

Reviewed Preprint

v1 • December 29, 2025

Not revised

Reviewed Preprint

v2 • July 14, 2026

Revised by authors

For correspondence:

J.Guy@ed.ac.uk

A.Bird@ed.ac.uk

Author notes: See [page 33](#)

Funding: See [page 32](#)

Reviewing editor: Lisa M Monteggia, Vanderbilt University, United States

© 2025, Guy et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Translational reading frame determines the pathogenicity of C-terminal frameshift deletions in MeCP2: an alternative therapeutic approach

Jacky Guy¹✉, Elena Hein¹, Beatrice Alexander-Howden¹, Timur von Bock und Polach¹, Tricia Mathieson¹, Benjamin P Kleinstiver^{2,3}, Huda Zoghbi^{4,5,6}, Adrian Bird¹✉

¹Institute of Cell Biology, University of Edinburgh, Edinburgh, United Kingdom • ²Center for Genomic Medicine and Department of Pathology, Massachusetts General Hospital, Boston, United States • ³Department of Pathology, Harvard Medical School, Boston, United States • ⁴Department of Molecular and Human Genetics and Neuroscience, Baylor College of Medicine, Houston, United States • ⁵Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, United States • ⁶Howard Hughes Medical Institute, Houston, United States

eLife Assessment

This study offers **important** insight into the pathogenic basis of intragenic frameshift deletions in the carboxy-terminal domain of MECP2, which account for some Rett syndrome cases, yet similar variants also appear in unaffected individuals. Using base editing and mouse models, the authors present **convincing** evidence supporting the pathogenicity of select deletion variants, with potential implications for therapeutic development.

<https://doi.org/10.7554/eLife.109170.2.sa3>

Abstract

Mutations in the *MECP2* gene cause the severe neurological disorder Rett syndrome. A cluster of frameshift-causing C-terminal deletions (CTDs) lead to loss of ~100 amino acids at the C-terminus of the MeCP2 protein, and account for approximately 10% of RTT-causing mutations. The pathogenicity of C-terminal deletions (CTDs) is unexpected, as this C-terminal domain is non-essential in mice. Utilising databases of pathogenic and benign human *MECP2* mutations, we find that some individuals with apparently typical CTDs do not exhibit Rett syndrome, confirming that C-terminal truncations are not intrinsically pathogenic. Using human DNA sequence data and mouse models, we demonstrate that pathogenicity results from a drastic reduction in MeCP2 levels and is determined by the presence of the short amino acid motif proline-proline-stop (-PPX) at the C-terminus, which results from a shift to the +2 reading frame. Individuals with CTDs that shift to the +1 frame avoid this motif and do not develop Rett syndrome. Mutating the stop codon of the PPX motif to tryptophan rescues MeCP2 expression and RTT-like phenotypes in a CTD mouse model. Finally, we demonstrate that an adenine base editor can efficiently introduce this tryptophan substitution in cultured cells. Overall, our findings uncover a simple and reliable prognostic distinction between benign and pathogenic CTDs and provide proof-of-concept for an editing strategy that potentially corrects all disease-causing CTD mutations.

Introduction

Rett syndrome (RTT, OMIM #312750) is a severe neurological disorder caused by heterozygous mutations in the X-linked *MECP2* gene (Amir et al., 1999 [\[1\]](#)). RTT mainly affects girls, with an occurrence of around 1:10,000 births. At approximately 18 months of age normal development stagnates and acquired skills such as speech and purposeful hand use are lost. RTT is a devastating condition, affecting speech, feeding, digestion, breathing, heart rhythm, ambulation, hand use and susceptibility to seizures (Hagberg et al., 1983 [\[2\]](#); Hagberg, 2002 [\[3\]](#)). MeCP2 protein is found in the nucleus of all cell types but is most highly expressed in neurons (Lewis et al., 1992 [\[4\]](#); Ross et al., 2016 [\[5\]](#); Skene et al., 2010 [\[6\]](#)) where it has been implicated in a range of functions, including splicing, microRNA processing, transcriptional activation and repression (Lyst and Bird, 2015 [\[7\]](#)). Analysis of patient mutations and experimental data from mouse models support a role in transcriptional repression (Gabel et al., 2015 [\[8\]](#); Kinde et al., 2016 [\[9\]](#)) although its loss leads to both increases and decreases in gene expression (Bajikar et al., 2025 [\[10\]](#)). Two functional domains have been characterised in MECP2: the methyl-binding domain (MBD), through which MeCP2 binds to 5-methyl cytosine in DNA (Nan et al., 1993 [\[11\]](#)), and an NCoR interacting domain (NID) (Lyst et al., 2013 [\[12\]](#); Nan et al., 1998 [\[13\]](#)). Through the NID, MeCP2 binds to TBL1X/TBL1XR1, subunits of the nuclear corepressor complex NCoR (Kruusvee et al., 2017 [\[14\]](#); Lyst et al., 2013 [\[12\]](#)), indicating that MeCP2 acts as a bridge, bringing the corepressor complex to methylated sites in the genome.

RTT can be caused by a range of inactivating mutations, including large deletions, missense and nonsense point mutations, and frameshifting deletions and insertions resulting in a truncated protein. Missense RTT mutations cluster within the MBD and NID and, with a few exceptions (Guy et al., 2018 [\[15\]](#)), are not found elsewhere in the protein. On the contrary, neutral polymorphisms found in the gnomAD database tend to avoid the two functional domains but are commonly found throughout the rest of the protein. Alpha missense pathogenicity scores correlate with the presence of RTT mutations and absence of gnomAD variants in the *MECP2* coding sequence (Figure 1A [\[16\]](#)). Using knock-in mouse models it has been shown that the parts of MeCP2 outside the MBD and NID are non-essential, resulting in a “minigene”, ΔNIC, that retains MeCP2 function (Tillotson et al., 2017 [\[17\]](#)) (Figure 1A [\[16\]](#)).

This study concerns C-terminal deletion (CTD) mutations, which comprise about 10% of all RTT-causing mutations, comparable to the most common missense and nonsense mutations (Bebbington et al., 2010 [\[18\]](#); Cuddapah et al., 2014 [\[19\]](#); Smeets et al., 2005 [\[20\]](#)). CTDs occur in a ‘deletion-prone region’ (CT-DPR) between approximately c.1110 and c.1210 (numbering according to transcript ENST00000303391.11, NM_004992, e2 isoform), and lead to truncated MeCP2 protein lacking around 100 amino acids from the C-terminus, with a missense tail of variable length (Bebbington et al., 2010 [\[18\]](#)). More than 50 different CTDs can be found in the database of RTT mutations, RettBASE (Christodoulou et al., 2003 [\[21\]](#); Krishnaraj et al., 2017 [\[22\]](#)), ranging in frequency from single occurrences to over 40 individuals with the most common deletions (Figure 1B [\[23\]](#)). Taken as a class, CTDs are at the milder end of the RTT clinical severity spectrum, comparable to the common R306C and R294X point mutations (Bebbington et al., 2010 [\[18\]](#); Cuddapah et al., 2014 [\[19\]](#)). Our previous CTD knock-in mouse models (Figure 1C [\[24\]](#)) revealed that the C-terminal region removed by the frameshift is not required for MeCP2 function, as mice that expressed normal levels of the truncated protein (CTD2) were not detectably different from wildtype. Instead, the disease phenotype results from a drastic reduction in the amount of MeCP2 protein present in the brain, as seen in CTD1 mice (Guy et al., 2018 [\[15\]](#)). Intriguingly, the gnomAD database (Chen et al., 2024 [\[25\]](#); Gudmundsson et al., 2022 [\[26\]](#)), which lists DNA sequence polymorphisms that are unlikely to be severely pathogenic in humans, includes several individuals exhibiting what appear to be typical frameshifting CTDs. Here we uncover the genetic basis of pathogenic and non-pathogenic CTDs by demonstrating that the difference is determined by the amino acid sequence at the extreme C-terminus of the truncated protein, specifically the presence of a –PPX motif. Our findings have important prognostic implications in predicting the clinical outcomes of *MECP2* mutations. Using the knowledge gained from studying both pathogenic and benign CTD mutations we have designed and tested a strategy to edit RTT CTD mutations in the genome using adenine

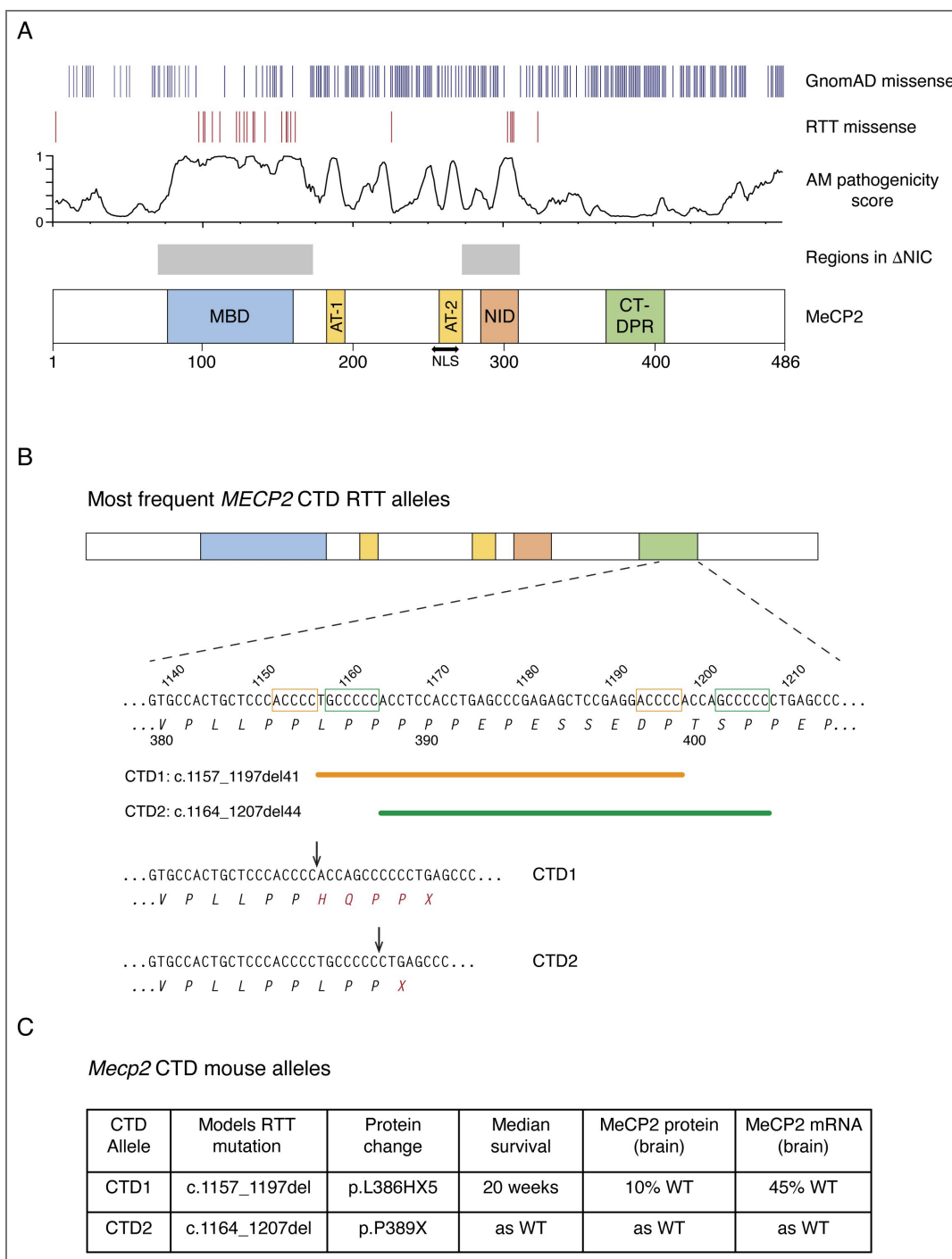


Figure 1. Frameshifting deletions in the C-terminal domain of MeCP2 cause Rett syndrome in humans and RTT-like phenotypes in mice.

(A) A schematic representation of human MeCP2 protein showing hemizygous missense mutations found in gnomAD v4.1.0; *de novo* Classical RTT mutations found in RettBASE; Alpha Missense pathogenicity score; regions found in an MeCP2 “minigene” (Δ NIC) and domains described in the literature: methyl-binding domain (MBD), AT-hooks (AT-1 and AT-2), nuclear localization signal (NLS), NCoR interacting domain (NID) and the C-terminal deletion-prone region (CT-DPR). (B) DNA and amino acid sequence found in the CT-DPR, numbered according to transcript ENST00000303391.11, e2 isoform. The two most common RTT deletions are shown, referred to as CTD1 and 2 for brevity. Microhomologies believed to recombine to cause the deletions are marked on the DNA sequence as orange and green boxes respectively, and the deleted sequences as corresponding lines below. The C-terminal amino acid sequences of CTD1 and CTD2 are shown, with the points of frameshift marked with arrows. (C) Summary of CTD1 and CTD2 knock-in mouse models.

base editing (ABE) technology, changing the C-terminal amino acid sequence and restoring WT levels of MeCP2 protein. This provides a potential therapeutic route which could treat the entirety of this diverse class of mutations using a single ABE/guide RNA combination.

Results

The MeCP2 deletion-prone region shows a high level of neutral variation in the human population

To investigate the effect of C-terminal amino acid sequence on the level of CTD MeCP2 we utilised the wealth of available human mutation data. RTT mutations were collated in RettBASE (Christodoulou et al., 2003 [↗](#); Krishnaraj et al., 2017 [↗](#)). Although RettBASE is no longer publicly available, the data it contained was downloaded and used in this study (Supplementary Tables 1 [↗](#)–4 [↗](#)). RettBASE data has now been incorporated into ClinVar (Landrum et al., 2016 [↗](#)) in an abridged form. In addition, the ExAC and gnomAD databases hold a large set of genetic variants found within the general population (Chen et al., 2024 [↗](#); Lek et al., 2016 [↗](#)). Individuals who contributed to the database are not all without clinical problems, but the inclusion of a variant suggests that it is tolerated and can be classed as neutral. It should be noted that as *MECP2* is an X-linked gene, it is possible that pathogenic variants could be present in healthy females due to extreme skewing of X-inactivation towards that allele. Indeed, family pedigrees with RTT mutations in multiple generations are possible due to this phenomenon (Augenstein et al., 2009 [↗](#); Ravn et al., 2011 [↗](#)) and were used to link mutations in *MECP2* to RTT (Amir et al., 1999 [↗](#)). On the other hand, the presence of a mutation in a hemizygous male in gnomAD strongly indicates that it is a benign variant that will not cause RTT.

GnomAD data shows that missense variants are extremely common in the C-terminal deletion-prone region (CT-DPR). Between p.370 and p.405 (c.1108 and c.1215) almost every amino acid is tolerant of change, many with multiple alternative substitutions (Fig. 2 [↗](#) SF1A, Supplementary Table 5 [↗](#)). Several of these variants are present in hundreds or thousands of individuals, with E397K represented over 5000 times. E397K is so common in the general population that it occurs in the homozygous state in 11 females, and was initially classed as a RTT-causing mutation (Yusufzai and Wolffe, 2000 [↗](#)) due to its recurrent appearance in RTT patients who were later found to have additional pathogenic *MECP2* mutations (Moncla et al., 2002 [↗](#); Wan et al., 1999 [↗](#)). Also of note, many apparently healthy individuals display in-frame deletions that remove segments of the C-terminal domain (Figure 2A [↗](#), Fig. 2 [↗](#) SF1B, Supplementary Table 6 [↗](#)). These findings strongly suggest that the amino acid sequence of this region is not important for human MeCP2 function, in agreement with phenotypic data from a number of mouse models (Guy et al., 2018 [↗](#); Tillotson et al., 2017 [↗](#)). The frequency of benign missense mutations in the CT-DPR contrasts strongly with the MBD, where the protein sequence is known to be critical for MeCP2 function (Fig. 2 [↗](#) SF1C, Supplementary Table 7 [↗](#)). This difference is also apparent in the AlphaMissense pathogenicity scores across these regions of the human protein (Figure 1A [↗](#)).

MeCP2 C-terminal frameshifting deletions may be pathogenic or benign

Intriguingly, a substantial number of individuals in gnomAD, including hemizygous males, have frameshifting deletions that result in truncated MeCP2 protein similar to that found in RTT patients (Supplementary Table 8 [↗](#)). To ensure that pathogenic RTT and non-pathogenic gnomAD mutations were reliably annotated, only “high confidence” sets of mutations that conformed to stringent criteria were selected for further examination. We stipulated that selected RTT mutations should be associated with a diagnosis of classical Rett syndrome and be absent in both parents (*de novo*) in at least one individual with that mutation (Supplementary Table 4 [↗](#)). GnomAD variants were considered when present in at least one hemizygous male, to exclude the possibility of a pathogenic mutation being masked by skewing of X-inactivation in some heterozygous females (Supplementary Table 8 [↗](#)). A comparison of these high confidence sets of benign frameshifts and pathogenic RTT CTD mutations does not reveal any obvious difference in the location of the

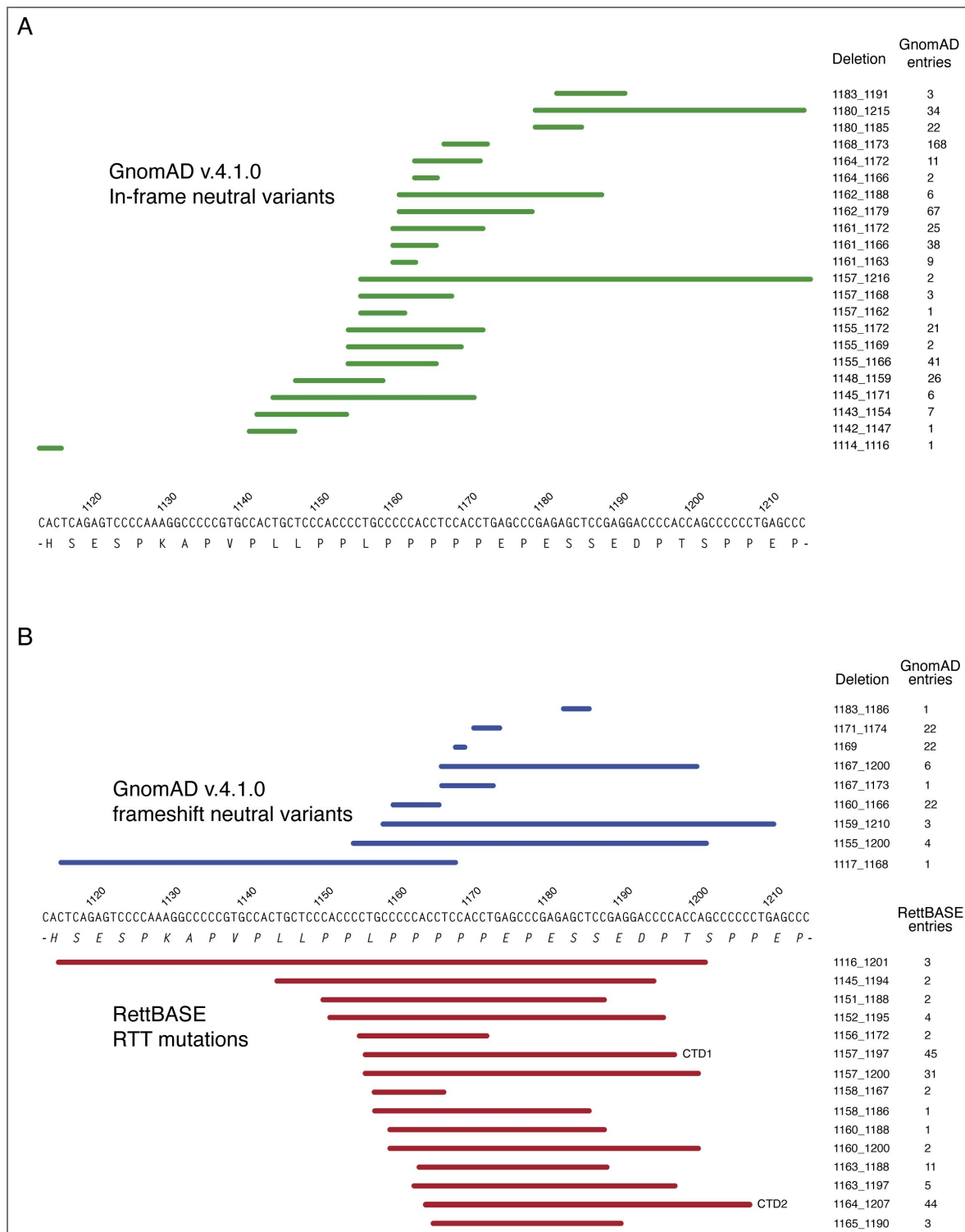


Figure 2. Genomic deletions in the CT-DPR of *MECP2* (high confidence sets).

(A) All hemizygous in-frame deletions in the region found in gnomAD v4.1.0 are indicated with green lines above the DNA sequence. The deletion co-ordinates (numbered according to ENST00000303391.11) and number of individuals with each deletion are shown alongside. (B) Frameshifting deletions in the CT-DPR. Above the genomic sequence, hemizygous deletions from gnomAD v4.1.0 are shown in blue. Deletions from RettBASE are shown below in red, with each deletion found in at least one individual with a *de novo* mutation and a diagnosis of classical RTT. Coordinates and number of individuals with each mutation are shown, with the two most common RTT mutations, CTD1 and CTD2 indicated.

frameshifts, with similar truncated proteins resulting in very different clinical outcomes (Figure 2B). Our previous study using CTD knock-in mice implied that the detailed amino acid sequence close to the C-terminus of CTD mutants can affect the level of MeCP2 (Fig. 2 SF2). We therefore compared the C-termini of our sets of RTT and gnomAD CTD alleles from the deletion site to the stop codon.

We found that all high confidence benign frameshifts ended in the codons –SPRTX due to a shift from the wildtype 0 frame to the +1 frame, caused by a deletion of $3n+1$ nucleotides. Almost all +1 frameshifts in this region generate this ending, regardless of the exact position of the deletion. On the other hand, all but one of the high confidence pathogenic CTDs involved a shift to the +2 frame ($3n+2$ nucleotides deleted) resulting in most cases in the ending –HQPPX (Figure 3A, B). For both frameshifts the sequence prior to the extreme C-terminus varies depending on the size and location of the deletion. By comparing pathogenic with benign frameshifting deletions, we uncovered a striking difference between their C-terminal amino acid sequences. RTT-causing CTDs shifted to the +2 frame to give a common ending –PPX (with one exception, c.1158_1167del, CTD3; see below), whereas all benign CTDs shifted to the +1 frame to end in –SPRT (Figure 3B).

We expanded our enquiry to look at all CTDs in gnomAD, including heterozygotes. The most recent gnomAD release, v.4.1.0, contains sequence data from over 800,000 individuals, with a total of 103 *MECP2* CTD alleles carrying 27 different mutations (Supplementary Table 8). In line with the expectation that the pathogenic frameshift would be severely under-represented in the general population, we observed 96 individuals with 21 different alleles that were shifted to the “benign” +1 frame, whereas only 7 individuals had the potentially pathogenic +2 frame (Fig. 3 SF1A). On closer examination, all but two of the +2 frameshifts avoided the –PPX ending due to deletions which ended prior to an earlier stop codon (Fig. 3 SF1B). The two exceptional mutations were each found in a single heterozygous female individual, with one carrying the recurring c.1164_1207del RTT allele (CTD2). A possible explanation for the rare presence of these pathogenic mutations, so far untested, is that skewed X-inactivation masks their deleterious effect.

In contrast to the predominance of +1 frameshifts found in gnomAD, when all CT-DPR frameshifts in RettBase were considered, +2 frameshifts were present in a large majority of RettBASE entries (Fig. 3 SF2A, Supplementary Table 2). Once individuals without a RTT diagnosis had been excluded, there were 25 cases with +1 and 185 with +2 frameshifts (Fig. 3 SF2B).

A case study illustrating the importance of differentiating benign versus pathogenic C-terminal deletions

A pregnant woman approached one of us because prenatal testing revealed an *MECP2* c.1159_1210 deletion in the female foetus. The diagnostic company indicated this is a pathogenic deletion. The deletion was present in the unaffected father which inspired the family to reach for a second opinion. The deletion was in 100% of the father’s cells and a detailed interview with a focus on neuropsychiatric history revealed a healthy adult male. To ensure that his healthy phenotype is not due to a second site suppressor mutation, the DNA of his healthy mother and brother was analysed. Both carried the same CTD. Of note, the brother who carries the deletion has two healthy daughters who are expected to be obligatory carriers of the mutation (Figure 3C). Lastly, the infant was born and continues to be healthy at ten months of age. According to gnomAD, identical mutations have been found in 3 further individuals, including one hemizygous male (Figure 2B, Supplementary Table 8). In agreement with our prior analysis, this mutation results in a switch to the +1 reading frame leading to a benign deletion with an –SPRTX ending (Figure 3D).

Testing the effect of the codons adjacent to the translational stop on MeCP2 abundance and function

Our analysis of human *MECP2* DNA sequences with radically different clinical symptoms suggested that pathogenicity is determined by codons immediately adjacent to the stop codon. Specifically, the ending –SPRTX is associated with almost all benign human CTDs that involve frameshifts. Unexpectedly this ending is also reported in one of the stringent set of RTT CTDs,

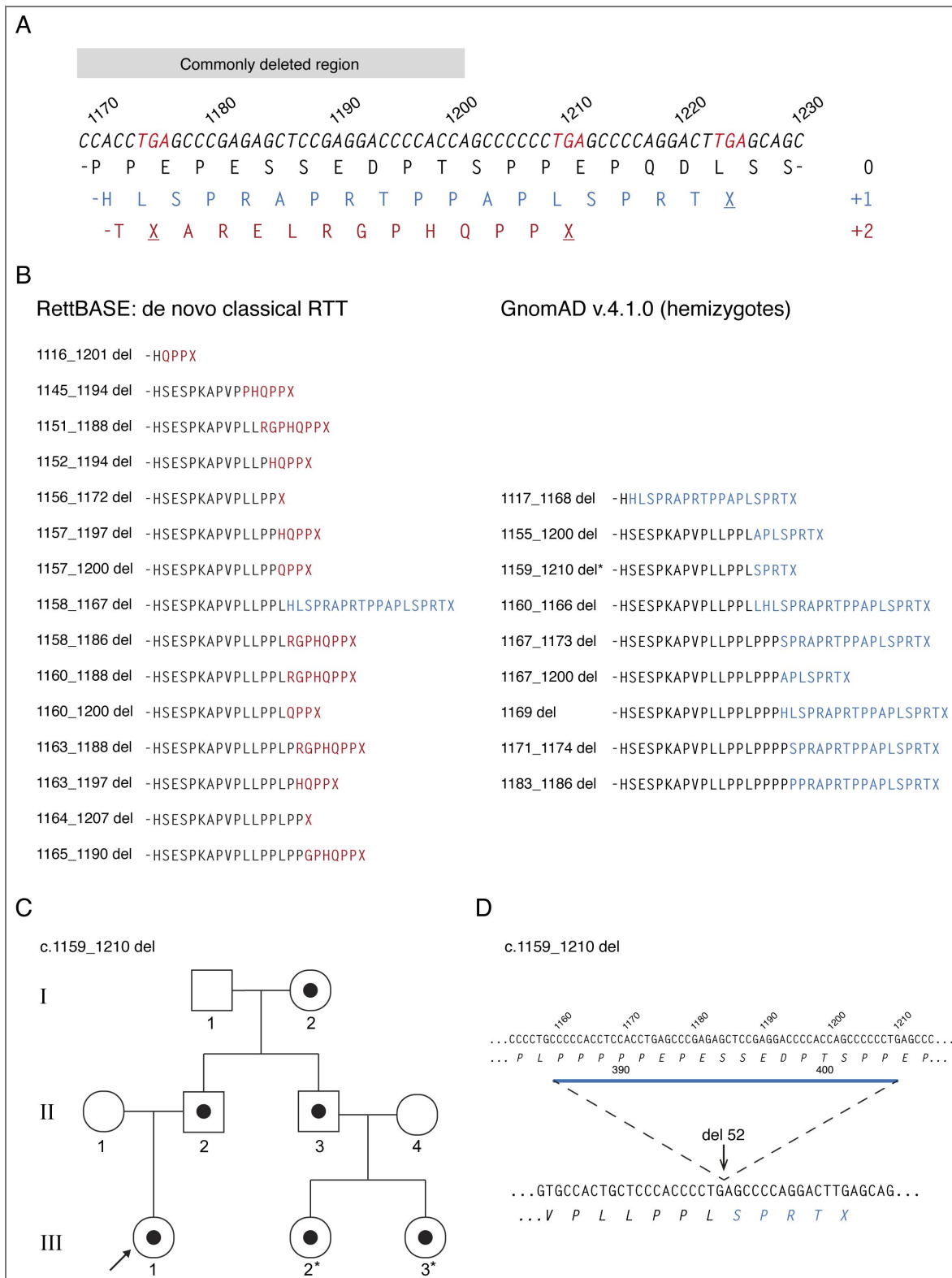


Figure 3. C-terminal amino acid sequences of pathogenic and benign MeCP2 CTDs (high confidence sets).

(A) The three possible reading frames after frame shifts in the CT-DPR. WT genomic sequence is shown, with all possible stop codons in red. The amino acid sequence of the WT reading frame (0) is shown in black, with +1 frame in blue and +2 in red. (B) C-terminal amino acid sequences of frameshifting deletions shown in Figure 2. Sequence after frame shift is shown in blue (+1 frame) or red (+2 frame). (C) Family pedigree showing three generations from a family with a c.1159_1210 del *MECP2* mutation (black filled symbols). (D) Genomic DNA sequence and amino acid sequence showing c.1159_1210 deletion site and molecular consequences.

c.1158_1167del (CTD3) (Bienvenu et al., 2000 [↗](#); Philippe et al., 2006 [↗](#)). Our hypothesis predicted that this mutation had mistakenly been classified as pathogenic and is in fact benign. To test this prediction, we created a CTD3 knock-in mouse allele (Fig. 3 [↗](#) SF3A). The resulting male hemizygous CTD3 mice appeared like WT littermates, did not develop RTT-like phenotypes (Fig. 3 [↗](#) SF3B, C) and had 100% survival at one year. CTD3 MeCP2 protein and RNA were expressed at WT levels in hemizygous mouse brain (Fig. 3 [↗](#) SF3D-F). In contrast, CTD1 mice modelling the c.1157_1197 RTT mutation developed RTT-like phenotypes, had a median survival of 20 weeks and had reduced levels of MeCP2 protein and mRNA in brain, as previously described (Guy et al., 2018 [↗](#)). Thus, despite the CTD3 frameshift occurring only one nucleotide downstream of that in the RTT model CTD1, CTD3 hemizygous mice did not display phenotypic defects, strongly suggesting that the CTD3 mutation is benign. The result confirms our expectation that the +1 frame ending –SPRTX is not pathogenic.

These findings conflict with the presence of two RettBASE entries found in the high confidence CTD set (Figure 2B [↗](#), 3B). Both cases presented a *de novo* c.1158_1167del mutation (Bienvenu et al., 2000 [↗](#); Philippe et al., 2006 [↗](#)). Our findings predict that these two exceptions are in fact benign mutations and therefore not responsible for the observed clinical phenotype. It would be interesting to revisit these cases to identify alternative causal mutations, either within the *MECP2* gene or elsewhere.

Our previous modelling of two recurrent RTT-causing CTD mutations, c.1157_1197del (CTD1) and c.1164_1207del (CTD2) (Guy et al., 2018 [↗](#)), already pointed to C-terminal codons within the truncated transcript as the root cause of MeCP2 deficiency. Building on this information, we found that although both protein and mRNA were consistently reduced in CTD1 mouse brain, the primary transcript was produced at WT levels (Fig. 3 [↗](#) SF3F), pointing to a post-transcriptional mechanism for MeCP2 loss.

There is considerable evidence that consecutive prolines within a messenger RNA are associated with increased stalling of the ribosome during translational elongation, both in pro- and eukaryotes. Indeed, known translation factors function to circumvent this problem (Gutierrez et al., 2013 [↗](#); Huter et al., 2017 [↗](#); Schuller et al., 2017 [↗](#)). Although less well studied, there is also evidence that the presence of two or more prolines preceding a stop codon may also interfere with translational termination (Hayes et al., 2002 [↗](#); Janzen et al., 2002 [↗](#); Matheisl et al., 2015 [↗](#)). Our findings suggest that this may also be the case with RTT-causing CTDs. The presence of two or more prolines immediately preceding the stop codon interferes with translational termination, thereby triggering a process similar to nonsense mediated decay (NMD). This process would differ from canonical NMD in that the premature stop codon resulting from the frameshift occurs in the final exon of *MECP2*, and is also dependent on the amino acid sequence immediately before the stop codon, rather than its location relative to splice junctions. Salvage of the stalled ribosome would result in the loss of both MeCP2 protein and messenger RNA, with the amounts of primary transcript unaffected, as is the case in CTD1 mouse brain (Fig. 3 [↗](#), SF3F).

A genome editing strategy for the CTD class of RTT mutations

Despite their highly variable start and end points, our findings point to common amino acid sequence features shared by all CTDs that give rise to RTT. We used this knowledge to devise an editing strategy that would apply broadly across this mutation category. Current trials of genetic therapies to treat RTT involve delivery of *MECP2* gene constructs (ClinicalTrials.gov [↗](#) IDs NCT05606614, NCT05898620) using Adeno-Associated Viral (AAV) vectors. A disadvantage of this approach is that the dosage to neurons is variable and must not stray into over-expression, which is toxic (Collins et al., 2004 [↗](#); Gadalla et al., 2017 [↗](#); Koerner et al., 2018 [↗](#); Ramocki et al., 2010 [↗](#); Sinnett et al., 2017 [↗](#)). Whilst regulatory strategies to prevent MeCP2 overexpression have been devised (Ross et al., 2025 [↗](#); Sadhu et al., 2023 [↗](#)), an attractive alternative approach is to edit the endogenous *MECP2* gene so that the mutational defect is permanently corrected. This should result in WT expression levels due to the preservation of all endogenous regulatory elements. Several of the common RTT missense mutations can potentially be reverted to WT using adenine or cytosine base editors (ABEs or CBEs) (Gaudelli et al., 2017 [↗](#); Komor et al., 2016 [↗](#)) but

correcting the more complex CTDs would require Prime editing (Anzalone et al., 2019; Chen and Liu, 2023), which is currently less efficient than base editing and would require thorough optimisation of a specific prime editing guide RNA (pegRNA) for the correction of each mutation. Knowing that the C-terminus is not necessary for MeCP2 function and that the –PPX ending is responsible for the reduction in MeCP2 levels, we devised a simple strategy to edit the TGA stop codon found in CTD alleles to TGG (tryptophan) using an ABE (Figure 4A). In this case the next downstream stop codon would result in a –QGASX ending which is expected to be benign (Figure 4B). An advantage of our strategy is that the obligate presence of the pathogenic –PPX ending in all RTT CTD alleles means that a single sgRNA could be used to treat all mutations in the class, amounting to about 10% of all RTT cases.

Conversion of A to G in the stop codon will install a tryptophan codon in its place. To test the consequences of this replacement, we made a new CTD1 X>W knock-in allele in mouse embryonic stem cells (mESCs) that contained tryptophan at this position, thereby mimicking the outcome of the proposed edit (Fig. 4 SF1A, B). The truncated protein is slightly larger than MeCP2 CTD1 and CTD2 due to the extended missense tail (Figure 4B) and we found that ESC-derived neurons carrying this CTD1 X>W allele restored the CTD MeCP2 protein to WT levels (Fig. 4 SF2A). To demonstrate that this approach could be applied to multiple CTD alleles, we generated a corresponding CTD2hu X>W allele and again observed recovery of wildtype levels of MeCP2 protein in mESC-derived neurons (Fig. 4 SF2B). Importantly, CTD1 X>W hemizygous mice were indistinguishable from WT littermates, did not develop RTT-like phenotypes (Figure 4C, Fig. 4 SF3A) and survived to at least one year, in contrast to the original CTD1 mouse line (Figure 4D). As with the mESC-derived neurons, the X>W change restored MeCP2 protein and mRNA levels (Figure 4E, F, Fig. 4 SF3B). We conclude that introduction of a novel tryptophan and extended missense tail at this position has no discernible effect on protein function. In order to test the efficacy of base editing reagents to install the X>W allele, we next sought an experimental system that reproduced the relative MeCP2 protein levels produced from different mouse knock-in alleles. To express the protein at consistent physiological levels, we used the HEK cell-based Flp-In T-REx system (Invitrogen), previously found to perform better than transient transfections for NMD studies (Gerbracht et al., 2017). As MeCP2 is usually found at low levels in dividing cells and is only highly expressed in post-mitotic differentiated cells, tetracycline-inducible cDNA constructs were used so that transgene expression could be kept off during standard cell culture and induced for a limited period to examine mRNA and protein expression. Constructs based on the major e1 *MECP2* isoform were recombined into a single genomic locus using Flp recombinase to create an isogenic series of single-copy *MECP2* transgenic cell lines. CTD mutations replicating mouse alleles (CTD1, CTD2hu and CTD2mo) and patient mutations (c.1157_1197del, CTD1 and c.1164_1207del, CTD2) were introduced into mouse and human full-length cDNAs (Figure 5A, Fig. 5 SF1A) and integrated into the cells. Upon induction of transgene expression, these cell lines recapitulated the reduced MeCP2 protein and mRNA levels previously seen for the knock-in alleles (Figure 5B–D, Fig. 5 SF1B, C). The finding that the effect of the CTD mutations on protein and mRNA levels is well reproduced using expression constructs with a heterologous promoter, 3' UTR and polyadenylation signal, and no intronic sequences, reinforces the notion that this is a post-transcriptional defect.

The most commonly used ABEs consist of an evolved Tad deaminase domain(s) fused to a nicking *S. pyogenes* (Sp) Cas9, requiring an NGG PAM following the 20 nucleotide protospacer sequence (Gaudelli et al., 2017; Richter et al., 2020). Due to the cytosine-rich nature of the C-terminal deletion-prone region, there were no suitable PAMs available that would appropriately position the ABE to edit the target A (Figure 6A). Engineered SpCas9 PAM variant enzymes have been developed with relaxed PAM requirements to yield SpG that requires a single G PAM (NGN), and SpRY which has a relaxed sequence requirement but works most efficiently with either a single G or A (NG/AN) (Walton et al., 2020). We designed two gRNAs which would work with either SpG or SpRY ABEs (guides 1, A7 and 2, A4), and one with SpRY only (guide 5, A5) (Figure 6A). Plasmids encoding U6 promoter-expressed guide 1 or 2 and CMV promoter ABEmax-SpG (Walton et al., 2020) were transfected into T-REx cells with a mouse CTD1 (moCTD1) transgene, but no editing of an amplicon spanning the target site was detected by Sanger sequencing. As the two

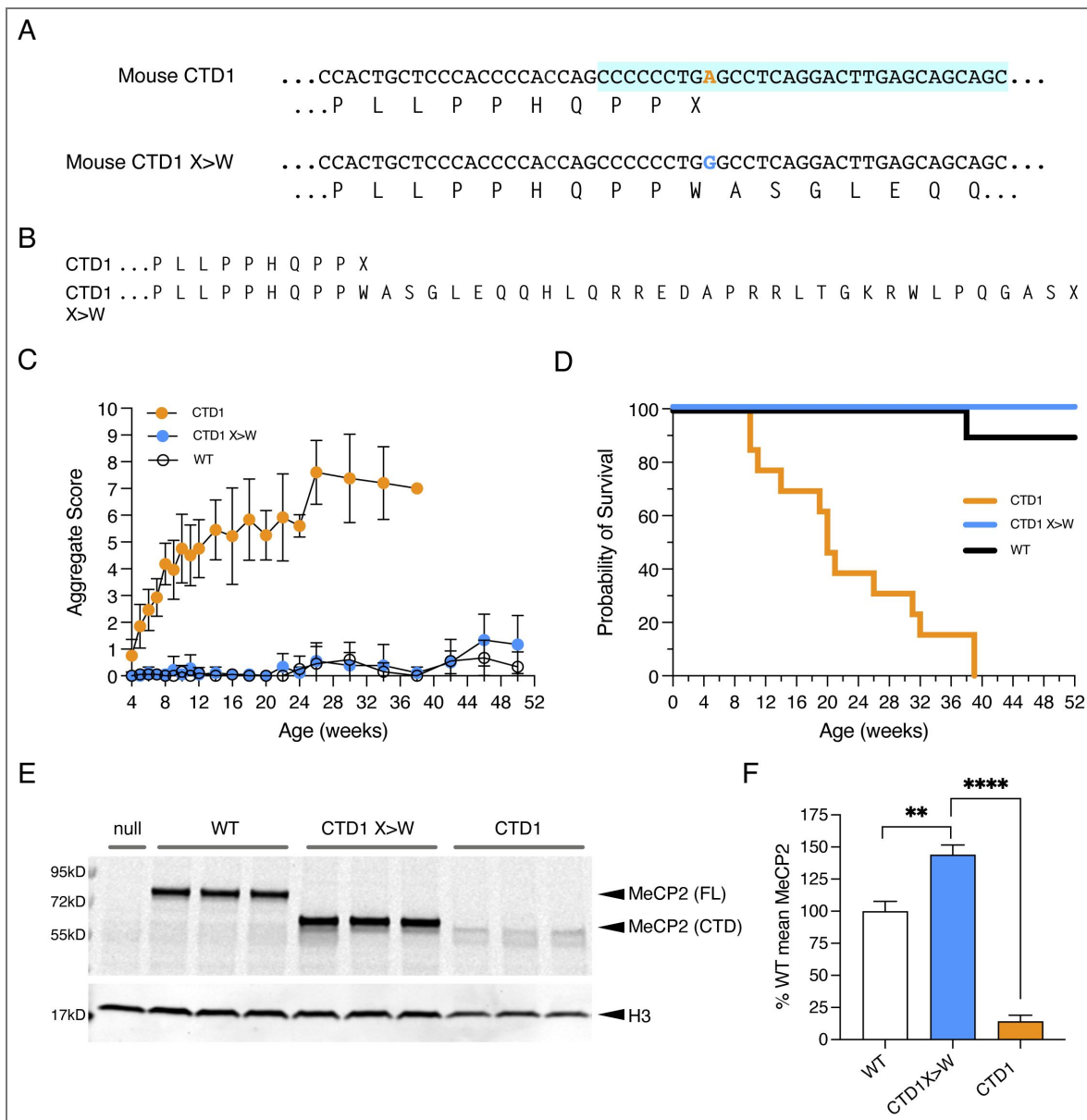


Figure 4. A CTD1 X>W knock-in mouse models adenine base editing of the CTD1 stop codon.

(A) The genomic sequence of mouse CTD1 and CTD1 X>W alleles. The mutated adenine is shown in orange, and the sequence common to all mouse CTD alleles is shaded in turquoise. (B) The amino acid consequences of alleles in (A). (C) Phenotypic scoring of hemizygous male mice with CTD1 (n=13) and CTD1 X>W (n=9) knock-in alleles, and WT littermates of CTD1 X>W animals (n=10). Mean $\pm$ standard deviation (sd). (D) Kaplan-Meier plot of survival of animals shown in (C). (E) Western blot of whole brain protein from 6 week old male mice hemizygous for *Mecp2*-null, CTD1, CTD1 X>W and WT alleles. Full-length (FL) and C-terminally deleted (CTD) MeCP2 proteins are indicated. Histone H3 is used as a loading control. (F) Quantification of (E). N=3 per genotype, mean $\pm$ sd. Unpaired two-tailed t-test: CTD1 X>W vs CTD1 $P < 0.0001$ (****), CTD1 X>W vs WT $P = 0.0021$ (**).

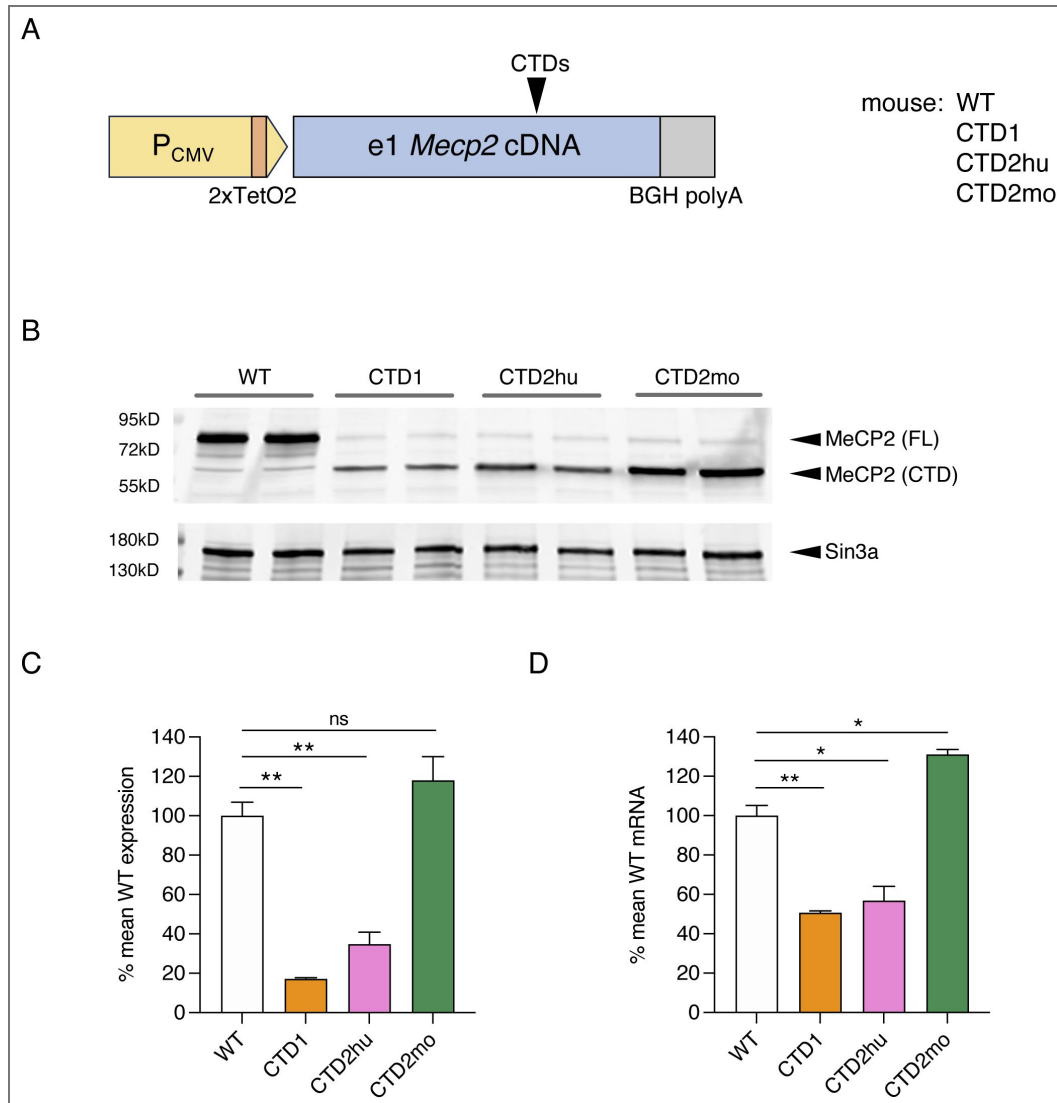


Figure 5. FLP-In T-REx cell lines reproduce the reduction in MeCP2 protein and mRNA seen with CTD knock-in mouse alleles.

(A) Schematic of *Mecp2* transgenes in FLP-In T-REx cell lines. Deletions are introduced into a full-length e1 *Mecp2* cDNA, with a bovine growth hormone (BGH) polyadenylation signal and tetracycline-inducible CMV promoter. (B) Western blot with whole cell lysates from independent FLP-In T-REx clones carrying mouse cDNA transgenes (24 hours tetracycline induction). Sin3a loading control. (C) Quantification of MeCP2 protein expression from (B). N=2 clones per genotype, mean $\pm$ sd. Unpaired two-tailed t-test: WT vs CTD1 $P=0.0035$ (**), WT vs CTD2hu $P=0.0099$ (**), WT vs CTD2mo $P=0.2101$ (ns). (D) Quantification of *Mecp2* transgene mRNA from the same experiment as (B) and (C). N=2 clones per genotype, mean $\pm$ sd. Unpaired two-tailed t-test: WT vs CTD1 $P=0.0057$ (**), WT vs CTD2hu $P=0.0208$ (*), WT vs CTD2mo $P=0.0171$ (*).

gRNAs position the target A7 at the edge of the optimal editing window for ABE_{max} (Gaudelli et al., 2017 [DOI](#)), next we tested ABE8e-SpG and ABE8e-SpRY constructs containing a further evolved TadA domain with increased editing efficiency and a wider editing window (Richter et al., 2020 [DOI](#)) (Figure 6B [DOI](#)). These ABEs were able to edit the moCTD1 transgene with high efficiency using all gRNAs tested (Figure 6C [DOI](#)), reaching 50-60% editing at the target A without selection of transfected cells. Editing with ABE8e-SpG was somewhat more efficient than with ABE8e-SpRY. Although editing at the target site was similarly efficient for all gRNAs, bystander editing at two nearby As (+6 and +9 relative to the target A) differed, with guide 1 giving minimal bystander editing, whereas using guides 2 and 5 resulted in significant bystander editing, particularly at the +6 position. Guide 1 extends furthest upstream (target A7), placing the bystander As furthest away from the editing window. Insertion or deletion mutations (indels) were low for all combinations (Fig. 6 [DOI](#) SF1A).

To determine whether editing of the moCTD1 allele led to an increase in MeCP2 protein, expression of the transgene was induced for 24 hours five days after transfection of the ABE constructs and guides. The significant editing seen in Figure 6C [DOI](#) was accompanied by a concomitant increase in the amount of MeCP2, and a size shift due to the increased length of the missense tail (Figure 6D [DOI](#), Fig. 6 [DOI](#) SF1B, C). We anticipated that a single gRNA could be applied to the editing of different CTD mutations. To test this, we used ABE8e-SpG guide 1 to edit a mouse CTD2hu allele in T-REx cells. Editing efficiency was indeed comparable to that of the CTD1 allele (Fig. 6 [DOI](#) SF1D, E).

As a final proof-of-concept experiment, we asked whether these reagents also edited the human mutation. There is one nucleotide difference between mouse and human sequence in the guide region, so we tested human-specific gRNAs, guides 1 and 2, on a huCTD1 T-REx line (Fig. 6 [DOI](#) SF2A). The human sequence-targeted gRNAs performed in a very similar way to their mouse equivalents (Fig. 6 [DOI](#) SF2B, C), with the highest editing efficiency of almost 80% achieved using ABE8e-SpG and human guide 1. The mouse guide 1, with one mismatch to the human target, was also tested on the huCTD1 transgene, resulting in only a modest decrease in editing efficiency. We conclude that our DNA base editing strategy offers promise for future development as a therapeutic for RTT caused by CTD mutations.

Discussion

The present study provides a simple genetic explanation for the initially puzzling observation that C-terminal deletions in the human *MECP2* gene cause RTT in some individuals but appear to be benign in others. In-frame deletions remove only a small number of amino acids from a region of MeCP2 that appears, from human genomic sequencing data, to be highly tolerant of mutation. CTDs causing shifts to the +1 and +2 frames both remove the entire C-terminus after the deletion site. Based on previous work in mice indicating that the detailed amino acid sequence close to newly created C-termini influences the level of MeCP2 expression, we compared the C-terminal sequences in databases of RTT patients and the general population. It emerged that RTT truncations terminated with the codons –PPX, whereas apparently benign truncations terminated in –SPRTX (Figure 7A [DOI](#)). Using mouse models, we verified that the two prolines prior to the stop codon are essential contributors to the resulting severe deficiency of MeCP2 protein (Figure 7B [DOI](#)). Cell lines containing CTD transgenes replicated these findings and were used to show that DNA base editing of the stop codon following the two prolines leads to readthrough to the next stop codon, preventing loss of MeCP2 protein (Figure 7C [DOI](#)). Such genome editing could potentially be used to treat RTT patients with CTD mutations.

The reason why mutations in the disordered C-terminal domain of MeCP2 should cause RTT has been the subject of speculation (Cheng et al., 2014 [DOI](#); Li et al., 2020 [DOI](#)). This part of the protein is remote from the well characterised MBD and NID domains of MeCP2, and mouse models in which this entire region of the protein is missing have no RTT-like phenotype (Tillotson et al., 2017 [DOI](#)), suggesting that it is dispensable. Our previous work with mouse models of CTD mutations made clear that the truncations themselves do not interfere significantly with MeCP2 function. In particular, CTD2 mice in which the –PPX terminus is replaced by –SPX are like wildtype littermates

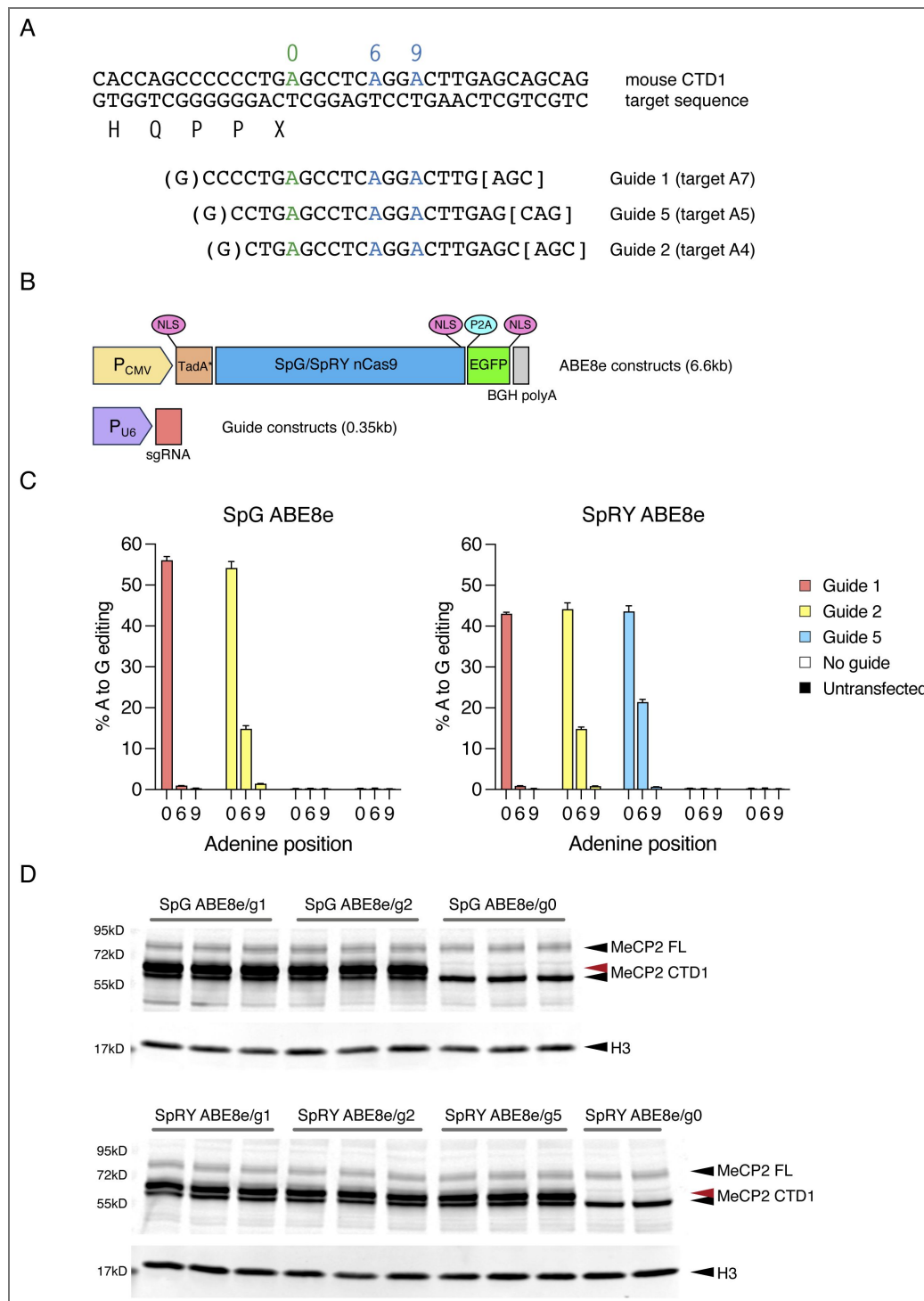


Figure 6. Base editing of mouse CTD transgenes in Flp-In T-REx cell lines.

(A) Target genomic sequence and sgRNAs. The target A (position 0) is shown in green, with two bystander As within the guide sequence shown in blue (positions +6 and +9). Target site protospacer sequences are shown along with the PAM; the gRNA spacers are flanked by an additional G added to promote RNA polIII transcription. (B) ABE and gRNA constructs used for transfection experiments. (C) Editing efficiency following transfection of mouse CTD1 Flp-In T-REx cells with ABE8e-SpG or ABE8e-SpRY base editors and gRNA expression plasmids. Editing efficiency at the target and bystander As is quantified by amplicon sequencing (n=3 transfections per ABE/guide combination). (D) western blot showing MeCP2 protein levels from the experiment in (C) after 24 hours induction of transgene expression. The red arrow indicates MeCP2 CTD1 protein after editing (CTD1 X>W).

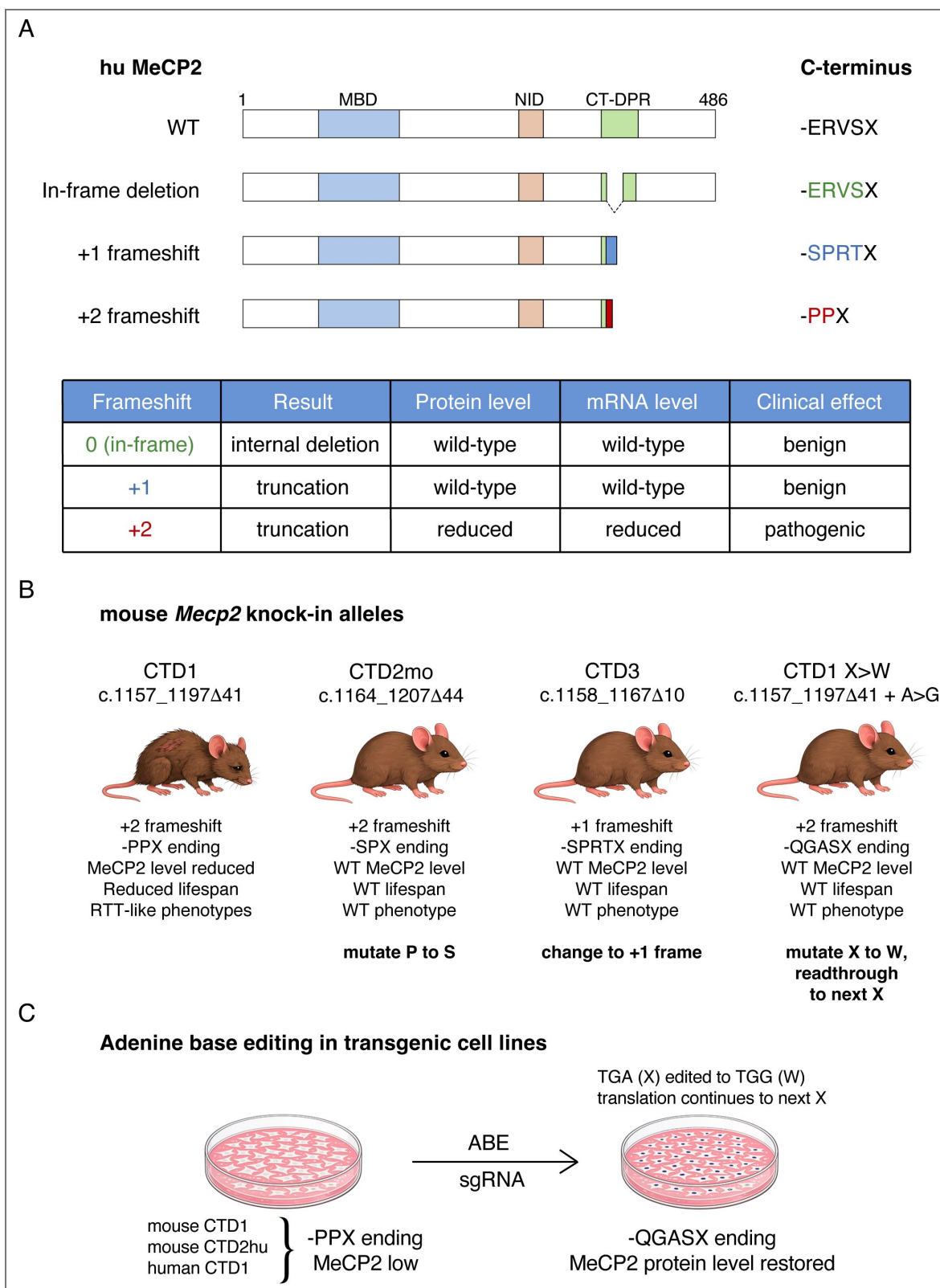


Figure 7. Summary of CTD alleles used in this study

(A) Human WT MeCP2 protein and the effect of the three possible reading frames remaining after deletions in the CT-DPR (B) *Mecp2* knock-in mouse models. The phenotype of the hemizygous CTD1 mouse can be ameliorated by altering the C-terminal amino acid sequence in several ways. All of these lead to non-PPX endings. (C) Efficient X>W editing of CTD transgenes using ABEs and sgRNAs in transfected cells. This replicates the CTD1 X>W allele and demonstrates the potential for therapeutic use of adenine base editing for RTT patients with CTDs.

despite persistence of the severe truncation. We show here that it is greatly reduced abundance of truncated MeCP2, presumably triggered by translational stalling at the –PPX codons, that drives the RTT phenotype.

A likely explanation for MeCP2 deficiency in CTD mutants is that the combination of consecutive prolines juxtaposed with a stop codon interferes with translation by the ribosome. Runs of multiple prolines are known to be problematic for translational elongation, likely due to the atypical geometry of this amino acid, with the translation factors eIF5A in mammals and EF-P in bacteria playing an important role in facilitating their translation (Gutierrez et al., 2013 [↗](#); Huter et al., 2017 [↗](#); Schuller et al., 2017 [↗](#)). The effect of multiple prolines on translational termination is less well understood, although there is evidence from bacteria and eukaryotic cells that they may hinder efficient termination (Hayes et al., 2002 [↗](#); Janzen et al., 2002 [↗](#); Matheisl et al., 2015 [↗](#)). In *E. coli* certain C-terminal sequences can trigger SsrA tagging, whereby the nascent polypeptide and translating RNA are targeted for degradation, thereby rescuing the stalled ribosome (Hayes et al., 2002 [↗](#)). Interestingly, a sequential pair of prolines prior to the stop codon showed the strongest effect and, whereas seven *E. coli* proteins ending in PPX would be expected by chance, there are in fact none. In human cells, translation of a stalling peptide from human cytomegalovirus (hCMV) was used to trap the translational termination complex for structural studies. This effect was found to be dependent on the presence of two proline residues at the C-terminus of the peptide and the translational termination factor eRF1 (Matheisl et al., 2015 [↗](#)).

Our findings have important clinical implications. We have shown here a contemporary pedigree containing several individuals with the frameshifting C-terminal deletion c.1159_1210, p.P387SX5. The CTD mutation was first identified after prenatal screening of a female foetus and was assigned as disease-causing, with the family informed that she was likely to develop Rett syndrome. After further investigation the same mutation was found in several healthy family members, including the child's father, and the family were reassured that the mutation was likely to be benign due to the +1 frameshift and –SPRTX ending. Our results promise to resolve the discrepancy between genetic inference and clinical manifestation in these cases. Genetic testing can now incorporate this information to provide reassurance in cases where the pathogenic –PPX terminus is absent. The outcome of deletions in the CT-DPR can be predicted based on the exact location and size of the deletion (Figure 8 [↗](#)). On a cautionary note, despite clinical examples arguing that these are phenotypically neutral mutations, it will be wise to maintain vigilance. While benign CTDs will not give rise to RTT, the possibility that less severe consequences accompany these mutations deserves future scrutiny.

The shared nucleotide sequence features of pathogenic CTDs raise the possibility that genome editing can be therapeutically beneficial. Using an SpCas9-based DNA base editor, we have shown that one guide RNA is potentially capable of targeting all pathogenic CTD mutations with the –PPX ending and can edit CTD transgenes efficiently in human cell lines. The edits replace the –PPX terminal sequence by –PPW– and add a short stretch of out-of-frame amino acid sequence before encountering the next stop codon. Modelling in mice indicates that the introduction of tryptophan plus downstream amino acids completely rescues the effects of the pathogenic mutation. Next steps will involve testing for phenotypic rescue after DNA base editing in an animal model, and addressing the issue of AAV-mediated delivery, which is relatively inefficient with most currently available vectors. The ultimate goal is to provide a unified therapeutic approach to this varied but unexpectedly homogeneous category of RTT-causing mutations.

Materials and Methods

Human mutation data

RTT patient mutation data was taken from RettBASE (Christodoulou et al., 2003 [↗](#)), formerly hosted at <http://mecp2.chw.edu.au/> [↗](#) and downloaded 17/08/2017 (Supplementary Table 1 [↗](#)). The data held in RettBASE has now been incorporated into ClinVar (RRID:SCR_006169 [↗](#)) in an abbreviated form (<https://www.ncbi.nlm.nih.gov/clinvar/submitters/504290/> [↗](#)). High confidence

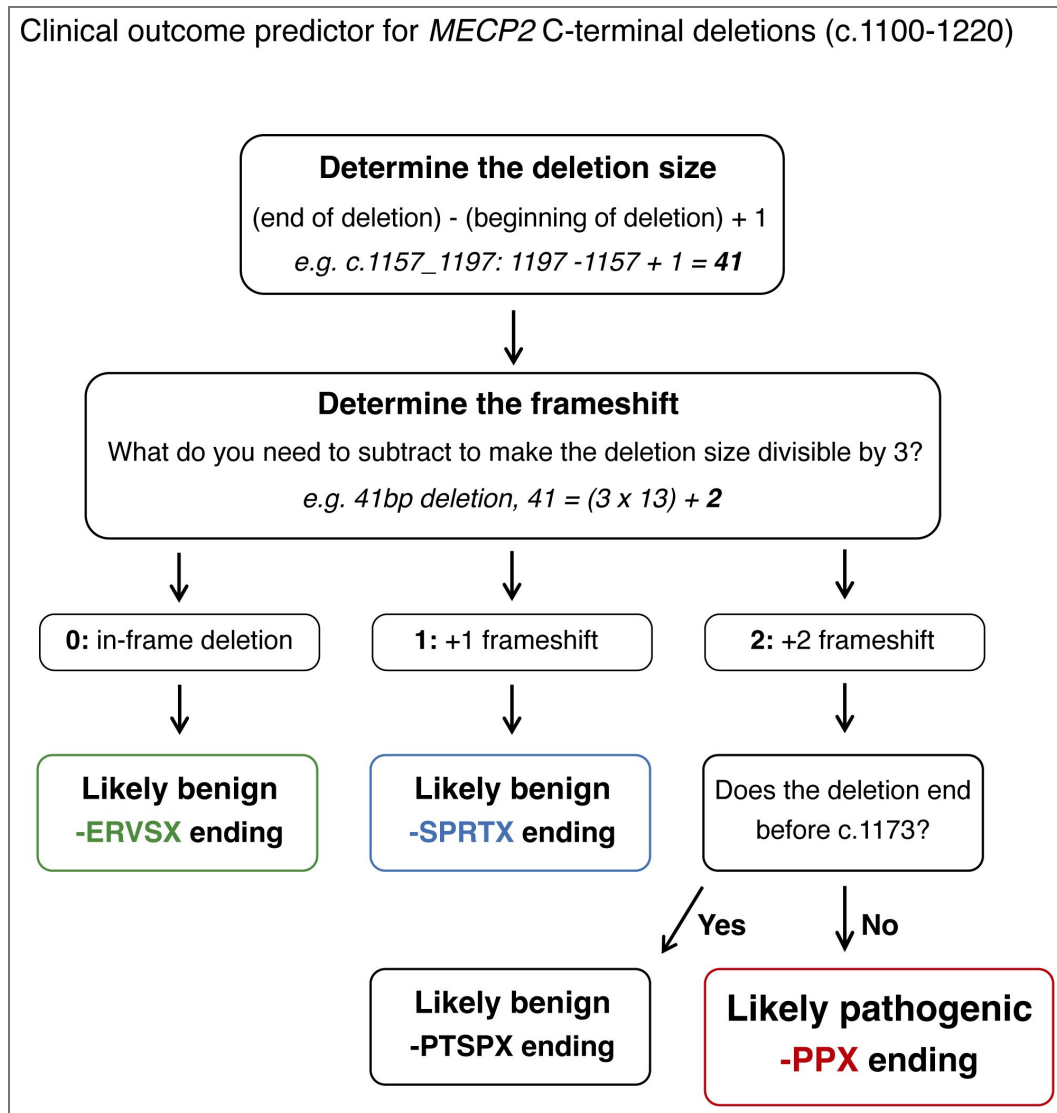


Figure 8. Flow chart for predicting the likely clinical prognosis of deletions in the CT-DPR of human *MECP2* based on the findings in this study.

RettBASE CTD alleles were those where at least one individual carrying that mutation had been diagnosed with classical Rett syndrome and carried a *de novo* *MECP2* mutation (i.e. not present in either unaffected parent).

GnomAD data was taken from release v4.1.0, *RRID:SCR_014964* <https://gnomad.broadinstitute.org/news/2024-04-gnomad-v4-1/> on 21/11/2024 (Supplementary Tables 5 [5](#)–8 [8](#)). *MECP2* mutations throughout are numbered according to transcript ENST00000303391.11 and amino acids are numbered according to the e2 MeCP2 isoform, as commonly used for RTT patient mutations. High confidence gnomAD CTD alleles were defined as those where at least one hemizygous (male) individual carried that mutation.

The family with the c1159-1210 deletion reached out to HZ for advice and this led to a detailed assessment of the medical history and the recommendation for expanded genetic testing in other family members.

AlphaMissense data

Mean pathogenicity scores at each amino acid position of human MeCP2 (e2 isoform) were obtained from AlphaMissense (<https://alphamissense.hegelab.org>) (Cheng et al., 2023). The mean value for a sliding window of five amino acids was plotted at each position.

Generation of *Mecp2* CTD mutations in mouse ES cells

Mecp2 CTD knock-in alleles were generated in male mouse ES cell line JU09 (A gift from Joe Mee, University of Edinburgh), derived from 129/Ola E14Tg2a cells (*RRID:CVCL_3505*). Cells were transfected with a CRISPR/Cas9 plasmid based on pX330-U6-Chimeric_BB-CBh-hSpCas9, a gift from Feng Zhang (<http://n2t.net/addgene:42230>; *RRID:Addgene_42230*) (Cong et al., 2013) containing a sgRNA guide with protospacer sequence CCCCTGAGCCTCAGGACTTG within the deleted region for each allele, along with a 190nt single stranded oligodeoxynucleotide (ssODN) containing the desired DNA sequence changes, and a puromycin resistance plasmid for transient selection of transfected cells. For each knock-in allele 4×10^5 ES cells were transfected with 500ng CRISPR/Cas9 plasmid, 200ng pPGK-Puro-pA plasmid and 10pmol ssODN using Lipofectamine2000 (ThermoFisher). After 24 hours 0.8µg/ml puromycin selection was added and maintained for 48 hours. Selected cells were plated at clonal density without selection and clones were picked after approximately 10 days.

ssODNs

CTD3:

AGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCCCAAA GGCG
CCCGTGCCACTGTCTCCACCCCTCCACCTGAGCCCGGGCTCCGAGGACCCACCAGCCCC CTGAGCCCC
AGGACTTGAGCAGCAGCATCTGCAAAGAAGAGAAGATGCCCCGAGG

CTD1 X>W:

AGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCCCAA AGGCC
CCCGTGCCACTGTCTCCACCCCTCAGCCCCCTGGGCCTCAGGACTTGAGCAGCAGCATCT GCAAAGAAGA
GAAGATGCCCCGAGGAGGCTCACTGGAAAGCGATGGCTGCCCAAG

CTD2 X>W:

AGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCCCAA AGGCC
CCCGTGCCACTGTCTCCACCCCTGCCCCCTGGGCCTCAGGACTTGAGCAGCAGCATCTGCA AAGAAGAGA
AGATGCCCCGAGGAGGCTCACTGGAAAGCGATGGCTGCCCAAGGAG

Clones were screened for hemizygous knock-in alleles by PCR on genomic DNA (forward primer: A GCATCAGAAGGTGTTTCAGG, reverse primer: CCATAGGCTGAGTCTTAGCTGG), followed by cutting with SacII (CTD3, site gained) or SacI (CTD1 X>W, CTD2 X>W, site lost), and sequencing. Correctly edited clones were karyotyped and mycoplasma tested before further use.

Differentiation of mouse ES cells into neurons

Verified ES cell knock-in clones were differentiated into neurons. At least two independently targeted clones were used for each genotype. Differentiation was carried out using a 4-/4+ retinoic acid (RA) procedure as previously described (Bain et al., 1995 [↗](#); Li et al., 1998 [↗](#)). Neural precursor cells were seeded onto 6cm dishes at a density of 1.5×10^5 cells/cm² and harvested after 7 days by scraping into ice cold phosphate-buffered saline (PBS), pelleting the cells and snap freezing.

Establishment of *Mecp2* mutant mouse lines

Wildtype C57BL/6J mice (RRID:IMSR_JAX:000664 [↗](#)) were purchased from Charles River Laboratories. CTD3 and CTD1 NS mouse lines were generated by injection of knock-in ES cells into mouse C57BL/6J blastocysts using standard methods. Chimeric males were crossed with C57BL/6J females and germline transmission from the ES cells was identified by coat colour. All agouti female pups were obligate heterozygotes for the *Mecp2* mutant allele due to its location on the X chromosome. This (N=1) and subsequent generations were bred by crossing *Mecp2*-mutant heterozygote females with C57BL/6J WT males. The genotypes of mutant lines were confirmed by sequencing. Routine genotyping was performed on ear biopsies using custom assays developed by Transnetyx, details available on request.

Mice were bred in a specific-pathogen-free facility in individually ventilated cages, with wood chippings, tissue bedding and environmental enrichment and were given *ad libitum* access to food and water. Rooms were maintained on a 12-hour light/dark cycle at 45-65% humidity at 20-24°C. Procedures were carried out by certified persons, licensed by the UK Home Office and according to the Animals (Scientific Procedures) Act 1986 under project licences (PPLs) 60/4547 and PP4326006. Ethical review of the PPLs was carried out by the University of Edinburgh Animal Welfare and Ethical Review Board (AWERB).

Mouse phenotyping

Cohorts of hemizygous mutant male mice and WT littermates (N=3 backcross to C57BL/6J) were weighed and scored for a range of RTT-like phenotypes (mobility, gait, hindlimb clasping, tremor, breathing and general condition) to give an aggregate score between 0 and 12 as previously described (Cheval et al., 2012 [↗](#); Guy et al., 2007 [↗](#)). Cohorts of at least 9 mutants and 9 wild-type littermates were scored for each knock-in line, with higher numbers being used where available to allow for any losses unrelated to the mutation, such as fighting. Sample sizes were selected based on our previous studies of *Mecp2* mutant mice and practical considerations of animal availability; no formal power calculation was performed. Animals were allocated to experimental groups without formal randomisation, as group assignment was determined by genotype; however, littermates were used where possible to minimise confounding effects. Males which would by necessity have to be singly housed during the study period were excluded. Scoring was carried out blind to genotype and to previous scores. Animals receiving a maximum score of 2 for tremor, breathing or general condition or which had lost 20% of their body weight had reached the severity limit of the experiment according to the Home Office license and were humanely culled. These animals were counted as having ‘died’ for the purposes of survival data. Animals of any genotype which were culled for reasons not linked to the mutation, such as fighting with cage mates, were removed from survival plots at that point (censored data). Scoring was carried out for one year, from 4 to 50 weeks of age, starting weekly (4-12 weeks), then every two weeks (12-26 weeks) and then every four weeks (26-50 weeks).

Preparation of RNA and quantitative RT-PCR

Total cellular RNA was prepared from mESC-derived *in vitro* differentiated neurons, Flp-In T-REx cell lines and mouse brain using TriReagent (Sigma), according to the manufacturer’s protocol. cDNA was prepared from 1µg total RNA per reaction using a QuantiTect kit (Qiagen) and amplified

in a qPCR reaction using SensiMix SYBR and Fluorescein Master Mix (Bioline) in a LightCycler 480 (Roche). Primers for qPCR were as follows:

Mouse *Mecp2* mRNA: Exon3 F – ACCTTGCCTGAAGGTTGGAC, Exon4 R – GCAATCAATTCTACTTTAGA GCGAAAA

Mouse *Mecp2* primary transcript: Intron1 F – ACATGGCCGACAGAGTGC, Intron1 R – GCACCTGAGGA AGCAAACC

Flp-In T-REx *Mecp2* transgene (mouse and human): moMe2TREx F – GAGAGGAGCCTGTGGACAGC, BGHpA R – CACAGTCGAGGCTGATCAGC

Cyclophilin A mRNA (control): CypA F – TCGAGCTCTGAGCACTGGAG, CypA R – CATTATGGCGTGTAAGTCACCA

Samples were run in triplicate and the amount of *Mecp2* cDNA or primary transcript was calculated for each biological replicate after correction of CT values for PCR efficiency (using a dilution series to plot a standard curve) and normalisation to the control transcript Cyclophilin A. Each biological replicate was normalised to the mean of all WT samples for that transcript.

Western blotting

In vitro differentiated ES cell-derived neurons, Flp-In T-REx cells and whole brain samples were prepared for western blotting as previously described (Ross et al., 2016 [DOI](#)). Extracts were run on TGX 4-20% gradient gels (BioRad) loading extract equivalent to approximately 5×10^5 nuclei per well. For quantification, samples were run on duplicate gels and transferred to nitrocellulose membrane overnight in the cold at 25V. Western blots were processed as described previously (Guy et al., 2018 [DOI](#); Ross et al., 2016 [DOI](#)) using the following antibodies: anti-MeCP2 (N-terminus): mouse monoclonal Men-8, *RRID:AB_477235* [DOI](#) (Sigma), anti-NeuN: rabbit polyclonal ABN78, *RRID:AB_10807945* [DOI](#) (Millipore), anti-histone H3: rabbit polyclonal ab1791, *RRID:AB_302613* [DOI](#) (Abcam), anti mSin3a: rabbit polyclonal ab3479, *RRID:AB_303922* [DOI](#) (Abcam). Western blots were developed with IR-dye secondary antibodies (IRDye 800CW donkey anti-mouse, *RRID:AB_621847* [DOI](#) and IRDye 680LT donkey anti-rabbit, *RRID:AB_10956166* [DOI](#), LI-COR Biosciences) and scanned using a LI-COR Odyssey machine. Images were quantified using Image Studio Lite software version 6.0.0.28 (LI-COR Biosciences). The ratio of NeuN:histone H3 signals for each lane was used to check for equal neuronal *in vitro* differentiation and clones that did not differentiate well were discarded.

MeCP2 levels were normalised to the histone H3 (neurons, whole brain) or mSin3a signal (Flp-In-TREx cells) for each lane to compare the amount of MeCP2/nucleus between samples.

Production and characterisation of Flp-In T-REx cells

Flp-In™ T-REx™ 293 cells (*RRID:CVCL_U427* [DOI](#)) were obtained from Invitrogen and maintained according to the manufacturer's instructions. Cells were maintained under appropriate selection and routinely tested for mycoplasma. Tetracycline-inducible MeCP2 expression constructs were made by cloning human and mouse e1 cDNAs into plasmid pcDNA5/FRT/TO, Cat#V6520920 (Invitrogen). These were then inserted into the FRT site of Flp-In™ T-REx™ 293 cells according to the manufacturer's instructions. Single-cell clones were verified by sequencing and single integration of the plasmid into the FRT site was confirmed by PCR or Southern blotting. Transgene expression was induced by addition of 0.5µg/ml tetracycline to the growth medium for 24 hours prior to cell harvesting and snap freezing of cell pellets.

Transfection with ABEs/guides and quantification of DNA base editing

Plasmids ABE8e-SpG and ABE8e-SpRY were derived from pCMV-T7-ABEmax(7.10)-SpG-P2A-EGFP (Addgene plasmid # 140002; <http://n2t.net/addgene:140002> [DOI](#); *RRID:Addgene_140002* [DOI](#)) and pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP (Addgene plasmid #140003; <http://n2t.net/addgene:140003> [DOI](#); *RRID:Addgene_140003* [DOI](#)) (Walton et al., 2020 [DOI](#)) by exchanging the ABEmax adenine deaminase

domain for ABE8e (Alves et al., 2024 [↗](#)). A separate plasmid, pGuide, a gift from Kiran Musunuru (Addgene plasmid # 64711; <http://n2t.net/addgene:64711> [↗](#); [RRID:Addgene_64711](#) [↗](#)) was used to express sgRNAs using a human U6 promoter (Ding et al., 2013 [↗](#)). Protospacer sequences were cloned into a BbsI site in pGuide using standard methods.

Flp-In T-REx cells were transfected with an equimolar mixture of ABE and guide plasmids. “No guide” controls received ABE + pGuide plasmid with no protospacer sequence (g0, empty guide). Cells were plated at 4×10^5 cells per well in 6-well plates. The next day each well was transfected with 1000ng plasmid DNA and 6μl Lipofectamine 2000. Five days after transfection transgene expression was induced with 0.5μg/ml tetracycline. After 24 hours each well was harvested by trypsinisation, and the cells were split and snap frozen as two pellets for preparation of genomic DNA and protein for western blotting.

Genomic DNA was prepared from cell pellets using a Puregene Cell kit (Qiagen). Editing was quantified by MiSeq sequencing of a 190bp PCR amplicon spanning the target site: primers Me2 F6 – AGAAGGAGCACCATCATCACC, Me2 R2 – CCATAGGCTGAGTCTTAGCTGG. Amplicons were purified using DNA Clean and Concentrator-5 columns (Zymo Research) and sequencing libraries were made using NEBNext Ultra II DNA library prep kit for Illumina and NEBNext Multiplex Oligos for Illumina (New England Biolabs). Libraries were sequenced using a MiSeq Reagent Nano Kit v2 (300 cycles) (Illumina). Sequencing data was analysed using CRISPResso2, <https://crispresso2.pinellolab.org/submission> [↗](#), [RRID:SCR_021939](#) [↗](#) (Clement et al., 2019 [↗](#)).

Statistical analysis

All statistical analysis was performed using GraphPad Prism version 11.0.0, [RRID:SCR_001066](#) [↗](#) (GraphPad Software). Datasets were compared using Student’s t-tests (unpaired, two-tailed, with or without Welch’s correction for unequal variance, as appropriate).

Study approval

All animal experiments were performed under United Kingdom Home Office project licences (PPLs 60/4547 and PP4326006).

Supplementary figures

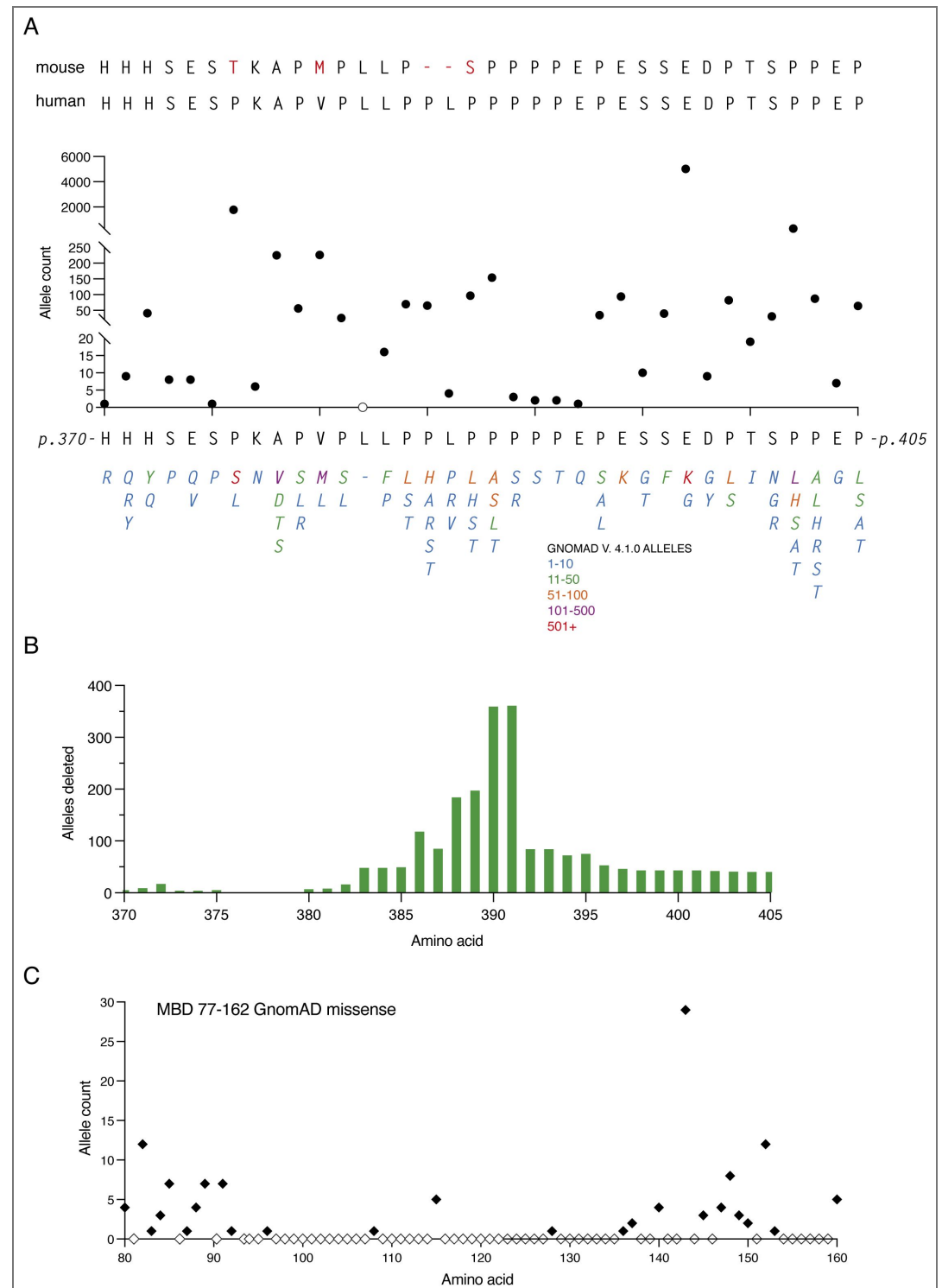


Figure 2, Supplementary Figure 1. GnomAD v4.1.0 data from the CT-DPR (A) GnomAD missense mutations. A comparison of mouse and human amino acid sequence in the CT-DPR, followed by a plot of the number of individuals in gnomAD with missense mutations present (filled circles) or absent (open circles) at each position. All amino acid changes found are listed below, colour-coded according to frequency. Hemizygous, heterozygous and homozygous mutations are included. **(B)** Plot of the total number of gnomAD alleles with in-frame deletions at

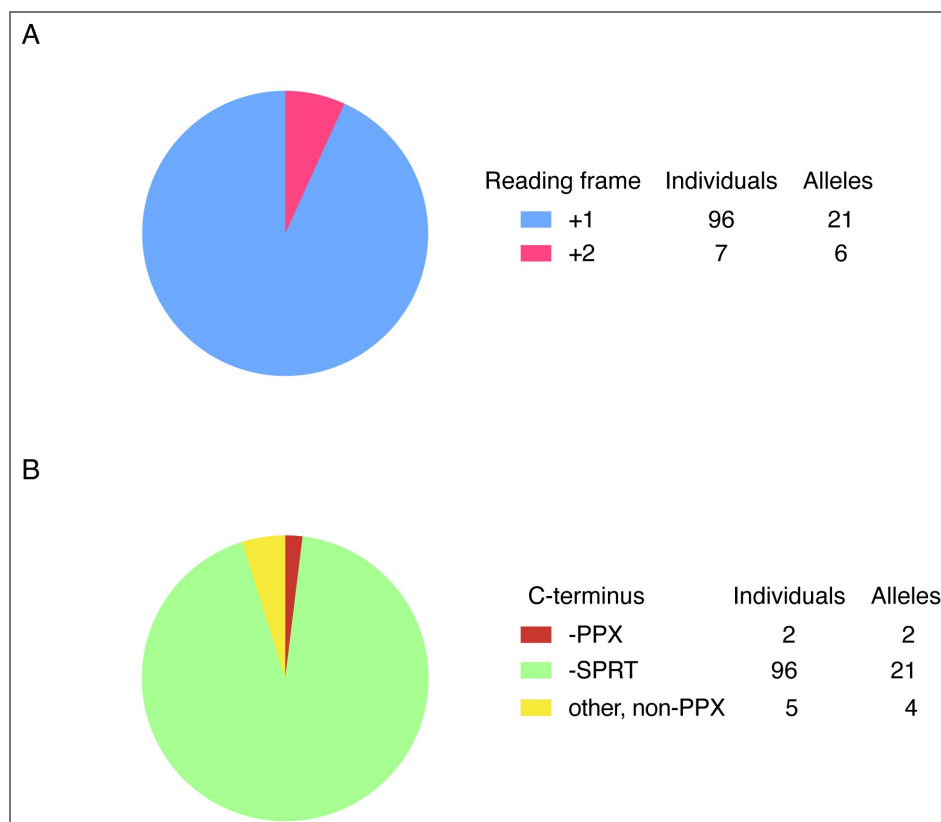
each amino acid position in the CT-DPR (hemizygous and heterozygous individuals). (C) Plot of the gnomAD missense allele count in the MBD for comparison with (A). Mutations present (filled diamonds), no gnomAD changes (empty diamonds).

Figure 2, Supplementary Figure 2. Summary of CTD C-terminal amino acid sequences and experimental MeCP2 protein levels for each.

																	MeCP2 level	Experimental system				
CTD2																						
human RTT mutation	-	H	S	E	S	P	K	A	P	V	P	L	L	P	P	L	P	P	X	low	LUHMES-derived neurons	
mouse allele	-	H	S	E	S	T	K	A	P	M	P	L	L	P	P	L	S	P	X	WT	mouse brain, ESC-derived neurons	
humanized mouse allele	-	H	S	E	S	P	K	A	P	V	P	L	L	P	P	L	P	P	X	low	ESC-derived neurons	
CTD1																						
human RTT mutation	-	H	S	E	S	P	K	A	P	V	P	L	L	P	P	H	Q	P	P	X	low	LUHMES-derived neurons
mouse allele	-	H	S	E	S	T	K	A	P	M	P	L	L	P	P	H	Q	P	P	X	low	mouse brain, ESC-derived neurons
humanized mouse allele	-	H	S	E	S	P	K	A	P	V	P	L	L	P	P	H	Q	P	P	X	low	ESC-derived neurons

Figure 3, Supplementary Figure 1. All frameshifting deletion alleles in gnomAD v4.1.0.

(A) Number of individuals with each type of frameshift and number of different CTD alleles. **(B)** GnomAD entries shown in (A) classified by C-terminal amino acid sequence for individuals and CTD alleles.



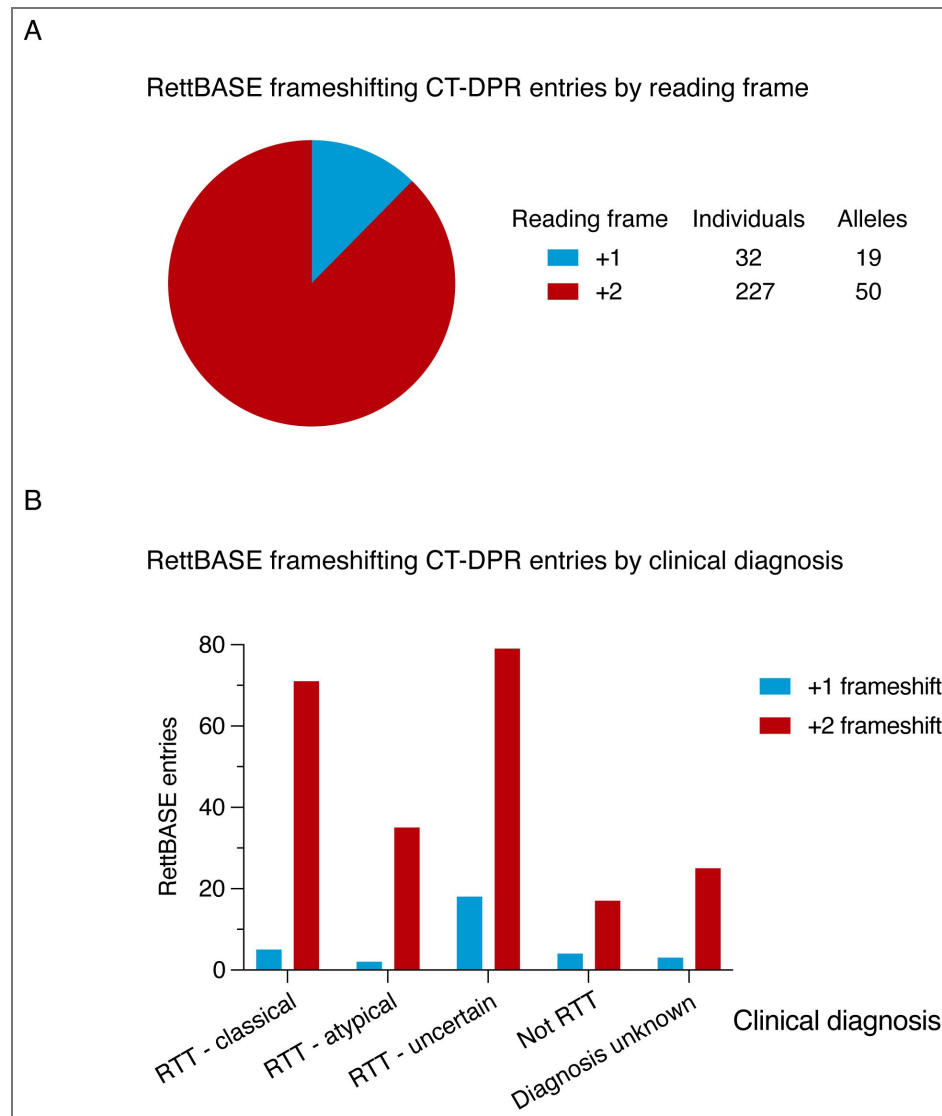


Figure 3, Supplementary Figure 2. All frameshifting deletion alleles in RettBASE.

(A) Number of individuals with each type of frameshift and number of different CTD alleles. (B) Entries shown in (A) grouped by clinical diagnosis displayed in RettBASE.

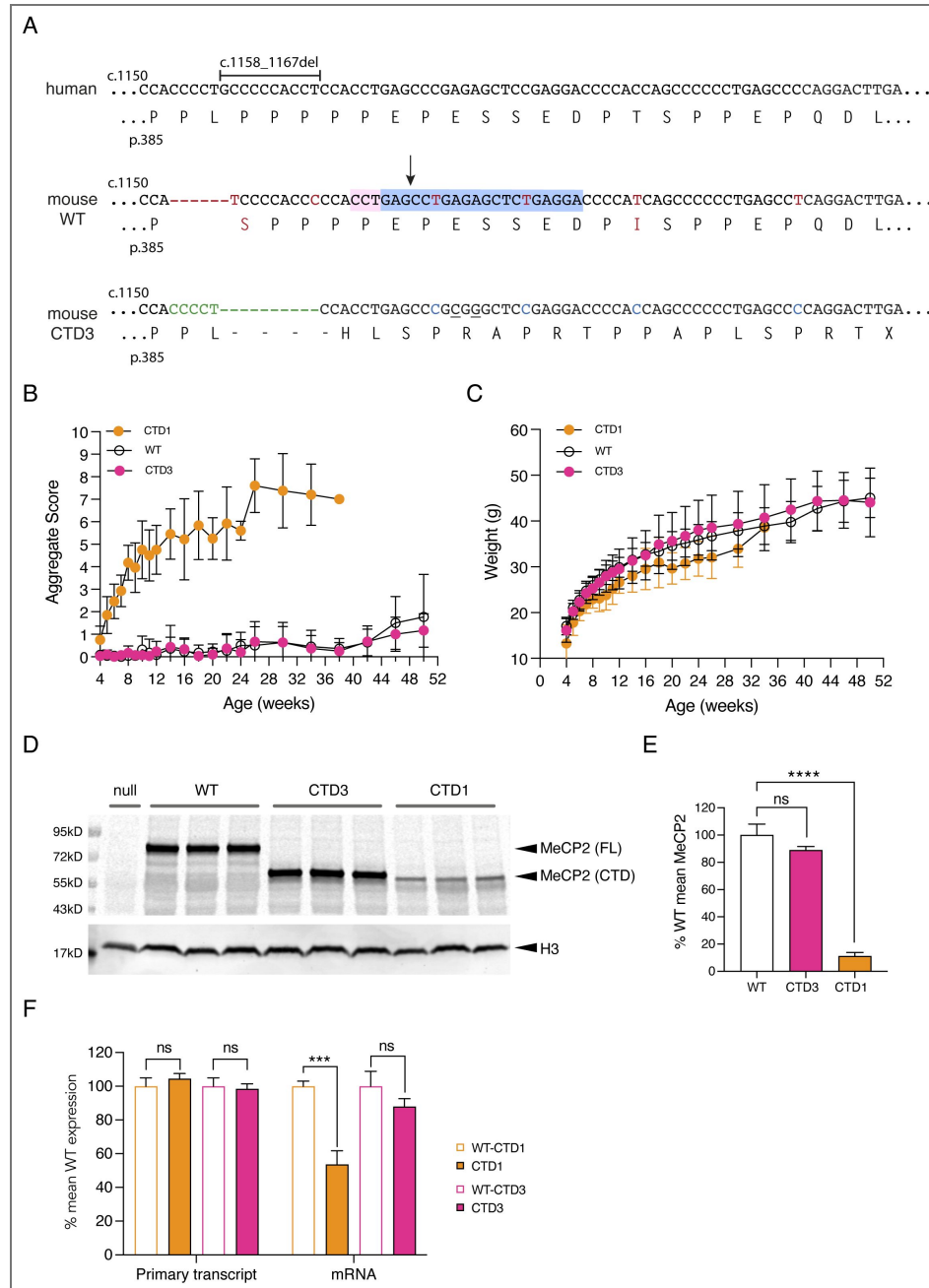


Figure 3, Supplementary Figure 3. CTD3 mouse knock-in allele.

(A) Human genomic and amino acid sequence showing location of CTD3 deletion (1158_1167del). Mouse WT genomic and amino acid sequences. Differences to human sequence are shown in red. The protospacer sequence (blue) and PAM sequence (pink) used to cut the WT allele for CRISPR editing are shown, with the cut site marked with an arrow. The mouse CTD3 knock-in allele is shown with nucleotide additions and deletions to the WT allele shown in green. Changes made to the mouse sequence to reproduce the human missense tail are in blue and two silent changes which introduce a diagnostic SacII site are underlined. (B) Phenotypic scoring of hemizygous male mice with CTD1 (n=13) and CTD3 (n=15) knock-in alleles, and WT male littermates of CTD3 animals (n=11). Mean $\pm$ sd. (C) Body weights of animals shown in (B). Mean $\pm$ sd. (D) Western blot of whole brain protein from 6 week old male mice hemizygous for *MeCP2*-null, CTD1, CTD3 and WT alleles. Full-length (FL) and C-terminally deleted (CTD) MeCP2 proteins are indicated. Histone H3 is used as a loading control. (E) Quantification of (D). N=3 per genotype, mean $\pm$ sd. Unpaired two-tailed t-test: CTD3 vs WT P=0.079 (ns), CTD1 vs WT P<0.0001 (****). (F) Quantification of *MeCP2* primary transcript and mRNA in whole brain of 6 week old male mice as in (D). N=3 brains per genotype. Mean $\pm$ sd. Unpaired two-tailed t-test: Primary transcript CTD3 vs WT littermates P=0.6687 (ns), CTD1 vs WT P=0.2585 (ns), mRNA CTD3 vs WT P=0.1055 (ns), CTD1 vs WT P=0.0008 (***).

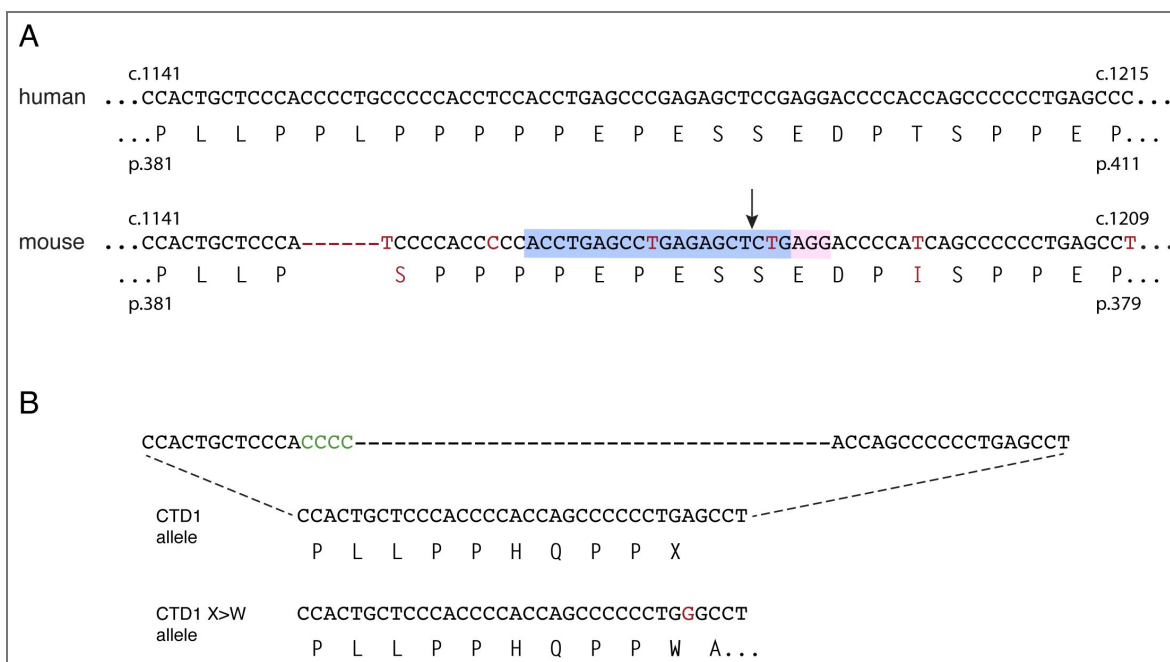


Figure 4, Supplementary Figure 1. Structure of the mouse CTD1 X>W allele.

(A) Human and mouse CT-DPR genomic and amino acid sequences. Differences between mouse and human sequence are shown in red. The protospacer sequence (blue) and PAM sequence (pink) used to cut the WT allele for CRISPR editing are shown, with the cut site marked with an arrow. (B) The mouse CTD1 knock-in allele compared to the CTD1 X>W allele. The single nucleotide A to G change is shown in red.

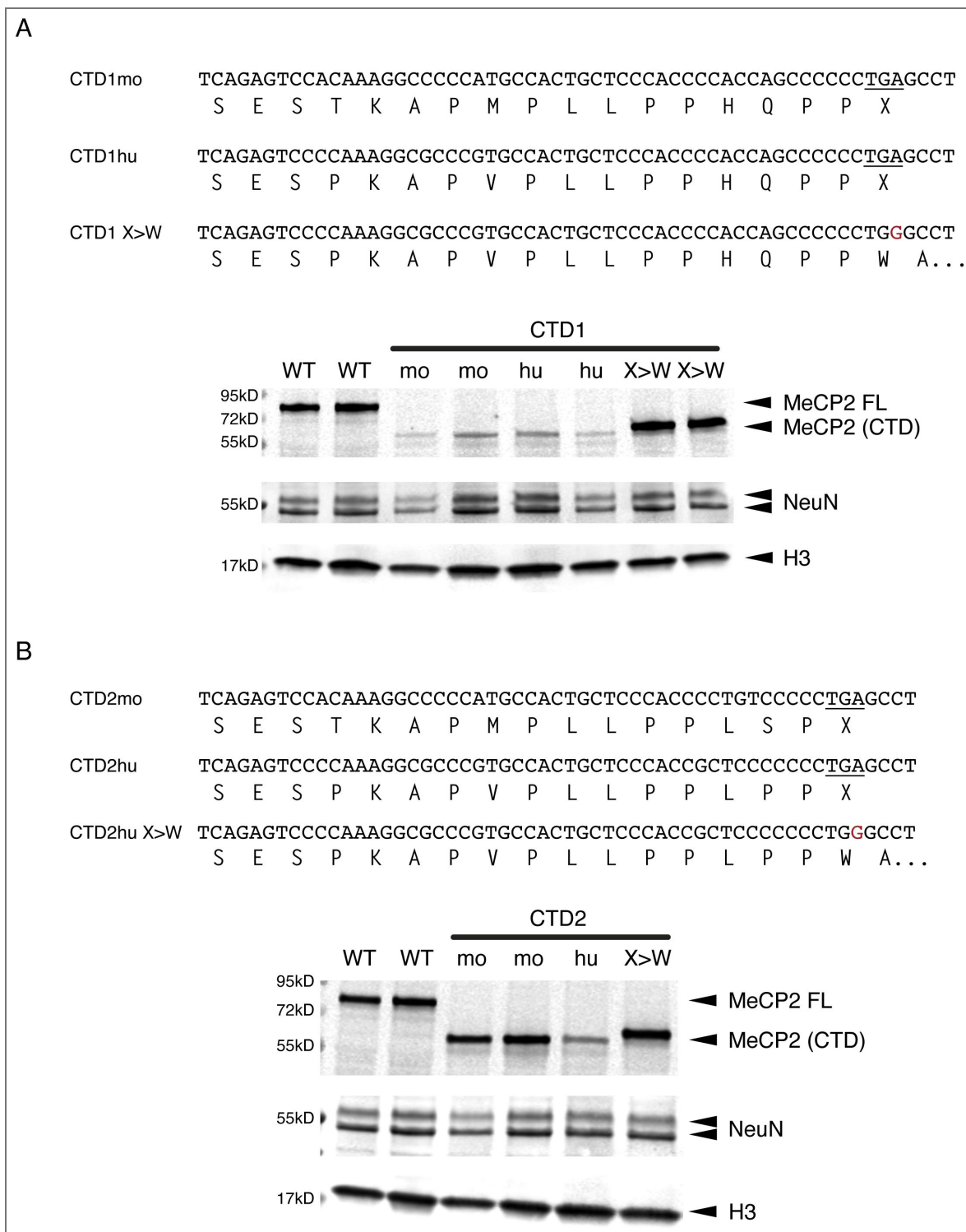


Figure 4, Supplementary Figure 2. CTD mouse alleles: mESC-derived neurons.

(A) genomic and amino acid sequences of 3 CTD1 mouse knock-in alleles and western blot of MeCP2 protein from mESC-derived neurons 7 days after plating neuronal progenitors. Two independent clones per genotype, histone H3 loading control, NeuN control for differentiation status. (B) genomic and amino acid sequences of 3 CTD2 mouse knock-in alleles and western blot of MeCP2 protein from mESC-derived neurons 7 days after plating neuronal progenitors.

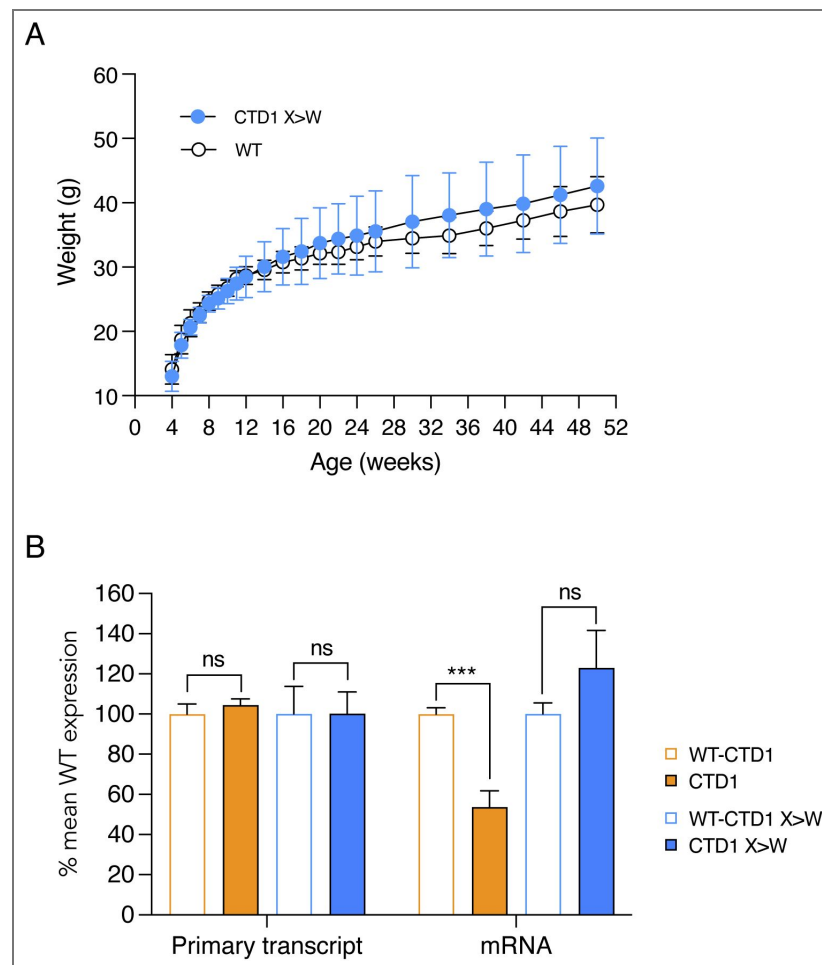


Figure 4, Supplementary Figure 3. CTD1 X>W knock-in mice: weights and brain RNA levels.

(A) Body weights of animals shown in Figure 4C and D. Mean $\pm$ sd. CTD1 ($n=13$) and CTD1 X>W ($n=9$), WT male littermates of CTD1 X>W animals ($n=10$). **(B)** Quantification of *Mecp2* primary transcript and mRNA in whole brain of 6 week old male mice. $N=3$ brains per genotype. Mean $\pm$ sd. Unpaired two-tailed t-test: Primary transcript CTD1 X>W vs WT littermates $P>0.9999$ (ns), CTD1 vs WT $P=0.2585$ (ns), mRNA CTD1 X>W vs WT $P=0.1419$ (ns), CTD1 vs WT $P=0.0008$ (***).

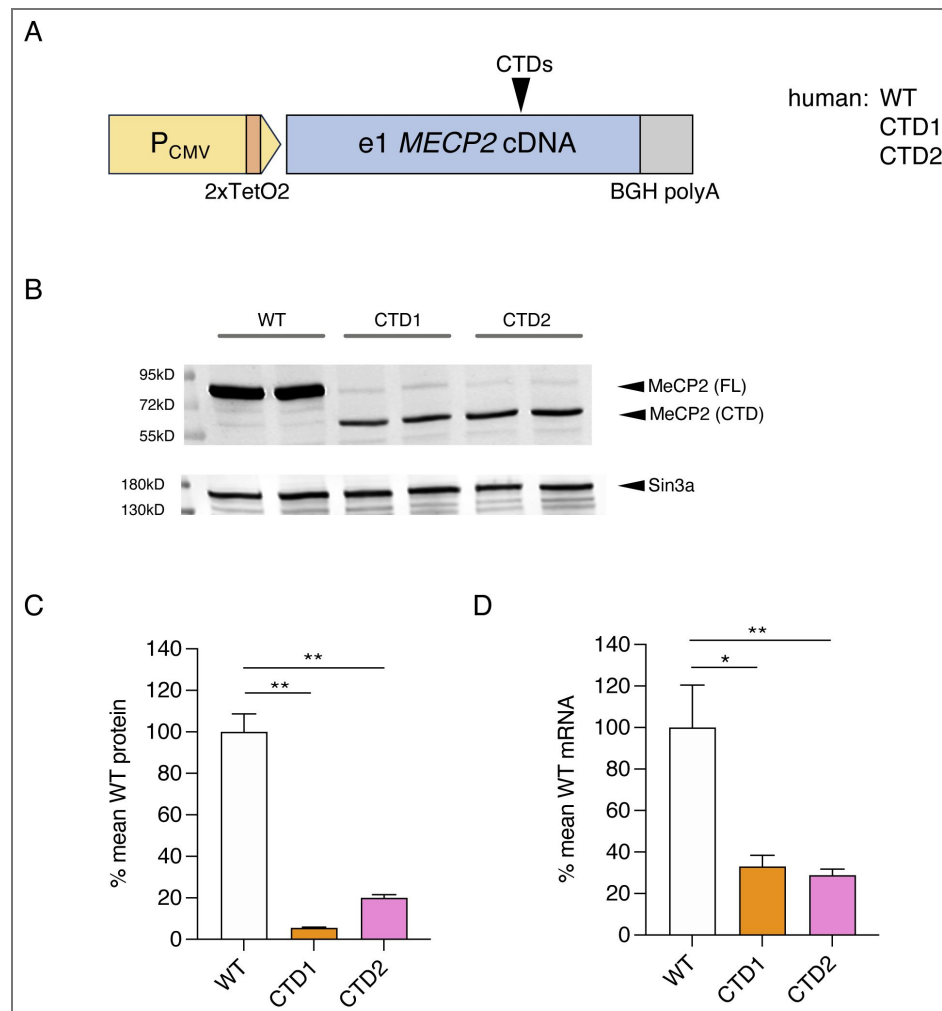


Figure 5, Supplementary Figure 1. Flp-In T-REx cell lines with human *MECP2* transgenes.

(A) Schematic of human *MECP2* transgenes in Flp-In T-REx cell lines. (B) Western blot with whole cell lysates from independent Flp-In T-REx clones carrying human cDNA transgenes (24 hours tetracycline induction). Sin3a loading control. (C) Quantification of MeCP2 protein expression from (B). N=2 clones per genotype, mean $\pm$ sd. Unpaired two-tailed t-test: WT vs CTD1 $P=0.0042$ (**), WT vs CTD2 $P=0.0060$ (**). (D) Quantification of *Mecp2* transgene mRNA from the same experiment as (B) and (C). N=2 clones per genotype, mean $\pm$ sd. Unpaired two-tailed t-test: WT vs CTD1 $P=0.0126$ (*), WT vs CTD2 $P=0.0099$ (**), WT vs CTD2mo $P=0.0171$ (*).

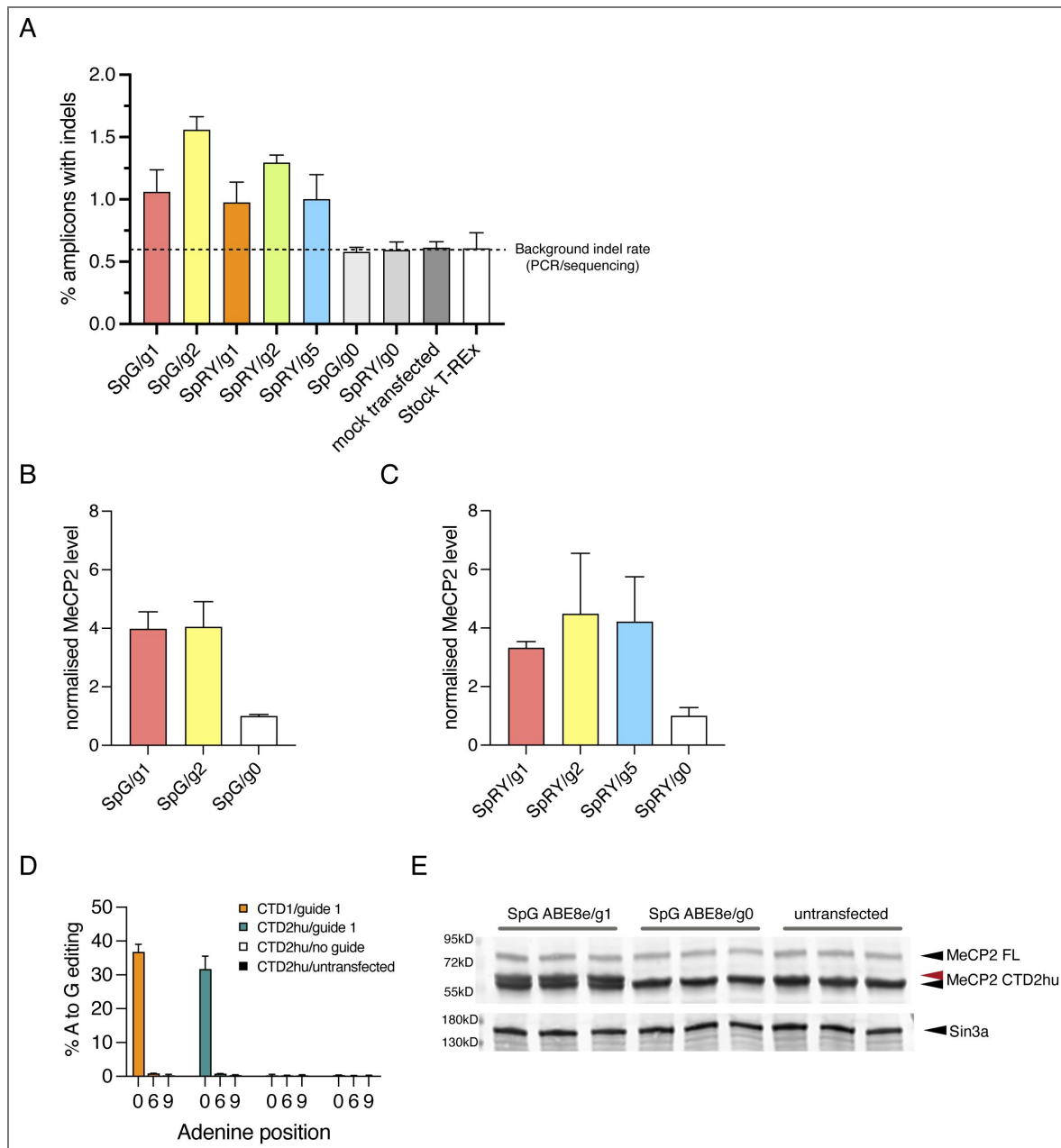


Figure 6, Supplementary Figure 1. Base editing of CTD transgenes in Flp-In T-REx cell lines.

(A) mouse CTD Flp-In T-REx cells treated with ABE8e/guide RNAs as shown in Figure 6C. Percentage of mapped reads with indels. Background indel rate is indicated based on the % indels found in amplicons from mock- or untransfected cells. (B) Quantification of total CTD MeCP2 levels from western blot in Figure 6D, SpG ABE8e. (C) Quantification of total CTD MeCP2 levels from western blot in Figure 6D, SpRY ABE8e. (D) Editing efficiency of SpG ABE8e/mouse guide 1 at target and bystander As in mouse CTD1 and CTD2hu Flp-In T-REx cells. Mean $\pm$ sd, n=3 transfections per ABE8e/guide combination. (E) western blot showing MeCP2 protein levels from the mouse CTD2hu cells in (D) after 24 hours induction of transgene expression. The red arrow indicates MeCP2 CTD1 protein after editing (CTD1 X>W).

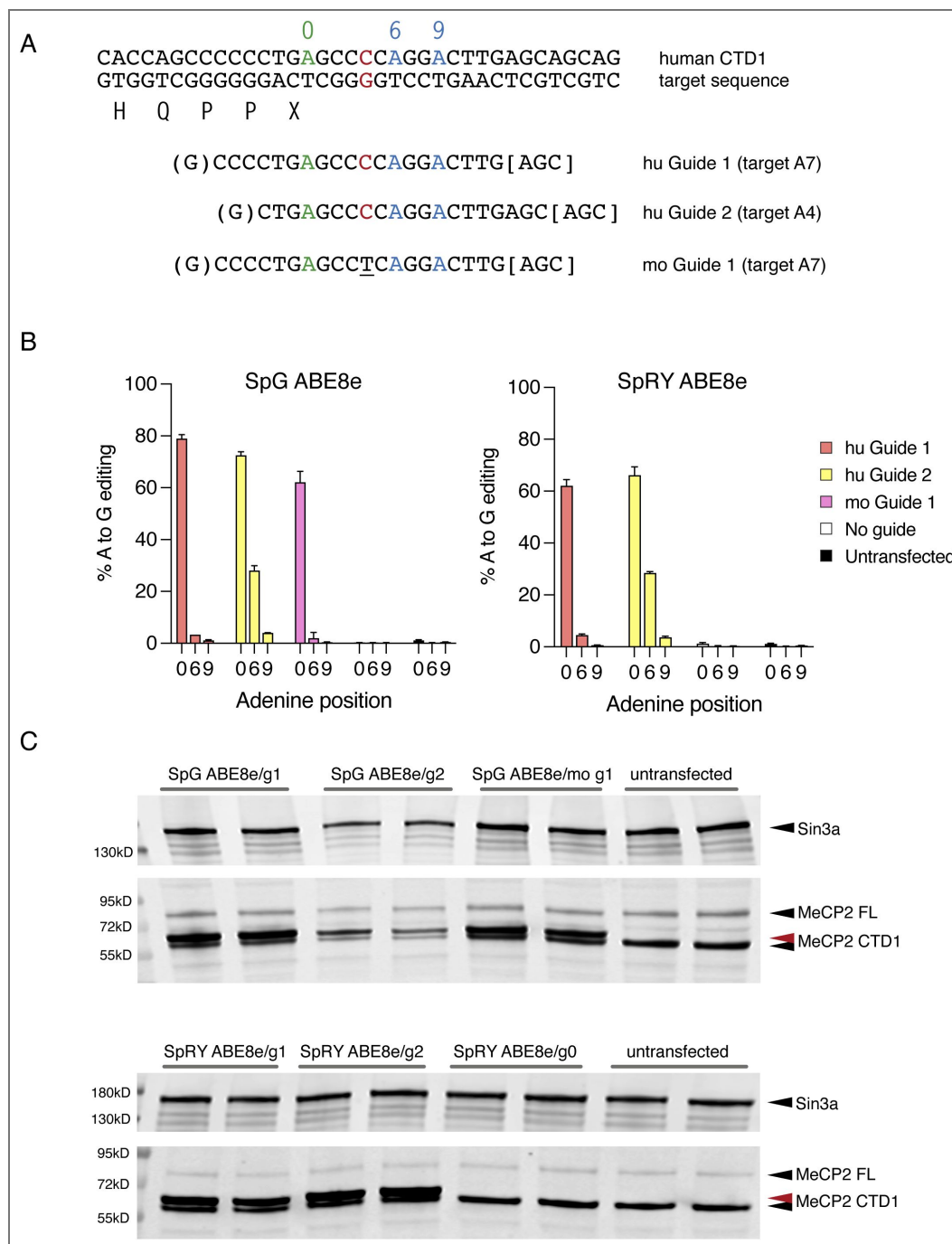


Figure 6, Supplementary Figure 2. Editing of human *MECP2* transgenes in Flp-In T-REx cells.

(A) Target genomic sequence and sgRNAs. Target A (position 0) is shown in green, with two bystander As within the guide sequence shown in blue (positions +6 and +9). The difference between human (red) and mouse (underlined) sequences is indicated. (B) Editing efficiency following transfection of human CTD1 Flp-In T-REx cells with ABE8e-SpG or SpRY base editors and guide RNA plasmids shown in (A). Editing efficiency at the target and bystander As is quantified by amplicon sequencing (n=3 transfections per ABE/guide combination). (C) western blot showing MeCP2 protein levels from the experiment in (B) after 24 hours induction of transgene expression. The red arrow indicates MeCP2 CTD1 protein after editing (CTD1 X>W). Sin3a loading control.

Data availability

Source data used to create panels Figures 1-3 and associated supplementary figures is contained in Supplementary Tables 1-8.

Acknowledgements

We would like to thank Emma Allen and the University of Edinburgh Central Transgenic Core (CTC) for production of mouse lines by blastocyst injection, Amanda Morris for MiSeq sequencing, Matthew Lyst, Jim Selfridge and Dirk-Jan Kleinjan for critical reading of the manuscript and John Christodoulou for helpful revision suggestions. This work was supported by Wellcome (Investigator Award 222507/Z/21/Z to AB), the Rett Syndrome Research Trust (Grant 13669477_13669480 to AB and JG), National Institutes of Health grant DP2CA281401 (B.P.K.), and the Simons Initiative for the Developing Brain (SIDB). A.B is a member of SIDB at the University of Edinburgh.

Additional files

Supplementary Table 1 All *MECP2* mutations held in RettBASE listed by location of first nucleotide change (4664 entries). Data downloaded 17/08/2017.

Supplementary Table 2 RettBASE frameshifting mutations in the CT-DPR (c.1110-1210). All entries, listed by start of frameshift.

Supplementary Table 3 RettBASE frameshifting alleles in the CT-DPR (c.1110-1210) listed by start of frameshift. The number of individuals with each allele is shown. Entries have been annotated with further details from RettBASE which were used to select a high confidence set of RTT CTD alleles (highlighted in pink): at least one individual with a diagnosis of Classical Rett syndrome and at least one individual with evidence for a *de novo* *MECP2* mutation. Frameshifts caused by insertions are also shown, with high confidence mutations highlighted in lilac. Alleles which are also found in gnomAD are indicated.

Supplementary Table 4 Set of 15 high confidence RTT CTD alleles selected from RettBASE.

Supplementary Table 5 gnomAD *MECP2* missense mutations in the CT-DPR listed by location. gnomAD release v4.1.0 downloaded 21/11/2024, nucleotide numbering according to transcript ENST00000303391.11.

Supplementary Table 6 gnomAD *MECP2* in-frame deletions in the CT-DPR. High confidence mutations found in at least one hemizygous individual are highlighted in green.

Supplementary Table 7 gnomAD *MECP2* missense mutations in the MBD listed by location (c.239-479).

Supplementary Table 8 gnomAD *MECP2* frameshifting deletion alleles in the CT-DPR. Nine high confidence alleles which are present in at least one hemizygous individual are highlighted in blue. Two single mutations with a -PPX ending are highlighted in yellow.

Additional information

Funding

Funder	Grant reference number	Author
Wellcome	https://doi.org/10.35802/222507	Adrian Bird
Rett Syndrome Research Trust (RSRT)	13669477_13669480	Adrian Bird Jacky Guy
HHS National Institutes of Health (NIH)	DP2CA281401	Ben Kleinstiver
Simons Initiative for the Developing Brain (SIDB)	13800235_1380239	Adrian Bird

Author ORCID iDs

Jacky Guy:  <https://orcid.org/0000-0002-8440-5667>

Huda Zoghbi:  <https://orcid.org/0000-0002-0700-3349>

Author notes

Competing interests: B.P.K. is an inventor on patents or patent applications filed by Mass General Brigham that describe genome engineering technologies, including the SpG and SpRY enzymes. B.P.K. has consulted for EcoR1 capital, Novartis Venture Fund, Foresite Labs, and Jumble Therapeutics, and is on the scientific advisory boards of Acrigen Biosciences, Life Edit Therapeutics, and Prime Medicine. B.P.K. has a financial interest in Prime Medicine, Inc., a company developing therapeutic CRISPR-Cas technologies for gene editing. B.P.K.'s interests were reviewed and are managed by MGH and MGB in accordance with their conflict-of-interest policies.

References

1. **Alves CRR**, Ha LL, Yaworski R, Sutton ER, Lazzarotto CR, Christie KA, Reilly A, Beauvais A, Doll RM, de la Cruz D, *et al.* (2024) Optimization of base editors for the functional correction of SMN2 as a treatment for spinal muscular atrophy. *Nature Biomedical Engineering* **8**:118-131 <https://doi.org/10.1038/s41551-023-01132-z> | PubMed
2. **Amir RE**, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**:185-8 <https://doi.org/10.1038/13810> | PubMed
3. **Anzalone AV**, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149-157 <https://doi.org/10.1038/s41586-019-1711-4> | PubMed
4. **Augenstein K**, Lane JB, Horton A, Schanen C, Percy AK (2009) Variable phenotypic expression of a MECP2 mutation in a family. *Journal of Neurodevelopmental Disorders* **1**:313 <https://doi.org/10.1007/s11689-009-9034-7> | PubMed
5. **Bain G**, Kitchens D, Yao M, Huettner JE, Gottlieb DI (1995) Embryonic stem cells express neuronal properties in vitro. *Dev Biol* **168**:342-57 <https://doi.org/10.1006/dbio.1995.1085> | PubMed
6. **Bajikar SS**, Zhou J, O'Hara R, Tirumala HP, Durham MA, Trostle AJ, Dias M, Shao Y, Chen H, Wang W, *et al.* (2025) Acute MeCP2 loss in adult mice reveals transcriptional and chromatin changes that precede neurological dysfunction and inform pathogenesis. *Neuron* **113**:380-395.e8. <https://doi.org/10.1016/j.neuron.2024.11.006> | PubMed
7. **Bebbington A**, Percy A, Christodoulou J, Ravine D, Ho G, Jacoby P, Anderson A, Pineda M, Ben Zeev B, Bahi-Buisson N, *et al.* (2010) Updating the profile of C-terminal MECP2 deletions in Rett syndrome. *Journal of Medical Genetics* **47**:242-248 <https://doi.org/10.1136/jmg.2009.072553> | PubMed
8. **Bienvenu T**, Carrie A, de Roux N, Vinet MC, Jonveaux P, Couvert P, Villard L, Arzimanoglou A, Beldjord C, Fontes M, *et al.* (2000) MECP2 mutations account for most cases of typical forms of Rett syndrome. *Hum Mol Genet* **9**:1377-84 <https://doi.org/10.1093/hmg/9.9.1377> | PubMed
9. **Chen PJ**, Liu DR (2023) Prime editing for precise and highly versatile genome manipulation. *Nature Reviews Genetics* **24**:161-177 <https://doi.org/10.1038/s41576-022-00541-1> | PubMed
10. **Chen S**, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, *et al.* (2024) A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**:92-100 <https://doi.org/10.1038/s41586-023-06045-0> | PubMed
11. **Cheng J**, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, *et al.* (2023) Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**:eadg7492 <https://doi.org/10.1126/science.adg7492> | PubMed

12. Cheng T-L, Wang Z, Liao Q, Zhu Y, Zhou W-H, Xu W, Qiu Z (2014) MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex. *Developmental Cell* **28**:547-560 <https://doi.org/10.1016/j.devcel.2014.01.032> | PubMed
13. Cheval H, Guy J, Merusi C, De Sousa D, Selfridge J, Bird A (2012) Postnatal inactivation reveals enhanced requirement for MeCP2 at distinct age windows. *Human Molecular Genetics* **21**:3806-3814 <https://doi.org/10.1093/hmg/dds208> | PubMed
14. Christodoulou J, Grimm A, Maher T, Bennetts B (2003) RettBASE: The IRSA MECP2 variation database—a new mutation database in evolution. *Hum Mutat* **21**:466-72 <https://doi.org/10.1002/humu.10194> | PubMed
15. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, et al. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* **37**:224-226 <https://doi.org/10.1038/s41587-019-0032-3> | PubMed
16. Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, Noebels JL, David Sweatt J, Zoghbi HY (2004) Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Human Molecular Genetics* **13**:2679-2689 <https://doi.org/10.1093/hmg/ddh282> | PubMed
17. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**:819-23 <https://doi.org/10.1126/science.1231143> | PubMed
18. Cuddapah VA, Pillai RB, Shekar KV, Lane JB, Motil KJ, Skinner SA, Tarquinio DC, Glaze DG, McGwin G, Kaufmann WE, et al. (2014) Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *J Med Genet* **51**:152-8 <https://doi.org/10.1136/jmedgenet-2013-102113> | PubMed
19. Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, Musunuru K (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**:393-394 <https://doi.org/10.1016/j.stem.2013.03.006> | PubMed
20. Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**:89-93 <https://doi.org/10.1038/nature14319> | PubMed
21. Gadalla KKE, Vudhironarit T, Hector RD, Sinnett S, Bahey NG, Bailey MES, Gray SJ, Cobb SR (2017) Development of a Novel AAV Gene Therapy Cassette with Improved Safety Features and Efficacy in a Mouse Model of Rett Syndrome. *Molecular Therapy Methods & Clinical Development* **5**:180-190 <https://doi.org/10.1016/j.omtm.2017.04.007> | PubMed
22. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**:464-471 <https://doi.org/10.1038/nature24644> | PubMed
23. Gerbracht JV, Boehm V, Gehring NH (2017) Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. *Scientific Reports* **7**:10616 <https://doi.org/10.1038/s41598-017-10847-4> | PubMed
24. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, Genome Aggregation Database Consortium, Rehm HL, MacArthur DG, O'Donnell-Luria A (2022) Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation* **43**:1012-1030 <https://doi.org/10.1002/humu.24309> | PubMed
25. Gutierrez E, Shin B-S, Woolstenhulme CJ, Kim J-R, Saini P, Buskirk AR, Dever TE (2013) eIF5A promotes translation of polyproline motifs. *Molecular Cell* **51**:35-45 <https://doi.org/10.1016/j.molcel.2013.04.021> | PubMed
26. Guy J, Alexander-Howden B, FitzPatrick L, DeSousa D, Koerner MV, Selfridge J, Bird A (2018) A mutation-led search for novel functional domains in MeCP2. *Human Molecular Genetics* **27**:2531-2545 <https://doi.org/10.1093/hmg/ddy159> | PubMed
27. Guy J, Gan J, Selfridge J, Cobb S, Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315**:1143-1147 <https://doi.org/10.1126/science.1138389> | PubMed

28. Hagberg B (2002) Clinical manifestations and stages of Rett syndrome. *Mental Retardation and Developmental Disabilities Research Reviews* **8**:61-65 <https://doi.org/10.1002/mrdd.10020> | PubMed
29. Hagberg B, Aicardi J, Dias K, Ramos O (1983) A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann Neurol* **14**:471-9 <https://doi.org/10.1002/ana.410140412> | PubMed
30. Hayes CS, Bose B, Sauer RT (2002) Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *The Journal of Biological Chemistry* **277**:33825-33832 <https://doi.org/10.1074/jbc.M205405200> | PubMed
31. Huter P, Arenz S, Bock LV, Graf M, Frister JO, Heuer A, Peil L, Starosta AL, Wohlgemuth I, Peske F, *et al.* (2017) Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Molecular Cell* **68**:515-527.e6. <https://doi.org/10.1016/j.molcel.2017.10.014> | PubMed
32. Janzen DM, Frolova L, Geballe AP (2002) Inhibition of translation termination mediated by an interaction of eukaryotic release factor 1 with a nascent peptidyl-tRNA. *Molecular and Cellular Biology* **22**:8562-8570 <https://doi.org/10.1128/MCB.22.24.8562-8570.2002> | PubMed
33. Kinde B, Wu DY, Greenberg ME, Gabel HW (2016) DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc Natl Acad Sci U S A* **113**:15114-15119 <https://doi.org/10.1073/pnas.1618737114> | PubMed
34. Koerner MV, FitzPatrick L, Selfridge J, Guy J, De Sousa D, Tillotson R, Kerr A, Sun Z, Lazar MA, Lyst MJ, *et al.* (2018) Toxicity of overexpressed MeCP2 is independent of HDAC3 activity. *Genes & Development* **32**:1514-1524 <https://doi.org/10.1101/gad.320325.118> | PubMed
35. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**:420-424 <https://doi.org/10.1038/nature17946> | PubMed
36. Krishnaraj R, Ho G, Christodoulou J (2017) RettBASE: Rett syndrome database update. *Human Mutation* **38**:922-931 <https://doi.org/10.1002/humu.23263> | PubMed
37. Kruusvee V, Lyst MJ, Taylor C, Tarnauskaite Z, Bird AP, Cook AG (2017) Structure of the MeCP2-TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. *Proc Natl Acad Sci U S A* **114**:E3243-E3250 <https://doi.org/10.1073/pnas.1700731114> | PubMed
38. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research* **44**:D862-868 <https://doi.org/10.1093/nar/gkv1222> | PubMed
39. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**:285-291 <https://doi.org/10.1038/nature19057> | PubMed
40. Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**:905-14 [https://doi.org/10.1016/0092-8674\(92\)90610-o](https://doi.org/10.1016/0092-8674(92)90610-o) | PubMed
41. Li CH, Coffey EL, Dall'Agnese A, Hannett NM, Tang X, Henninger JE, Platt JM, Oksuz O, Zamudio AV, Afeyan LK, *et al.* (2020) MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* **586**:440-444 <https://doi.org/10.1038/s41586-020-2574-4> | PubMed
42. Li M, Pevny L, Lovell-Badge R, Smith A (1998) Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr Biol* **8**:971-4 [https://doi.org/10.1016/s0960-9822\(98\)70399-9](https://doi.org/10.1016/s0960-9822(98)70399-9) | PubMed
43. Lyst MJ, Bird A (2015) Rett syndrome: a complex disorder with simple roots. *Nature Reviews. Genetics* **16**:261-275 <https://doi.org/10.1038/nrg3897> | PubMed
44. Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, *et al.* (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nature Neuroscience* **16**:898-902 <https://doi.org/10.1038/nn.3434> | PubMed

45. Matheisl S, Berninghausen O, Becker T, Beckmann R (2015) Structure of a human translation termination complex. *Nucleic Acids Research* **43**:8615-8626 <https://doi.org/10.1093/nar/gkv909> | PubMed
46. Moncla A, Kpebe A, Missirian C, Mancini J, Villard L (2002) Polymorphisms in the C-terminal domain of MECP2 in mentally handicapped boys: implications for genetic counselling. *Eur J Hum Genet* **10**:86-9 <https://doi.org/10.1038/sj.ejhg.5200761> | PubMed
47. Nan X, Meehan RR, Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* **21**:4886-4892 <https://doi.org/10.1093/nar/21.21.4886> | PubMed
48. Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**:386-389 <https://doi.org/10.1038/30764> | PubMed
49. Philippe C, Villard L, De Roux N, Raynaud M, Bonnefond JP, Pasquier L, Lesca G, Mancini J, Jonveaux P, Moncla A, et al. (2006) Spectrum and distribution of MECP2 mutations in 424 Rett syndrome patients: a molecular update. *Eur J Med Genet* **49**:9-18 <https://doi.org/10.1016/j.ejmg.2005.04.003> | PubMed
50. Ramocki MB, Tavyev YJ, Peters SU (2010) The MECP2 duplication syndrome. *Am J Med Genet A* **152A**:1079-88 <https://doi.org/10.1002/ajmg.a.33184> | PubMed
51. Ravn K, Roende G, Duno M, Fuglsang K, Eiklid KL, Tümer Z, Nielsen JB, Skjeldal OH (2011) Two new Rett syndrome families and review of the literature: expanding the knowledge of MECP2 frameshift mutations. *Orphanet Journal of Rare Diseases* **6**:58 <https://doi.org/10.1186/1750-1172-6-58> | PubMed
52. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, et al. (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology* **38**:883-891 <https://doi.org/10.1038/s41587-020-0453-z> | PubMed
53. Ross PD, Gadalla KKE, Thomson SR, Selfridge J, Bahey NG, Benito J, Burstein SR, McMinn R, Bolon B, Hector RD, et al. (2025) Self-regulating gene therapy ameliorates phenotypes and overcomes gene dosage sensitivity in a mouse model of Rett syndrome. *Science Translational Medicine* **17**:eadq3614 <https://doi.org/10.1126/scitranslmed.adq3614> | PubMed
54. Ross PD, Guy J, Selfridge J, Kamal B, Bahey N, Tanner KE, Gillingwater TH, Jones RA, Loughrey CM, McCarroll CS, et al. (2016) Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. *Hum Mol Genet* **25**:4389-4404 <https://doi.org/10.1093/hmg/ddw269> | PubMed
55. Sadhu C, Lyons C, Oh J, Jagadeeswaran I, Gray SJ, Sinnott SE (2023) The Efficacy of a Human-Ready miniMECP2 Gene Therapy in a Pre-Clinical Model of Rett Syndrome. *Genes* **15**:31 <https://doi.org/10.3390/genes15010031> | PubMed
56. Schuller AP, Wu CC-C, Dever TE, Buskirk AR, Green R (2017) eIF5A Functions Globally in Translation Elongation and Termination. *Molecular Cell* **66**:194-205.e5. <https://doi.org/10.1016/j.molcel.2017.03.003> | PubMed
57. Sinnott SE, Hector RD, Gadalla KKE, Heindel C, Chen D, Zaric V, Bailey MES, Cobb SR, Gray SJ (2017) Improved MECP2 Gene Therapy Extends the Survival of MeCP2-Null Mice without Apparent Toxicity after Intracisternal Delivery. *Molecular Therapy Methods & Clinical Development* **5**:106-115 <https://doi.org/10.1016/j.omtm.2017.04.006> | PubMed
58. Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, Andrews R, Bird AP (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol Cell* **37**:457-68 <https://doi.org/10.1016/j.molcel.2010.01.030> | PubMed
59. Smeets E, Terhal P, Casaer P, Peters A, Midro A, Schollen E, van Roozendaal K, Moog U, Matthijs G, Herbergs J, et al. (2005) Rett syndrome in females with CTS hot spot deletions: a disorder profile. *Am J Med Genet A* **132A**:117-20 <https://doi.org/10.1002/ajmg.a.30410> | PubMed

60. Tillotson R, Selfridge J, Koerner MV, Gadalla KKE, Guy J, De Sousa D, Hector RD, Cobb SR, Bird A (2017) Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. *Nature* **550**:398-401 <https://doi.org/10.1038/nature24058> | PubMed
61. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP (2020) Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**:290-296 <https://doi.org/10.1126/science.aba8853> | PubMed
62. Wan M, Lee SS, Zhang X, Houwink-Manville I, Song HR, Amir RE, Budden S, Naidu S, Pereira JL, Lo IF, *et al.* (1999) Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *American Journal of Human Genetics* **65**:1520-1529 <https://doi.org/10.1086/302690> | PubMed
63. Yusufzai TM, Wolffe AP (2000) Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Research* **28**:4172-4179 <https://doi.org/10.1093/nar/28.21.4172> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

The authors scrutinized difference in C terminal region variant profiles between Rett syndrome patients and healthy individuals and pinpointed that subtle genetic alternation can cause benign or pathogenic output, which harbors important implication in Rett syndrome diagnosis and proposing therapeutic strategy. This work will be beneficial to clinicians and basic scientists who work on Rett syndrome and carries potential to be applied to other Mendelian rare diseases.

Strengths:

Well-designed genetic and molecular experiments, translating genetic differences into functional and clinical changes. This is a unique study resolving subtle changes in sequences give rise to dramatic phenotypic consequences.

Comments on revised version.

Improvements were made during the revision.

<https://doi.org/10.7554/eLife.109170.2.sa2>

Reviewer #2 (Public review):

Summary:

This study by Guy and Bird and colleagues is a natural follow-up to their 2018 Human Molecular Genetics paper, further clarifying the molecular basis of C-terminal deletions (CTDs) in MECP2 and how they contribute to Rett syndrome. The authors combine human genetic data with well-designed experiments in embryonic stem cells, differentiated neurons, and knock-in mice to explain why some CTD mutations are disease-causing while others are harmless. They show that pathogenic mutations create a specific amino acid motif at the C-terminus, where +2 frameshifts produce a PPX ending that greatly reduces MeCP2 protein levels (likely due to translational stalling) whereas +1 frameshifts generating SPRTX endings are well tolerated.

Strengths:

This is a comprehensive and rigorous study that convincingly pinpoints the molecular mechanism behind CTD pathogenicity, with strong agreement between the cell-based and

animal data. The authors also provide a proof of principle that modifying the PPX termination codon can restore MeCP2-CTD protein levels and rescue symptoms in mice. In addition, they demonstrate that adenine base editing can correct this defect in cultured cells and increase MeCP2-CTD protein levels. Overall, this is a well-executed study that provides important mechanistic and translational insight into a clinically important class of MECP2 mutations.

Weaknesses:

The adenine base editing to change the termination codon is shown feasible in generated cell lines, but yet to be shown in vivo in animal models.

Comments on revised version.

The authors have addressed all of my questions and comments.

<https://doi.org/10.7554/eLife.109170.2.sa1>

Author response:

The following is the authors' response to the original reviews.

This is a summary of the changes that have been made to the Reviewed Preprint:

- (1) The data from RettBASE which was analysed in the manuscript has been added in the form of four supplementary tables. Supplementary Table 1 contains the download of all MeCP2 mutations contained in RettBASE. Supplementary Tables 2-4 contain subsets of this data which were used in Figure 2B and Figure 3 SF2. Supplementary Table 4 also has the HGVS nomenclature for both e1 and e2 isoforms and the ClinVar Variation ID for each allele. Wording has been changed to clarify that analysis in the manuscript used this data from RettBASE and not the information that was deposited in ClinVar.
- (2) Similarly, Supplementary Tables 5-8 contain the gnomAD data that was used in the preparation of Figure 2, Figure 2 SF1 and Figure 3 SF1. The "high confidence" alleles in Supplementary Table 8 have been annotated with their HGVS names.
- (3) The criteria for selecting "high confidence" RettBASE and gnomAD alleles have been more explicitly stated in both the Results and Materials and Methods sections.
- (4) An additional "high confidence" RTT mutation (c.1152_1195) has been added to Figures 2B and 3B.
- (5) Figure 3 Supplementary Figure 2 has been added to show reading frame data for all frameshifting deletions in the C-terminal deletion-prone region (CT-DPR), showing that +2 frameshifts predominate in this larger data set, not just in the "high confidence" set. This has necessitated changing the previous Fig. 3 SF2 to Fig. 3 SF3.
- (6) A summary of the genetic alterations described in the manuscript, and their outcomes, has been added as Figure 7.
- (7) A simple flow chart which assists in the classification of human CTDs as "likely benign" or "likely pathogenic" has been added as Figure 8. This will aid future assessment of novel mutations in this region.
- (8) Additions have been made to the Materials and Methods section to comply with reporting guidelines.

(9) Minor changes have been made to the text to correct typographical errors and to clarify meaning.

Public Reviews:

Reviewer #1 (Public review):

Summary:

The authors scrutinized differences in C-terminal region variant profiles between Rett syndrome patients and healthy individuals and pinpointed that subtle genetic alternation can cause benign or pathogenic output, which harbors important implications in Rett syndrome diagnosis and proposes a therapeutic strategy. This work will be beneficial to clinicians and basic scientists who work on Rett syndrome, and carries the potential to be applied to other Mendelian rare diseases.

Strengths:

Well-designed genetic and molecular experiments translate genetic differences into functional and clinical changes. This is a unique study resolving subtle changes in sequences that give rise to dramatic phenotypic consequences.

Weaknesses:

There are many base-editing and protein-expression changes throughout the manuscript, and they cause confusion. It would be helpful to readers if authors could provide a simple summary diagram at the end of the paper.

We have added a summary diagram, as suggested (Figure 7). We have also provided a flowchart which shows how to classify human CTDs as “likely benign” or “likely pathogenic” based on their location.

Reviewer #2 (Public review):

Summary:

This study by Guy and Bird and colleagues is a natural follow-up to their 2018 Human Molecular Genetics paper, further clarifying the molecular basis of C-terminal deletions (CTDs) in MECP2 and how they contribute to Rett syndrome. The authors combine human genetic data with well-designed experiments in embryonic stem cells, differentiated neurons, and knock-in mice to explain why some CTD mutations are disease-causing while others are harmless. They show that pathogenic mutations create a specific amino acid motif at the C-terminus, where +2 frameshifts produce a PPX ending that greatly reduces MeCP2 protein levels (likely due to translational stalling) whereas +1 frameshifts generating SPRTX endings are well tolerated.

Strengths:

This is a comprehensive and rigorous study that convincingly pinpoints the molecular mechanism behind CTD pathogenicity, with strong agreement between the cell-based and animal data. The authors also provide a proof of principle that modifying the PPX termination codon can restore MeCP2-CTD protein levels and rescue symptoms in mice. In addition, they demonstrate that adenine base editing can correct this defect in cultured cells and increase MeCP2-CTD protein levels. Overall, this is a well-executed study that provides important mechanistic and translational insight into a clinically important class of MECP2 mutations.

Weaknesses:

The adenine base editing to change the termination codon is shown to be feasible in generated cell lines, but has yet to be shown in vivo in animal models.

This work is the obvious next step and is in progress. However, with the rise in pre- and neonatal genetic testing we felt it was important to disseminate our findings as soon as possible. The family pedigree in Figure 3C is a clear illustration of this need

Reviewer #3 (Public review):

Summary:

Guy et al. explored the variation in the pathogenicity of carboxy-terminal frameshift deletions in the X-linked MECP2 gene. Loss-of-function variants in MECP2 are associated with Rett syndrome, a severe neurodevelopmental disorder. Although 100's of pathogenic MECP2 variants have been found in people with Rett syndrome, 8 recurrent point mutations are found in ~65% of disease cases, and frameshift insertions/deletions (indels) variants resulting in production of carboxy-terminal truncated (CTT) MeCP2 protein account for ~10% of cases. Many of these occur in a "deletion prone region" (DPR) between c.1110-1210, with common recurrent deletions c.1157-1197del (CTD1) and c.1164_1207del (CTD2). While two major protein functional domains have been defined in MeCP2, the methyl-binding domain (MBD) and the NCoR interacting domain (NID), the functional role of the carboxy-terminal domain (CTD, beyond the NID, predicted to have a disordered protein structure) has not been identified, and previous work by this group and others demonstrated that a Mecp2 "minigene" lacking the CTD retains MeCP2 function suggesting that the CTD is dispensable. This raises an important question: If the CTD is dispensable, what is the pathogenic basis of the various CTT frameshift variants? Prior work from this group demonstrated that genetically engineered mice expressing the CTD1 variant had decreased expression of Mecp2 RNA and MeCP2 protein and decreased survival, but those expressing the CTD2 variant had normal Mecp2 RNA and protein and survival. However, they noted that differences between the mouse and human coding sequences resulted in different terminal sequences between the two common CTD, with CTD1 ending in -PPX in both mouse and human, but CTD2 ending in -PPC in human but -SPX in mouse, and in the previous paper they demonstrated in humanized mouse ES cells (edited to have the same -PPX termination) containing the CTD2 deletion resulted in decreased Mecp2 RNA and protein levels. This previous work provides the underlying hypotheses that they sought to explore, which is that the pathological basis of disease causing CTD relates to the formation of truncated proteins that end with a specific amino acid sequence (-PPX), which leads to decreased mRNA and protein levels, whereas tolerated, non-pathogenic CTD do not lead to production of truncated proteins ending in this sequence and retain normal mRNA/protein expression.

In this manuscript, they evaluate missense variants, in-frame deletions, and frame shift deletions within the DPR from the aggregated Genome Aggregated Database (gnomAD) and find that the "apparently" normal individuals within gnomAD had numerous tolerated missense variants and in-frame deletions within this region, as well as frameshift deletions (in hemizygous males) in the defined region. All of the gnomAD deletions within this region resulted in terminal amino acid sequences -SPRTX (due to +1 frameshift), whereas nearly all deletion variants in this region from people with Rett syndrome (from the Clinvar copy of the former RettBase database) had a terminal -PPX sequence, due to a +2 frameshift. They hypothesized that terminal proline codons causing ribosomal stalling and "nonsense mediated decay like" degradation of mRNA (with subsequent decreased protein expression) was the basis of the specific pathogenicity of the +2 frameshift variants, and that utilizing adenine base editors (ABE) to convert the termination codon to a tryptophan could correct this issue. They demonstrate this by engineering the change into mouse embryonic stem cell lines and

mouse lines containing the CTD1 deletion and show that this change normalized Mecp2 mRNA and protein levels and mouse phenotypes. Finally, they performed an initial proof-of-concept in an inducible HEK cell line and showed the ability of targeted ABE to edit the correct adenine and cause production of the expected larger truncated Mecp2 protein from CTD1 constructs.

The findings of this manuscript provide a level of support for their hypothesis about the pathogenicity versus non-pathogenicity of some MECP2 CTT intragenic deletions and provide preliminary evidence for a novel therapeutic approach for Rett syndrome; however, limitations in their analysis do not fully support the broader conclusions presented.

Strengths:

(1) Utilization of publicly available databases containing aggregated genetic sequencing data from adult cohorts (gnomAD) and people with Rett syndrome (Clinvar copy of RettBase) to compare differences in the composition of the resulting terminal amino acid sequences resulting from deletions presumed to be pathogenic (n+2) versus presumed to be tolerated (n+1).

(2) Evaluation of a unique human pedigree containing an n+1 deletion in this region that was reported as pathogenic, with demonstration of inheritance of this from the unaffected father and presence within other unaffected family members.

(3) Development of a novel engineered mouse model of a previously assumed n+1 pathogenic variant to demonstrate lack of detrimental effect, supporting that this is likely a benign variant and not causative of Rett syndrome.

(4) Creation and evaluation of novel cell lines and mouse models to test the hypothesis that the pathogenicity of the n+2 deletion variants could be altered by a single base change in the frameshifted stop codon.

(5) Initial proof-of-concept experiments demonstrating the potential of ABE to correct the pathogenicity of these n+2 deletion variants.

Weaknesses:

(1) While the use of the large aggregated gnomAD genetic data benefits from the overall size of the data, the presence of genetic variants within this collection does not inherently mean that they are "neutral" or benign. While gnomAD does not include children, it does include aggregated data from a variety of projects targeting neuropsychiatric (and other conditions), so there is information in gnomAD from people with various medical/neuropsychiatric conditions. The authors do make some acknowledgement of this and argue that the presence of intragenic deletion variants in their region of interest in hemizygous males indicates that it is highly likely that these are tolerated, non-pathogenic variants. Broadly, it is likely true that gnomAD MECP2 variants found in hemizygous males are unlikely to cause Rett syndrome in heterozygous females, it does not necessarily mean that these variants have no potential to cause other, milder, neuropsychiatric disorders. As a clear example, within gnomAD, there is a hemizygous male with the rs28934908 C>T variant that results in p.A140V (p.A152V in e1 transcript numbering convention). This pathogenic variant has been found in a number of pedigrees with an X-linked intellectual disability pattern, in which males have a clear neurodevelopmental disorder and heterozygous females have mild intellectual disability (see PMIDs 12325019, 24328834 as representative examples of a large number of publications describing this). Thus, while their claim that hemizygous deletion variants in gnomAD are unlikely to cause Rett syndrome, that cannot make the definitive statement

that they are not pathogenic and completely benign, especially when only found in a very small number of individuals in gnomAD.

We have included the possibility that mutations found in gnomAD may give rise to less severe neurological conditions in the discussion.

(2) The authors focus exclusively on deletions within the "DPR", they define as between c.1110-1210 and say that these deletions account for 10% of Rett syndrome cases. However, the published studies that are the basis for this 10% estimate include all genetic variants (frameshift deletions, insertions, complex insertion/deletions, nonsense variants) resulting in truncations beyond the NID. For example, Bebbington 2010 (PMID: 19914908), which includes frameshift indels as early as c.905 and beyond c.1210. Further specific examples from RettBase are described below, but the important point is that their evaluation of only frameshift variants within c.1110-1210 is not truly representative of the totality of genetic variants that collectively are considered CTT and account for 10% of Rett cases.

The vast majority of C-terminal truncating mutations do occur within the "CT-DPR", likely due to its C-rich nucleotide sequence and the presence of microhomologies within the region. Looking at frameshifting deletions in RettBASE that start after the NID, a large proportion of these end within the CT-DPR and result in a -PPX ending. We decided to restrict our analysis to the c.1110-1210 region to avoid including the rarer examples that may have a different reason for their pathogenicity. We do not assert that all C-terminal truncations are pathogenic due to this mechanism, but current evidence suggests that most are.

(3) The authors say that they evaluated the putative pathogenic variants contained within RettBase (which is no longer available, but the data were transferred to Clinvar) for all cases with Classic Rett syndrome and de novo deletion variants within their defined DPR domain. Looking at the data from the Clinvar copy of RettBase, there are a number (n=143) of c-terminal truncating variants (either frameshift or nonsense) present beyond the NID, but the authors only discuss 14 deletion frameshift variants in this manuscript. A number of these variants have molecular features that do not fall into the pathogenic classification proposed by the authors and are not addressed in the manuscript and do not support the generalization of the conclusions presented in this manuscript, especially the conclusion that the determination of pathogenicity of all c-terminal truncating variants can be determined according to their proposed n+2 rule, or that all of the 10% of people with Rett syndrome and c-terminal truncating variants could be treated by using a base editor to correct the -PPX termination codon.

It is important to state here that we did not use the data in ClinVar for our analysis, but the original information that was held in RettBASE. We have clarified this in the manuscript and have now included supplementary tables containing the data we downloaded from both RettBASE and gnomAD. Table 1 contains all the RettBASE entries with *MECP2* mutations, while Tables 2-4 contain subsets of this data pertaining to CTDs. We have extended our "high confidence" set of RTT alleles to contain one more that was previously overlooked (c.1152_1195) to bring the total number of alleles to 15. Taken together these alleles account for 158 individual entries in RettBASE. We have now included an analysis of all frameshifting mutations in the CT-DPR (Supplementary Table 3, Fig. 3 SF2) which covers 69 different mutations and 260 individual entries. Of these, +2 frameshifts make up the large majority, in contrast to the gnomAD data shown in Supplementary Table 8 and Fig. 3 SF1.

(4) The HEK-based system utilized is convenient for doing the initial experiments testing ABE; however, it represents an artificial system expressing cDNA without splicing. Canonical NMD is dependent on splicing, and while non-canonical "NMD-like" processes are less well understood, a concern is whether the artificial system used can adequately predict efficacy in a native setting that includes introns and splicing.

We disagree with this opinion. We show that the loss of protein and mRNA seen with knock in mouse and human alleles is recapitulated when using a cDNA-based transgene in the HEK system, demonstrating that the mechanism of loss does not involve factors bound at splice junctions etc. We also demonstrate the effect of the A to G change at the stop codon is the same whether we do this by base editing our cDNA transgene in T-REx cells or by making the CTD1 X>W knock-in mouse. Both result in increased levels of a slightly extended but still truncated protein.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

The phrase in the title, "an alternative therapeutic approach" is only insinuated in the manuscript, making it rather inappropriate to be in the title.

In this study we use adenine base editing to modify RTT-causing CTD mutations in T-REx cells which is clearly a precursor to developing a therapy, utilising the new findings in this study. We therefore feel that the use of "an alternative therapeutic approach" can be justified.

Reviewer #2 (Recommendations for the authors):

I have a few minor comments for the authors to consider:

(1) Please double-check Figure 2 Supplementary Figure 1, as the allele count for E394K does not appear to be in the thousands; rather, E397K seems to be the variant shown in the graph.

Yes, this was an error and has been corrected to E397K in the text. Thank you for spotting it.

(2) On page 12, the phrase "common DNA sequence features shared by all CTDs that give rise to RTT" might be better described as "amino acid sequence features."

This has been altered in the text as suggested.

(3) On page 3, the sentence "analysis of patient mutations and experimental data from mouse models support a role in transcriptional repression" cites Gabel et al. 2015 and Kinde et al. 2016, which focus on null alleles but not patient mutations. It would be appropriate to also cite Johnson et al. 2017, which analyzed MeCP2 T158M and R106W patient mutations.

This has now been cited.

(4) In the same sentence, Bajikar et al. 2025 are described as studying the "acute loss of MeCP2," but gene expression was analyzed after a week or longer period of time, not minutes to hours as in degron-mediated degradation systems.

The word "acute" has been removed from the text.

(5) On page 7, the statement that "E394K is common... who were later found to have additional pathogenic MECP2 mutations" should include a supporting reference.

We have now included the references Moncla et al (2002) and Wan et al (1999) to address this.

(6) Similarly, on page 10, the sentence "these findings question the validity of two cases where individuals presented with classical Rett..." is missing a reference to the case report mentioned.

The references Bienvenu et al (2000) and Philippe et al (2006) have been added.

(7) While the manuscript is well written and full of detail, if space is an issue, the authors might consider tightening sections that reiterate findings from their 2018 HMG paper.

A section discussing the CTD2 allele from the 2018 HMG paper has been removed from the results section.

Reviewer #3 (Recommendations for the authors):

(1) Overall, the manuscript is rather dense and potentially challenging to follow easily, especially for a non-expert reader.

Minor edits have been made to the text which will hopefully make it easier to follow. We have also added two new figures (Figures 7 and 8) to summarize the different alleles and edits which appear in the paper, and to show how to determine whether a CTD in the region is likely to be benign or pathogenic.

(2) The introduction of data presented in Figure 1 within the manuscript introduction seems inappropriate and should be moved to the results section.

We would say that this is unconventional rather than inappropriate, and is referred to in the introduction, so we would prefer to leave it as it is.

(3) Providing specific, common nomenclature for genetic repository variants (rs numbers, gnomAD IDs, etc) somewhere would be beneficial. This is an issue because of the complexity of numbering (either coding or protein) for MECP2 due to the different transcript-based numbering systems.

This nomenclature is now included in the supplementary tables of data from RettBASE and gnomAD.

(4) As described in the public comments, there are a number of MECP2 genetic variants listed in the Clinvar copy of RettBase, resulting in c-terminal truncations that are not mentioned or discussed within the manuscript. Without the level of detail present in the original form of RettBase (number of events, de novo, etc) in the currently available Clinvar iteration, it is unclear why a number of variants, even within the limited DPR region, were not mentioned. A supplementary file including the more complete information from the RettBase version, with a complete listing of all c-terminal truncating variants, and an explicit rationale for the exclusion of variants would be helpful.

We have now included supplementary tables with our download of all MECP2 mutations which were held in RettBASE. We have further added tables with the subset of mutations that we have analysed and have more explicitly stated our criteria for defining the "high confidence" sets of mutations. We did not download the data relating to "evidence of pathogenicity" (ie de novo?, absent from parents etc) from RettBASE, but annotated our list of CTDs with this information while RettBASE was still available. This was used in Supplementary Table 3.

(5) A discussion of the limitations, notably that the fact that the focus exclusively on deletion variants within a restricted region (c.1110-1210) does not truly represent all genetic variants that cumulatively account for 10% of Rett cases, is needed. Furthermore,

as pointed out, not all frameshift variants, even those that are $n+2$, result in the -PPX termination that is presented as the pathogenic basis of c-terminal truncations and amenable to correction by ABE. This should be noted in the discussion, as well as consideration of the late nonsense variants that cause c-terminal truncations (some of which would be very similar to the deletion variants discussed but without the proposed primary pathogenic driver, -PPX).

We do not claim to explain the pathogenic mechanism of all C-terminal frameshift mutations found in cases of Rett syndrome. There will certainly be some that do not fit our explanation. However, we believe we have shown evidence that a large proportion of CTDs in RTT will be amenable to the therapy we propose.

(6) Regarding point 3 in the public review, specifically:

(a) $n=7$ nonsense variants (S360X, K363X, E397X, R453X, E455X) that do not carry the destabilizing -PPX sequence.

(b) $n=136$ frameshift indel variants beyond the NID.

(i) $n=11$ that have indels that extend past the native stop codon, $n=4$ of which start within the DPR domain (c.1110-1210) but would have a different terminal sequence than their proposed pathogenic -PPX sequence.

(ii) $n=125$ frameshift indels with terminal breakpoint before the native stop codon

(c) $n=89$ that have start or stop points within c.1110-1210

(d) $n=72$ not mentioned within the manuscript.

(e) $n=27$ are $n+1$, with 26/27 having what the authors term as the "tolerated" -SPRTX ending, but 1/27 having a frameshift beyond this region (c.1133_1361)

(f) $n=45$ are $n+2$, with 32/45 ending in -PPX (supporting authors conclusion), but 10/45 will use the frameshift stop codon preceding the -PPX and have a different terminal sequence, and 3/45 result in a frameshift termination beyond the -PPX sequence.

(g) $n=36$ have breakpoints either before c.1110 or after c.1210

(h) $n=16$ start before c.1110, with 11/16 ending before c.1110. 4 of these 11 are $n+2$, but would use the earlier frameshift stop codon and not have -PPX terminal sequence. For 5/16, the indel extends past c.1210, with the $n+2$ leading to frameshift termination codons beyond the -PPX sequence.

(i) $n=20$ indels start beyond c.1210, with 8/20 being $n+1$ and 12/20 being $n+2$, with neither leading to the -SPRTX or --PPX termination sequences characterized in this manuscript.

As mentioned previously, we have used data taken from RettBASE, not from ClinVar. Both RettBASE and ClinVar will contain *MECP2* mutations found in cases of RTT which are not the causative mutation. Databases of this kind contain sequencing errors and mutations that have been mistakenly assigned as causative. It is therefore imperative that the publicly available information is screened to only include mutations that meet stringent criteria. This is why we chose to start by looking at high confidence sets of mutations, with our conclusions supported by analysis of all such mutations in our region of study.

As mentioned in response to point 5 above, we do not claim to explain every mutation in the region, but believe this study reveals an important disease mechanism for a large proportion of CTDs, leading to a potential therapy. It also contains significant information for predicting the likely prognosis of individuals with CTDs, who may remain healthy but are currently

informed that their mutation is likely pathogenic. At present it seems that this is often based solely on the presence of a frameshifting mutation with similarities to *bona fide* RTT CTDs, without strong evidence of pathogenicity.

<https://doi.org/10.7554/eLife.109170.2.sa0>