Mutations in \textit{HECW2} are associated with intellectual disability and epilepsy

Jonatan Halvardson,\textsuperscript{1} Jin J Zhao,\textsuperscript{1} Ammar Zaghool,\textsuperscript{1} Christian Wentzel,\textsuperscript{1} Patrik Georgii-Hemming,\textsuperscript{1,2} Else Månssson,\textsuperscript{3} Helena Ederth Sävmarker,\textsuperscript{4} Göran Brandberg,\textsuperscript{2} Cecilia Soussi Zander,\textsuperscript{1} Ann-Charlotte Thuresson,\textsuperscript{1} Lars Feuk\textsuperscript{1}

\textbf{ABSTRACT}

\textbf{Background} De novo mutations are a frequent cause of disorders related to brain development. We report the results of screening patients diagnosed with both epilepsy and intellectual disability (ID) using exome sequencing to identify known and new causative de novo mutations relevant to these conditions.

\textbf{Methods} Exome sequencing was performed on 39 patient–parent trios to identify de novo mutations. Clinical significance of de novo mutations in genes was determined using the American College of Medical Genetics and Genomics standard guidelines for interpretation of coding variants. Variants in genes of unknown clinical significance were further analysed in the context of previous trio sequencing efforts in neurodevelopmental disorders.

\textbf{Results} In 39 patient–parent trios we identified 29 de novo mutations in coding sequence. Analysis of de novo and inherited variants yielded a molecular diagnosis in 11 families (28.2%). In combination with previously published exome sequencing results in neurodevelopmental disorders, our analysis implicates \textit{HECW2} as a novel candidate gene in ID and epilepsy.

\textbf{Conclusions} Our results support the use of exome sequencing as a diagnostic approach for ID and epilepsy, and confirm previous results regarding the importance of de novo mutations in this patient group. The results also highlight the utility of network analysis and comparison to previous large-scale studies as strategies to prioritise candidate genes for further studies. This study adds knowledge to the increasingly growing list of causative and candidate genes in ID and epilepsy and highlights \textit{HECW2} as a new candidate gene for neurodevelopmental disorders.

\textbf{INTRODUCTION}

Intellectual disability (ID) has a prevalence of 1%–3% and is defined by an IQ <70 with an onset before the age of 18.\textsuperscript{3} It has been estimated that 20%–30% of patients with ID also have epilepsy, pointing to a drastic over-representation of epilepsy in patients with ID compared with the general population (prevalence of 0.5%–1%).\textsuperscript{2} The prevalence of epilepsy is even higher with increased severity of ID, and epilepsy co-occurring with ID is also more commonly treatment resistant and displays a higher mortality rate than epilepsy in the general population.\textsuperscript{3,4}

In both ID and epilepsy it is well established that a large fraction of cases have a genetic cause, and there are numerous genetic syndromes where ID and epilepsy are part of the phenotype. These facts together indicate that there is a strong genetic correlation between ID and epilepsy and gives incentive to further identify and investigate genes with causative mutations in patients with both conditions.

Investigations into the genetic aetiology of ID and epilepsy have primarily been performed using chromosomal microarray analysis (CMA) as a first genetic test, resulting in clinically significant findings in 15%–20% of patients.\textsuperscript{5} With the rapid development of high-throughput sequencing technologies, exome sequencing of trios to identify de novo mutations (DNMs) has been introduced in genetic diagnostics, typically resulting in clinically significant findings in 20%–30% of patients already screened by CMA.\textsuperscript{6–9} The limitations in clinical yield include a lack of molecular understanding, resulting in many variants labelled as being of uncertain clinical significance. In addition, genetic interaction effects and environmental causes may play roles in a significant fraction of patients. To address the problem of lacking molecular understanding, major exome sequencing projects, such as the Deciphering Developmental Disorders (DDD), have screened large cohorts in order to identify novel disease-causing genes.\textsuperscript{10} By screening 1133 cases the DDD project identified 12 novel disease genes, increasing the proportion of cases with a molecular diagnosis by 10%.\textsuperscript{11} Exome sequencing in epileptic encephalopathies has also implicated DNMs as a major cause and has led to identification of several new candidate genes.\textsuperscript{7} This underlines the importance of adding to the growing list of causative genes. For frequently co-occurring conditions such as ID and epilepsy, an expanded list of causative genes may also greatly help to understand the pathophysiology and genetic aetiology as well as the connection between these conditions.

In this study we used exome sequencing in 39 patient–parent trios, where the patients have ID in combination with epilepsy. We report the identification of 29 DNMs and one pathogenic inherited single nucleotide variant (SNV) in coding sequence of 23 trios, of which 16 were found in genes previously known to cause epilepsy and/or ID. For 11 families we identified variants determined to be pathogenic, giving a clinical yield of 28.2% in this cohort. Our results also lend further support to previously identified candidate genes in ID and epilepsy, and highlight \textit{HECW2} as a novel candidate gene.
gene in neurodevelopmental disorders based on network analysis and a combined analysis with previous exome sequencing efforts.

METHODS

Study design and patients

The participating patients and parents were recruited between 2012 and 2015 in collaboration with the Genetic Diagnostics Unit at Uppsala University Hospital. Ethical approval for exome sequencing was received from the Uppsala Ethical Review Board and informed consent was received from the parents of all patients. The selection criteria for patients included ID and epilepsy, while parents had to be healthy with no family history of neurodevelopmental disorders. All patients had previously been screened with CMA (250K Nsp Array, Genome-Wide SNP Array V6.0 or CytoScan HD (Affymetrix, Santa Clara, California, USA)) and no pathogenic CNVs had been detected. Genomic DNA was extracted from peripheral blood leucocytes according to standard procedures.

Sequencing

Exome enrichment was performed using SureSelect (Agilent) versions 2–5 and samples were sequenced on either SOLiD, Illumina or IonProton platforms. The sequencing was performed to achieve at least 30× coverage of the captured regions. Mapping of SOLiD reads was performed using Biosoce (Life Technologies) until the release of LifeScope (Life Technologies), which was then used. Illumina reads were mapped using Burrows-Wheeler Aligner (BWA) and IonProton reads were mapped using the Torrent suit software (Life Technologies). All reads were mapped to the Hg19 version of the human reference genome. Programs used for mapping were run using default settings.

Data analysis

After alignment of SOLiD and Illumina reads, variants were called using the Genome Analysis Toolkit (GATK) HaplotypCaller and the standard GATK workflow (Broad Institute). For IonProton, variants were called using the Torrent suit software (LifeTechnologies) and standard settings. To identify DNMs all called SNVs were filtered against our in-house database containing previously identified variants from 170 exomes and the Database of SNP (dbSNP) V.42 (non-database containing previously identiﬁed variants).12 To compile the list of genes previously associated to both ID and epilepsy, all genes that were categorized as conﬁrmed ID genes by the DDD project were collected. After this the Human Phenotype Ontology (HPO) terms associated with the ﬁndings in each of these genes were ﬁltered so that only genes with at least one HPO term associated with epileptic seizures was included (terms included were: HP:0002184, HP:0010818, HP:0011171, HP:0002384, HP:0002373, HP:0007294, HP:0006902, HP:0006869, HP:0007075, HP:0007284, HP:0007202, HP:0002123, HP:0002306, HP:0002182, HP:0002348, HP:0001275, HP:0002466, HP:0002125, HP:0002417, HP:0010520, HP:0006997, HP:0002391, HP:0002437, HP:0002434, HP:0001303, HP:0002479, HP:0002432, HP:0002279, HP:0006907, HP:0002306, HP:0002184, HP:0010818, HP:0002437, HP:0002434, HP:0001303, HP:0002479, HP:0002432, HP:0002279, HP:0002430, HP:0002431, HP:0002794, HP:0006869). To measure the deleteriousness of the identiﬁed DNMs we calculated combined annotation-dependent depletion (CADD) scores for all mutations.25 CADD scores are to be interpreted as a relative measurement of pathogenicity of genetic variants, and higher CADD scores indicate higher pathogenicity. The CADD scores showed a distribution where the synonymous mutations all showed a low score (<10), while stopgains mutations and nonsynonymous variants (with one exception) had scores >10 (see online supplementary table S2). Approximately half of the identiﬁed DNMs had a CADD score higher than 20. It has previously been shown that disease-causing variants in the OMIM database are enriched for CADD scores higher than 20.

Validation and comparison to previous studies

DNMs were validated by Sanger sequencing using standard protocols. Each validated variant was interpreted using the American College of Medical Genetics and Genomics (ACMG) guidelines.15 For genes where a DN M was validated, the number of DNMs identiﬁed in cases (ID, epilepsy and autism) and controls in a selected set of previous exome sequencing studies were counted.6–16–23

Network generation

Network generation was performed using GeneMania adding all genes where DNMs were found in the present study together with a compiled list of genes reported to be associated to both ID and epilepsy.24 To compile the list of genes previously associated to both ID and epilepsy, all genes that were categorized as conﬁrmed ID genes by the DDD project were included. After this the Human Phenotype Ontology (HPO) terms associated with theﬁndings in each of these genes were ﬁltered so that only genes with at least one HPO term associated with epileptic seizures was included (terms included were: HP:0002184, HP:0010818, HP:0011171, HP:0002384, HP:0002373, HP:0007294, HP:0006902, HP:0006869, HP:0007075, HP:0007284, HP:0007202, HP:0002123, HP:0002306, HP:0002182, HP:0002348, HP:0001275, HP:0002466, HP:0002125, HP:0002417, HP:0010520, HP:0006997, HP:0002391, HP:0002437, HP:0002434, HP:0001303, HP:0002479, HP:0002432, HP:0002279, HP:0006907, HP:0002306, HP:0002184, HP:0010818, HP:0002437, HP:0002434, HP:0001303, HP:0002479, HP:0002432, HP:0002279, HP:0002430, HP:0002431, HP:0002794, HP:0006869). The network was then compiled using only protein–protein and pathway interactions, and only genes with at least one connection to any other gene was included in the ﬁnal network.

RESULTS

Trio sequencing

By exome sequencing of 39 trio families we identiﬁed a total of 29 DNMs within protein-coding regions. All variants were validated using Sanger sequencing. The number of DNMs ranged from 0 to 3 per trio and DNMs were identiﬁed in 22 of the 39 families. Among the coding DNMs identiﬁed four were stopgain mutations, 20 were non-synonymous and ﬁve were synonymous mutations. Out of the 29 genes with DNMs, 13 have previously been associated both with ID and epilepsy, one associated with ID only and one only with epilepsy (table 1). Three of the DNMs were identiﬁed in genes previously known to cause recessive or x-linked forms of ID (AARS, MED12 and CERS1). In none of these cases could a second mutation or CNV be identiﬁed despite careful review of alignments and array probe intensities across the genes. Using the ACMG guidelines for classiﬁcation of variants, we identiﬁed pathogenic and likely pathogenic DNMs in known causative genes in 10 families.13 This results in a diagnostic yield of 25.6% based on DNMs. The full list of patient phenotypes and the mutations identiﬁed in each patient are described in online supplementary table S1. Each parent–offspring trio was also investigated for homozygous and compound heterozygous SNVs of clinical relevance in order to identify recessive candidate genes. This analysis led to the identiﬁcation of one additional gene (ADSL) determined to be causative. Including both DNMs and inherited variants, we thus identify pathogenic variants in 11 of 39 families (28.2%).

To measure the deleteriousness of the identiﬁed DNMs we calculated combined annotation-dependent depletion (CADD) scores for all mutations.25 CADD scores are to be interpreted as a relative measurement of pathogenicity of genetic variants, and higher CADD scores indicate higher pathogenicity. The CADD scores showed a distribution where the synonymous mutations all showed a low score (<10), while stopgains mutations and nonsynonymous variants (with one exception) had scores >10 (see online supplementary table S2). Approximately half of the identiﬁed DNMs had a CADD score higher than 20. It has previously been shown that disease-causing variants in the OMIM database are enriched for CADD scores higher than 20.

The participating patients and parents were recruited between 2012 and 2015 in collaboration with the Genetic Diagnostics Unit at Uppsala University Hospital. Ethical approval for exome sequencing was received from the Uppsala Ethical Review Board and informed consent was received from the parents of all patients. The selection criteria for patients included ID and epilepsy, while parents had to be healthy with no family history of neurodevelopmental disorders. All patients had previously been screened with CMA (250K Nsp Array, Genome-Wide SNP Array V6.0 or CytoScan HD (Affymetrix, Santa Clara, California, USA)) and no pathogenic CNVs had been detected. Genomic DNA was extracted from peripheral blood leucocytes according to standard procedures.
To investigate the magnitude of the impact of the DNMs, the CADD scores of the DNMs were compared with the CADD scores of 1000 randomly chosen SNVs from the ExAC database, indicating a clear shift towards more deleterious variants identified in this study compared with the ExAC database. Of the 29 genes with DNMs in this study, 15 genes were found to have DNMs reported in patients in previous studies. In total, these 15 genes contained 63 previously described disease-associated genes reported in neurodevelopmental disorders. To further context, the mutations the level of conservation was assessed using genomic evolutionary rate profiling (GERP) scores for the non-synonymous and stopgain mutations. These scores measure the level of constraint of each base. The results showed that 62% of the mutations could be considered to be in positions that are subjected to evolutionary constraint (GERP >3) (see online supplementary table S2). To complement the CADD scores described above, other commonly used prioritisation scores (SIFT, PolyPhen2 and MutationTaster) are also listed (see online supplementary table S2).

Due to the genetic heterogeneity of neurodevelopmental disorders, DNMs in causative genes are expected to be individually rare, and gathering data from several studies may therefore be one way to find further support for the involvement of candidate genes. To evaluate the potential pathogenicity of the DNMs identified in our families in the context of previous exome sequencing studies in neurodevelopmental disorders we collated data from 13 studies in ID, epilepsy, autism spectrum disorder and control trios (see Methods). The patient categories were chosen as considerable overlap has been shown in the genes implicated in these disorders, and based on the fact that there is commonly overlap in phenotype between these patient groups. In total, this amounted to 5338 patient trios and 2181 control trios. Of the 29 genes with DNMs in this study, 15 genes were found to have DNMs reported in patients in previous studies. In total, these 15 genes contained 63 previously described disease-associated genes reported in neurodevelopmental disorders.

**Table 1** A list of disease-associated genes with DNMs identified in this study.

| Gene | Position | Family | Mutation type | ID | Epilepsy | Associated disorder (OMIM designation and number) | Inheritance |
|------|----------|--------|--------------|---|----------|-----------------------------------------------|-------------|
| CDKL5 | chr1:18598085:C/T | Fam2 | Stopgain | Yes | Yes | Epileptic encephalopathy, early infantile, 2, MIM:300672 | XD |
| KCNQ2 | chr20:62044879:C/A | Fam3 | Non-synonymous | Yes | Yes | Epileptic encephalopathy, early infantile, 7, MIM:613720, Seizures, benign neonatal, 1, MIM:121200 | AD |
| SYNGAP1 | chr6:33400477:C/T | Fam4 | Stopgain | Yes | Yes | Mental retardation, autosomal dominant 5, MIM:612621 | AD |
| SETD5 | chr3:9490126:G/T | Fam5 | Stopgain | Yes | Yes | Mental retardation, autosomal dominant 23, MIM:615761 | AD |
| SMC1A | chrX:162428:G/A | Fam6 | Stopgain | Yes | Yes | Cornelia de Lange syndrome 2, MIM:300590 | XD |
| ZMYND11 | chr10:298399:C/T | Fam7 | Non-synonymous | Yes | No | Mental retardation, autosomal dominant 30, MIM:616083 | AD |
| EFTUD2 | chr17:42931953:T/G | Fam7 | Non-synonymous | Yes | Yes | Mandibulo-facial dysostosis, Guion-Almeida type, MIM:610536 | AD |
| AAAS | chr12:53702981:C/T | Fam8 | Non-synonymous | Yes | Yes | Achalasia-addisonianism-aklakmia syndrome, MIM:231550 | AR |
| GABRG2 | chr5:161576159:G/A | Fam8 | Non-synonymous | No | Yes | Epilepsy, generalised, with febrile seizures plus, type 3, MIM:611277, Epilepsy, childhood absence, susceptibility to, 2, MIM:607681 | AD |
| GRIN1 | chr9:140053150::AC | Fam9 | Non-synonymous | Yes | Yes | Mental retardation, autosomal dominant 8, MIM:614254 | AD |
| SCN2A | chr2:166166923:C/T | Fam10 | Non-synonymous | Yes | Yes | Epileptic encephalopathy, early infantile, 11, MIM:613721, Seizures, benign familiar infantile, 3, MIM:607745 | AD |
| STS | chr1:8752629:G/C | Fam11 | Non-synonymous | Yes | Yes | Mental retardation, MIM:140750 | AD |
| KCNA1 | chr12:5021751:C/T | Fam12 | Non-synonymous | Yes | Yes | Episodic ataxia/myokymia syndrome, MIM:160120 | AD |
| CERS1 | chr19:18990105:G/T | Fam13 | Non-synonymous | Yes | Yes | Epilepsy, progressive myoclonic, 8, MIM:616230 | AR |
| MED12 | chrX:18598085:C/T | Fam2 | Stopgain | Yes | Yes | Epileptic encephalopathy, early infantile, 2, MIM:300895, Opitz-Kaveggia syndrome, MIM:305450 | AD |
| MIM:300895, Opitz-Kaveggia syndrome, MIM:305450 | AD |
| MED12 | chrX:18598085:C/T | Fam2 | Stopgain | Yes | Yes | Epileptic encephalopathy, early infantile, 2, MIM:300895, Opitz-Kaveggia syndrome, MIM:305450 | AD |
| MED12 | chrX:18598085:C/T | Fam2 | Stopgain | Yes | Yes | Epileptic encephalopathy, early infantile, 2, MIM:300895, Opitz-Kaveggia syndrome, MIM:305450 | AD |
| MED12 | chrX:18598085:C/T | Fam2 | Stopgain | Yes | Yes | Epileptic encephalopathy, early infantile, 2, MIM:300895, Opitz-Kaveggia syndrome, MIM:305450 | AD |

For each gene it is noted if it has been associated with ID, epilepsy or both, as well as OMIM IDs for each specific disease it has been associated with. AD, autosomal dominant; AR, autosomal recessive; ID, intellectual disability; XD, X-linked dominant; XR, X-linked recessive.

**Figure 1** A histogram showing the distribution of combined annotation-dependent depletion (CADD) scores from the Exome Aggregation Consortium (ExAC) project, with the CADD score of de novo mutations found in this study shown as coloured circles at the top of the figure.
De novo mutations in genes previously associated with ID or epilepsy

Among the 15 genes with DNMs identified in this study and reported to be involved in ID and/or epilepsy, the mutations detected in SETD5, CDKL5, SYNGAP1 and SMC1A were stop-gains, and the remaining genes carried non-synonymous mutations. Five of the nonsense and non-synonymous mutations identified had previously been reported as causative in dbSNP (ZMYND11, SCN2A, SETD5, GABRG2, CDKL5). In addition, the mutation in KCNQ2 was found in the same position as a previously reported causative mutation, but with a different base change leading to another amino acid substitution. The identification of DNMs already present in public databases are in line with recently published data from a deep sequencing of 10 trios where 3.5% of the identified DNMs were already present in dbSNP.

The symptoms reported for the patients carrying mutations in SYNGAP1, EFTUD2, KCNQ2, GRIN1, SMC1A and ADSL in this study all mirrored the phenotypes previously reported for patients with mutations in these genes. The patient carrying a mutation in EFTUD2 also had a second

Although detailed phenotype descriptions are not available for most of these patients, we note that one of the autism cases also had a low IQ (<65), while the second patient had febrile seizures reported. The HECW2 gene is a HECT-type ubiquitin ligase and is known to regulate the stability of p73. The DNMs in our study was identified in the Homologous to the E6-AP Carboxyl Terminus (HECT) domain of the protein. Out of the five previously reported mutations, four were located in exons associated with the HECT domain, which displays a lower number of non-synonymous mutations in the general population (figure 2).

Network categorisation

To further categorise the genes with DNMs in this study in the context of previously established causative genes, we extracted all genes with causative DNMs in patients with ID and epilepsy from the DDD project data. The genes extracted from DDD were then used together with the genes identified in this study to construct a network based on protein–protein interactions and known pathways (figure 3). The resulting network shows that 48% of the genes with DNMs in this study interact with at least one other gene in the network, while the remaining DNMs showed no connections. Of the genes with interactions, PTCHD2, TMOD2, BAZ1A, PAN2 and HECW2 have not previously been shown to be associated with ID and/or epilepsy. The variants detected in PTCHD2 and TMOD2 were silent mutations and therefore not considered candidate causative mutations in this study. The BAZ1A gene codes for a chromatin remodelling factor, providing another potential candidate gene to the list of known causative chromatin remodelling genes in ID and epilepsy. The PAN2 protein interacts with several DNA-binding proteins, and is a subunit of the PAN with the function to shorten the poly(A)-tails of RNA. Mice homozygous for pan2 mutations exhibit embryonic lethality, while seizures have been reported in mice carrying a heterozygous deletion of pan2. In our network analysis PAN2 was connected to eight other genes, making it one of the most highly interconnected genes among the genes found in this study. The network analysis also points to a central role for HECW2, showing interaction with nine other known causative genes, further strengthening the candidacy of HECW2 as a new causative gene in neurodevelopmental disorders.

Table 2 Showing the number of DNMs identified in cases and controls in previous exome sequencing studies, as well as the OMIM designation number, for genes with DNMs found in this study

| Gene   | Cases | Controls | ID | Epilepsy | Associated disorder (OMIM designation and number) |
|--------|-------|----------|----|----------|---------------------------------------------------|
| SCN2A  | 20    | 0        | Yes| Yes      | Epileptic encephalopathy, MIM:613721, seizures, MIM:607745 |
| SYNGAP1| 12(1) | 1(1)     | Yes| Yes      | Mental retardation, MIM:612621                      |
| SETD5  | 6     | 1        | Yes| No       | Mental retardation, MIM:615761                      |
| HECW2  | 5     | 0        | No | No       | None                                               |
| CDKL5  | 3     | 0        | Yes| Yes      | Epileptic encephalopathy, MIM:300672               |
| KCNQ2  | 3(1)  | 0        | Yes| Yes      | Epileptic encephalopathy, MIM:613720, Myokymia, MIM:121200 |
| ZMYND11| 2     | 0        | Yes| No       | Mental retardation, MIM:160683                      |
| KIAA1244| 1    | 1        | No | No       | None                                               |
| TBC1D4 | 1(1)  | 0        | No | No       | None                                               |
| KCNA1  | 1     | 0        | Yes| Yes      | Episodic ataxia/myokymia syndrome, MIM:160120      |
| BAZ1A  | 1     | 1(1)     | No | No       | None                                               |
| ERC2   | 1(1)  | 1        | No | No       | None                                               |
| GABRG2 | 1     | 0        | Yes| Yes      | Epilepsy, MIM:611277, MIM:607681                    |
| GRIN1  | 1     | 1        | Yes| Yes      | Mental retardation, MIM:614254                     |
| SMC1A  | 1     | 0        | Yes| Yes      | Cornelia de Lange syndrome, MIM:300590             |

The number of synonymous DNMs for each gene and category is noted in parenthesis. One DNM in each of these genes was found in this study, including four stop-gains (SYNGAP1, SETD5, CDKL5, SMC1A), ten non-synonymous (SCN2A, HECW2, KCNQ2, ZMYND11, TBC1D4, KCNA1, BAZ1A, ERC2, GABRG2, GRIN1) and one synonymous (KIAA1244). Variants identified in the present study are not included in this table (listed in table 1).

ID, intellectual disability; DNM, de novo mutation.
DNM in ZMYND11. The two mutations were determined to be likely pathogenic (EFTUD2) and pathogenic (ZMYND11).

The ST5 non-synonymous mutation occurred in a region outside any known motifs or domains present in the ST5 protein. The mutation was found in a patient presenting with ID, seizures, delayed speech, slight dysmorphic features, frequent infections and a benign teratoma. The function of the ST5 protein is relatively unknown; however, studies show that the protein can function as a tumour suppressor in cultured cells. To our knowledge, only a single patient has previously been reported to carry a translocation interrupting the ST5 gene. This patient presented with a similar phenotype, including ID, epilepsy, recurrent infections and a partially overlapping facial gestalt, although more severely affected. Altogether this adds convincing evidence for ST5 being causative in our patient, strengthening the evidence of ST5 as causative in ID and epilepsy. An interesting feature, however, is the benign teratoma in infancy present in our patient, as ST5 has been described as a tumour suppressor.

The de novo stopgain mutation in SETD5 was discovered in a patient with ID and myoclonic seizures. Recent calculations show that loss-of-function mutations in SETD5 might explain up to 0.7% of ID cases identified. Seizures have been reported in a subset of patients. It is therefore interesting to note that the patient with an SETD5 mutation also carries a DNM in ERC2. The ERC2 gene encodes a protein with a central role in the presynaptic active zone. In mice, conditional knockout of ERC2 has been shown to lead to a large increase in inhibitory synaptic strength by increasing the size of releasable vesicles at inhibitory neurons. It is therefore possible that the mutation in ERC2, potentially in concert with the mutation in SETD5, further exacerbates the myoclonus phenotype in this patient.

**DISCUSSION**

Our exome study identified clinically significant DNMs in 10 of 39 patients with ID and epilepsy. In one case we found a recessive cause for the patient phenotype. The diagnostic yield (28.2%) is similar to previous exome sequencing studies in ID, reporting a diagnostic yield ranging from 16% to 29% with the majority explained by DNMs. Of the genes with DNMs that we identified, approximately half have previously been associated with both ID and epilepsy, indicating that the patients selected for this study represent a genetically well-defined group. In one trio (2.5% of patients) we identified a pathogenic recessive mutation. The number of recessive mutations identified in previous exome sequencing projects differ significantly and range from 0 in one study investigating 245 families to 20% in a recent study investigating 45 patients. In several of the known causative genes the specific mutations we identify have not been previously reported, adding to the catalogues of clinically relevant mutations in these genes. The identification of mutations in previously reported genes also adds further evidence to their causative nature and contributes to the description of the clinical spectrum of mutation carriers.

Our analysis shows that many of the genes with DNMs are interconnected through protein–protein interactions or exist in the same pathway together with genes previously linked to ID and epilepsy. As networks are constructed using proof and knowledge from previous studies, it is interesting to notice that several of the genes independently linked to ID and epilepsy are also interconnected in the networks generated. This indicates that the knowledge of gene interactions accumulated to date are sufficient to identify pertinent connections between the genes identified in studies where the patients are selected by a well-defined and delimited set of symptoms. From the network analysis it is interesting to note that genes with DNMs, not previously linked to ID or epilepsy, are connected to several other genes in the network. Even though interaction cannot be considered a proof of clinical significance this makes them interesting candidates for further study and shows the strength of network analysis as a tool for prioritisation of candidate genes.

Two genes, PAN2 and HECW2, stand out in the network analysis by showing connections to multiple other known causative ID and epilepsy. Of these, HECW2 is the most interesting as five DNMs in HECW2 have been identified previous exome sequencing projects in neurodevelopmental disorders, including ID and epilepsy. We show that this represents a significantly higher number of DNMs than would be expected in the number of trios included in the survey. Using residual variation intolerance scores we further notice that the HECW2 gene is among the 0.98% genes most intolerant to functional variation in the human genome, which is also evident from the plot of...
synonymous and non-synonymous variants reported in the ExAC database. Looking into distribution of variation across the exons of the gene we see that most DNMs reported cluster in the exons that are depleted in coding variation, with five of six reported DNMs located within (n=3) or immediately adjacent to (n=2) the HECT domain of the HECW2 protein.

Taken together, these lines of evidence indicate that DNMs in HECW2 are associated with neurodevelopmental phenotypes. Interestingly, a search of social media led to the discovery of additional patients with de novo HECW2 mutations, displaying overlapping phenotypes. Knockout mice for this gene do not show a similar phenotype, with partial preweaning lethality, lean body mass and lowered mean platelet volume as major symptoms. In light of the mouse knockout phenotype, it is important to note that all DNMs reported are non-synonymous, potentially pointing to a gain of function or dominant negative role. Using BrainSpan data, we find that HECW2 is expressed at moderate levels in the brain throughout development, with the highest expression in frontal cortex. Interestingly, the gene that shows the highest correlation in expression in frontal cortex during brain development is CDKL5, a well-established causative gene in ID and epilepsy. The expression pattern therefore lends further support to HECW2 as a highly interesting novel candidate gene. Our results show that the systematic use of interaction data can be used as an effective tool for candidate prioritisation. However, results must also be interpreted with caution, as the effectiveness of this strategy will depend on the criteria used in the patient selection and to what extent the gene in question has been studied previously.

The identification of mutations in known causative genes provides the opportunity to refine and expand on previous reports on associated clinical symptoms. In our data, we identify several patients that provide potential new insight into the genotype-phenotype correlation. For example, the patient carrying a GABRG2 mutation found in this study had several phenotypes not present in previously confirmed carriers. At the same time, these results must be interpreted with caution, as it is possible that the additional symptoms are the result of a second mutation. For example, in the patient with the EFTUD2 mutation a second pathogenic mutation was found in the ZMYND11 gene, making it probable that both genes contribute to phenotype. The finding of a second causative mutation is in accordance to a recent study where it was calculated that about 1.4% of patients had a second mutation contributing to the phenotype. The fact that second mutations may have an impact on the resulting phenotype is further highlighted by the stopgain SETD5 mutation. In this patient a second mutation was found in ERC2, a gene known to be involved in the presynaptic active zone where it has an effect on inhibitory synaptic strength. Previous studies show that even modest changes in synaptic plasticity at inhibitory neurons may trigger epileptic activity. This raises the possibility that ERC2 contributes to the epileptic phenotype in our patient, but additional patients with ERC2 mutations or further functional studies are needed to confirm its involvement in epilepsy aetiology.

An explanation for additional symptoms identified in a patient with mutation in a known causative gene may be that the disease phenotype is poorly defined due to a limited
number of patients. In such cases the identification of additional patients is crucial, underlining the importance of this and similar studies. It is also important to point out that epileptic episodes have in many cases been observed to cause brain damage, and when investigating patients with both ID and epilepsy it is possible that the ID reported may be a consequence of an early epileptic episode.

One drawback of our study is that patients have been run sequentially over a longer time period, with a concurrent development in technology and analysis tools. Trios have therefore been sequenced with different capture kits and different sequencing approaches. We do not find statistically significant differences between these different technologies due to the limited size of our study, but there is a trend towards identification of more DNMs and more likely pathogenic mutations in more recent analyses. Still, we find an average of 0.79 DNMs per trio, which is similar or better than several previous large-scale trio exome sequencing studies, but lower than studies that have performed much deeper sequencing. It is therefore likely that several causative DNMs have been missed, especially in the trios sequenced first. Future whole-genome resequencing of these patients will hopefully provide a molecular diagnosis for additional families.

In summary, we identify variants likely to be pathogenic in 11 genes previously linked to ID and/or epilepsy, resulting in a molecular diagnostic yield of 28%. We also identified several mutations that point to candidate causative genes such as the PAN2, HECC2 and ERC2 genes. HECC2 is the strongest novel candidate as DNMs affecting a specific domain of the protein and have been identified in several studies in closely related disorders. Additional patients, better clinical phenotype information or functional studies will be required to conclusively determine the potential role of HECC2 in brain development. All in all this study underlines the potential and possibilities of using exome sequencing as a tool for identification of disease genes in a stringently selected group of patients, and the utility of using previous knowledge of protein interaction and biological pathways to prioritise candidate genes.

Acknowledgements
We are very grateful to the participating families for their cooperation. Sequencing was performed using the SciLifeLab National Genomics Infrastructure at Uppsala Genome Center and the Uppsala SNP & Seq Facility. Computational analyses were performed by personnel provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Contributors
JH, A-CT and LF wrote the manuscript. All authors read, commented on and approved the manuscript.

Funding
This work was supported by grants from the Föreningen Margarethahemmet, the Sävstaholm Society and the Regionala forskningsrådet (to A-CT), and the European Research Council ERC Starting Grant Agreement no. 282330 and the Swedish Medical Research Council (to LF). We are very grateful to the participating families for their cooperation. Sequencing was performed using the SciLifeLab National Genomics Infrastructure at Uppsala Genome Center and the Uppsala SNP & Seq Facility. We are very grateful to the participating families for their cooperation. Sequencing was performed using the SciLifeLab National Genomics Infrastructure at Uppsala Genome Center and the Uppsala SNP & Seq Facility.

Competing interests
None declared.

Ethics approval
Uppsala Ethical Review Board.

Provenance and peer review
Not commissioned; externally peer reviewed.

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES
1. Rogers HH. Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet 2010;11:161–87.
2. Bell GS, Sander JW. The epidemiology of epilepsy: the size of the problem. Seizure 2002;11(Suppl A):306–14; quiz 315–16.
3. Lhooq SD, Sander JNAS. The epidemiology of epilepsy and learning disability. Epilepsia 2001;42(Suppl 1):9–69; discussion 19–20.
4. McGrother CW, Bhaumik S, Thorp CF, Hauck A, Branford D, Watson JM. Epilepsy in adults with intellectual disabilities: prevalence, associations and service implications. Seizure 2006;15:376–86.
5. Miller DT, Adam MP, Aradhya S, Biethler AB, Carteau C, Chan C, Chirife M, Chishti H, Chueh A, Chopra U, Chua L, Chukwuka C, Churilov L, Cicek A, Cicerone K, Cifuentes-Rodríguez E, Cifuentes-Pescador J, Cisin G, Classen J, Clopper A, Cordero P, Corda N, Cornejo P, Correa E, Cucchiara J, Cui L, Cuculich J, Cuq PM, Dąbrowska M, Daniel M, Daniel M, Davidson P, Davidoff J, Davison K, Davis N, Davies K, Davison K, Davies N, De Davenport A, De Davenport B, De Davenport C, De Davenport D, De Davenport E, De Davenport F, De Davenport G, De Davenport H, De Davenport J, De Davenport K, De Davenport L, De Davenport M, De Davenport N, De Davenport O, De Davenport P, De Davenport Q, De Davenport R, De Davenport S, De Davenport T, De Davenport U, De Davenport V, De Davenport W, De Davenport X, De Davenport Y, De Davenport Z, De Davenport A, De Davenport B, De Davenport C, De Davenport D, De Davenport E, De Davenport F, De Davenport G, De Davenport H, De Davenport J, De Davenport K, De Davenport L, De Davenport M, De Davenport N, De Davenport O, De Davenport P, De Davenport Q, De Davenport R, De Davenport S, De Davenport T, De Davenport U, De Davenport V, De Davenport W, De Davenport X, De Davenport Y, De Davenport Z, De Davenport A, De Davenport B, De Davenport C, De Davenport D, De Davenport E, De Davenport F, De Davenport G, De Davenport H, De Davenport J, De Davenport K, De Davenport L, De Davenport M, De Davenport N, De Davenport O, De Davenport P, De Daven

New loci
