Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study

Anita Rauch*, Dagmar Wieczorek*, Elisabeth Graf†, Thomas Wieland*, Sabine Endele, Thomas Schwarzmayer, Beate Albrecht, Deborah Bartholdi, Jasmin Beyo, Nataliya Di Donato, Andreas Dufké, Kirsten Cremer, Maja Hempel, Denise Horn, Juliane Hoyer, Pascal Joset, Albrecht Röpke, Ute Moog, Angelika Riess, Christian T Thiel, Andreas Tschach, Antje Wiesener, Eva Wohleber, Christiane Zweier, Arif B Ekici, Alexander M Zink, Andreas Rump, Christa Meisinger, Harald Grallert, Heinrich Sticht, Annette Schenck, Hartmut Engels, Gudrun Rappold, Evelin Schröck, Peter Wierack, Olof Riess, Thomas Mettinger, André Reis†, Tim M Strom†

Summary

Background The genetic cause of intellectual disability in most patients is unclear because of the absence of morphological clues, information about the position of such genes, and suitable screening methods. Our aim was to identify de-novo variants in individuals with sporadic non-syndromic intellectual disability.

Methods In this study, we enrolled children with intellectual disability and their parents from ten centres in Germany and Switzerland. We compared exome sequences between patients and their parents to identify de-novo variants. 20 children and their parents from the KORA Augsburg Diabetes Family Study were investigated as controls.

Findings We enrolled 51 participants from the German Mental Retardation Network. 45 (88%) participants in the case group and 14 (70%) in the control group had de-novo variants. We identified 87 de-novo variants in the case group, with an exomic mutation rate of 1.71 per individual per generation. In the control group we identified 24 de-novo variants, which is 1-2 events per individual per generation. More participants in the case group had loss-of-function variants than in the control group (20/51 vs 2/20; p=0.022), suggesting their contribution to disease development. 16 patients carried de-novo variants in known intellectual disability genes with three recurrently mutated genes (STXBP1, SYNAP1, and SCN2A). We deemed at least six loss-of-function mutations in six novel genes to be disease causing. We also identified several missense alterations with potential pathogenicity.

Interpretation After exclusion of copy-number variants, de-novo point mutations and small indels are associated with severe, sporadic non-syndromic intellectual disability, accounting for 45–55% of patients with high locus heterogeneity. Autosomal recessive inheritance seems to contribute little in the outbred population investigated. The large number of de-novo variants in known intellectual disability genes is only partially attributable to known non-specific phenotypes. Several patients did not meet the expected syndromic manifestation, suggesting a strong bias in present clinical syndrome descriptions.

Funding German Ministry of Education and Research, European Commission 7th Framework Program, and Swiss National Science Foundation.

Introduction

Intellectual disability is defined as substantial impairment of cognitive and adaptive functions that has onset in childhood and has an estimated prevalence of 1·5–2·0%. Whereas frequency estimates of mild intellectual disability—the large reproductive disadvantage of the disease, which hampers use of linkage analysis and subsequent positional cloning. The introduction of array-based copy-number analysis led to the identification of de-novo microdeletions and duplications present in several genes in roughly 14% of patients with intellectual disability. Sequencing of candidate genes and use of next-generation sequencing techniques showed that a large proportion of sporadic cases might be caused by de-novo point mutations and small insertions or deletions. Our aim was to identify de-novo variants by exome sequencing in patients with intellectual disability.

Methods

Study design and patients

Between February and November, 2011, we enrolled patients with severe non-syndromic intellectual disability...
and both healthy, non-consanguineous parents from ten centres of the German Mental Retardation Network. Inclusion criteria were severe intellectual disability with grossly preserved motor function, absence of malformations, absence of syndrome-specific minor anomalies, absence of specific neurological signs, absence of causative copy number variants by high resolution molecular karyotyping, non-consanguineous parents, and sporadic disease. The study was approved by the ethics committee of each participating centre and written informed consent was obtained from the guardians of all patients.

Patients with non-specific minor anomalies (eg, single transverse crease of palm) and neurological signs were not excluded from the classification of non-syndromic intellectual disability because most patients with severe intellectual disability have such anomalies.

All patients were pre-screened with genome-wide high-resolution arrays and those with de-novo copy-number variations and known disease-causing copy-number variations were excluded. 20 control trios were recruited from the KORA Augsburg Diabetes Family Study on type 2 diabetes.

### Procedures

We extracted DNA from peripheral blood leucocytes. We enriched exomes in solution provided by the manufacturer and indexed them with SureSelect XT Human All Exon 50 Mb kits (Agilent, Santa Clara, CA, USA). We sequenced samples as 100 bp paired-end runs on a HiSeq2000 system (Illumina, San Diego, CA, USA). Pools of 12 indexed libraries were sequenced on four lanes. To identify putative de-novo variants, we did read mapping, variant calling, and variant annotation of affected individuals and their parents (appendix). To exclude false positives, we investigated the identified de novo variants manually with the Integrative Genomics Viewer.

To assess the sensitivity of variant detection, we compared nonreference single nucleotide polymorphisms established by Affymetrix 6.0 array with respective sites in exome data.9–11

All de-novo, rare compound heterozygous, homozygous, and X chromosome variants detected by exome sequencing were validated by Sanger sequencing (appendix). Possible mosaicism was investigated for 26 variants from 13 participants who could be re-contacted, with buccal smear as the comparison tissue. We validated a 23 bp deletion in MECP2 by melting curve analyses with two independent SYBR Green real-time PCR assays of the patient, her parents, and four controls. We used PolyPhen2 for computational prediction of the functional effect of missense mutations. We used a list of genes with their probability of being haploinsufficient for the prediction of their involvement in autosomal dominant disease.12 We used phylp scores—provided by the UCSC Genome Browser—to analyse evolutionary sequence conservation.

The study aims were to compare mutation rates in cases and controls, to characterise the possible role in intellectual disability of genes carrying de-novo variants by functional predictions, and to identify possible disease-causing variants.

### Statistical analysis

We used the Mann-Whitney U test to compare the haploinsufficiency and phylp scores between genes carrying a de-novo variant and all genes, and to compare the number of de-novo variants between cases and controls. We used Fisher exact test to compare the proportion of loss-of-function variants between case and control groups. Data were analysed with 2-sided tests. The analyses were done with R (version 2.10.0).

### Role of funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Results

We enrolled 51 patients from the German Mental Retardation Network, 32 girls and 19 boys. All but three had non-specific intellectual disability with an intelligence quotient of less than 60, and all but five had height, weight, or head circumference within 2·5 SDs of the mean. 17 patients had a history of seizures and 15 had autistic behaviour. 36 patients had cerebral MRI scans, of whom 17 had non-specific minor anomalies such as mild brain atrophy, delayed myelination, and cysts.

To identify variants, we sequenced exomes of the 213 participants to high depth (median 112-times, at least 63-times), resulting in about 90% of nucleotides in the target region covered by at least 20-times. By comparing the sequencing data with the reference sequence (human genome assembly hg19), we detected on average 10 500 synonymous and 9600 non-synonymous variants, applying the same criteria as for detection of de-novo variants. We estimated the sensitivity to detect any single nucleotide variation to be 97.5% by comparing

<table>
<thead>
<tr>
<th>Cases (n=51)</th>
<th>Controls (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>16</td>
</tr>
<tr>
<td>Ts vs Tv</td>
<td>42 vs 14</td>
</tr>
<tr>
<td>CpG</td>
<td>15</td>
</tr>
<tr>
<td>Nonsense</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>3 vs 2</td>
<td>0 v 0</td>
</tr>
<tr>
<td>Splice site</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0 vs 0</td>
<td>1</td>
</tr>
<tr>
<td>Frameshift</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>0 vs 0</td>
<td>1</td>
</tr>
<tr>
<td>Synonymous</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>10 vs 1</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>20</td>
</tr>
<tr>
<td>55 vs 17</td>
<td>24</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1: De novo variants

For PolyPhen2 see [http://genetics.bwh.harvard.edu/pph2/](http://genetics.bwh.harvard.edu/pph2/)

For the UCSC Genome Browser see [http://genome.ucsc.edu](http://genome.ucsc.edu)
heterozygous non-reference single nucleotide polymorphisms assessed by array hybridisation with the corresponding sites in the exome data of 26 participants. This sensitivity is much the same as that from whole-genome sequencing and shows that our approach enabled highly sensitive variant detection. In the entire sample we defined 111 de-novo variants that were confirmed by Sanger sequencing (table 1, appendix).

We used the occurrence of de-novo point mutations (silent, missense, nonsense, and conserved splice sites) as an approximation for the genome-wide mutation rate and compared it with previous studies with exome data13–15 or whole genome data.16 The average number of de-novo point mutations per person was 1·41 in the case group and 1·15 in the control group. Most of these mutations were transitions (table 1). We estimated that the point mutation rate in coding regions is 29% higher than in the genome (appendix).

Mutation rates could be inflated because of somatic mosaicism. Four variants in the case group and three in the control group had a small peak on Sanger sequencing and were represented by less than 25% of reads in the exome sequencing data, so we deemed them as likely to be a result of somatic mosaicism. These variants were excluded from the analyses and calculations of mutation rates.

The four variants in the case group were detected by Sanger sequencing of DNA from buccal smears, but again had only a small peak. We analysed 22 additional de-novo variants from 13 patients by buccal smear in the case group; we detected no mosaicism. These data suggest that the proportion of possible somatic mosaics is small.

To detect disease-causing mutations we compared the frequencies and characteristics of de-novo variants in patients and controls. 45 (88%) participants in the case group and 14 (70%) participants in the control group had de-novo variants (figure 1). The synonymous mutation rate was lower in cases compared with controls, whereas the average number of protein-altering (missense, nonsense, frameshift, and splice site) variants was significantly higher in the case group than in the control group (table 2). Specifically, a significantly higher number of people in the case group than in the control group had loss-of-function variants (20 of 51 [39%] vs two of 20 [10%; p=0·022). The role of a large proportion of de-novo variants in causing intellectual disability is also

![Figure 1: Number of de-novo variants per person](image)

Number of variants shown in the case group (A) and the control group (B). The average number of de-novo variants was 1·68 per person in the case group and 1·2 per person in the control group.

<table>
<thead>
<tr>
<th></th>
<th>Sequence length (bases)</th>
<th>Protein-altering</th>
<th>Missense</th>
<th>Nonsense</th>
<th>Synonymous</th>
<th>Protein-altering/ synonymous</th>
<th>Missense/ synonymous*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases (n=51)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutations</td>
<td>2·90×10⁷</td>
<td>3 8×10⁷</td>
<td>2·90×10⁷</td>
<td>9 03×10⁷</td>
<td>6 90</td>
<td>5·09</td>
<td></td>
</tr>
<tr>
<td>Mutations per person</td>
<td>1·49</td>
<td>1·41</td>
<td>1·10</td>
<td>0·22</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Mutation rate (mutations per person per generation)</td>
<td>2·57×10⁻⁴</td>
<td>1·86×10⁻⁴</td>
<td>1·89×10⁻⁴</td>
<td>1·19×10⁻⁴</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>Controls (n=20)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutations</td>
<td>17</td>
<td>23</td>
<td>35</td>
<td>7</td>
<td>2·43</td>
<td>2·14</td>
<td></td>
</tr>
<tr>
<td>Mutations per person</td>
<td>0·85</td>
<td>1·15</td>
<td>0·75</td>
<td>0·35</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Mutation rate (mutations per person per generation)</td>
<td>1·47×10⁻⁴</td>
<td>1·51×10⁻⁴</td>
<td>1·29×10⁻⁴</td>
<td>1·94×10⁻⁴</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>p value‡</td>
<td>0·008</td>
<td>0·46</td>
<td>0·23</td>
<td>0·15</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Protein-altering variants are defined as the sum of missense, nonsense, splice site, and frameshift variants. Point mutations are defined as the sum of synonymous, missense, nonsense, and point mutations at the conserved splice sites. The expected ratio (2·23) is substantially different from that of the case group.‡ See appendix for details of calculation of sequence lengths of changes. †For number of mutations in cases vs controls.

Table 2: Mutations
**Sex** | **Gene** | **Online Mendelian Inheritance in Man reference** | **Type** | **Genomic change** | **Protein change** | **Haploinsufficiency index (%)** | **PolyPhen2 category (score)**
--- | --- | --- | --- | --- | --- | --- | ---
E10-0275 | Male | IQSEC2 | Nonsense | X chromosome: g.53277315G→A | Arg855* | 9.8% | --
BH17109 | Female | MECP2 | Frameshift | X chromosome: g.152936903_152936915del | Pro401Argfs*8 | 24.3% | --
ZH58769 | Male | NAA10 | Missense | X chromosome: g.153197564G→A | Arg116Trp | -- | Benign (0.233)
ERS2725 | Female | SATB2 | Missense | Chromosome 2: g.200213455A→C | Val381Gly | 4.3% | Probably damaging (1)
ER84940 | Male | SCN2A | Frameshift | Chromosome 2: g.166179821_166179822delCT | Leu611Valfs*35 | 12.7% | --
MS111684 | Male | SCN2A | Frameshift | Chromosome 2: g.166172100_166172101insA | Asn503Lysfs*19 | 12.7% | --
ZH60991 | Male | SCN2A | Missense | Chromosome 2: g.166201311C→T | Arg937Cys | 12.7% | Probably damaging (1)
ER12988 | Male | SCN2A | Missense | Chromosome 12: g.52001200G→A | Arg1617Gln | 12.7% | Probably damaging (1)
BO2210 | Female | SETBP1 | Nonsense | Chromosome 18: g.42531079A→T | Lys592* | 9.9% | --
PL111540 | Male | SLC2A1 | Missense | Chromosome 1: g.43396356G→A | Arg331Cys | 24.1% | Possibly damaging (1)
E5070046 | Female | STXB1P | Missense | Chromosome 9: g.130422388G→C | Ala101Pro | 8.7% | Possibility damaging (0.860)
MR-NET001 | Female | STXB1P | Splice | Chromosome 9: g.130422388delC | Aberrant splicing predicted | 8.7% | --
P4726 | Female | STXB1P | Missense | Chromosome 9: g.130422065G→A | Gln598Lys | 8.7% | Probable damaging (0.994)
BO14109 | Female | SYNGAP1 | Frameshift | Chromosome 6: g.33410959A, 334109590insT | Thr878Aspfs*60 | 23.6% | --
ERS3894 | Male | SYNGAP1 | Frameshift | Chromosome 6: g.33405939A, 334059390insT | Lys418Argfs*54 | 23.6% | --
TUBA080997 | Female | TCF4 | Missense | Chromosome 18: g.52070275G→A | Ser110Leu | 1.9% | Benign (0.073)

*Molecular modelling suggests that the bulky Trp116 side chain interferes with coenzyme A binding, thereby affecting enzymatic activity. †The crystal structure indicates that Ala101 is located at the N-terminal region of a sheet structure. The γ angle of -158° is not possible for proline, and molecular modelling suggests that this mutation destabilizes the structure and probably also hampers ligand binding (appendix).

**Figure 2:** Comparisons of probability of haploinsufficiency (A) and phyloP score (B) between cases, controls, and exome average.

The probabilities of haploinsufficiency (n=17 070) and the phyloP scores (n=19 592) of most genes in the exome were compared with the scores of genes carrying de-novo variants in cases and controls. The genes in the case group were further subdivided into genes known to be associated with intellectual disability, novel genes with loss-of-function variants (nonsense, splice site, and frameshift mutations) and novel genes with missense variants. We calculated phyloP scores for entire genes as the average of all single nucleotides. The whiskers extend to the most extreme datapoint that is no more than 1.5-times the IQR from the box. Circles are single values of the corresponding genes.

**Table 3:** Missense, nonsense, frameshift, and splice site de-novo variants in genes associated with intellectual disability in each patient-parent trio.
supported by a high probability of haploinsufficiency and high evolutionary sequence conservation of the affected genes (figure 2). The probability of haploinsufficiency was significantly higher for genes carrying a de-novo variant (median=0.358) than for all genes in the genome (0.165; Mann-Whitney U test p<0.0001). Likewise, the average phyloP score of the genes carrying a de-novo variant (1.81) was higher than that of all genes (1.11; p<0.0001).

We detected 16 de-novo variants in genes known to be associated with intellectual disability (table 3). Eight were loss-of-function variants and eight were missense variants. With the exception of two, the missense variants were predicted to be possibly or probably damaging by

<table>
<thead>
<tr>
<th>Sex</th>
<th>Gene</th>
<th>Genomic change (reference sequence hg19)</th>
<th>Mutation type</th>
<th>Haploinsufficiency index (%)</th>
<th>Smallest region of overlap</th>
<th>Brain expression</th>
<th>Mouse phenotype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>AR8H1</td>
<td>Chromosome 15: g.7284768deleT</td>
<td>Frameshift</td>
<td>24.4%</td>
<td>--</td>
<td>Ubiquitous</td>
<td>--</td>
<td>Ubiquitin-conjugating enzyme E2 binding protein</td>
</tr>
<tr>
<td>Female</td>
<td>CHD2</td>
<td>Chromosome 15: g.93498742deleG</td>
<td>Frameshift</td>
<td>15.8%</td>
<td>Deletion 3 genes 472 bp</td>
<td>Ubiquitous</td>
<td>--</td>
<td>The CHD gene family is characterised by chromatin organisation modifier and SNF2-related helicase and ATPase domains</td>
</tr>
<tr>
<td>Female</td>
<td>HIVEP2</td>
<td>Chromosome 6: g.145081688delC</td>
<td>Frameshift</td>
<td>16.2%</td>
<td>Deletion 24 genes 702 bp</td>
<td>Very high</td>
<td>MGI:1338076; homozygous knock-out mice have abnormal thymus anatomy</td>
<td>A transcription factor that interacts with TCF4</td>
</tr>
<tr>
<td>Female</td>
<td>SETD5</td>
<td>Chromosome 3: g.94902707C→T</td>
<td>Nonsense</td>
<td>21.3%</td>
<td>Deletion 7 genes 702 bp</td>
<td>Ubiquitous</td>
<td>--</td>
<td>Uncharacterised protein that contains a SET domain; SET7 is a histone H3 lysine methyltransferase</td>
</tr>
<tr>
<td>Female</td>
<td>SLC6A1</td>
<td>Chromosome 3: g.10060365delT</td>
<td>Frameshift</td>
<td>53.6%</td>
<td>Deletion 10 genes 1.5 Mb</td>
<td>Very high</td>
<td>MGI:95627; homozygous hypomorphic mice have abnormal γ-aminobutyric-acid uptake and release</td>
<td>γ-aminobutyric-acid transporter</td>
</tr>
<tr>
<td>Female</td>
<td>SYNCRIP</td>
<td>Chromosome 6: g.86328428T→G</td>
<td>Frameshift</td>
<td>1.5%</td>
<td>Duplication 5 genes 698 bp</td>
<td>Ubiquitous</td>
<td>--</td>
<td>Nuclear ribonucleoprotein, involved in mRNA processing</td>
</tr>
<tr>
<td>Male</td>
<td>CUX2</td>
<td>Chromosome 12: g.131748934G→A</td>
<td>Missense</td>
<td>40.0%</td>
<td>Deletion 54 genes 4.8 Mb</td>
<td>Very high</td>
<td>MGI:107321; homozygotes for a targeted null mutation have various neural defects</td>
<td>Belongs to the CUT homoeobox family, thought to be a transcription factor involved in neural specification</td>
</tr>
<tr>
<td>Female</td>
<td>DEAF2</td>
<td>Chromosome 11: g.656817T→G</td>
<td>Missense</td>
<td>58.6%</td>
<td>No deletion</td>
<td>Very high</td>
<td>MGI:1858496; many homozygotes have exencephaly</td>
<td>Reported by Vissers and colleagues</td>
</tr>
<tr>
<td>Female</td>
<td>E1F2C1</td>
<td>Chromosome 1: g.35393331T→C</td>
<td>Missense</td>
<td>84.5%</td>
<td>Deletion 22 genes 1.1 Mb</td>
<td>Ubiquitous</td>
<td>--</td>
<td>Reported by Sanders and colleagues</td>
</tr>
<tr>
<td>Female</td>
<td>KCNQ3</td>
<td>Chromosome 8: g.1333192493G→A</td>
<td>Missense</td>
<td>7.3%</td>
<td>Deletion 14 genes 3.3 Mb</td>
<td>High</td>
<td>MGI:1336181; mice homozygous for a knock-in allele have spontaneous seizures and premature death; defects in KCNQ3 cause benign neonatal epilepsy type 2 (EBN2) [MIM:221201] and epileptic encephalopathy</td>
<td>M channel, a slowly activating and deactivating potassium channel that has a critical role in regulation of neuronal excitability</td>
</tr>
<tr>
<td>Female</td>
<td>STAG1</td>
<td>Chromosome 3: g.136240090T→C</td>
<td>Missense</td>
<td>6.8%</td>
<td>ISCA msv577920 deletion 22 genes 3.9 Mb</td>
<td>High</td>
<td>--</td>
<td>Encodes a component of cohesion, functionally related to SMC1 and SMC3 causing Cornelia de Lange syndrome</td>
</tr>
<tr>
<td>Female</td>
<td>ZNF238</td>
<td>Chromosome 1: g.244218559C→G</td>
<td>Missense</td>
<td>2.6%</td>
<td>Deletion 4 genes 1 Mb</td>
<td>Very high</td>
<td>MGI:1253609; homozygous knock-out mice have neonatal lethality, cortical and hippocampal hypoplasia, laminar disorganisation, and abnormal neuron apoptosis</td>
<td>CHD2-type zinc finger protein, possibly involved in chromatin assembly; within microdeletion region</td>
</tr>
</tbody>
</table>

Table 4: Probable disease-causing de-novo variants in each patient-parent trio

*Numbers in brackets are the probability of the mutation being damaging according to polyPhen (>0.85 probably damaging, 0.15–0.85 possibly damaging). Modelling of DEAF1 suggests that Gln264 forms a hydrogen bond with Lys253, which is part of the DNA-binding interface. This hydrogen bond cannot form in the Gln264Pro mutant, which is thought to destabilise the protein’s structure and affect its DNA-binding properties (appendix).
PolyPhen2. Although two variants were classified as benign, the phenotypic features of the affected patient and predictions based on protein structure suggest that the NAA10 variant has a causal effect (appendix). Three genes—STXBP1, SCN2A, and SYNGAP1—had more than one de-novo variant. For one mutation, the results of Sanger and exome sequencing were discordant: a girl with symptoms suggestive of Rett syndrome was negative for MECP2 mutations in the initial diagnostic investigation, but exome sequencing detected a 23 bp deletion in the C-terminal part of MECP2. Sanger sequencing produced a small, ambiguous peak. Deleted alleles had small peaks on electropherograms, but two different alleles could clearly be seen in melting curve analysis (data not shown).

A similar range of mutations probably contributes to the pathogenesis of intellectual disability in novel genes. Of the 27 cases that had at least one de-novo variant, 11 had a de-novo loss-of-function variant. However, no novel genes had multiple de-novo variants, which suggests high non-allelic heterogeneity. We assessed the possible pathogenicity of variants with several criteria (table 4). We judged six loss-of-function mutations to cause disease in the genes ARHI1, CDH2, HIVEP2, SETD5, SLC6A1, and SYNGRP. Functional knowledge, gene homology or mouse models, and findings from previous studies suggest that some of the remaining genes affected by missense mutations might also cause intellectual disability (KCNQ3, CUX2, ZNF238, STAG1, DEAF1, and EIF2C).

We tested whether the 12 candidate genes had any loss-of-function mutations in roughly 1600 control exomes and did a post-hoc test in the 6500 exomes from the Exome Variant Server (version 0.0.14), and in 179 low-coverage genomes of the 1000 Genomes project. The Exome Variant Server dataset does not contain small deletions or insertions. With the exception of SYNGRP, for which a single nonsense variant was present in the Exome Variant Server dataset, the candidate genes did not carry any loss-of-function variant, supporting the likely pathogenicity of the candidate de-novo variants identified.

We investigated possible autosomal recessive and X-linked inheritance including all variants absent from our control patients and the 1000 Genomes dataset (appendix). In our study, we excluded individuals with known parental consanguinity. Accordingly, we recorded only one rare homozygous missense variant, located in CDK18. However, the same patient carried a DEAF1 de-novo variant, which seems more likely to cause disease because it has been associated with intellectual disability previously. We also identified rare compound heterozygous variants in 18 genes (appendix), but the affected genes have not previously been implicated in intellectual disability, with the exception of ACY1. Variants in three possible candidates (TUBAL3, BDP1, ACY1) seem unlikely to cause intellectual disability because the participants also had a de-novo mutation in a gene known to cause intellectual disability and, in the case of ACY1, increased urinary excretion of N-acetylated aminoacids was excluded. Too little is known about two other candidates (FKBP1 and DLG3) to make conclusions about their role in intellectual disability. These results suggest that the contribution of autosomal recessive inheritance is small in outbred populations.

Aside from the two X chromosome de-novo variants in genes known to cause intellectual disability (IQSEC2 in a boy and MECP2 in a girl), we recorded only one maternally inherited nonsense variant, located in DLG3—a known X-linked intellectual disability gene—as a likely mutation in a boy. Thus, the contribution of X-linked inheritance to intellectual disability in our study was roughly 10% in male participants. 38 more rare X chromosome inherited variants were missense variants, and one nonsense mutation existed in a possible pseudogene, ATXN3L. A maternally inherited missense mutation in the fragile X-E syndrome gene AFF2 was detected in a boy who also carried a de-novo SLC2A1 missense mutation. Because the SLC2A1 mutation is sufficient to explain the phenotype, the AFF2 mutation was deemed a rare polymorphism.

Discussion

The mutation rate for coding variants was high in cases compared with controls and with patients included in previous studies of autism spectrum disorders. Including variants both in genes known to cause intellectual disability and in new candidate genes, we estimate that—after exclusion of copy number variations—severe intellectual disability could be caused by de-novo variants in approximately 45–55% of patients. Because of the high locus heterogeneity, final conclusions about the pathogenicity of each individual mutation cannot be made. Although truncating mutations in genes associated with intellectual disability are very likely to cause disease, the phenotypic effects of variants in novel genes are uncertain until more patients are studied and in-depth functional studies have been done. Nevertheless, our results show that de-novo point mutations and small indels have the potential to be a major cause of severe, sporadic non-syndromic intellectual disability whereas X-linked and autosomal recessive inheritance seem to be much rarer in a non-consanguineous population. This finding accords with an empirical risk of recurrence of 8–4% for non-syndromic intellectual disability and an estimated proportion of 10–12% of X-linked recessive inheritance among boys with the disease.

Locus heterogeneity of intellectual disability is high—as shown by targeted Sanger sequencing of specific genes—which impedes the complete characterisation of the genetics of intellectual disability. However, de-novo mutations in three genes, SYNGAP1, STXBP1, and SCN2A were present in 16% of patients. Although SYNGAP1 mutations have been reported in 2% of patients with...
non-specific intellectual disability and STXB1P1 mutations in 3%.12,13 The high prevalence of SCN2A mutations was unexpected. The high heterogeneity in intellectual disability is supported by the small overlap of genes with de-novo mutations in our study and other reports on intellectual disability9 and autism spectrum disorder.11–12,24

Overlap occurred in genes known to be associated with intellectual disability (SCN2A and SETBP1) and in novel genes (CHD2, CTTNB2, DEAF1, EIF2C1, GPRASPI, SETD5, and SLC6A4). One of the three patients with SCN2A mutations and the patient with an SETBP1 truncating mutation also had signs of autism, further supporting the overlapping causes of some neuro-developmental disorders. Conversely, de-novo missense mutations have been reported in CHD2, SETD5, and SLC6A4 in autistic patients but our three patients with frameshift mutations in these genes did not have autism. These findings suggest that missense mutations in these genes might cause autism whereas truncating mutations in the same genes might cause severe intellectual disability without autism. In some cases the sensitivity of exome sequencing might exceed that of targeted Sanger sequencing, as suggested by the MECP2 mutation that was not detected by Sanger sequencing in our study.

The high number of variants in genes known to be associated with intellectual disability is only partially attributable to the non-specific nature of the described clinical phenotypes. Several mutations did not cause the expected phenotype, which shows that only unbiased large-scale sequencing is capable of characterising the complete clinical range associated with mutations in specific genes.

In the absence of an understanding of the molecular causes of intellectual disability, diagnostic classification is often based on specific phenotypes. However, several genes in which mutations have first been associated with a specific phenotype have been subsequently shown to have a broader phenotypic spectrum. This finding is consistent with our study. For example, SCN2A mutations have been shown to cause early-onset infantile epileptic encephalopathy,21 but none of the three patients in this study who had such mutations had a history of seizures. Instead, they all had the same clinical signs with severe intellectual disability, autoaggressive behaviour, and similar facial features (appendix). We also identified a nonsense mutation in SETBP1 that is located in the critical region of (18)(q12.2q21.1).22 SETBP1 missense mutations clustering in an 11 bp region can cause Schinzel-Giedion syndrome, probably by a dominant negative or gain-of-function mechanism.23 The patient in our study did not have the typical features associated with this syndrome, suggesting that loss-of-function mutations cause a different non-syndromic phenotype as has been suggested previously.15 We also detected a nonsense mutation in SATB2 (reported to cause intellectual disability with cleft palate)44 and missense mutations in NAA10 (reported to cause Ogden syndrome45) and SATB2 mutation that was not detected by Sanger sequencing in our study.

The high number of variants in genes known to be associated with intellectual disability is only partially attributable to the non-specific nature of the described clinical phenotypes. Several mutations did not cause the expected phenotype, which shows that only unbiased large-scale sequencing is capable of characterising the complete clinical range associated with mutations in specific genes.

### Panel: Research in context

**Systematic review**

We searched for reports in PubMed with the terms “exome sequencing” and “de novo exome”. For each gene we studied, all alternative names were searched for in the OMIM database and PubMed, and in reports we found. We included reports published up until Aug 13, 2012.

**Interpretation**

Similar exome sequencing studies of de-novo variants in patient–parent trios have been done for autism spectrum disorders.31–33,46,47 Previously published studies of small numbers of patients’ or of candidate genes’ suggest a significant contribution of de-novo mutations to intellectual disability. Our study is the first of a large cohort of patients investigated with whole exome sequencing to report a significantly higher number of de-novo mutations in patients with non-syndromic severe intellectual disability than in controls. We also show, in an unbiased study of patients with intellectual disability, that present syndrome descriptions have a strong bias towards particular clinically recognised phenotypes and that sequencing is needed to understand fully the complex relation between genotype and phenotype. Our findings therefore suggest that large-scale sequencing should become a first-tier clinical diagnostic test for patients without a recognisable condition. Making diagnosis easier will also release resources for in-depth genotype-phenotype and natural history studies. Combination of such studies with research into mechanisms of disease should lead to improved patient care and novel treatment strategies.
involved in pathways previously associated with intellectual disability; for example, ubiquitin conjugation (ARIH1), chromatin modification (CDH2) or methylation (SETD5), γ-aminobutyric-acid transport (SLC6A1), and mRNA processing and transport to dendrites (SYNCRIP). The products of two of these genes interact with gene products from genes associated with intellectual disability. The Drosophila homologue of ARIH1 (ari-1) interacts with and regulates NRII3 (Ecr), a gene mutated in patients with Kleefstra syndrome spectrum disorders. Furthermore, the transcription factor HIVEP2, also named MIBP1 or MBP2, is coexpressed and interacts with TCF4—a gene associated with intellectual disability—when somatostatin receptors are expressed in the frontal cortex and hippocampus.

Some of the remaining genes affected by missense mutations (KCNQ3, CUX2, ZNF238, STAG1, DEAF1, and EIF2C1) might also cause intellectual disability. For example, KCNQ2 and KCNQ3 mutations, which cause benign familial neonatal epilepsy, are also responsible for recurrent seizures or other neuronal signs including intellectual deficits in 15% of patients. Furthermore, KCNQ2 de-novo mutations cause neonatal epileptic encephalopathy with substantial intellectual disability. The KCNQ3 de-novo missense mutation reported in our study might contribute to the phenotype of patient TUTLN, who has severe intellectual disability and multifocal epileptic activity according to electroencephalogram. Further candidates are the homeobox transcription factor CUX2, which regulates dendrite branching, spine development, and synapse formation in layer 2–3 neurons of the cerebral cortex, ZNF238, which is implicated in chromatin assembly and is a candidate gene for intellectual disability with corpus callosum hypogenesis, and STAG1, which codes for a protein that interacts with the cohesion complex. This complex contains NIPBL, SMC1, and SMC3, which are mutated in Cornelia de Lange syndrome. Finally, we identified de-novo missense variants in DEAF1 and EIF2C1 that have been linked with intellectual disability or autism spectrum disorder. The variant in DEAF1 is close to the previously reported variant. EIF2C1 encodes a member of the Argonaute family of proteins, which have a role in RNA interference and are enriched in dendritic spines and postsynaptic densities.

We did not consider the potential pathogenicity of mosaic mutations in this study, but the higher sensitivity of exome sequencing for such mutations compared with Sanger sequencing will help to clarify their contribution to intellectual disability in future studies. In view of the high diagnostic yield of exome sequencing in our study, this technique could be used for diagnosis of unexplained intellectual disability and even as part of a first-line standard diagnostic approach. However, to fully exploit the diagnostic potential of exome sequencing, more knowledge about the frequency and pathogenicity of sequence variants is needed.

Contributors
ARa, DW, BA, DB, NDD, AD, KC, MH, DH, JH, UM, ARi, CTT, AT, AW, EW, CZ, PJ, JB, ARö, ABE, AMZ, ARu, AS, SE, HE, GR, ES, PW, OR, and AR contributed to clinical genetics investigations, microarray analyses, and confirmation of exome sequencing findings. CM and HG provided control samples and data. HS analysed the protein structures. ARa and DW coordinated assessment of the clinical data. EG, TW, TS, and TMS were responsible for exome sequencing and data analysis. SE coordinated the confirmation of exome sequencing findings. ARa, DW, ARö, TW, TM, and TMS interpreted the data and wrote the manuscript. All authors approved the final version.

Conflicts of interest
We declare that we have no conflicts of interest.

Acknowledgements
We thank all patients and their families for participating in this study. This work was supported by the German Ministry of Education and Research (01GS08160, 01GR0802, and 01GM0867), the European Commission 7th Framework Program (261123, EUVADIS), and the Swiss National Science Foundation (120030_135669).

References


