RettBASE: Rett Syndrome Database Update

Rahul Krishnaraj¹, Gladys Ho¹, ²John Christodoulou¹, ², ³, ⁴

¹NSW Centre for Rett Syndrome Research, Western Sydney Genetics Program, Children’s Hospital at Westmead, Sydney, Australia

²Disciplines of Child and Adolescent Child Health and ³Genetic Medicine, Sydney Medical School, University of Sydney, Sydney, Australia

⁴Neurodevelopmental Genomics Research Group, Murdoch Childrens Research Institute and Department of Paediatrics, Melbourne Medical School, University of Melbourne, Melbourne, Australia

Address for correspondence:

Professor John Christodoulou, Neurodevelopmental Genomics Research Group, Murdoch Childrens Research Institute, 50 Flemington Rd, Parkville, Victoria 3052, Australia

Phone: 613 9936 6515

Fax: 613 9348 1391

Email: john.christodoulou@mcri.edu.au

Running title: RettBASE: an update

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/humu.23263.

This article is protected by copyright. All rights reserved.
Abstract
Rett syndrome (RTT) is an X-linked progressive neurodevelopmental disorder that primarily affects females. Mutations in the MECP2 gene have been attributed as the major genetic cause of Rett syndrome. Recently, mutations in CDKL5 and FOXG1 genes have also been suggested to give rise to Rett syndrome, although subsequent more extensive studies suggest that diseases resulting from mutations in these two genes should be considered as distinct clinical entities. While the genetic basis for the Rett syndrome has been recognized, so far there is no effective cure for the disease and the treatments available are mainly aimed at ameliorating clinical problems associated with the disorder. The swift identification of the mutations in children is crucial for pursuing the best therapeutic care. RettBASE was created in 2002 as a MECP2 variant database and has grown to become a comprehensive variant database for Rett syndrome and related clinical phenotypes, containing a curated collection of variants for MECP2, CDKL5 and FOXG1 genes. Here we describe the development and growth of RettBASE after its inception in 2001. Currently, RettBASE holds a total of 4668 variants in MECP2, 498 variants in CDKL5 and 64 variants in FOXG1.

Introduction
Rett syndrome (RTT; OMIM – 312750) is an X-linked progressive neurodevelopmental disorder that affects females with an approximate incidence rate of 1:10,000 (Laurvick et al., 2006). Mutations in the X-linked methyl-CpG-binding protein 2 gene (MECP2; OMIM# 300005) are the major cause of Rett syndrome (Amir et al., 1999, Amano et al., 2000, Kim and Cook, 2000). The phenotypic spectrum of Rett syndrome is highly variable. The most
recent classification has identified typical and variant forms of Rett syndrome (Neul et al., 2010). MECP2 mutations has been attributed to be the single largest genetic cause of RTT phenotypes with recent data suggesting over 95% of classical RTT cases and over 75% of variant RTT cases resulted from mutations in this gene (Cuddapah et al., 2014, Neul et al., 2014). Further studies into this orphan disorder have identified mutations in the X-linked cyclin-dependent kinase-like 5 gene (CDKL5; OMIM – 300203) as a cause for the early-onset seizure variant of Rett syndrome (Grosso et al., 2007, Weaving et al., 2004). Subsequently, mutations in the FOXG1 gene (FOXG1; OMIM – 164874), a DNA-binding transcription factor containing the fork-head-binding domain, have been implicated in the congenital variant of Rett syndrome (Mencarelli et al., 2010). More recently, it has been suggested that individuals with CDKL5 and FOXG1 mutations may represent distinct clinical entities (Kortum et al., 2011, Fehr et al., 2013).

The clinical features of classical Rett syndrome are characterized by a period of apparently normal development, quickly superseded by a period of severe developmental regression leading to diminished intellectual, motor, communication and hand movement skills (Ellaway and Christodoulou, 2001). In the classical variant clinical clues become apparent at 6-18 months, with progressive failure thereafter. In human females, the phenomenon of X-chromosome inactivation, where there is transcriptional silencing of one of the alleles of genes on the X-chromosome (Clerc and Avner, 2006), can influence disease severity in X-linked disorders. This can influence disease severity in carriers of MECP2 mutations, in whom the phenotypic severity may be greatly ameliorated or even absent due to marked skewing of X-inactivation favouring the wildtype allele as the active one. Hemizygous males who are affected by this disease seldom survive past infancy, while for females the severity depends upon the localisation and the nature of the mutations with genetic (Ben Zeev et al., 2009) and epigenetic (Weaving et al., 2003) factors contributing to disease severity. Some cases of males with mutations in MECP2 and a RTT or RTT-like phenotype surviving beyond infancy have been noted (Reichow et al., 2015). In addition, some males may
exhibit a classical RTT phenotype because of X-chromosome aneuploidy (Schwartzman et al., 2001) or somatic mosaicism (Armstrong et al., 2001) or may have a milder phenotype in the context of specific mutations.

The MeCP2 protein is a part of the methyl-CpG-binding domain (MBD) family of proteins and is abundantly expressed in brain. It binds to methylated DNA through its methyl-binding domain (MBD) (Meehan et al., 1992). It is shown to act as a transcriptional regulator of other genes and plays a major role in neuronal development, growth and maturation (Chahrour et al., 2008, Tsujimura et al., 2015). The molecular mechanisms that lead to RTT, although obscure, were thought to be a consequence of dysregulation of neuronal function, with mutations in MECP2 predicted to alter expression of various genes that lead to disruption of the normal neurodevelopmental processes (Ellaway and Christodoulou, 2001). Interestingly, a number of recent studies have shown compelling evidence that mutant glial cells, including astrocytes and microglia, may be significant contributors to the RTT neurological phenotype (Ballas et al., 2009, Lioy et al., 2011, Nguyen et al., 2013).

Human variant databases have played a key role in improving health care for patients. The identification and understanding of the type of the mutation gives valuable information for clinicians in diagnosis and prognosis of the disease. Acquiring data relating to repeat instances of mutations are particularly difficult due to lack of novelty in publishing such variants. However, the significance of these repeat variants cannot be underestimated as these form valuable mutation prevalence data for the disorder, as well as assisting in discerning genotype-phenotype relationships. RettBASE was established with the aim of collecting data relating to all known instances of MECP2 variations, funded by the International Rett Syndrome Association, now known as Rettsyndrome.org (Christodoulou et al., 2003). The database has been running for over 14 years and contains more than 4500 variants pertaining to Rett syndrome and related disorders, systematically collected from researchers, diagnostic laboratories and clinicians and curated. RettBASE not only contains novel variants pertaining to genes associated with Rett syndrome, it keeps the record of any
newly reported repeat instances of the variants including non-pathogenic variants. This paper is a follow up to our initial RettBASE development paper and details the changes to the database, addition of new functionalities and the incorporation of new variants.

Database Features

Database architecture

The database runs on the open-source MySQL database system. This relational database management system is a fast, efficient database manipulation system which works on standard SQL and is easily manoeuvrable and accessible. The website also utilizes the highly powerful PHP scripting language coupled with HTML to provide the end user a fast and user friendly website experience.

Database content

RettBASE (http://mecp2.chw.edu.au/) is hosted at the Kid’s Research Institute, Children’s Hospital, Westmead, Sydney, Australia. Initially, the website stored variant data only on MECP2, which at the time was the major gene implicated with Rett syndrome. Subsequently, CDKL5 and FOXG1 mutants were also shown to be responsible for Rett syndrome and related phenotypes, and the website expanded to incorporate these variants as well.

As of 5th June 2016 the database holds 4668 entries for MECP2, 498 entries for CDKL5, 64 entries for FOXG1. The variants have been identified from systematic searches in published reports of all available English-speaking journals and are manually curated before uploading to the database. We are also constantly in touch with clinicians and researchers who generate quality variant data, which are unpublished for a variety of reasons, asking them to contribute to the database. Finally, submissions have come from our local diagnostic laboratory and international diagnostic laboratory collaborators also involved in Rett syndrome screening, generating potentially useful variant data which are also integrated to the database. The database is continuously updated to include data on Rett syndrome.
variants in an effort to keep RettBASE a useful platform for users to extract information on this debilitating disease.

A web form in downloadable Microsoft Word™ format or Adobe PDF™ format is available from the website for researchers, diagnostic laboratories or clinicians wishing to submit any unpublished data to RettBASE. Alternately, if there is a need for bulk submission of variant data, there is an Excel spreadsheet that is formatted to assist in such an activity. The data can be submitted either through email to one of the curators or send through a secure fax line.

To prevent the incorporation of duplicates in the database, as much information as allowable pertaining to the variant is collected and recorded. Variants directly submitted are given a unique patient identifier encrypted code, generated using an algorithm that converts personal identification information into a de-identified code specific to a patient. This encrypted code is generated through an Excel spreadsheet containing the required algorithm, easily accessible from the RettBASE website, enabling submitters to submit variant data without compromising on patient anonymity. This process also prevents the submission of the same variant again to the database at a later time and minimizes redundancy of the data. The variant data is also protected behind the internal firewall and only curators are allowed access to modify or manipulate the primary data.

**Database functionality**

Users can view variant data on the MECP2, CDKL5 and FOXG1 genes. Users can either choose to search and query specific variants from one of these databases or view the total variant collection in each of the databases. For searching of specific variants, users are able to educe information on the nucleotide and amino acid change, the type of mutation, the affected protein domain, the broad patient phenotype, as well as the predicted pathogenicity of the observed variant. Useful additional information, such as the links to the publication
where the variant was first reported, the familial carrier status of the individual variant, chromosomal studies carried out and also SNP IDs where available can also be obtained from the databases.

**MECP2 database**

A “simple search” selection is available in the MECP2 database, which enables users to search the database based on few basic fields. Also, the simple search allows users to query MECP2 variant data based on only one of the search criteria. The fields available in this selection for users to query include nucleotide change, amino acid change, type of sequence change (silent, missense, nonsense etc.), domain change location, Rett syndrome phenotype classification (classical, atypical) and also on the gender of the proband submitted. An “advanced search” is also available in the MECP2 database, where users can query variants with a flexible selection of fields and also customize the displayed results. The advanced search option allows for displaying results based on any combination of fields selected and also contains additional fields not available in “simple search” that point to the specific publication where appropriate, the predicted pathogenicity of the variant, the X-inactivation data and mutation entry date.

The “table of all entries” reports all the variants submitted to the MECP2 database, displayed in a table format for the perusal of users (Fig 1). The “graph of all mutations” generates the frequency of mutations at each amino acid in the MECP2 protein in a column-based graphical format, where the height of each column equates to the number of instances of variations at that particular amino acid along the length of the protein. Moving the mouse over a column reveals information about variations occurring at that amino acid. The information includes the wild type amino acid, the amino acid residue number, a mutation description either indicating the new amino acid or describing the mutation in full, and the frequency of the mutation. In cases where different mutations occur at the same location, each mutation and their respective frequencies are shown, separated by a slash.
The graph can also be rescaled according to user requirements. Users are also able to view graphs of selected types of mutations as well as on the whole set of mutations. The “Summary of mutations” extracts summarized variant information relating to the MECP2 variants. Users can view the “summary by frequency”, which groups the mutations by frequency in descending order, with the most frequent mutation first and the rarest towards the end of the table. Users are also able to determine the most common protein mutations associated with the disease by selecting “summary by amino acid”. Additional more general information related to Rett syndrome is also available for users in the useful information table in the website. These are a collection of resources that users of RettBASE may find of interest, such as information relating to reference sequences for Rett syndrome-related genes and guidelines for mutation nomenclature.

**CDKL5 & FOXG1 databases**

The CDKL5 and FOXG1 databases have been developed recently and are based on a similar structural framework with the exception of the removal of X-chromosome inactivation data for FOXG1, as FOXG1 is an autosomal gene. Their search and query functionalities are slightly different from the MECP2 database. There is an additional functionality within the CDKL5 and FOXG1 databases, where users can search using two formats; “variant” or “proband”. The variant section holds data on unique variations submitted to the database while the probands section holds data on all the instances of the mutations. The “proband” style search is comparable to the existing MECP2 RettBASE search functions, and includes basic information (such as phenotype, familial/control screening results, X-inactivation results, etc.) (Fig.2). The “variant” search on the other hand allows users to view only the information pertaining to the variant, and may be useful for those interested in the pathogenicity classification or functional effects of a particular variant. Users are able to view the “Table of variants” and “Table of probands”, which displays the whole list of variants and probands stored in the database respectively. In addition to a table display, users may also
view the complete comprehensive data for any particular variant/proband in a form view. This information is accessible when users click on the cDNA nomenclature (for the variant form) or the proband ID number (for the proband form) in any of the results in the tabular format. This feature of the website allows users to access all of the information available on a specific variant. The variant id, the systematic name, the protein name, alternate name, the type of mutation, the domain affected, the predicted pathogenicity of variants, the dbSNP ID and also references to the publication on the variant are displayed. There is also a comments section which the curator may use to report additional data such as in silico analysis, pathogenicity classification information or any useful notes associated with the particular variant.

Database Curation

Nomenclature of variants

The variants in the FOXG1 and CDKL5 databases have been designated with nomenclature based on the most recent HGVS guidelines (http://www.hgvs.org/mutnomen/)(den Dunnen and Antonarakis, 2001). Being an older database, the MECP2 database still follows slightly older nomenclature standards, however efforts are ongoing to update these variants to reflect the latest recommended guidelines for nomenclature. One of the tools used to check the variant nomenclature as recommended in the HGVS guidelines is Mutalyzer (https://mutalyzer.nl/) (Wildeman et al., 2008). We also routinely submit all the variants in the database through the batch checker functionality in Mutalyzer in order to minimize nomenclature errors.

MECP2 has two alternatively spliced isoforms, MECP2_e1 and MECP2_e2. MECP2_e1 is formed with the exon 1,3 and 4 with the start codon in exon 1 and encodes a protein of 498 amino acids. MECP2_e2 has the start codon in exon 2 and contains the exons 1, 2, 3 and 4 that encodes a 486 amino acid long protein. The MECP2_e2 isoform is the most
characterised isoform, although MECP2_e1 is more abundant in human brain (Mnatzakanian et al., 2004). Due to historical reasons, MECP2 variants in RettBASE are named according to MECP2_e2, except for those directly affecting the e1 isoform (which includes the prefix MECP2_e1). Plans are also under way to expand the MECP2 database to describe variants according to both isoforms. Table 1 presents the details on the genomic sequences of the MECP2, CDKL5 and FOXG1 genes and their corresponding cDNA and protein sequence information. Two RefSeq transcripts are available for CDKL5. As they encode the same protein, only one transcript is listed in the table.

Pathogenicity of variants

All the variants that are submitted to RettBASE are assigned a pathogenic classification. The variants in the CDKL5 and FOXG1 database are classified based on the 5 class pathogenic classification: C5 = pathogenic; C4 = likely pathogenic; C3 = variant of uncertain significance (VOUS); C2 = likely benign; C1 = benign. The variants in the MECP2 database still follow the old convention where the variants are classified into one among ‘Mutation associated with disease’, ‘Polymorphism not causing disease’, ‘Silent polymorphism’ and ‘Unknown’.

The variants are scored based on a variety of factors which contribute to the final pathogenicity classification. There is no absolute rule in terms of classification, but it is often determined based on any number of supporting pieces of available evidence. Any available in vitro and in vivo studies of the variant are highly useful in predicting the potential pathogenicity of the variant. The nature of the mutation, the domain affected, the evolutionary conservation of the canonical amino acid, the prediction of the likely effect of the protein change on function and the frequency of the allele variant in the general population also contribute towards the final classification. Patient specific details such as whether the variant was de novo or inherited and any X-inactivation studies are also collected and taken into account to classify the pathogenicity of the variant. In silico analysis is used to evaluate functional effects of the variants, and aids in supporting the classification.
of the variants with higher confidence. All the variants are subjected to analysis with these tools before they are submitted to RettBASE. SIFT (Ng and Henikoff, 2001) (http://sift.jcvi.org/) is a tool that can predict the pathogenicity of missense variants. It predicts if the variant affects protein function and assigns a score for the variant. Amino acid changes with probabilities less than 0.5 are predicted to be deleterious. Tools such as Polyphen-2 (Adzhubei et al., 2010) (http://genetics.bwh.harvard.edu/pph2/) and MutationTaster (Schwarz et al., 2014) (http://www.mutationtaster.org/) are also used to elicit valuable information on the variant pathogenicity, especially missense variants, as these are the most difficult to assign with high confidence. These tools predict if the substitution is possibly damaging and disease causing respectively. Splicing prediction tools are used to generate reports on the likelihood of intronic variants to affect splicing. NNSplice (http://www.fruitfly.org/seq_tools/splice.html), MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) and Human Splicing Finder (http://www.umd.be/HSF/) are splicing predictors each having their own advantages and disadvantages.

Overview of Database Content

Over 4600 MECP2 variations have been recorded in RettBASE. These variations are seen distributed throughout the MECP2 coding region and all major types of mutations are observed (Fig 3). The majority of mutations are point mutations (missense and nonsense) which constitute ~65% of the total mutations. Frameshift insertions/deletions make up 14%, while the rest are comprised of silent, intronic, in-frame insertions/deletions and others as described in Fig 3. Most of the pathogenic mutations are located within the methyl binding domain (MBD; 25%) and the transcriptional repression domain (TRD; 34%) (Fig 4.a). Eight major mutations account for ~47% of all the mutations in MECP2. These mutations are p.Arg106Trp, p.Arg133Cys, p.Thr158Met, Arg168*, p.Arg255*, p.Arg270*, p.Arg294* and p.Arg306Cys, which are distributed between the MBD and TRD domain (Fig. 4.b).
Furthermore, the analysis of disease causing mutations in MECP2 database shows that 71% of all the variations (including repeat instances) in MECP2 are pathogenic (Fig.5). Notably, 98% of the mutations falling in the nuclear localisation signal (NLS) domain are considered pathogenic. Similarly a high proportion of variants seen in the TRD (88%) and MBD (84%) have been assigned pathogenic status, while only (52%) variations in the C-terminal region are noted as pathogenic. Like MECP2, CDKL5 is highly expressed in human brain and has important functions in neuronal development (Mari et al., 2005). It consists of 20 exons and codes for a 1030 a.a. long protein. Up until now just over 160 pathogenic CDKL5 variations have been recorded in RettBASE. Most pathogenic variants in CDKL5 were frameshift insertion and/or deletion mutations. Although missense variants represent a large proportion of variants in CDKL5, only 27% of these missense variations are considered to be pathogenic or likely pathogenic variants. It has to be noted here that there will be a time bias influencing the number of pathogenic variants associated with MECP2 compared to CDKL5 and FOXG1. Early MECP2 variants in RettBASE were classified based on the pathogenicity guidelines available at the time, which were not as strict as the guidelines that are currently applied for classifying variants. Variants associated with CDKL5 and FOXG1 on the other hand, being later additions to the database, were scrutinized with stricter rules relating to pathogenicity compared to early MECP2 variants. However there are plans in progress to update the MECP2 database to reflect the latest guidelines and also to improve the structure of the database. Details on the distribution of sequence variations in the RettBASE categorized by their clinical phenotype are reported in Table 2.

Discussion

From the inception of RettBASE in 2002, the main aim of the database was to provide a platform to record and report all the available variations related to MECP2 gene. With the recent association of the CDKL5 and FOXG1 with Rett syndrome and related phenotypes, the database was upgraded to report these variants as well. Each variant is formatted and
deposited in RettBASE such that maximum information regarding the variant is extracted, either from the published documents or through correspondence with submitters. We also take utmost effort towards standardizing the variants according to recommended guidelines for ease of reporting and sharing data.

Being a comprehensive Rett syndrome variant repository, RettBASE forms a platform to not only extract RTT variant information but also identify useful functional data on the variants. There is substantial evidence that the phenotypic severity of patients suffering from Rett syndrome caused by MECP2 mutations depends on the nature and the location of the mutation. For instance, the relatively common p.Arg133Cys mutation has been shown to be of least severity compared to other common MECP2 mutations. The nonsense mutations p.Arg168* and p.Arg255* seen in young children are observed to be amongst the most clinically severe (Pidcock et al., 2016, Neul et al., 2008). Early truncating and large deletion mutations in MECP2 gene have been implicated in delayed early development and inability to walk. Nonsense mutations p.Arg255* and p.Arg270* are seen in a large proportion of cases in children, leading to impairment in mobility, hand movement and early development. Of the missense mutations, p.Arg106Try has been reported to be very severe (Neul et al., 2008), potentially due to its reduced binding affinity for methylated cytosines (Kudo et al., 2001).

It is not surprising that individuals with severe mutations are clinically diagnosed earlier, whilst less severe mutations are recognised much later. In addition, the early diagnosis of the atypical variants of Rett syndrome are generally difficult as clinical symptoms may be quite subtle initially (Bisgaard et al., 2015). An early diagnosis is particularly important because it gets patients off a potentially long and arduous diagnostic treadmill, restores reproductive confidence and potentially gives patients access to the supportive therapy tailored to their needs. In addition, a genetic diagnosis will provide opportunities for patients to gain access to targeted therapies as they become available.
An interesting observation relating to MECP2 mutations is that although the frequency of mutations observed in the NLS domain is lower compared to those in MBD and TRD, they lead to very severe phenotypes, emphasizing the importance of the NLS domain for MECP2 functionality. On the other hand, C-terminal mutations affecting the MECP2 gene tend to show milder phenotypes, indicating that the loss of C-terminal is not as critical for the gene compared to upstream areas.

One of the factors hypothesised to influence the clinical phenotype in Rett syndrome patients is X-chromosome inactivation (XCI) (Archer et al., 2007). Generally, an increased proportion of the mutant allele being active would be predicted to result in a more severe clinical phenotype. Our database holds data on XCI status of individual patients (where available) and although these data may not be a standalone accurate predictor of prognosis of patients, further clarity on the phenotype-genotype relation of the variant may be possible. Patients with CDKL5 mutations show a clinical profile different to Rett syndrome, with in the main severe neurological problems within the first few months of birth (Fehr et al., 2013).

Phenotypic studies of mutations affecting CDKL5 have described that mutations seen along exon 2-12 are more severe (Castren et al., 2011). The catalytic domain is a major conserved domain in the CDKL5 protein and a large number of pathogenic missense mutations are seen in this domain, which suggests a crucial role of the kinase domain in the normal neurodevelopmental processes. Although mutations leading to exonic deletions in CDKL5 are generally considered pathogenic, variations in exons 19-21 of are suggested to be non-pathogenic or at most mildly pathogenic mutations (Ho et al., 2012, Psoni et al., 2010).

FOXG1 is a 489 amino acid long protein which acts as a DNA-transcription factor, and plays a major role in development of the telencephalon in humans. It is an autosomal gene and almost all pathogenic mutations affecting FOXG1 have been de novo cases. Analysis of FOXG1 mutations has also shown that over 90% of identified mutations are a combination of frameshift, insertion/deletion and point mutations. There were 32 reported single nucleotide variations that predominantly seem to affect the forkhead-binding domain leading to
Another interesting inference from the mutations affecting the FOXG1 gene is that there is a predominance of frameshift mutations observed within the early part of the protein (1-180 aa), while there is a higher prevalence of missense mutations in the fork-head binding domain and towards the C-terminal region of the protein. Due to the smaller sample size of variants reported for FOXG1, a distinct correlation of genotype-phenotype relationships has been difficult.

Phenotype-genotype correlations are likely to be highly useful in the evaluation of efficacy of therapeutics in Rett syndrome. More informed genetic counselling can be provided to families through analysis of the phenotypic outcome of the mutation. RettBASE holds data that can reveal the pathogenic severity and previous clinical instances of the variation. This data can be utilized to identify trends in patient disease prognosis, especially the likely phenotypic profile and severity. While the clinical severity of an individual patient may be different from RTT, a case of best and worst case scenarios of severity for specific mutations can be predicted from the data.

**Conclusion**

There is no definitive mode of treatment available for Rett syndrome sufferers. Being a rare disorder, any information assisting the diagnosis of the disease is vital for the clinical and research communities. This locus-specific database provides a comprehensive collection of variant data on the three principal genes associated with Rett syndrome and related phenotypes, and has proved to be a useful knowledgebase for clinicians, researchers, patients and families of patients. As a result of the careful curation of the database, we are confident that accurate prevalence data for disease-causing variants of Rett syndrome can be provided and we hope discerning this information will continue to be beneficial in the understanding of the pathogenesis of Rett syndrome. Moreover, through continual addition of data and systematic improvements to the database, we hope to ensure the most accurate
and up-to-date genetic and functional data on RTT variants are made available to clinicians and laboratorians in a timely and clinically useful manner.

Acknowledgements

The curators at RettBASE would like to thank all the support given by the patients and their families who were instrumental in the establishment of and contributions to the database. We would also like to thank our previous curators Dr Gladys Ho, Andrew Grimm, as well as Dr Sarah Williamson, Dr Linda Weaving and Prof Bruce Bennett for their assistance in the curation process. The construction, maintenance and upgrade of RettBASE have been made possible by the very generous financial support of the RettSyndrome.org. Finally, we would also like to convey our appreciation to all the contributors to the database, some of who are listed in our RettBASE contributor's page.

References


ARCHER, H., EVANS, J., LEONARD, H., COLVIN, L., RAVINE, D., CHRISTODOULOU, J., WILLIAMSON, S., CHARMAN, T., BAILEY, M. E.,


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


TSUJIMURA, K., IRIE, K., NAKASHIMA, H., EGASHIRA, Y., FUKAO, Y., FUJIWARA, M., ITOH, M., UESAKA, M., IMAMURA, T., NAKAHATA, Y., YAMASHITA, Y., ABE, T., TAKAMORI, S. & NAKASHIMA, K. 2015. miR-199a Links MeCP2 with...


Figure Legends:

Figure 1: MECP2 database website search functionality, summary of data and additional information.

The selections available for users in the MECP2 database. Querying of variants is possible through a simple search or advanced search. Summary information of all the mutations can be generated in the summary section.
Figure 2: Search and display functionality in CDKL5 database.

The search option allows users to query the variants based on several criteria. This particular search example looks at pathogenic missense variants in CDKL5 and displays the 15 entries associated with the query. The result is viewed in tables and there is an option to customize the fields of output to display.
### CDKL5 Variants Search Results

<table>
<thead>
<tr>
<th>cDNA name</th>
<th>Protein name</th>
<th>Mutation type</th>
<th>Pathogenicity class</th>
<th>First cited reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.125A&gt;G</td>
<td>p.Lys42Arg</td>
<td>missense</td>
<td>pathogenic variant</td>
<td>Directly submitted</td>
</tr>
</tbody>
</table>

Displaying 15 entries
There are 4668 variant entries in the MECP2 database. The major variants observed were missense mutations (1779) followed by nonsense mutations (1286). The CDKL5 database currently contains 498 proband entries of which 233 are point mutations. In addition, there were a total of 64 FOXG1 proband entries in RettBASE.
**Figure 4: Overview of MECP2 variants and domains associated.**

a) The distribution of variants in the MECP2 gene is shown above. The major domains of the gene are listed in the horizontal axis and the number of variants is plotted in the vertical axis. The total variants affecting each domain are numerically marked above the column. MBD – Methyl Binding Domain; TRD - Transcription Repression Domain; NLS – Nuclear Localization Signal; b) The most commonly observed variants in the MECP2 gene and their proportion in the total variants are highlighted. The major domains of the gene are shown in the horizontal axis. The numbers 1–487 refers to the amino acid distribution in the MECP2 gene.
This article is protected by copyright. All rights reserved.
Figure 5: Schematic illustration of distribution of pathogenicity of variants in MECP2, CDKL5 and FOXG1 genes.

The graphs detail the Rett variants in MECP2, CDKL5 and FOXG1 genes (including repeat instances) according to their pathogenicity.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Genomic sequence</th>
<th>cDNA sequence</th>
<th>Protein sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MECP2 (isoform e2)</td>
<td>AF030876</td>
<td>NM_004992</td>
<td>NP_004983</td>
</tr>
<tr>
<td>MECP2 (isoform e1)</td>
<td>NG_007107</td>
<td>NM_001110792</td>
<td>NP_001104262</td>
</tr>
<tr>
<td>CDKL5</td>
<td>NG_008475</td>
<td>NM_003159</td>
<td>NP_003150</td>
</tr>
<tr>
<td>FOXG1</td>
<td>NG_009367</td>
<td>NM_005249</td>
<td>NP_005240</td>
</tr>
</tbody>
</table>
Table 2: RettBASE variants and clinical phenotypes associated with them (as of June 2016)

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Number of entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rett syndrome – classical</td>
<td>1582</td>
</tr>
<tr>
<td>Rett syndrome – atypical</td>
<td>307</td>
</tr>
<tr>
<td>Rett syndrome – congenital onset</td>
<td>27</td>
</tr>
<tr>
<td>Rett syndrome – forme fruste</td>
<td>28</td>
</tr>
<tr>
<td>Rett syndrome – male variant</td>
<td>26</td>
</tr>
<tr>
<td>Rett syndrome – preserved speech</td>
<td>51</td>
</tr>
<tr>
<td>Rett syndrome – not certain</td>
<td>1521</td>
</tr>
<tr>
<td>Angelman syndrome</td>
<td>14</td>
</tr>
<tr>
<td>Non Rett syndrome control</td>
<td>20</td>
</tr>
<tr>
<td>Non-progressive encephalopathy of neonatal onset</td>
<td>1</td>
</tr>
<tr>
<td>Progressive encephalopathy of neonatal onset</td>
<td>14</td>
</tr>
<tr>
<td>Sporadic mental retardation</td>
<td>51</td>
</tr>
<tr>
<td>X-linked mental retardation</td>
<td>50</td>
</tr>
<tr>
<td>Unaffected family member</td>
<td>257</td>
</tr>
<tr>
<td>Autism only</td>
<td>27</td>
</tr>
<tr>
<td>Not known</td>
<td>13</td>
</tr>
<tr>
<td>MECP2</td>
<td>Useful Information</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>Simple search</td>
<td>Reference sequence information</td>
</tr>
<tr>
<td>Advanced search</td>
<td>Deletions causing Rett syndrome + related disorders</td>
</tr>
<tr>
<td>Table of all entries</td>
<td>General resources</td>
</tr>
<tr>
<td>Graph of all mutations</td>
<td>Submitting mutations to RettBASE</td>
</tr>
<tr>
<td>Summary of MECP2 mutations:</td>
<td>Contributors</td>
</tr>
<tr>
<td>By frequency</td>
<td>References for MECP2 mutations</td>
</tr>
<tr>
<td>By nucleotide</td>
<td></td>
</tr>
<tr>
<td>By amino acid</td>
<td></td>
</tr>
</tbody>
</table>
# CDKL5 Variants Search Results

<table>
<thead>
<tr>
<th>cDNA name</th>
<th>Protein name</th>
<th>Mutation type</th>
<th>Pathogenicity class</th>
<th>First cited reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.125A&gt;G</td>
<td>p.Lys42Arg</td>
<td>missense</td>
<td>pathogenic variant</td>
<td>Directly submitted</td>
</tr>
</tbody>
</table>

Displaying 15 entries
CDKL5 Variants Search Results

<table>
<thead>
<tr>
<th>cDNA name</th>
<th>Protein name</th>
<th>Mutation type</th>
<th>Pathogenicity class</th>
<th>First cited reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.125A&gt;G</td>
<td>p.Lys42Arg</td>
<td>missense</td>
<td>pathogenic variant</td>
<td>Directly submitted</td>
</tr>
</tbody>
</table>
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Krishnaraj, R; Ho, G; Christodoulou, J

Title:
RettBASE: Rett syndrome database update

Date:
2017-08-01

Citation:

Persistent Link:
http://hdl.handle.net/11343/293011