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 Schwarzmayr, Beate Albrecht, Deborah Bartholdi, Jasmin Beygo, Nataliya Di Donato, Andreas Dufke, Kirsten Cremer, Maja Hempel, Denise Horn, Juliane Hoyer, Pascal Joset, Albrecht Röpke, Ute Moog, Angelika Riess, Christian T Thiel, Andreas Tzschach, Antje Wiesener, Eva Wohlleber, 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 Wieacker, Olaf Riess, Thomas Meitinger, André Reis†, Tim M Strom†*

Lancet 2012; 380: 1674–82

Published Online September 27, 2012 http://dx.doi.org/10.1016/S0140-6736(12)61480-9

See Comment page 1630

*These authors contributed equally

†These authors contributed equally

Institute of Medical Genetics (Prof A Rauch MD, D Bartholdi MD, P Joset PhD), **Neuroscience Center Zurich** (Prof A Rauch), **Zurich Center of Integrative Human Physiology** (Prof A Rauch), **University of Zurich, Schwerzenbach-Zurich, Switzerland; Institute of Human Genetics, University of Erlangen-Nuremberg, Erlangen, Germany** (Prof A Rauch, S Endele PhD, J Hoyer MD, C T Thiel MD, A Wiesener MD, C Zweier MD, A B Ekici PhD, Prof A Reis MD); **Institute of Human Genetics, Universitätsklinikum Essen, Essen, Germany** (D Wieczorek MD, B Albrecht MD, J Beygo MSc, K Cremer MD); **Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany** (E Graf MSc, T Wieland MSc, T Schwarzmayr MSc, Prof T Meitinger MD, T M Strom MD); **Institute of Clinical Genetics, Technische Universität Dresden, Dresden, Germany** (N Di Donato MD, A Rump PhD, Prof E Schröck MD); **Institute of Human Genetics, Eberhard Karls Universität Tübingen, Tübingen, Germany** (A Dufke MD, A Riess MD, A Tzschach MD, Prof O Riess MD); **Institute of Human Genetics, Technische Universität München, Munich, Germany** (M Hempel MD, Prof T Meitinger,T 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*, *SYNGAP1*, 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[1] and has an estimated prevalence of 1·5–2·0%.[2] Whereas frequency estimates of mild intellectual disability differ between studies, most researchers agree that severe intellectual disability (an intelligence quotient of <50) has a prevalence of 0·3–0·4%.[2] Studies of genetic changes in children with intellectual disability have led to the identification of mutations in many genes on the X chromosome[3] and some genes on autosomal chromosomes.[4,5] However, the genetic basis of the disorder is still unclear in most affected children, especially those with non-syndromic intellectual disability, in whom there are no physical signs.[6] The difficulty of establishing the genetics of intellectual disability could be because of the high locus heterogeneity and—for autosomal dominant inheritance—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.[7] 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.[4,8] 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 leuckocytes. 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 phyloP 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 phyloP 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

	Cases (n=51)			Controls (n=20)		
	Variants	Ts *vs* Tv	CpG	Variants	Ts *vs* Tv	CpG
Missense	56	42 *vs* 14	16	15	11 *vs* 4	3
Nonsense	5	3 *vs* 2	2	0	0 *vs* 0	0
Splice site	2	0 *vs* 0	0	1	0 *vs* 1	0
Frameshift	13	0 *vs* 0	0	1	0 *vs* 0	0
Synonymous	11	10 *vs* 1	2	7	6 *vs* 1	3
Total	87	55 *vs* 17	20	24	17 *vs* 6	6

Ts *vs* Tv=transition *vs* transversion. CpG=variant located in CpG dinucleotide.

Table 1: **De novo variants**

Institute of Medical Genetics, Charité, Universitätsmedizin Berlin, Berlin, Germany (D Horn MD)**; Institute of Human Genetics, Westfälische Wilhelms-Universität Münster, Münster, Germany** (A Röpke PhD, Prof P Wieacker MD)**; Institute of Human Genetics, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany** (U Moog MD, Prof G Rappold PhD)**; Institute of Human Genetics, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany** (E Wohlleber MD, A M Zink MSc, H Engels PhD)**; Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany** (C Meisinger MD)**; Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany** (H Grallert PhD)**; Institute of Biochemistry, University of Erlangen-Nuremberg, Erlangen, Germany** (H Sticht PhD)**; and Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands** (A Schenck PhD)

Correspondence to:
Dr Tim M Strom, Institute of Human Genetics, Helmholtz Zentrum München, Ingolstädter Landstr 1, 85764 Neuherberg, Germany
timstrom@helmholtz-muenchen.de

See **Online** for appendix

For the **Integrative Genomics Viewer** see http://www.broadinstitute.org/igv/

For **PolyPhen2** see http://genetics.bwh.harvard.edu/pph2/

For the **UCSC Genome Browser** see 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 data[13–15] or whole genome data.[9,16] 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

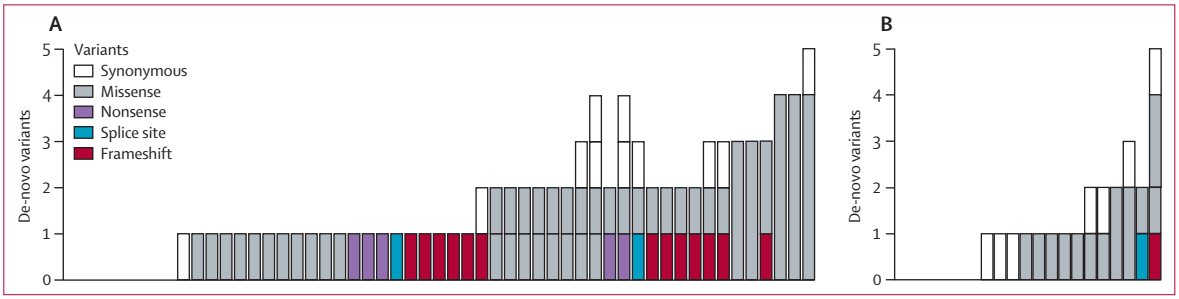

Figure 1: **Number of de-novo variants per person**
Number of variants shown in the case group (A) and the control group (B). The average number of de-novo variants was 1·68 per person in the case group and 1·2 per person in the control group.

	Protein-altering	Point	Missense	Synonymous	Protein-altering/ synonymous	Missense/ synonymous*
Sequence length (bases)†	$2{\cdot}90\times10^{7}$	$3{\cdot}8\times10^{7}$	$2{\cdot}90\times10^{7}$	$9{\cdot}03\times10^{6}$	..	..
Cases (n=51)						
Mutations	76	72	56	11	6·90	5·09
Mutations per person	1·49	1·41	1·10	0·22	..	..
Mutation rate (mutations per person per generation)	$2{\cdot}57\times10^{-8}$	$1{\cdot}86\times10^{-8}$	$1{\cdot}89\times10^{-8}$	$1{\cdot}19\times10^{-8}$	..	..
Controls (n=20)						
Mutations	17	23	15	7	2·43	2·14
Mutations per person	0·85	1·15	0·75	0·35	..	..
Mutation rate (mutations per person per generation)	$1{\cdot}47\times10^{-8}$	$1{\cdot}51\times10^{-8}$	$1{\cdot}29\times10^{-8}$	$1{\cdot}94\times10^{-8}$	..	..
p value‡	0·008	0·46	0·23	0·15	..	..

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.[8] †See appendix for details of calculation of sequence lengths of changes. ‡For number of mutations in cases *vs* controls.

Table 2: **Mutations**

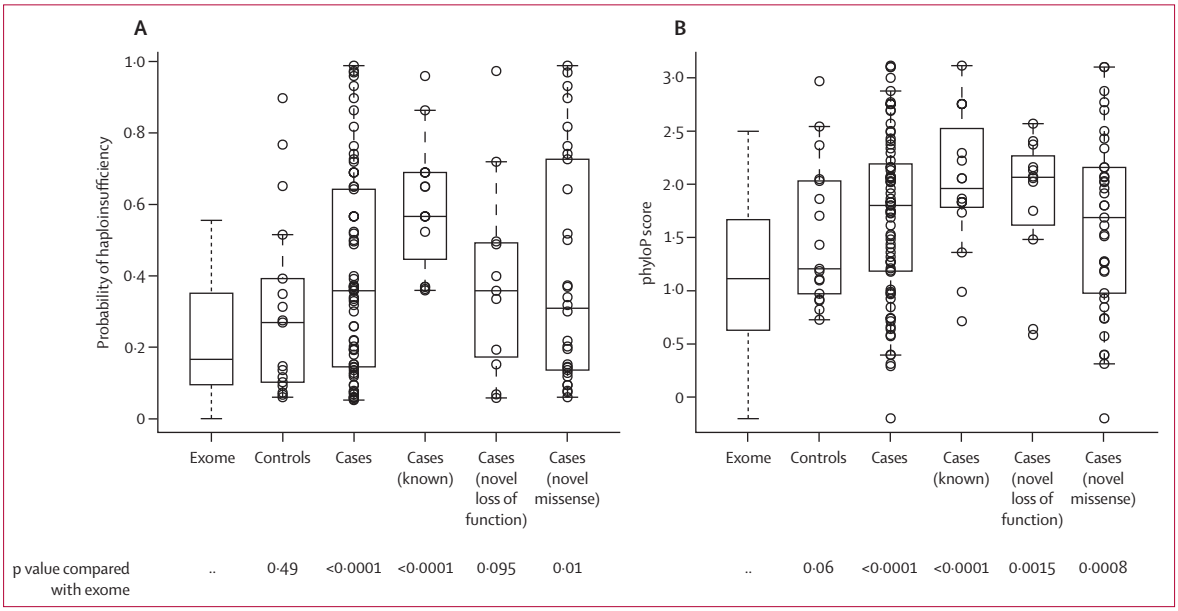

Figure 2: **Comparisons of probability of haploinsufficiency (A) and phyloP score (B) between cases, controls, and exome average**
The probabilities of haploinsufficiency (n=17 070) and the phyloP scores (n=19 592) of most genes in the exome were compared with the scores of genes carrying de-novo variants 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.

	Sex	Gene	Online Mendelian Inheritance in Man reference	Type	Genomic change	Protein change	Haploinsufficiency index (%)	PolyPhen2 category (score)
E10-0275	Male	*IQSEC2*	309530	Nonsense	X chromosome: g.53277315G→A	Arg855*1	9·8%	..
BO17/09	Female	*MECP2*	312750	Frameshift	X chromosome: g.153296093_153296115del	Pro401Argfs*8	24·3%	..
ZH58769	Male	*NAA10*	300855	Missense	X chromosome: g.153197564G→A	Arg116Trp	..	Benign (0·233)*
ER52725	Female	*SATB2*	608148	Missense	Chromosome 2: g.200213455A→C	Val381Gly	4·3%	Probably damaging (1)
ER8490	Male	*SCN2A*	613721	Frameshift	Chromosome 2: g.166179821_166179822delCT	Leu611Valfs*35	12·7%	..
MS111684	Male	*SCN2A*	613721	Frameshift	Chromosome 2: g.166172100_166172101insA	Asn503Lysfs*19	12·7%	..
ZH60991	Female	*SCN2A*	613721	Missense	Chromosome 2: g.166201311C→T	Arg937Cys	12·7%	Probably damaging (1)
ER12988	Female	*SCN8A*	614558, 614306	Missense	Chromosome 12: g.52200120G→A	Arg1617Gln	14·5%	Probably damaging (1)
BO22/10	Female	*SETBP1*	269150	Nonsense	Chromosome 18: g.42531079A→T	Lys592*	9·9%	..
PL111540	Male	*SLC2A1*	606777, 612126	Missense	Chromosome 1: g.43396356G→A	Arg153Cys	24·1%	Probably damaging (1)
ES07E0046	Female	*STXBP1*	612164	Missense	Chromosome 9: g.130422363G→C	Ala101Pro	8·7%	Possibly damaging (0·860)†
MR-NET001	Female	*STXBP1*	612164	Splice	Chromosome 9: g.130422308delC	Aberrant splicing predicted	8·7%	..
P4276	Female	*STXBP1*	612164	Missense	Chromosome 9: g.130420659G→A	Glu59Lys	8·7%	Probably damaging (0·994)
BO14/09	Female	*SYNGAP1*	612621	Frameshift	Chromosome 6: g.33410958_33410959insT	Thr878Aspfs*60	23·6%	..
ER53899	Male	*SYNGAP1*	612621	Frameshift	Chromosome 6: g.33405934_33405935delAA	Lys418Argfs*54	23·6%	..
TUBA080997	Female	*TCF4*	610954	Missense	Chromosome 18: g.53070725G→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 ϕ angle of –158° is not possible for proline, and molecular modelling suggests that this mutation destabilises the structure and probably also hampers ligand binding (appendix).

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).[12] Likewise, the average phyloP score[17] 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

	Sex	Gene	Genomic change (reference sequence hg19)	Mutation type*	Haploinsufficiency index (%)	Smallest region of overlap	Brain expression	Mouse phenotype	Description
ER10924	Male	*ARIH1*	Chromosome 15: g.72847668delT	Frameshift	24·4%	..	Ubiquitous	..	Ubiquitin-conjugating enzyme E2 binding protein
MS134	Female	*CHD2*	Chromosome 15: g.93498742delG	Frameshift	15·8%	Deletion 3 genes 472 kp	Ubiquitous	MGI:2448567; heterozygous gene trap mice have postnatal lethality, growth retardation, and multiorgan defects	The *CHD* gene family is characterised by chromatin organisation modifier and SNF2-related helicase and ATPase domains
BO63/11	Female	*HIVEP2*	Chromosome 6: g.143081688delC	Frameshift	16·2%	Deletion 24 genes	Very high	MGI:1338076; homozygous knock-out mice have abnormal thymus anatomy	A transcription factor that interacts with TCF4
ER14209	Female	*SETD5*	Chromosome 3: g.9490270C→T	Nonsense	21·3%	Deletion 7 genes 702 kp	Ubiquitous	..	Uncharacterised protein that contains a SET domain; SET7 is a histone H3 lysine methyltransferase
ZH50743	Female	*SLC6A1*	Chromosome 3: g.11060365delT	Frameshift	53·6%	Deletion 10 genes 1·5 Mp	Very high	MGI:95627; homozygous hypomorphic mice have abnormal inhibitory postsynaptic currents and abnormal γ-aminobutyric-acid uptake and release	γ-aminobutyric-acid transporter
NS0908	Female	*SYNCRIP*	Chromosome 6: g.86324828_86324828insG	Frameshift	1·5%	Duplication 5 genes 698 kp	Ubiquitous	..	Nuclear ribonucleoprotein, involved in mRNA processing
07E0967	Male	*CUX2*	Chromosome 12: g.111748354G→A	Missense (0·999)	40·0%	Deletion 54 genes 4·8 Mb	Very high	MGI:107321; homozygotes for a targeted null mutation have various neural defects	Belongs to the CUT homoeobox family, thought to be a transcription factor involved in neural specification
ER52808	Female	*DEAF1*	Chromosome 11: g.686871T→G	Missense (0·997)†	58·6%	No deletion	Very high	MGI:1858496; many homozygotes have exencephaly	Reported by Vissers and colleagues[8]
DD15852	Female	*EIF2C1*	Chromosome 1: g.36359331T→C	Missense (1·000)	84·5%	Deletion 22 genes 1·1 Mb	Ubiquitous	..	Reported by Sanders and colleagues[13]
TUTLN	Female	*KCNQ3*	Chromosome 8: g.133192493G→A	Missense (0·999)	7·3%	Deletion 14 genes 3·3 Mb	High	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:121201] and epileptic encephalopathy	M channel, a slowly activating and deactivating potassium channel that has a critical role in regulation of neuronal excitability
ER51232	Male	*STAG1*	Chromosome 3: g.136240090T→C	Missense (0·692)	6·8%	ISCA nssv577920: deletion 22 genes 3·9 Mb	High	..	Encodes a component of cohesion, functionally related to SMC1 and SMC3 causing Cornelia de Lange syndrome
MS047	Female	*ZNF238*	Chromosome 1: g.244218559C→G	Missense (0·972)	2·6%	Deletion 4 genes 1 Mb	Very high	MGI:1353609; homozygous knock-out mice have neonatal lethality, cortical and hippocampal hypoplasia, laminar disorganisation, and abnormal neuron apoptosis	C2H2-type zinc finger protein, possibly involved in chromatin assembly; within microdeletion region[19]

*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).

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

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*.[18] 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 *ARIH1*, *CDH2*, *HIVEP2*, *SETD5*, *SLC6A1*, and *SYNCRIP*. 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 *EIF2C1*).

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.[20] The Exome Variant Server dataset does not contain small deletions or insertions. With the exception of *SYNCRIP*, 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.[8] 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 (*FKBPL* and *DLG5*) 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.[13,14,21] 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[22] and an estimated proportion of 10–12% of X-linked recessive inheritance among boys with the disease.[3]

For the **Exome Variant Server** see http://evs.gs.washington.edu/EVS/

Locus heterogeneity of intellectual disability is high—as shown by targeted Sanger sequencing of specific genes[4,23]—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 *STXBP1* mutations in 3%,[4,23] 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 disability[8] and autism spectrum disorder.[13–15,24] Overlap occurred in genes known to be associated with intellectual disability (*SCN2A* and *SETBP1*) and in novel genes (*CHD2*, *CTTNBP2*, *DEAF1*, *EIF2C1*, *GPRASP1*, *SETD5*, and *SLC6A1*). 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 *SLC6A1* 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.

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.[9,11,16,33] Previously published studies of small numbers of patients[8] or of candidate genes[4] 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.

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 reported to cause early-onset infantile epileptic encephalopathy,[25] 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).[26] *SETBP1* missense mutations clustering in an 11 bp region can cause Schinzel-Giedion syndrome, probably by a dominant negative or gain-of-function mechanism.[27] 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[28]) and missense mutations in *NAA10* (reported to cause Ogden syndrome[29]), *TCF4* (reported to cause Pitt-Hopkins syndrome[30]), and *SCN8A* (reported to cause intellectual disability with pancerebellar atrophy and ataxia[31]). In our study, the patient with the *SATB2* mutation had no cleft palate, but did have bifid uvula. The children in our study did not have cardiac anomalies associated with Ogden syndrome, the typical facial features of Pitt-Hopkins syndrome, or the ataxia described in children with *SCN8A* mutations.

Mutation rates are higher in coding sequences than in non-coding sequences.[10,32] This difference can be explained by a higher content of CpG sites—which have higher mutation rates—in coding sequences. We estimated that the point mutation rate in coding regions is 29% higher than in the genome (appendix). Using this proportion and a previously published estimate of the genome-wide mutation rate of $1{\cdot}18\times10^{-8}$, we expect an exomic mutation rate of $1{\cdot}52\times10^{-8}$, very close to that in control participants and in good agreement with previous estimates from exome sequencing in autism spectrum disorders (panel).[13–15] Additionally, the 2·14 missense variants per synonymous variant in controls accords with previous studies.[34] We estimate that our measurements of mutation rate are conservative for three reasons: first, the target region of 38×10^{6} bp includes non-unique sequences and duplicated regions where single nucleotide variant calling is hindered; second, we could have missed a small proportion of de-novo variants in low coverage regions; and third, some mutations might be lethal.

The genes *ARIH1*, *CDH2*, *HIVEP2*, *SETD5*, *SLC6A1*, and *SYNCRIP* had loss-of-function mutations that probably cause disease. These genes are predicted to be

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*).[35] 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 *NR1I3* (*EcR*), a gene mutated in patients with Kleefstra syndrome spectrum disorders.[36,37] 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.[38]

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.[39] 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 homoeobox transcription factor *CUX2*, which regulates dendrite branching, spine development, and synapse formation in layer 2–3 neurons of the cerebral cortex,[40] *ZNF238*, which is implicated in chromatin assembly and is a candidate gene for intellectual disability with corpus callosum hypogenesis,[19] 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[8] or autism spectrum disorder.[13] 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.[33]

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 ARe 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, ARe, 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.

Acknowledgments

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, GEUVADIS), and the Swiss National Science Foundation (320030_135669).

References

1 Desk Reference to the Diagnostic Criteria From DSM-IV-TR. Washington: American Psychiatric Association, 2000.
2 Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* 2002; **8:** 117–34.
3 Ropers HH. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* 2010; **11:** 161–87.
4 Hamdan FF, Gauthier J, Spiegelman D, et al. Mutations in *SYNGAP1* in autosomal nonsyndromic mental retardation. *N Engl J Med* 2009; **360:** 599–605.
5 Najmabadi H, Hu H, Garshasbi M, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011; **478:** 57–63.
6 Rauch A, Hoyer J, Guth S, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A* 2006; **140:** 2063–74.
7 Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. *Nat Genet* 2011; **43:** 838–46.
8 Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. *Nat Genet* 2010; **42:** 1109–12.
9 Conrad DF, Keebler JE, DePristo MA, et al. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 2011; **43:** 712–14.
10 Awadalla P, Gauthier J, Myers RA, et al. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet* 2010; **87:** 316–24.
11 Li Y, Vinckenbosch N, Tian G, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 2010; **42:** 969–72.
12 Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010; **6:** e1001154.
13 Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485:** 237–41.
14 Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485:** 242–45.
15 O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485:** 246–50.
16 Roach JC, Glusman G, Smit AF, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010; **328:** 636–39.
17 Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010; **20:** 110–21.
18 Buschdorf JP, Stratling WH. A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. *J Mol Med (Berl)* 2004; **82:** 135–43.

19 Ballif BC, Rosenfeld JA, Traylor R, et al. High-resolution array CGH defines critical regions and candidate genes for microcephaly, abnormalities of the corpus callosum, and seizure phenotypes in patients with microdeletions of 1q43q44. *Hum Genet* 2012; **131:** 145–56.

20 MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; **335:** 823–28.

21 O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011; **43:** 585–89.

22 Van Naarden Braun K, Autry A, Boyle C. A population-based study of the recurrence of developmental disabilities—Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1991–94. *Paediatr Perinat Epidemiol* 2005; **19:** 69–79.

23 Hamdan FF, Piton A, Gauthier J, et al. De novo STXBP1 mutations in mental retardation and nonsyndromic epilepsy. *Ann Neurol* 2009; **65:** 748–53.

24 Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74:** 285–99.

25 Kamiya K, Kaneda M, Sugawara T, et al. A nonsense mutation of the sodium channel gene SCN2A in a patient with intractable epilepsy and mental decline. *J Neurosci* 2004; **24:** 2690–98.

26 Buysse K, Menten B, Oostra A, Tavernier S, Mortier GR, Speleman F. Delineation of a critical region on chromosome 18 for the del(18)(q12.2q21.1) syndrome. *Am J Med Genet A* 2008; **146A:** 1330–34.

27 Hoischen A, van Bon BW, Gilissen C, et al. De novo mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat Genet* 2010; **42:** 483–85.

28 Leoyklang P, Suphapeetiporn K, Siriwan P, et al. Heterozygous nonsense mutation *SATB2* associated with cleft palate, osteoporosis, and cognitive defects. *Hum Mutat* 2007; **28:** 732–38.

29 Rope AF, Wang K, Evjenth R, et al. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet* 2011; **89:** 28–43.

30 Zweier C, Sticht H, Bijlsma EK, et al. Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. *J Med Genet* 2008; **45:** 738–44.

31 Trudeau MM, Dalton JC, Day JW, Ranum LP, Meisler MH. Heterozygosity for a protein truncation mutation of sodium channel SCN8A in a patient with cerebellar atrophy, ataxia, and mental retardation. *J Med Genet* 2006; **43:** 527–30.

32 Subramanian S, Kumar S. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res* 2003; **13:** 838–44.

33 Lugli G, Larson J, Martone ME, Jones Y, Smalheiser NR. Dicer and eIF2c are enriched at postsynaptic densities in adult mouse brain and are modified by neuronal activity in a calpain-dependent manner. *J Neurochem* 2005; **94:** 896–905.

34 Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 2007; **80:** 727–39.

35 Duning K, Buck F, Barnekow A, Kremerskothen J. SYNCRIP, a component of dendritically localized mRNPs, binds to the translation regulator BC200 RNA. *J Neurochem* 2008; **105:** 351–59.

36 Gradilla AC, Mansilla A, Ferrus A. Isoform-specific regulation of a steroid hormone nuclear receptor by an E3 ubiquitin ligase in *Drosophila melanogaster*. *Genetics* 2011; **189:** 871–83.

37 Kleefstra T, Kramer JM, Neveling K, et al. Disruption of a novel EHMT1-associated chromatin modification module causes intellectual disability. *Am J Hum Genet* 2012; **91:** 73–82.

38 Dorflinger U, Pscherer A, Moser M, Rummele P, Schule R, Buettner R. Activation of somatostatin receptor II expression by transcription factors MIBP1 and SEF-2 in the murine brain. *Mol Cell Biol* 1999; **19:** 3736–47.

39 Weckhuysen S, Mandelstam S, Suls A, et al. KCNQ2 encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. *Ann Neurol* 2012; **71:** 15–25.

40 Cubelos B, Sebastian-Serrano A, Beccari L, et al. Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron* 2010; **66:** 523–35.