].

Chemical screening using zebrafish might also help the development of new HSCR therapeutic strategies. Autologous cell transplantation has recently become the focus of research on novel therapies for HSCR[69,70]. The idea is to isolate ENS stem cells from the gut and, after *in vitro* expansion, transplant these cells back into the patient's aganglionic distal colon. However, HSCR-causing genetic variants may adversely affect the ability of the ENS stem cells to re-populate the intestine. The gut microenvironment of HSCR patients could also be suboptimal for the transplanted cells to colonize the intestine. One possible solution is to pre-treat the stem cells before transplantation with pharmacological compounds that can facilitate re-population by promoting cell migration, proliferation, or differentiation. Through chemical screening using human pluripotent stem cell-derived ENS progenitor cells in culture, pepstatin A was identified to be capable of improving colonization of gut by transplanted cells in vitro and in vivo[71]. Using existing mutant zebrafish lines that exhibit a HSCR-like phenotype, similar chemical screens could be conducted and the results may be more informative as the complex ENS developmental process would be more accurately mimicked *in vivo* in zebrafish.

PHENOTYPIC ANALYSIS OF ZEBRAFISH ENS

Here we provide a snapshot of different assays for screening the zebrafish ENS and for the study of homologs of human genes identified to be associated with HSCR or other ENS defects. These tools can be utilized to gain a better insight of

the pathogenic mechanisms and might help to establish the connection between the genetic mutations identified in HSCR patients with that of the functionality of the ENS.

Neuronal count and microscopy

Enteric neurons must be generated in correct numbers for the development of normal gut function. In the developing zebrafish, the enteric precursors migrate along the gut and start differentiating into enteric neurons by 2dpf as shown by expression of the pan neuronal marker HuC/D[8,39]. The migration of enteric precursors in the zebrafish gut is complete by 3dpf[72]. During ENS development, defects in the migration, differentiation or survival of enteric precursors could lead to alteration in neuronal number and distribution resulting in abnormal intestinal motility. Studies from several laboratories have shown that zebrafish ENS can be analyzed by comparing the enteric neuronal count of control embryos with that of treated or mutant embryos[32,45,47].

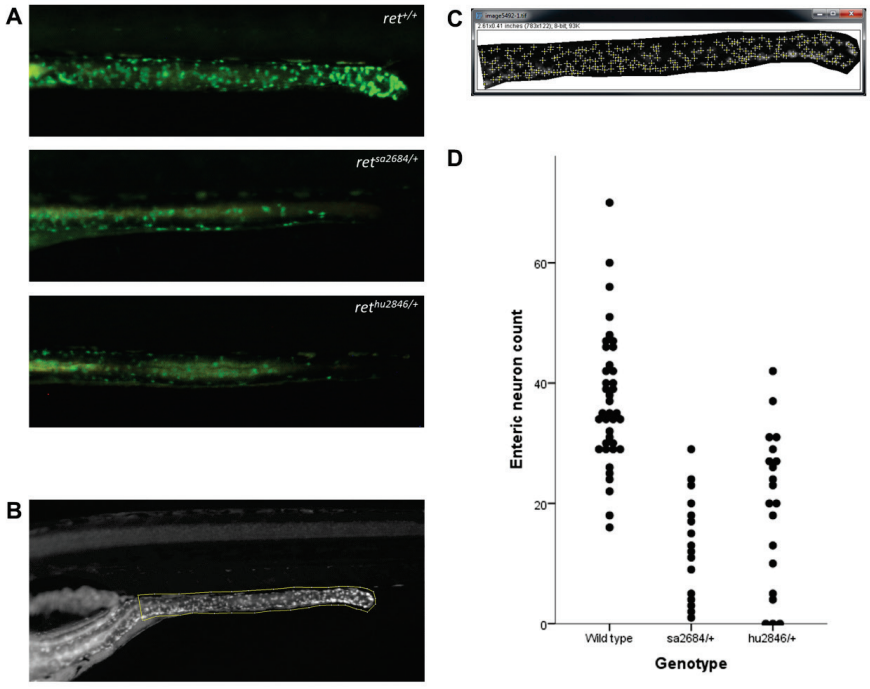


Figure 3. Enteric neuronal count in zebrafish embryos (A) Transgenic reporter line embryos displaying enteric neurons in the gut, ret mutants display aganglionosis and hypoganglionosis. (B, C) In order to count the enteric neurons in the zebrafish gut a selection can be made and image analysis tools can be used for counting. (D) Comparative analysis of neuronal count in wild type and to *ret* mutant lines, which display fewer enteric neurons in the most distal part of intestine.

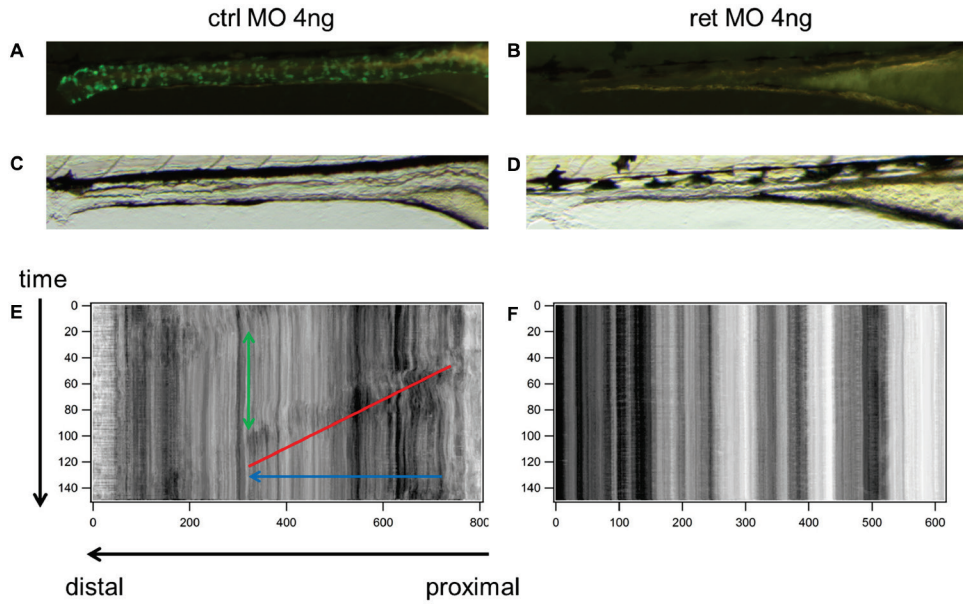


Figure 4. Gastrointestinal motility assay performed in (A) Control zebrafish gut and (B) ret knockdown. . In vivo video recording of zebrafish gut from control (C) and ret knockdown (D) was performed followed by generating spatio temporal map from the video sequences. (E) Anterograde contraction seen in control zebrafish. (F) No anterograde contraction in ret knockdown. The lacks of enteric neurons in ret knockdown zebrafish affects the gut motility patterns as seen in the spatiotemporal map.

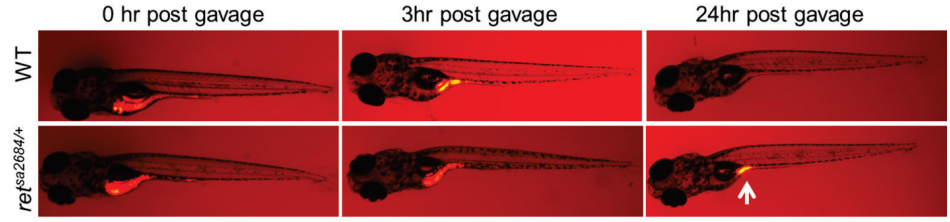


Figure 5. Transit assay in zebrafish using microgavage. Shown in the panel is a comparison between WT and ret mutant zebrafish embryos at 7dpf. Fluoroscent beads were injected directly into the intestinal lumen and screened under fluoroscent microscope at 3 time points (0,3 and 24 hours post gavage). At 24 hours post gavage, the ret mutant still retains the fluoroscent beads (shown by arrow), while the WT fish don't have any beads left.

Manual counting of enteric neurons in the zebrafish intestine is laborious and time consuming. To overcome the difficulty of counting the enteric neurons in whole mount zebrafish intestine, Simonson et al. developed a semi- automated cell counting program based on MATLAB to effectively count and analyze co-expression of different neuronal markers using 3D confocal image stacks[73]. We

have also developed a straightforward way to count the number of enteric neurons of zebrafish intestine in a semi-automated fashion using digital images captured by fluorescent microscopy. In-house developed algorithms with image analysis software from FIJI can be used for quantitative scoring of enteric neurons on a selected portion of intestine (Figure 3 B, C).

Gut motility

Gut motility is controlled by the ENS and modulated by different neurotransmitters. Functional and immunohistochemical studies demonstrated that the zebrafish intestine expresses a range of neurotransmitters including 5HT, NKA, VIP, PACAP, NOS and CGRP[40,41]. Gastrointestinal motility comprises a range of processes including (1) standing contraction; mixing food in stomach; (2) peristaltic movements in anterograde (oral to anal) and retrograde (vomiting or regurgitation) directions and (3) phasic contraction of sphincters[74]. The zebrafish offers the possibility of imaging gut motility patterns *in vivo* using real time video microscopy[41,47]. Erratic and spontaneous contraction waves are observed by 3dpf (before the onset of feeding) and later (4dpf-7dpf), distinct anterograde, retrograde and rectal contractions are observed[39]. The zebrafish is stomach less prominent, so the retrograde contractions in the anterior intestine may take over the function of food mixing, whereas retrograde and anterograde contractions spread in both directions from mid intestine to mainly transport contents along the gut[39].

These zebrafish gut movements can be recorded and analyzed from video recordings to investigate the functioning control of gut motility patterns and the data shown as spatiotemporal maps of gut movement[75]. As an example, we show the gut motility assay in WT and ret knockdown zebrafish embryo (Figure 4).

Gut transit assay

The functional consequence of coordinated contractions of the zebrafish intestine can be measured by the intestinal transit assay[76]. Field et al. developed a method to visualize intestinal transit in zebrafish larvae in real time by feeding fish with normal larval feed mixed with fluorescent microspheres, which are non-absorbable, non-digestible and thus traceable using fluorescent microscopy. This method is performed by immersing zebrafish larvae in a solution containing feed-coated microspheres. Although individual larvae exhibit differences in their feeding such that they may ingest different amounts of food, pre-sorting is performed before an experiment to reduce variability. To overcome this shortfall, a newly established method known as microgavage was performed, which

utilizes the microinjection of microspheres directly into the lumen of the anterior intestine of larval zebrafish[77]. In both the methods, larvae are screened for the fluorescent microspheres in the intestine at different time points under fluorescent microscopy and the transit time measured by the time taken to expel the fluorescent beads (Figure 5).

The microgavage method appears to be more consistent compared to the former, but it is equally time consuming and labor intensive. Nevertheless GI motility can be monitored in zebrafish larvae by using either of these methods and a comparison can be made between wild type and mutant/morphant zebrafish to identify and assess the effect of any gene on gut motility and transit time. Using such assays, it has already been shown that the degree of enteric neuron loss in zebrafish larvae correlates with the extent of intestinal transit deficits[61,76].

Cell tracking and live imaging

The high-resolution study of embryonic development and disease modeling requires analysis of individual cells in context of an organism. The zebrafish provides an excellent model to study cell fate and track individual cell types during their cell division and progressive development. Genetic manipulation to express fluorescent or even photoconversible proteins by labeling specific cells allows cell lineage tracing for a particular cell type and live imaging in a particular organ or tissue in real time. Using two-photon confocal microscopy, it is possible to differentially label any cell and to fate map the photoactivated cell[78]. Transgenic reporter lines used for studying the ENS development of zebrafish *in vivo* have provided the ability to track any particular cell using photoconversible proteins[44]. Live image profiling of neural crest lineages in zebrafish using transgenic lines have been also performed and could be implemented for other transgenic lines as well as to better understand cell dynamics[78]. Altogether, such approaches can be used to visualize the migrating NCC population in time until they colonize the zebrafish gut.

CONCLUSIONS AND FUTURE PROSPECTS

Despite many years of research on ENS development, we still lack a complete understanding of the genetic basis of HSCR. Human genetic studies using NGS approaches have revealed many new HSCR genes and the list of candidate genes is increasing. To better understand the pathogenicity of mutations in the novel genes and how these genes control ENS development demands rapid and robust methods to functionally validate their effects on ENS development. The zebrafish

is an excellent model for the study of HSCR as rapid transgenic techniques, high-resolution fluorescent in vivo imaging, and well-characterized promoters for tissue-specific expression already exist. Furthermore, the zebrafish provides a medium throughput system for the assay of potential treatment strategies using genome-editing technologies. Zebrafish genes can be customized to study their effects on the ENS by evaluating the neuronal count, gut motility and intestinal transit time.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the lab members of Hofstra group for their critical reading and their comments. This study was supported by research grant from ZonMW (TOP-subsidie 40-00812-98-10042) and the Maag Lever Darm stichting (WO09-62).

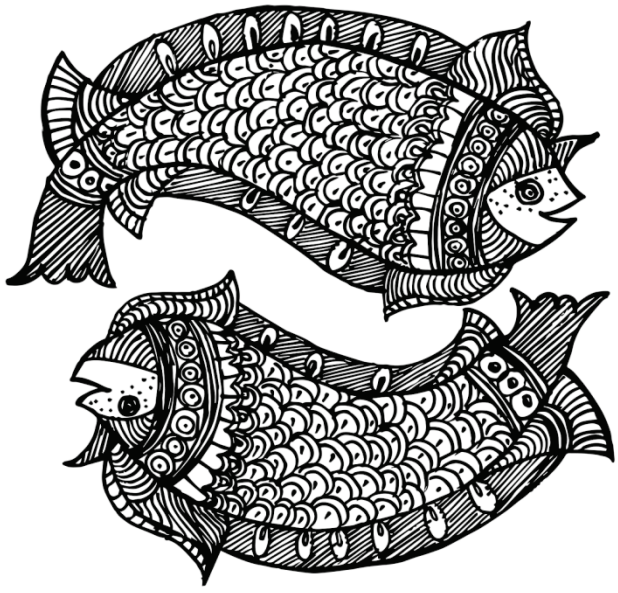
REFERENCES

1. Gershon, M. *The Second Brain: A Groundbreaking New Understanding of Nervous Disorders of the Stomach and Intestine*, (HarperCollins, 1999).
2. Furness, J.B. & Wiley, I. The enteric nervous system. (Blackwell Pub., Malden, Mass., 2006).
3. Le Douarin, N. *The neural crest*, (Cambridge University Press, Cambridge [Cambridgeshire]; New York, 1982).
4. Burns, A.J. & Douarin, N.M. The sacral neural crest contributes neurons and glia to the post-umbilical gut: spatiotemporal analysis of the development of the enteric nervous system. *Development* **125**, 4335-47 (1998).
5. Wang, X., Chan, A.K., Sham, M.H., Burns, A.J. & Chan, W.Y. Analysis of the sacral neural crest cell contribution to the hindgut enteric nervous system in the mouse embryo. *Gastroenterology* **141**, 992-1002 e1-6 (2011).
6. Gershon, M.D., Chalazonitis, A. & Rothman, T.P. From neural crest to bowel: development of the enteric nervous system. *J Neurobiol* **24**, 199-214 (1993).
7. Wallace, A.S. & Burns, A.J. Development of the enteric nervous system, smooth muscle and interstitial cells of Cajal in the human gastrointestinal tract. *Cell Tissue Res* **319**, 367-82 (2005).
8. Shepherd, I.T., Pietsch, J., Elworthy, S., Kelsh, R.N. & Raible, D.W. Roles for GFRalpha1 receptors in zebrafish enteric nervous system development. *Development* **131**, 241-9 (2004).
9. Heanue, T.A. & Pachnis, V. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat Rev Neurosci* **8**, 466-79 (2007).
10. Badner, J.A., Sieber, W.K., Garver, K.L. & Chakravarti, A. A genetic study of Hirschsprung disease. *Am J Hum Genet* **46**, 568-80 (1990).
11. Amiel, J. *et al.* Hirschsprung disease, associated syndromes and genetics: a review. *J Med Genet* **45**, 1-14 (2008).
12. Burzynski, G., Shepherd, I.T. & Enomoto, H. Genetic model system studies of the development of the enteric nervous system, gut motility and Hirschsprung's disease. *Neurogastroenterol Motil* **21**, 113-27 (2009).

13. Heanue, T.A. & Pachnis, V. Ret isoform function and marker gene expression in the enteric nervous system is conserved across diverse vertebrate species. *Mech Dev* **125**, 687-99 (2008).
14. Carrasquillo, M.M. *et al.* Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. *Nat Genet* **32**, 237-44 (2002).
15. Gabriel, S.B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* **31**, 89-93 (2002).
16. Emison, E.S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857-63 (2005).
17. de Pontual, L. *et al.* Epistasis between RET and BBS mutations modulates enteric innervation and causes syndromic Hirschsprung disease. *Proc Natl Acad Sci U S A* **106**, 13921-6 (2009).
18. Sribudiani, Y. *et al.* Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology* **140**, 572-582 e2 (2011).
19. Gonzaga-Jauregui, C., Lupski, J.R. & Gibbs, R.A. Human genome sequencing in health and disease. *Annu Rev Med* **63**, 35-61 (2012).
20. Enguix-Riego, M.V. *et al.* Identification of different mechanisms leading to PAX6 down-regulation as potential events contributing to the onset of Hirschsprung disease. *Sci Rep* **6**, 21160 (2016).
21. 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**, 581-96 (2015).
22. Shin, J.G. *et al.* Potential association of VAMP5 polymorphisms with total colonic aganglionosis in Hirschsprung disease. *Neurogastroenterol Motil* (2016).
23. Lee, J.S. *et al.* Association Analysis of SLC6A20 Polymorphisms With Hirschsprung Disease. *J Pediatr Gastroenterol Nutr* **62**, 64-70 (2016).
24. Kim, L.H. *et al.* Genetic variants of IL-11 associated with risk of Hirschsprung disease. *Neurogastroenterol Motil* **27**, 1371-7 (2015).
25. Liu, J.A. *et al.* Identification of GLI Mutations in Patients With Hirschsprung Disease That Disrupt Enteric Nervous System Development in Mice. *Gastroenterology* **149**, 1837-1848 e5 (2015).
26. Wang, Y. *et al.* Common genetic variations in Patched1 (PTCH1) gene and risk of hirschsprung disease in the Han Chinese population. *PLoS One* **8**, e75407 (2013).
27. Bisgrove, B.W., Raible, D.W., Walter, V., Eisen, J.S. & Grunwald, D.J. Expression of c-ret in the zebrafish embryo: potential roles in motoneuronal development. *J Neurobiol* **33**, 749-68 (1997).
28. Marcos-Gutierrez, C.V., Wilson, S.W., Holder, N. & Pachnis, V. The zebrafish homologue of the ret receptor and its pattern of expression during embryogenesis. *Oncogene* **14**, 879-89 (1997).
29. Shepherd, I.T., Beattie, C.E. & Raible, D.W. Functional analysis of zebrafish GDNF. *Dev Biol* **231**, 420-35 (2001).
30. Dutton, K.A. *et al.* Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development* **128**, 4113-25 (2001).
31. Kelsh, R.N. & Eisen, J.S. The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. *Development* **127**, 515-25 (2000).
32. Elworthy, S., Pinto, J.P., Pettifer, A., Cancela, M.L. & Kelsh, R.N. Phox2b function in the enteric nervous system is conserved in zebrafish and is sox10-dependent. *Mech Dev* **122**, 659-69 (2005).
33. Streisinger, G., Walker, C., Dower, N., Knauber, D. & Singer, F. Production of clones of homozygous diploid zebra fish (Brachydanio rerio). *Nature* **291**, 293-6 (1981).
34. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498-503 (2013).
35. Lieschke, G.J. & Currie, P.D. Animal models of human disease: zebrafish swim into view. *Nat Rev Genet* **8**, 353-67 (2007).
36. Santoriello, C. & Zon, L.I. Hooked! Modeling human disease in zebrafish. *J Clin Invest* **122**, 2337-43 (2012).

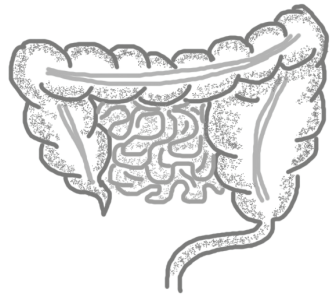
37. Wallace, K.N., Akhter, S., Smith, E.M., Lorent, K. & Pack, M. Intestinal growth and differentiation in zebrafish. *Mech Dev* **122**, 157-73 (2005).
38. Wallace, K.N. & Pack, M. Unique and conserved aspects of gut development in zebrafish. *Dev Biol* **255**, 12-29 (2003).
39. Holmberg, A., Schwerte, T., Fritsche, R., Pelster, B. & Holmgren, S. Ontogeny of intestinal motility in correlation to neuronal development in zebrafish embryos and larvae. *Journal of Fish Biology* **63**, 318-331 (2003).
40. Olsson, C., Holmberg, A. & Holmgren, S. Development of enteric and vagal innervation of the zebrafish (Danio rerio) gut. *J Comp Neurol* **508**, 756-70 (2008).
41. Holmberg, A., Schwerte, T., Pelster, B. & Holmgren, S. Ontogeny of the gut motility control system in zebrafish Danio rerio embryos and larvae. *J Exp Biol* **207**, 4085-94 (2004).
42. Shepherd, I. & Eisen, J. Development of the zebrafish enteric nervous system. *Methods Cell Biol* **101**, 143-60 (2011).
43. Uribe, R.A. & Bronner, M.E. Meis3 is required for neural crest invasion of the gut during zebrafish enteric nervous system development. *Mol Biol Cell* **26**, 3728-40 (2015).
44. Harrison, C., Wabbersen, T. & Shepherd, I.T. In vivo visualization of the development of the enteric nervous system using a Tg(-8.3bphox2b:Kaede) transgenic zebrafish. *Genesis* **52**, 985-90 (2014).
45. Pietsch, J. *et al.* lessen encodes a zebrafish trap100 required for enteric nervous system development. *Development* **133**, 395-406 (2006).
46. Uyttebroek, L. *et al.* The zebrafish mutant lessen: an experimental model for congenital enteric neuropathies. *Neurogastroenterol Motil* **28**, 345-57 (2016).
47. Kuhlman, J. & Eisen, J.S. Genetic screen for mutations affecting development and function of the enteric nervous system. *Dev Dyn* **236**, 118-27 (2007).
48. Summerton, J. Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim Biophys Acta* **1489**, 141-58 (1999).
49. Draper, B.W., Morcos, P.A. & Kimmel, C.B. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis* **30**, 154-6 (2001).
50. Dutton, K., Dutton, J.R., Pauliny, A. & Kelsh, R.N. A morpholino phenocopy of the colourless mutant. *Genesis* **30**, 188-9 (2001).
51. Cheng, W.W. *et al.* Depletion of the IKBKAP ortholog in zebrafish leads to hirschsprung disease-like phenotype. *World J Gastroenterol* **21**, 2040-6 (2015).
52. Eisen, J.S. & Smith, J.C. Controlling morpholino experiments: don't stop making antisense. *Development* **135**, 1735-43 (2008).
53. Kok, F.O. *et al.* Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell* **32**, 97-108 (2015).
54. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230-3 (2015).
55. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-23 (2013).
56. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-21 (2012).
57. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
58. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6 (2013).
59. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227-9 (2013).
60. Jao, L.E., Wente, S.R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A* **110**, 13904-9 (2013).
61. Bernier, R. *et al.* Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158**, 263-76 (2014).
62. Shah, A.N., Davey, C.F., Whitebirch, A.C., Miller, A.C. & Moens, C.B. Rapid reverse genetic screening using CRISPR in zebrafish. *Nat Methods* **12**, 535-40 (2015).
63. Varshney, G.K. *et al.* High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res* **25**, 1030-42 (2015).
64. Irion, U., Krauss, J. & Nusslein-Volhard, C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* **141**, 4827-30 (2014).

65. Rosen, J.N., Sweeney, M.F. & Mably, J.D. Microinjection of zebrafish embryos to analyze gene function. *J Vis Exp* (2009).
66. Soret, R. *et al.* A collagen VI-dependent pathogenic mechanism for Hirschsprung's disease. *J Clin Invest* **125**, 4483-96 (2015).
67. Lake, J.I., Tusheva, O.A., Graham, B.L. & Heuckeroth, R.O. Hirschsprung-like disease is exacerbated by reduced de novo GMP synthesis. *J Clin Invest* **123**, 4875-87 (2013).
68. Schill, E.M. *et al.* Ibuprofen slows migration and inhibits bowel colonization by enteric nervous system precursors in zebrafish, chick and mouse. *Dev Biol* **409**, 473-88 (2016).
69. Burns, A.J. & Thapar, N. Neural stem cell therapies for enteric nervous system disorders. *Nat Rev Gastroenterol Hepatol* **11**, 317-28 (2014).
70. Hotta, R. *et al.* Delivery of enteric neural progenitors with 5-HT4 agonist-loaded nanoparticles and thermosensitive hydrogel enhances cell proliferation and differentiation following transplantation in vivo. *Biomaterials* **88**, 1-11 (2016).
71. Fattahi, F. *et al.* Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease. *Nature* **531**, 105-9 (2016).
72. Olden, T., Akhtar, T., Beckman, S.A. & Wallace, K.N. Differentiation of the zebrafish enteric nervous system and intestinal smooth muscle. *Genesis* **46**, 484-98 (2008).
73. Simonson, L.W., Ganz, J., Melancon, E. & Eisen, J.S. Characterization of enteric neurons in wild-type and mutant zebrafish using semi-automated cell counting and co-expression analysis. *Zebrafish* **10**, 147-53 (2013).
74. Olsson, C. & Holmgren, S. The control of gut motility. *Comp Biochem Physiol A Mol Integr Physiol* **128**, 481-503 (2001).
75. Holmberg, A., Olsson, C. & Hennig, G.W. TTX-sensitive and TTX-insensitive control of spontaneous gut motility in the developing zebrafish (Danio rerio) larvae. *J Exp Biol* **210**, 1084-91 (2007).
76. Field, H.A., Kelley, K.A., Martell, L., Goldstein, A.M. & Serluca, F.C. Analysis of gastrointestinal physiology using a novel intestinal transit assay in zebrafish. *Neurogastroenterol Motil* **21**, 304-12 (2009).
77. Cocchiaro, J.L. & Rawls, J.F. Microgavage of zebrafish larvae. *J Vis Exp*, e4434 (2013).
78. Kwak, J. *et al.* Live image profiling of neural crest lineages in zebrafish transgenic lines. *Mol Cells* **35**, 255-60 (2013).



CHAPTER 7

General Discussion and Future Perspectives



Hirschsprung disease (HSCR) is a congenital intestinal motility disorder characterized by a lack of enteric ganglia (aganglionosis) of variable lengths of the gastrointestinal tract. Aganglionosis in HSCR is caused by a failure of enteric neural crest cells (NCC) to migrate, differentiate, proliferate or survive and form a functional network of neurons and glia, called the enteric nervous system (ENS)[1]. HSCR is a complex inherited disease and can be caused by (combinations of) mutations in distinct genes all affecting ENS development. Numerous studies have been performed to find genes involved in HSCR development, linkage studies and genome wide association studies (GWAS) have revealed a number of common disease-associated genes and variants. All the mutations and variations identified in HSCR cases explain approximately 30% of the overall genetic risk[2]. However, the vast majority of (sporadic) HSCR cases cannot yet be explained by the identified mutations or associations. This has led to the hypothesis that, in patients with HSCR, the majority of the disease risk may be explained by combinations of rare coding or non-coding variants in the identified and other unknown genes.

GENOMIC APPROACHES TO STUDY GENETICS OF HSCR

Since only a part of the overall genetic risk of HSCR can be explained by identified variants the question is how to proceed next? Should GWAS studies be performed further in the search for the missing heritability? GWAS and subsequent meta-analysis has identified many robustly replicating common variants associated with many complex diseases. In HSCR, three GWAS have been performed mainly on sporadic HSCR cases from different ethnicities to identify additional HSCR genetic loci that could contribute to the disease risk[3-5]. However, in general the variants identified by GWAS usually explain only a small fraction of complex disease susceptibility[6].

One could indeed increase the number of patients and aim to identify additional loci. This is a common approach for common diseases, and provides good results. However, we have learned from these studies that the number of patients needed for such meta-studies are at a magnitude of what has been done so far for HSCR. Indeed thousands of patients are needed to obtain some additional loci[6]. Moreover, these additional loci will add only slightly to the overall genetic risk. In addition, as HSCR is a rare disease collecting thousands of patients is extremely difficult. Therefore, it is questionable whether this should be the focus of research for the missing heritability.

Should we then focus on rare disease associated variants? Variants have been categorized on the basis of allele frequencies as well as effect sizes, namely very common (5 – 50%), less common (1 -5%), rare (<1%) and private (restricted to probands and immediate relatives)[7]. It was postulated that rare mutations of severe effect can explain a substantial portion of complex human diseases[8]. These rare variants cannot be identified via GWAS studies, but with the advent of next generation sequencing (NGS) technology they are detectable in a large scale manner by looking at the exome or whole genome sequencing data. There are multiple examples of rare and low frequency variants associated with complex diseases as already summarized by Schork et al. in 2009[9]. Assuming that indeed rare variants are involved it is worthwhile to search for them to decipher the complexity of HSCR.

However such causative rare variant research poses difficulties as well. Finding rare inherited variants can be performed by exome or whole genome sequencing on individual patients. However, the difficulty is in how to confirm that the identified variants are associated with or causative for the disease? As in GWAS studies one relies on statistics. Commonly, burden tests are performed to determine whether identified variants in gene or gene networks are overrepresented in patients compared to the controls. As for GWAS studies, this requires large numbers of patients and controls. The numbers needed depend largely on the frequency of the variants in genes. For less common variants (1 - 5%) or rare (<1%) variants the numbers are within the hundreds and for private variants the numbers increase to thousands. Although possible, these types of studies are difficult to perform as well. An often-accepted argument for causality is the finding of a second family with a variant in the same gene. Whether this is always 'statistically sound proof' remains questionable. The same holds true for the search for *de novo* mutations. The chance of finding *de novo* mutations in genes connected to the disease process is low and finding these is often seen as additional evidence for disease causality. However, the fact that a variant is *de novo* is, on its own, not sufficient proof for causality. In chapter 2 (and the next paragraph) we describe how additional evidence might be collected to make causality more likely.

Identification of *de novo* mutations in HSCR

Over the last few years, de novo mutations have been implicated in the etiology of several complex genetic diseases, including intellectual disability, autism and schizophrenia[10]. The rarest variants are *de novo* mutations and analysis of whole exome sequencing data from unaffected parents-affected child trios can identify such *de novo* variants. L-HSCR is highly heritable and follows mostly a dominant

mode of inheritance with incomplete penetrance[11]. Based on this fact, we exome sequenced 24 sporadic, non-syndromic HSCR patients to identify rare variants that had large contribution in manifestation of L-HSCR phenotype. (**Chapter 2**) In total, we found 28 *de novo* mutations distributed among 21 genes and 8 of them were found in *RET*, the major HSCR gene, corroborating the contributions made by *de novo* mutations in the etiology of L-HSCR. We carried out an unbiased *in silico* analysis of these newly identified *de novo* mutated genes for their prospective role in ENS development. *In silico* analysis revealed that *RET* and *CKAP2L* were enriched for rare variants in HSCR patients compared to controls, but only *RET* was confirmed in an independent cohort. None of the genes, besides *RET,* were linked to known ENS signaling pathways, although all of them were expressed by mouse ENCCs (E14.5) or NCC derived from human iPS cells.

As mentioned above, a *de novo* appearance of a variant is on its own can't prove causality. Therefore, we first searched for inherited mutations in our HSCR cohorts. This initial search did not give us statistical proof for causality (larger studies are ongoing). Did this mean that the variants found were not disease associated? We hypothesize that some might be associated, based on expression studies and we therefore decided to add functional proof instead of statistical proof. We tested all the *de novo* mutated genes (N=12) with a zebrafish orthologue by knockdown experiments in zebrafish. Morpholino (MO) mediated knockdown was done in *Tg (-8.3phox2b:Kaede)* transgenic zebrafish embryos that express the fluorescent protein Kaede in enteric neuron precursors and differentiated enteric neurons[12]. We found that 4 genes (*DENND3*, *NCLN*, *NUP98* and *TBATA*) displayed aganglionosis in the zebrafish intestine mimicking HSCR like phenotype. Is this proof enough? There is an ongoing debate on concordance between MO knockdown and mutant phenotypes in zebrafish[13,14]. In order to rule out this issue, we have been able to recapitulate similar phenotypes using the Crispr/Cas9 gene editing system for some of these genes (Cheng unpublished data).

After finding functional proof that 4 genes might well be disease associated we are still left with the question whether we have enough convincing evidence to call these genes to be disease causing and associated with HSCR. Is our functional evidence comparable with the finding of a second family with a mutated gene? All the evidence makes us propose that the 4 genes found contribute to disease development. Nonetheless, their contribution remains to be determined. We do think that our approach, genetics in combination with *in vivo* modelling, could be considered in the diagnostics of other rare diseases in which a private variant is found. Besides that, we plan to interrogate the role of these

four genes in ENCCs migration, proliferation, differentiation or survival using in-vitro assays to establish their specific roles during ENS development.

Family-based study to identify new HSCR genes

To circumvent the problem of large numbers of patients, one might perform linkage analysis on (large) multigenerational families. Linkage analysis has again emerged as an extremely useful method for the genomic analysis of complex traits[15]. Both linkage analysis on families with Mendelian inheritance as well as sibpair analysis on smaller families has already been performed for HSCR. Linkage analysis resulted in the identification of the major HSCR associated loci and genes, *RET* and EDNRB[16-20]. Moreover, linkage and sibpair analysis resulted in the identification of additional HSCR susceptibility loci at 9q31, 3p21, 19q12, 13q21 and 4q31.3-q32.3 respectively[21-24].

Nowadays these studies are not often performed any more as most families have already been used in such studies. However, the current technology to perform such studies is much better than 10 years ago. Linkage analysis is easier with the development of high density SNP arrays, and finding mutations has also become easier with NGS. Therefore, it may be worthwhile to rescreen unsolved families. This is what we did with a family in which a previous study revealed HSCR linkage to 4q31.3-q32.3[24]. (**Chapter 3**) We performed exome sequencing and variant prioritization to determine the genetic cause of HSCR in this family. We identified one main candidate in the linkage interval in exon 20 of *LRBA.* This variant was present in all five affected family members. Functional proof of the *LRBA* variant could not be established and we can't ignore the fact that there might be another variant present in the non-coding DNA within the linkage region. We do show that *MAB21L2*, *a* gene embedded within an intron of *LRBA,* is important for enteric neural crest cell (ENCC) differentiation during ENS development in a zebrafish model. Whether *MAB21L2* is the gene that contributes to disease development is however yet unclear. It shows that although exome sequencing facilitates finding mutations, pinpointing the real disease-causing gene is still difficult, and we are still not sure if we have identified the disease contributing variant. One could even argue that there is no such variant as linkage still is a matter of chance.

The incomplete penetrance of the disease in the family suggests the involvement of other (rare) variants elsewhere in the genome. Therefore, we searched for variants present especially in known HSCR genes and genes associated with ENS development. We identified several candidates, among which was a rare *RET* coding variant in one branch of the family and inherited *IHH*, *GLI3* variants along with a *de novo* mutation in *GDNF* (RET ligand) in the

second branch of the family. Functional studies confirmed the pathogenic nature of the variants identified in *RET* and *IHH,* confirming the importance of RET and Hedgehog signalling for ENS development. This study demonstrates that rare variants in multiple genes could lead to the development of HSCR, further illustrating the complexity of HSCR genetics.

Finding multiple variants is due to the fact that we did a rigorous search for them. Many similar studies stop after finding a good candidate disease variant. It could be that for most diseases one could find multiple disease contributing variants when one would search better as we did in our studies.

Identification and integration of predictive regulatory sequences for HSCR

Until now, HSCR research has identified variants mainly in the protein coding DNA. In the human genome, protein coding genes are well studied sequences although they account for only 1.5% of the genome and 2% if untranslated regions (UTRs) are also included[25]. 80% of common human variants found in GWAS studies localize in the non-coding sequences[26]. Multiple studies outline the importance of the vast information hidden in the non-coding DNA, especially in the DNA regulatory elements (such as enhancers) for gene regulation at the level of transcription and translation.

For HSCR we also have been searching for such non-coding variants. Comparative genomics approaches have identified a multispecies conserved enhancer region in intron1 of *RET* and within this enhancer region two strongly disease associated SNPs (rs2435357 and rs2506004) negatively influence the enhancer activity leading to lower RET expression[27,28]. These findings show that non-coding DNA variants play a role in the etiology of HSCR. So far the number of non-coding variants involved in HSCR is limited. This is due to the small sample size used for the GWAS analysis performed which makes finding common variants in these regions impossible and because we have not examined non-coding regions for rare or private variants. The reason for this is that we do not know where to look and how to deal with the possible non-coding variants. As a first step in the identification of rare and private non-coding variants which contribute to HSCR, we searched for enhancer regions in a set of known and candidate genes for HSCR.

The integrative analysis of the human epigenomic landscape for primary cells and tissues has revealed a genome-wide map of regulatory regions. Approximately 5% of each reference epigenome has enhancer and promoter signatures[29]. These signatures were made for many different tissues and for different developmental stages.

In **chapter 4**, we extracted gut-specific active enhancers from the human epigenome atlas for a set of genes known to be involved in ENS development. These gut specific enhancer regions for the most important ENS genes can be further investigated for mutation screening in HSCR patients. Moreover, we also searched for significantly enriched TF binding sites within these enhancer regions. We did this by using data for the sigmoid colon and from fetal large Intestine and compared this to control datasets (fetal thymus and fetal lung). Furthermore, in order to identify enriched functions for the identified TFs for the colon and fetal large intestine, we analyzed these genes using the Ingenuity Pathway Analysis (IPA) tool. An overlap between TF binding sites, identified from gut specific enhancers and upstream regulators of ENS genes, resulted in identification of the TFs JUN and TCF7L2 as significantly ($p<0.005$) enriched. These transcription factors, JUN and TCFL2, are associated with ERK/JNK and Wnt signaling pathways, respectively. Both these pathways have a role in ENS development.

Having identified the enhancers for the most important ENS associated genes makes it possible to screen for non-coding DNA variations in HSCR patients. It should be noted however that the gut specific enhancer dataset is derived from data sets of only two individuals (one adult and one fetal). It would be good to include more individuals at different stages of ENS development to be sure that we identified all important enhancers. The sequencing of these regions can be utilized by targeted sequencing followed up by reporter assays/ChIP/ATAC-seq as described by[28], to determine possible functional effects of the identified non-coding variants.

MODELLING HSCR: *IN VIVO* STUDIES

In parallel to linkage and association studies, syndromic diseases have been crucial in identifying disease associated genes and understanding disease processes. HSCR occurs as an isolated trait in 70% of the cases, associated with a chromosomal abnormality in 12% of cases, and the occurrence of additional congenital anomalies in 18% of the cases[30]. Trisomy 21 leading to Down syndrome (DS) is the most frequent chromosomal abnormality associated with HSCR. The overall incidence of DS ranges from 2-10% in all HSCR cases[30-32]. DS patients have >100 fold higher risk of developing HSCR than the general population[30].

None of the established HSCR genes are localized on chromosome 21 suggesting that overexpression of one or more genes on chromosome 21 may

have a substantial contribution to HSCR development. In **chapter 5**, we overexpressed mRNA of selected, highly conserved genes of chromosome 21 into a transgenic zebrafish reporter line *Tg (-8.3phox2b:Kaede)* embryos that express the fluorescent protein Kaede in ENS progenitors and found that elevated levels of one of the chromosome 21 genes, *ATP5O* resulted in altered ENS development and an HSCR-like phenotype. Moreover, we show that *ATP5O* is expressed in the zebrafish gut and in the myenteric and submucosal ganglia of human postnatal colon sections. This is the first demonstration of altering the expression of chromosome 21 genes in a zebrafish model to investigate their potential role in ENS development. Moreover, we also carried out a study to find any epistatic interaction of *ATP5O* with *ret* in zebrafish to investigate their potential interaction during ENS development, however our zebrafish data suggests that they act independently.

Our data suggest that *ATP50* might well be the link between DS and HSCR. However, strong evidence of *ATP5O* involvement with HSCR in humans is still lacking. We have not found any mutations in *ATP50* gene in isolated HSCR cases nor did we find any association of a variant to the disease. However, *ATP50* is localized in the HSCR-Down syndrome critical region.

Functional analysis of HSCR genetic data

As mentioned already, proving that the variants found by NGS are disease associated is difficult. A combination of *in vitro* and *in vivo* work can greatly help in proving causality. The use of animal models is possible as the pathways and mechanisms involved in ENS development are highly conserved[33,34].

Within our studies we opted for the zebrafish as our model system. Zebrafish (*Danio rerio*) emerged as an attractive vertebrate model for the geneticists almost 3 decades ago[35]. It has proven to be a powerful tool due to high fecundity, *ex utero* development, transparent embryos and the ease of genome manipulation and editing (Morpholinos and Crispr/Cas9). In **chapter 6**, we review the use of the zebrafish model in HSCR research. We discuss the use of zebrafish mutants, transgenic reporter lines and different available methods for conducting ENS functional screens using reverse genetics. Nearly all the HSCR genes involved with HSCR are required for ENS development in zebrafish as well[36]. *In silico* programs can predict the pathogenicity of the variant and *in vitro* assays also help to a certain extent, but the use of *in vivo* models is indispensable in understanding disease (HSCR) pathogenesis.

FUTURE PERSPECTIVE

Although previous human genetic studies and recent implementation of NGS has greatly transformed our understanding on HSCR, validating causality still remains challenging as is the identification of the genes mutated in the yet unexplained HSCR cases. We propose future research to focus on:

Coding mutations and Non-coding mutations

By screening larger cohorts by exome or whole genome sequencing one will be able to statistically prove the involvement of coding variants and non-coding variants in new genes. We hypothesize that non-coding mutations are far more important than we currently can prove. The identified gut regulatory regions (enhancers) as described in **Chapter 4** can be screened for non-coding DNA variants. The sequencing of the active enhancers of these ENS genes, to identify rare variants in the non-coding DNA, can greatly improve our understanding of noncoding genetic contributions to HSCR.

Somatic mutations

Besides coding and non-coding mutations it can be hypothesized that somatic mutations also play a significant role. So far, there is no comprehensive study on somatic mutations, mutations only present in the progenitors of the ENS. Comparison of exome sequencing data derived from blood DNA compared to exome sequencing data derived from ENCC of HSCR patients can reveal such somatic mutations.

Pathway disturbances

In a complex disease such as HSCR, disease development is caused by multiple genes in concert. During ENS development, NCC undergo extensive migration to colonize the most distal part of the colon and potentially these progenitor cells interact with many different molecular pathways, which may all influence their migratory potential. Genes, proteins and small molecules could interact to form a complex network known as modules or subnetworks. We can study these interacting partners using a network biology based approach to identify dysregulated pathways affecting ENS development. Functional annotation and validation of HSCR genes using a zebrafish model provides a quick and efficient method to determine their pathogenic effect on ENS development (as described in chapter 5, 6). We can also design similar experiments for epistatic interaction of newly identified genes to get an insight into disrupted pathways/mechanisms involved with ENS development.

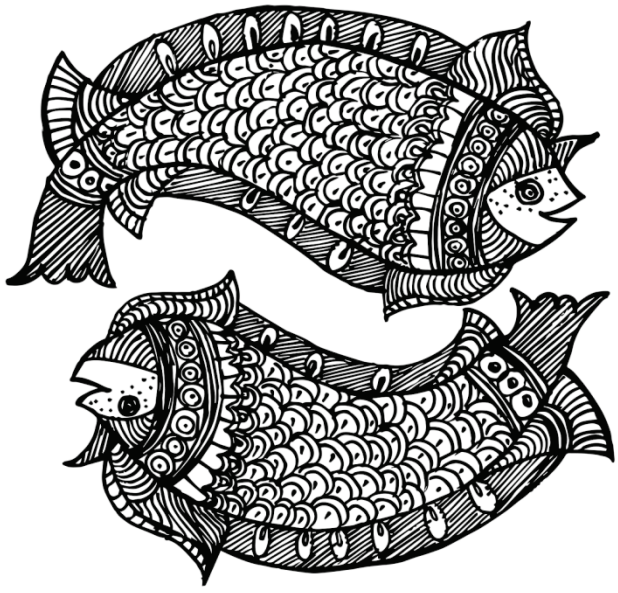
CONCLUSIONS

In this thesis we describe various approaches which we undertook to better understand and unravel the complexity of HSCR genetics: implementation of NGS in HSCR research; the contribution of rare *de novo* genetic variants in HSCR; non-coding DNA variants associated with HSCR using *in silico* approaches; reverse genetic approaches in zebrafish and a possible genetic link between HSCR and DS. Taken together the ultimate goal of deciphering the complexity of HSCR is far from complete and requires further investigations using a wide range of study designs. HSCR is no longer a lethal condition due to the implementation of modern surgical procedures, but a better understanding of HSCR genetics and the mechanisms involved in the disease pathogenesis would improve diagnosis, prevention and future treatment. This will benefit patients directly when they undergo genetic screening for their condition and will potentially offer scientists/clinicians new targets for the development of novel therapies to treat HSCR.

REFERENCES

1. Heanue, T.A. & Pachnis, V. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat Rev Neurosci* **8**, 466-79 (2007).
2. Alves, M.M. *et al.* Contribution of rare and common variants determine complex diseases-Hirschsprung disease as a model. *Dev Biol* **382**, 320-9 (2013).
3. Garcia-Barcelo, M.M. *et al.* Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc Natl Acad Sci U S A* **106**, 2694-9 (2009).
4. 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**, 581-96 (2015).
5. Kim, J.H. *et al.* A genome-wide association study identifies potential susceptibility loci for Hirschsprung disease. *PLoS One* **9**, e110292 (2014).
6. Hu, J.X., Thomas, C.E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* (2016).
7. Cirulli, E.T. & Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**, 415-25 (2010).
8. McClellan, J. & King, M.C. Genetic heterogeneity in human disease. *Cell* **141**, 210-7 (2010).
9. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* **19**, 212-9 (2009).
10. Veltman, J.A. & Brunner, H.G. De novo mutations in human genetic disease. *Nat Rev Genet* **13**, 565-75 (2012).
11. Badner, J.A., Sieber, W.K., Garver, K.L. & Chakravarti, A. A genetic study of Hirschsprung disease. *Am J Hum Genet* **46**, 568-80 (1990).
12. Harrison, C., Wabbersen, T. & Shepherd, I.T. In vivo visualization of the development of the enteric nervous system using a Tg(-8.3bphox2b:Kaede) transgenic zebrafish. *Genesis* **52**, 985-90 (2014).
13. Kok, F.O. *et al.* Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell* **32**, 97-108 (2015).

14. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230-3 (2015).
15. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**, 275-84 (2015).
16. Lyonnet, S. *et al.* A gene for Hirschsprung disease maps to the proximal long arm of chromosome 10. *Nat Genet* **4**, 346-50 (1993).
17. Angrist, M. *et al.* A gene for Hirschsprung disease (megacolon) in the pericentromeric region of human chromosome 10. *Nat Genet* **4**, 351-6 (1993).
18. Edery, P. *et al.* Long segment and short segment familial Hirschsprung's disease: variable clinical expression at the RET locus. *J Med Genet* **31**, 602-6 (1994).
19. Romeo, G. *et al.* Point mutations affecting the tyrosine kinase domain of the RET proto-oncogene in Hirschsprung's disease. *Nature* **367**, 377-8 (1994).
20. Puffenberger, E.G. *et al.* A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. *Cell* **79**, 1257-66 (1994).
21. Bolk, S. *et al.* A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus. *Proc Natl Acad Sci U S A* **97**, 268-73 (2000).
22. Gabriel, S.B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* **31**, 89-93 (2002).
23. Puffenberger, E.G. *et al.* Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum Mol Genet* **3**, 1217-25 (1994).
24. Brooks, A.S. *et al.* A novel susceptibility locus for Hirschsprung's disease maps to 4q31.3-q32.3. *J Med Genet* **43**, e35 (2006).
25. Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M.B. Annotating non-coding regions of the genome. *Nat Rev Genet* **11**, 559-71 (2010).
26. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
27. Emison, E.S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857-63 (2005).
28. Sribudiani, Y. *et al.* Variants in RET associated with Hirschsprung's disease affect binding of transcription factors and gene expression. *Gastroenterology* **140**, 572-582 e2 (2011).
29. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
30. Amiel, J. *et al.* Hirschsprung disease, associated syndromes and genetics: a review. *J Med Genet* **45**, 1-14 (2008).
31. Moore, S.W. & Johnson, A.G. Hirschsprung's disease: genetic and functional associations of Down's and Waardenburg syndromes. *Semin Pediatr Surg* **7**, 156-61 (1998).
32. Friedmacher, F. & Puri, P. Hirschsprung's disease associated with Down syndrome: a meta-analysis of incidence, functional outcomes and mortality. *Pediatr Surg Int* **29**, 937-46 (2013).
33. Burzynski, G., Shepherd, I.T. & Enomoto, H. Genetic model system studies of the development of the enteric nervous system, gut motility and Hirschsprung's disease. *Neurogastroenterol Motil* **21**, 113-27 (2009).
34. Heanue, T.A., Shepherd, I.T. & Burns, A.J. Enteric nervous system development in avian and zebrafish models. *Dev Biol* (2016).
35. Streisinger, G., Walker, C., Dower, N., Knauber, D. & Singer, F. Production of clones of homozygous diploid zebra fish (Brachydanio rerio). *Nature* **291**, 293-6 (1981).
36. Shepherd, I. & Eisen, J. Development of the zebrafish enteric nervous system. *Methods Cell Biol* **101**, 143-60 (2011).



Appendix

Summary

The enteric nervous system (ENS) innervates the wall of gastrointestinal (GI) tract and governs many functions such as gut motility, local blood flow, and mucosal transport. It is composed of a meshwork of neurons and glial cells that are organized in ganglia, which are found throughout the entire gut. The ENS is embryologically derived from the neural crest and failure of neural crest cells (NCC) to migrate, differentiate, proliferate or survive and thereby form a functional ENS leads to Hirschsprung disease (HSCR), also known as intestinal megacolon. HSCR is a congenital disease characterized by life threatening constipation, abdominal distention and vomiting or neonatal enterocolitis. The incidence of HSCR is approximately 1:5000 live births and varies with ethnicity.

HSCR is a complex genetic disorder that shows clinical variability, and Mendelian and non-Mendelian inheritance. In most cases HSCR occurs as an isolated trait, however, it is found associated with many other diseases or syndromes in 30% of cases. In ~90% of all cases the disease it not present in any family member (sporadic form), although in around 10% of all cases the disease occurs more often in a family (familial form). Many different genetic studies have been performed and have resulted in the identification of mutations in 16 genes and in 5 HSCR susceptibility loci. Mutations is *RET* (the major HSCR gene) have been found in around 50% of familial patients and 10-15% of sporadic patients. Mutations in genes other than *RET* occur predominantly in syndromic cases. However, all the mutations and variations identified in HSCR cases explain approximately 30% of the overall genetic risk. This has led to our hypothesis that there are probably several other yet unknown genes contributing to HSCR development. Thus the aim of this thesis is to finding novel HSCR genes, mutation within these genes and other non-coding regulatory regions that could explain part of the missing heritability.

Chapter 2 of this thesis describes the identification of novel genes harbouring *de novo* mutations in sporadic HSCR patients. In recent years, *de novo* mutations have been implicated in the etiology of several complex genetic diseases, including intellectual disability, autism and schizophrenia. Exome sequencing studies of 24 sporadic, non-syndromic HSCR patients led to the identification of 28 *de novo* mutations among 21 genes and 8 of them were present in *RET*. None of the newly identified genes, besides *RET,* were linked to known ENS signaling pathways or present in our HSCR cohorts. For functional validation of these genes, we tested them using morpholino mediated knockdown in *Tg (-8.3phox2b:Kaede)* transgenic zebrafish embryos that express the

fluorescent protein Kaede in ENS progenitors. Knockdown of 4 genes (*DENND3*, *NCLN*, *NUP98* and *TBATA*) resulted in gut aganglionosis in the zebrafish intestine mimicking a HSCR-like phenotype. In short, we have identified 4 candidate genes for HSCR that are known to be involved in CNS development and our studies now suggest that they are also involved in ENS development.

Historically, most of the HSCR genes and loci have been identified using linkage studies in multigenerational families. Such studies are even more suited in cases of rare diseases, where it is difficult to perform large scale GWAS studies due to the requirement of thousands of cases and controls. Linkage analysis has again emerged as an extremely useful approach for the genomic analysis of complex traits. A linkage study previously performed on a large Dutch multi-generational HSCR family revealed linkage to 4q31.3-q32.3.

Chapter 3 describes our exome sequencing study used to determine the genetic cause of HSCR in this family and the identification of a possible disease-causing mutation(s) in the linkage region. We identified one main candidate in the linkage interval in exon 20 of the *LRBA* gene. *In silico* prediction suggested an effect on mRNA splicing, but functional assays did not confirm this. We postulated that it could be an enhancer mutation for *MAB21L2* (a gene embedded within an intron of *LRBA*), but could not confirm that either. However, *mab21l2* is important for ENS development of zebrafish. Incomplete penetrance of HSCR in this family suggests the involvement of other (rare) variants elsewhere in the genome. Therefore, we searched for variants present especially in HSCR genes and genes associated with ENS development. We identified several candidates, among which were a rare *RET* coding variant in one branch of the family and inherited *IHH*, *GLI3* variants along with a *de novo* mutation in *GDNF* (RET ligand) in the second branch of the family. Functional studies confirmed the pathogenic nature of the variants identified in *RET* and *IHH,* confirming the importance of RET and Hedgehog signalling for ENS development. This study demonstrates that rare variants in multiple genes lead to the development of HSCR, further illustrating the complexity of HSCR genetics.

Most of the HSCR research has focused on the identification of coding variants in the genes associated with HSCR, although coding mutations do not explain more than ~30 % of all cases. The identification of non-coding DNA variants linked to complex diseases has been difficult, although many non-coding SNPs have been linked to human diseases. **Chapter 4** describes our strategy to identify novel regulatory elements involved in ENS development and potentially with HSCR. We used existing epigenome atlas datasets in conjunction with a computational approach to catalogue all the active enhancers of the sigmoid colon and the fetal large intestine. We mapped them to a set of 115 candidate

HSCR genes. We identified 20 transcription factors (TF) that are overrepresented in gut enhancers compared to the control dataset. Upstream regulator analysis of the ENS genes and an overlap with highly enriched TFs on gut enhancers resulted in identification of JUN and TCF7L2 as highly enriched TFs. Our studies offer datasets for screening non-coding mutations in regulatory regions of candidate HSCR genes. These datasets can be utilized for functional follow-up of GWAS loci, including fine mapping of GWAS signal(s).

HSCR occurs as an isolated trait in 70% of the cases, associated with a chromosomal abnormality in 12% of cases, and the occurrence of additional congenital anomalies in 18% of the cases. Trisomy 21, leading to Down syndrome (DS) is the most frequent chromosomal abnormality associated with HSCR. None of the established HSCR genes are localized on chromosome 21 suggesting that overexpression of one or more genes on chromosome 21 may contribute to HSCR development. **Chapter 5** describes our study to test this hypothesis. We overexpressed mRNA of selected, highly conserved chromosome 21 genes into a transgenic zebrafish reporter model *Tg (-8.3phox2b:Kaede).* We prioritized 21 genes and overexpressed them by microinjecting *in vitro* transcribed capped mRNAs into 1-cell-stage zebrafish embryos and scored the phenotypes at 5 days post fertilization (dpf). We showed that overexpression of *ATP5O (*ATP synthase, H^+ transporting, mitochondrial F1 complex, O subunit*)* leads to a disturbed ENS with a reduced number of enteric neurons, strongly implicating *ATP5O* as a contributor to a HSCR phenotype. *ATP5O* gene is highly expressed in mouse enteric NCC (E14.5), zebrafish gut and in enteric ganglia of human post-natal colon sections. Our findings suggest that an extra copy of this gene may contribute to HSCR development in patients with DS.

Use of NGS in solving the complex genetics of HSCR provides a list of many variants present in many different genes. Pinpointing the functional relevance of newly identified genetic variants in HSCR is rather difficult. Use of animal models has greatly helped in unravelling the function of many genes in development and disease. The zebrafish has emerged as a prominent vertebrate model for studying development processes and modeling human diseases and we therefore chose this model animal for our studies. In **chapter 6**, we review the use of the zebrafish model in HSCR research. We discuss the development of gut and ENS in zebrafish, the use of transgenic reporter lines and zebrafish mutants in understanding HSCR genetics. The zebrafish genome can be easily manipulated to knock-out target genes or induce point mutations. We also discuss the different available methods for conducting phenotypic analysis of the ENS in the zebrafish gut. A range of assays, such as enteric neuronal counts, gut motility studies,

intestinal transit and live cell tracking and imaging can be used to screen and assess the ENS.

Finally in **chapter 7**, we discuss the inference drawn from the research described in this thesis. We emphasize the role of rare coding and non-coding DNA variants in the etiology of HSCR and the use of a zebrafish model for studying ENS development. We conclude that although deciphering the genetic complexity of HSCR is not easy, our work, and that of others, has increased our understanding of the genetics of this disease.

Samenvatting

Het enterische zenuwstelsel (EZS) innerveert de wand van het maagdarmkanaal en reguleert vele functies zoals darmmotiliteit, lokale doorbloeding en mucosale transport. Het bestaat uit een vlechtwerk van neuronen en gliacellen die door ganglia georganiseerd zijn, deze bevinden zich overal in de gehele darm. Het EZS is embryologisch afgeleid van de neurale lijst en het falen van neurale cellen (NCC) bij het migreren, differentiëren, prolifereren en overleven waardoor er een functionele EZS leidt tot de ziekte van Hirschsprung (HSCR), ook wel bekend als intestinale megacolon. HSCR is een aangeboren ziekte die wordt gekenmerkt door levensbedreigende verstopping, opgezette buik en braken of neonatale enterocolitis. De incidentie van HSCR is ongeveer 1:5000 levendgeborenen en varieert met etniciteit.

HSCR is een complexe genetische aandoening die klinische variabiliteit laat zien en Mendeliaanse en niet-Mendeliaanse overerving. In de meeste gevallen treedt HSCR als een geïsoleerd eigenschap, maar het wordt vaak geassocieerd met vele andere ziekten of syndromen, in 30% van de gevallen. In ~ 90% van alle gevallen is de ziekte niet aanwezig in een familielid (sporadische vorm), hoewel de ziekte bij ongeveer 10% van alle gevallen vaker voorkomt bij een gezin (familiale vorm). Veel verschillende genetische studies zijn uitgevoerd en hebben geleid tot de identificatie van mutaties in 16 genen en in 5 HSCR gevoelige loci. In het *RET* gen (het major HSCR gen) zijn er mutaties gevonden in ongeveer 50% van familiale patiënten en in 10-15% van sporadische patiënten. Mutaties in andere genen dan *RET* zijn voornamelijk gevonden in syndromatische gevallen. Echter, alle mutaties en variaties die in HSCR gevallen zijn gevonden verklaren ongeveer 30% van het totale genetische risico. Dit heeft geleid tot onze hypothese dat er waarschijnlijk andere verschillende en tot nog toe onbekende genen bijdragen aan de ontwikkeling van HSCR. Het doel van dit onderzoek is het vinden van nieuwe HSCR genen, mutatie in deze genen en andere niet-coderende regulerende gebieden die de ontbrekende erfelijkheid deels kunnen verklaren.

Hoofdstuk 2 van dit proefschrift beschrijft de identificatie van nieuwe genen die de novo mutaties havens in sporadische HSCR patiënten. De laatste jaren zijn *de novo* mutaties betrokken geweest in de etiologie van verscheidene complexe genetische ziekten, waaronder verstandelijke beperking, autisme en schizofrenie. Exome sequencing studies van 24 sporadische, non-syndromale HSCR patiënten heeft geleid tot de identificatie van 28 *de novo* mutaties in 21 genen en waarvan 8 aanwezig waren in *RET*. Geen van de nieuw geïdentificeerde genen naast *RET*, waren gekoppeld aan bekende EZS signaalwegen of waren

aanwezig in onze HSCR cohorten. Voor de validatie van deze genen, hebben we ze getest met behulp van morfolino gemedieerde knock-down in *Tg (-8.3phox2b:Kaede)* transgene zebravis embryo's die het fluorescente eiwit Kaede produceren in EZS voorlopers. Knockdown van 4 genen (*DENND3*, *NCLN*, *NUP98* en *TBATA*) resulteerde in aganglionosis in de darm van de zebravis waarin het HSCR-achtige fenotype werd nagebootst. Kortom, we hebben 4 kandidaatgenen geïdentificeerd voor HSCR waarvan bekend is dat ze betrokken zijn bij de ontwikkeling van het centrale zenuwstelsel en onze studies suggereren nu dat ze ook betrokken zijn bij de ontwikkeling van EZS.

Historisch gezien zijn de meeste HSCR genen en loci geïdentificeerd via linkage studies in multi generatie families. Dergelijke studies zijn zelfs geschikter in het geval van zeldzame ziekten, waarbij het moeilijk is om grootschalige GWAS studies te verrichten vanwege het vereiste van duizenden patiënten en controles. Linkage analyse is weer naar voren gekomen als een zeer nuttige benadering voor de genomische analyse van ingewikkelde eigenschappen. In een linkage studie, eerder uitgevoerd op een grote Nederlandse multi-generatie HSCR familie, is er een koppeling gevonden van 4q31.3-q32.3.

Hoofdstuk 3 beschrijft onze exome sequencing studie die gebruikt is om de genetische oorzaak van HSCR in deze familie en de identificatie van een mogelijke ziekte-veroorzakende mutatie(s) in de linkage regio te bepalen. We hebben een belangrijke kandidaat geïdentificeerd in de koppeling interval in exon 20 van het gen *LRBA*. *In silico* voorspelling suggereerde een effect op mRNA-splitsing, maar functionele testen hebben dit niet bevestigd. We veronderstelden dat het een enhancer mutatie voor *MAB21L2* (een gen ingebed in een intron van *LRBA*) zou kunnen zijn, maar dat konden we niet bevestigen. Echter is *mab21l2* belangrijk voor EZS ontwikkeling bij de zebravis. Incomplete penetratie van HSCR in deze familie suggereert de betrokkenheid van andere (zeldzame) varianten elders in het genoom. Daarom zochten we naar varianten die vooral aanwezig zijn in HSCR genen en genen die geassocieerd zijn met EZS ontwikkeling. We identificeerden een aantal kandidaten, waaronder een zeldzame *RET* coderende variant in een tak van de familie en geërfde *IHH*, *GLI3* varianten samen met een de novo mutatie in *GDNF* (RET ligand) in de tweede tak van de familie. Functionele studies bevestigden de pathogene aard van de varianten die in *RET* en *IHH* zijn geïdentificeerd, waarmee het belang van de RET en Hedgehog signalering voor EZS ontwikkeling werd bevestigd. Deze studie toont aan dat zeldzame varianten in meerdere genen tot de ontwikkeling van HSCR leiden, die de complexiteit van HSCR genetica illustreren.

Het meeste HSCR onderzoek richt zich op de identificatie van coderende varianten in genen geassocieerd met HSCR, hoewel coderende mutaties niet meer

dan ~30% van alle gevallen kunnen verklaren. De identificatie van niet-coderend DNA varianten gekoppeld aan complexe ziekten is moeilijk, alhoewel veel niet-coderende SNPs zijn gekoppeld aan menselijke ziekten. **Hoofdstuk 4** beschrijft onze strategie om nieuwe regulerende elementen te identificeren die betrokken zijn bij de ontwikkeling van EZS en mogelijk met HSCR. We hebben bestaande epigenome atlas datasets gebruikt in combinatie met een computationele benadering om alle actieve enhancers van de sigmoïde colon en foetale dikke darm te catalogiseren. We hebben hen toegewezen aan een set van 115 kandidaat-HSCR genen. We hebben 20 transcriptiefactoren (TF) geïdentificeerd die overgerepresenteerd zijn in de darm enhancers vergeleken met de controle dataset. Opwaartse regulator analyse van de EZS genen en een overlap met hoogverrijkt TF's op darm enhancers leidde tot de identificatie van JUN en TCF7L2 als hoogverrijkt TF. Onze studies hebben datasets voor het screenen van niet-coderende mutaties in regulerende gebieden van kandidaat HSCR genen. Deze datasets kunnen worden gebruikt voor functionele follow-up van GWAS loci, inclusief het in kaart brengen van GWAS signaal.

HSCR komt voor als een geïsoleerde eigenschap in 70% van de gevallen, gepaard met een chromosomale afwijking in 12% van de gevallen en het voorkomen van bijkomende aangeboren afwijkingen bij 18% van de gevallen. Trisomie 21, wat leidt tot het syndroom van Down (DS), is de meest voorkomende chromosomale afwijking geassocieerd met HSCR. Geen van de vastgestelde HSCR genen zijn gelokaliseerd op chromosoom 21, suggererend dat overexpressie van één of meer genen op chromosoom 21 zouden kunnen bijdragen aan de ontwikkeling van HSCR. **Hoofdstuk 5** beschrijft onze studie om deze hypothese te testen. mRNA van geselecteerde, sterk geconserveerd chromosoom 21 genen in een transgeen zebravis reporter model *Tg (-8.3phox2b: Kaede)* werden in overexpressie gebracht. We prioriteerden 21 genen en brachten ze in overexpressie door *in vitro* getranscribeerd afgetopte mRNAs te injecteren in 1-cel-stadium zebravis embryo's en scoorden de fenotypes in 5 dagen na de bevruchting). We hebben aangetoond dat overexpressie van *ATP5O* (ATP synthase, H + transport, mitochondriaal F1 complex, O subunit) leidt tot een verstoord EZS met een verminderd aantal enterische neuronen en sterk betrokken *ATP5O* als bijdrage aan een HSCR fenotype. Het *ATP5O* gen komt hoog tot expressie in het enterische NCC (E14.5) van de muis, darm van de zebravis en in enterische ganglia van menselijke postnatale colon secties. Onze resultaten suggereren dat een extra kopie van dit gen kan bijdragen aan HSCR ontwikkeling bij patiënten met DS.

Het gebruik van NGS in het oplossen van de complexe genetica van HSCR geeft een overzicht van de vele varianten die aanwezig zijn in vele verschillende

genen. Het aanwijzen van de functionele relevantie van nieuw geïdentificeerde genetische varianten in HSCR is nogal moeilijk. Het gebruik van dierlijke modellen heeft enorm geholpen in het ontrafelen van de functie van vele genen in de ontwikkeling en ziekte. De zebravis heeft zich ontpopt als een prominent gewerveld model voor het bestuderen van de ontwikkeling van processen en het modelleren van ziekten bij de mens en daarom hebben we dus voor dit diermodel gekozen voor onze studies. In **hoofdstuk 6** bespreken we het gebruik van het zebravis model in HSCR onderzoek. We bespreken de ontwikkeling van darmen en EZS in zebravissen, het gebruik van transgene reporter lijnen en zebravismutanten voor het begrijpen van de HSCR genetica. Het zebravis genoom kan gemakkelijk worden gemanipuleerd om knock-out doelwit genen of puntmutaties te induceren. We bespreken ook de verschillende beschikbare methoden voor het uitvoeren van een fenotypische analyse van het EZS in de darmen van een zebravis. Een reeks assays, zoals enterische neuronale tellingen, darmmotiliteit studies, darmpassage en live cell tracking en imaging kunnen worden gebruikt voor het screenen en beoordelen van het EZS.

Tot slot in **hoofdstuk 7**, bespreken we de conclusie die is getrokken uit de in dit proefschrift beschreven onderzoek. Wij benadrukken de rol van zeldzame coderende en niet-coderende DNA varianten in de etiologie van HSCR en het gebruik van een zebravis model voor de studie van de ontwikkeling van het EZS. We concluderen dat, hoewel het ontcijferen van de genetische complexiteit van HSCR niet gemakkelijk is, heeft ons werk en dat van anderen ons begrip voor de genetica van deze ziekte vergroot.

Curriculum Vitae

Name	Rajendra Kumar Chauhan
Address	Spanjaardstraat 30 C, Rotterdam, Netherlands 3025 TR
Phone	+31652561711
E-Mail	rajchauhan65@gmail.com, r.chauhan@erasmumc.nl
D.O.B.	25/01/1984

Education

2012-2016	PhD student, Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands
2006-2009	MSc Bioinformatics, Birla Institute of Technology, Mesra, Ranchi, India
2002-2005	BSc (H) Botany, Delhi University, New Delhi, India

Professional experience

2012-2016	PhD student, Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands
2011-2012	PhD student, Department of Clinical Genetics, UMCG, Groningen, The Netherlands
2009-2011	Research Assistant, Institute of Genomics and Integrative Biology, New Delhi, India (Dr. Sridhar Sivasubbu)

List of Publications

Oligogenic inheritance in Hirschsprung disease: implications of RET and Hedgehog signaling in ENS development

Y. Sribudiani, **R.K.Chauhan**, L. Petrova, C. Harrison, T.D. Wabbersen, E. Brosens, B.M. de Graaf, G.Burzynski, R.W.W. Brouwer, W. F.J. van IJcken, A. de Klein, J. Osinga, B.J.L. Eggen, M.M. Alves, A.S. Brooks, I.T. Shepherd, R.M.W. Hofstra
2016 (Manuscript in preparation)

Identification of predictive regulatory sequences for gut development

R.K. Chauhan, M. van der Sijde, C. Paul, J. Fu, R.M.W. Hofstra, B.J.L. Eggen, A. Ijpma
2016 (Manuscript in preparation)

Overexpression of the chromosome 21 gene *ATP5O* results in fewer enteric neurons: the missing link between Down syndrome and Hirschsprung disease?

R.K. Chauhan, R. Lasabuda, D. Schriemer, W.W. Cheng , Z. Azmani, B.M. de Graaf , A.S. Brooks, S. Edie, R.H. Reeves, B.J.L. Eggen, A.J.Burns, I.T. Shepherd, R.M.W. Hofstra
2016 (Manuscript in preparation)

Functional analysis of HSCR genetic data using the zebrafish model

Rajendra K. Chauhan, Bart J. L. Eggen, Alan J. Burns, Iain T. Shepherd, Robert M.W. Hofstra, William. W. Cheng
2016 (Manuscript in preparation)

De novo mutations in Hirschsprung patients link central nervous system genes to the development of enteric nervous system

Hongsheng Gui, Duco Schriemer, William W.C. Cheng, **Rajendra K. Chauhan**, Guillermo Antiňolo, Courtney Berrios, Marta Bleda, Alice S. Brooks, Rutger W.W. Brouwer, Alan J. Burns, Stacey S. Cherny, Joaquin Dopazo, Bart J.L. Eggen, Paola Griseri, Binta Jalloh, Thuy-Linh Le, Vincent C.H. Lui, Berta Luzón-Toro, Ivana Matera, Elly S.W. Ngan, Anna Pelet, Macarena Ruiz-Ferrer, Pak C. Sham, Iain T. Shepherd, Man-Ting So, Yunia Sribudiani, Clara S.M. Tang, Mirjam C.G.N. van den Hout, Wilfred F.J. van IJcken, Joke B.G.M. Verheij, Jeanne Amiel, Salud Borrego, Isabella Ceccherini, Aravinda Chakravarti, Stanislas Lyonnet, Paul K.H. Tam, Maria-Mercè Garcia-Barceló & Robert M.W. Hofstra
2016 (Manuscript submitted)

Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of Mycobacterium tuberculosis pangenome

Periwal V, Patowary A, Vellarikkal SK, Gupta A, Singh M, Mittal A, Jeyapaul S, **Chauhan RK**, Singh AV, Singh PK, Garg P, Katoch VM, Katoch K, Chauhan DS, Sivasubbu S, Scaria V.,
PLoS One. 2015 Apr 8;10(4):e0122979. doi: 10.1371/journal.pone.0122979.

The Zebrafish GenomeWiki: a crowdsourcing approach to connect the long tail for zebrafish gene annotation

Singh M, Bhartiya D, Maini J, Sharma M, Singh AR, Kadarkaraisamy S, Rana R, Sabharwal A, Nanda S, Ramachandran A, Mittal A, Kapoor S, Sehgal P, Asad Z, Kaushik K, Vellarikkal SK, Jagga D, Muthuswami M, **Chauhan RK**, Leonard E, Priyadarshini R, Halimani M, Malhotra S, Patowary A, Vishwakarma H, Joshi P, Bhardwaj V, Bhaumik A, Bhatt B, Jha A, Kumar A, Budakoti P, Lalwani MK, Meli R, Jalali S, Joshi K, Pal K, Dhiman H, Laddha SV, Jadhav V, Singh N, Pandey V, Sachidanandan C, Ekker SC, Klee EW, Scaria V, Sivasubbu S.
Database (Oxford). 2014 Feb 26;2014:bau011. doi: 10.1093/database/bau011

A sequence-based variation map of zebrafish
Patowary A, Purkanti R, Singh M, **Chauhan R**, Singh AR, Swarnkar M, Singh N, Pandey V, Torroja C, Clark MD, Kocher JP, Clark KJ, Stemple DL, Klee EW, Ekker SC, Scaria V, Sivasubbu S.
Zebrafish. 2013 Mar;10(1):15-20. doi: 10.1089/zeb.2012.0848.

Reverse genetics screen in zebrafish identifies a role of miR-142a-3p in vascular development and integrity.
Lalwani MK, Sharma M, Singh AR, **Chauhan RK**, Patowary A, Singh N, Scaria V, Sivasubbu S.
PLoS One. 2012;7(12):e52588. doi: 10.1371/journal.pone.0052588. Epub 2012 Dec 21.

Systematic analysis and functional annotation of variations in the genome of an Indian individual
Patowary A, Purkanti R, Singh M, **Chauhan RK**, Bhartiya D, Dwivedi OP, Chauhan G, Bharadwaj D, Sivasubbu S, Scaria V.
Hum Mutat. 2012 Jul;33(7):1133-40. doi: 10.1002/humu.22091. Epub 2012 Apr 16.

De novo identification of viral pathogens from cell culture hologenomes
Patowary A, **Chauhan RK**, Singh M, Kv S, Periwal V, Kp K, Sapkal GN, Bondre VP, Gore MM, Sivasubbu S, Scaria V.
BMC Res Notes. 2012 Jan 6;5:11. doi: 10.1186/1756-0500-5-11.

Proteomic analysis of zebrafish (*Danio rerio*) embryos exposed to cyclosporine A
Ponnudurai RP, Basak T, Ahmad S, Bhardwaj G, **Chauhan RK**, Singh RA, Lalwani MK, Sivasubbu S, Sengupta S.
J Proteomics. 2012 Jan 4;75(3):1004-17. doi: 10.1016/j.jprot.2011.10.016. Epub 2011 Oct 29.

FishMap Zv8 update-a genomic regulatory map of zebrafish
Bhartiya D, Maini J, Sharma M, Joshi P, Laddha SV, Jalali S, Patowary A, Purkanti R, Lalwani M, Singh AR, **Chauhan R**, Singh N, Bhardwaj A, Scaria V, Sivasubbu S.
Zebrafish. 2010 Jun;7(2):179-80. doi: 10.1089/zeb.2009.0624.

Appendix

PhD Portfolio

Summary of PhD training and teaching

Name: PhD student: Rajendra Kumar Chauhan Erasmus MC Department: Clinical Genetics Research School: MGC	PhD period: 2012-2016 Promotor(s): Prof.dr. R.M.W. Hofstra Supervisor: Prof.dr. R.M.W. Hofstra	
1. PhD training		
	Year	**Workload (ECTS)**
General courses		
- Genetics course	2012	3
- Biochemistry and Biophysics	2012	3
- Cell and Developmental Biology	2012	3
- Literature Course	2012	2
- Laboratory animal science	2013	3
- Biomedical English Writing and Communication	2014	4
Specific courses (e.g. Research school, Medical Training)		
- Safely working in the laboratory, Leiden	2012	0.5
- Next Generation Sequencing (NGS) Data analysis, UMC, Utrecht	2013	1.5
- Functional Imaging and Super Resolution Microscopy course (OIC), Rotterdam	2013	3
- Epigenetic regulation in health and disease, Leiden	2012	1.5
- Microbiological Safety (VMT) course, Groningen	2014	1.5
- BD Flow Cytometry Course 2015 , Rotterdam	2016	0.5
Seminars and workshops		
- 19th MGC PhD Workshop, Dusseldorf, Germany	2012	1
- Photoshop/Illustrator CS5 Workshop (1 day)	2014	0.5
- 21st MGC PhD Workshop, Munster, Germany	2014	1
- 22nd MGC PhD Workshop, Maastricht	2015	1
- SURE Symposium 2015-The Next Step: Career Perspectives for Young Biomedical Scientists, Erasmus MC, Rotterdam	2015	0.5
Presentations		
- Poster presentation at "The 3rd International Symposium on development of enteric nervous system, cells and genes, Hongkong"	2012	1
- Poster presentation at 21st MGC PhD Workshop, Munster, Germany	2014	1
- Poster presentation Sophia Research day, Rotterdam, The Netherlands	2014	1
- Poster presentation at "The NVHG Autumn Symposium 2014, Papendal, Arnhem" The Netherlands	2014	1
- Poster presentation at "The 4th International Symposium on development of enteric nervous system, cells, signals, genes and therapy, Rotterdam" (19th to 22nd April 2015) The Netherlands	2015	1
- Oral presentation at 22nd MGC PhD Workshop, Maastricht, The Netherlands	2015	0.5
- Oral presentation at Sophia Research day 2015, Rotterdam, The Netherlands	2015	0.5
- Oral presentation at American Society of Human Genetics(ASHG), Annual Meeting, Baltimore, USA	2015	1

- Dept. Clinical Genetics Research Meeting	2012-2015	4
(Inter)national conferences - Dutch Stem Cell Meeting , Amsterdam, The Netherlands - The 3rd International Symposium on Development of ENS, Cells and Genes, Hongkong - The NVHG Autumn Symposium 2014, Papendal, Arnhem, The Netherlands - The 4th International Symposium on development of enteric nervous system, cells, signals, genes and therapy, Rotterdam, The Netherlands - American Society of Human Genetics(ASHG), Annual Meeting, Baltimore, USA	 2012 2012 2014 2015 2015	 0.5 2 1 2 2
Other -		
2. Teaching		
	Year	**Workload (Hours/ECTS)**
Supervising students Rizky Lasabuda's 6 month internship (Student, HAN University of Applied Sciences, Nigmegen, The Netherlands)	2014	2

Total ECTS **51**

Acknowledgements

The journey of my PhD had tremendous contribution from my colleagues, friends and family. Looking back, it seems that the time just flew and it was all your constant support, guidance and love, which made me accomplish my PhD. The time spent with you all would be cherished in my memory forever.

First and foremost, my appreciation goes to my supervisor and PhD promoter **Prof. dr. R.M.W. Hofstra**. Dear **Robert**, Thanks for giving me the opportunity to work with you and offering invaluable advice throughout my PhD. You are a superb supervisor and a great human. I have learnt and transformed my scientific understanding to another level with your guidance and the opportunities provided by you. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist.

I am also indebted to my co-promoter **Dr. B.J.L. Eggen**. Dear **Bart**, thanks for your constant support and advice throughout these years. Many thanks for allowing me to work in your lab and helping me to enhance my knowledge on epigenetics. My second co-promoter **Dr. Alan J. Burns**, Dear **Alan**, I would also like to thank you for happily sharing your knowledge and going through my manuscripts. Your suggestions and inputs on every project were very helpful.

I am very grateful to **Prof. dr. Dick Tibboel**, **Prof. Dr. Rob Willemsen** and **Dr. Iain Shepherd** for being part of my PhD thesis assessment committee. Many thanks for reviewing my thesis and for your invaluable comments. Dear **Rob**, thanks for all the help around the lab, zebrafish imports etc. Dear **Iain**, Thanks a lot for all the collaborative work and coming over from Atlanta for my PhD defence. Thanks to **Dr. Robbert Rottier** and **Dr. Alice Brooks** for being part of the committee. Dear **Robbert**, It was always nice of you to chair the GI meetings and exchange our research in the field. Dear **Alice**, Thanks a lot for your critical views and making me explain about the importance of clinical genetic findings in HSCR research.

Special thanks to my amazing paranymphs **Shami** and **Katherine**. Dear **Shami**, thanks for dutch translation. You are always very helpful, understanding and full of joy. I wish you good luck for everything. Dear **Katherine**, my best wishes for your PhD work and thanks for help with the cover. You are such a cheerful person in our group. I would like to thank you both for your immense help in organizing my PhD defence at its best.

Many thanks to my former and current colleagues of GI motility research group. Dear **Yunia**, thanks a lot for sharing your knowledge and guiding me in the beginning of my PhD for the experiments all the way from Groningen to Rotterdam. Dear **Jan**, thanks for always being so cheerful and ready to help with

&

experiments during my stint at UMCG. I am grateful to **Bianca**, for all the help with experiments in different projects. I hope you will enjoy travelling Nepal and sometimes we will surely meet in India. Dear **Erwin**, your expertise in exome data analysis had been a tremendous help for both the Chr 4 paper and ongoing somatic project. Dear **William,** thanks for sharing your knowledge on zebrafish and helping out with finishing up my PhD thesis. Hopefully, we would keep on meeting at some zebrafish meetings. Dear **Duco**, thanks for the collaborative work on my thesis and glad to know that you are enjoying your new position. Dear **Rhiana**, it was nice working with you on the somatic project and best wishes with your PhD thesis. Dear **Maria**, thanks for the nice discussions and for your experimental inputs. Thanks to **Kaushal** for nice talks and hope you are enjoying your work in UK. I would miss the lunch session with you all. **Danny**, my friend! , thanks for being a nice colleague and officemate since last 4 years. Best wishes for your PhD defence.

I recognize that this research would not have been possible without the financial support from **Clinical genetic department** of Erasmus MC, research grants from **ZonMW** (TOP-subsidie 40-00812-98-10042) and the **Maag lever Darm stichting** (W009-62). Sincere gratitude to all the families involved in the study presented in this thesis.

Thanks to all the members of **The International Hirschsprung Disease Consortium**, it was nice to work together with you all. Thanks to our collaborators at UCL, London especially **Dr. Nikhil Thapar**, **Dipa**, **Julie**, **Jean-Marie** and **Conor** for making my stay comfortable at UCL and helping me with the mice experiments. I am also thankful to our collaborator **Prof. Roger Reeves** and **Sarah Edie** from Johns Hopkins University, USA. I would like to thank **Arne Ijpma** (Bioinformatics Department, Erasmus MC) for helping me out with Enhancer work. Thanks to **Jing** and **Marijke** from UMCG, Groningen for their help on enhancer work.

I am thankful to my student **Rizky** for working together on Chr21 work and best wishes for your Masters study. Thanks to **Zakia** for helping with zebrafish work and good luck with your NGS expedition.

I would like to express my gratitude to different facilities (**Biomics**, **Imaging**, **FACS**, **IT** and **Animal**) within the Erasmus MC. Thanks to the diagnostics group of clinical genetics department for their help with DNA. Thanks to **Dr. Wilfred van Ijcken**, **Rutger**, **Frank**, **Mirjam**, **Christel**, **Frank** and **Zehlina** for your help with next generation sequencing studies. Many thanks to **Dr. Alex Maas**, **Edwin**, **Judith** and **Dominique** for your help with mice and zebrafish work.

During my PhD, I also had pleasure to work on collaborative projects with **Maria Enguix** from Spain; **Veerle** and **Natalie** from UMC, Maastricht and thanks for discussions on those projects and it was pleasure working with you.

During last years, my officemates from Ee928, 930a and 930c, **Celine**, **Andrea**, **Atze**, **Erik**, **Ronald**, **Fenne**, **Roy**, **Judith**, **Laura**, **Nynke**, **Qiushi** and **Mike** it was really a great pleasure and fun sharing office with you all. Thanks for nice chats, laughs, comparing h-indexes and taking time to also go through Dutch correspondences now and then for me. Mijn Dutch docent **Ronald**, Bedankt voor helpen! Thanks for the enormous fun in MGC workshops with many of you. *"Whatever happens in MGC stays in MGC"*, so not gonna share it here.

Experience speaks louder than words! Big thanks to the experienced Lab gurus, **Guido**, **Herma**, **Liess-anne**, **Marianne** for all the help in lab related work. Many thanks to the former colleagues and current members of the Clinical Genetics Research Group at Erasmus MC for all the scientific discussions and suggestions, non-scientific conversations in the corridor, at coffee machine and especially during the borrels. Thanks to **Arnold**, **Marian**, **Aida**, **Mark**, **Renate**, **Pim**, **Tjakko**, **Vincenzo**, **Annelies**, **Grazia**, **Gerben**, **Wim**, **Merel**, **Mark**, **Josja**, **Simone**, **Christian**, **Marialuisa**, **Roy**, **Adriana**, **Monica**, **Leontine**, **Rachel**, **Michelle**, **Ingeborg**, **Helen**, **Stijn**, **Joon** and **Tom** for creating nice working atmosphere.

Special thanks go to **Jeannette**, **Bep** and **Job** for their secretarial support and making my life easy in Netherlands. You are so kind and helpful all the time. I would also like to thank **Marjoleine** for introducing me to MGC and with arranging all the courses. Dear **Tom**, many thanks for fine-tuning of my cover along with all the photographic support.

I would also like to thank other colleagues (Reproduction and biology) of 9th floor, who often used to meet at the courses, cell culture labs, corridors and parties for having all the fun together.

I would also like to thank our past and new neighbor cum friends, **Shimriet** & **Kjell**, **Chaitanya** & **Vivek**, **Vaibhav**, **Noorie** & **Prasanth.** It was nice to know you guys. We really enjoyed your company and had good time over dinners. In my time at Erasmus MC, I got to know some people from various departments; those became good friends, **Daphne** & **Ivo**, **Simar** and **Sarvanan** thanks for all the nice chats and discussions.

Overseas relocation is hectic and some friends and families have always helped me in overcoming that hurdle of isolation. Thanks to all the Indian friends from Groningen, especially **Suresh**, **Abhishek**, **Khayum**, **Kushi**, **Gaurav**, **Shiva**, **Pramod.** Big thanks to GISA board members for organizing different functions

&

together and thereafter as well. Heartfelt thanks to **Tushar** & **Jasmine** for always be welcoming.

Moving from Groningen to Rotterdam was a big change and gradually got to know some Indian families here that helped me in many ways. Thanks to **Reddy family**, **Ank** & **Jim**, **Anette & Jason**, and **Srinivas** & **Sujata** for lovely get-togethers. You have guided me in many ways and made me feel like home in Rotterdam.

I would also like to express my sincere gratitude to my previous supervisor, **Dr. Sridhar Sivasubbu** and **Vinod Scaria** at IGIB for introducing me to zebrafish biology and next generation sequencing. Thanks to my friends from IGIB, **Prateek** & **Ramya**, **Jayant**, **Sunny**, **Ashok** & **Meghna** for all the support and making sure that we still meet at different places in the world. Many thanks to the faculties of my BSc (Zakir Husain College, New Delhi) and MSc (Birla Institute of Technology, Mesra) for their teachings. Thanks to my close friend **Nandu** for always ready to help me whenever required.

I would like to thank my family in India, who have provided unconditional support and encouraged me with everything. Special thanks to the whole **Chauhan Family**. I owe gratitude to **Mummy ji** & **Daddy Ji** and my sister **Kavita** for always helping me and showering your love and blessings. I understand that it would be difficult to come here for my PhD defence, but I would see you soon in India. I would also like to thank **Paul family**, **Maa** and **Bapi**, my parents-in-law for all the love, care and blessings. Thanks for believing in me and providing strength all the time. Thanks to **Bulu** da and **Apu** da (**Dutta's**) for all the help and always standing by in need.

Last but not the least; I would like to thank my better half, **Chaitali** for joining me here in Rotterdam and transforming my life. You have helped me throughout my PhD in every possible way. Thanks for the work done on the Enhancer chapter and designing the lovely cover and layout of the thesis. You made me feel this journey shorter and lovable. Thanks for your support and encouragement.

This was an attempt to thank most of you and I apologize, if I have forgotten to mention anyone.

Rajendra