

# Title Page

## Characterisation of *Cdkl5* Transcript Isoforms in Rat

Ralph D. Hector <sup>a</sup>, Owen Dando <sup>b,c,d</sup>, Tuula E. Ritakari <sup>a</sup>, Peter C. Kind <sup>b,c,d</sup>, Mark E.S. Bailey <sup>e</sup>, Stuart R. Cobb <sup>a\*</sup>

<sup>a</sup> Institute of Neuroscience and Psychology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

<sup>b</sup> Patrick Wild Centre, University of Edinburgh, Edinburgh, UK

<sup>c</sup> Centre for Integrative Physiology, University of Edinburgh, Edinburgh, UK

<sup>d</sup> Centre for Brain Development and Repair, The Institute for Stem Cell Biology and Regenerative Medicine, Bangalore, India

<sup>e</sup> School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, UK

Author email addresses: Ralph D. Hector (ralph.hector@glasgow.ac.uk); Owen Dando (owen.dando@ed.ac.uk); Tuula E. Ritakari (tuula.ritakari@gmail.com); Peter C. Kind (pkind@exseed.ed.ac.uk); Mark E.S. Bailey (Mark.Bailey@glasgow.ac.uk)

\* Corresponding author at Institute of Neuroscience and Psychology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. Email:

[stuart.cobb@glasgow.ac.uk](mailto:stuart.cobb@glasgow.ac.uk).

Abbreviations: CDKL5, cyclin-dependent kinase-like 5; UTR, untranslated region; KO, knockout; RACE, rapid amplification of cDNA ends; TSS, transcription start site.

# Abstract

CDKL5 deficiency is a severe neurological disorder caused by mutations in the X-linked Cyclin-Dependent Kinase-Like 5 gene (*CDKL5*). The predominant human *CDKL5* brain isoform is a 9.7 kb transcript comprised of 18 exons with a large 6.6 kb 3'-untranslated region (UTR). Mammalian models of CDKL5 disorder are currently limited to mouse, and little is known about *Cdkl5* in other organisms used to model neurodevelopmental disorders, such as rat. In this study we characterise, both bioinformatically and experimentally, the rat *Cdkl5* gene structure and its associated transcript isoforms. New exonic regions, splice sites and UTRs are described, confirming the presence of four distinct transcript isoforms. The predominant isoform in the brain, which we name *rCdkl5\_1*, is orthologous to the human *hCDKL5\_1* and mouse *mCdkl5\_1* isoforms and is the most highly expressed isoform across all brain regions tested. This updated gene model of *Cdkl5* in rat provides a framework for studies into its protein products and provides a reference for the development of molecular therapies for testing in rat models of CDKL5 disorder.

## Keywords:

CDKL5

CDKL5 disorder

Neurological

Neurodevelopmental

Expression

# 1. Introduction

CDKL5 deficiency (Mendelian Inheritance in Man, MIM: 300203) is a rare, X-linked genetic disorder that results in early onset seizures and neurodevelopmental impairment. A range of phenotypes are associated with the disorder, and are a result of mutations in the Cyclin-Dependent Kinase-Like 5 gene, *CDKL5*, located on the X-chromosome (Xp22.13) (Kalscheuer et al., 2003; Fehr et al., 2013). Early-onset seizures are common, and phenotypes may also include other features, such as stereotypic hand movements, severe psychomotor impairment, general hypotonia and visual impairments (Kalscheuer et al., 2003; Tao et al., 2004; Artuso et al., 2010; Intusoma et al., 2011; Moseley et al., 2012). Many of these features, such as impaired motor control, altered behaviour, abnormal eye tracking and general hypoactivity are recapitulated in the two *Cdkl5* knockout (KO) mouse models that have been reported (Wang et al., 2012; Amendola et al., 2014). However, spontaneous infantile seizures, a defining feature of the disorder in patients, are not observed in mice. Rats offer an alternative model in the study of neurodevelopmental disorders, and recent work using a rat model of Fragile X Syndrome has allowed a number of key aspects of the disorder to be investigated, such as complex cognitive and social functions (Till et al., 2015). These biological and technical advantages, coupled with the fact that mice fail to recapitulate the spontaneous seizures that define CDKL5 disorder, suggests that the development of another rodent and larger species models of CDKL5 disorder may be beneficial to the field. In advance of this however, it is important to assess differences in the *Cdkl5* gene between species.

In a recent study we characterised the structure of the human *CDKL5* and mouse *Cdkl5* genes, identifying a number of transcript isoforms resulting from alternative splicing (Hector et al., 2016). In the rat, two different transcript isoforms have been identified to date; CDKL5a and CDKL5b (GenBank accession numbers FJ807484 and GU351881,

respectively) (Chen et al., 2010). Although these transcripts were thought to be expressed primarily in the brain, neither is orthologous to the predominant brain isoform of *CDKL5* in human and mouse, a large 9.7 kb transcript comprised of 18 exons with a large 6.6 kb 3'-UTR. At present the rat *Cdkl5* gene structure, transcript set and protein isoforms are not fully understood. In this study we sought to characterise the gene structure of *Cdkl5* in rat using bioinformatic analyses and molecular methods, to identify all transcript isoforms and the predicted suite of resultant protein isoforms, and to compare rat *Cdkl5* with its human and mouse orthologues.

## 2. Materials and Methods

### *2.1 RNA-seq data analysis*

Bioinformatic and molecular analyses were performed as described previously to investigate the transcript isoforms of rat *Cdkl5* (Hector et al., 2016). In brief, RNA-seq datasets were analysed (rat brain GEO sample ID's GSM1020666, GSM1020675 and GSM1020684; rat testis GEO sample ID's GSM1020674, GSM1020683, GSM1020692 (Merkin et al., 2012)) and mapped to the rat genome using the STAR read aligner, version 2.4.2a.

### *2.2 RNA isolation*

Rat total RNA was isolated from tissues obtained from 3 month old wild-type male Sprague Dawley rats using the RNeasy Lipid Tissue Kit (Qiagen). Animal samples were collected in accordance with the European Communities Council Directive (86/609/EEC) and with the terms of a project license under the UK Scientific Procedures Act (1986). The quality and quantity of isolated RNA was analysed using the RNA 6000 kit on a 2100 Bioanalyzer (Agilent).

### *2.3 RT-PCR and Quantitative RT-PCR*

Total RNA was reverse transcribed using Superscript III (Life Technologies), according to the manufacturer's protocol, in 20 µl reactions containing 200 ng of RNA template and 1 µM random hexamers. Reactions were then incubated with 2 units of RNaseH (Life Technologies) at 37°C for 20 min). End-point PCR was performed using Maxima Hot Start Green (Thermo Scientific), according to the manufacturer's protocol, in 50 µl reactions containing 500 nM gene-specific primers, and products visualised on agarose gels. PCR was performed under the following cycling conditions: an initial denaturation at 95°C for 2

min, then 35 cycles of 95°C for 30 s, 58°C for 30 s and 72°C for 60 s, followed by a final extension of 72°C for 7 min. SYBR Green PCR reactions were carried out using the PerfeCTa SYBR kit (Quanta Bioscience), according to the manufacturer's protocol, in 20 µl reactions using 1/20th of the first-strand synthesis reaction and 300 nM gene-specific primers. PCR was performed under the following cycling conditions on an Mx3005P thermocycler (Agilent Technologies): an initial denaturation at 95°C for 30 s, then 40 cycles of 95°C for 10 s and 60°C for 60 s, followed by a dissociation curve. Appropriate controls were included as recommended by the MIQE guidelines. The primers used in this study are provided in Figure S1.

#### *2.4 Rapid amplification of cDNA ends (RACE)*

Rapid Amplification of cDNA Ends was performed using the 3'-RACE & 5'-RACE System for Rapid Amplification of cDNA ends (Life Technologies), according to the manufacturer's protocol. PCR amplicons were cloned using the TOPO TA Cloning Kit (Life Technologies), individual colonies were grown in L-Broth, and plasmids were purified using the PureYield Plasmid Miniprep System (Promega) and then sequenced (Source BioScience). Sequence comparisons were performed using BLASTN (<http://blast.ncbi.nlm.nih.gov/>).

## 3. Results

### 3.1. Rat *Cdkl5* transcript isoforms

In order to identify all *Cdkl5* splicing events, rat tissue-specific RNA-seq datasets were analysed and sensitive alignment tools were used to detect reads mapping across all potential splice junctions. To validate these exon boundaries and splice junctions experimentally, RT-PCR was used to generate products spanning multiple exons and PCR products sequenced. Using this approach, 23 discrete exons were detected, including two novel exons at the 5' end of the gene, 1a and 1b (Fig. 1). Exon boundaries and cryptic splice sites were confirmed, and the composition of each specific isoform was elucidated, so isoform-specific RT-PCR assays could be designed (design details are provided in Fig. S1). Combined, the RNA-seq and RT-PCR data confirmed the existence of four distinct transcript isoforms containing distinct protein-coding regions (Fig. 1A, Table 1). In accordance with new nomenclature recommendations (Hector et al., 2016), we have named these *rCdkl5\_1*, *rCdkl5\_2*, *rCdkl5\_9*, and *rCdkl5\_10*. The composition of each transcript, the exon boundaries and chromosomal sequence coordinates are detailed in Table 2.

RNA-seq analysis also predicted two rare transcripts containing additional exons that were detected and confirmed by sequencing of RT-PCR amplicons (data not shown). Mammalian *Cdkl5* is unusual in having an embedded gene, *Gja6*, located within one of its introns and transcribed in the same direction, and one of these additional *rCdkl5* transcripts incorporates a 125 bp portion of exon 2 of *Gja6*. The second rare transcript incorporates a 117 bp exon, located in *rCdkl5* intron 12, which is predicted to be part of a Long Interspersed Nuclear Element-1 (LINE-1) retrotransposon (analysis was carried out using BLASTN and RepeatMasker: <http://www.repeatmasker.org/>). As such, it was concluded that both of these rare transcripts resulted from aberrant splicing events and they were excluded from further

analysis. In addition, no rat orthologues of *hCDKL5\_3* and *hCDKL5\_4* were identified. The cryptic splice donor site in exon 11 used by these human isoforms, and present in the rat genomic sequence, is absent in mouse due to a single nucleotide difference. Despite its presence in rat, however, we found no evidence, either by RNA-seq analysis or by RT-PCR, that it is ever utilised.

### 3.2. Expression of *Cdkl5* in adult tissues and across brain regions

*rCdkl5\_1* and *rCdkl5\_2* are the major isoforms detected in the rat brain and are orthologous to human and mouse isoforms *\_1* and *\_2*, respectively (Fig. 1B, Table 1). RT-PCR and quantitative RT-PCR were performed to assess transcript expression across a range of tissues. *rCdkl5\_1* is expressed most abundantly within the brain, but is also detected in all of the adult tissues tested, showing a similar pattern of expression to that of the mouse, and human (Fig. 2) (Hector et al., 2016). *rCdkl5\_2*, which incorporates the highly conserved exon 17, is also expressed primarily in the brain, but can only be detected at lower levels in some other peripheral tissues (Fig. 2). This also reflects the pattern of expression seen in mouse and human (Hector et al., 2016). Of the two isoforms expressed in the brain, *rCdkl5\_1* is the major isoform, as detected by qRT-PCR (Fig. 3) and RNA-seq. *rCdkl5\_1* is expressed at different levels across different brain regions, with the highest levels detected in the neocortex (Fig. 3). In contrast, *rCdkl5\_2* displayed lower but consistent levels of expression across the different brain regions, with modestly elevated levels only in the neocortex (Fig. 3). Analysis by qRT-PCR indicates that levels of *rCdkl5\_2* constitute approximately 20% of *Cdkl5* transcripts in the cerebellum and brain stem, but only 5% in the hippocampus, neocortex and striatum (Fig. 3). Analysis of read count data in RNA-seq datasets estimates that transcripts incorporating exon 17 (*rCdkl5\_2*) constitute approximately 10% of *Cdkl5* transcripts expressed in whole brain (Fig. S2).

### 3.3. Analysis of 5' and 3' regions of *Cdkl5* transcripts



The coding regions of *rCdkl5\_9* and *rCdkl5\_10* are not orthologous to human or mouse isoforms, due to the use of species-specific exons at the 3' end of the gene (exons 20 and 21; Fig. 1B). The composition of the 3'-ends of these transcripts had been reported in a previous study by Chen et al. (Chen et al., 2010) (where they were known as CDKL5a and CDKL5b), and are confirmed in this study by 3'-RACE (Fig. 4A). This results in 3'-UTRs of 208 and 162 bases, respectively. 3'-RACE was also used to characterise the 3'-UTRs of *rCdkl5\_1* and *rCdkl5\_2*. Both of these transcripts were shown to use the same canonical polyadenylation signal identified in human and mouse (Hector et al., 2016), located 6.8 kb downstream of the stop codon in exon 19 (Fig. 4A). This yields transcripts of approximately 9.9 and 10 kb in the rat, respectively. A minority of transcripts in the brain were also found to use a more proximal, non-canonical (TATAAA) polyadenylation signal which lies 23 bp upstream of a polyadenylation site at position 35726228 (RGSC 6.0/rn6 genomic reference sequence coordinates). This results in a 3'-UTR of 843 bases.

At the 5' end of the gene, 5'-RACE was used to characterise the TSSs and their contributions to transcript isoforms *rCdkl5\_1* and *\_2* (Fig. 4B). The major TSS (mapped to position 35536407 at the 5'-end of exon 1; rn6, as above) is the same as that identified in orthologous human and mouse transcripts (Hector et al., 2016), resulting in a highly conserved 5'-UTR of approximately 260 bases. Isoforms *rCdkl5\_9* and *\_10*, which are predominantly expressed in testis in the adult, use alternative first exons, primarily exon 1a but also 1b (situated approximately 15 kb downstream of exon 1a) (Fig. 4B). Exon 1a is orthologous to mouse exon 1a and highly conserved; exon 1b, however, is not and evidence from BLAST analysis suggests this to be novel and specific to rat. For exon 1a, a number of TSSs were identified, clustering between positions 35554473 and 35554494 (rn6), yielding transcripts with a 5'-UTR of approximately 360 bases. The major TSS in exon 1b mapped to position 35570401 (rn6), yielding a 5'-UTR of 223 bases. Although three different initial exons are utilised (1, 1a and 1b), all transcripts subsequently splice from these exons to

exon 2. This results in a 167 base region of 5'-UTR, immediately upstream of the ATG start codon, that is conserved in all rat *Cdk15* transcripts (Fig. 4B).

## 4. Discussion

In this study we demonstrate a greater diversity in rat *Cdkl5* transcripts than previously realised by identifying and validating all splicing events that occur across a diverse panel of tissues. After elucidating all common and rare splicing events, the majority of the rat *Cdkl5* coding region was found to be orthologous and well-conserved with human and mouse, particularly with mouse (Fig. 1). The diversity arises from alternative first exon usage and alternative splicing to species-specific exons at the 3' end of the gene (Fig. 1).

In total, four different coding transcript isoforms in rat *Cdkl5* were identified and characterised, and named accordingly with the new nomenclature proposed in our previous study (Hector et al., 2016). The predominant brain isoform is *rCdkl5\_1*, which is orthologous with the human and mouse isoforms *hCDKL5\_1* and *mCdkl5\_1*, respectively. The TSS, exon composition, 3'-UTR, polyadenylation signal and site are all well conserved. This contrasts with previous reports, where CDKL5a and CDKL5b were thought to be the major brain isoforms in rat (Chen et al., 2010). The anti-Cdkl5 antibody used in that study is able to detect all four isoforms identified in this study, so it is possible that the isoforms detected in the brain by western blot were *rCdkl5\_1* and *rCdkl5\_2* and not *rCdkl5\_9* and *rCdkl5\_10*. Our analysis indicates that *rCdkl5\_9* and *rCdkl5\_10* are in fact predominantly expressed in the testis, in a similar pattern to *hCDKL5\_5* in human, and *mCdkl5\_6*, *mCdkl5\_7* and *mCdkl5\_8* in mouse. The C-terminal exons in *rCdkl5\_9* and *rCdkl5\_10* (named exons 20 and 21) are not clearly orthologous to any conserved human, mouse or other mammalian genomic sequence downstream of the conserved exon 19, according to analysis using BLAST, and it appears that they are specific to rat. The presence of exon 17 in *rCdkl5\_2* accounts for the only difference between this isoform and *rCdkl5\_1*. The inclusion of this exon adds 41 amino acids to the protein, but although this exon is extremely well conserved between human, mouse and rat, it contains no identified functional elements.

It is apparent that the putative catalytic region of CDKL5, a highly conserved serine/threonine kinase domain in the N-terminal half of the protein, is preserved in all known CDKL5 isoforms in rat, mouse and human (Hector et al., 2016). The large 3'-UTR (> 6.6 kb) of the major brain isoforms, encoded by exon 19, is also well conserved between human, rat and mouse. Although a number of conserved putative miRNA binding sites in the 3'-UTR can be identified using bioinformatics tools, little else is known about its function. Long 3'-UTRs have been associated with the dendritic trafficking of mRNAs, such as *BDNF* (An et al., 2008), and it is possible that the long 3'-UTR of *Cdkl5* plays a similar function, as mouse *Cdkl5* mRNA has been predicted to have a role in local Cdkl5 synthesis at the dendrites in the adult brain (La Montanara et al., 2015).

Species-specific diversity is found in the C-terminal region of CDKL5, in coding regions from exons downstream of the highly conserved exon 19. These exons are found in regions of low genomic sequence conservation between rat, mouse and human, and are present only in transcript isoforms that are expressed predominantly in testis (Fig. 1b). This C-terminal region was originally implicated in the regulation of CDKL5 subcellular distribution (Rusconi et al., 2008); however, more recent studies have downplayed the role of these exons in pathogenicity, and therefore function (Diebold et al., 2014).

Characterising the complexity of *Cdkl5* transcripts in more species will aid the understanding of CDKL5 biology in novel animal models as well as facilitating the development of effective, isoform-specific antibodies, the lack of which hampers current research. Careful consideration of this transcript complexity will also be required to inform the development of gene- and protein-based therapies for CDKL5 disorders.

## **Acknowledgements**

N/A

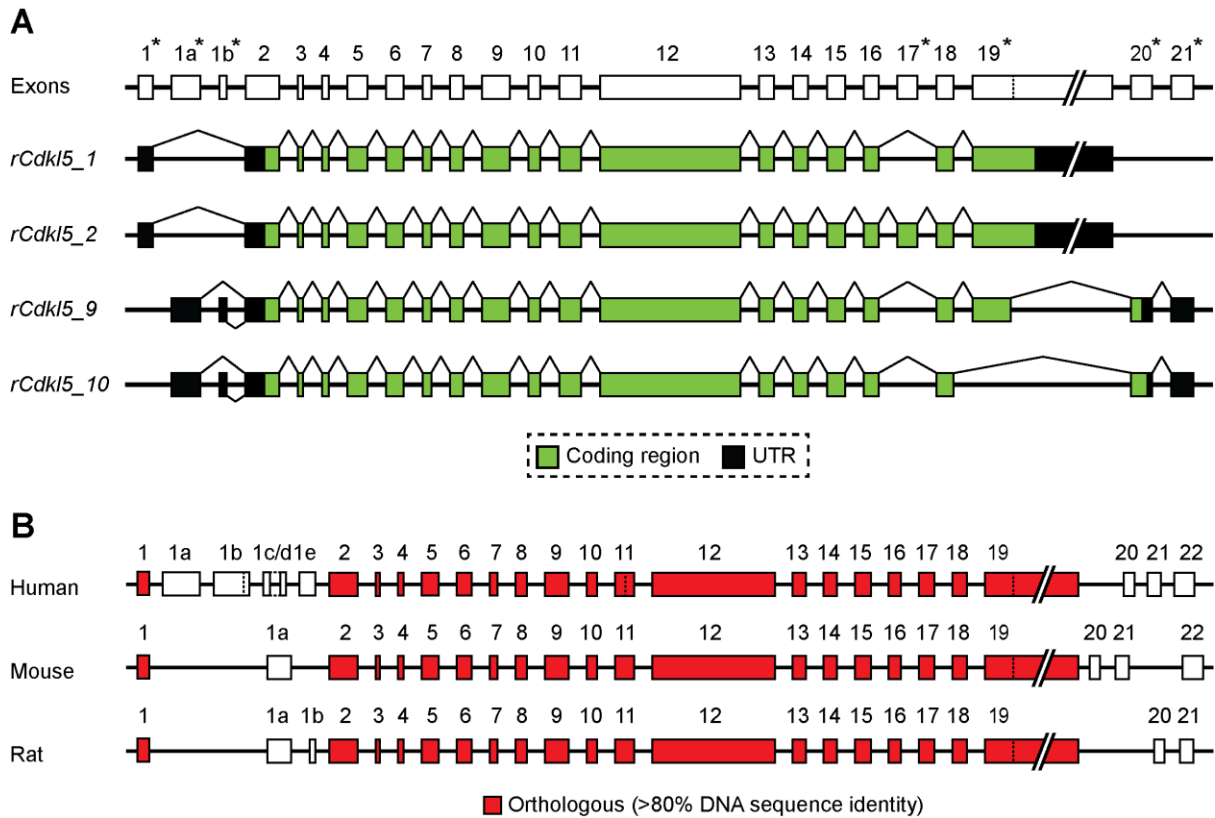
## **Funding Information**

This work was supported by the Chief Scientist Office of the Scottish Government (grant ETM/334 provided funding to SRC and MESB) and CDKL5 UK.

## References

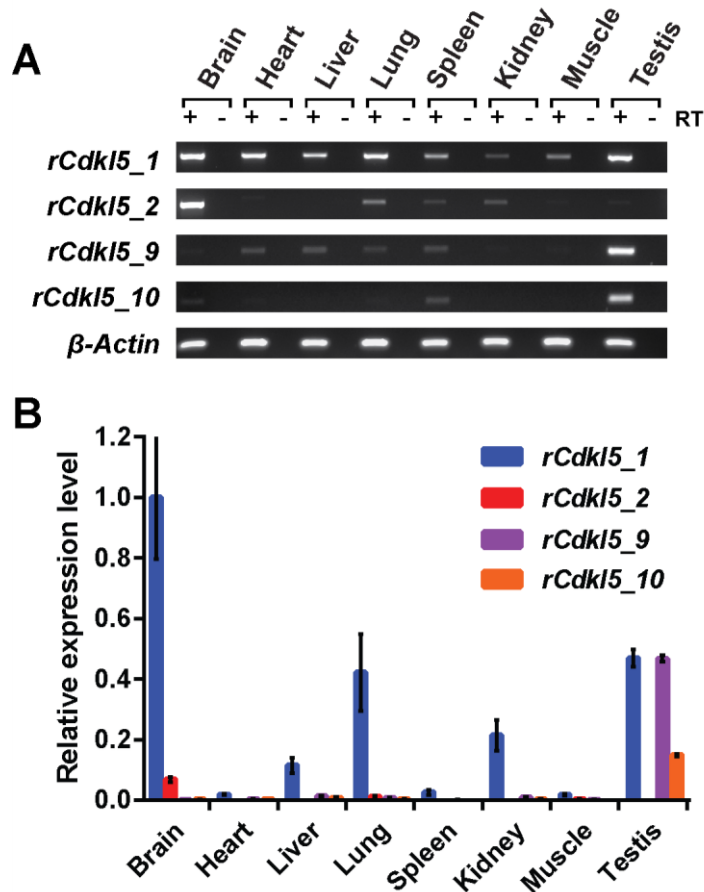
- Amendola, E., Zhan, Y., Mattucci, C., Castroflorio, E., Calcagno, E., Fuchs, C., Lonetti, G., Silingardi, D., Vyssotski, A.L., Farley, D., Ciani, E., Pizzorusso, T., Giustetto, M. and Gross, C.T., 2014. Mapping pathological phenotypes in a mouse model of CDKL5 disorder. *PLoS One* 9, e91613.
- An, J.J., Gharami, K., Liao, G.Y., Woo, N.H., Lau, A.G., Vanevski, F., Torre, E.R., Jones, K.R., Feng, Y., Lu, B. and Xu, B., 2008. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* 134, 175-87.
- Artuso, R., Mencarelli, M.A., Polli, R., Sartori, S., Ariani, F., Pollazzon, M., Marozza, A., Cilio, M.R., Specchio, N., Vigevano, F., Vecchi, M., Boniver, C., Dalla Bernardina, B., Parmeggiani, A., Buoni, S., Hayek, G., Mari, F., Renieri, A. and Murgia, A., 2010. Early-onset seizure variant of Rett syndrome: definition of the clinical diagnostic criteria. *Brain Dev* 32, 17-24.
- Chen, Q., Zhu, Y.C., Yu, J., Miao, S., Zheng, J., Xu, L., Zhou, Y., Li, D., Zhang, C., Tao, J. and Xiong, Z.Q., 2010. CDKL5, a protein associated with rett syndrome, regulates neuronal morphogenesis via Rac1 signaling. *J Neurosci* 30, 12777-86.
- Diebold, B., Delepine, C., Gataullina, S., Delahaye, A., Nectoux, J. and Bienvenu, T., 2014. Mutations in the C-terminus of CDKL5: proceed with caution. *Eur J Hum Genet* 22, 270-2.
- Fehr, S., Wilson, M., Downs, J., Williams, S., Murgia, A., Sartori, S., Vecchi, M., Ho, G., Polli, R., Psoni, S., Bao, X., de Klerk, N., Leonard, H. and Christodoulou, J., 2013. The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. *Eur J Hum Genet* 21, 266-73.
- Hector, R.D., Dando, O., Landsberger, N., Kilstrup-Nielsen, C., Kind, P.C., Bailey, M.E. and Cobb, S.R., 2016. Characterisation of CDKL5 Transcript Isoforms in Human and Mouse. *PLoS One* 11, e0157758.
- Intusoma, U., Hayeeduereh, F., Plong-On, O., Sripo, T., Vasiknanonte, P., Janjindamai, S., Lusawat, A., Thammongkol, S., Visudtibhan, A. and Limprasert, P., 2011. Mutation screening of the CDKL5 gene in cryptogenic infantile intractable epilepsy and review of clinical sensitivity. *Eur J Paediatr Neurol* 15, 432-8.
- Kalscheuer, V.M., Tao, J., Donnelly, A., Hollway, G., Schwinger, E., Kubart, S., Menzel, C., Hoeltzenbein, M., Tommerup, N., Eyre, H., Harbord, M., Haan, E., Sutherland, G.R., Ropers, H.H. and Gecz, J., 2003. Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. *Am J Hum Genet* 72, 1401-11.
- La Montanara, P., Rusconi, L., Locarno, A., Forti, L., Barbiero, I., Tramarin, M., Chandola, C., Kilstrup-Nielsen, C. and Landsberger, N., 2015. Synaptic Synthesis, Dephosphorylation, and Degradation: A NOVEL PARADIGM FOR AN ACTIVITY-DEPENDENT NEURONAL CONTROL OF CDKL5. *J Biol Chem* 290, 4512-27.
- Merkin, J., Russell, C., Chen, P. and Burge, C.B., 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593-9.
- Moseley, B.D., Dhamija, R., Wirrell, E.C. and Nickels, K.C., 2012. Historic, clinical, and prognostic features of epileptic encephalopathies caused by CDKL5 mutations. *Pediatr Neurol* 46, 101-5.
- Rusconi, L., Salvatoni, L., Giudici, L., Bertani, I., Kilstrup-Nielsen, C., Broccoli, V. and Landsberger, N., 2008. CDKL5 expression is modulated during neuronal development and its subcellular distribution is tightly regulated by the C-terminal tail. *J Biol Chem* 283, 30101-11.
- Tao, J., Van Esch, H., Hagedorn-Greiwe, M., Hoffmann, K., Moser, B., Raynaud, M., Sperner, J., Fryns, J.P., Schwinger, E., Gecz, J., Ropers, H.H. and Kalscheuer, V.M., 2004. Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene

- are associated with severe neurodevelopmental retardation. *Am J Hum Genet* 75, 1149-54.
- Till, S.M., Asiminas, A., Jackson, A.D., Katsanevaki, D., Barnes, S.A., Osterweil, E.K., Bear, M.F., Chattarji, S., Wood, E.R., Wyllie, D.J. and Kind, P.C., 2015. Conserved hippocampal cellular pathophysiology but distinct behavioural deficits in a new rat model of FXS. *Hum Mol Genet* 24, 5977-84.
- Wang, I.T., Allen, M., Goffin, D., Zhu, X., Fairless, A.H., Brodtkin, E.S., Siegel, S.J., Marsh, E.D., Blendy, J.A. and Zhou, Z., 2012. Loss of CDKL5 disrupts kinome profile and event-related potentials leading to autistic-like phenotypes in mice. *Proc Natl Acad Sci U S A* 109, 21516-21.

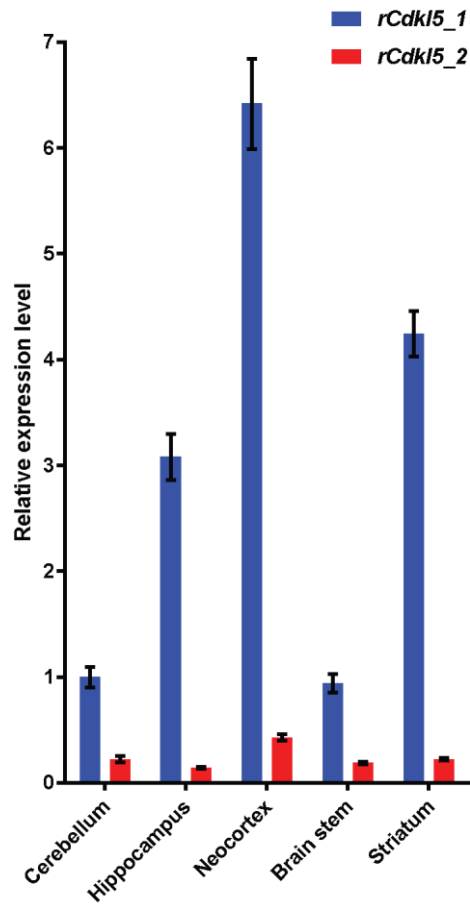


**Fig. 1.** Rat *Cdkl5* gene and transcript isoforms. **A.** Diagram depicting the structure of the rat *Cdkl5* gene and the exon composition of the four different coding isoforms. Angled lines linking exons indicate splicing events. Exon 1a or 1b may act as the initial exon in *rCdkl5\_9* or *rCdkl5\_10*, but these transcripts are otherwise identical. Asterisks next to exon numbers indicate where differences are found between different transcript isoforms. Dotted lines within exons indicate alternative splice sites. Introns and the 3'-UTR portion of exon 19 are not drawn to scale. **B.** Comparison of human, mouse and rat *CDKL5* gene structures.

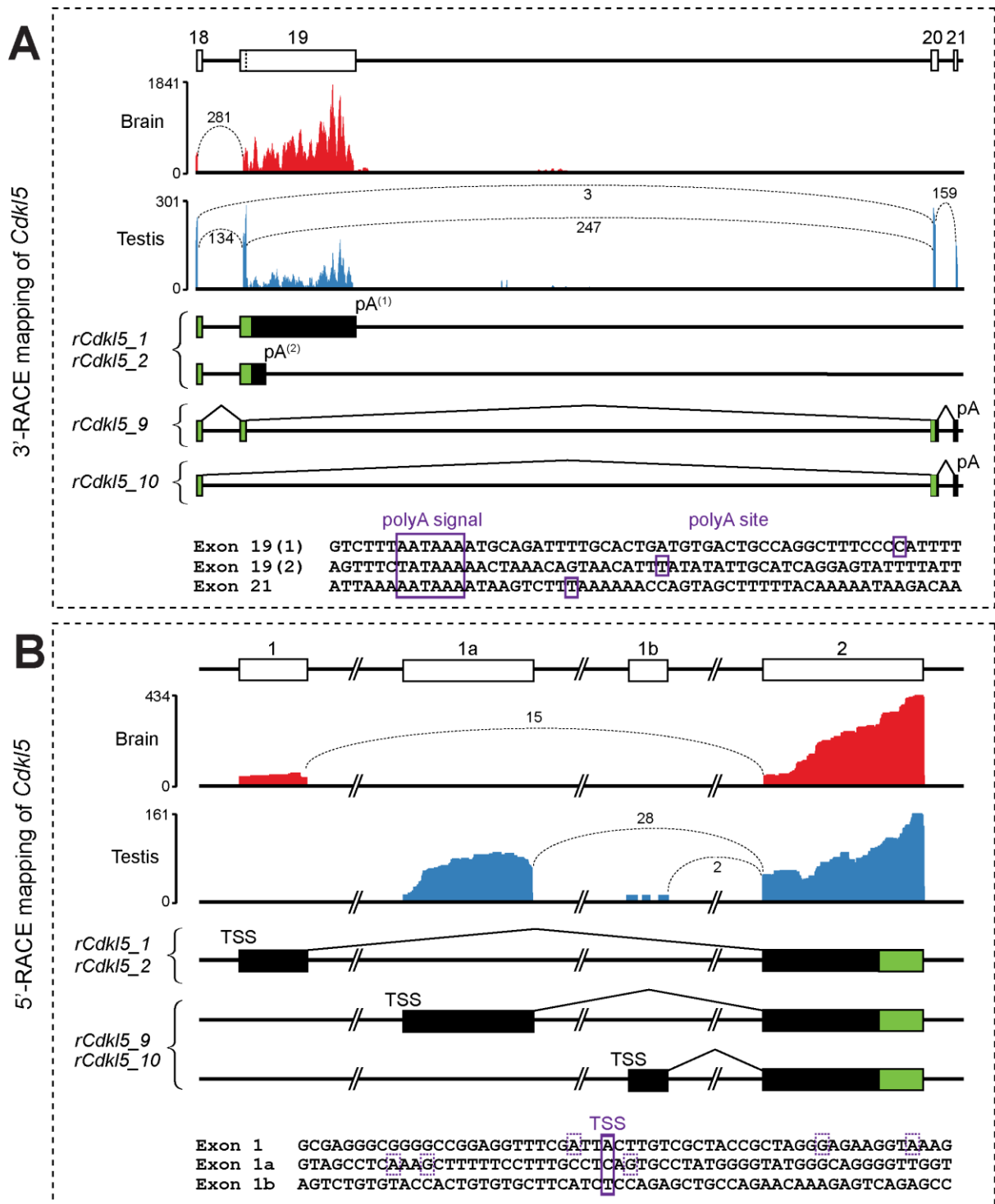




**Fig. 2.** *CDKL5* isoform expression in rat tissues. (A) RT-PCR and (B) quantitative RT-PCR analysis of *Cdkl5* isoforms in a panel of adult tissues. Comparative expression levels of isoforms as revealed by  $2^{-\Delta Ct}$  analysis of quantitative RT-PCR. All values are shown relative to the *rCdkl5\_1* brain sample. All qRT-PCR assays were normalised to *Hprt*. Data plotted are mean  $\pm$  standard error of the mean. Each data point represents two biological replicates analysed as technical triplicates.



**Fig. 3.** *Cdk15* expression levels in brain regions. Comparative expression levels of the two isoforms as revealed by  $2^{-\Delta Ct}$  analysis of quantitative RT-PCR. All values are shown relative to the *rCdk15\_1* cerebellum sample. All qRT-PCR assays were normalised to  $\beta$ -actin. Data plotted are mean  $\pm$  standard error of the mean. Each data point represents three biological replicates analysed as technical triplicates.



**Fig. 4.** RNA-seq and 5' and 3'-RACE mapping of rat *Cdkl5* transcripts. **A.** Upper panels: RNA-seq data from brain (red) and testis (blue) datasets show reads mapping to the 3' end of *Cdkl5*; (the y-axis indicates read count across the analysed region). Indicative numbers of RNA-seq reads spanning each exon junction are also shown, indicated by values and dotted

lines joining exon boundaries. Middle panels: the exon composition and splicing patterns at the 3' end of each rat isoform is shown. Coding regions are indicated by green colouring and 5'-UTRs by black colouring. Lower panel: sequences around each of the three polyadenylation signals and sites (pA) are shown; each was confirmed by sequencing of 3'-RACE mapping. **B.** Upper panels: RNA-seq data from brain and testis datasets show reads mapping to the 5' end of *Cdkl5*; exon boundary-spanning red counts are also shown, as in A, above. Middle panels: boxes representing each exon at the 5' end of the gene are shown, aligned with those in the upper panels. Transcription Start Sites (TSSs) and splice events upstream of exon 2 are indicated. Colouring as in A, above. Lower panel: exonic sequences for each first exon are shown. TSSs, confirmed by sequencing of 5'-RACE products, are indicated by boxes; the major TSS is indicated by a solid box, minor TSSs are indicated by hatched boxes.

Rat isoform	Coding region sizes		Ensembl transcript	Former name	Orthologous isoform		Descriptive name
	(bp)	(aa)			Human	Mouse	
<i>rCdkl5_1</i>	2883	960	-	-	<i>hCDKL5_1</i>	<i>mCdkl5_1</i>	<i>rCdkl5_2-19Δ17</i>
<i>rCdkl5_2</i>	3006	1001	-	-	<i>hCDKL5_2</i>	<i>mCdkl5_2</i>	<i>rCdkl5_2-19</i>
<i>rCdkl5_9</i>	2805	932	Cdkl5-201	CDKL5a	-	-	<i>rCdkl5_2-21Δ17s19</i>
<i>rCdkl5_10</i>	2634	877	Cdkl5-202	CDKL5b	-	-	<i>rCdkl5_2-21Δ17Δ19</i>

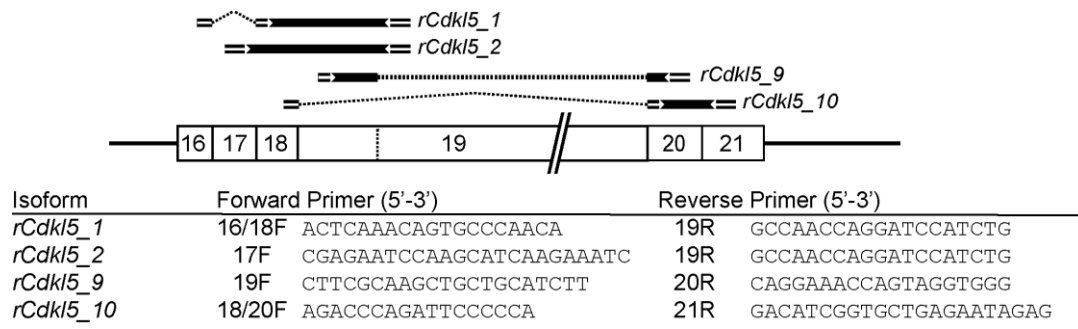
**Table 1.** Summary of *rCdkl5* nomenclature. The new proposed nomenclature of *rCdkl5* isoforms, former names, equivalent Ensembl transcript model and coding region sizes. The descriptive name indicates, in addition to the major coding isoform, the variable exon usage associated with each transcript: Δ indicates that an exon is not used; s indicates that the short form of the exon is used, where an alternative splice site is present.

Rat							
Exon	Size (bp)	X-chromosome coordinates		<i>Cdkl5</i> isoform			
		Exon start	Exon end	<u>1</u>	<u>2</u>	<u>9</u>	<u>10</u>
1	89	<i>35536407</i>	35536495	+	+	+	+
1a <sup>#</sup>	189	<i>35554492</i>	35554680			+	+
1b <sup>§</sup>	56	<i>35570401</i>	35570456			+	+
2	231	<i>35599124</i>	35599354	+	+	+	+
3	35	<i>35602829</i>	35602863	+	+	+	+
4	46	<i>35654903</i>	35654948	+	+	+	+
5	137	<i>35665038</i>	35665174	+	+	+	+
6	121	<i>35669920</i>	35670040	+	+	+	+
7	60	<i>35672087</i>	35672146	+	+	+	+
8	91	<i>35674708</i>	35674798	+	+	+	+
9	190	<i>35678867</i>	35679056	+	+	+	+
10	81	<i>35687604</i>	35687684	+	+	+	+
11	152	<i>35693420</i>	35693571	+	+	+	+
12	967	<i>35699096</i>	35700062	+	+	+	+
13	102	<i>35703854</i>	35703955	+	+	+	+
14	106	<i>35704227</i>	35704332	+	+	+	+
15	124	<i>35709063</i>	35709726	+	+	+	+
16	100	<i>35716514</i>	35716613	+	+	+	+
17 <sup>#</sup>	123	<i>35720481</i>	35720603		+	+	
18	120	<i>35721822</i>	35721941	+	+	+	+
19 <sup>#</sup>	217	<i>35724998</i>	35725214			+	+
	1231	<i>35724998</i>	<u>35726228</u>		+	+	
	7248	<i>35724998</i>	<u>35732245</u>	+	+		
20 <sup>§</sup>	145	<i>35771511</i>	35771655			+	+
21 <sup>§</sup>	155	<i>35773044</i>	<u>35773198</u>			+	+

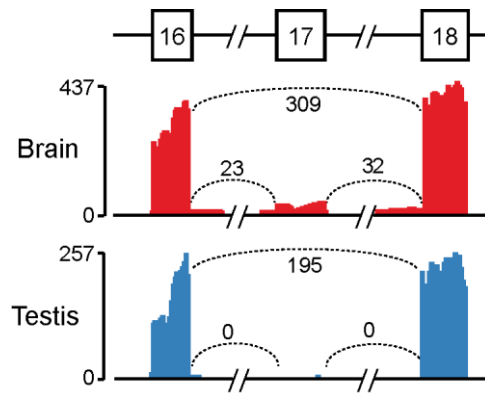
<sup>#</sup> novel exon or UTR, not previously characterised

<sup>§</sup> rat-specific, not orthologous to human or mouse

**Table 2.** Summary of *rCdkl5* transcript isoforms. Exon numbers, sizes and co-ordinates for the rat *Cdkl5* gene (RGSC 6.0/rn6 assembly). Exon composition of each transcript isoform is indicated. Starting coordinates for the main TSS of each initial exon are italicised; coordinates of polyadenylation sites in each terminal exon are underlined. Hatched boxes indicate exons that contain multiple splice sites.



**Fig. S1.** Primer locations and sequences. PCR Primer pairs specific for each *Cdkl5* isoform are listed, and the diagrams indicate the coverage of the resultant amplicons.



**Fig. S2.** RNA-seq data from brain and testis datasets show reads mapping to exons 16, 17 and 18. All reads contributing to these data span a maximum of two exons.