Noncoding *RET* variants explain the strong association with Hirschsprung disease in patients without rare coding sequence variant

Valtteri B. Virtanen^{a,b,*}, Perttu P. Salo^{c,d,e}, Jia Cao^f, Anna Löf-Granström^{f,g}, Lili Milani^h, Andres Metspalu^h, Risto J. Rintala^a, Outi Saarenpää-Heikkiläⁱ, Tiina Paunio^j, Tomas Wester^{f,g}, Agneta Nordenskjöld^{f,g}, Markus Perola^{c,d,e,h}, Mikko P. Pakarinen^{a,b}

^a Section of Pediatric Surgery, Hospital for Children and Adolescents, University of Helsinki, Finland

^b Pediatric Liver and Gut Research Group, University of Helsinki, Finland

^c National Institute for Health and Welfare, Helsinki, Finland

^d Institute for Molecular Medicine, Finland (FIMM), University of Helsinki, Finland

^e Diabetes and Obesity Research Program, University of Helsinki, Finland

^f Department of Women's and Children's Health, Karolinska Institutet and Center of Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden

^g Pediatric Surgery, Astrid Lindgren Children Hospital, Karolinska University Hospital, Stockholm, Sweden

^h The Estonian Genome Center, University of Tartu, Estonia

ⁱ Department of Pediatrics, Tampere University Hospital, Tampere, Finland

^j Department of Health, National Institute for Health and Welfare and Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Finland

ARTICLE INFO

Keywords:

Hirschsprung disease
RET
Genetics
GWAS

ABSTRACT

The pathogenesis of Hirschsprung disease is complex. Although the *RET* proto-oncogene is the most frequently affected gene in Hirschsprung disease, rare coding sequence variants explain only a small part of Hirschsprung disease cases. We aimed to assess the genetic background of Hirschsprung disease using a genome-wide association analysis combined with sequencing all *RET* exons in samples from 105 Hirschsprung disease cases (30 familial and 75 sporadic) and 386 controls.

As expected, variants in or near *RET* showed the strongest overall association with Hirschsprung disease and the most statistically significant association was observed when using a recessive genetic model (rs2435357, NC_000010.10:g.43582056T > C; genotype TT, OR = 17.31, $P = 1.462 \times 10^{-21}$). Previously published associations in variants in *SEMA* (rs11766001, NC_000007.13:g.84145202A > C; allele C, OR = 2.268, $P = 0.009533$) and *NRG1* (rs4541858, NC_000008.10:g.32410309A > G; allele G, OR = 1.567, $P = 0.015$; rs7835688, NC_000008.10:g.32411499G > C; allele C, OR = 1.567, $P = 0.015$) were also replicated in the genome-wide association analysis. Sequencing revealed a total of 12 exonic *RET* rare variants. Of these, eight amino acid changing rare variants and two frameshift variants caused or possibly caused Hirschsprung disease.

Only a minority of the Hirschsprung disease cases (9/30 familial; 7/75 sporadic) carried one of the rare variants. Excluding the rare variant carriers from the genome-wide association analysis did not appreciably change the association of rs2435357 with Hirschsprung disease. We estimate that approximately two thirds of the sporadic cases may be statistically attributed to the recessive action of the common non-coding *RET* variants. Thus, even though most cases do not carry rare *RET* variants, combinations of rare variants and the common non-coding *RET* variant cause the majority of the cases in our population.

1. Introduction

Hirschsprung disease (HD) is the most common congenital intestinal motility disorder present in 1/4000 European newborns (Best et al., 2014). It is characterized by the absence of ganglion cells in the

myenteric and submucosal plexuses in the distal bowel (Heanue and Pachnis 2007). HD is thought to be caused by disordered migration and differentiation of neural crest cells during embryonic development and is thereby regarded as a neurocristopathy. The timing of the disruption in neural cell migration defines the length of the aganglionic segment,

* Corresponding author. Section of Pediatric Surgery, Hospital for Children and Adolescents, University of Helsinki, Stenbäckinkatu 11, PL 281, FI-00029 HUS, Helsinki, Finland.

E-mail address: valtteri.virtanen@helsinki.fi (V.B. Virtanen).

<https://doi.org/10.1016/j.ejmg.2018.07.019>

Received 24 February 2018; Received in revised form 6 June 2018; Accepted 17 July 2018

Available online 18 July 2018

1769-7212/ © 2018 Elsevier Masson SAS. All rights reserved.

which most commonly reaches the rectosigmoid level (80%) but sometimes extends to the more proximal colon (15–20%) or to the small intestine (5%). HD is treated surgically by removing the aganglionic part of the intestine. Disorders of anorectal and bowel function are common after operative treatment and a significant proportion of patients with extended aganglionosis remain dependent on parenteral nutrition without intestinal transplantation (Pakarinen et al., 2013). If untreated, HD leads to intestinal obstruction, enteritis, severe malnutrition and even death.

HD displays a complex inheritance pattern. Family history of the disease is reported in 7%–20% of cases, being the most common in total colonic aganglionosis (Amiel et al., 2008). So far, over 20 genes have been associated with HD including *RET*, *EDNRB*, *SOX10*, *SEMA3C*, *SEMA3D*, and other genes involved in the NTF-3/TRKC, prokineticin, NRG, and SEMA signalling pathways (Fernández et al., 2009, Garcia-Barcelo et al., 2009, Jiang et al., 2012, Ruiz-Ferrer et al., 2008, Wang et al., 2011, Yang et al., 2013). Rare coding sequence variants of these genes are often reported in HD patients with rare phenotypes such as familial disease and long segment aganglionosis, and female patients (Garcia-Barcelo et al., 2009). Most HD cases, however, do not appear to carry rare coding sequence variants in any of the genes associated with HD. The majority of HD cases is thus attributed to the disease alleles with incomplete penetrance.

The *RET* proto-oncogene is the most frequently affected gene in HD. Depending on study cohort, rare function affecting variants in the coding region of *RET* are found in approximately 40% of the familial and 20% of the sporadic cases (Emison et al., 2010). The penetrance of *RET* variants is incomplete and only about half of children with inactivating *RET* variants display the HD phenotype (Amiel et al., 2008). Common low-penetrance variants in the non-coding regions in or near *RET* are also associated with risk for HD. The variant rs2435357, located within a conserved enhancer element in the first intron of *RET*, has been shown to associate with HD and to alter the expression of *RET* in the gastrointestinal tract (Miao et al., 2010). The incomplete penetrance of *RET* variants and the identification of novel HD candidate genes suggest that factors other than those affecting *RET* coding sequence play an important role in the development of HD. Here, we sought to comprehensively assess the contribution of genetic variation in or near *RET* to the risk for developing HD in a well-characterized population using a genome-wide association analysis (GWAS) and *RET* exon sequencing.

2. Materials and methods

2.1. Patients and background data

A total of 105 patients treated for HD at our tertiary center volunteered to participate during the study period between 2007 and 2013. Healthy representatives of the general population ($n = 386$) from the Child Sleep Study were used as controls. Data abstracted from the completely reviewed patient records included indication to surgery, type of the operation, operative details, length of the aganglionic segment and associated disorders. The diagnosis of HD was based on the disease history, operative findings and the histology of the resected colonic specimen featuring the absence of ganglion cells together with an increased acetylcholinesterase staining.

Replication of the identified genome-wide significantly associated variants for new candidate genes was performed with the co-operation of Karolinska Institutet. Their independent population included 154 HD patients and 177 healthy Caucasian controls. All controls were collected anonymous as placenta samples with data on ethnicity of the parents. It was also confirmed that the child was without any known malformation at birth. In addition, the replication of previously reported variants in and near *SEMA* and *NRG1* region was proofed in this study.

2.2. Genotyping and sequencing

DNA was extracted from whole blood samples during 2007–2013. All samples ($n = 491$) were genotyped using the Illumina HumanOmniExpress BeadChip using standard protocols at the Core Facility of the Estonian Genome Center, University of Tartu, Estonia. We used GRCh37 (hg19) as the reference genome throughout the study. We removed variants and individuals with genotyping rate $< 95\%$, variants with $P < 5 \times 10^{-3}$ for an exact test of Hardy-Weinberg equilibrium in controls, variants with $P < 0.05$ for a test of differential missingness between cases and controls (PLINK –test-missing), and variants violating Mendelian inheritance in any of the genotyped pedigrees. Then we pre-phased and imputed the data with shapeIT v2. r644 and Impute v2.3.0 using the 1000 Genomes Project (1000G) integrated phase1 v3 reference haplotypes. Finally, we removed variants with minor allele frequency (MAF) < 0.01 or Impute “info” metric < 0.9 .

From each HD patient as well as from five healthy first-degree relatives from familial HD patients all 21 exons of *RET* cDNA were PCR amplified and nucleotide sequences were resolved with the ABI3730xl DNA Analyzer capillary electrophoresis instrument (Applied Biosystems) using standard methods. The remaining seven healthy first-degree relatives were genotyped if necessary based on the sequencing results. The reference sequence for *RET* was NC_000010.10. The rare variants have been submitted to the ClinVar database (<http://www.ncbi.nlm.nih.gov/clinvar/>).

The replication was performed using DNA of 154 Swedish HD patients and 177 healthy Caucasian controls. The samples were genotyped using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) and the ABI 3730 DNA Sequencer (Applied Biosystems, Foster City, CA, USA) with standard procedure; the sequences were analyzed by CodonCode Aligner V3.71 (CodonCode Co., USA).

2.3. Variant classification

Identified variants were classified into five categories based on their predicted effect on the gene function using the Ensembl Variant Effect Predictor (<http://grch37.ensembl.org/info/docs/tools/vep/index.html>): affects function, probably affects function, unknown, probably does not affect function and does not affect function. The effects of amino acid substitutions were predicted using SIFT (either tolerated or deleterious) and PolyPhen2 (probably damaging, possibly damaging, benign or unknown). The difference in the allele frequency between cases and controls was also used in the classification. For the variants identified by sequencing HD cases, the allele frequencies in the general (control) population were obtained from the Exome Aggregation Consortium database (ExAC; <http://exac.broadinstitute.org>) and the allele frequencies were compared to both Finnish and non-Finnish Europeans. Variants were classified based on minor allele frequency (MAF) in the general population as common (MAF $> 5\%$), low-frequency (MAF 1–5%), or rare (MAF $< 1\%$).

2.4. Statistical analyses

Capillary sequencing traces were analyzed using the ABI Variant Reporter and Sequencer software packages. PLINK v1.90b3w was used for data management and association testing. The association between genetic variants and disease status was tested using additive, recessive, and dominant genetic models with a logistic regression model for variants with MAF > 0.10 and Fisher's exact test for the rarer variants. For GWAS, the P-value limit for statistical significance was set at 5×10^{-8} . For the replication, the P-value limit for the statistical significance was set at 0.05. Population attributable fractions (PAF, the proportion of the disease in the population that is attributable to exposure) were calculated using the R Epi package.

Table 1
Subject characteristics.

	Familial HD (n = 30)	Sporadic HD (n = 75)	P-value ^a
Females	6 (20%)	15 (20%)	1.00
Level of aganglionosis			
Rectosigmoid	22 (73%)	69 (92%)	
Long segment	4 (13%)	3 (4%)	0.08 LS vs. RS
TCA	4 (13%)	0	0.005 TCA vs. RS
Unclear	0	3 (4%)	

HD, Hirschsprung disease; TCA, total colon aganglionosis; LS, long segment. RS, rectosigmoid.

^a Fisher exact test.

2.5. Ethics

The study protocol was approved by the Ethics committee of the Hospital for Children and Adolescents, University of Helsinki, by the Pirkanmaa Hospital District and by the Ethics Committee at Karolinska Institutet, Stockholm.

3. Results

3.1. Subject characteristics

A total of 105 HD patients participated including 30 familial and 75 sporadic HD cases (Table 1). Of the 30 familial HD patients 25 had affected first degree relatives. There was a statistically significant difference in the length of aganglionosis between familial and sporadic HD patients (Table 1), total colon aganglionosis being more common in familial disease ($P = 0.005$ Fisher exact test). One sporadic female HD patient with short segment disease had Down syndrome. One patient had cartilage hair hypoplasia, one was paraplegic and deaf, and one was blind and had Marfan syndrome. In addition, one patient had been treated for medullary thyroid carcinoma and one for multiple endocrine neoplasia 2A.

3.2. GWAS of the whole cohort

After genotyping and imputation, a total of 8 432 870 variants passed quality control. As expected, variants in or near *RET* showed the strongest overall association with HD and the most statistically significant association was observed when using a recessive genetic model (lead variant rs2505991, NC_000010.10:g.43566360A > G; genotype AA, OR = 22.63, $P = 6.873 \times 10^{-23}$). The previously described common regulatory *RET* variant rs2435357, NC_000010.10:g.43582056T > C showed a similar association (genotype TT, OR = 17.31, $P = 1.462 \times 10^{-21}$) under the recessive model (Fig. 1a). For the additive and dominant models, the corresponding

associations of rs2435357 with HD were considerably weaker (allele T, OR = 5.747, $P = 1.391 \times 10^{-16}$ and OR = 5.356, $P = 4.438 \times 10^{-7}$, respectively) (Fig. 1b). The genotype relative risk (GRR) for TT vs. CC genotypes was substantially higher than for CT vs CC (21.8 vs 1.54) (Table 2). Previous studies had similar findings with the strongest association under the recessive model (Table 2). In total, nearly a hundred variants in or near *RET* were genome-wide significantly associated with HD, but all were statistically non-significant ($P > 0.05$) after the conditional analysis using rs2435357 as a covariate.

3.3. Replication of newly identified genome-wide significantly associated variants and previously reported variants

The association of previously reported variants near *SEMA* and *NRG1* with HD was also nominally statistically significant (rs11766001, NC_000007.13:g.84145202A > C in *SEMA*, allele C, OR = 2.268, $P = 0.0005328$ and rs4541858, NC_000008.10:g.32410309A > G in *NRG1*, allele G, OR = 1.567, $P = 0.01503$) using logistic regression and the additive genetic model (Table 3) (Gunadi et al., 2014, Jiang et al., 2015). Much stronger effect was seen using recessive model (rs11766001, allele A, OR = 18.75, $P = 0.007853$ and rs4541858, allele A, OR = 1.282, $P = 0.0299$). Given the prior information on the association of these variants with HD, their association with the disease thus nominally replicated in this study even if the variants did not meet the pre-specified limit of genome-wide significance.

In addition, three variants outside of the *RET* locus showed genome-wide significant association. Two variants (rs2347320, NC_000003.11:g.16666312C > A in chromosome 3 and rs4922081, NC_000008.10:g.19536790C > A in chromosome 8) were genome-wide significantly associated but their association was not supported by nearby SNPs indicating possible genotyping errors (Fig. 1b). Upon manual inspection of the probe intensities the genotype calls were deemed erroneous due to poor clustering. Rs11120336, NC_000001.10:g.214660748C > A on chromosome 1 was genome-wide significantly associated (allele A, OR = 3.053, $P = 1.752 \times 10^{-8}$) when using the additive genetic model. To validate this association, the SNP rs11120336 was genotyped in an independent sample of 154 HD patients and 177 Caucasian controls and tested for association using logistic regression and the additive genetic model. These HD patients included 18 familial (12%), 12 long segment (7,8%) and 41 females (27%) cases. The SNP did not replicate in the independent data set (allele A, OR = 1.282, $P = 0.20$).

3.4. Coding *RET* variants

A total of 12 rare variants (RVs) were identified and they were divided among 16 of 105 patients (Table 4). These included one synonymous variant, eight missense variants, one non-sense variant, one frameshift insertion and one frameshift deletion. All RVs were traced from the ExAC database and the allele frequencies were compared to

Fig. 1. a. Manhattan plot for all HD patients; recessive model. b. Manhattan plot for all HD patients; additive model.

Table 2

Genotype distributions of rs2435357 in cases and controls in our study and in previous studies.

Author	Year	Genotype	Cases/Controls	Penetrance ^a	Genotype relative risk ^b	
Virtanen all HD patients	2018	CC	12/140	6.03×10^{-5}	1.54	CT vs. CC
		CT	17/129	9.27×10^{-5}		
		TT	56/30	0.00131		
Virtanen sporadic not affecting function	2018	CC	6/140	3.83×10^{-5}	2.35	CT vs. CC
		CT	13/129	9.00×10^{-5}		
		TT	49/30	0.00146		
Kapoor	2015	CC	95/345	9.81×10^{-5}	1.65	CT vs. CC
		CT	109/240	0.000162		
		TT	148/42	0.001254		
Zhang	2015	CC	1/16	1.25×10^{-5}	8.53	CT vs. CC
		CT	16/30	0.000107		
		TT	42/13	0.000646		
Gunadi	2014	CC	4/24	6.56×10^{-5}	1.24	CT vs. CC
		CT	14/68	8.10×10^{-5}		
		TT	42/26	0.000635		
Miao	2010	CC	22/95	4.79×10^{-5}	1.66	CT vs. CC
		CT	65/169	7.96×10^{-5}		
		TT	228/62	0.000761		
Pini Prato	2009	CC	5/50	7.73×10^{-5}	1.87	CT vs. CC
		CT	6/32	0.000145		
		TT	11/3	0.00283		
Arnold	2009	CC	16/18	8.60×10^{-5}	1.64	CT vs. CC
		CT	16/11	0.000141		
		TT	30/1	0.0029		
Zhang	2007	CC	14/41	9.11×10^{-5}	33.7	CT vs. CC
		CT	28/62	0.00012		
		TT	57/29	0.000524		

^a Penetrance calculated from the reported genotype frequencies using Bayes' theorem assuming 1/5000 as the population prevalence.^b Genotype relative risk calculated from the penetrances using the least penetrant genotype as reference.**Table 3**Allele frequencies for main single nucleotide polymorphisms^a.

SNP	Chr.	Position	Location	Alleles (Ref/Alt)	Allele frequencies cases/controls	P- Value	Odds Ratio	R-squared ^c
rs2505991 ^b	10	43 566 360	RET	G/A	0.7470/0.2905	5.546×10^{-17}	5.913	
rs2435357	10	43 582 056	RET	C/T	0.7588/0.3161	1.391×10^{-16}	5.747	0.9360
rs2506030	10	43 447 847	RET	A/G	0.6353/0.4103	1.719×10^{-6}	2.344	0.1897
rs1974511 ^b	7	83 991 521	SEMA3	A/C	0.4000/0.2475	0.0002035	1.971	
rs11766001	7	84 145 202	SEMA3	A/C	0.2176/0.1130	0.0005328	2.268	0.3902
rs12707682	7	84 443 356	SEMA3	T/C	0.2866/0.1915	0.008261	1.737	0.2365
rs4445183 ^b	8	32 562 443	NRG1	C/T	0.4353/0.3106	0.001686	1.832	
rs4541858	8	32 410 309	NRG1	A/G	0.3647/0.4684	0.01503	1.567	0.0671
rs7835688	8	32 553 981	NRG1	G/C	0.3647/0.4684	0.01503	1.567	0.0671

SNP, single nucleotide polymorphism; Chr., Chromosome; Ref, Reference; Alt, Alternate.

hg19 was used as the genomic reference sequence.

^a For this table were chosen our strongest SNP from each of these genes and the best SNPs from previous studies.^b Our strongest SNP in this gene, not previously reported.^c R-squared of the strongest SNP in the GWAS and the previously reported HD-associated variants in the same gene.

Finnish and European (Non-Finnish) (Table 4). Of these RVs all but the synonymous variant V125V (c.375C > A) and one missense variant R982C (c.2944C > T) were affecting or probably affecting the gene function. The affecting function of the RV R982C remains unknown because it has also been discovered from healthy controls and the frequency of the rarer allele is relatively high in some populations (Svensson et al., 1998). RVs tended to occur more frequently in patients with familial (9/30; 30%) than sporadic (7/75; 9%) disease ($P = 0.0141$, Fisher exact test). RVs which were affecting the gene function included a novel non-sense variant W37* (c.111G > A) in exon 2 and a novel missense variant F147S (c.440T > C) in exon 3. A novel frameshift-introducing insertion c.2864_2865insT was identified in exon 17 and a novel frameshift introducing deletion c.3143delT in exon 19. Five remaining coding sequence rare variants were C609R (c.1825T > C), C611R (c.1831T > C), C620R (c.1858T > C) in exon 10, Y791F (c.2372A > T) in exon 13 and R813W (c.2437C > T) in exon 14. In addition, one *de novo* RV V202M (c.604G > C) was found in exon 3 from one familial HD patient with small intestinal

aganglionosis. Using capillary sequencing, this was confirmed to be absent from his healthy mother and, surprisingly from his father suffering from rectosigmoid aganglionosis and from his brother suffering from total colon aganglionosis. The identity of the DNA sample was confirmed using the GWAS genotypes showing the correct relatedness metrics with the father and mother (PI HAT ~ 0.5). There are probably multiple genetic factors causing HD in this family.

3.5. GWAS after exclusion of patients with causative rare variants

Rare function affecting *RET* variants explained 16 of the 105 HD cases. In addition, two familial patients had a frameshift variant in *EDNRB* identified in parallel performed whole exome sequencing. Four familial patients had a deletion covering the whole *PHOX2B* gene identified in parallel performed multi-locus sequence analysis (MLSA), leaving 68 sporadic and 15 familial HD cases without identified rare function affecting variants. To explore the association of common SNPs with HD in the absence of rare high-impact variants, we next excluded

Table 4
Observed rare variants in all exons of RET with their clinical correlations in HD patients.

RV	Chr.	location	Exon	Alleles (Ref/Alt)	Allele frequencies (cases/controls Fin/controls Eur)	Type of variant	Affects function	Age (years)	Familial HD	Sex	Level of aganglionosis
c.111G > A (n = 2) ²	43	595 944	2	G/A	0.00952/NA/NA	nonsense	yes	19 ³ , 46 ³	yes	M, F	RS, RS
c.375C > A (n = 2) ²	43	597 827	3	C/A	0.00952/0.00363/0.00973	synonymous	no	4 ¹ , 34 ⁴	yes, yes	M, M	RS, RS
c.440T > C (n = 2) ²	43	597 892	3	T/C	0.00952/NA/NA	missense	yes	5 ⁵ , 33 ^{1,5}	yes, yes	M, M	TCA, RS
c.604G > C (n = 1) ²	43	598 056	3	G/A	0.00476/NA/NA	missense	yes	4	yes	M	TCA
c.1825T > C (n = 1) ²	43	609 069	10	T/C	0.00476/NA/NA	missense	yes	31	no	M	RS
c.1831T > C (n = 1) ²	43	609 075	10	T/C	0.00476/NA/NA	missense	yes	35	yes	M	RS
c.1858T > C (n = 1) ²	43	609 102	10	T/C	0.00476/NA/NA	missense	yes	37	no	M	RS
c.2372A > T (n = 3) ²	43	613 908	13	A/T	0.0143/0.00579/0.00264	missense	yes	38, 45, 50	all no	F, M, M	all RS
c.2437C > T (n = 1) ²	43	615 023	14	C/T	0.00476/NA/1.991 × 10 ⁻⁵	missense	yes	51	no	M	RS
c.2864_2865insT (n = 3) ²	43	619 181	17	C/T	0.0143/NA/NA	frameshift	yes	18 ^{1,6} , 23 ^{1,6} , 46 ^{1,6}	all yes	F, M, M	RS, LS, LS
c.2944C > T (n = 5) ²	43	620 335	18	C/T	0.0238/0.00635/0.01840	missense	unknown	33 ^{1,5} , 40 ¹ , 18 ^{1,6} , 23 ^{1,6} , 46 ^{1,6}	yes, no, yes, yes	M, M, F, M, M	RS, RS, RS, LS, LS
c.3143delT (n = 1) ²	43	622 126	19		0.00476/NA/NA	frameshift	yes	22	no	M	RS

¹ Identical age corresponding same patient.

² Number of identified variants in cases.

^{3,4,5,6} Each number indicates one family.

Reference sequence NC_000010.10.

Chr., Chromosomal; Fin, Finnish; Eur, European (Non-Finnish); RV, Rare variant; HD, Hirschsprung disease.

M, male; F, female; RS, rectosigmoid; TCA, total colon aganglionosis; LS, long segment.

NA; not applicable (not found from the ExAC database).

all HD patients with function affecting RVs and the remaining familial cases and repeated the GWAS. The SNP rs2435357 showed similar association as before (genotype TT, OR = 23.12, P = 2.9 × 10⁻²¹) under the recessive model but the genotype relative risk was stronger than in the whole cohort, the GRR for rs2435357 was 38.1 (TT vs. CC) and 2.35 (CT vs. CC) (Table 3).

3.6. Calculation of the population attributable fraction for rs2435357

PAF was calculated for sporadic cases to estimate the fraction of cases explained by common non-coding RET variation in absence of other identified causal factors. For this analysis, all familial cases were excluded as well as sporadic cases with possible gene function affecting variants. PAF for rs2435357 was calculated using different genetic models. The dominant genetic model produced the largest value for the PAF estimate (81%), followed by the additive (73%) and the recessive models (69%).

4. Discussion

We combined GWAS with RET exon sequencing to simultaneously assess the contribution of genome-wide common variants and rare coding RET variants to the risk for developing HD. Sequencing all 21 exons of RET revealed 10 RVs which were affecting the gene function in 16 of the 105 (15%) HD patients in our sample. GWAS in the same patient cohort identified numerous common variants in or near RET that were significantly associated with the risk for HD. Common variants tagged the previously reported regulatory variant rs2435357, located in the first intron of RET. Previously published associations of SNPs in SEMA and NRG1 replicated also in our sample. The population-attributable fraction estimate of rs2435357 varied between 69% and 81% for idiopathic sporadic HD cases depending on the genetic model. As these sporadic cases formed 64% of our patients, about half of the cases in our entire sample may be statistically attributed to the common non-coding RET variants. Thus, approximately two out of three HD cases in Finland may be attributed to either rare RET coding region variants or by the common noncoding variant affecting RET.

The common RET variant associate with typical disease forms in HD whereas rare coding sequence variants are more often associated with rare phenotypes such as familial disease and long segment aganglionosis, and female patients (Garcia-Barcelo et al., 2009). As expected, rare coding sequence variants of RET were more common among familial HD patients. On the other hand, their frequency showed no difference between males and females (P = 1.0) or between rectosigmoid and long segment aganglionosis.

The strong association between common RET variants and HD is evident (Arnold et al., 2009; Gunadi et al., 2014; Kapoor et al., 2015; Miao et al., 2010; Pini Prato et al., 2009; Zhang et al., 2007; Zhang et al., 2015). Variants located in the promoter region and intron 1 of RET have been most studied, especially rs2435357. Practically in each study the strongest association for rs2435357 have been reported for TT vs. CC which is against the common suggestion for additive or dominant genetic models (Table 3). These common SNPs in or near RET rather seem to have the strongest effect in a homozygote when homozygous, following a recessive mode of inheritance. This has been rarely discussed previously (Burzynski et al., 2004). In our study the GRR of rs2435357 (TT vs. CC) was strong among sporadic patients with no known variant which affects function as well as in the whole cohort (GRR 38.1 vs. 21.8). This underlines the role of common noncoding RET variants in the pathogenesis of HD in patients with no coding sequence variants affecting the gene function.

Previous GWAS studies have discovered few candidate genes for HD, including SEMA and NRG1 (Gunadi et al., 2014; Jiang et al., 2015). In our study the association with HD was replicated in both genes. The dys-regulation of SEMA3C and SEMA3D without any existing RET variant is considered to cause enteric aganglionosis whereas the effect

of *NRG1* genotypes are considered to depend on prevailing *RET* genotypes (Gunadi et al., 2014; Jiang et al., 2015).

It is crucial to clarify the affect function of the examined *RET* variants for accurate patient counselling and planning of future studies. Two of the detected frameshift variants had a damaging effect on the protein and were therefore affecting the gene function in four HD patients. The novel W37* (c.111G > A) nonsense variant was found in a family with two affected patients causing premature stop codon and was as well affecting the gene function. The novel F147S (c.440T > C) missense variant was found in another family with two affected patients. Because of a possibly damaging effect of the variant it was also considered to affect or probably affect the gene function. Of the identified variants, C609R (c.1825T > C), C611R (c.1831T > C), C620R (c.1858T > C) and Y791F (c.2372A > T) are known to associate with both HD and medullary thyroid cancer (Virtanen et al., 2013; Romeo et al., 1998). An isolated variant R813W (c.2437C > T) was found in one sporadic HD patient. This variant has been considered as a probably damaging variant in previous studies and it remains unclear if it affects the gene function (Griseri et al., 2002). Four of the familial HD patients with rare affecting function *RET* variants also carried the R982C (c.2944C > T) variant (Table 4). R982C might also be a modifying variant and its pathogenic significance in one affected sporadic HD patients without any other variants is uncertain. The affect function of R982C has been widely discussed and without any other RV it may not cause HD (Svensson et al., 1998). The *de novo* V202M (c.604G > C) variant in a familial HD patient might contribute to HD alone or it might be a modifying variant contributing to more severe HD phenotype (Julies et al., 2001). There is probably some major variant causing HD in other affected family members (father and son) such as a frameshift variant in an intronic region of *RET* or a variant with affecting function in other genes. The RV V125V (c.375C > A) was a silent variant.

Despite extensive research genetic etiology of HD remains unclear in the most of patients. Our study confirms the key role of *RET* in HD and demonstrates the importance of non-coding sequence variants of *RET*. Exploration of an oligogenic trait such as HD with GWAS tools also gives us an important lesson regarding the complexity we are facing when dissecting the genetic background of more complex traits in the post GWAS era.

Conflicts of interest

There are no conflicts of interest.

Acknowledgements

This work was financially supported by the Helsinki University Central Hospital Grant, the Sigrid Jusélius Foundation, the Finnish Pediatric Research Foundation, the EU FP7 under grant agreements nr. 313010 (BBMRI-LPC), nr. 305280 (MIMOmics) and nr. 261433 (BioSHaRE-EU), the Finnish Academy grant no. 269517, the Yrjö Jahnsson Foundation, Academy of Finland, grant nr. 134880 and nr. 253346, Signe and Ane Gyllenberg Foundation, the Juho Vainio Foundation, the Foundation Frimurarna, Stockholm, Sweden, IUT20-60 by Estonian Research Council and by the EU through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012).

References

Amiel, J., Sproat-Emission, E., Garcia-Barcelo, M., Lantieri, F., Burzynski, G., Borrego, S., Pelet, A., Arnold, S., Miao, X., Griseri, P., Brooks, A.S., Antinolo, G., de Pontual, L., Clement-Ziza, M., Munnich, A., Kashuk, C., West, K., Wong, K.K., Lyonnet, S., Chakravarti, A., Tam, P.K., Ceccherini, I., Hofstra, R.M., Fernandez, R., 2008. Hirschsprung disease, associated syndromes and genetics: a review. *J. Med. Genet.* 45, 1–14.

Arnold, S., Pelet, A., Amiel, J., Borrego, S., Hofstra, R., Tam, P., Ceccherini, I., Lyonnet, S., Sherman, S., Chakravarti, A., 2009. Interaction between a chromosome 10 *RET* enhancer and chromosome 21 in the Down syndrome-Hirschsprung disease association. *Hum. Mutat.* 30, 771–775.

Best, K.E., Addor, M.C., Arriola, L., Balku, E., Barisic, I., Bianchi, F., Calzolari, E., Curran, R., Doray, B., Draper, E., Garne, E., Gatt, M., Haeusler, M., Bergman, J., Khoshnood, B., Klungsoyr, K., Martos, C., Materna-Kiryluk, A., Matias Dias, C., McDonnell, B., Mullaney, C., Nelen, V., O'Mahony, M., Queisser-Luft, A., Randrianaivo, H., Rissmann, A., Rounding, C., Sipek, A., Thompson, R., Tucker, D., Wellesley, D., Zymak-Zakutnia, N., Rankin, J., 2014. Hirschsprung's disease prevalence in Europe: a register based study. *Birth Defects Res A Clin Mol Teratol* 100, 695–702.

Burzynski, G.M., Nolte, I.M., Osinga, J., Ceccherini, I., Twigg, B., Maas, S., Brooks, A., Verheij, J., Plaza Menacho, I., Buys, C.H., Hofstra, R.M., 2004. 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, 604–612.

Emison, E.S., Garcia-Barcelo, M., Grice, E.A., Lantieri, F., Amiel, J., Burzynski, G., Fernandez, R.M., Hao, L., Kashuk, C., West, K., Miao, X., Tam, P.K., Griseri, P., Ceccherini, I., Pelet, A., Jannot, A.S., de Pontual, L., Henrion-Cau, A., Lyonnet, S., Verheij, J.B., Hofstra, R.M., Antinolo, G., Borrego, S., McCallion, A.S., Chakravarti, A., 2010. Differential contributions of rare and common, coding and noncoding ret mutations to multifactorial Hirschsprung disease liability. *Am. J. Hum. Genet.* 87, 60–74.

Fernández, R.M., Sánchez-Mejías, A., Mena, M.D., Ruiz-Ferrer, M., López-Alonso, M., Antinolo, G., Borrego, S., 2009. A novel point variant in *NTRK3*, R645C, suggests a role of this gene in the pathogenesis of Hirschsprung disease. *Ann. Hum. Genet.* 73, 19–25.

Garcia-Barcelo, M.M., Tang, C.S., Ngan, E.S., Lui, V.C., Chen, Y., So, M.T., Leon, T.Y., Miao, X.P., Shum, C.K., Liu, F.Q., Yeung, M.Y., Yuan, Z.W., Guo, W.H., Liu, L., Sun, X.B., Huang, L.M., Tou, J.F., Song, Y.Q., Chan, D., Cheung, K.M., Wong, K.K., Cherny, S.S., Sham, P.C., Tam, P.K., 2009. Genome-wide association study identifies *NRG1* as a susceptibility locus for Hirschsprung's disease. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2694–2699.

Griseri, P., Pesce, B., Patrone, G., Osinga, J., Puppo, F., Sancandi, M., Hofstra, R., Romeo, G., Ravazzolo, R., Devoto, M., Ceccherini, I., 2002. A rare haplotype of the *RET* proto-oncogene is a risk-modifying allele in Hirschsprung disease. *Am. J. Hum. Genet.* 71, 969–974.

Gunadi, Kapoor, A., Ling, A.Y., Rochadi, Makhmudi, A., Herini, E.S., Sosa, M.X., Chatterjee, S., Chakravarti, A., 2014. Effects of *RET* and *NRG1* polymorphisms in Indonesian patients with Hirschsprung disease. *J. Pediatr. Surg.* 49, 1614–1618.

Heanue, T.A., Pachnis, V., 2007. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nat. Rev. Neurosci.* 8, 466–479.

Jiang, Q., Turner, T., Sosa, M.X., Rakha, A., Arnold, S., Chakravarti, A., 2012. Rapid and efficient human mutation detection using a bench-top next-generation DNA sequencer. *Hum. Mutat.* 33, 281–289.

Jiang, Q., Arnold, S., Heanue, T., Kilambi, K.P., Doan, B., Kapoor, A., Ling, A.Y., Sosa, M.X., Guy, M., Jiang, Q., Burzynski, G., West, K., Bessling, S., Griseri, P., Amiel, J., Fernandez, R.M., Verheij, J.B., Hofstra, R.M., Borrego, S., Lyonnet, S., Ceccherini, I., Gray, J.J., Pachnis, V., McCallion, A.S., Chakravarti, A., 2015. Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with *ret* are critical to Hirschsprung disease liability. *Am. J. Hum. Genet.* 96, 581–596.

Julies, M.G., Moore, S.W., Kotze, M.J., du Plessis, L., 2001. Novel *RET* mutations in Hirschsprung's disease patients from the diverse South African population. *Eur. J. Hum. Genet.* 9, 419–423.

Kapoor, A., Jiang, Q., Chatterjee, S., Chakravarti, P., Sosa, M.X., Berrios, C., Chakravarti, A., 2015. Population variation in total genetic risk of Hirschsprung disease from common *RET*, *SEMA3* and *NRG1* susceptibility polymorphisms. *Hum. Mol. Genet.* 15, 2997–3003.

Miao, X., Leon, T.Y., Ngan, E.S., So, M.T., Yuan, Z.W., Lui, V.C., Chen, Y., Wong, K.K., Tam, P.K., Garcia-Barcelo, M., 2010. Reduced *RET* expression in gut tissue of individuals carrying risk alleles of Hirschsprung's disease. *Hum. Mol. Genet.* 19, 1461–1467.

Pakarinen, M.P., Kurvinen, A., Koivusalo, A.I., Ruuska, T., Mäkelä, H., Jalanko, H., Rintala, R.J., 2013. Surgical treatment and outcomes of severe pediatric intestinal motility disorders requiring parenteral nutrition. *J. Pediatr. Surg.* 48, 333–338.

Pini Prato, A., Musso, M., Ceccherini, I., Mattioli, G., Giunta, C., Ghiggeri, G.M., Jasonni, V., 2009. Hirschsprung disease and congenital anomalies of the kidney and urinary tract (CAKUT): a novel syndromic association. *Medicine (Baltimore)* 88, 83–90.

Romeo, G., Ceccherini, I., Celli, J., Priolo, M., Betsos, N., Bonardi, G., Seri, M., Yin, L., Lerone, M., Jasonni, V., Martucciello, G., 1998. Association of multiple endocrine neoplasia type 2 and Hirschsprung disease. *J. Intern. Med.* 243, 515–520.

Ruiz-Ferrer, M., Fernandez, R.M., Antinolo, G., Lopez-Alonso, M., Borrego, S., 2008. *NTF-3*, a gene involved in the enteric nervous system development, as a candidate gene for Hirschsprung disease. *J. Pediatr. Surg.* 43, 1308–1311.

Svensson, P.J., Anvret, M., Molander, M.L., Nordenskjöld, A., 1998. Phenotypic variation in a family with mutations in two Hirschsprung-related genes: *RET* and endothelin receptor B. *Hum. Genet.* 103, 145–148.

Virtanen, V.B., Pukkala, E., Kivisaari, R., Salo, P.P., Koivusalo, A., Arola, J., Miettinen, P.J., Rintala, R.J., Perola, M., Pakarinen, M.P., 2013. Thyroid cancer and co-occurring *RET* mutations in Hirschsprung disease. *Endocr. Relat. Canc.* 20, 595–602.

Wang, L.L., Fan, Y., Zhou, F.H., Li, H., Zhang, Y., Miao, J.N., Gu, H., Huang, T.C., Yuan, Z.W., 2011. Semaphorin 3A expression in the colon of Hirschsprung disease. *Birth Defects Res A Clin Mol Teratol* 91, 842–847.

Yang, J., Duan, S., Zhong, R., Yin, J., Pu, J., Ke, J., Lu, X., Zou, L., Zhang, H., Zhu, Z., Wang, D., Xiao, H., Guo, A., Xia, J., Miao, X., Tang, S., Wang, G., 2013. Exome sequencing identified *NRG3* as a novel susceptible gene of Hirschsprung's disease in a Chinese population. *Mol. Neurobiol.* 47, 957–966.

Zhang, X.N., Zhou, M.N., Qiu, Y.Q., Ding, S.P., Qi, M., Li, J.C., 2007. Genetic analysis of *RET*, *EDNRB*, and *EDN3* genes and three SNPs in MCS + 9.7 in Chinese Patients with isolated Hirschsprung disease. *Biochem. Genet.* 45, 523–527.

Zhang, Z., Jiang, Q., Li, Q., Cheng, W., Qiao, G., Xiao, P., Gan, L., Su, L., Miao, C., Li, L., 2015. Genotyping analysis of 3 *RET* polymorphisms demonstrates low somatic mutation rate in Chinese Hirschsprung disease patients. *Int. J. Clin. Exp. Pathol.* 8, 5528–5534.