7

GENERAL DISCUSSION
Hirschsprung disease (HSCR) is a congenital disorder in which the muscles in a variable segment of the distal colon are unable to relax. This results in constipation, vomiting and distention of the bowel proximal to the affected segment. The lack of colonic relaxation in HSCR stems from an incomplete development of the enteric nervous system (ENS). During embryonic development almost all enteric neural crest cells (ENCCs) that form the ENS enter the foregut and proliferate, survive and migrate towards the distal hindgut, before they differentiate into neurons and glial cells\(^1\). A distortion of any of these processes leads to an incomplete colonization of the colon by ENCCs, lack of neuronal innervation, and hence failure of the colonic smooth muscles to relax.

HSCR is an inherited disease for which mutations in over 20 genes and 5 associated genomic loci have been found. Collectively they explain ~25% of the observed heritability\(^2\)–\(^4\). The majority of these genes has been identified by linkage analysis and sequencing of candidate genes. However, over the past decade genetic research has been greatly facilitated by technological developments in genome-wide association studies (GWAS) and next-generation sequencing (NGS). Three GWAS on predominantly sporadic HSCR patients have been conducted\(^5\)–\(^7\). In addition, rare variants in the associated loci and in yet undiscovered HSCR genes will likely contribute to the genetic risk of this complex genetic disease. These hypotheses can be tested by exome- or genome sequencing of (large) cohorts of HSCR patients.

**NEXT-GENERATION SEQUENCING IN COMPLEX GENETIC DISEASES**

**Study designs to identify rare, pathogenic variants**
Common variants have successfully been associated with disease in GWAS, but have a low penetrance. Rare disease-associated variants on the other hand usually have a higher penetrance and are therefore more informative for understanding disease etiology and in genetic counseling\(^8\). Rare variants can be identified on a genome/exome-wide scale by NGS, but large-scale rare variant association studies are far less abundant than GWAS. The main reason for this is the high cost per sample in NGS compared to GWAS. Genetic power calculations show that rare variant associations studies require similar numbers of cases and controls as GWAS, meaning that NGS studies have less statistical power than equally expensive GWAS\(^9\). However, the cost of NGS is declining rapidly, thereby opening new
opportunities for rare variant association studies. However, even at reducing costs sequencing thousands of cases and controls remains an expensive ordeal and patient numbers may not be available for diseases that are rare. For these reasons the question arises whether it is worthwhile to sequence small numbers of patients (and controls) in complex genetic diseases.

Case-control analysis on a limited number of sporadic patients will not have sufficient statistical power to find genome-wide significance, but other study designs may be successful with small numbers of individuals. Multiplex families remain a valuable source for genetic testing. For example, exome sequencing of 29 patients and 11 controls from 14 late-onset Alzheimer disease pedigrees successfully identified rare variants in \textit{PLD3} in two families\textsuperscript{10}.

If sporadic patients are being studied, several approaches have been suggested to increase the power of a rare variant association study. In \textbf{chapter 5} we applied these strategies to analyze exome sequencing data from 48 sporadic HSCR cases and 212 controls. We maximized the power of the study by prioritizing long-segment HSCR cases (which have the highest heritability), selecting rare, pathogenic variants, collapsing all variants per gene and performing a meta-analysis on data from three sequencing centers. Since we found no genome-wide associations, we analyzed whether the highest associated genes (by nominal p-value) are good candidate genes for HSCR. Gene prioritization tools consistently predicted \textit{CELSR1}, \textit{CLOCK} and \textit{FASN} as the most likely disease-causal genes. In addition, \textit{GHDC} is a poorly characterized gene for which no connection to the ENS was found, but the gene was highly expressed by ENCCs and contained disrupting variants.

Predicting the most likely disease-causal genes in an underpowered genetic study based on gene function or expression is yet no evidence for involvement in the disease. Even when assuming that there are no real associations with HSCR, some genes will have more significant p-values than others and these genes could be of interest in the disease context just by chance. The gene prioritization and gene expression data do therefore not identify new disease-associated genes, but should be regarded as means to identify the best candidate genes for follow up studies on a limited number of genes. Moreover, small rare variant association studies can be used to assess the frequency of rare variants in the candidate genes in cases and controls, which can be used to calculate odds ratios. These are important for power calculations in follow-up studies.
Identification of *de novo* mutations in NGS data

As said small studies are almost all underpowered. However, NGS can be successfully applied on relatively small numbers of sporadic disease patients for the identification of *de novo* mutations (chapter 2). Exome sequencing studies have revealed a role for *de novo* mutations in the genetic etiology of rare congenital syndromes and more common neurodevelopmental disorders\(^\text{11}\). However, the identification of *de novo* mutations in exome sequencing data is not trivial, since the number of true *de novo* mutations is much lower than the number of false positives that arise from sequencing errors and incorrectly called genotypes. Joint genotype calling of offspring-parent trios\(^\text{12}\) and statistical modelling efforts\(^\text{13–17}\) have improved *de novo* mutation detection significantly. In chapter 4 we analyzed if and how *de novo* mutations can be detected by applying filtering thresholds for reference allele ratio, sequencing depth and Genotype Quality score. We found that by removing the maximum assigned value of the Genotype Quality score, *de novo* mutations could be detected with high sensitivity and specificity in our training data and in a second, independent dataset.

We restricted our analysis to three sequencing parameters (reference allele ratio, sequencing depth and Genotype Quality score) that were reported to efficiently filter errors from NGS data\(^\text{18,19}\). However, other parameters such as calling quality, mapping quality, quality/depth ratio and [depth in proband]/[depth in parent] ratio have also been proposed as discriminating parameters for *de novo* mutation detection\(^\text{17}\). A combined approach where the unmaximized Genotype Quality score and other sequencing parameters are incorporated in a statistical model may be most powerful in detecting *de novo* mutations.

**GENETICS OF HSCR**

**Identification of four new HSCR genes in a *de novo* mutation screen**

In our exome sequencing on sporadic, non-syndromic HSCR patients, we predominantly included patients with long-segment HSCR, since this subtype has the highest heritability and follows a dominant mode of inheritance with incomplete penetrance\(^\text{20}\). Rare variants with a large effect size are therefore expected to make a large contribution to long-segment HSCR. In chapter 2 we tested the hypothesis that rare variants in sporadic, long-segment HSCR occur *de*
We identified 28 de novo mutations in 24 HSCR patients. Eight of these de novo mutations were found in RET, the major HSCR gene, supporting a role for de novo mutations in the etiology of long-segment HSCR. Moreover, this finding, and the overrepresentation of loss of function de novo mutations, suggests that de novo mutations in other genes may contribute to HSCR as well.

To overcome the mentioned problems one encounters when sequencing small cohorts of patients, we carried out unbiased in silico and in vivo analyses for the 20 genes (besides RET) that harbor de novo mutations to determine whether they are involved in the development of the ENS. RET and CKAP2L were enriched for rare variants in HSCR patients compared to controls, but only RET could be replicated in an independent cohort. The 20 genes with de novo mutations could not be linked to the signaling pathways in HSCR, but all were expressed by either ENCCs from E14.5 mouse embryos or by human iPS cell-derived neural crest cells. Since our in silico analyses did not yield convincing evidence for a role of the de novo mutated genes in HSCR, we decided to functionally test our candidate genes using the transgenic Tg(-8.3phox2b:Kaede) zebrafish model. Morpholino-mediated knockdown of the 12 genes that had non-synonymous de novo mutations and an orthologue in the zebrafish resulted in HSCR-like aganglionosis for four genes (DENND3, NCLN, NUP98 and TBATA).

To find a disease-mimicking phenotype for four out of 12 candidate genes is a great success. However, questions have been raised about the specificity of morpholinos to study gene function, as up to 70% of morpholino-induced phenotypes cannot be reproduced by genome-editing techniques. Does this mean that the HSCR-like phenotypes we observed are false positives? We think not. Second, non-overlapping morpholinos were used to confirm the results for all genes that showed a phenotype in the initial screen and appropriate 5-nucleotide mismatch controls were used. The morpholino experiments were repeated independently in a second laboratory and phenotypes were reproducible. Moreover, morpholinos were co-injected with a morpholino against p53 to test whether the phenotype was caused by p53-induced apoptosis. Since the results from all control experiments are in line with a specific morpholino-induced phenotype, we confidently conclude that the four identified HSCR genes are indispensable for ENS development. Nevertheless, confirmation of the phenotype by a genetically engineered mutant would further strengthen this conclusion.

Mutations in the four novel HSCR genes were infrequent in our replication cohort and not significantly associated to HSCR. A larger replication cohort is
required to assess to what extent mutations in these genes contribute to HSCR. This question is currently being addressed by the International Hirschsprung Disease Consortium. DNA from approximately 300 HSCR patients and controls will be analyzed for mutations in the four genes and other candidate genes by targeted NGS. This study will be informative for the frequency of mutations in the four novel genes. It is also unclear how mutations in the four genes contribute to aganglionosis in the HSCR patients that carry the mutations. Functional studies on how these genes are involved in the migration, proliferation, survival and differentiation of ENCCs are therefore warranted. Such studies could be performed in vivo in animal models, for example in the zebrafish model that was used to screen the candidate genes, or in transgenic mouse models. Alternatively, in vitro model systems, such as primary ENCCs, iPS cell-derived neural crest cells, neural crest-like cell lines or cultured gut explants can be applied to delineate the role of DENND3, NCLN, NUP98 and TBATA in ENS development.

**How to identify additional HSCR-associated variants?**

Functionally testing candidate genes identified in a genetic study, as we applied in chapter 2, resulted in the identification of four new disease genes that would not have been identified by statistical evidence only and which were not connected to the known signaling pathways in HSCR. This emphasizes the importance of unbiased functional testing. Although there should be no doubt that the plethora of bioinformatic predictions that is available has contributed massively to our knowledge of biology, one should realize that our understanding of cellular processes is limited. Using functional analysis, although costly and time consuming, we were able to unveil new and unexpected genes for HSCR, that would not have been found when solely relying on bioinformatic predictions. One might conclude that we should perform more functional assays to include or exclude identified variants. Clearly this is not realistic for all identified variants. Burden tests might help but as mentioned in small cohorts these tests are not informative. An alternative might be the screening of a selection of genes based on gene function and expression.

**Gene expression analysis in the developing ENS**

Another approach to deal with low statistical power in genetic studies is to select a number of candidate genes, rather than performing a genome-wide study. The low statistical power of genome-wide studies stems from the large multiple testing
correction and this correction is considerably lower in studies on limited numbers of genes. Candidates genes for HSCR can be prioritized based on genetic pilot studies (chapter 5), or on their expression in the developing ENS. Previous gene expression studies on the ENS found that known HSCR genes are expressed by ENCCs and uncovered new genes and pathways in ENS development. Using a transgenic mouse model to specifically isolate ENCCs and a novel analytical approach to identify regulatory genes, we analyzed the gene expression profile of ENCCs in more detail in chapter 3. Gene Ontology analysis showed that ENCCs expressed markers of early and late neuronal differentiation and activation of the RET receptor by its ligand GDNF induced neuronal maturation of ENCCs. Moreover, we applied an Upstream Regulator analysis to identify genes that are predicted to regulate the expression of differentially expressed genes. The known HSCR genes Sox10 and Ret were predicted regulators of gene expression in ENCCs, suggesting the Upstream Regulator analysis is a powerful approach to identify relevant signaling molecules. Other predicted regulators, Bdnf, App, Mapk10 and Timp1, are therefore likely to be important for ENS development as well and may play a role in HSCR. The predicted Upstream Regulators could be linked together in a molecular network underlying ENS development. Likewise, Adora2a and Npy2r (mapping to the HSCR susceptibility locus 4q31.3-q32.3) were near-significant regulators of terminal neuronal differentiation in GDNF-treated ENCCs.

Our study resulted in a list of good candidate genes, however this list is far from complete. In our gene expression study ENCCs were isolated from mouse embryos at developmental stage E14.5 and were cultured in vitro. The single time point of analysis and in vitro culturing are limitations of this study. ENS development in mice takes place between E9 and E15. ENCCs invade the gut, proliferate and migrate along the intestinal wall, and differentiate into neurons and glia. ENCCs make up a heterogeneous cell population at any moment during ENS development, where cells in the proximal regions start differentiating when migratory ENCCs have not yet reached the hindgut. At E14.5, the time point of our expression analysis, ENCCs have reached the hindgut and we found that markers of early and late neuronal differentiation were expressed. At earlier stages of ENS development there will be more migratory and fewer differentiating ENCCs and the corresponding expression profile would likely be different from that at E14.5. A temporal expression analysis between E9 and E15 would give more insight in the changes that ENCC undergo as they colonize the gut. Also the GDNF-responsive genes will likely be different at earlier time points, as GDNF is
important for the proliferation, migration, survival and neuronal differentiation of ENCCs, depending on the developmental stage and position in the gut. Moreover, as HSCR originates from incomplete colonization of the gut by ENCCs, it would be interesting to compare gene expression profiles of ENCCs at the migratory wavefront and non-migrating cells behind the wavefront. Such an experiment would provide information about the genes involved in the migration of ENCCs and the gene expression changes that are associated with a halt in migration.

Moreover, not all genes that are important for ENS development are highly expressed by ENCCs. For example, the known HSCR genes GDNF, NRTN, EDN3, ECE1 and SEMA3A are ligands that are secreted by the gut mesenchyme. Other ligands, such as BMPs, NT-3 and SHH contribute to ENS development as well, but are not highly expressed by ENCCs. So genes important for ENS development may have been missed by exclusively looking at expression in ENCCs. Furthermore, these data suggest that when new receptors are identified, it is worthwhile to also study the function of its ligand(s).

Ideally, isolated ENCCs should not be cultured prior to RNA isolation, as cell culture conditions may affect gene expression. Nevertheless, in our gene expression study we cultured isolated ENCCs to be able to study the effect of GDNF on gene expression. In addition, we needed to expand our ENCC cultures in vitro to obtain sufficient amounts of RNA for microarray analysis. Currently, microarrays are being replaced by RNA sequencing to quantify gene expression and RNA sequencing requires a lot less input material. In vitro culturing of isolated ENCCs can therefore be circumvented in future gene expression experiments. In comparison to microarrays, RNA sequencing is also more sensitive in detecting low-abundance transcripts, can discriminate between different transcripts of a gene, allows for the identification of genetic variants, and can be used to compare the relative expression of different genes within a sample.

Non-coding variants and disease
In addition to selection of candidate genes, gene expression data can be used in disease gene discovery by expression quantitative trait locus (eQTL) mapping. It has been suggested that eQTL mapping has greater power of finding risk SNPs than investigating the association of SNPs with the presence of absence of disease. In HSCR it is assumed that reduced expression levels of RET and other HSCR genes impair colonization of the bowel and common variants in intron 1 of RET have
been shown to affect \textit{RET} expression\textsuperscript{38,39}. A genome-wide survey of common or rare variants affecting the expression of known ENS genes is therefore likely to identify novel genes underlying ENS development and HSCR.

**Low penetrant mutations**

There is a relatively large contribution of common variants to short-segment HSCR. The two published GWAS by the International Hirschsprung Disease Consortium analyzed 371 patients and 856 controls and 629 HSCR trios, respectively\textsuperscript{5,7}. A third GWAS on HSCR included 123 patients and 432 controls\textsuperscript{6}. Meta-analysis of these three GWAS showed that 34\pm11\% of the HSCR risk is attributable to common variation, but only 6\% can be explained by the three associated loci (\textit{RET}, \textit{NRG1} and \textit{SEMA3}) (Tang et al., unpublished data). This suggests that there are undiscovered HSCR loci that may not have reached genome-wide significance. Increasing patient cohorts in GWAS generally leads to more associated loci\textsuperscript{40–42}. Increasing cohort sizes in GWAS on HSCR to several thousand cases and controls will therefore uncover additional associations and lead to the identification of new HSCR genes. However, such an undertaking is expensive and the collection of large numbers of HSCR patients may be troublesome.

It must be noted that although the GWAS on HSCR uncovered new signaling pathways in ENS development\textsuperscript{5–7}, common variants explain only part of the heritability in HSCR. The missing heritability may come from rare variants in both previously associated and as yet undiscovered HSCR genes.

**A molecular link between the CNS and the ENS**

The four newly identified HSCR genes were expressed not only in the gut during zebrafish development, but also in the anterior central nervous system (CNS). This suggests a role for these genes in the development of both the ENS and the CNS. Although the developmental origins of the CNS (neural tube) and ENS (neural crest) are different, there are many similarities between the two nervous systems. Like the CNS, the ENS is composed of efferent, afferent and interneurons that use largely the same repertoire of neurotransmitters\textsuperscript{43,44}. Moreover, the genes involved in late ENS development show a large overlap with those expressed in the developing CNS (\textbf{chapter 3}). Various CNS pathologies, such as autism, stroke, Parkinsonian disease, multiple sclerosis and spinal cord injury can be associated with constipation, suggesting that these conditions affect the proper function of the
ENS\textsuperscript{45,46}. Vice versa, the HSCR genes \textit{KBP}, \textit{SOX10}, \textit{NRG1}, \textit{SEMA3A}, \textit{IKBKAP} and \textit{ZEB2} all contribute to CNS pathologies\textsuperscript{47–53}. Further evidence for the link between the ENS and CNS comes from the HSCR-associated syndromes, most of which show neurological involvement\textsuperscript{2}. The most prevalent HSCR-associated syndrome is Down syndrome (DS, or Trisomy 21) with an incidence of 7.3\% among HSCR patients\textsuperscript{54}. Conversely, DS patients have a 130-fold elevated risk of HSCR compared to the general population\textsuperscript{54}). This suggests that the triplication of one or more genes on chromosome 21 poses a risk for developing HSCR. In \textbf{chapter 6} we explored the genetic link between DS and HSCR by overexpressing selected human chromosome 21 genes into the \textit{Tg(-8.3bphox2b:Kaede)} reporter zebrafish model. We found that injection of \textit{ATP5O} mRNA resulted in hypoganglionosis, predominantly in the distal intestine. \textit{atp5o} was expressed in the zebrafish gut and cerebellum between day 1 and 5 post-fertilization. In post-natal human colon \textit{ATP5O} was expressed in the myenteric and submucosal plexuses, as well as in the epithelium. These data suggest that \textit{ATP5O} is relevant for ENS development and that triplication of \textit{ATP5O} likely contributes to HSCR in DS patients.

\textbf{Implications for genetic counseling}

The revolution in genetic screening, due to the introduction of NGS, has improved genetic counseling enormously. For example in patients with intellectual disability 20\% more causative mutations are being found, which more or less doubles the number of diagnoses that can be made.

For rare diseases, such as HSCR, however, this is slightly more complicated. The study presented in \textbf{chapter 2} nicely illustrates the problems and opportunities one is facing when screening a rare disease. On the one hand NGS improved genetic counseling as some \textit{RET} mutations that were missed in the original screening were detected (not mentioned in the paper). However, what to do with the identified \textit{de novo} mutations? Do we have sufficient evidence that the mutations identified are really disease causing? We did show in a zebrafish model that knocking down the expression of the gene resulted in an HSCR-like phenotype, but is this sufficient in combination with \textit{in silico} prediction and the fact that then mutations were \textit{de novo}? In our case one could argue that the combination of all data should suffice. But what to do when no functional assays are or can be performed? Is \textit{in silico} prediction in combination with a \textit{de novo} appearance enough? Criteria should be made how to deal with these \textit{de novo} findings.
For de novo mutations these criteria might be relatively easy but what to do with a gene in which an inherited mutation is found only once (even truncating)? In particular when one of the healthy parents is also carrying the mutation? Burden tests will likely not give statistical evidence for involvement. Should these all be functionally tested? This is not realistic. One could decide, based on function and mutation type to use the mutation in counseling, but what is known about the penetrance of the mutation? Again criteria need to be set how on what to do with this type of mutations in rare diseases.

Clearly, we need rules on how to work with these mutations as this type of findings will become more and more common practice.

**CONCLUSIONS**

The work in this thesis describes for the first time the application of NGS to study the contribution of rare, coding variants to sporadic HSCR. We show that a rare variant association study has limited power to find genome-wide associations, but does provide information on mutation frequencies per gene and odds ratios. NGS is more powerful when applied to identify de novo mutations. Although the identification of de novo mutations in NGS data is technically challenging, combining de novo mutation screening and functional genomics shows that the identified de novo mutations make a large contribution to sporadic, long-segment HSCR. This finding is of importance for genetic counseling in HSCR and underlines the value of unbiased functional analysis of candidate genes identified in a genetic study.
REFERENCES
