A homozygous missense mutation in *SLCO1C1* may cause a novel progressive encephalopathy syndrome

A student thesis in medicine by Anette Torgersbråten
Acknowledgements
Eirik Frengen\textsuperscript{1,2} and Doriana Misceo\textsuperscript{1,2} have been my main supervisors in the research project and in the process of writing this student thesis. Petter Strømme\textsuperscript{3} is the head clinician within the Progressive Encephalopathies project. He recruited the patient described in this thesis, performed the clinical examinations and administrated the hospital tests. Asbjørn Holmgren\textsuperscript{1,2} performed the whole exome sequencing. In addition I would like to thank Tuva, Christeen, Shalini, Thilini, Marit, Siri, Silje, Ellen and Nadia for their help, support, smiles and supervision during my time at the Department of Medical Genetics.

\textsuperscript{1} Faculty of Medicine University of Oslo, \textsuperscript{2} Department of Medical Genetics, Oslo University Hospital, \textsuperscript{3} Department of Pediatrics, Oslo University Hospital.
Preface
The work for this student thesis was performed in the autumn of 2013 and in the spring of 2014. As a student researcher at the Department of Medical Genetics, Oslo University Hospital, Ullevål, I was involved in an ongoing project focusing on Progressive Encephalopathies (PEs). In the PE project, I analyzed whole exome sequencing data from one family and identified the putative disease causing genetic variant which is the subject of this thesis. This included literature studies. I then performed Sanger sequencing of DNA from the patient, parents and a cohort of controls. I participated in the clinical examination of the patient and in communication with her mother. I also presented whole exome sequencing as a method to diagnose rare neurologic diseases at the Norwegian Research School in Neuroscience (NRSN) National PhD Conference in Neuroscience at Stiklestad in September 2013.
Summary
The subject of this thesis is a 12 year old girl with progressive dementia, ataxia, apraxia and upper motor neuron signs, in which all genetic analyses performed prior to inclusion in the PE project had not provided a molecular diagnosis. Whole exome sequencing of DNA from the patient and her parents identified a possible cause of the disease. The child was homozygous for the single nucleotide variant (SNV) chr12.hg19.g.20870143G>A, which is a missense mutation in the gene SLCO1C1, substituting a highly conserved aspartic acid at amino acid position 252 for an asparagine. The variant was not found in publicly available databases of genetic variants (in September 2013), nor in an in house collection of 181 exomes. The variant was shown by Sanger sequencing to segregate with the disease in the family, and it was not found in 176 controls. SLCO1C1 encodes a thyroid hormone (TH) T4 and T3 transporter, also called Organic Anion Transporter 1C1 (OATP1C1). In vitro assays performed by collaborators showed decreased T4 transport activity in cells ectopically expressing the mutant protein, as compared to cells expressing the wild type transporter (unpublished data). Loss of function of another TH transporter, the monocarboxylate transporter 8 (MCT8, encoded by SLC16A2) causes Allan-Herndon-Dudley syndrome (AHDS, OMIM #300523), a disease presenting with clinical features partially overlapping with the clinical observations in our patient. However, disease causing mutations in SLCO1C1 have not previously been reported. The results presented here suggest that the homozygous SLCO1C1 p.D252N variant causes the disease in the patient of this thesis.
Background
Childhood onset progressive encephalopathies (PE) is a heterogeneous group of diseases with a cumulative incidence of 0.6 per 1000 live births (1, 2). PE usually presents within the first year of life, and the majority of cases are due to inborn errors of metabolism (1). Other subtypes of PE include neurodegenerative and infectious/inflammatory origin. The incidence of each etiological diagnosis can be extremely rare, for example <1:1000.000, but due to a high degree of comorbidity and mortality, the disease burden is significant, both on an individual and on a societal level. Finding the molecular diagnosis is important even if there is no treatment available, as it can improve follow up and give some indication as to the prognosis. Diagnosis enables carrier and prenatal testing and, finally, provides a sense of understanding and emotional solace to the family members (3).

The process of finding a genetic diagnosis has traditionally proceeded in the phenotype to genotype direction. With this approach, the phenotype guides the decisions as to what genetic tests should be performed in order to determine the diagnosis. These tests have traditionally included karyotyping, array Comparative Genomic Hybridization (aCGH), Multiplex ligation-dependent probe amplification (MLPA) and single gene sequencing methods useful for identifying gene dosage imbalances and DNA sequence variations with resolutions ranging from whole chromosomes down to single base pairs.

An emerging trend over the last few years has been the pursuit of genetic diagnosis in the genotype to phenotype direction, using high throughput sequencing (HTS) (4-9). Whole exome sequencing (WES) is a common HTS approach, allowing for sequencing of near all coding regions of a genome at single nucleotide resolution, providing vast amounts of sequence information with the use of considerably less time and resources than the traditional approach of single gene sequencing. In WES, all known exons in a genome are first captured and then sequenced. The rationale is that even though the protein coding regions comprise less than 2% of the genome, it is estimated that ~85% of disease causing variants are located in these regions (10, 11). Exome sequencing is less costly and less time consuming than whole genome sequencing, and the diagnostic yield with WES has shown to be in the range of 25-68% of cases (4-9, 12) as compared to 15-20% by microarrays for neurodevelopmental disorders (13). An important advantage of WES is that it is a so called “hypothesis free” method. This means it is not dependent on a clinician formulating the correct clinical diagnosis, a task which can be very challenging in rare diseases or diseases with atypical clinical presentations, and not feasible in previously undescribed syndromes. Currently WES is not an efficient method to detect copy number variations (CNVs) and other structural variation, but CNVs can be detected by aCGH.

In some instances, mostly in diagnostic settings, the analysis of the WES data can be restricted to the analysis of a list of genes known to cause a certain phenotypic feature or constellation of such features, compatible with the clinical presentation of the patient. Analyzing a finite number of genes makes interpretation of WES data less demanding and also considerably reduces the risk of incidental findings.
In research settings, it is more common to analyse the entire list of variants generated in a WES experiment, without limiting the study to a gene panel. This approach is useful when there is no clear clinical hypothesis. A common approach is the “trio WES”, which entails sequencing of all exons in the patient and in both parents, and then performing bioinformatic analysis to identify possible disease causing variants. Comparison of the proband exome with the parent exomes is necessary in the filtering process. It eases the post sequencing analysis by making it possible to identify and exclude some of the harmless variants that run in the family without causing a disease. If the affected proband and his/her healthy parents are all homozygous for a rare, genetic variant, the variant is unlikely to be disease causing. By the same logic, trio WES makes it possible to identify de novo disease causing genetic variants, based on the rationale that if they are plausible variants present in the sick child and not present in her healthy parents, they are possibly disease causing. In this thesis, trio WES was performed, aiming to identify the disease causing genetic variant in a patient with PE.

**Materials and methods**

**Clinical description**

The patient is a 12 year old Norwegian girl, born to non-consanguineous parents, both originating from a valley in south-east Norway. She is the oldest of three siblings. Her two sisters are healthy.

The patient was born to term after an uneventful pregnancy and developed normally until the age of 10 months, at which point she started to lag behind her peers. The patient’s mother recalls the patient as ”clumsier”, and she was falling over more often than other children her age. At the age of five, she had a 25 word vocabulary and was showing stereotypical behaviour compatible with autism spectrum disorders or possible Rett syndrome. There are no reports of febrile or afebrile seizures. Her mother describes her as intolerant to cold, requiring layers of thick clothing even on warm days.

At the age of 10,5 years her height was 138 cm (< 50th percentile for height-for-age), BMI 16,4 (< 50th percentile) and she had a small head (circumference 51,2 cm) (14). She has smooth nasolabial folds and high arched eyebrows.

On examination at 12 years of age, her clinical picture was characterized by dementia, ataxia and apraxia. She communicated in one-syllable sounds. She ambulated slowly and unsteadily and there were signs of developing limb contractures, probably due to spasticity, which was also evident. At this point bulbar involvement was also suspected, due to profuse drooling.

**DNA**

After obtaining written informed consent, peripheral blood samples were collected from the patient and her parents for DNA preparation from leukocytes. Genomic DNA from 176 controls of matched ethnicity was also available at the department and used in this study.
Reference genome
Positions on the human genome and mapping of sequence data refer to the GRCh37/hg19 produced by the Reference Genome Consortium.

Whole exome sequencing
DNA was sheared using a Covaris sonicator (Covaris, Woburn, MA) to produce fragments with an average size of 100 bp. Paired-end Illumina adapters (Illumina, Inc., San Diego, CA) were ligated to the fragments according to the manufacturer’s recommendations. Exome capture was performed with the SureSelect Human All Exon kit v2 (Agilent Technologies). The final amplified exome captured library was quantified using a Qubit Fluorometer (Life Technologies) and qPCR using primers binding in adapter sequence and Power SYBR Green PCR Master Mix (Life Technologies). Illumina PhiX control kit v2.0 DNA (Illumina, Inc., San Diego, CA) was used for standard curve generation. Fragment size distribution of the input library was measured using a 2100 Bioanalyzer and Agilent High Sensitivity DNA Chip (Agilent Technologies).

The exome captured library was sequenced on an Illumina HiSeq2000 with 100 bp paired-end reads. Reads that did not pass Illumina’s chastity filter were removed prior to alignment. The remaining reads were aligned to the reference human genome (hg19) using Novoalign (novocraft.com/main/index.php). The alignment was refined by the Genome Analysis Toolkit (GATK), and PCR duplicates were marked by Picard (picard.sourceforge.net). 98% of the reads mapped to the reference sequence, yielding an average of 80x coverage per targeted base. Variant calling was performed using GATK. Variants were annotated by ANNOVAR (2012 May 25).

Data analysis
The three variant call files (vcfs) generated for the family trio were simultaneously analyzed using FILTUS (15). Variants were discarded if they were previously detected more than twice in 181 exome sequences available in house and variants with allelic frequency above 0.01 according to 1000 genome (1000genomes.org/), dbSNP142 (ncbi.nlm.nih.gov/SNP), Exome Aggregate Consortium database (exac.broadinstitute.org/) and ExomeVariant Server (evs.gs.washington.edu/EVS). Variants considered benign/tolerable for protein function according to the in silico prediction software SIFT and PolyPhen2 (SIFT cutoff > 0.05, PolyPhen2 cutoff < 0.07) were also discarded. The attention was focused on variants resulting in missense, nonsense, frameshift, and small insertions/deletions. WES data was analyzed with autosomal dominant and autosomal recessive (homozygous and compound heterozygous) modes of inheritance.
Sanger sequencing
The variant found in SLCO1C1 by WES was confirmed with Sanger sequencing after PCR amplification of the selected region. Primers for PCR amplification were designed using Primer3 (16). Forward primer SLCO1C1-F0056: CTTGGTGCGAGGTGTTGTGTG (Chr12:20870056-20870075)
Reverse primer SLCO1C1-R0239: AATTTCCACACCTGGTCCCATTA (Chr12:20870218-20870239).

Sanger sequencing was performed using an ABI 3730xl DNA analyzer and ABI BigDye dye terminator cycle-sequencing kits (Life Technologies). Sequences were aligned using EMBL-EBI Emboss Needle pairwise alignment tool for nucleotides (ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html).

Literature studies
Searches were performed for all of the candidate variants detected after filtering of WES data using PubMed, ClinVar, Online Mendelian Inheritance in Man (OMIM) and the Human Gene Mutation Database (HGMD (17)). Literature searches were performed for all of the listed candidate genes with gene name/s in combination with the search terms variant, mutation, disease, human disease, neurology, neurogenetic, syndrome, phenotype, ataxia, apraxia, dementia and spasticity.

Results

Diagnostic tests
Previously karyotyping and aCGH had been performed with normal results. FMRI1, MECP2, CDKL5, FOXG1 and UBE3A were sequenced and no clinically significant variants were found. An MLPA test was performed for Angelman and Rett syndromes showing normal results. Fibroblast studies revealed no findings compatible with lysosomal storage or mitochondrial diseases.

Cerebral Magnetic Resonance Imaging (MRI) at the ages of 7 and 11 years revealed cortical, subcortical and periventricular atrophy, progressing from discrete to obvious.

A CSF biomarker assay (Table 1) revealed abnormally high levels of neurofilament light protein (NFL) at 11 years, indicating axonal degeneration (18). Glial fibrillar acidic protein (GFAP) was also increased, which has been shown to correlate with dementia (19). Beta-Amyloid and Tau were normal. T4 in CSF was slightly elevated and T3 level was normal compared to 10 randomly selected controls, as national reference values for CSF thyroid hormone levels does not exist. Serum TSH, T3, and T4 were normal (Table 2).
Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Result at 11 years</th>
<th>Reference values</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFAP</td>
<td>280 ng/L</td>
<td>&lt; 175 ng/L</td>
</tr>
<tr>
<td>NFL</td>
<td>2600 ng/L</td>
<td>&lt; 380 ng/L</td>
</tr>
<tr>
<td>Tau</td>
<td>243 ng/L</td>
<td>&lt; 250 ng/L</td>
</tr>
<tr>
<td>T4</td>
<td>11.06 ng/L</td>
<td>Elevated*</td>
</tr>
<tr>
<td>T3</td>
<td>1,10 ng/L</td>
<td>Normal*</td>
</tr>
</tbody>
</table>

Table of results from lumbar puncture. *No reference values available.

Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Result at 12 years</th>
<th>Reference values</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-free thyroxine</td>
<td>15,3 pmol/L</td>
<td>10-18 pmol/L</td>
</tr>
<tr>
<td>S-TSH</td>
<td>1,96 mIU/L</td>
<td>0,5-4,9 mIU/L</td>
</tr>
<tr>
<td>S-free T3</td>
<td>6,2 pmol/L</td>
<td>0,0-10,0 pmol/L</td>
</tr>
<tr>
<td>P-PTH</td>
<td>3,8 pmol/L</td>
<td>9-52 pmol/L</td>
</tr>
<tr>
<td>S-Calcium</td>
<td>2,45 mmol/L</td>
<td>2,15-2,65 mmol/L</td>
</tr>
</tbody>
</table>

Table of results. Blood tests. Biochemical tests performed on peripheral blood samples drawn at 12 years of age.

**Whole exome sequencing results**

After exome sequencing of the family trio, base calling, alignment to the reference genome, variant calling and annotation, information about SNPs, indels and structural variants was compiled in vcf files. There were ~72,000 variants in ~15,500 genes from each of the three exomes. The mean read depth (mean read/base) and the exome coverage (% of bases read >10x) indicated that the WES results were of high quality (Table 3).

Table 3. Statistics from the whole exome sequencing.

<table>
<thead>
<tr>
<th>Exome</th>
<th>Total reads</th>
<th>Mean read/base</th>
<th>% of bases above 10x coverage</th>
<th>Number of variants</th>
<th>Number of genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>5,050,441,182</td>
<td>81,35</td>
<td>92,3</td>
<td>71981</td>
<td>15485</td>
</tr>
<tr>
<td>M</td>
<td>5,764,541,704</td>
<td>92,85</td>
<td>93,2</td>
<td>72075</td>
<td>15623</td>
</tr>
<tr>
<td>P</td>
<td>5,376,785,200</td>
<td>86,6</td>
<td>93,1</td>
<td>72188</td>
<td>15541</td>
</tr>
</tbody>
</table>

**VCF filtering**

**Dominant model**
As the proband was born to non-consanguineous parents and she was the only affected child in the family, the disease could have been due to a *de novo* variant with autosomal dominant inheritance. Therefore search criteria in the WES data of the family trio were set up to include only variants that were *de novo* in the patient. This means that the variant should be present in the patient exome data at a given genomic position and not present in the mother or father exome data at the same genomic position. The majority of these variants do not cause disease. *De novo* variants are usually heterozygous, as the likelihood of the same *de novo* mutation occurring in both alleles of a gene at the same time is small. *De novo* variants are termed dominant when they cause expression of a phenotype, e.g. by causing a disease phenotype.

The filter “/ID does not contain rs/” was used to exclude variants already described in variant databases with a minor allele frequency MAF > 0.01. The filter “/FILTER equal to PASS/” was used to set a quality threshold for the called variants with the aim of reducing the risk of false positives. The filter “/SNPEFF_IMPACT_INFO equal to MODERATE or HIGH/” was used to limit the search to variants predicted by *in silico* analysis to have a moderate or high impact on protein function. During annotation, a score of HIGH is typically given to frame shift variants, splice acceptor and splice donor variants, start lost, and stop gained. A score of MODERATE is typically given to coding sequence variants, in-frame insertions and deletions, missense variants and splice region variants.

The vcf(s) were analysed using these filters, resulting in seven variants in seven genes. They were all SNVs, six of them resulting in missense mutations, one of them resulting in loss of a start codon. The variants are displayed in Table 4 with genomic position and predicted effect on protein function. Where the *in silico* prediction column states “No data”, no output was given in Filtus.

**Table 4.** *De novo* variants in proband, results according to a dominant model of inheritance:

<table>
<thead>
<tr>
<th>Genomic position (GRCh37)</th>
<th>Gene name (HGNC)</th>
<th>Consequence (AA substitution)</th>
<th>Polyphen-2 / SIFT prediction</th>
<th>Disease associated gene (Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:17109490</td>
<td>PLD6</td>
<td>A11T</td>
<td>No data / No data</td>
<td>N</td>
</tr>
<tr>
<td>22:51159094</td>
<td>SHANK3</td>
<td>S931P</td>
<td>Possibly damaging (0.744) / Tolerated (0.3)</td>
<td>Y</td>
</tr>
<tr>
<td>17:42290219</td>
<td>UBTF</td>
<td>E210K</td>
<td>Probably damaging (0.906) / Tolerated (0.14)</td>
<td>N</td>
</tr>
<tr>
<td>19:1230042</td>
<td>C19orf26</td>
<td>E438D</td>
<td>Benign (0.003) / Tolerated (0.54)</td>
<td>N</td>
</tr>
<tr>
<td>9:130517366</td>
<td>SH2D3C</td>
<td>Q25K</td>
<td>Benign (0.009) / Tolerated (0.51)</td>
<td>N</td>
</tr>
<tr>
<td>7:148921022</td>
<td>ZNF282</td>
<td>I433M</td>
<td>Benign (0.009) / Tolerated (0.22)</td>
<td>N</td>
</tr>
<tr>
<td>10:74928099</td>
<td>FAM149B1</td>
<td>M1R (start lost)</td>
<td>Possibly damaging (0.858) / Deleterious (0)</td>
<td>N</td>
</tr>
</tbody>
</table>
Literature searches were performed for all of the above genes. *PLD6* (Phospholipase D family, member 6, OMIM #614960) is a mitochondrial phospholipase enzyme. Male *PLD6-KO* mice are infertile (20). *UBTF* (Upstream binding transcription factor, OMIM #600673) is a transcription factor required for human ribosomal RNA synthesis. *C19orf26* (also known as *CBARP*) is a voltage dependent calcium channel beta subunit associated regulatory protein, negatively regulating calcium channels at the plasma membrane. It was found to be previously associated with Meibomian cysts and Morgagni cataract (source Genecard/Malacard (21), no original source listed). *SH2D3C* (also known as *SHEP1*, OMIM #604722) is a SH2 domain-containing protein, important in signaling networks coordinating marginal zone B-cell development and function in mice (22). *ZNF282* (zinc finger protein 282, also known as *HUB1*, OMIM #603397) is a transcriptional repressor which has been shown *in vitro* to repress human T-cell leukemia virus type 1 by binding to its U5 repressive element (23). The protein product of *Fam149B1* (Family with sequence similarity 149, member B1) is unknown. None of these genes have previously been found to be associated with a human disease relevant for the phenotype of the patient.

*SHANK3* is a scaffolding protein involved in the structural organization of dendritic spines (OMIM 606230). Phelan-McDermid syndrome can be caused by mutations in *SHANK3* (OMIM 606232). This is a syndrome characterized by neonatal hypotonia, global developmental delay, possible growth disturbances, absent or severely delayed speech, autistic behavior and minor dysmorphic features (24). *SHANK3* seemed like a promising candidate, but Sanger sequencing did not confirm the variant, documenting that it was a false positive WES finding.

### Recessive model

As none of the variants identified using the dominant disease model were convincing, a new search was performed, this time using a recessive model of inheritance. Recessive genetic disease results from disease causing variants in both alleles of a gene. These variants can be identical, giving a homozygote genotype, or they can be distinct, giving a compound heterozygote genotype. To search for candidate variants within a recessive model of inheritance, filters were set up to identify only variants that could be found in homozygous or compound heterozygous state in patient exome data, and also in heterozygous state in mother and father exome data.

As in the dominant model, the filter “/ID does not contain rs/” was used to exclude variants already described in variant databases with a minor allele frequency MAF > 0.01, and “/FILTER equal to PASS/” was used to set a quality threshold for the called variants with the aim of reducing the risk of false positives. “/SNPEFF_IMPACT_INFO equal to MODERATE or HIGH/” was used to identify variants predicted to have a moderate or high impact on protein function. Five variants in three genes passed these filters. One of the variants was a codon deletion, one variant was a synonymous SNV, the three remaining variants were missense SNVs (Table 5).
Table 5. List of recessive variants in the proband after filtering vcf files from trio WES according to a recessive model of inheritance:

<table>
<thead>
<tr>
<th>Genomic position (GRCh37)</th>
<th>Gene name (HGNC)</th>
<th>Reference / alternative sequence</th>
<th>Amino acid change (AC)</th>
<th>cDNA position (AC)</th>
<th>PolyPhen / SIFT Prediction</th>
<th>Disease associated Y/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:179436256</td>
<td>TTN</td>
<td>acaata/ata</td>
<td>T/15800/ (codon deletion) (2)</td>
<td>47780-47782 (2)</td>
<td>No data / No data</td>
<td>Y</td>
</tr>
<tr>
<td>2:179575619</td>
<td>TTN</td>
<td>C/T</td>
<td>R8158H (2)</td>
<td>24473 (2)</td>
<td>Unknown / No data</td>
<td></td>
</tr>
<tr>
<td>2:179640319</td>
<td>TTN</td>
<td>T/C</td>
<td>Q2045R (2)</td>
<td>6272 (2)</td>
<td>Unknown / Tolerated (0.21)</td>
<td></td>
</tr>
<tr>
<td>17:17109490</td>
<td>PLD6</td>
<td>C/T</td>
<td>P37P (2)</td>
<td>111 (2)</td>
<td>No data / No data</td>
<td>N</td>
</tr>
<tr>
<td>12:20870143</td>
<td>SLOC1C1</td>
<td>G/A</td>
<td>D252N (4)</td>
<td>754 (4)</td>
<td>Probably damaging (0.998) / Deleterious (0)</td>
<td>N</td>
</tr>
</tbody>
</table>

AC: allele count; theoretical allele frequency corresponding to the genotype call.

TTN encodes the protein titin, or connectin, a muscle protein which is expressed predominantly in cardiac and skeletal muscle. TTN mutations cause dilated cardiomyopathy (OMIM #604145) (25) and tibial muscular dystrophy (OMIM #600334) (26), but no patients with a neurological phenotype have been described with TTN mutations. PLD6 is not previously associated with human disease and the clinical impact of synonymous SNVs in general is uncertain.

In the gene solute carrier organic anion transporter family, member 1C1 (SLOC1C1), encoding the thyroid hormone (TH) transporter organic anion transporter protein 1C1 (OATP1C1), the patient was found to harbor a homozygous SNV, substituting a guanine (G) for an adenine (A) in the position g.12:20870143G>A.c.754G>A.p.D252N (reference sequence NM_001145946.1). WES data from the parents indicated that they were heterozygous for the same variant in the same position. Homozygosity mapping in parents’ exome data performed using FILTUS confirmed homozygosity by descent. The in silico bioinformatic prediction tool PolyPhen-2 predicted the variant in SLOC1C1 to be ”probably damaging” (probability score 0.998).

According to Ensembl Genome Browser data, SLOC1C1 has 10 transcripts, five of which are protein coding, and the c.754G>A variant is present in all coding transcripts.

The identified SNV in SLOC1C1 results in the substitution of a conserved (Figure 1), electrically charged aspartic acid for a polar aspargine. The substitution is located in a transmembrane protein domain (27).
The variant in *SLCO1C1* was considered a plausible candidate disease causing variant and investigated further. This choice was based also on previous descriptions of the X-linked Allan-Herndon-Dudley syndrome (AHDS), presenting with spastic paraplegia and intellectual disability due to central hypothyroidism. AHDS is a syndrome caused by mutation in the gene *SLC16A2*, encoding the central nervous system TH transporter monocarboxylate transporter MCT8 (22). The AHDS phenotype has similarities to the phenotype of the patient in this thesis, and missense SNVs are among the types of mutations associated with AHDS, supporting the hypothesis that the substitution detected in our patient can result in a severe neurological phenotype (28).

**Verification of the SLCO1C1 variant with Sanger sequencing**

Sanger sequencing of the target area in *SLCO1C1* was performed on DNA from the patient and her family using the primers SLCO1C1-F0056 and SLCO1C1-R0239, confirming that the patient was homozygous and the parents heterozygous for the variant in *SLCO1C1* g.12:20870143 (Figure 2).
Figure 2. Sanger sequencing results:

Table: Sanger sequencing results. I-1: mother is heterozygote G/A, I-2 father is heterozygote G/A, II-1 proband is homozygote A/A.

Sequencing of target area in 176 controls of matched ethnicity
DNA from 176 healthy individuals was available at the department (blood was supplied from Blodbanken). The SLCO1C1 target sequence was PCR amplified and Sanger sequenced using the primers SLCO1C1-F0056 and SLCO1C1-R0239. The g.12:20870143G>A variant was not found in the controls. 169/176 controls were homozygous WT (G/G). 7/176 controls were heterozygous for G/T (thymine) or G/C (cytosine).

Discussion
Whole exome sequencing has revolutionized the field of medical genetics, but the method does have some limitations. WES is not efficiently identifying deletions and insertions, as it is not possible to discern these from inadequate or increased coverage. WES cannot efficiently identify epigenetic changes, e.g. abnormal methylation patterns. Variations in noncoding regions and regulatory areas go unnoticed due to selected capture of protein coding regions only. Sequencing of mitochondrial DNA requires a different sequencing technology and DNA from leukocytes is not the best source to identify mutations in mitochondria, which means that mitochondrial DNA variants would be missed with the approach described in this thesis. In highly penetrant Mendelian diseases, the diagnostic yield from WES has still shown to be satisfactory (4, 5). In our patient karyotyping, MLPA and single gene sequencing for selected genes have been performed with normal results.

Downstream of the WES, there are both software and human limitations to data interpretation. A disease causing variant might be located in an area of the genome which was not captured. Disease causing variants can be missed due to imperfect variant calling, or it might have been discarded prior to or during vcf filtering due to poor quality (e.g. poor coverage). One or more disease causing variant/s can be discarded post vcf filtering due to lack of knowledge about
the affected protein product/s. The latter point is perhaps the greatest challenge with the genotype to phenotype process of diagnosis, making it difficult to draw conclusions about novel findings.

The patient in this thesis is a 12 year old girl with a progressive neurological phenotype characterized by dementia, ataxia, apraxia and upper motor neuron signs. With trio WES she was found to be homozygous for the single nucleotide variant g.12:20870143G>A in the gene SLCO1C1, encoding the thyroid hormone transporter OATP1C1. Her mother and father were both found to be heterozygous for the same variant. The in silico prediction software Polyphen-2 and SIFT predict the variant to be damaging to the protein. It seems plausible that the variant impacts on protein function, due to the differing characteristics of aspargine vs aspartic acid. No previous description of the variant was found in literature and databases. The variant was not found when searching an in house database of 181 exomes available at the Department of Medical Genetics or with Sanger sequencing of DNA samples from 176 healthy controls of matched ethnicity.

OATP1C1 is a thyroid hormone transporter with high affinity for T4 and the T3 degradation product rT3 in particular (29). It exhibits bidirectional transport of T4 (30). It also accepts T3 as substrate. In rats and mice, OATP1C1 is highly expressed in brain endothelial cells and in choroid plexus structures (30). In the human hypothalamus, OATP1C1 is expressed in glial cells throughout, and the expression is particularly high in the supraoptic, paraventricular and infundibular nuclei. Regarding OATP1C1 expression in the human choroid plexus, results vary between studies (31, 32).

T4 transport assay performed on H4 glioma cells transiently expressing the mutant OATP1C1 (D252N) transporter demonstrated 30 % reduced T4 transport capacity, as compared to cells expressing wild type OATP1C1 (personal communication: Theo Visser, Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands). This supports pathogenicity of the variant, consistent with the Polyphen-2 and SIFT in silico predictions.

Thyroid hormones (TH) are essential for the normal development and function of the central nervous system. Absence of TH during the pre- and early postnatal period has irreversible detrimental effects on neuroanatomical development like migration and myelination (33, 34).

Thyroid hormones are lipophilic and were traditionally believed to freely diffuse across plasma membranes to reach their nuclear receptor. This view was first challenged by Rao et al in 1976 (35), and in 2004 came the first case report of a disease causing mutation in a gene encoding a thyroid hormone transporter, causing the Allan-Herndon-Dudley syndrome (AHDS) (36). AHDS is an X-linked dominant neurological and endocrinological syndrome, characterized by hypotonia of the newborn, evolving into spastic paraplegia during the first few years of life. Affected children also display severely delayed psychomotor development. Free and total serum T4 is decreased in AHDS patients, T3 is increased and thyroid-
stimulating hormone is normal or slightly elevated (37). Some patients exhibit signs of what could represent a peripheral thyrotoxicosis, like severe muscle hypoplasia and elevated tissue lactate and ammonium (38, 39). AHDS is caused by mutations in SLC16A2, encoding the TH transporter MCT8. In the human hypothalamus, MCT8 is expressed by neurons in the paraventricular, supraoptic and infundibular nuclei (40).

The human MCT8 and OATP1C1 expression patterns are different. But there are clinical similarities between the AHDS phenotype and the phenotype of the patient in this thesis: psychomotor affection has been apparent since an early age, and she is currently displaying spasticity of the lower extremities and severe intellectual disability. Although her serum values of TH are within reference range, she demonstrates elevated levels of CSF-T4, indicating an imbalance in transport and distribution of TH.

In a study by Trajkovic et al (2007), Slc16a2-KO mice replicated the abnormal serum TH levels of AHDS but displayed no overt neurological phenotype (41), leading the researchers to propose that a different TH transporter than MCT8 was more important for TH transport in the murine brain. An essential role was hypothesized for OATP1C1. In following studies with Slco1c1-KO mice, these were indeed found to display findings indicative of central hypothyroidism, with decreased brain levels of T4 and T3 and deiodinase levels in accordance with a hypothyroid state (D2 increased and D3 levels decreased). The mice did not, however, display any overt neurological phenotype (42). It was still hypothesized that the CNS specific hypothyroidism, albeit without an overt effect upon mouse reproduction or behavior, was due to impaired OATP1C1-transport of T4 across the blood-brain-barrier, indicating a central role for this transporter in supplying the murine CNS with thyroid hormone. At the human blood brain barrier, OATP1C1 expression has been found to be low (43). Recent studies performed on mice with double KO of Slc16a2 and Slco1c1 (Slc16a2/Slco1c1 DKO mice) resulted in a striking phenotype of CNS hypothyroidism with markedly reduced uptake of T4 and abnormal neurodevelopment (44), further supporting the physiological importance of these transporters in maintaining a proper TH level in the CNS.

The serum TH levels in the patient in this thesis were within reference range, but there is evidence that the intracellular availability of TH is not solely dependent on serum TH levels (45). Additional regulation of TH levels takes place on an intracellular level, with the deiodinase 2 enzyme being up- and downregulated under conditions of hypo- and hyperthyroidism, respectively. In vitro studies with COS1 cells have shown that OATP1C1 expression is rate limiting in intracellular iodothyronine metabolism by the deiodinases (46). This suggests that even with adequate D2 expression, the cellular availability of active T3 is ultimately dependent on a functional transport protein feeding T4 into the cell. Alkemade et al (2011) has proposed a model in which T4 is taken up by astrocytes via OATP1C1 and converted to T3 by D2 before further transport into neurons (31). This model is based in part on the coexpression of OATP1C1 and D2 in glial cells in the periventricular nuclei (PVN) area of the hypothalamus and the lack of D2 expression in neurons in the same area (40). While the serum concentrations of TH are regulated by the hypothalamic–pituitary–thyroid (HPT) axis, the mechanisms and findings above provide an additional possibility for TH
regulation at a cellular and therefore tissue specific level, a mechanism which is seemingly more dependent on the expression levels of transporters MCT8 and OATP1C1 and the deiodinase enzymes than on the serum concentration of TH (30). This evidence supports the hypothesis that serum TH levels within reference range does not exclude the possibility of a tissue specific TH imbalance.

The thesis patient displays progressive cortical, subcortical and periventricular atrophy on serial MRIs, and she has increased levels of CSF-NFL and -GFAP, biomarkers indicative of axonal degeneration and dementia, respectively. Cerebral and cerebellar atrophy has previously been associated with congenital and central hypothyroidism (36, 47). The mechanism behind this neurodegeneration is not elucidated. The pathophysiology is complex, as TH positively and negatively regulates the expression levels of several genes with differing sensitivity towards TH. One TH-regulated gene, \textit{SLC16A1}, is expressed in the spinal cord, cerebral cortex and the corpus callosum. It encodes the lactate/pyruvate transporter MCT1, abundant in the CNS. MCT1 deficiency produces axon damage and neuron loss in cell culture and animal models (48).

OATP1C1 mediates bidirectional transport of T4 in transfected HEK293 cells, perhaps an indication of a dual role of this transporter in TH homeostasis (30). OATP1C1 also mediates efflux of rT3 (29). Pizzagalli et al (2002) proposes that this ability of OATP1C1 to mediate efflux of the T3 degradation product rT3 might provide a mechanism for protecting tissue from TH excess.

Results support that the phenotype of progressive encephalopathy in the thesis patient could be caused by an intraneuronal imbalance of TH levels due to decreased transport capacity by the TH transporter OATP1C1. With the information currently available on the expression and function of the OATP1C1 transport protein in the human CNS, it is not possible to conclude if the high levels of CSF-T4 in the thesis patient is a sign of central thyrotoxicosis or central hypothyroidism.

Further research is needed on the described variant in \textit{SLCO1C1} in order to establish causality regarding the phenotype of the thesis patient. Further functional studies should include transgenic D252N mice or another animal model for \textit{in vivo} evaluation of the variant. Discovering more patients harboring \textit{SLCO1C1} mutations and displaying phenotypes similar to the thesis patient would further strengthen the hypothesis that \textit{SLCO1C1} dysfunction causes disease in humans and that the related phenotype is a severe progressive encephalopathy.
References


