Citation for published version (APA):
Microdeletions of *ELP4* are associated with language impairment, autism spectrum disorder and epilepsy.
| Key Words: | Copy Number Variation (CNV), Epilepsy and seizures, Developmental, Neurology, ELP4 |
Microdeletions of *ELP4* are associated with language impairment, autism spectrum disorder and epilepsy


1: Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
2: Neuroscience Discovery Research, Eli Lilly and Company, Erl Wood, Surrey, UK
3: Department of Cytogenetics, Guy's and St Thomas' NHS Foundation Trust, London, UK
4: Department of Biostatistics and NIHR BRC for Mental Health, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
5: Departments of Psychiatry, and Genetics and Genomic Sciences, Seaver Autism Center, The Mindich Child Health & Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
6: Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada.
7: Utah Autism Research Program, Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT, USA
8: Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
9: Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
10: Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Perth, WA, Australia.
11: School of Medicine and Pharmacology, University of Western Australia, Nedlands, WA, Australia.
12: Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
13: Institut de Génétique Médicale, Hopital Jeanne de Flandre, CHRU de Lille, France
14: Centre de Génétique Chromosomique, GHICL, Hôpital Saint Vincent de Paul, Lille, France
15: Applied Human Molecular Genetics, Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
16: Dept of Pediatrics, Child Neurology and Psychiatry, Sapienza Università di Roma, Rome, Italy
17: Institute of Human Genetics, Medical University of Graz, Graz, Austria
18: Department of Clinical Genetics, Helsinki University Hospital, Helsinki, Finland
19: Department of Molecular Genetics, Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada
20: McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
21: Program in Genetics and Genome Biology, The Hospital for Sick Children; Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

**Correspondence:** Prof. Deb K Pal, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 1-4 Windsor Walk, London SE5 8AF, London, UK. Tel: +44 20 7848 0636; Fax: +44 20 7848 0988; E-mail: deb.pal@kcl.ac.uk
ABSTRACT

Copy number variations (CNV) are important in the aetiology of neurodevelopmental disorders and show broad phenotypic manifestations. We compared the presence of small CNVs disrupting the ELP4-PAX6 locus in 4,092 U.K. individuals with a range of neurodevelopmental conditions, clinically referred for array comparative genomic hybridisation (aCGH), with WTCCC controls (n=4,783). The phenotypic analysis was then extended using the DECIPHER database. We followed up association using an autism patient cohort (n=3,143) compared with six additional control groups (n=6,469). In the clinical discovery series we identified eight cases with ELP4 deletions, and one with a partial duplication of ELP4 and PAX6. These cases were referred for neurological phenotypes including language impairment, developmental delay, autism and epilepsy. Six further cases with a primary diagnosis of ASD and similar secondary phenotypes were identified with ELP4 deletions, as well as another six (out of 9) with neurodevelopmental phenotypes from DECIPHER. CNVs at ELP4 were only present in 1/11,252 controls. We found a significant excess of CNVs in discovery cases compared with controls, p=7.5x10^{-3}; as well as for autism, p=2.7x10^{-3}. Our results suggest ELP4 deletions are highly likely to be pathogenic, predisposing to a range of neurodevelopmental phenotypes from ASD to language impairment and epilepsy.

Key Words: Copy Number Variation (CNV), Epilepsy and seizures, Developmental, Neurology
INTRODUCTION

Copy number variation (CNV) plays an important role in the aetiology of neurodevelopmental and psychiatric disorders. Both recurrent de novo and rare segregating CNVs have begun to explain the overlap of diverse phenotypes in individual cases and families (Cooper, et al., 2011; Malhotra and Sebat, 2012). Copy number variation is a strong risk factor in both focal and generalized epilepsies, and they are also found in 8% of patients with epileptic encephalopathies (Mefford, et al., 2010; Mefford, et al., 2011). Recent findings in the rare epileptic encephalopathies illustrate the connection between epilepsy, language impairment and autism spectrum disorder (ASD) through overrepresentation of novel CNVs containing cell adhesion genes, (e.g. cadherins and contactins) (Lesca, et al., 2012). However, there are also differences between disorders: for example, specific language impairment cases, whilst having an increased burden of CNVs, do not in general show enrichment for novel or de novo events (Simpson, et al., 2015), whereas rare copy number variation is an important source of risk in ASD (Pinto, et al., 2014).

The examples above indicate that a given genomic alteration can sustain broad susceptibility to several phenotypes depending on the genetic background of the subject. So called ‘hotspot’ CNVs also manifest this phenotypic variability. The recurrent 15q13.3 microduplication increases the risk for intellectual disability, idiopathic generalised epilepsy, ASD and schizophrenia and (Helbig, et al., 2009; Poot, et al., 2011), and deletions at 16p13.11 contribute to a diverse spectrum of epilepsy disorders (Heinzen, et al., 2010). The 16p11.2 hotspot is also pleiotropic; deletions are common in ASD and developmental delay, (Marshall, et al., 2008) and duplications have been associated with seizures and speech delay (Shinawi, et al., 2010). Other notable examples of pleiotropy are CNVs of the CNTNAP2
gene, which are implicated in ASD, Gilles de la Tourette syndrome, schizophrenia and
epilepsy, and AUTS2 with ASD and mental retardation. Interestingly AUTS2 and CNTNAP2
may interact with each other on a molecular level, (Poot, et al., 2011), indicating emerging
convergent pathways for neurodevelopment. A recent report of a deletion of the ELP4 gene at
11p13, and adjacent 3’ PAX6 enhancer elements has been described in a case with aniridia,
autism and mental retardation. This case differs from the ‘classical’ PAX6 gene deletions
cause aniridia alone, as only the 3’ enhancer elements are deleted in this case and ELP4 is
included (Davis, et al., 2008). ELP4 has previously been associated with the
electroencephalographic (EEG) signature of the common childhood epilepsy Rolandic
epilepsy (RE) (Strug, et al., 2009), and such EEG abnormalities as well as epilepsy are well
established in autism spectrum disorder (ASD) and language impairments, (Nasr, et al., 2001;
Parmeggiani, et al., 2010). These examples illustrate again that genomic alterations can show
broad phenotypic manifestations during neurodevelopment, as well as incomplete penetrance.

In the present study we report the presence of a number of deletions of ELP4 and the
regulatory elements of PAX6 in the U.K. database of individuals with a childhood onset
developmental condition referred for clinical genetic testing (BB-GRE). We test the
hypothesis that the burden of ELP4 CNVs is increased in those with neurodevelopmental
conditions compared to controls. This phenotypic analysis is then extended using the
DECIPHER database of chromosomal imbalances in over 10,000 cases of developmental
disorders. Using a CNV-led approach we then further expand the phenotype associated with
ELP4 microdeletions to cases with ASD and varying comorbidities, and carry out a second
case-control analysis of frequency. This data supports our hypothesis that disruption of ELP4
and the regulatory regions of PAX6 contained within its introns, lead to a range of
neurodevelopmental conditions.
METHODS

Study Design

We used a three-stage design; first testing the hypothesis of CNV enrichment at \textit{ELP4} in a clinical discovery sample of developmental disorders (Brain and Body Genetics Research Exchange (BBGRE, \url{https://bbgre.iop.kcl.ac.uk}) and control dataset (WTCCC), both from the U.K.; second, extending the phenotypic analysis to a larger dataset of developmental disorders (DECIPHER) and finally, replicating the association to neurodevelopmental disorders in ASD cases (Autism Genome Project and two Canadian ASD cohorts) compared to a large multi-centre control sample set.

Samples

U.K. Clinical Dataset – BB-GRE

4092 children referred to Guy’s and St Thomas NHS Foundation Trust, southeastern UK from paediatricians and regional hospitals, \url{https://bbgre.iop.kcl.ac.uk}. Individuals referred for array-CGH testing for a range of developmental problems including developmental delay (DD), ASD, speech or language delay or congenital defects. Individuals had clinical diagnoses made prior to genetic testing, which was part of standard clinical care. Genomic data and referral phenotype information were anonymised and recorded in a clinical database, (63% males; August 2014).

Global Clinical Dataset - DECIPHER

We performed a search in the DECIPHER database (Firth, et al., 2009) in order to identify additional cases with small CNVs which included and/or disrupted \textit{ELP4}. DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources,
http://decipher.sanger.ac.uk) is an interactive web-based database of over 10,000 cases which enables clinical scientists to maintain records of phenotype and chromosome rearrangement, to aid patient diagnosis by linking to other bioinformatics resources and interactive tools, and to share this information with the clinical research community.

**Canadian ASD Samples**

The cohort contained 349 probands previously published (Lionel, et al., 2011) and 350 additional patients diagnosed with ASD from Canada described below, totaling 699. Individuals were recruited from four different Canadian sites: The Hospital for Sick Children, Toronto; McMaster University, Hamilton; Memorial University, St. John's, Newfoundland and University of Alberta, Edmonton. All had a clinical diagnosis of ASD, using the Autism Diagnostic Interview-Revised (ADI-R) and/or Autism Diagnostic Observation Schedule (ADOS).

**Autism Genome Project (AGP) Samples**

2,147 European ASD cases were genotyped as part of a study by the AGP Consortium for rare CNVs affecting autism and are formally described in the following reference, (Pinto, et al., 2014). All cases had a clinical diagnosis of autism rated using the ADI-R and/or the ADOS.

**Control Populations**

A total sample of 11,252 controls from six different datasets were included this study. Group 1 was compared with the BBGRE cases, and groups 2-6 with the AGP cases: (1) WTCCC, Wellcome Trust Case Control Consortium controls - 4,783 population controls from the UK (Consortium, et al., 2010) (2) Ottawa Heart Institute (OHI) controls - A cohort of 1,234
control individuals collected as part of a large case control GWA study (Stewart, et al., 2009); (3) German POPGEN controls – a sample of 1,123 individuals of northern German origin (Schleswig-Holstein) (Krawczak, et al., 2006); (4) Ontario Population Genomics Platform (OPGP) controls – a Canadian sample of 416 control individuals of European ancestry [http://www.tcag.ca/facilities/cyto_population_control_DNA.html]; (5) HapMap3 controls - a sample of 1,056 individuals from populations from around the world from the International HapMap Project [http://www.hapmap.org/]; (6) Controls from the AGP project – consisting of 2,640 of European ancestry assembled from three studies in which subjects had no obvious psychiatric history: 'Study of Addiction Genetics and Environment (SAGE)', 'Ontario Colorectal Cancer case-control study (OC)', and the 'Health, Aging, and Body Composition (HABC)'.

**Genotyping and CNV Analysis**

**Array CGH analysis of BB-GRE Samples**

Array CGH testing was carried out at the Guys and St Thomas’ Services cytogenetics CPA accredited laboratory. We have previously described the protocols, analysis and interpretation using an Agilent oligonucleotide array 60K platform (AMAID 028469 and 017457) and a patient vs. patient hybridization strategy and 3-probe minimum aberration call in (Ahn, et al., 2013; Ahn, et al., 2010). The average probe density over $ELP4$ is 8.5Kb, giving a limit of around 25Kb for detection. CNVs in this population are available by application to BB-GRE; [https://bbgre.iop.kcl.ac.uk/](https://bbgre.iop.kcl.ac.uk/).

**Canadian ASD and Control Groups 1-5**

Canadian ASD cases, and control populations 1-5, were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 with standard protocols. Arrays meeting Affymetrix
quality control guidelines of Contrast QC > 0.4 were further analysed. Raw data analysis was carried out using a multiple-algorithm approach to maximize sensitivity and specificity of CNV calling, as described previously (Lionel, et al., 2011; Silversides, et al., 2012). Briefly, arrays were analyzed for CNVs with Birdsuite (Korn, et al., 2008) iPattern (Pinto, et al., 2011) and Affymetrix Genotyping Console and merged into a single dataset. A CNV call was considered high confidence if it was detected by at least two of the calling algorithms and spanned at least 10kb and >5 consecutive array probes. Average probe density over \textit{ELP4} was 2.0Kb, giving a limit of around 10Kb for detection.

**Autism Genome Project Samples and Control Group 6**

2,147 ASD cases and 2,640 controls were genotyped with the Illumina Infinium 1M SNP microarray. CNV calling was performed using a multi-algorithm approach incorporating PennCNV, iPattern and QuantiSNP (Pinto, et al., 2010). Subsequent analyses focused on those CNVs spanning five or more array probes and detected by at least two algorithms. The analysis is formally described in (Pinto, et al., 2014). The average probe density over \textit{ELP4} was 2.1Kb, giving a limit of around 10Kb for detection.

**Association analysis**

A two-tailed Fishers exact test was used to compare frequencies of \textit{ELP4} CNVs in the 4,092 cases in BB-GRE with the 4,783 controls from the WTCCC. Subsequently another two-tailed Fishers exact test was used to compare the frequency in 2,845 unrelated ASD cases compared to 6,469 controls from control sets 2-6 combined.

**Limitations**
A limitation of this study is that the CNVs were not identified on the same platform or by the same analysis method between the sample sets. Therefore there is a chance of false CNV enrichment related to probe density, data quality and analysis methods. However, all platforms are high-density, with probe coverage shown in Figure 1. This ensures \textit{ELP4} and the surrounding region is well covered, and indeed the control data is generated on higher density platforms than the BBGRE cases, resulting in a higher CNV detection power for controls. We have also ensured that all reported CNVs can be called using all three methodologies. The Canadian ASD cases and control groups 1-5, also use the same platform and analysis methods as each other. Cases and controls from the AGP study also used the same array and analysis methods as each other. To reduce the chance of error, all of the CNV calling methods from each data set employ published, rigorous quality control measures as detailed above, ensuring that CNVs called are highly unlikely to be false positives. All of the ASD CNVs have also been validated by orthogonal methods such as qPCR. By also using data sets with different platforms, we have shown that our results are consistent even between the different methods used.
RESULTS

Copy Number Variation of *ELP4* in the BB-GRE Database

Out of 4092 individuals referred for neurodevelopmental disorders, we identified nine patients with small (<1Mb) CNVs disrupting *ELP4* that could also have be detected by the other array methods. Eight CNVs were deletions, Figure 1 and Table 1, varying in size from 26Kb to 101Kb. The ninth CNV was a 232Kb duplication of the first 7 exons of *ELP4* and the *PAX6* gene. This patient also carried a ‘hotspot’ deletion of 1.2Mb at 16p13.11, which is also implicated in several neuropsychiatric disorders (Heinzen, et al., 2010). One deletion (117374) is intronic, but does however disrupt regulatory enhancers of *PAX6* and so is included in our analysis. Four of the deletions were maternally inherited and two were paternally inherited, one arose *de-novo* and one had unknown inheritance. The inheritance pattern of the duplication was also unknown. Clinical information was not available for the parents as it is not collected for BBGRE and referring clinicians cannot be contacted. Three of the deletion patients carried a second CNV, Table 1, none of which are predicted to affect the phenotype; 119460 had a deletion of unknown inheritance of 77.6Kb at 5q21 with no genes present in the region, and 112601 had a maternally inherited 226Kb duplication at 5q15 disrupting *FAM172A*, a potential tumor suppressor. Patient 130693 carried a maternally inherited duplication of 23Kb at 6p22.2, disrupting the MHC-associated genes *BTN3A3* and *BTN2A1*.

All cases were diagnosed with a neurodevelopmental phenotype; five had speech and language delay or disorder, with one also diagnosed with epilepsy, two had social communication difficulties and two had a diagnosis of autism, with one further case showing emerging autistic traits. Six of the patients also had a range of cognitive delays, Table 1.
Unfortunately we do not know the age at last neurological assessment for BBGRe cases; only age at aCGH testing is recorded. Therefore some cases may be too young for some phenotypes to manifest and be reported, e.g. 117003.

Only one CNV involving ELP4 was found in the WTCCC control set; a 221Kb microdeletion, Supplementary Information Table 1. On comparison of the U.K. BB-GRE samples with the WTCCC controls, the difference in CNV frequency disrupting ELP4 was significant; p-value=7.5x10^{-3}.

Microdeletions of ELP4 in the DECIPHER Database

We identified nine individuals with a small (< 1Mb) CNV encompassing ELP4 in the DECIPHER database (https://decipher.sanger.ac.uk) (Firth, et al., 2009). All were deletions, Table 2, with at least one breakpoint within the gene. Detailed phenotypic information was available for eight of the nine patients; six individuals were diagnosed with developmental delay or intellectual disability. Several cases had speech delay; two had behavioral disorders, one was diagnosed with a pervasive developmental disorder (PDD), most likely ASD, and one further case had ASD. Another case was also diagnosed with ADHD and epilepsy (257614). Three cases were too young at the age of last clinical visit (263619, 265704 and 287341) for a full assessment of neurodevelopmental phenotypes such as ASD.

Two cases, 289275 and 270752, had deletions that disrupted PAX6 exons. Most likely due to PAX6 enhancer or exon disruption, these two cases, as well as 265704 and 263741, also have aniridia, an abnormality of the iris. Two others cases, 257614 and 249728, have congenital eye malformations but deletion breakpoints much further from PAX6. The remaining three
cases, whilst having breakpoints very similar to those with aniridia, do not share that phenotype, indicating a complex genotype-phenotype relationship.

Two cases carried a second CNV: 289275 had an intronic duplication of $ZNF674$ at Xp11.3, and 261471 a 245Kb deletion at 10p11.21, disrupting $CUL2$, $CREM$ and $CCNY$, genes not involved in neuronal development.

**Microdeletions of $ELP4$ in Autism Cases**

Given that several individuals from BB-GRE and DECIPHER have ASD, PDD or social communication difficulties, we decided to investigate the prevalence of $ELP4$ CNVs in two autism cohorts. Out of 2,446 cases from the Autism Genome Project, AGP, (Pinto, et al., 2014) three had microdeletions of $ELP4$, Table 2. All three fulfilled the criteria for a strict definition of autism, and were verbal (verbal IQ >70), but experienced language delay of first words and phrases. None of the cases had a history of seizures or epilepsy. Case 8596_201 also carried a 500Kb maternally inherited duplication disrupting the collagen gene $COL27A1$ that is highly unlikely to contribute to the neurological phenotype.

Three out of 699 individuals from the Canadian autism study also carried $ELP4$ deletions, Table 3. An affected sister pair both had a 112Kb deletion of half of the gene, and a male case carried a 130Kb deletion of the 3’ (but proximal due to reverse gene orientation) part of $ELP4$ and neighboring $IMMP1L$. Again, all three had speech and language delay and the sister-pair had mild developmental delay. Interestingly, the sisters also both carried a deletion of one copy of exon 2 of $TMLHE$, an enzyme involved in carnitine biosynthesis, on Xq28.
A case-control analysis of the frequency of *ELP4* CNVs from unrelated individuals in these 2,845 ASD cases compared with 6,469 control individuals from groups 2-6, where no *ELP4* CNVs were found, yielded a highly significant p value of 2.7x10^{-3}.
DISCUSSION

In this study we have found a strong and consistent pleiotropic association between CNVs disrupting \textit{ELP4} and neurodevelopmental conditions over several experimental platforms. We have described CNVs that can be captured and called from all three high density platforms/methods. We have also addressed the potential problem of enrichment bias of CNVs in cases, as all control data is generated on higher density platforms than the BBGRE cases, or the same as ASD/DECIPHER cases, resulting in a higher CNV detection power for controls. CNVs disrupting \textit{ELP4} appear to be rare in the general population given that we found only one CNV in the six control groups studied (total n = 11,252), and that there are no regions of segmental duplication around the gene (UCSC Segmental Duplication track, (Bailey, et al., 2002)). \textit{ELP4} now joins the growing list of genes such as \textit{CNTNAP2}, \textit{SHANK3} and \textit{NRXN1}, where heterozygous copy number are repeatedly associated with a wide range of neuropsychiatric disorders (Gregor, et al., 2011; Lesca, et al., 2012; Poot, et al., 2011).

We have extended the phenotype associated with disruptions of \textit{ELP4} from the EEG signature of Rolandic epilepsy and speech sound disorder (Pal, et al., 2010; Strug, et al., 2009) to ASD, social communication difficulties, developmental delay, and epilepsy. This corroborates the findings of Davis et al, who found a deletion of \textit{ELP4} and \textit{PAX6} enhancer elements in a patient with autism, aniridia and mental retardation that was inherited from an affected mother (Davis, et al., 2008). The \textit{ELP4} locus may influence the development of language function, as a frequent trait across almost half of the 24 patients described here are speech and language difficulties. There appears to be a genetic crossroads between childhood epilepsy, autism, and speech and language disorders. Several genes and pathways provide a common link such as the cell adhesion genes cadherins and catenins, glutamate receptors \textit{GRIN2A} and \textit{2B}, brain-
expressed nuclear proteins such as AUTS2, and the transcription factor FOXP2 (Graham and Fisher, 2012; Lesca, et al., 2013; Lesca, et al., 2012; Poot, et al., 2011).

ELP4 is one of six subunits (ELP1-6) of the Elongator complex, which plays a role in transcriptional elongation (Wittschieben, et al., 1999), tRNA modification and polarized exocytosis (Huang, et al., 2005). This complex also regulates the migration of multiple cell types; e.g. ELP1 co-localises with filamin A in membrane ruffles, and when depleted creates a disorganised actin cytoskeleton, contributing to motility defects (Johansen, et al., 2008). Impairment of Elongator may be involved in several different neurological disorders (Nguyen, et al., 2009) e.g. variants within ELP3 are associated with cases of sporadic ALS, a progressive motor-neuron disease (Simpson, et al., 2009). Furthermore, mutations of ELP1 cause familial dysautonomia (Slaugenhaupt, et al., 2001), a neurodevelopmental and neurodegenerative disorder with EEG abnormalities and seizures, characterized by defects in neuronal development and survival. Elongator also underlies the migration and branching of cortical projection neurons during development and memory consolidation (Creppe, et al., 2009). Thus there are several mechanisms through which disruption of ELP4 could result in altered neuronal development and migration, as well as the balance of neuronal excitatory and inhibitory circuits. These changes may disrupt Elongator function in a temporal and regional manner depending on cellular context and the different array of Elongator targets available.

It is of note that the large intronic regions between exons 9 and 12 of ELP4 are ultraconserved. They contain long-range cis-regulatory enhancers for downstream PAX6, which are tissue- or developmental stage specific in their expression (McBride, et al., 2011). PAX6 is a transcription factor crucial for the correct development of the eyes, spinal cord, several brain regions and other organs. Deletions of PAX6 with WT1 cause Wilms tumor,
aniridia, genital anomalies, and intellectual disability (WAGR syndrome). Loss-of-function mutations in PAX6 also cause aniridia. A rare case of duplication of PAX6 and the last two exons/introns of ELP4 has been reported with frontotemporal neonatal seizures, developmental delay, microcephaly and minor ocular findings, (Aradhya, et al., 2011).

Recently PAX6 has also been proposed as the foremost transcription factor governing glutamatergic neuronal differentiation (Kim, et al., 2014), linking it with the major idiopathic focal epilepsy gene glutamate receptor GRIN2A. Therefore disruption of PAX6 and/or its regulatory elements within ELP4 and its link via the glutamatergic neurotransmission system described above also make it a prime candidate for involvement in the neurodevelopmental disorders described in some cases here. However, whilst it is difficult to untangle which gene is causing which phenotype in a few cases, for the majority, PAX6 is not disrupted, indicating the phenotypic consequences of ELP4 disruption alone.

The genetic model of disease described here is clearly not monogenic: in 14/24 patients the ELP4 CNVs were inherited (13 unrelated events due to the ASD sister pair); four occurred de novo and six were of unknown inheritance. The phenotypic status of most parents is unknown and therefore a precise estimation of penetrance will require further segregation studies. However, presuming that the majority of the parents are unaffected, these inherited CNVs are unlikely to cause a phenotype by reduced expression from haploinsufficiency alone. It is most likely that an interacting model of disease is in action and screening of the second allele of ELP4 and its regulatory regions can rule out the unmasking of recessive mutations. We note that sequencing of ELP4 exons has failed to find mutations within RE patients (Reinthaler, et al., 2014) and postulate that disruption of the regulatory elements of ELP4 and/or of PAX6 within its introns could be causal in the developmental disorders described here.
A two-hit hypothesis can explain CNVs that are non-syndromic, i.e. those that are associated with variable phenotypes and not always inherited, such as the deletions described here, (Girirajan and Eichler, 2010). One hit may reach a threshold to induce some sub-clinical features and create a sensitized background, onto which the second hit (mutation or second CNV) occurs producing a more severe phenotype. If we assume that these disorders share common neurodevelopmental pathways, the final disease outcome will then differ depending on the combination of genes affected. Interestingly, a sister pair with ASD in our study who shared the same ELP4 microdeletion, also shared a microdeletion of exon 2 of the carnitine biosynthesis enzyme gene TMLHE, on Xq28. Deletions of TMLHE are important in non-dysmorphic autism in male-male multiplex families, although with low penetrance (Celestino-Soper, et al., 2012). However the significance in females is unclear. It is possible that deletion of the only copy of TMLHE is enough of a risk factor for some males to develop ASD, but for females (who normally have two copies of TMLHE), further ‘hits’ are necessary, such as the loss of ELP4 in these sisters. Several other patients also carry a second CNV, as described earlier, but it is unlikely that these specific CNVs contribute to the neurological phenotype. Exome sequencing of the patients without a second causal CNV may uncover coding mutations that would contribute to the developmental burden of ELP4 loss.

The predominance in our datasets of deletions verses duplications is unlikely to be a platform bias as both aCGH and SNP arrays were used. Deletion enrichment could be a consequence of undiagnosed duplications, but as this study was not driven by a particular diagnosis this is unlikely. When CNVs are generated by non-allelic homologous recombination (NAHR) between low-copy repeats, a deletion and reciprocal duplication are generated (Malhotra and Sebat, 2012). A possibility is that the duplications could be selected against due to negative genetic selection, i.e. a lower viability or fecundity of carriers. However, since the breakpoints
for ELP4 CNVs differ between cases and there are no low-copy repeats that could explain the
generation of CNVs, this mechanism is also unlikely. Instead, there is more in common with
deletions seen at NRXN1, which may occur by a mechanism involving inverted repeats of
variable sizes, or a significantly higher AT nucleotide content at the breakpoints, generating a
rearrangement hotspot of genome instability. These non-recurrent breakpoints could be
generated by a non-homologous end joining mechanism of double strand breaks or by
replication errors and may be influenced by the genomic architecture of a region in particular

Examination of the data from the copy number variation morbidity map of developmental
delay (Cooper, et al., 2011) shows four microdeletions (<1Mb) with breakpoints within ELP4
(n=15,767), and five microduplications, Supplementary Information Table 1. All duplication
cases had neurological deficits, and two deletion cases and one duplication had ASD.
However, seven microdeletions of ELP4 were also found in the 8,329 control individuals, one
of which is the WTCCC sample reported here. This increase in frequency of smaller CNVs
among controls compared to all of the other control datasets used in our study, indicates that
they may be due to an artifact from the less dense Illumina arrays used by Cooper et al,
compared to the more rigorous platform and methods used to analyse their cases. Indeed, the
authors commented that their detection power is substantially higher for cases, the reverse of
our study, and will manifest itself as false positive enrichment for CNVs in controls.
However, more information (not publicly available) is needed about the specific arrays used
for each control with an ELP4 deletion, their LRR and BAF images and probe coverage over
ELP4 to draw further conclusions about potential false positives and array bias in their
investigation.
Future work will focus on the functional consequences of the *ELP4* deletions by investigation of expression levels of the gene in these cases. Work with cellular and animal models with *ELP4* deletions will help to cement the role of *ELP4* in neurodevelopment through identification of altered interaction networks and developmental pathways such as neuronal migration, branching and survival.
ACKNOWLEDGEMENTS

We wish to thank Thomy de Ravel at Leuven University Hospitals who provided phenotype information on DECIPHER patients 289275 and 249728. We also wish to thank Bertrand Isidor at CHU Nantes for helpful discussion on the manuscript. This project and the BBGRE database (http://bbgre-dev.iop.kcl.ac.uk) were funded through a strategic partnership of the South London and Maudsley Trust NIHR specialist Biomedical Research Centre and the Guys and St Thomas Trust NIHR comprehensive Biomedical Research Centre. This study also makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the DECIPHER data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the DECIPHER project was provided by the Wellcome Trust. Those who carried out the initial data collection for DECIPHER bear no responsibility for its further analysis.

David Collier is an employee of, and Laura Addis a contractor for, Eli Lilly and Company. The authors gratefully acknowledge funding from: Ali Paris Fund for Landau-Kleffner Syndrome Research and Education (DKP), Waterloo Foundation (DKP; LA), National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London (DKP), Charles Sykes Epilepsy Research Trust (DKP), European Union Marie Curie International Reintegration Award of the Seventh Framework Programme PIRG05-GA-2009-248866 (DKP).

Compliance with Ethical Standards

The BBGRE project was approved by the Cambridgeshire Central Research Ethics Committee. In the UK, DECIPHER has been approved by the Eastern MREC 04/MRE05/50 and the project has also been notified to the Information Commissioner in accordance with the
Data Protection Act. The DECIPHER Ethical Framework is detailed here:

https://decipher.sanger.ac.uk/assets/pdfs/decipher_ethical_framework.pdf

Ethical statements for the AGP and control populations and information on informed consent can be found on their published references. Informed consent was obtained from all individual participants included in this study.
REFERENCES


FIGURE LEGEND

**Figure 1** Deletions (red) and a duplication (blue) identified over the *ELP4-PAX6* locus on 11p13 in 9 patients from the BB-GRE clinical genetic database with neurodevelopmental phenotypes, 6 patients with autism from the AGP and Canadian ASD resource, and 9 patients with neurodevelopmental phenotypes from the DECIPHER database. Hg19 ([http://genome.ucsc.edu/](http://genome.ucsc.edu/)). Tracks showing positions of probes genotyped from the Illumina 1M Single Array, the Affymetrix GenomeWide Human SNP6 Array and the Custom Agilent oligonucleotide array used for BBGRE patients are above the UCSC gene tracks. Alternatively spliced gene transcripts are shown.
Figure 1 Deletions (red) and a duplication (blue) identified over the ELP4-PAX6 locus on 11p13 in 9 patients from the BB-GRE clinical genetic database with neurodevelopmental phenotypes, 6 patients with autism from the AGP and Canadian ASD resource, and 9 patients with neurodevelopmental phenotypes from the DECIPHER database. Hg19 (http://genome.ucsc.edu/). Tracks showing positions of probes genotyped from the Illumina 1M Single Array, the Affymetrix GenomeWide Human SNP6 Array and the Custom Agilent oligonucleotide array used for BBGRE patients are above the UCSC gene tracks. Alternatively spliced gene transcripts are shown.
Table 1. Microdeletions and a microduplication of *ELP4* on Chr11 identified in 9 patients from the BB-GRE clinical genetic database, ([http://bbgre-dev.iop.kcl.ac.uk](http://bbgre-dev.iop.kcl.ac.uk)). AgeAtTest indicates age at arrayCGH testing.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>AgeAtTest</th>
<th>Phenotype</th>
<th>hg19Start</th>
<th>hg19Stop</th>
<th>Size bp</th>
<th>Inheritance</th>
<th>CNV</th>
<th>Other CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>129016</td>
<td>F</td>
<td>3 years</td>
<td>Developmental delay (progressing), microcephaly, poor balance</td>
<td>31,573,422</td>
<td>31,674,789</td>
<td>101,368</td>
<td>Unknown</td>
<td>X1</td>
<td></td>
</tr>
<tr>
<td>119460</td>
<td>M</td>
<td>2 years</td>
<td>Social communication difficulties, speech and language delay</td>
<td>31,561,220</td>
<td>31,625,448</td>
<td>64,229</td>
<td>Maternal</td>
<td>x1</td>
<td>Deletion Chr5:97,302,377-97,380,022. No genes.</td>
</tr>
<tr>
<td>116589</td>
<td>M</td>
<td>3 years</td>
<td>PDD: Social interaction difficulties, language disorder, behaviour problems.</td>
<td>31,584,329</td>
<td>31,642,325</td>
<td>57,997</td>
<td>Maternal</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>108970</td>
<td>M</td>
<td>5 years</td>
<td>Severe cognitive delay (IQ 20-34) speech &amp; language disorder, reading &amp; spelling development disorder, autism spectrum disorder, epilepsy &gt;24 months at age of onset</td>
<td>31,495,260</td>
<td>31,546,276</td>
<td>51,017</td>
<td>Paternal</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>117374</td>
<td>M</td>
<td>20 years</td>
<td>Autism, learning difficulties</td>
<td>31,705,076</td>
<td>31,747,631</td>
<td>42,556</td>
<td>Maternal</td>
<td>x1</td>
<td>Duplication Chr5: 93,197,999-93,424,468. Disrupts FAM172A.</td>
</tr>
<tr>
<td>112601</td>
<td>F</td>
<td>1 year</td>
<td>Developmental delay, speech &amp; language disorder, microcephaly (&lt;5th centile), mild cognitive delay, motor skills development disorder</td>
<td>31,691,270</td>
<td>31,722,740</td>
<td>31,471</td>
<td>Paternal</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Gender</td>
<td>Age</td>
<td>Clinical Manifestations</td>
<td>Chromosome</td>
<td>Genomic Coordinates</td>
<td>Allelic State</td>
<td>Copy Number</td>
<td>Maternal / Parental Origins</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>-------------------------------------------------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>130693</td>
<td>M</td>
<td>5 years</td>
<td>Moderate developmental delay mainly affecting language, emerging autistic traits</td>
<td></td>
<td>31,760,904 - 31,786,914</td>
<td>26,010</td>
<td>De novo x1</td>
<td></td>
<td>Deletion Chr16: 15,048,750-16,305,736 16p13.11 hotspot.</td>
</tr>
<tr>
<td>112031</td>
<td>F</td>
<td>12 years</td>
<td>Developmental delay, hypotonia</td>
<td></td>
<td>31,616,889 - 31,849,574</td>
<td>232,686</td>
<td>Unknown x3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Microdeletions of *ELP4* on Chr11 identified in 9 individuals from the DECIPHER database. Age indicates age at last clinical assessment.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age</th>
<th>Phenotype</th>
<th>hg19Start</th>
<th>hg19Stop</th>
<th>Size bp</th>
<th>Inheritance</th>
<th>CNV</th>
<th>Other CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>265704</td>
<td>M</td>
<td>&lt;1yr</td>
<td>Aniridia. No further information.</td>
<td>31,172,410</td>
<td>31,775,457</td>
<td>603,047</td>
<td>De novo</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>249728</td>
<td>F</td>
<td>24yrs</td>
<td>Rieger anomaly.</td>
<td>31,118,027</td>
<td>31,710,576</td>
<td>592,549</td>
<td>De novo</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>257614</td>
<td>M</td>
<td>7yrs</td>
<td>Epilepsy – partial complex seizures with secondary generalization due to cortical dysplasia, mild developmental delay, ADHD, neurinomas, congenital malformation in left eye, fine motor dyspraxia.</td>
<td>30,991,456</td>
<td>31,564,708</td>
<td>573,252</td>
<td>Parent</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>292869</td>
<td>F</td>
<td>15yrs</td>
<td>Severe intellectual disability, muscle hypotrophy with severe hypotonia and absent gross motor and fine adaptive motor development; no language; severe dysphagia requiring tube feeding; craniofacial abnormalities.</td>
<td>31,597,322</td>
<td>31,802,120</td>
<td>204,798</td>
<td>De novo</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>287341</td>
<td>M</td>
<td>2yrs</td>
<td>Partial aniridia. Currently no signs of neurological impairment.</td>
<td>31,605,859</td>
<td>31,783,590</td>
<td>177,731</td>
<td>Maternal</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>258970</td>
<td>M</td>
<td>4 yrs</td>
<td>Developmental delay, behavioral disturbances, regression of language at 18mo to absent at age 4, pervasive developmental disorder</td>
<td>31,605,859</td>
<td>31,775,457</td>
<td>169,598</td>
<td>Unknown</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>Study ID</td>
<td>Sex</td>
<td>Age</td>
<td>Clinical Features</td>
<td>Genomic Positions</td>
<td>Analysis</td>
<td>Relationship</td>
<td>Abnormality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>261471</td>
<td>M</td>
<td>4yrs</td>
<td>Behavioral and speech disorders, mild mental retardation.</td>
<td>31,625,389 - 31,775,457</td>
<td>Parent</td>
<td>x1</td>
<td>Deletion Chr10:35360169-35605506 disrupting CUL2, CREM, CCNY.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>289275</td>
<td>M</td>
<td>24yrs</td>
<td>Aniridia, global developmental delay, autistic behaviour.</td>
<td>31,742,075 - 31,870,603</td>
<td>Unknown</td>
<td>x1</td>
<td>Duplication ChrX:46389227-46396390 disrupting intron of ZNF674.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>270752</td>
<td>F</td>
<td>9yrs</td>
<td>Aniridia, congenital cataract. Mild developmental delay due to processing speed deficiencies largely due to visual impairment.</td>
<td>31,735,689 - 31,825,698</td>
<td>Paternal</td>
<td>x1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Microdeletions of *ELP4* on Chr11 identified in 3 patients from the Autism Genome Project and 3 from a Canadian study of autism. * sibling sister pair.

<table>
<thead>
<tr>
<th>Autism Study ID</th>
<th>Sex</th>
<th>Phenotype</th>
<th>hg19Start</th>
<th>hg19Stop</th>
<th>Size bp</th>
<th>Inheritance</th>
<th>CNV</th>
<th>Other CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>3617_3</td>
<td>M</td>
<td>Strict autism; no seizures, verbal, language delay; delayed first words (at 24 mo), delayed first phrases (at 36 mo); verbal IQ &gt;70.</td>
<td>31460506</td>
<td>31655108</td>
<td>194,602</td>
<td>Paternal</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>NA0285</td>
<td>M</td>
<td>Autism, language delay; delayed first words (at 32 mo), no seizures</td>
<td>31518924</td>
<td>31649475</td>
<td>130,551</td>
<td>Maternal</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>8596_201</td>
<td>M</td>
<td>Strict autism, high functioning; no seizures, verbal, language delay; delayed first words (at 25 mo), typical first phrases (at 25 mo); verbal IQ &gt;70.</td>
<td>31488890</td>
<td>31607986</td>
<td>119,096</td>
<td>Maternal</td>
<td>x1</td>
<td>Duplication, chr9:115994263-116495631 disrupting <em>COL27A1</em></td>
</tr>
<tr>
<td>MM1259-003*</td>
<td>F</td>
<td>Autism, language delay; delayed first words (at 21 mo), mild developmental delay, motor delay, no seizures</td>
<td>31652219</td>
<td>31764393</td>
<td>112,174</td>
<td>Unknown</td>
<td>x1</td>
<td>Deletion chrX:154772341-154775951 disrupting <em>TMHLE</em></td>
</tr>
<tr>
<td>MM1259-004*</td>
<td>F</td>
<td>Autism, language delay; delayed first words (at 18 mo), delayed first phrases (at 36 mo), expressive language problems, mild developmental delay, motor delay, no seizures</td>
<td>31652219</td>
<td>31764393</td>
<td>112,174</td>
<td>Unknown</td>
<td>x1</td>
<td>Deletion chrX:154772341-154775951 disrupting <em>TMHLE</em></td>
</tr>
<tr>
<td>20130_6005001</td>
<td>M</td>
<td>Strict autism; no seizures, verbal, language delay; delayed first words (at 36 mo), delayed first phrases (at 48 mo); verbal IQ &gt;70, coordination problems</td>
<td>31576768</td>
<td>31653568</td>
<td>76,800</td>
<td>Maternal</td>
<td>x1</td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 1** Microdeletions and duplications disrupting *ELP4* reported in cases and controls from the copy number variation morbidity map of developmental delay (Cooper et al., 2011).

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Variant ID</th>
<th>hg19Start</th>
<th>hg19Stop</th>
<th>Size bp</th>
<th>CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9890931</td>
<td>nsv540979</td>
<td>31,118,026</td>
<td>31,790,388</td>
<td>672,363</td>
<td>x3</td>
</tr>
<tr>
<td>9882508</td>
<td>nsv540980</td>
<td>31,152,003</td>
<td>31,751,699</td>
<td>599,697</td>
<td>x1</td>
</tr>
<tr>
<td>9896715</td>
<td>nsv540981</td>
<td>31,401,095</td>
<td>31,656,511</td>
<td>255,417</td>
<td>x1</td>
</tr>
<tr>
<td>9908285</td>
<td>nsv540986</td>
<td>31,747,371</td>
<td>32,063,394</td>
<td>315,979</td>
<td>x3</td>
</tr>
<tr>
<td>9883063</td>
<td>nsv540987</td>
<td>31,775,599</td>
<td>31,804,354</td>
<td>28,756</td>
<td>x1</td>
</tr>
<tr>
<td>9882119</td>
<td>nsv540982</td>
<td>31,602,061</td>
<td>31,875,238</td>
<td>273,178</td>
<td>x3</td>
</tr>
<tr>
<td>9899815</td>
<td>nsv540983</td>
<td>31,656,450</td>
<td>31,857,797</td>
<td>201,348</td>
<td>x3</td>
</tr>
<tr>
<td>9883029</td>
<td>nsv540984</td>
<td>31,703,100</td>
<td>31,751,699</td>
<td>48,600</td>
<td>x1</td>
</tr>
<tr>
<td>9889844</td>
<td>nsv540985</td>
<td>31,703,100</td>
<td>31,929,503</td>
<td>226,404</td>
<td>x3</td>
</tr>
<tr>
<td><strong>Controls:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTCCC</td>
<td>nsv553949</td>
<td>31,317,835</td>
<td>31,539,587</td>
<td>221,753</td>
<td>x1</td>
</tr>
<tr>
<td>nsv553989</td>
<td>31,727,232</td>
<td>31,796,560</td>
<td>69,329</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>HGDP00106</td>
<td>nsv5553987</td>
<td>31,488,890</td>
<td>31,696,336</td>
<td>207,447</td>
<td>x1</td>
</tr>
<tr>
<td>nsv553986</td>
<td>31,488,890</td>
<td>31,679,048</td>
<td>190,159</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>nsv553988</td>
<td>31,535,355</td>
<td>31,651,189</td>
<td>115,835</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>nsv553984</td>
<td>31,454,975</td>
<td>31,654,406</td>
<td>199,432</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>nsv553985</td>
<td>31,488,890</td>
<td>31,654,406</td>
<td>165,517</td>
<td>x1</td>
<td></td>
</tr>
</tbody>
</table>