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

Jacky Guy[1,*], Elena Hein[1], Bea Alexander-Howden[1], Timur von Bock und Polach[1], Tricia Mathieson[1], Benjamin P. Kleinstiver[2,3], Huda Zoghbi[4,5,6] and Adrian Bird[1,*]

[1]University of Edinburgh, Institute of Cell Biology, Michael Swann Building, Max Born Crescent, Edinburgh, EH9 3BF, UK.

[2]Center for Genomic Medicine and Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA.

[3]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA.

[4]Department of Molecular and Human Genetics and Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA.

[5]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX 77030, USA

[6]Howard Hughes Medical Institute, Houston, TX 77030, USA

*Corresponding authors: J.Guy@ed.ac.uk, A.Bird@ed.ac.uk

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). 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; Hagberg, 2002). MeCP2 protein is found in the nucleus of all cell types but is most highly expressed in neurons (Lewis et al., 1992; Ross et al., 2016; Skene et al., 2010) where it has been implicated in a range of functions, including splicing, microRNA processing, transcriptional activation and repression (Lyst and Bird, 2015). Analysis of patient mutations and experimental data from mouse models support a role in transcriptional repression (Gabel et al., 2015; Kinde et al., 2016) although its acute loss leads to both increases and decreases in gene expression (Bajikar et al., 2025). 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) , and an NCoR interacting domain (NID) (Lyst et al., 2013; Nan et al., 1998). Through the NID, MeCP2 binds to TBL1X/TBL1XR1, subunits of the nuclear corepressor complex NCoR (Kruusvee et al., 2017; Lyst et al., 2013), 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) , 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). 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) (Figure 1A).

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; Cuddapah et al., 2014; Smeets et al., 2005). CTDs occur in a 'deletion-prone region' (CT-DPR) between approximately c.1110 and c.1210 (numbering according to transcript ENST00000303391.11, 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). More than 50 different CTDs can be found in the database of RTT mutations, RettBASE (Christodoulou et al., 2003; Krishnaraj et al., 2017), ranging in frequency from single occurrences to around 50 individuals with the most common deletions. 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; Cuddapah et al., 2014). Our previous CTD knock-in mouse models (Figure 1B, C) 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).

Intriguingly, the GnomAD database (Chen et al., 2024; Gudmundsson et al., 2022), 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 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 have been collated in RettBASE (Christodoulou et al., 2003; Krishnaraj et al., 2017). Although this database is no longer publicly available, abridged data has now been incorporated into ClinVAR (Landrum et al., 2016). 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 neutral variant.

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). Several of these variants are present in hundreds or thousands of individuals, with E394K represented over 5000 times. E394K 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. Also of note, many apparently healthy individuals display in-frame deletions that remove segments of the C-terminal domain (Figure 2A, Fig. 2, SF1B). 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). This difference is also apparent in the Alpha missense 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. A comparison of these benign frameshifts with pathogenic RTT CTD mutations does not reveal any obvious difference in the location of the 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. To ensure that pathogenic RTT and non-pathogenic GnomAD mutations were reliably annotated, only RTT mutations associated with classical Rett syndrome and absent in both parents (*de novo*) were included. Furthermore, GnomAD variants were only

considered when present in hemizygotes, to exclude the possibility of a pathogenic mutation being masked by skewing of X-inactivation in some heterozygous females.

We found that all 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 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, including heterozygotes, in the most recent GnomAD release, v.4.1.0, which contains sequence data from over 800,000 individuals, with a total of 103 *MECP2* CTD alleles carrying 27 different mutations. In line with the expectation that the pathogenic frameshift would be severely under-represented in the general population, we observed 96 individual alleles that were shifted to the “benign” +1 frame, whereas only 7 used 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 included this sequence (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.

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 a 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 were 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. This mutation will result in a switch to the +1 reading frame leading to an -SPRTX ending, thus a benign deletion which does not lead to a reduction in MeCP2 protein levels (Figure 3D).

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

Analysis of human *MECP2* DNA sequences implicated codons adjacent to the stop codon in pathogenicity. Specifically, the ending -SPRTX is associated with almost all benign human CTDs that involve frameshifts. This ending is also found as an exception in one of the stringent set of RTT CTDs, c.1158_1167del (CTD3) (Bienvenu et al., 2000). As the

consequences of this MeCP2 terminus had not been analysed in previous mouse models, we created a CTD3 knock-in mouse allele (Fig. 3 SF2A) to determine whether the +1 -SPRTX ending does result in WT levels of MeCP2 protein, and to test whether this mutation may have been classified as pathogenic in error. The resulting male hemizygous CTD3 mice appeared like WT littermates, did not develop RTT-like phenotypes (Fig. 3, SF2B, C) and had 100% survival at one year. CTD3 MeCP2 protein and RNA was expressed at WT levels in hemizygous mouse brain (Fig. 3 SF2D-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. Therefore, our expectation that the +1 frame ending -SPRTX should not lead to reduced MeCP2 protein or mRNA levels is supported by this evidence.

These findings question the validity of two cases where individuals presented with classical Rett syndrome despite a shift to the proposed benign +1 reading frame, giving a C-terminal sequence of codons highly similar to neutral variants found in GnomAD (see Figure 3B). Both cases presented a *de novo* c.1158_1167del mutation. 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_1210del (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 SF2F), pointing to a post-transcriptional mechanism for MeCP2 loss. Our CTD2 mouse model unexpectedly had a wildtype phenotype, which was due to the presence of normal levels of truncated MeCP2 protein in brain (Guy et al., 2018). Due to differences in sequence between mouse and human *MECP2*, the mouse allele ended in the sequence -SPX, whereas the human CTD2 protein ends -PPX. Humanisation of the mouse CTD2 terminal sequence (CTDhu) in ESC-derived neurons restored the -PPX ending and led to MeCP2 protein and mRNA depletion (see Fig. 2, SF2). This further supports the idea that the -PPX ending (at least in the context of *MECP2*) is sufficient for protein depletion.

There is considerable evidence that consecutive prolines within a messenger RNA are associated with increased stalling at 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 dependent on the amino acid sequence immediately before the stop codon. 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, SF2F.

A genome editing strategy for the CTD class of RTT mutations

Despite their highly variable start and end points, our findings point to common DNA 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. 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) 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 SF1A-C). The finding that the effect of the CTD mutations on protein and mRNA levels is well reproduced using expression constructs with heterologous promoter, 3' UTR and polyadenylation signal 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 gRNAs position the target A7 at the edge of the optimal editing window for ABEmax (Gaudelli et al., 2017), 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) (Figure 6B). These ABEs were able to edit the moCTD1 transgene with high efficiency using all gRNAs tested (Figure 6C), 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, 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 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, Fig. 6 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 SF1 D, E).

As a final proof-of-concept experiment, we asked whether these reagents also edited the human mutation. As there is only one nucleotide difference between mouse and human sequence in the guide region, we tested human-specific gRNAs, guides 1 and 2, on a huCTD1 T-REx line (Fig. 6 SF2A). The human sequence-targeted gRNAs performed in a very similar way to their mouse equivalents (Fig. 6 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 nucleotide 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. 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.

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; Li et al., 2020). 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), 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 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 are 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, C-terminal proline-proline showed the strongest effect, and while statistically there would be expected to be seven *E. coli* proteins ending in PPX, there are 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. On a cautionary note, despite clinical examples arguing that these are phenotypically neutral mutations, it will be wise to maintain vigilance. While it is clear that benign CTDs will not give rise to RTT, the possibility that milder 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 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

Data was taken from GnomAD release v4.1.0 (downloaded from https://gnomad.broadinstitute.org/news/2024-04-gnomad-v4-1/ on 21/11/2024). *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.

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. The data held in RettBASE has now been incorporated into ClinVar in an abbreviated form (https://www.ncbi.nlm.nih.gov/clinvar/submitters/504290/). High confidence 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 parent).

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.

Alpha Missense data

Mean pathogenicity scores at each amino acid position of human MeCP2 (e2 isoform) were obtained from Alpha Missense (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, derived from 129/Ola E14Tg2a cells (A gift from Joe Mee, University of Edinburgh). 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 x 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
GGCGCCCGTGCCACTGCTCCCACCCCTCCACCTGAGCCCGCGGGCTCCGAGGACCCCACCAGCCCCC
CTGAGCCCCAGGACTTGAGCAGCAGCATCTGCAAAGAAGAGAAGATGCCCCGAGG

CTD1 X>W:

AGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCCCAA

AGGCCCCCGTGCCACTGCTCCCACCCCATCAGCCCCCCTGGGCCTCAGGACTTGAGCAGCAGCATCT

GCAAAGAAGAGAAGATGCCCCGAGGAGGCTCACTGGAAAGCGATGGCTGCCCCAAG

CTD2 X>W:

AGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCCCAA

AGGCCCCCGTGCCACTGCTCCCACCCCTGCCCCCCTGGGCCTCAGGACTTGAGCAGCAGCATCTGCA

AAGAAGAGAAGATGCCCCGAGGAGGCTCACTGGAAAGCGATGGCTGCCCCAAGGAG

Clones were screened for hemizygous knock-in alleles by PCR on genomic DNA (forward primer: AGCATCAGAAGGTGTTCAGG, 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 $1.5x10^5$ cells/cm^2 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

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 C57BL6/J 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.

Mouse phenotyping

Cohorts of hemizygous mutant male mice and WT littermates (N=3 backcross to C57BL/6J) were weighed and scored weekly 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. 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, 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 Fluoroscein Master Mix (Bioline) in a LightCycler 480 (Roche). Primers for qPCR were as follows:

Mouse *Mecp2* mRNA: Exon3 F - ACCTTGCCTGAAGGTTGGAC, Exon4 R - GCAATCAATTCTACTTTAGAGCGAAAA

Mouse *Mecp2* primary transcript: Intron1 F - ACATGGCCGACAGAGTGC, Intron1 R - GCACCTGAGGAAGCAAACC

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

Cyclophilin A mRNA (control): CypA F - TCGAGCTCTGAGCACTGGAG, CypA R - CATTATGGCGTGTAAAGTCACCA

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). Extracts were run on TGX 4-20% gradient gels (BioRad) loading extract equivalent to approximately 5 x 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; Ross et al., 2016) using the following antibodies: anti-MeCP2 (N-terminus): mouse monoclonal Men-8 (Sigma), anti-NeuN: rabbit polyclonal ABN78 (Millipore), anti-histone H3: rabbit polyclonal ab1791 (Abcam), anti mSin3a: rabbit polyclonal ab3479 (Abcam). Western blots were developed with IR-dye secondary antibodies (IRDye 800CW donkey anti-mouse, IRDye 680LT donkey anti-rabbit, LI-COR Biosciences) and scanned using a LI-COR Odyssey machine. Images were quantified using Image Studio Lite software (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 were obtained from Invitrogen and maintained according to the manufacturer's instructions. Tetracycline-inducible MeCP2 expression constructs were made by cloning human and mouse e1 cDNAs into plasmid pcDNA5/FRT/TO (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 and pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP (Walton et al., 2020) 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 x 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 (Clement et al., 2019).

Statistical analysis

All statistical analysis was performed using GraphPad Prism (GraphPad Software). Datasets were compared using Student's t-tests (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).

Data availability

Data from this work are available from the authors.

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 and Matthew Lyst, Jim Selfridge and Dirk-Jan Kleinjan for critical reading of the

manuscript. This work was supported by Wellcome (Investigator Award 222507/Z/21/Z to AB), the Rett Syndrome Research Trust (Grant no. 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.

Competing Interests Statement

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

Alves CRR, Ha LL, Yaworski R, Sutton ER, Lazzarotto CR, Christie KA, Reilly A, Beauvais A, Doll RM, de la Cruz D, Maguire CA, Swoboda KJ, Tsai SQ, Kothary R, Kleinstiver BP. 2024. Optimization of base editors for the functional correction of SMN2 as a treatment for spinal muscular atrophy. *Nat Biomed Eng* **8**:118–131. doi:10.1038/s41551-023-01132-z

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. *Nature genetics* **23**:185–8. doi:10.1038/13810

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149–157. doi:10.1038/s41586-019-1711-4

Augenstein K, Lane JB, Horton A, Schanen C, Percy AK. 2009. Variable phenotypic expression of a MECP2 mutation in a family. *J Neurodev Disord* **1**:313. doi:10.1007/s11689-009-9034-7

Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI. 1995. Embryonic stem cells express neuronal properties in vitro. *Developmental biology* **168**:342–57. doi:10.1006/dbio.1995.1085

Bajikar SS, Zhou J, O'Hara R, Tirumala HP, Durham MA, Trostle AJ, Dias M, Shao Y, Chen H, Wang W, Yalamanchili HK, Wan Y-W, Banaszynski LA, Liu Z, Zoghbi HY. 2025. Acute MeCP2 loss in adult mice reveals transcriptional and chromatin changes that precede neurological dysfunction and inform pathogenesis. *Neuron* **113**:380-395.e8. doi:10.1016/j.neuron.2024.11.006

Bebbington A, Percy A, Christodoulou J, Ravine D, Ho G, Jacoby P, Anderson A, Pineda M, Ben Zeev B, Bahi-Buisson N, Smeets E, Leonard H. 2010. Updating the profile of C-terminal MECP2 deletions in Rett syndrome. *Journal of Medical Genetics* **47**:242–248. doi:10.1136/jmg.2009.072553

Bienvenu T, Carrié A, de Roux N, Vinet M-C, Jonveaux P, Couvert P, Villard L, Arzimanoglou A, Beldjord C, Fontes M, Tardieu M, Chelly J. 2000. MECP2 mutations account for most cases of typical forms of Rett syndrome. *Human Molecular Genetics* **9**:1377–1384. doi:10.1093/hmg/9.9.1377

Chen PJ, Liu DR. 2023. Prime editing for precise and highly versatile genome manipulation. *Nat Rev Genet* **24**:161–177. doi:10.1038/s41576-022-00541-1

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, Poterba T, Wilson MW, Tarasova Y, Phu W, Grant R, Yohannes MT, Koenig Z, Farjoun Y, Banks E, Donnelly S, Gabriel S, Gupta N, Ferriera S, Tolonen C, Novod S, Bergelson L, Roazen D, Ruano-Rubio V, Covarrubias M, Llanwarne C, Petrillo N, Wade G, Jeandet T, Munshi R, Tibbetts K, O'Donnell-Luria A, Solomonson M, Seed C, Martin AR, Talkowski ME, Rehm HL, Daly MJ, Tiao G, Neale BM, MacArthur DG, Karczewski KJ. 2024. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**:92–100. doi:10.1038/s41586-023-06045-0

Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, Schneider RG, Senior AW, Jumper J, Hassabis D, Kohli P, Avsec Ž. 2023.

Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**:eadg7492. doi:10.1126/science.adg7492

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. doi:10.1016/j.devcel.2014.01.032

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. *Hum Mol Genet* **21**:3806–3814. doi:10.1093/hmg/dds208

Christodoulou J, Grimm A, Maher T, Bennetts B. 2003. RettBASE: The IRSA MECP2 variation database-a new mutation database in evolution. *Human mutation* **21**:466–72. doi:10.1002/humu.10194

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L. 2019. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**:224–226. doi:10.1038/s41587-019-0032-3

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. doi:10.1093/hmg/ddh282

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, NY* **339**:819–23. doi:10.1126/science.1231143

Cuddapah VA, Pillai RB, Shekar KV, Lane JB, Motil KJ, Skinner SA, Tarquinio DC, Glaze DG, McGwin G, Kaufmann WE, Percy AK, Neul JL, Olsen ML. 2014. Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *Journal of medical genetics* **51**:152–8. doi:10.1136/jmedgenet-2013-102113

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. doi:10.1016/j.stem.2013.03.006

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. doi:10.1038/nature14319

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. *Mol Ther Methods Clin Dev* **5**:180–190. doi:10.1016/j.omtm.2017.04.007

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. doi:10.1038/nature24644

Gerbracht JV, Boehm V, Gehring NH. 2017. Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. *Sci Rep* **7**:10616. doi:10.1038/s41598-017-10847-4

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. *Hum Mutat* **43**:1012–1030. doi:10.1002/humu.24309

Gutierrez E, Shin B-S, Woolstenhulme CJ, Kim J-R, Saini P, Buskirk AR, Dever TE. 2013. eIF5A promotes translation of polyproline motifs. *Mol Cell* **51**:35–45. doi:10.1016/j.molcel.2013.04.021

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. *Hum Mol Genet* **27**:2531–2545. doi:10.1093/hmg/ddy159

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. doi:10.1126/science.1138389

Hagberg B. 2002. Clinical manifestations and stages of Rett syndrome. *Ment Retard Dev Disabil Res Rev* **8**:61–65. doi:10.1002/mrdd.10020

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. *Annals of neurology* **14**:471–9.

Hayes CS, Bose B, Sauer RT. 2002. Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *J Biol Chem* **277**:33825–33832. doi:10.1074/jbc.M205405200

Huter P, Arenz S, Bock LV, Graf M, Frister JO, Heuer A, Peil L, Starosta AL, Wohlgemuth I, Peske F, Nováček J, Berninghausen O, Grubmüller H, Tenson T, Beckmann R, Rodnina MV, Vaiana AC, Wilson DN. 2017. Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Mol Cell* **68**:515-527.e6. doi:10.1016/j.molcel.2017.10.014

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. *Mol Cell Biol* **22**:8562–8570. doi:10.1128/MCB.22.24.8562-8570.2002

Kinde B, Wu DY, Greenberg ME, Gabel HW. 2016. DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proceedings of the National Academy of Sciences of the United States of America* **113**:15114–15119. doi:10.1073/pnas.1618737114

Koerner MV, FitzPatrick L, Selfridge J, Guy J, De Sousa D, Tillotson R, Kerr A, Sun Z, Lazar MA, Lyst MJ, Bird A. 2018. Toxicity of overexpressed MeCP2 is independent of HDAC3 activity. *Genes Dev* **32**:1514–1524. doi:10.1101/gad.320325.118

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. doi:10.1038/nature17946

Krishnaraj R, Ho G, Christodoulou J. 2017. RettBASE: Rett syndrome database update. *Human Mutation* **38**:922–931. doi:10.1002/humu.23263

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. *Proceedings of the National Academy of Sciences of the United States of America* **114**:E3243–E3250. doi:10.1073/pnas.1700731114

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**:D862-868. doi:10.1093/nar/gkv1222

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do

R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**:285–291. doi:10.1038/nature19057

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.

Li CH, Coffey EL, Dall'Agnese A, Hannett NM, Tang X, Henninger JE, Platt JM, Oksuz O, Zamudio AV, Afeyan LK, Schuijers J, Liu XS, Markoulaki S, Lungjangwa T, LeRoy G, Svoboda DS, Wogram E, Lee TI, Jaenisch R, Young RA. 2020. MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* **586**:440–444. doi:10.1038/s41586-020-2574-4

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.

Lyst MJ, Bird A. 2015. Rett syndrome: a complex disorder with simple roots. *Nat Rev Genet* **16**:261–275. doi:10.1038/nrg3897

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME, Bird A. 2013. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci* **16**:898–902. doi:10.1038/nn.3434

Matheisl S, Berninghausen O, Becker T, Beckmann R. 2015. Structure of a human translation termination complex. *Nucleic Acids Res* **43**:8615–8626. doi:10.1093/nar/gkv909

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.

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.

Ramocki MB, Tavyev YJ, Peters SU. 2010. The MECP2 duplication syndrome. *American journal of medical genetics* **152A**:1079–88.

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 J Rare Dis* **6**:58. doi:10.1186/1750-1172-6-58

Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, Doudna JA, Liu DR. 2020. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38**:883–891. doi:10.1038/s41587-020-0453-z

Ross PD, Gadalla KKE, Thomson SR, Selfridge J, Bahey NG, Benito J, Burstein SR, McMinn R, Bolon B, Hector RD, Cobb SR. 2025. Self-regulating gene therapy ameliorates phenotypes and overcomes gene dosage sensitivity in a mouse model of Rett syndrome. *Science Translational Medicine* **17**:eadq3614. doi:10.1126/scitranslmed.adq3614

Ross PD, Guy J, Selfridge J, Kamal B, Bahey N, Tanner KE, Gillingwater TH, Jones RA, Loughrey CM, McCarroll CS, Bailey ME, Bird A, Cobb S. 2016. Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. *Hum Mol Genet* **25**:4389–4404. doi:10.1093/hmg/ddw269

Sadhu C, Lyons C, Oh J, Jagadeeswaran I, Gray SJ, Sinnett SE. 2023. The Efficacy of a Human-Ready miniMECP2 Gene Therapy in a Pre-Clinical Model of Rett Syndrome. *Genes (Basel)* **15**:31. doi:10.3390/genes15010031

Schuller AP, Wu CC-C, Dever TE, Buskirk AR, Green R. 2017. eIF5A Functions Globally in Translation Elongation and Termination. *Mol Cell* **66**:194-205.e5. doi:10.1016/j.molcel.2017.03.003

Sinnett 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. *Mol Ther Methods Clin Dev* **5**:106–115. doi:10.1016/j.omtm.2017.04.006

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. *Molecular cell* **37**:457–68. doi:10.1016/j.molcel.2010.01.030

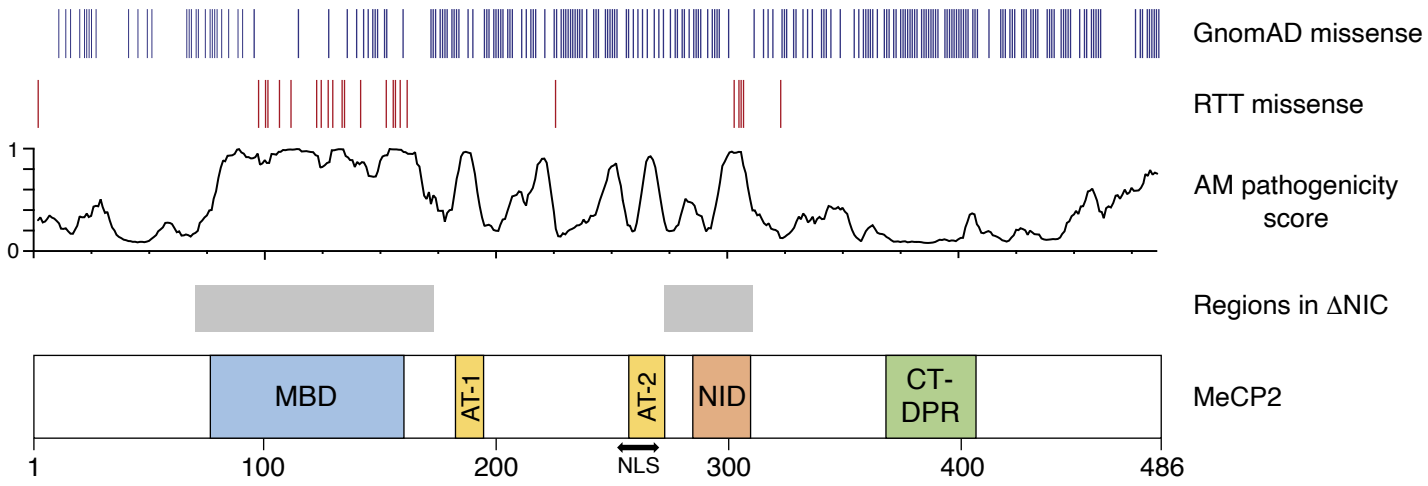
Smeets E, Terhal P, Casaer P, Peters A, Midro A, Schollen E, van Roozendaal K, Moog U, Matthijs G, Herbergs J, Smeets H, Curfs L, Schrander-Stumpel C, Fryns JP. 2005. Rett syndrome in females with CTS hot spot deletions: a disorder profile. *American journal of medical genetics* **132A**:117–20. doi:10.1002/ajmg.a.30410

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. doi:10.1038/nature24058

Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. 2020. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**:290–296. doi:10.1126/science.aba8853

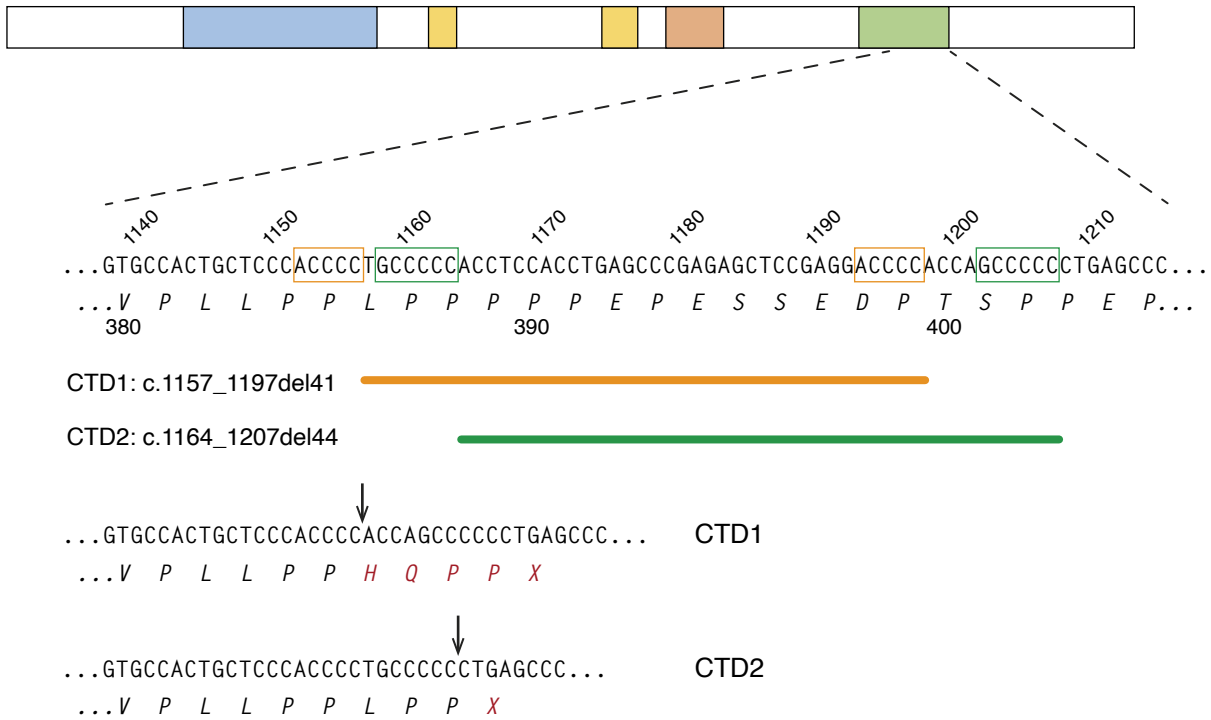
Yusufzai TM, Wolffe AP. 2000. Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Res* **28**:4172–4179. doi:10.1093/nar/28.21.4172

A



B

MECP2 CTD RTT alleles



C

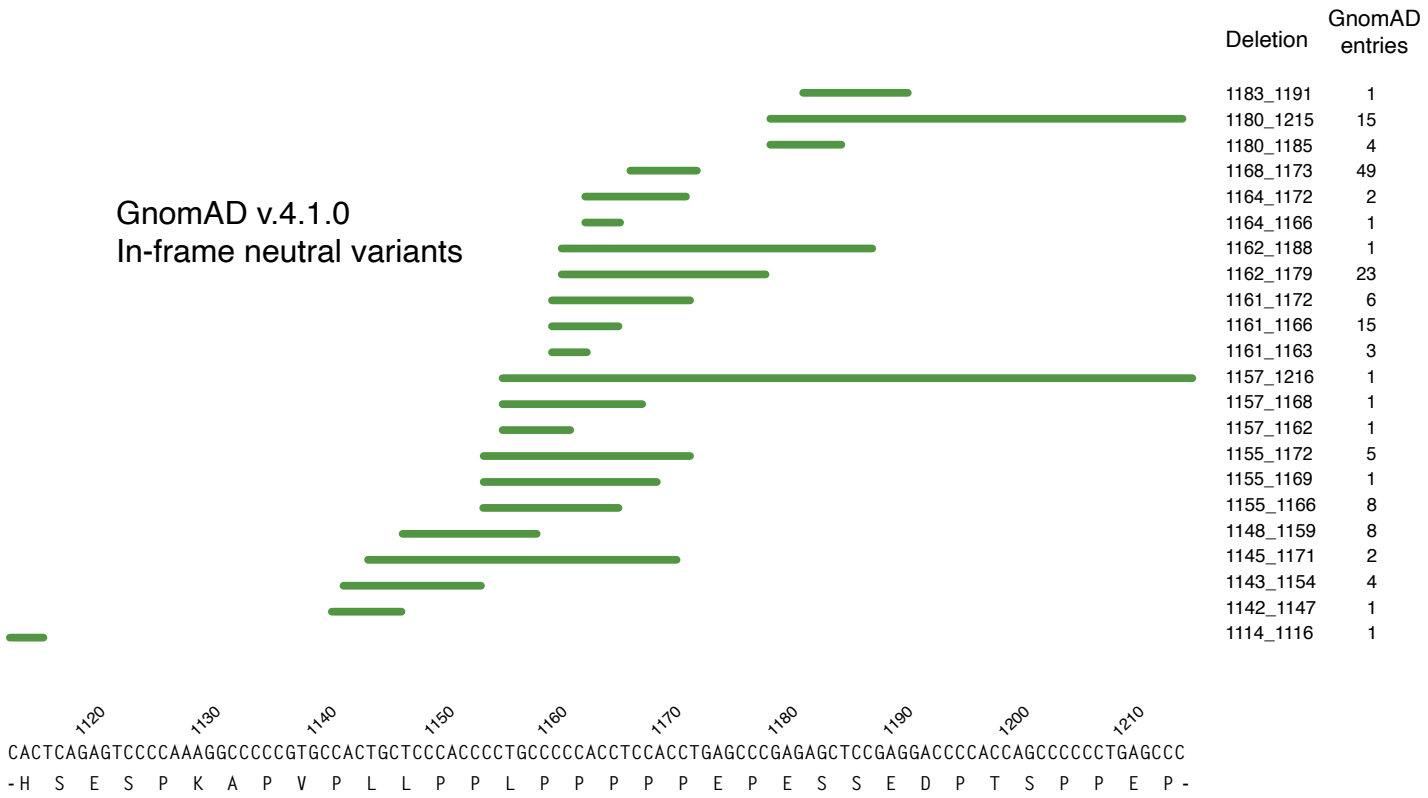
Mecp2 CTD mouse alleles

CTD Allele	Mutation	Protein change	Median survival	MeCP2 protein (brain)	MeCP2 mRNA (brain)
CTD1	c.1157_1197del	p.L386HX5	20 weeks	10% WT	45% WT
CTD2	c.1164_1207del	p.P389X	as WT	as WT	as WT

Figure 1

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" (ΔNIC) 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**) Human 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.

A


Deletion
GnomAD entries
GnomAD v.4.1.0
In-frame neutral variants
1183_1191 1
1180_1215 15
1180_1185 4
1168_1173 49
1164_1172 2
1164_1166 1
1162_1188 1
1162_1179 23
1161_1172 6
1161_1166 15
1161_1163 3
1157_1216 1
1157_1168 1
1157_1162 1
1155_1172 5
1155_1169 1
1155_1166 8
1148_1159 8
1145_1171 2
1143_1154 4
1142_1147 1
1114_1116 1
1120 1130 1140 1150 1160 1170 1180 1190 1200 1210
CACTCAGAGTCCCCAAAGGCCCCCGTGCCACTGCTCCCACCCCTGCCCCCACCTCCACCTGAGCCCGAGAGCTCCGAGGACCCCACCAGCCCCCCTGAGCCC
-H S E S P K A P V P L L P P L P P P P P E P E S S E D P T S P P E P-

B

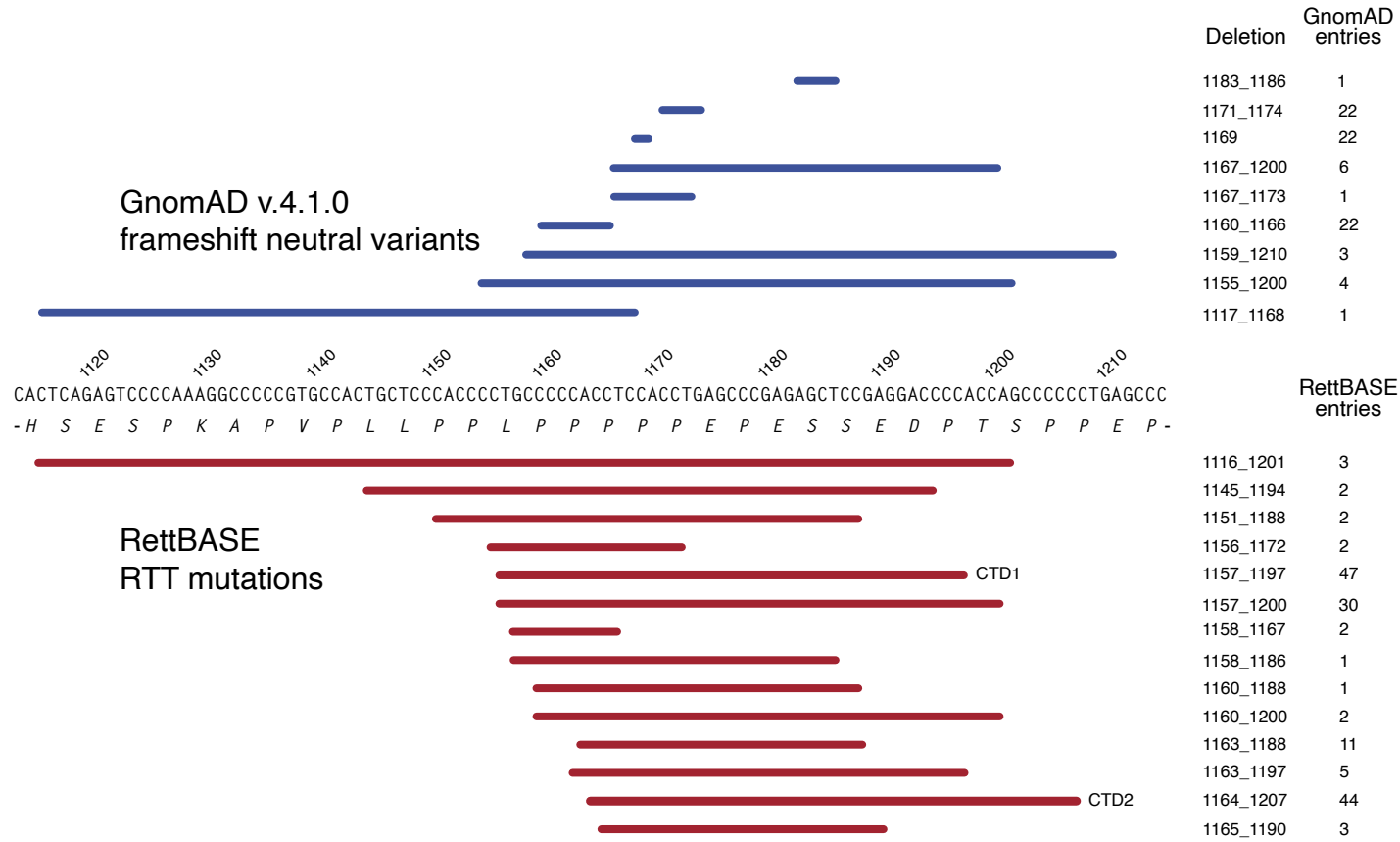
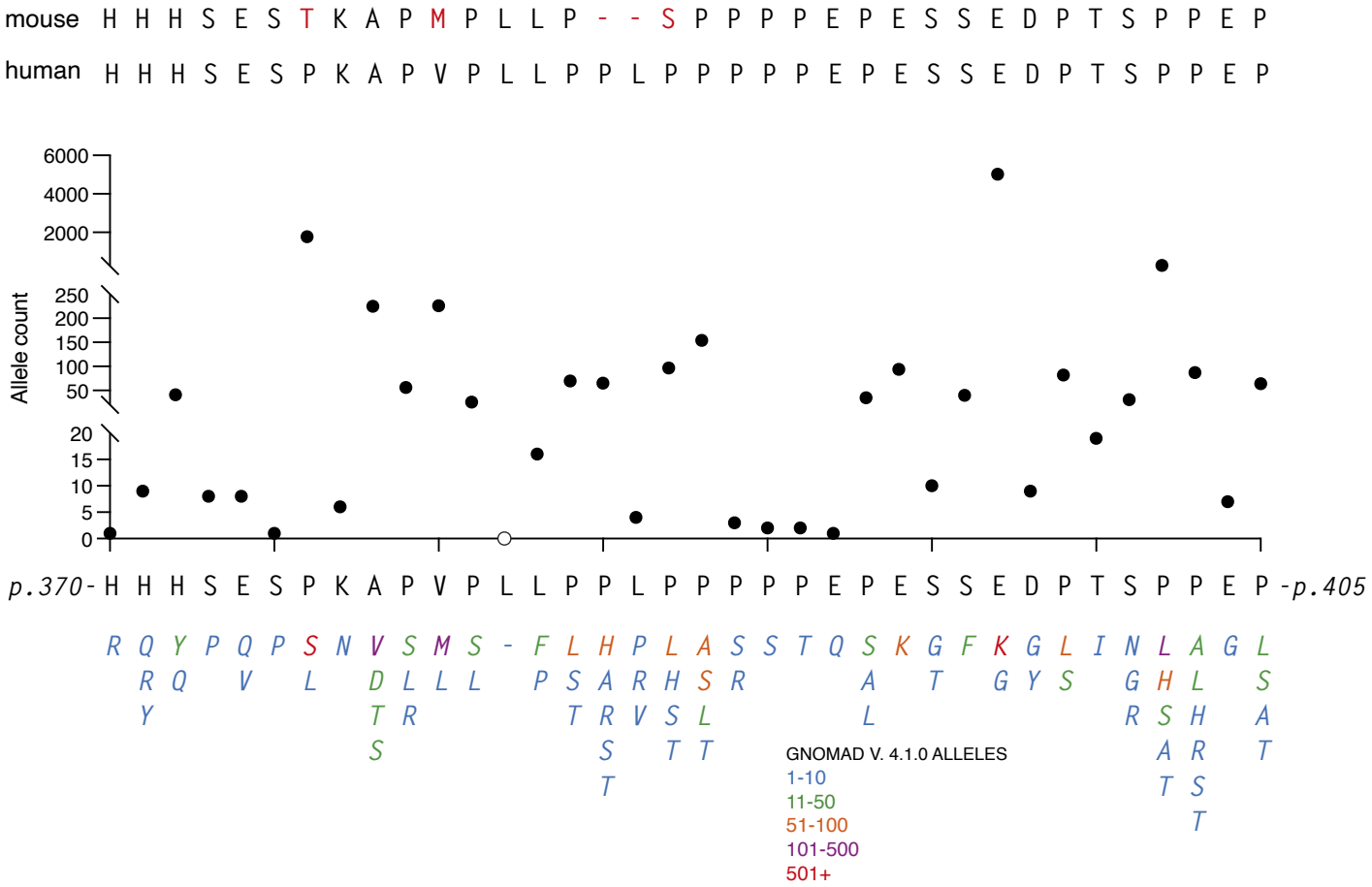

Deletion
GnomAD entries
GnomAD v.4.1.0
frameshift neutral variants
1183_1186 1
1171_1174 22
1169 22
1167_1200 6
1167_1173 1
1160_1166 22
1159_1210 3
1155_1200 4
1117_1168 1
1120 1130 1140 1150 1160 1170 1180 1190 1200 1210
CACTCAGAGTCCCCAAAGGCCCCCGTGCCACTGCTCCCACCCCTGCCCCCACCTCCACCTGAGCCCGAGAGCTCCGAGGACCCCACCAGCCCCCCTGAGCCC
-H S E S P K A P V P L L P P L P P P P P E P E S S E D P T S P P E P-
RettBASE entries
RettBASE
RTT mutations
1116_1201 3
1145_1194 2
1151_1188 2
1156_1172 2
CTD1
1157_1197 47
1157_1200 30
1158_1167 2
1158_1186 1
1160_1188 1
1160_1200 2
1163_1188 11
1163_1197 5
CTD2
1164_1207 44
1165_1190 3

Figure 2

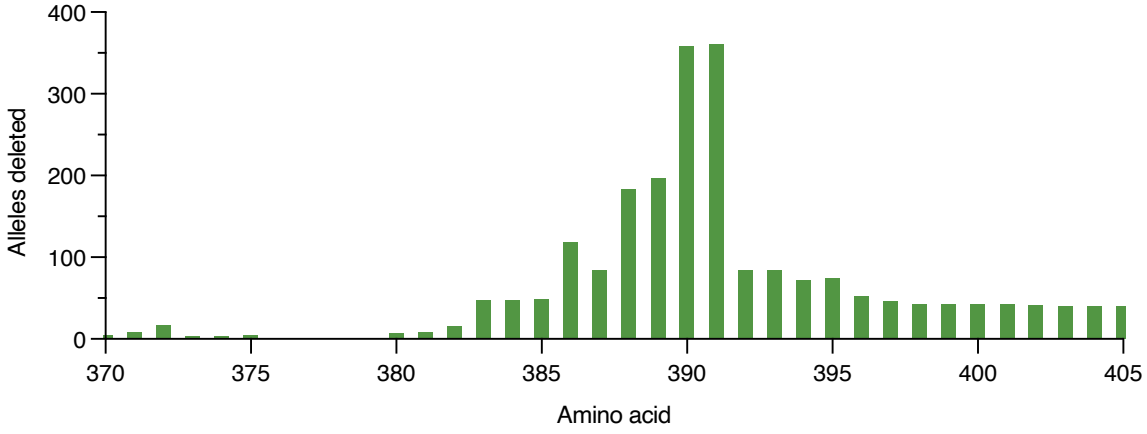
Figure 2. Genomic deletions in the CT-DPR of MECP2. (**A**) All hemizygous in-frame deletions in the region found in GnomAD v4.1.0 are indicated with green lines above the DNA sequence. The deletion co-ordinates (numbered according to ENST00000303391.11) and number of individuals with each deletion are 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.

Figure 2 SF1

A



B



C

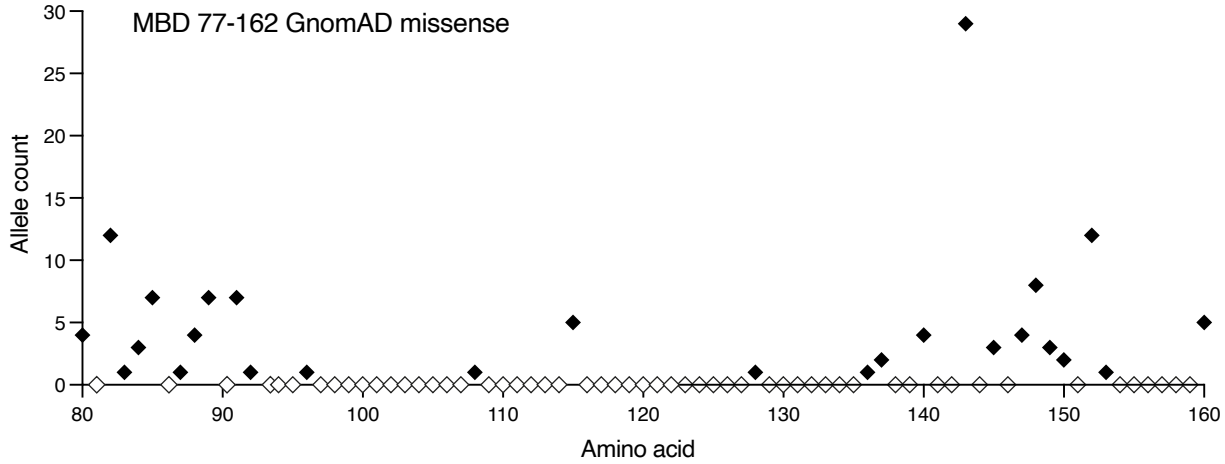


Figure 2, Supplementary Figure 1 GnomAD v4.1.0 data from the CT-DPR (**A**) GnomAD missense mutations. A comparison of mouse and human amino acid sequence in the CT-DPR, followed by a plot of the number of individuals in GnomAD with missense mutations present (filled circles) or absent (open circles) 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).

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

A

Commonly deleted region

1170 1180 1190 1200 1210 1220 1230

CCACCTGAGCCCGAGAGCTCCGAGGACCCCACCAGCCCCCCTGAGCCCCAGGACTTGAGCAGC

-P P E P E S S E D P T S P P E P Q D L S S- 0

-H L S P R A P R T P P A P L S P R T X +1

-T X A R E L R G P H Q P P X +2

B

RettBASE: de novo classical RTT

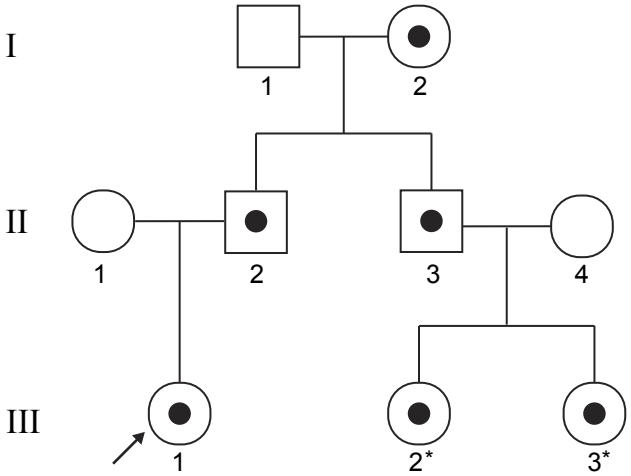
1116_1201 del	-HQPPX
1145_1194 del	-HSESPKAPVPPHQPPX
1151_1188 del	-HSESPKAPVPLLRGPHQPPX
1156_1172 del	-HSESPKAPVPLLPPX
1157_1197 del	-HSESPKAPVPLLPPHQPPX
1157_1200 del	-HSESPKAPVPLLPPQPPX
1158_1167 del	-HSESPKAPVPLLPPLHLSPRAPRTPPAPLSPRTX
1158_1186 del	-HSESPKAPVPLLPPLRGPHQPPX
1160_1188 del	-HSESPKAPVPLLPPLRGPHQPPX
1160_1200 del	-HSESPKAPVPLLPPLQPPX
1163_1188 del	-HSESPKAPVPLLPPLPRGPHQPPX
1163_1197 del	-HSESPKAPVPLLPPLPHQPPX
1164_1207 del	-HSESPKAPVPLLPPLPPX
1165_1190 del	-HSESPKAPVPLLPPLPPGPHQPPX

GnomAD v.4.1.0 (hemizygotes)

1117_1168 del	-HHLSPRAPRTPPAPLSPRTX
1155_1200 del	-HSESPKAPVPLLPPLAPLSPRTX
1159_1210 del*	-HSESPKAPVPLLPPLSPRTX
1160_1166 del	-HSESPKAPVPLLPPLLHLSPRAPRTPPAPLSPRTX
1167_1173 del	-HSESPKAPVPLLPPLPPPSPRAPRTPPAPLSPRTX
1167_1200 del	-HSESPKAPVPLLPPLPPPAPLSPRTX
1169 del	-HSESPKAPVPLLPPLPPPHLSPRAPRTPPAPLSPRTX
1171_1174 del	-HSESPKAPVPLLPPLPPPPSPRAPRTPPAPLSPRTX
1183_1186 del	-HSESPKAPVPLLPPLPPPPPRAPRTPPAPLSPRTX

C

c.1159_1210 del



D

c.1159_1210 del

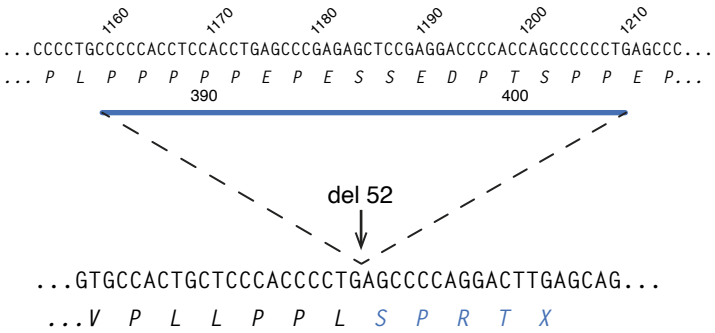


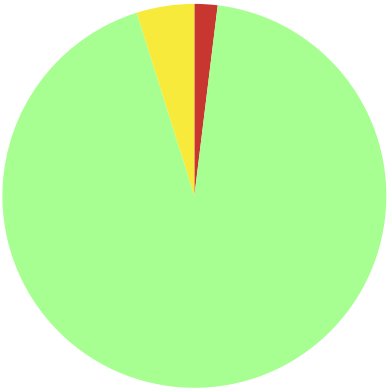
Figure 3

Figure 3. C-terminal amino acid sequences of pathogenic and benign MeCP2 CTDs. (**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 circles). (**D**) Genomic DNA sequence and amino acid sequence showing c.1159_1210 deletion site and molecular consequences.

A

Reading frame	Individuals	Alleles
+1	96	21
+2	7	6

B

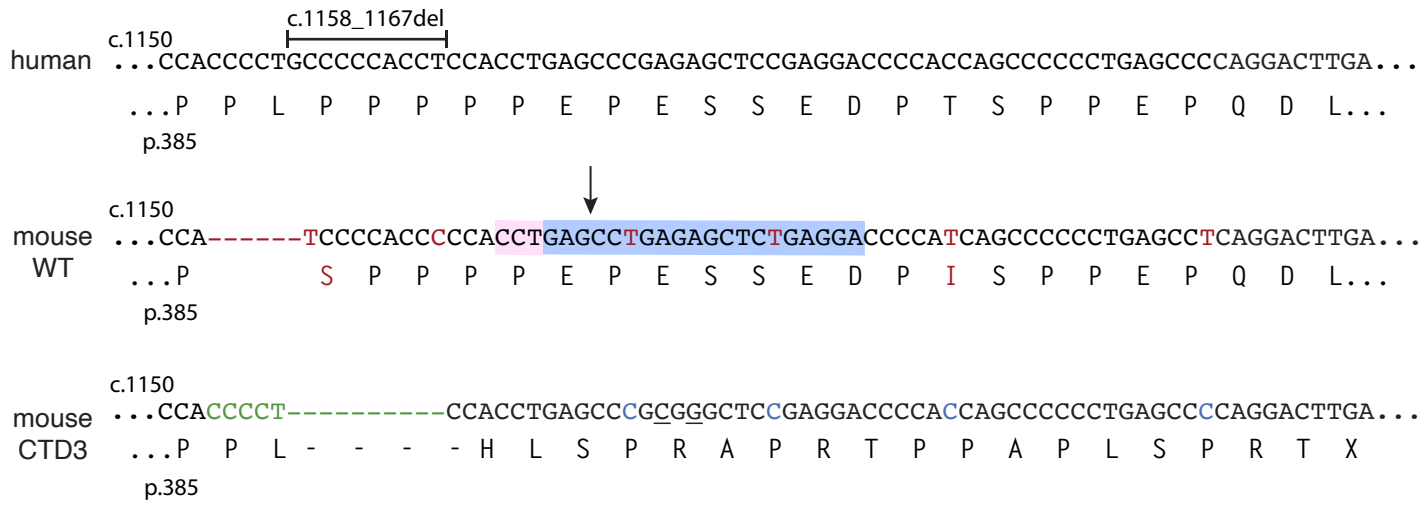


C-terminus	Individuals	Alleles
-PPX	2	2
-SPRT	96	21
other, non-PPX	5	4

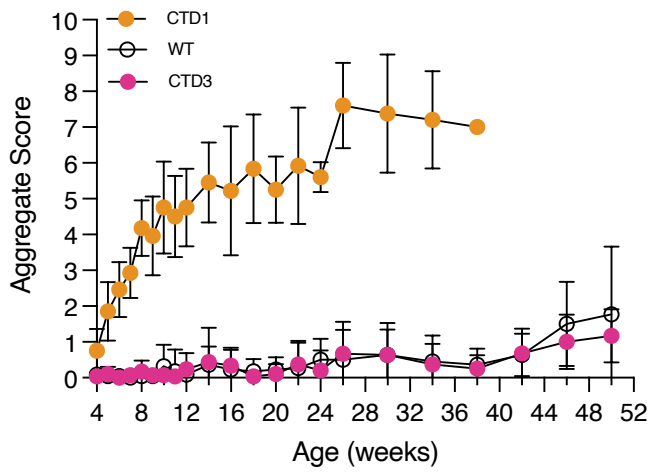
Figure 3, Supplementary Figure 1 GnomAD v4.1.0 frameshifting deletion alleles. (**A**) Number of individuals with each type of frameshift and number of different CTD alleles. (**B**) GnomAD entries shown in (A) classified by C-terminal amino acid sequence for individuals and CTD alleles.

Figure 3 SF2

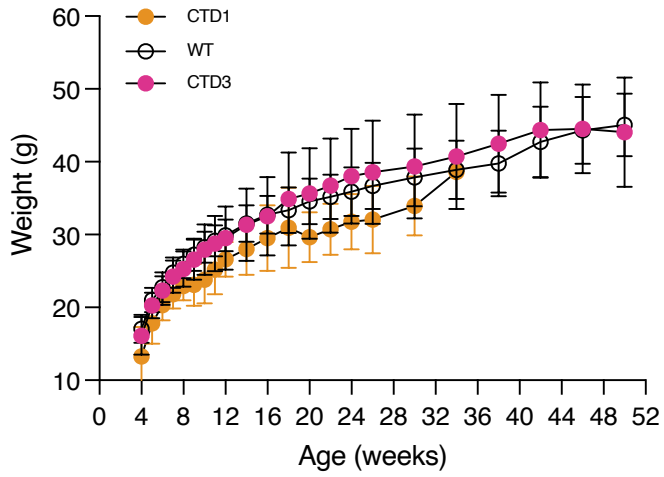
A



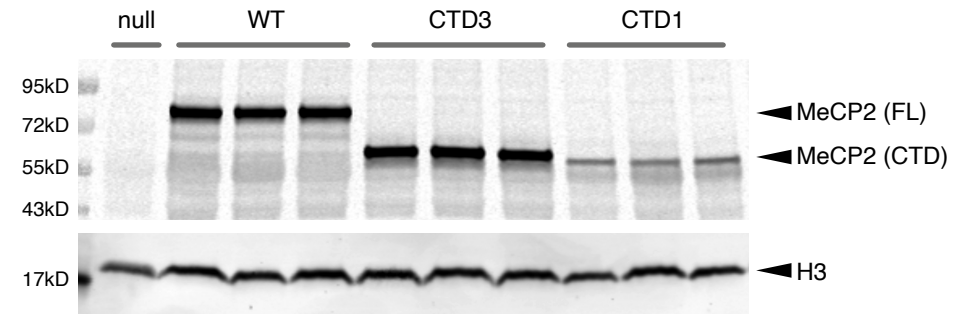
B



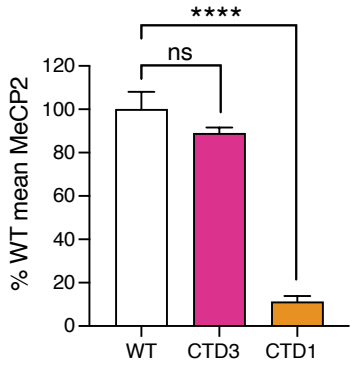
C



D



E



F

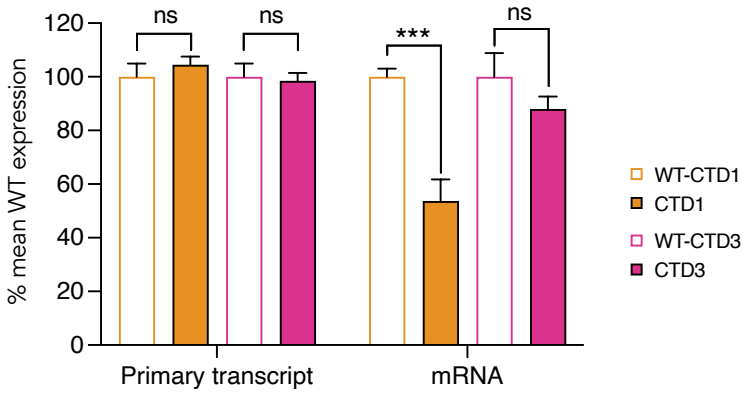


Figure 3, Supplementary Figure 2 CTD3 mouse knock-in allele. (**A**) Human genomic and amino acid sequence showing location of CTD3 deletion (1158_1167del). Mouse WT genomic and amino acid sequences. Differences to human sequence are shown in red. The protospacer sequence (blue) and PAM sequence (pink) 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 +/- sd. (**C**) Body weights of animals shown in (B). Mean +/- 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 +/- 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 +/- 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 (***).

A

Mouse CTD1 ...CCACTGCTCCCACCCCACCAGCCCCCCTGAGCCTCAGGACTTGAGCAGCAGC...
...P L L P P H Q P P X

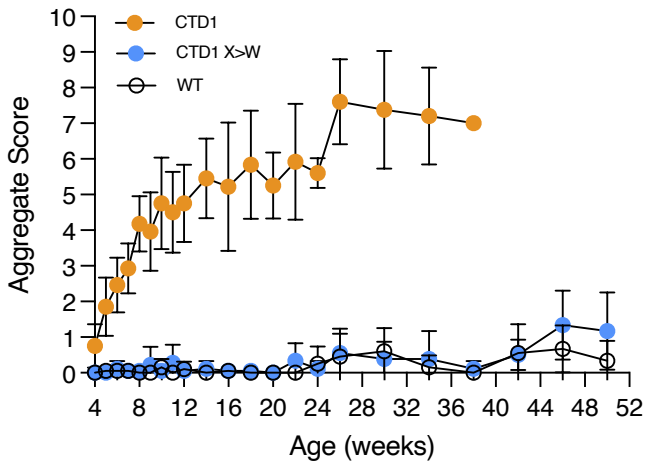
Mouse CTD1 X>W ...CCACTGCTCCCACCCCACCAGCCCCCCTGGGCCTCAGGACTTGAGCAGCAGC...
...P L L P P H Q P P W A S G L E Q Q...

B

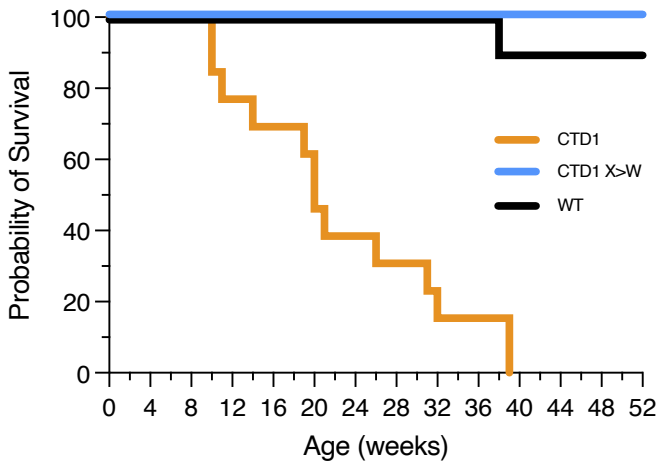
CTD1 ...P L L P P H Q P P X

CTD1 X>W ...P L L P P H Q P P W A S G L E Q Q H L Q R R E D A P R R L T G K R W L P Q G A S X

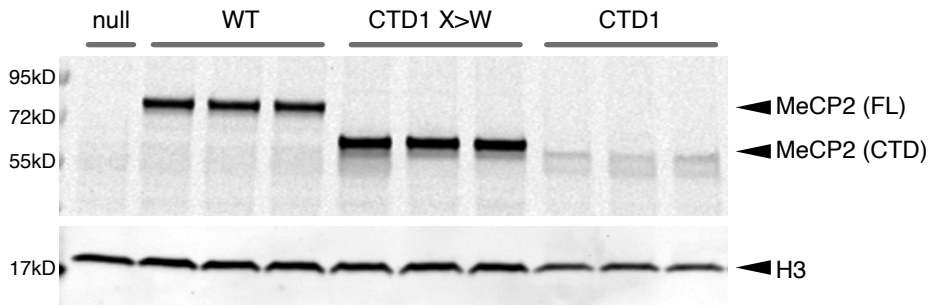
C



D



E



F

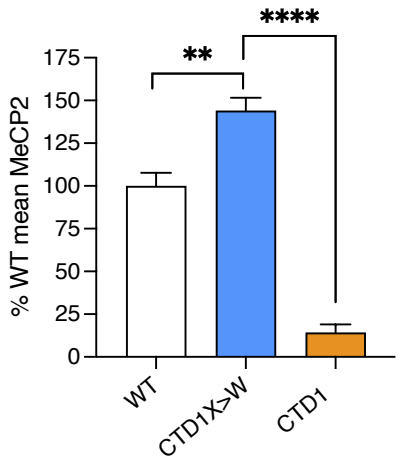


Figure 4. A CTD1 X>W knock-in mouse models adenine base editing of the CTD1 stop codon. (**A**) The genomic sequence of mouse CTD1 and CTD1 X>W alleles. The mutated adenine is shown in orange, and the sequence common to all mouse CTD alleles is shaded in turquoise. (**B**) The amino acid consequences 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 male littermates of CTD1 X>W animals (n=10). Mean +/- 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 +/- sd. Unpaired two-tailed t-test: CTD1 X>W vs CTD1 P<0.0001 (****), CTD 1 X>W vs WT

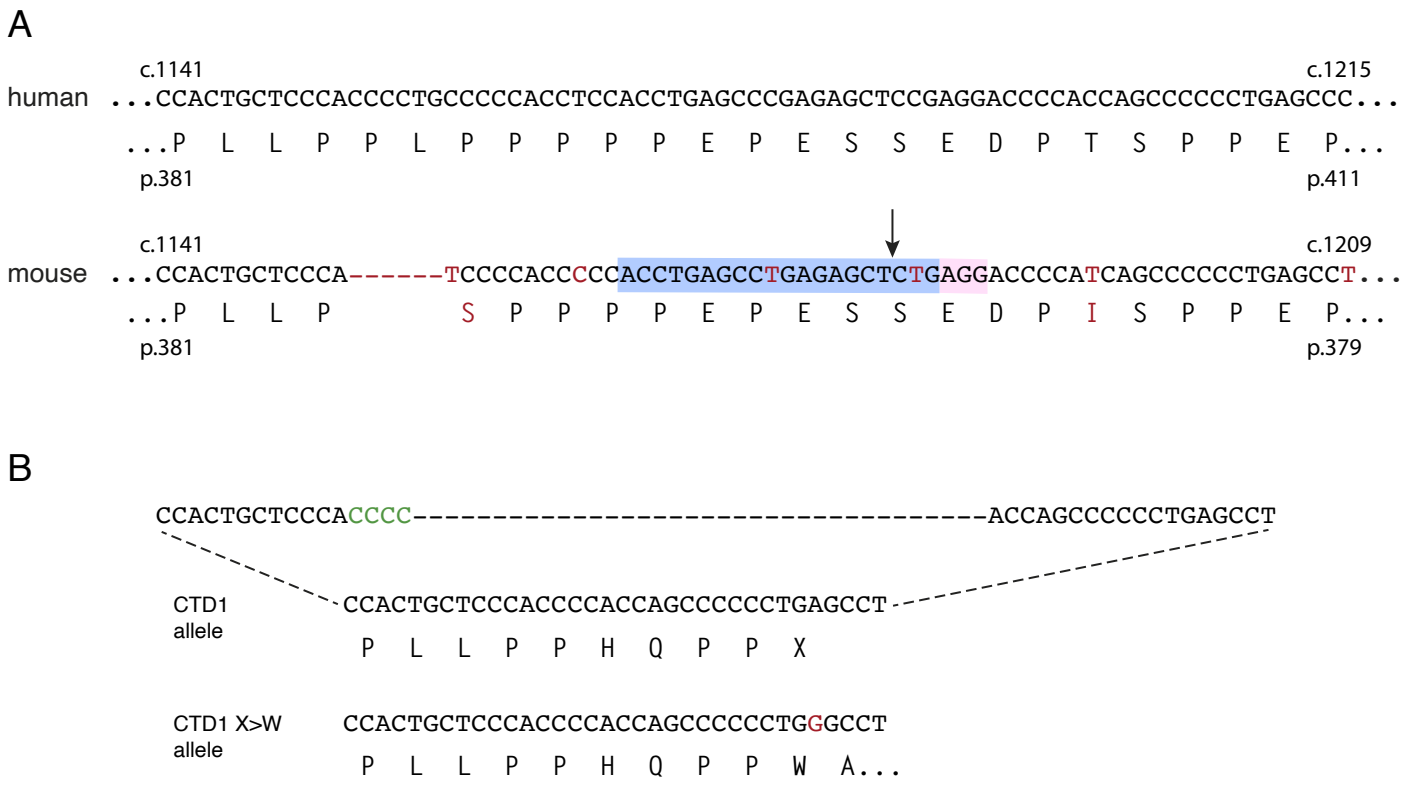


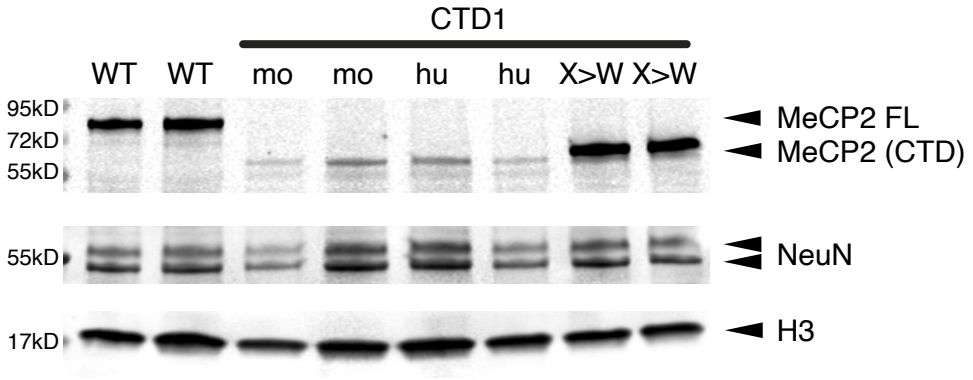
Figure 4, Supplementary Figure 1 Structure of the mouse CTD1 X>W allele. (**A**) Human and mouse CT-DPR genomic and amino acid sequences. Differences between mouse and human sequence are shown in red. The protospacer sequence (blue) and PAM sequence (pink) used to cut the WT allele for CRISPR editing 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.

A

CTD1mo TCAGAGTCCACAAAGGCCCCCATGCCACTGCTCCCACCCCACCAGCCCCCCTGAGCCT
S E S T K A P M P L L P P H Q P P X

CTD1hu TCAGAGTCCCCAAAGGCGCCCGTGCCACTGCTCCCACCCCACCAGCCCCCCTGAGCCT
S E S P K A P V P L L P P H Q P P X

CTD1 X>W TCAGAGTCCCCAAAGGCGCCCGTGCCACTGCTCCCACCCCACCAGCCCCCCTGGGCCT
S E S P K A P V P L L P P H Q P P W A...



B

CTD2mo TCAGAGTCCACAAAGGCCCCCATGCCACTGCTCCCACCCCTGTCCCCCTGAGCCT
S E S T K A P M P L L P P L S P X

CTD2hu TCAGAGTCCCCAAAGGCGCCCGTGCCACTGCTCCCACCGCTCCCCCCCTGAGCCT
S E S P K A P V P L L P P L P P X

CTD2hu X>W TCAGAGTCCCCAAAGGCGCCCGTGCCACTGCTCCCACCGCTCCCCCCCTGGGCCT
S E S P K A P V P L L P P L P P W A...

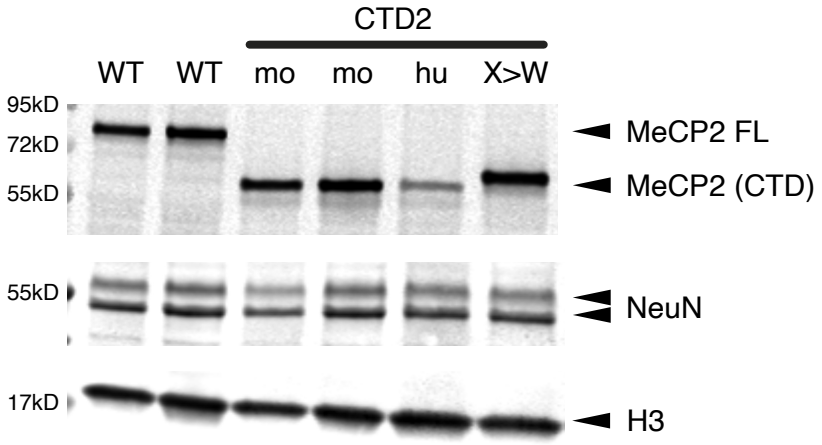
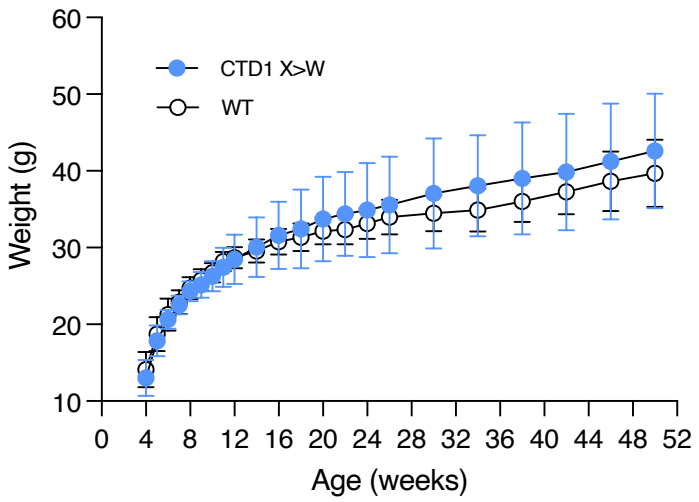


Figure 4, Supplementary Figure 2 CTD mouse alleles: mESC-derived neurons. (**A**) genomic and amino acid sequences of 3 CTD1 mouse knock-in alleles and western blot of MeCP2 protein from mESC-derived neurons 7 days after plating neuronal progenitors. Two independent clones per genotype, histone H3 loading 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.

A



B

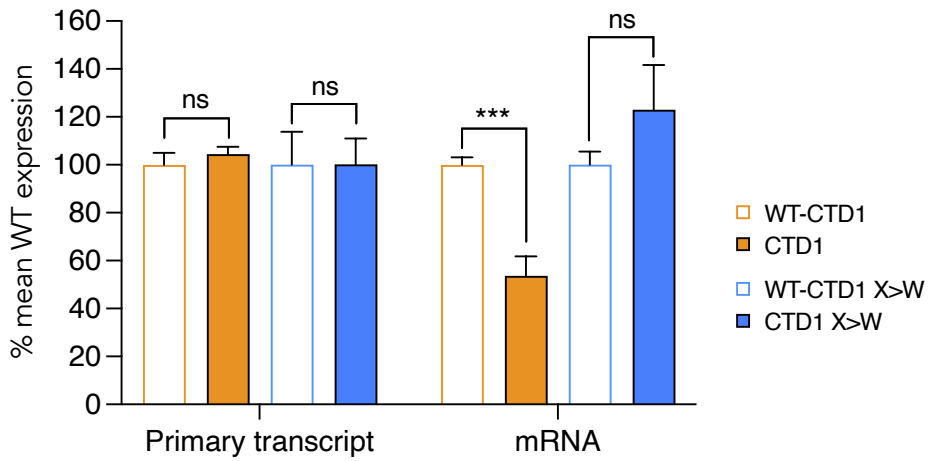
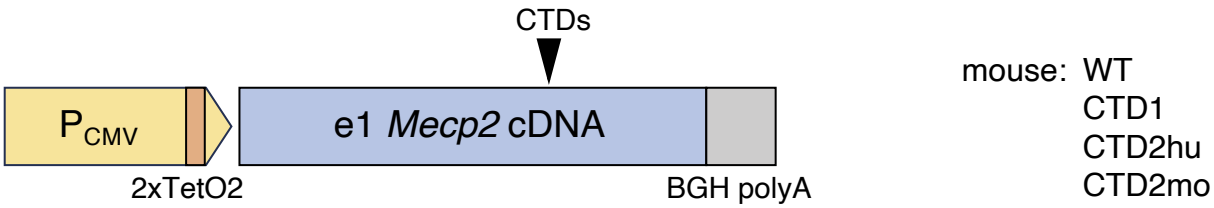
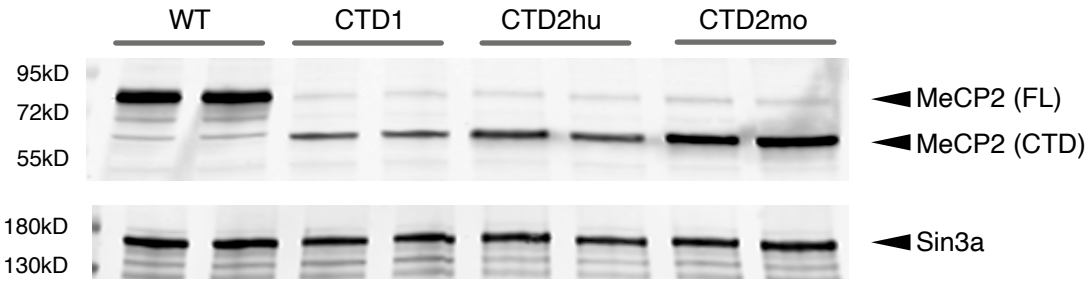


Figure 4, Supplementary Figure 3 CTD1 X>W knock-in mice: weights and brain RNA levels. (**A**) Body weights of animals shown in Figure 4C and D. Mean +/- sd. CTD1 (n=13) and CTD1 X>W (n=9), WT male littermates of CTD1 X>W animals (n=10). (**B**) Quantification of Mecp2 primary transcript and mRNA in whole brain of 6 week old male mice. N=3 brains per genotype. Mean +/- 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 (***).

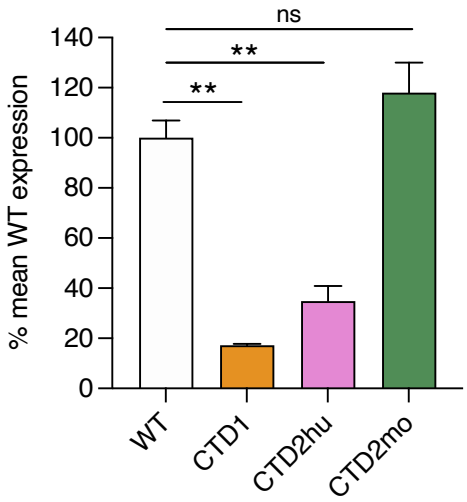
A



B



C



D

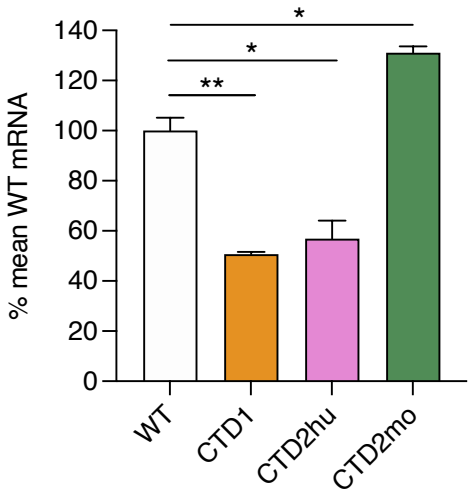


Figure 5. Flp-In T-REx cell lines reproduce the reduction in MeCP2 protein and mRNA seen with CTD knock-in mouse alleles. (**A**) Schematic of Mecp2 transgenes in Flp-In T-REx cell lines. Deletions are introduced into a full-length e1 Mecp2 cDNA, with a bovine growth hormone (BGH) polyadenylation signal 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 +/- 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 +/- sd. Unpaired two-tailed t-test: WT vs CTD1 P=0.0057 (**), WT vs CTD2hu P=0.0208 (*), WT vs CTD2mo P=0.0171 (*).

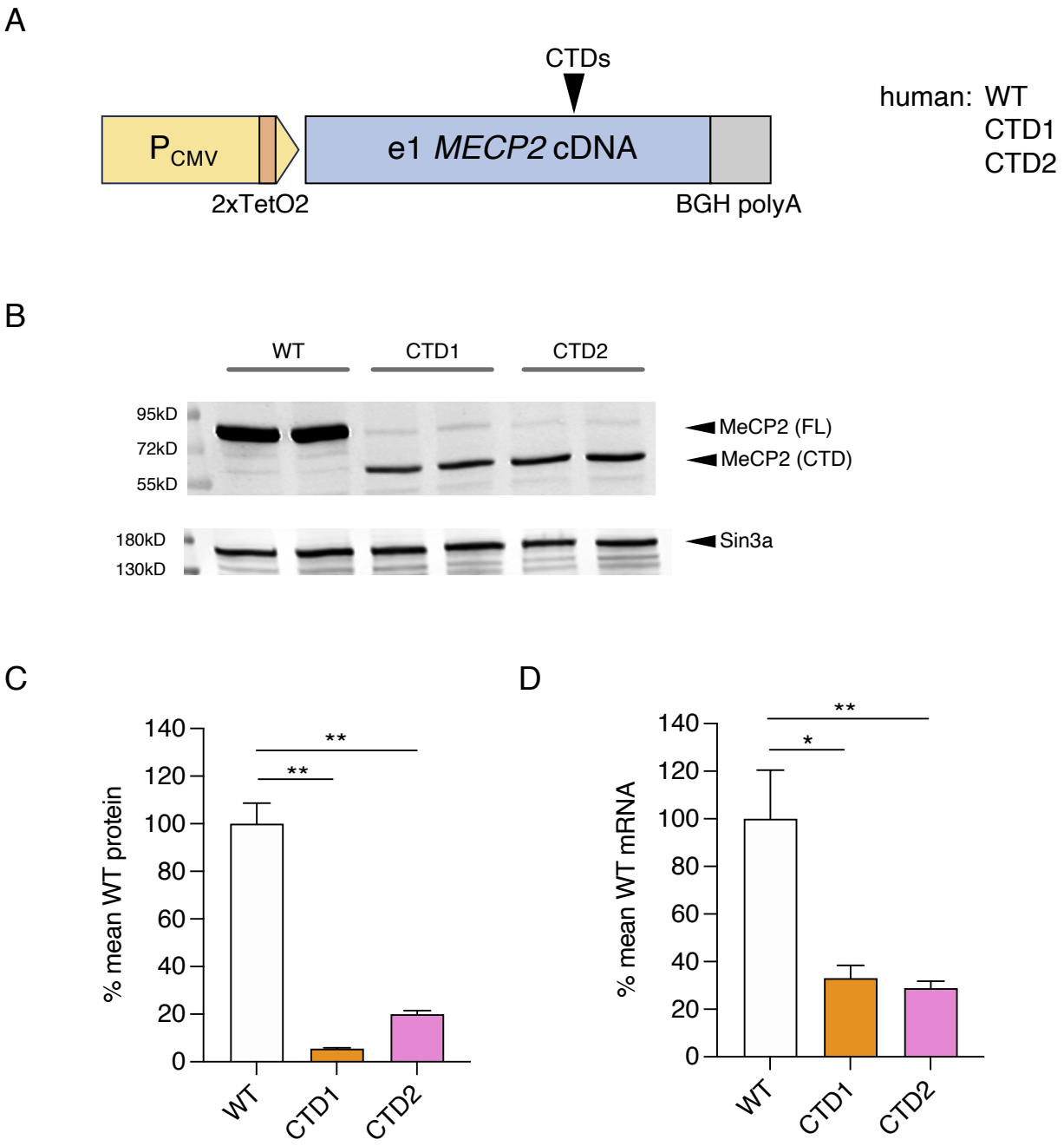


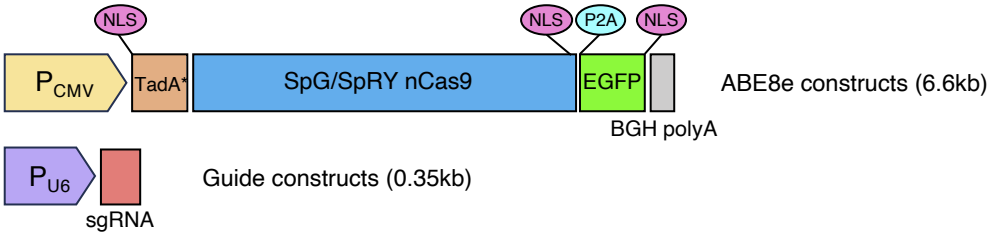
Figure 5, Supplementary Figure 1 Flp-In T-REx cell lines with human MECP2 transgenes. (**A**) Schematic of human MECP2 transgenes in Flp-In T-REx cell lines. (**B**) Western blot with whole cell lysates from independent Flp-In T-REx clones carrying human cDNA transgenes (24 hours tetracycline induction). Sin3a loading control. (**C**) Quantification of MeCP2 protein expression from (B). N=2 clones per genotype, mean +/- 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 +/- sd. Unpaired two-tailed t-test: WT vs CTD1 P=0.0126 (*), WT vs CTD2 P=0.0099 (**), WT vs CTD2mo P=0.0171 (*).

A

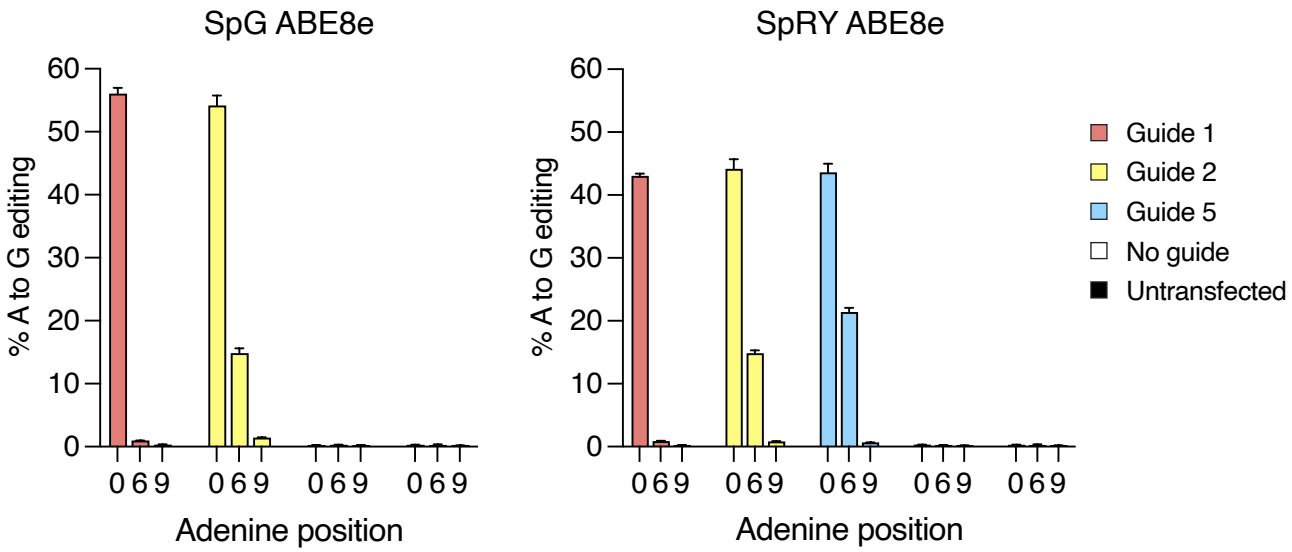
```
                    0       6   9
CACCAGCCCCCCTGAGCCTCAGGACTTGAGCAGCAG     mouse CTD1
GTGGTCGGGGGGACTCGGAGTCCTGAACTCGTCGTC     target sequence
 H   Q   P   P   X

      (G)CCCCTGAGCCTCAGGACTTG[AGC]       Guide 1 (target A7)
        (G)CCTGAGCCTCAGGACTTGAG[CAG]     Guide 5 (target A5)
         (G)CTGAGCCTCAGGACTTGAGC[AGC]    Guide 2 (target A4)
```

B



C



D

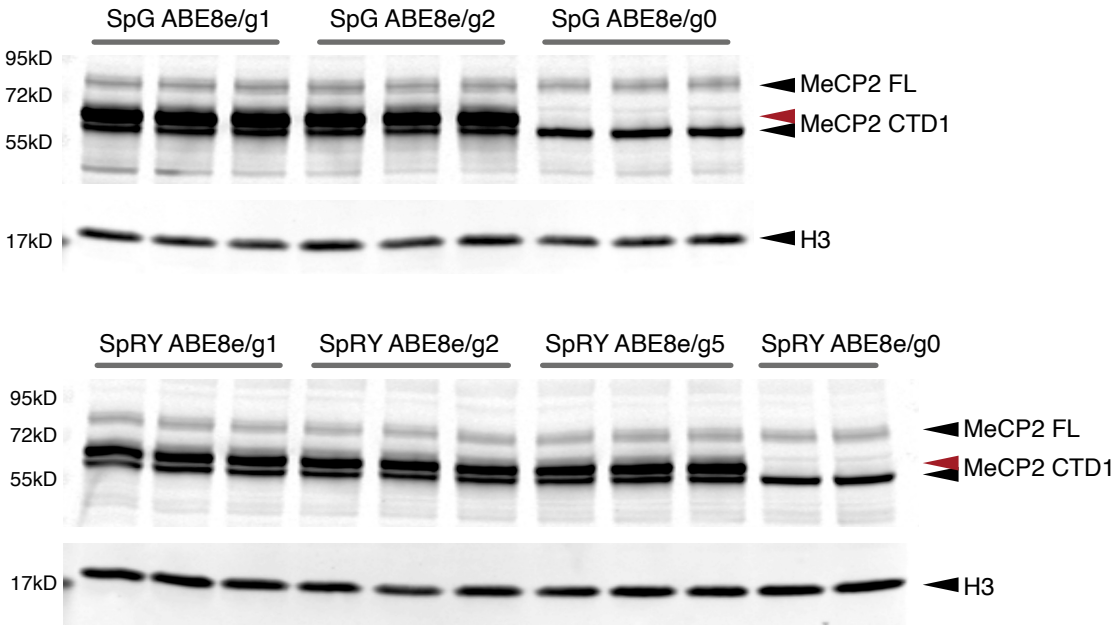


Figure 6

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).

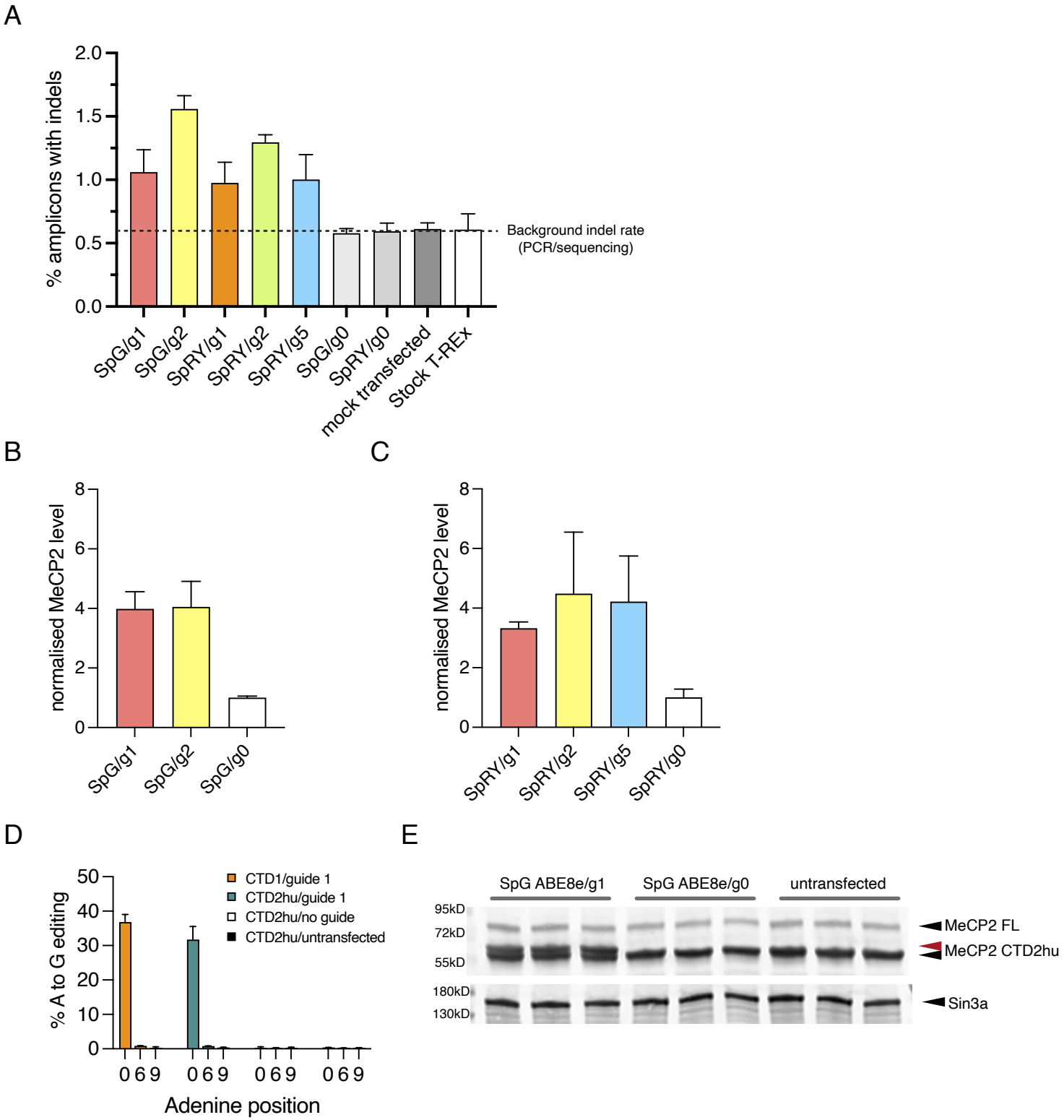


Figure 6, Supplementary Figure 1 Base editing of CTD transgenes in Flp-In T-REx cell lines. (**A**) mouse CTD Flp-In T-REx cells treated with ABE/guide RNAs as shown in Figure 6C. Percentage of mapped reads with indels. Background indel rate is indicated based on the % indels found in amplicons from 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 +/- sd, n=3 transfections per ABE/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).

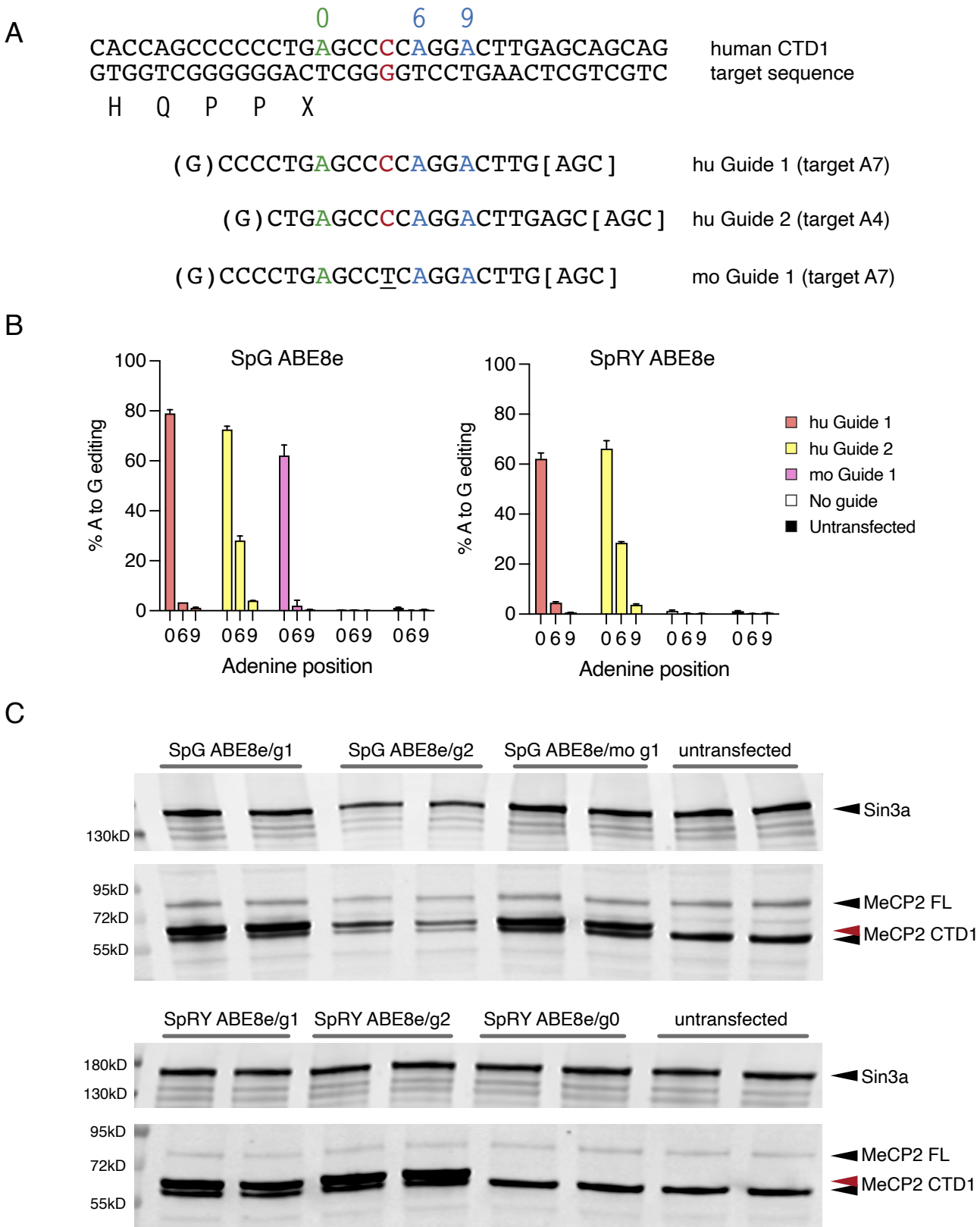


Figure 6, Supplementary Figure 2 Editing of human MECP2 transgenes in Flp-In T-REx cells . (**A**) Target genomic sequence and sgRNAs. Target A (position 0) is shown in green, with two bystander As within the guide sequence shown in blue (positions +6 and +9). The difference between human (red) and mouse (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.