

ARTICLE

Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4)

Lisa J Strug^{1,2}, Tara Clarke³, Theodore Chiang¹, Minchen Chien^{4,5}, Zeynep Baskurt¹, Weili Li¹, Ruslan Dorfman⁶, Bhavna Bali³, Elaine Wirrell⁷, Steven L Kugler⁸, David E Mandelbaum⁹, Steven M Wolf¹⁰, Patricia McGoldrick¹⁰, Huntley Hardison¹¹, Edward J Novotny¹², Jingyue Ju^{4,5}, David A Greenberg^{13,14}, James J Russo^{4,5} and Deb K Pal^{*,3,13,15}

¹Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; ²Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; ³Department of Epidemiology, Mailman School of Public Health, Columbia University Medical Center, New York, NY, USA; ⁴The Judith P. Sulzberger MD Columbia Genome Center, New York, NY, USA; ⁵Department of Chemical Engineering, Columbia University Medical Center, New York, NY, USA; ⁶Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; ⁷Division of Pediatric Neurology, Alberta Children's Hospital, University of Calgary, Calgary, Alberta, Canada; ⁸Division of Neurology, Children's Hospital of Philadelphia and University of Pennsylvania School of Medicine, Philadelphia, PA, USA; ⁹Department of Clinical Neurosciences, Hasbro Children's Hospital and the Warren Alpert Medical School of Brown University, Providence, RI, USA; ¹⁰Epilepsy Monitoring Unit, Beth Israel Medical Center, New York, NY, USA; ¹¹Division of Pediatric Neurology, St Christopher's Hospital for Children, Philadelphia, PA, USA; ¹²Department of Pediatrics, Yale University Medical Center, New Haven, CT, USA; ¹³Division of Statistical Genetics, Mailman School of Public Health, Columbia University Medical Center, New York, NY, USA; ¹⁴New York State Psychiatric Institute, New York, NY, USA; ¹⁵Department of Psychiatry, Columbia University Medical Center, New York, NY, USA

Rolandic epilepsy (RE) is the most common human epilepsy, affecting children between 3 and 12 years of age, boys more often than girls (3:2). Focal sharp waves in the centrotemporal area define the electroencephalographic (EEG) trait for the syndrome, are a feature of several related childhood epilepsies and are frequently observed in common developmental disorders (eg, speech dyspraxia, attention deficit hyperactivity disorder and developmental coordination disorder). Here we report the first genome-wide linkage scan in RE for the EEG trait, centrotemporal sharp waves (CTS), with genome-wide linkage of CTS to 11p13 (HLOD 4.30). Pure likelihood statistical analysis refined our linkage peak by fine mapping CTS to variants in Elongator Protein Complex 4 (*ELP4*) in two independent data sets; the strongest evidence was with rs986527 in intron 9 of *ELP4*, providing a likelihood ratio of 629:1 ($P=0.0002$) in favor of an association. Resequencing of *ELP4* coding, flanking and promoter regions revealed no significant exonic polymorphisms. This is the first report of a gene implicated in a common focal epilepsy and the first human disease association of *ELP4*. *ELP4* is a component of the Elongator complex, involved in transcription and tRNA modification. Elongator depletion results in the brain-specific downregulation of genes implicated in cell motility and migration. We hypothesize that a non-coding mutation in *ELP4* impairs brain-specific Elongator-mediated interaction of genes implicated in brain development, resulting in susceptibility to seizures and neurodevelopmental disorders.

European Journal of Human Genetics (2009) 17, 1171–1181; doi:10.1038/ejhg.2008.267; published online 28 January 2009

Keywords: linkage; neurodevelopmental traits; centrotemporal spikes; idiopathic partial epilepsy; association

*Correspondence: Dr DK Pal, Columbia University Medical Center, Mailman School of Public Health, 722 West 168th Street, 6th Floor, New York, NY 10032, USA. Tel: + 1 212 342 1237; Fax: + 1 212 342 8483; E-mail: dkp28@columbia.edu
Received 20 October 2008; revised 9 December 2008; accepted 10 December 2008; published online 28 January 2009

Introduction

Rolandic epilepsy (RE) (MIM 117100) is a neurodevelopmental disorder, affecting 0.2% of the population. It is characterized by classic focal seizures that recapitulate the functional anatomy of the vocal tract, beginning with guttural sounds at the larynx, sensorimotor symptoms then progressing up to the tongue, mouth and face, culminating with speech arrest. Seizures most often occur in sleep shortly before awakening. The disorder occurs more often in boys than in girls (3:2) and is diagnosed in one in five of all children with newly diagnosed epilepsy.¹ All patients exhibit the defining electroencephalographic (EEG) abnormality of centrotemporal sharp waves (CTS). The onset of seizures in childhood (3–12 years)² is frequently preceded by a constellation of developmental deficits including speech disorder, reading disability and attention impairment. These deficits have been noted to cluster in family members of RE patients who do not have epilepsy.^{3,4} None of these abnormalities are associated with major cerebral malformations visible on routine MRI.⁵ The seizures and the EEG abnormality of CTS spontaneously remit at adolescence, although the prognosis for developmental deficits is less clear. There is no known involvement of organs outside the nervous system.

RE belongs to a family of idiopathic epilepsies of childhood with focal sharp waves, some of which are characterized by more severe and varied types of seizures (atypical benign partial epilepsy or ABPE, MIM 604827), variable locations (benign occipital epilepsy, MIM 132090), acquired receptive aphasia (Landau–Kleffner syndrome, MIM 245570) and developmental regression (continuous spikes in slow-wave sleep). CTS are common in children (2–4%),⁶ have equal gender distribution and have been observed with increased frequency in developmental disorders, including speech dyspraxia,⁷ attention deficit hyperactivity disorder (ADHD)⁸ and developmental coordination disorder (DCD),⁹ suggesting that the EEG trait of CTS is not specific to epilepsy, but possibly a marker for an underlying subtle but more widespread abnormality of neurodevelopment.¹⁰

Despite the strong clustering of developmental disorders in RE families, RE itself has a low sibling risk of ~10%.¹¹ Several rare, phenotypically distinct Mendelian RE variants have been reported^{12–15} but the common form appears to have complex genetic inheritance. However, segregation analysis suggests that CTS in the common form of RE is inherited as an autosomal dominant trait.¹⁶ CTS were reported to link to 15q14 in a candidate gene study of families multiplex for RE and ABPE, but this locus has not been replicated and no genome-wide screen for CTS has been previously attempted.¹⁷ Understanding the mechanism of CTS could provide insight into the variety of common neurodevelopmental disorders in which CTS are observed. We therefore set out to genetically map the CTS trait in RE families.

We conducted genome-wide linkage analysis of the CTS trait in 38 US families singly ascertained through an RE proband. In 11 of the families, one additional sibling was known to carry the CTS trait, but the CTS status of individuals younger than 4 years or older than 16 years was unknown because of its age-limited expression. The maximum two-point and multipoint LOD scores for CTS were observed at 11p13. We designated a 13-cM linkage region, encompassing the area in which LOD scores >2.0, as our region of interest for fine mapping. We then tested for association of CTS with SNP markers distributed across genes in this region. We initially used a 'discovery' data set that included 68 cases and 187 controls group matched for ancestry and gender – 38 of these cases were included in the original linkage screen. In addition to case–control analysis, we used family-based analysis to guard against the potential for positive confounding due to population stratification. We took a pure likelihood approach to the statistical analysis of linkage and association^{18–21} that is explained in Methods. We then typed additional SNPs around genes that showed compelling evidence of association in the preliminary analysis. In a second, independent, 'replication' case–control data set, we typed a subset of the SNPs in our region of interest. The replication set included 40 RE cases and 120 controls from Western Canada; the two data sets were then jointly analyzed.

Methods

Subjects

Informed consent was obtained from all participants using procedures approved by institutional review boards at each of the clinical research centers collecting human subjects. The general methodology for the study has been detailed elsewhere.³ Briefly, cases with classic RE and their families were recruited for a genetic study from eight pediatric neurology centers in the northeastern United States (see Acknowledgements for referring physicians). Ascertainment was through the proband, with no other family member required to be affected with RE. All cases were centrally evaluated by a pediatric neurologist, as well as by one other study physician. Cases were enrolled if they met stringent eligibility criteria for RE, in accordance with the definition of the International League Against Epilepsy²² including (i) at least one witnessed seizure with typical features: nocturnal, simple partial seizures affecting one side of the body or on alternate sides; (ii) oro-facial-pharyngeal sensorimotor symptoms, with speech arrest and hypersalivation; (iii) age of onset between 3 and 12 years; (iv) no previous epilepsy type; (v) normal global developmental milestones; (vi) normal neurological examination; (vii) at least one interictal EEG with CTS and normal background, verified by two independent and blinded readers;¹⁶ and (viii) neuroimaging read by two independent and blinded board-certified neuroradiologists

that excluded an alternative structural, inflammatory or metabolic cause for the seizures.⁵ Thus, cases with unwitnessed episodes or with only secondary generalized seizures were excluded, even if the EEG was typical. Siblings between the ages of 4 and 15 years underwent sleep-deprived EEGs to assess their CTS status;¹⁶ EEGs were then evaluated blind to identity by two independent experts.

Cases had their first seizure at a median age of 8 years (range 3–12); most had less than 10 lifetime seizures; over a third had at least one secondary generalized seizure, but only two had a history of convulsive status epilepticus; and two-thirds had been treated with antiepileptic drugs. Table 1 shows the seizure characteristics of the cases. Cases were 60% male and 76% of European ancestry (see Supplementary Information, Supplementary Table 1). Details of EEG and imaging findings have been reported earlier.^{5,16} Affectedness data and DNA were collected from all potentially informative and consenting relatives of the proband. In most cases, this included at least both parents and all siblings over the age of 3 years.

One hundred and eighty-seven controls were recruited from the same geographic locations as the cases and were group matched for gender and ancestry (see Supplementary Information, Supplementary Table 1). Each potential control was screened for personal and family history of neuropsychiatric and developmental disorders: DNA from individuals with a history of seizures was excluded from the control panel. The lifetime CTS status of controls was

unknown because of their developmental expression, but assumed to be representative of the general population⁶ that is 2–4%; thus any observed association in case–control analysis should be conservative. This sample of independent cases and controls is referred to throughout as the discovery data set.

Forty cases and 120 controls were recruited from Calgary, Canada, according to the same eligibility criteria as in the discovery data set (see Supplementary Information, Supplementary Table 1), as a replication sample. The cases were 56% male and 83% of European ancestry, with median age of seizure onset at 7 years. Controls were also 56% male and 86% of European ancestry. Information regarding personal and family history of neuropsychiatric and developmental disorders was collected as above for possible exclusion from case–control analysis. The Calgary sample is referred to as the replication data set.

DNA collection

DNA was collected either by peripheral venous blood draw into 10 ml K-EDTA tubes (Fisher Scientific) or by salivary sample in ORAGENE (DNA Genotek, Ottawa) flasks.

STR genotyping

One hundred and ninety-four individuals from 38 RE families were genotyped using the deCODE 4cM STR marker panel. This panel contains approximately 1200 highly polymorphic STR markers. Amplified fragments were electrophoresed using ABI 3700 and ABI 3730 DNA analyzers with CEPH family DNA used as control. Alleles were checked for consistency with Hardy–Weinberg equilibrium and non-paternity. Errors were reconciled by resampling or by excluding genotypes.

Linkage analyses

Two-point and multipoint heterogeneity LOD scores were calculated in all 38 families combined. We used the MMLS approach to parametric linkage analysis,²³ which recommends a robust set of parameter values and an analytical approach.^{24,25} Briefly, one calculates LOD scores under both dominant and recessive modes of inheritance, specifying a dominant gene frequency of 0.01 and a recessive gene frequency of 0.14, a sporadic rate of 0.0002 and a penetrance of 0.50. In regions providing evidence for linkage, we then maximized over a grid of penetrance values from 0 to 1.0 by 0.05 increments. Marker allele frequencies were calculated from the data set. We followed up markers with two-point LOD scores greater than 2.0 with multipoint analysis using Genehunter,²⁶ again using MMLS but maximizing over penetrance and computing heterogeneity LOD scores. We used a sex-averaged map because the observed multipoint LOD scores should be conservative in the presence of linkage, if indeed there are male–female map differences.²⁷ Simulation results confirmed that differential male–female map distance has

Table 1 Clinical descriptors of RE cases

Feature	Discovery set (%)	Replication set (%)
Febrile seizures	4	17
Right handedness	86	93
<i>Usual laterality of seizure</i>		
Left	36	22
Right	32	17
Inconsistent	32	61
<i>Predominant EEG lateralization</i>		
Left	29	20
Right	53	41
Bilateral	11	32
<i>Lifetime seizure total</i>		
≤10	70	59
>10	30	41
<i>Relation of seizures to sleep or drowsiness</i>		
Exclusive	89	83
Not exclusive	11	17
Ever treated with antiepileptic drugs	70	66
Developmental speech delay	38	20
Reading disability	52	34
Migraine headaches	20	20

little effect on localization of the maximum LOD score (data not shown). Separate analyses were conducted in the European and non-European ancestral subgroups.

SNP markers

We first typed polymorphic SNP markers in the 11p13 linkage region, delimited by a LOD score of 1.0 on either side of the multipoint linkage peak. Thirty-six markers were distributed predominantly within known genes using Tagger, implemented in Haploview,²⁸ with $r^2=0.8$; eight additional SNPs were typed in the region of *ELP4* and *PAX6* where there was evidence of association. The 44 SNPs (Table 2) were placed in and between ESTs and genes annotated in Ensembl Release 46, from downstream to

upstream (see Figure 2): *DCDC5*, *DCDC1*, *DPH4*, *IMMP1L*, *ELP4* and *PAX6* between 30 819 214 and 31 780 205 bp (NCBI Build 36). In the replication data set, we typed a subset of 30 SNPs spanning 31 252 249–31 772 472 bp.

SNP genotyping

Genotyping was performed on the Nanogen platform at deCODE Genetics (Iceland). SNPs were analyzed by end-point scatter plot analysis utilizing the ABI 799HT Sequence Detection System. Sixty-eight cases, parents of 38 cases and 118 controls were successfully typed from the discovery set; all 38 cases and 138 controls were typed from the replication set. Only one SNP, rs10835810, had >5% missingness (30% missing rate, similar in cases and

Table 2 SNPs genotyped in this study

Marker number	dbSNP number	Alleles minor/major	Physical map location (bp)	bp to next marker	MAF in controls	Gene/type
1	rs1015541	A/C	30 811 481	0	0.323077	DCDC5 intron 31
2	rs1448938	T/C	30 849 400	37 919	0.444882	DCDC5 intron 28
3	rs273573	A/C	30 867 567	18 167	0.326923	DCDC5 intron 26
4	rs395032	A/G	30 883 776	16 209	0.324427	DCDC5 intron 20
5	rs163881	T/G	30 904 820	21 044	0.326923	DCDC5 intron 12
6	rs7117074	A/C	30 942 663	37 843	0.432	DCDC5 intron 10
7	rs290102	C/T	30 972 073	29 410	0.453125	DCDC5 intron 10
8	rs288458	G/C	31 007 585	35 512	0.096	DCDC5 intron 10
9	rs560395	G/A	31 044 369	36 784	0.454545	DCDC5 intron 8
10	rs621549	A/C	31 070 773	26 404	0.461538	DCDC5 intron 6
11	rs208068	G/A	31 108 520	37 747	0.392308	*
12	rs400964	A/T	31 133 836	25 316	0.386719	*
13	rs16921914	A/G	31 167 347	33 511	0.305344	*
14	rs286651	T/C	31 186 380	19 033	0.39313	*
15	rs7937421	C/T	31 252 249	65 869	0.205426	DCDC1 intron 7
16	rs2774403	A/T	31 277 106	24 857	0.392308	DCDC1 intron 6
17	rs12577026	A/G	31 304 419	27 313	0.169231	DCDC1 intron 3
18	rs1547131	C/T	31 343 175	38 756	0.257692	DCDC1 intron 1
19	rs483534	G/C	31 354 718	11 543	0.350806	DPH4 intron 2
20	rs578666	G/A	31 361 060	6342	0.383721	DPH4 intron 2
21	rs6484503	G/T	31 381 179	20 119	0.32	DPH4 intron 2
22	rs1223118	G/C	31 427 076	45 897	0.00384615	IMMP1L intron 5
23	rs1223068	G/T	31 436 925	9849	0.25	IMMP1L intron 4
24	rs1223098	T/G	31 463 483	26 558	0.472222	IMMP1L intron 1
25	rs509628	C/T	31 491 931	28 448	0.480159	ELP4 intron 1
26	rs502794	C/A	31 503 803	11 872	0.484375	ELP4 intron 2
27	rs2996470	T/C	31 516 234	12 431	0.247826	ELP4 intron 2
28	rs2973127	C/T	31 519 594	3360	0.251908	ELP4 intron 3
29	rs2104246	G/A	31 530 222	10 628	0.246094	ELP4 intron 3
30	rs2996464	C/T	31 545 775	15 553	0.265385	ELP4 intron 3
31	rs2146569	G/T	31 565 684	19 909	0.244186	ELP4 intron 3
32	rs10835793	T/A	31 575 426	9742	0.25	ELP4 intron 4
33	rs1232182	A/T	31 589 144	13 718	0.267176	ELP4 intron 5
34	rs986527	T/C	31 593 057	3913	0.425197	ELP4 intron 5
35	rs11031434	A/G	31 609 788	16 731	0.492308	ELP4 intron 6
36	rs1232203	A/C	31 622 784	12 996	0.25	ELP4 intron 7
37	rs964112	T/G	31 635 524	12 740	0.414634	ELP4 intron 9
38	rs2862801	A/G	31 652 912	17 388	0.248062	ELP4 intron 9
39	rs10835810	T/C	31 679 060	26 148	0.425532	ELP4 intron 9
40	rs12365798	C/T	31 704 334	25 274	0.244275	ELP4 intron 9
41	rs2863231	A/G	31 753 136	48 802	0.380952	ELP4 intron 9
42	rs3026411	A/T	31 758 120	4984	0.334615	ELP4 intron 9
43	rs1506	A/T	31 766 874	8754	0.223077	PAX6 3'
44	rs2239789	A/T	31 772 472	5598	0.480769	PAX6 intron 8

Asterisk (*) designates intergenic SNP.

Table 3 Single-SNP association results: pure likelihood and frequentist analyses at SNPs of interest in *ELP4*; *P*-values are unadjusted

SNP	Risk allele	Discovery analysis				Joint analysis			
		OR	1/32 LI	Max LR	P-value	OR	1/32 LI	Max LR	P-value
rs964112	G	2.04	1.15–3.80	156.95	0.0008	1.88	1.18–3.06	589.75	0.0002
rs11031434	G	1.80	1.05–3.16	57.94	0.0035	1.71	1.10–2.70	150.57	0.0013
rs986527	C	1.98	1.12–3.66	108.97	0.0013	1.88	1.18–3.06	628.85	0.0002

controls), and only rs2863231 was out of Hardy–Weinberg equilibrium in controls at the 0.001 level. All except two SNPs (rs1223118 and rs288458) had a minor allele frequency >0.15 (Table 2). For resequencing methods see Supplementary Information.

Association analysis

Pure likelihood vs frequentist analysis We conducted a pure likelihood analysis of the SNP data^{18,19} as well as calculating standard frequentist *P*-values for comparison. The two methods, in theory, provide the same ordering of importance for SNPs. However, they have different significance thresholds, different sample size requirements and different approaches to the adjustment for multiple hypothesis testing. We used pure likelihood analysis to determine our SNPs of interest for follow-up. We provide *P*-values for those unfamiliar with pure likelihood analysis, for comparison only. A pure likelihood display of the data provides a more visually informative understanding than standard plots of kb by $-\log_{10}$ (*P*-value). Moreover, a pure likelihood analysis is particularly well suited for joint analysis of multistage designs,²⁹ largely due to how pure likelihood analyses adjust for type I error inflation due to multiple hypothesis testing. Adjustments for multiple SNP tests are accomplished by following up signals from the first stage with additional samples analyzed in a joint analysis.²⁹ This is in contrast to standard *P*-value analysis approaches that require adjustment of *P*-values, for example Bonferroni, FDR.³⁰ For a discussion on pure likelihood multiple test adjustments, see Supplementary Information and reference 29.

Frequentist methods used We calculated a Cochran–Armitage test for trend in the case–control sample. We used a transmission disequilibrium test as implemented in FBAT³¹ in the subset of trios to ensure that any signal we found through case–control analysis was not itself due to population stratification. For multilocus analysis, we used multiple logistic regression of main effects and two-way interactions, coding the genotypes as -1 , 0 and 1 , with interaction being the product of the genotypes.

Pure likelihood methods used

In a pure likelihood analysis we report observed likelihood ratios (LRs) as well as provide figures of likelihood intervals

(LIs) for the odds ratio (OR), by base-pair position. For example, a 1/32 LI is defined as the set of OR values where the standardized likelihood function (divided by the likelihood evaluated at the maximum likelihood estimator) is greater than 1/32.¹⁸ LIs are analogous to confidence intervals in that they are comprised of all parameter values that are supported by the data. However, LIs do not require a long-run frequency interpretation, rather they reflect the evidence about the OR provided by the given data set. The pure likelihood analysis implemented assumes an additive disease model. We used profile likelihoods³² to construct the LRs and assess association at each SNP. We used LOD evidence of strength 1.5 as a criterion from the observed LRs to define an SNP of interest.

Results

CTS links to markers at 11p13

Only markers on chromosome 11 yielded two-point LOD scores exceeding 3.0. Markers in the region of chromosomal band 11p13 provided strong and compelling evidence for linkage to CTS. Marker D11S4102 yielded a two-point LOD score of 4.01, and seven other markers in the immediate region also exhibited LOD scores exceeding 2. Both European and non-European ancestry families contributed proportionally to the LOD score. The markers on chromosome 11 generally maximized at unequal male–female recombination fractions, because the male–female recombination map differs substantially in this region. For example, at D11S4102, the recombination rate for females is 1.70 cM/MB, whereas for males it is 0.48 cM/MB. Two-point LOD score maximization in this region of 11p most often occurred at 95% penetrance. Although single markers on chromosomes 5, 9, 10, 12 and 16 provided two-point LOD scores >2.0, the flanking marker information was not generally compelling. We did not observe significant evidence of linkage at markers previously reported for CTS at 15q14¹⁷ (D15S165 – maximum LOD score 0.1381), nor for a rare recessive variant of RE at 16p12–11.2¹³ (D16S3068 – maximum LOD score 0.2959), nor for X-linked rolandic seizures and cognitive deficit (MIM 300643)¹⁴ (DXS8020 – maximum LOD score 0.39). Similarly, we did not find evidence of linkage to 11p13 in an autosomal dominant variant of RE with speech dyspraxia and cognitive impairment.¹⁵

Figure 1 shows the heterogeneity ('HLOD') and homogeneity ('LOD') linkage results observed in the multipoint analysis of chromosome 11, for a dominant mode of inheritance with 50% penetrance. This analysis model resulted in the highest multipoint LOD scores: 4.30 at marker D11S914 (7.4 cM from the two-point maximum). There was no suggestion of heterogeneity ($\hat{\alpha}=1$) in the region of linkage. The region bounded by LOD scores >2.0 spans from 43.17 to 56.88 cM, with D11S914 located at 46.7 cM,³³ and includes the following annotated genes: *DCDC5*, *DCDC1*, *DPH4*, *IMMP1L*, *ELP4* and *PAX6*.

Association of CTS with SNPs in *ELP4*

We typed a total of 44 SNPs across the linkage region in 68 cases and 187 controls (discovery set). Here, we conducted a pure likelihood analysis as well as computing standard Cochran–Armitage trend test *P*-values for comparison. The pure likelihood analysis is particularly well suited to a joint analysis of discovery and replication samples,²⁹ and has been noted to be particularly appropriate for genetic data.^{20,34} The pure likelihood analysis plots OR on the *y* axis vs base-pair position on the *x* axis. Evidence for association at a given SNP is determined by calculating the LR; whether a calculated LR provides strong association evidence is interpreted by LOD score benchmarks: for example, a LOD >1.5 (equivalent to an LR >32) is interpreted as reasonably strong association evidence. We found no evidence of association with SNPs in *DCDC5*, *DCDC1*, *DPH4*, *IMMP1L* or *PAX6* as indicated by Figure 2 and with gray LIs on Figure 3a. The two long gray lines extending off the plot indicate lack of information, mainly due to a low minor allele frequency. However, we did find significant evidence of association with SNPs in *ELP4* with both the Cochran–Armitage trend test and the pure likelihood analysis. Most notably, the SNPs of interest identified by

the likelihood analysis were rs964112 in intron 9 ($P=0.0008$, significant after Bonferroni correction), rs11031434 in intron 6 ($P=0.003$) and rs986527 ($P=0.001$) in intron 5 (see Figures 3a and b, Table 3 for summary statistics) with estimated ORs 1.80–2.04 at these markers. We ensured that all SNPs that had an $r^2 > 0.8$ with rs964112 were genotyped, but none were identified as functionally significant. In the family-based *P*-value analysis using FBAT, only SNPs in *ELP4* provided evidence of association, with the smallest *P*-values observed at rs986527 ($P=0.06$) and rs1232182 ($P=0.04$) with 27 and 28 informative families, respectively. These results argue against population stratification as a positive confounder for the observed *ELP4* association.

In the pure likelihood joint analysis of discovery and replication samples, the replication sample confirms that SNPs in *ELP4* are highly associated with CTS (Figure 3b). Here the association evidence for all three SNPs of interest from the discovery set has increased after combination with the replication data set. The maximum LR at rs964112 is now 589.75 (formerly 156.95 in the discovery set), which is evidence equivalent to observing an LOD score of 2.77, and at rs986527 the maximum LR = 628.85 (LOD equivalent of 2.80). The estimated ORs represent a twofold increase in risk of CTS. The ORs, 1/32 LIs, maximum LRs and trend test-unadjusted *P*-values from the discovery and joint analyses are displayed in Table 3. It should be noted that, when analyzed on its own in a standard *P*-value analysis, the replication sample provided strong evidence of association in *ELP4*, with rs2104246 significant after Bonferroni correction (unadjusted $P=0.0006$). We have reported analysis of combined ancestry data, although the results are qualitatively similar when restricted to European ancestry data. The substantial increase in maximum LR from joint analysis of the two data sets provides

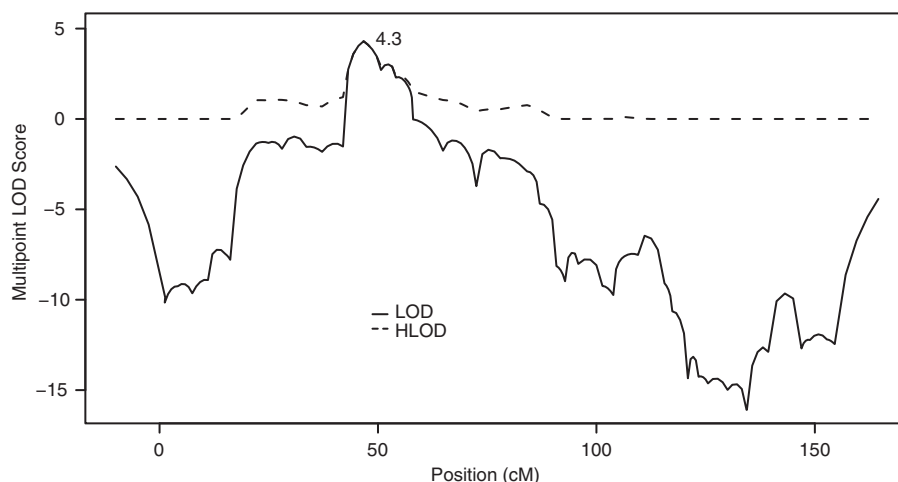


Figure 1 Multipoint LOD and heterogeneity LOD (HLOD) scores for CTS on chromosome 11: maximum HLOD = 4.3 at D11S914 under a dominant mode of inheritance with 50% penetrance.

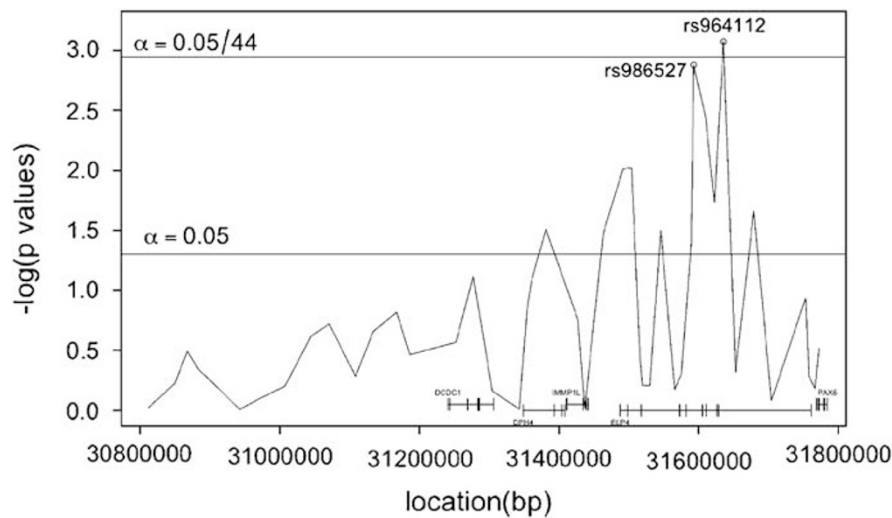


Figure 2 Cochrane–Armitage trend test of case–control association for CTS at the 11p13 locus in the discovery (New York) data set: Bonferroni critical value line displayed; significance criteria of 0.05/44 in discovery set corresponding to the 44 SNPs evaluated in the analysis.

compelling evidence that the *ELP4* variants, specifically rs986527 and rs964112, are indeed associated with CTS in RE families.

Multi-SNP analysis We used multiple logistic regression for multi-SNP analysis.³⁵ The SNPs of interest are in high LD with each other, which indicates that it is less likely we are detecting multiple independent variants in the region of *ELP4*. Multiple logistic regression analysis indicated that rs964112 was the best predictor of CTS, with no other SNP main effects or two-way interactions significant in the model; in the absence of rs964112, rs986527 played a similar predictive role. These SNPs were almost completely correlated.

Resequencing coding regions of *ELP4*

We resequenced the coding portions, exon–intron boundaries and 5' upstream region of the *ELP4* gene in 40 RE probands from the discovery set. The 274 kb *ELP4* gene is transcribed into a 1584 bp mRNA consisting of 12 exons, a 35 bp 5'-UTR and a 257 bp 3'-UTR. Alternative transcripts have been reported that include or exclude the last two exons. Primers were designed for direct sequencing of each of these 12 exons including some adjacent intronic sequence, as well as the putative promoter region; a list of these primers is included in Supplementary Table 2. The same primers were used for PCR and sequencing reactions. After alignment, all homozygous and heterozygous variants within the sequenced region were noted.

Three previously reported SNP variants were found in these 40 individuals: rs2295748 in the vicinity of the promoter; rs2273943 within intron 5 located 127 bases upstream of exon 6; and rs10767903, located within

exon 10. The genotypes and allele frequencies for these SNPs in these individuals were compared with those available through dbSNP. The minor allele for rs2295748 was slightly less common in the 40 RE cases (0.22) than in any of the AFD or CEPH populations, whereas the minor allele for rs2273943 occurred in these cases at approximately the same frequency (0.24) as in the Caucasian and Chinese CEPH populations. Frequency information was not available for comparison for rs10767903 so we typed 85 controls at this SNP. The T allele at rs10767903 is predicted to abolish an adjacent splice donor enhancer site that would result in skipping of alternative exons 10 and 11. Out of the 36 RE probands that we were able to type at this synonymous polymorphism, 34 carried the T allele (21 TT, 13 CT, 2 CC), whereas controls exhibited a similar genotypic distribution: 42 TT, 34 CT, 9 CC.

Discussion

Taken together, our results suggest that *ELP4* is associated with the pathogenesis of RE and has a strong effect on risk for CTS in RE families. This locus appears to be distinct from those discovered in rare Mendelian RE variants. The precise mutation that is presumably in linkage disequilibrium with the associated SNPs in *ELP4* remains to be determined. However, the data presented here suggest that the presumed mutation lies either in the non-coding regions of *ELP4*, or else possibly just beyond the gene. This finding represents the first susceptibility gene identified for a common idiopathic focal epilepsy and the first step in unlocking the complex genetics of RE and related

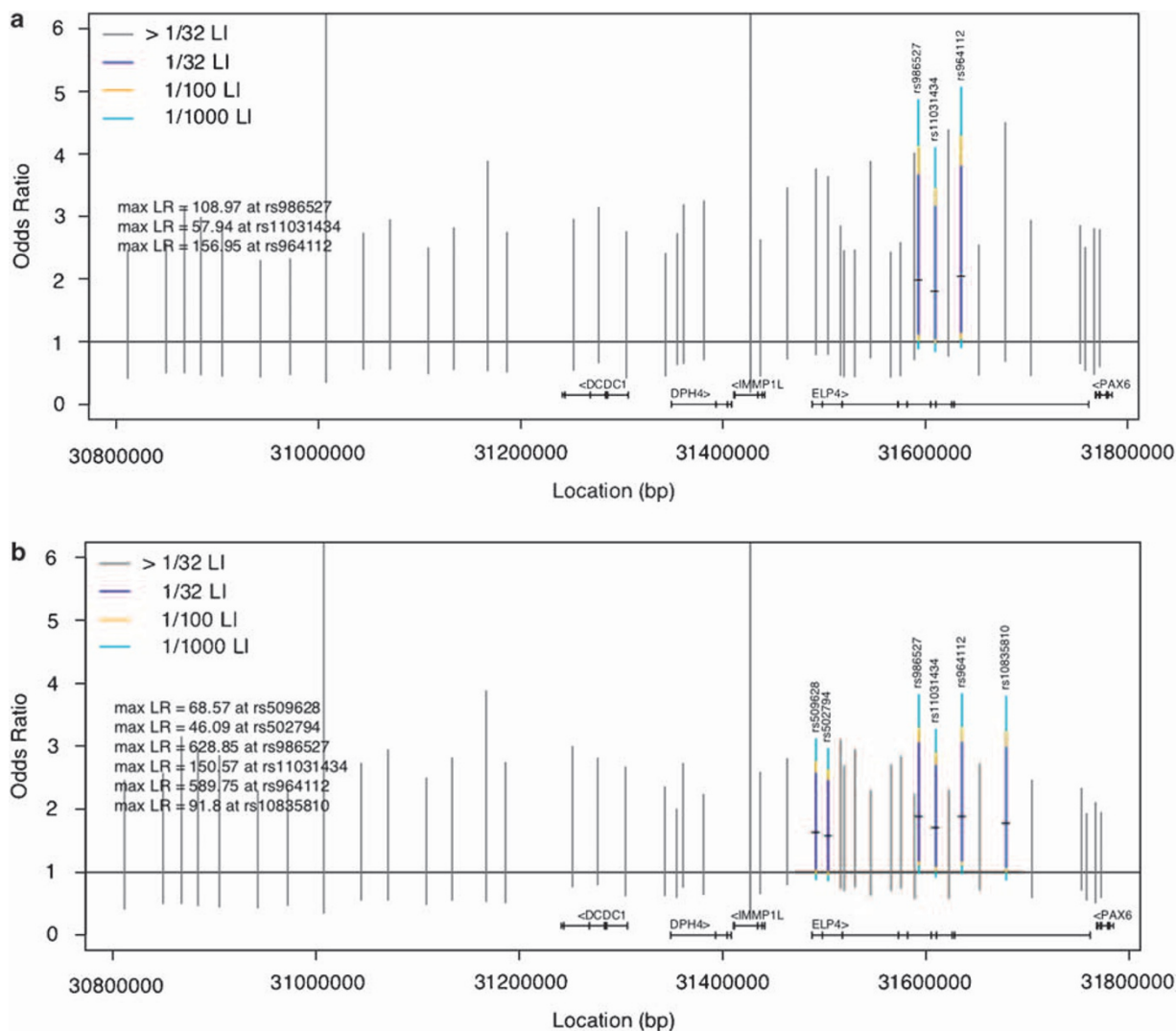


Figure 3 (a and b) Pure likelihood plot of association evidence in discovery set (a, top) and in joint analysis of data sets (b, bottom). This pure likelihood analysis plots odds ratio (OR) on the y axis and base-pair position on the x axis. Each vertical line represents a likelihood interval (LI) for the OR at a given SNP. The OR = 1 line is plotted as a solid black horizontal line, for reference. LIs in color are denoted as SNPs of interest, whereas a gray line indicates that the SNP is not of interest because the 1/32 LI for that SNP covers the OR = 1 line. The small horizontal tick on each LI is the maximum likelihood estimator for the OR. The portion of the colored LI that covers the OR = 1 horizontal line indicates the strength of the association information at that SNP. In particular, if the navy blue portion is above the OR = 1 line while the yellow portion of the LI covers the OR = 1 line, then the LOD evidence at that SNP is between 1.5 and 2 (ie, the 1/32 LI does not include the OR = 1 value, but the 1/100 LI does); similarly, if both the yellow and navy blue portions are above the OR = 1 line but the turquoise portion covers the line, then the LOD evidence is between 2 and 3 (ie, the 1/100 LI does not include OR = 1 as a plausible value but the 1/1000 LI does). The further the colored line is above the OR = 1 line, the stronger the association evidence. The max LR for each SNP in color is also provided as text in the plot, providing evidence not only of whether the LOD evidence is between 2 and 3, but also the exact value of the max LR.

childhood epilepsies. It is also the first reported disease association with *ELP4* in humans and offers possible insights into the etiology and kinship of associated developmental cognitive and behavioral disorders.

There are several reasons why these results are unlikely to be spurious. The localization of *ELP4* was conducted through genome-wide linkage analysis: only one area of the genome at 11p13 showed strong and compelling

evidence for linkage to CTS. Under that linkage peak, fine mapping evidence unambiguously pointed to the association of CTS with SNP markers in *ELP4*. SNPs in *ELP4* were associated with increased risk of CTS in both discovery and replication data sets, with evidence for association of the same SNPs in each data set. Furthermore, not only the same SNPs but also the same alleles were associated with increased risk of CTS in both data sets. Interestingly, we

found no evidence of locus or allelic heterogeneity based on ancestry in either linkage or association analyses. In addition, the association in the discovery set was consistent using FBAT, mitigating concerns about positive confounding due to population stratification. Thus, linkage and replicated association data offer compelling and consistent evidence for the role of *ELP4* in susceptibility to CTS. These results await independent replication.

The mapping of CTS to *ELP4* suggests that the common form of RE and rare variants of RE are genetically heterogeneous. Our data revealed little or no evidence of linkage to recessive (MIM 608105)¹³ or X-linked (MIM 300643)¹⁴ variants of RE, neither did a rare autosomal dominant form of RE with speech dyspraxia and cognitive impairment show linkage to 11p13.¹⁵ Thus, it seems that loci in Mendelian variants of RE may represent 'private' mutations. An earlier candidate gene study of CTS ascertained through northern European pedigrees multiplex for RE and ABPE did not test for linkage to chromosome 11.¹⁷ Instead, linkage was reported to the *EJM2* locus (MIM 604827), which may reflect genetic (locus) heterogeneity, or alternatively could be explained by shared susceptibility to myoclonic seizures, which feature in both juvenile myoclonic epilepsy and ABPE syndromes. There were no patients with myoclonic seizures in our samples.

This is the first reported disease association of *ELP4*. *ELP4* is one of the six subunits (*ELP1–ELP6*) of Elongator³⁶ that has both nuclear and cytoplasmic localization and two distinct but incompletely characterized roles in eukaryotic cells:³⁷ in transcription³⁸ and in tRNA modification.³⁹ Elongator plays a key role in transcription of several genes that regulate the actin cytoskeleton, cell motility and migration.⁴⁰ These functions are crucial in the nervous system for nerve cell growth cone motility, axon outgrowth and guidance, neuritogenesis and neuronal migration during development. Intriguingly, another Elongator subunit mutation has been implicated in human neurological disease. Riley–Day syndrome (MIM 223900) is an autosomal recessive, sensory and autonomic neuropathy, with EEG abnormalities and epilepsy.⁴¹ Riley–Day syndrome is caused by mutation in a splice site of the *hELP1* (or *IKAP*) gene, which causes tissue-specific exon skipping and expression of a truncated mRNA transcript,^{42,43} with highest ratios of mutant transcripts in the brain.⁴⁴ *hELP1* mutations abrogate Elongator function, not just *hELP1* expression,⁴⁰ because Elongator function is dependent on the integrity of all its subunits. Cells in Riley–Day patients have reduced motility, which can be rescued by wild-type *hELP1*.⁴⁰ *ELP4* mutations might also partially abrogate Elongator function in the central nervous system through its effect on multiple cell motility and actin cytoskeleton genes and/or proteins during development. Such a mechanism could plausibly explain the breadth of subtle developmental disorders that are associated with the

CTS trait,¹⁰ ranging from speech disorder⁷ and DCD⁹ to attention problems and ADHD.^{4,8} Proof of this hypothesis will require the genetic investigation of large cohorts of carefully phenotyped individuals and detailed functional characterization of *ELP4*.

Although we found association with SNPs across *ELP4*, regression analysis indicated that spread of association evidence could be explained by linkage disequilibrium around rs986527 in intron 9, LD that stretches to *IMMP1L* and the 3'-end of *ELP4*, but not to *PAX6*. Subsequent resequencing of the coding, boundary and promoter regions revealed no enrichment of *ELP4* exonic polymorphisms among probands. Exclusion of the coding sequences suggests that the genetic effector may lie in the non-coding regions of *ELP4*. It is less likely that the causative mutation lies in a distant gene beyond *IMMP1L* upstream or *PAX6* downstream there is a drop-off in linkage disequilibrium at subjacent markers. Interestingly, the non-coding regions between *ELP4* exon 9 and exon 12 are large (over 130 kb), and contain long range, tissue-specific, *cis*-regulatory elements for *PAX6*.⁴⁵ *PAX6* remains a candidate gene of interest because of its highly conserved, developmental regulatory role in the formation of the telencephalon.⁴⁵ Intron 9 of the canonical splice variant contains a recently inserted pseudogene that is included in the *ELP4* mRNA to produce a splice variant, with alternative exons 10 and 11, that is only found in higher primates. The alternative exon 10 contains multiple sequences that are consensus-binding sites for splicing enhancer-binding proteins. Although there are multiple transcripts of *ELP4* that contain alternatively spliced exons and encode different sized proteins, the functional and evolutionary significance of most is presently unclear. Expression analyses and resequencing of the non-coding regions may help to reveal the molecular mechanism of seizure susceptibility at this locus.

Substantiating *ELP4* as a risk locus for CTS is the first step in assembling the complex genetic model of RE. Additional genetic factors though, may need to be invoked to explain the occurrence of seizures and reading disability in RE. For example, although CTS is common in children,⁶ only an estimated 10% of children with the trait manifest clinical seizures.¹¹ At the same time, there is no evidence for an environmental contribution to RE. Thus, although CTS is mandatory for the definition of RE, additional genetic factors, which likely act in combination with the *ELP4* locus to cause the classic focal seizures of RE, remain to be elucidated. In summary, we report strong, replicated association between *ELP4* variants and the CTS trait. We hypothesize that an as-yet unidentified non-coding mutation exists that is in linkage disequilibrium with SNPs in *ELP4* intron 9. This hypothesized mutation impairs brain-specific Elongator function during brain development, possibly mediated through interaction with genes and proteins in cell migration and actin cytoskeleton pathways.

Acknowledgements

We thank the families participating in the IRELAND study (<http://www.rolandic.com>); Irina Gurvich PhD, Regina Santella PhD, Fengli Zhang MD, Frances Rhoads MD, Jerry Boxerman MD PhD, Jeffrey Rogg MD, Lewis Kull MS, Geoffrey Tremont PhD, Janessa Carvalho BS, Suzanne Foster BS, George Cholankeril, Doe West PhD, Kieran Hartsough BS, Lynne Capuano, Somaria Persaud; and the members of the IRELAND consortium: Cigdem Akman MD, William D Brown MD, Murray Engel MD, John Gaitanis MD, Frank Gilliam MD, Karameh Hawash MD, Huntley Hardison MD, Geoffrey Heyer MD, Steven L Kugler MD, Linda Leary MD, David E Mandelbaum MD PhD, Patricia McGoldrick RN, Edward Novotny MD, Steven M Wolf MD and Maria Younes MD. This study was supported by members of the Partnership for Pediatric Epilepsy Research, which includes the American Epilepsy Society, the Epilepsy Foundation, Anna and Jim Fantaci, Fight Against Childhood Epilepsy and Seizures (faces), Neurotherapy Ventures Charitable Research Fund and Parents Against Childhood Epilepsy (PACE) (DKP); the Epilepsy Foundation through the generous support of the Charles L Shor Foundation for Epilepsy Research Inc. (DKP); People Against Childhood Epilepsy (PACE) (DKP), National Institutes of Health grants NS047530 (DKP) and HG00-4314 (LJS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of Neurological Disorders And Stroke, the National Institutes of Health or any other sponsor. We acknowledge Professor Hermann Doose's intellectual contribution to the nosology of focal sharp wave-associated phenotypes in children. We thank Dr Susan E Hodge for critical reading of the paper; Dr Andreas Kottman and Dr Livio Pellizzoni for helpful discussions.

Web resources

Mendelian variants of RE are listed in Online Inheritance in Man <http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>. The Haploview application can be downloaded at <http://www.broad.mit.edu/mpg/haploview>. Information about marker location can be found at UniSTS at the NCBI website above and through the Ensembl genome browser at http://www.ensembl.org/Homo_sapiens/Info/Index. SNP frequencies were accessed at dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP>. The ESE Finder program was used to assess alternative exon 10 of *ELP4* and can be accessed at http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese_finder.cgi?process=home. For bioinformatic analyses of *ELP4* protein variants produced by alternative gene splicing we used PHI-BLAST, and accessed the following PANTHER and Bioinformatics websites: <http://www.pantherdb.org> and http://bioinformatics.weizmann.ac.il/blocks/blockmkr/www/make_blocks.html.

References

- Shinnar S, O'Dell C, Berg AT: Distribution of epilepsy syndromes in a cohort of children prospectively monitored from the time of their first unprovoked seizure. *Epilepsia* 1999; **40**: 1378–1383.
- Beaussart M: Benign epilepsy of children with Rolandic (centro-temporal) paroxysmal foci: a clinical entity. Study of 221 cases. *Epilepsia* 1972; **13**: 795–811.
- Clarke T, Strug LJ, Murphy PL *et al*: High risk of reading disability and speech sound disorder in rolandic epilepsy families: case-control study. *Epilepsia* 2007; **48**: 2258–2265.
- Kavros PM, Clarke T, Strug LJ, Halperin JM, Dorta NJ, Pal DK: Attention impairment in rolandic epilepsy: systematic review. *Epilepsia* 2008; **49**: 1570–1580.
- Boxerman J, Hawash K, Bali B, Clarke T, Rogg J, Pal DK: Is Rolandic epilepsy associated with abnormalities on cranial MRI? *Epilepsy Res* 2007; **75**: 180–185.
- Eeg-Olofsson O, Petersen I, Selden U: The development of the electroencephalogram in normal children from the age of 1 through 15 years. Paroxysmal activity. *Neuropadiatrie* 1971; **2**: 375–404.
- Echenne B, Cheminal R, Rivier F, Negre C, Touchon J, Billiard M: Epileptic electroencephalographic abnormalities and developmental dysphasias: a study of 32 patients. *Brain Dev* 1992; **14**: 216–225.
- Holtmann M, Becker K, Kentner-Figura B, Schmidt MH: Increased frequency of rolandic spikes in ADHD children. *Epilepsia* 2003; **44**: 1241–1244.
- Scabar A, Devescovi R, Blason L, Bravar L, Carrozzi M: Comorbidity of DCD and SLI: significance of epileptiform activity during sleep. *Child Care Health Dev* 2006; **32**: 733–739.
- Doose H, Neubauer B, Carlsson G: Children with benign focal sharp waves in the EEG – developmental disorders and epilepsy. *Neuropediatrics* 1996; **27**: 227–241.
- Heijbel J, Blom S, Rasmuson M: Benign epilepsy of childhood with centrotemporal EEG foci: a genetic study. *Epilepsia* 1975; **16**: 285–293.
- Scheffer IE, Jones L, Pozzebon M, Howell RA, Saling MM, Berkovic SF: Autosomal dominant rolandic epilepsy and speech dyspraxia: a new syndrome with anticipation. *Ann Neurol* 1995; **38**: 633–642.
- Guerrini R, Bonanni P, Nardocci N *et al*: Autosomal recessive Rolandic epilepsy with paroxysmal exercise-induced dystonia and writer's cramp: delineation of the syndrome and gene mapping to chromosome 16p12–11.2. *Ann Neurol* 1999; **45**: 344–352.
- Roll P, Rudolf G, Pereira S *et al*: SRPX2 mutations in disorders of language cortex and cognition. *Hum Mol Genet* 2006; **15**: 1195–1207.
- Kugler SL, Bali B, Lieberman P *et al*: An autosomal dominant genetically heterogeneous variant of rolandic epilepsy. *Epilepsia* 2008; **49**: 1086–1090.
- Bali B, Kull L, Strug L *et al*: Autosomal dominant inheritance of centrotemporal sharp waves in rolandic epilepsy families. *Epilepsia* 2007; **48**: 2266–2272.
- Neubauer BA, Fiedler B, Himmelein B *et al*: Centrotemporal spikes in families with rolandic epilepsy: linkage to chromosome 15q14. *Neurology* 1998; **51**: 1608–1612.
- Royall RM: *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall, 1997.
- Blume JD: Tutorial in biostatistics: likelihood methods for measuring statistical evidence. *Stat Med* 2002; **21**: 2563–2599.
- Strug LJ, Hodge SE: An alternative foundation for the planning and evaluation of linkage analysis. I. Decoupling 'error probabilities' from 'measures of evidence'. *Hum Hered* 2006; **61**: 166–188.
- Strug LJ, Rohde CA, Corey PN: An introduction to evidential sample size calculations. *Am Stat* 2007; **61**: 207–212.
- Commission on classification and terminology of the international league against epilepsy: proposal for revised classification of epilepsies and epileptic syndromes. *Epilepsia* 1989; **30**: 389–399.
- Hodge SE, Abreu P, Greenberg DA: Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. *Am J Hum Genet* 1997; **60**: 217–227.
- Abreu PC, Hodge SE, Greenberg DA: Quantification of type I error probabilities for heterogeneity LOD scores. *Genet Epidemiol* 2002; **22**: 159–169.
- Pal DK, Durner M, Greenberg DA: Effect of misspecification of gene frequency on the two-point lod score. *Am J Hum Genet* 2001; **9**: 855–859.
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- Daw EW, Thompson EA, Wijsman EM: Bias in multipoint linkage analysis arising from map misspecification. *Genet Epidemiol* 2000; **19**: 366–380.
- Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

- 29 Strug LJ, Hodge SE: An alternative foundation for the planning and evaluation of linkage analysis. II. Implications for multiple test adjustments. *Hum Hered* 2006; **61**: 200–209.
- 30 Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995; **57**: 289–300.
- 31 Laird NM, Horvath S, Xu X: Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 2000; **19** (Suppl 1): S36–S42.
- 32 Lindsey JK: *Parametric Statistical Inference*. Oxford: Clarendon Press, 1996.
- 33 Kong A, Gudbjartsson DF, Sainz J *et al*: A high-resolution recombination map of the human genome. *Nat Genet* 2002; **31**: 241–247.
- 34 Vieland VJ, Hodge SE: Review of statistical evidence: a likelihood paradigm by Royall, R. *Am J Hum Genet* 1998; **63**: 283–289.
- 35 Clayton D, Chapman J, Cooper J: Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol* 2004; **27**: 415–428.
- 36 Otero G, Fellows J, Li Y *et al*: Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol Cell* 1999; **3**: 109–118.
- 37 Svejstrup JQ: Elongator complex: how many roles does it play? *Curr Opin Cell Biol* 2007; **19**: 331–336.
- 38 Krogan NJ, Greenblatt JF: Characterization of a six-subunit holo–elongator complex required for the regulated expression of a group of genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* 2001; **21**: 8203–8212.
- 39 Esberg A, Huang B, Johansson MJ, Bystrom AS: Elevated levels of two tRNA species bypass the requirement for elongator complex in transcription and exocytosis. *Mol Cell* 2006; **24**: 139–148.
- 40 Close P, Hawkes N, Cornez I *et al*: Transcription impairment and cell migration defects in elongator-depleted cells: implication for familial dysautonomia. *Mol Cell* 2006; **22**: 521–531.
- 41 Niedermeyer E, McKusick VA, Brunt P, Mahloudji M: The EEG in familial dysautonomia (Riley–Day syndrome). *Electroencephalogr Clin Neurophysiol* 1967; **22**: 473–475.
- 42 Anderson SL, Coli R, Daly IW *et al*: Familial dysautonomia is caused by mutations of the IKAP gene. *Am J Hum Genet* 2001; **68**: 753–758.
- 43 Slaugenhaupt SA, Blumenfeld A, Gill SP *et al*: Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am J Hum Genet* 2001; **68**: 598–605.
- 44 Mezey E, Parmalee A, Szalayova I *et al*: Of splice and men: what does the distribution of IKAP mRNA in the rat tell us about the pathogenesis of familial dysautonomia? *Brain Res* 2003; **983**: 209–214.
- 45 Griffin C, Kleinjan DA, Doe B, van Heyningen V: New 3' elements control Pax6 expression in the developing pretectum, neural retina and olfactory region. *Mech Dev* 2002; **112**: 89–100.
- 46 Royall RM: On the probability of observing misleading statistical evidence (with discussion). *J Am Stat Assoc* 2000; **95**: 760–780.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)