Radically truncated MeCP2 rescues Rett syndrome-like neurological defects

Rebekah Tillotson[1], Jim Selfridge[1], Martha V. Koerner[1], Kamal K. E. Gadalla[2,3], Jacky Guy[1], Dina De Sousa[1], Ralph D. Hector[2], Stuart R. Cobb[2] and Adrian Bird[1,4]

[1] The Wellcome Centre for Cell Biology
University of Edinburgh
Michael Swann Building
King's Buildings
Max Born Crescent
Edinburgh
EH9 3BF
UK

[2] Institute of Neuroscience and Psychology,
College of Medical, Veterinary and Life Sciences,
University of Glasgow,
Glasgow
G12 8QQ
UK

[3] Pharmacology Department,
Faculty of Medicine,
Tanta University,
Tanta 31527,
Egypt

[4] Corresponding author

Heterozygous mutations in the X-linked *MECP2* gene cause the profound neurological disorder Rett syndrome (RTT)[1]. MeCP2 protein is an epigenetic reader whose binding to chromatin primarily depends on 5-methylcytosine (mC)[2,3]. Functionally, MeCP2 has been implicated in several cellular processes based on its reported interaction with >40 binding partners[4], including transcriptional co-repressors (e.g. the NCoR/SMRT complex[5]), transcriptional activators[6], RNA[7], chromatin remodellers[8,9], microRNA-processing proteins[10] and splicing factors[11]. Accordingly, MeCP2 has been cast as a multi-functional hub that integrates diverse processes that are essential in mature neurons[12]. At odds with the concept of broad functionality, missense mutations that cause RTT are concentrated in two discrete clusters coinciding with interaction sites for partner macromolecules: the Methyl-CpG Binding Domain (MBD)[13] and the NCoR/SMRT Interaction Domain (NID)[5]. Here, we test the hypothesis that the single dominant function of MeCP2 is to physically connect DNA with the NCoR/SMRT complex, by removing almost all amino acid sequences except the MBD and NID. We find that mice expressing truncated MeCP2 lacking both the N- and C-terminal regions (approximately half of the native protein) are phenotypically near-normal; and those expressing a minimal MeCP2 additionally lacking a central domain survive for over one year with only mild symptoms. This minimal protein is able to prevent or reverse neurological symptoms when introduced into MeCP2-deficient mice by genetic activation or virus-mediated delivery to the brain. Thus, despite evolutionary conservation of the entire MeCP2 protein sequence, the DNA and co-repressor binding domains alone are sufficient to avoid RTT-like defects and may therefore have therapeutic utility.

The amino acid sequence of MeCP2 is highly conserved throughout vertebrate species (Fig. 1a), suggesting that most of the protein is under evolutionary selection. Accordingly, full-length MeCP2 is reported to interact with multiple binding partners and has been implicated in several cellular pathways required for neuronal function[12,4]. RTT-causing missense mutations, however, are concentrated in the MBD and NID – a small minority of the protein – whereas the general population shows numerous polymorphisms elsewhere in the protein suggesting that other regions may be

dispensable (Fig. 1a). To test whether the MBD and NID might be sufficient for MeCP2 function, we generated mouse lines expressing a stepwise series of deletions of MeCP2. The three regions removed were sequences N-terminal to the MBD ('N'), C-terminal to the NID ('C') and the intervening amino acids between these domains ('I') (Fig. 1b). The *Mecp2* gene has four exons, with transcripts alternatively spliced to produce two isoforms that differ only at the extreme N-termini [14]. To conserve the *Mecp2* gene structure in the knock-in mice, exons 1 and 2 and the first 10 bp of exon 3 (splice acceptor site) were retained, resulting in the inclusion of 29 and 12 N-terminal amino acids from isoforms e1 and e2, respectively (Extended Data Fig. 1a-b, 3, 5). A C-terminal EGFP tag was added to facilitate detection and recovery (Fig. 1b). We defined the MBD as residues 72-173 and the NID as residues 272-312 (Extended Data Fig. 1c-d). The intervening region of the ΔNIC allele was replaced by a nuclear localisation signal (NLS) from SV40 virus, connected by a short flexible linker. The proportions of native MeCP2 protein sequence retained in ΔN, ΔNC and ΔNIC are 88%, 52% and 32%, respectively.

We tested whether the truncated MeCP2 proteins retained the ability to interact with methylated DNA and the NCoR/SMRT co-repressor complex using cell culture-based assays. They each immunoprecipitated endogenous NCoR/SMRT complex components when overexpressed in HeLa cells, whereas this interaction was abolished in the negative control NID mutant, R306C (Extended Data Fig. 2a). They also localised to mCpG-rich heterochromatic foci in mouse fibroblasts, which is dependent on both DNA methylation[2,16] and MBD functionality[17], whereas the negative control MBD mutant (R111G) was diffusely distributed (Extended Data Fig. 2b). Finally, we tested whether the truncated derivatives were able to recruit TBL1X, an NCoR/SMRT complex subunit that interacts directly with MeCP2[5,18], to heterochromatin. Transiently expressed TBL1X-mCherry accumulates in the cytoplasm, but it is efficiently recruited to heterochromatic foci in the presence of co-expressed WT MeCP2[5]. All three derivative proteins successfully bridged DNA with TBL1X-mCherry *in vivo* , whereas the negative control NID mutant (R306C) could not do so (Extended Data Fig. 2c). All

truncated proteins therefore retained the ability to bind methylated DNA and the NCoR/SMRT complex simultaneously.

We generated *ΔN* and *ΔNC* knock-in mice by replacing the endogenous *Mecp2* allele in ES cells, which were used to produce germline-transmitting chimaeras (Extended Data Fig. 3). Truncated proteins were expressed at approximately WT levels in brain and in neurons (Extended Data Fig. 4a-d). To assess phenotypes, knock-in mice were crossed onto a C57BL/6J background (for four generations) and cohorts underwent weekly phenotypic scoring[19,20] or behavioural analysis. Although heterozygous female mice are the genetic model for RTT, phenotypes develop late and are mild in the case of hypomorphic *Mecp2* mutations[21,15]. Hemizygous males provide a more sensitive assay of MeCP2 function: *Mecp2*-null males exhibit severe phenotypes that develop shortly after weaning and median survival is 9 weeks[21]. Both *ΔN* and *ΔNC* male mice were viable, fertile and showed phenotypic scores indistinguishable from *WT* littermates over one year (Fig. 2a-d). *ΔN* mice had normal body weight (Extended Data Fig. 4e), whereas *ΔNC* mice were slightly heavier than *WT* littermates (Extended Data Fig. 4f). This difference was absent in a more outbred cohort (Extended Data Fig. 4g), consistent with previous observations that body weight of *Mecp2* mutants is affected by genetic background[21].

At 20 weeks of age, cohorts were tested for RTT-like behaviours: hypoactivity, decreased anxiety and reduced motor abilities. Neither activity (distance travelled in an Open Field; Extended Data Fig. 4h) nor anxiety (time spent in the open arms of the Elevated Plus Maze; Fig. 2e) was abnormal in *ΔN* and *ΔNC* mice, although the latter did spend longer in the centre of the Open Field (Fig. 2f), indicative of mildly decreased anxiety. Motor coordination was assessed using the Accelerating Rotarod test over three days. Whereas mouse models of RTT show impaired performance that was most striking on the third day[22,15], *ΔN* and *ΔNC* mice were comparable to *WT* littermates throughout this test (Fig. 2g). Overall, the results suggest that contributions of the N- and C-terminal regions to MeCP2 function are at best subtle. The result is remarkable given the presence of a neurological phenotype

in male mice expressing a slightly more severe C-terminal truncation, which lacks residues beyond T308[23]. The difference may be explained by retention of full NID function in *ΔNC* mice, as loss of the extra four C-terminal amino acids (309-312) markedly reduces the affinity of truncated MeCP2 for the NCoR/SMRT co-repressor complex[5].

We next replaced the endogenous *Mecp2* gene with *ΔNIC*, a minimal allele comprising only the MBD and NID domains and retaining 32% of the full-length protein sequence (Fig. 1b, Extended Data Fig. 5). ΔNIC protein levels were reduced in whole brain (~50% of WT-EGFP controls; Fig. 3a-b) and in neurons (~40% of WT-EGFP controls; Fig. 3b). The presence of normal levels of mRNA in *ΔNIC* mice (Fig. 3c) suggests that deletion of the intervening region compromises protein stability. Despite low protein levels, male *ΔNIC* mice had a normal lifespan (Fig. 3e, Extended Data Fig. 6a). However, phenotypic scoring over one year detected mild neurological phenotypes (Fig. 3d), predominantly gait abnormalities and partial hind-limb clasping. These symptoms were relatively stable throughout the scoring period. *ΔNIC* mice also weighed ~40% less than their WT littermates (Extended Data Fig. 6b).

Behavioural analysis of a separate cohort at 20 weeks showed decreased anxiety in male *ΔNIC* mice, signified by reduced time spent in the closed arms of an Elevated Plus Maze (Fig. 3f), although this phenotype was not detected by the Open Field test (Fig. 3g). No activity phenotype was detected in the Open Field (Extended Data Fig. 6c), but, consistent with the gait defects detected by weekly scoring, *ΔNIC* mice displayed declining motor coordination on the Accelerating Rotarod over three days, culminating in a significantly impaired performance on the third day (Fig. 3h). It is noteworthy that *ΔNIC* animals are much less severely affected than male mice with the mildest common mutation found in RTT patients, R133C, which had a median lifespan of 42 weeks, higher phenotypic scores and a more pronounced reduction in body weight[15] (Extended Data Fig. 7). Reduced protein levels may contribute to the relatively mild phenotype, as mice with ~50% levels of full-length MeCP2 have neurological defects[24].

To further test ΔNIC functionality, we asked whether late genetic activation could reverse phenotypic defects in symptomatic MeCP2-deficient mice, as previously shown with the full-length protein[19]. Mice that were MeCP2-deficient through insertion of a floxed transcriptional STOP cassette in intron 2 of the *ΔNIC* gene (Extended Data Fig. 5, 8a-b) resembled *Mecp2*-nulls (Extended Data Fig. 8c-d). This line was crossed with mice carrying a C*reER*^{T} transgene (Cre recombinase fused to a modified estrogen receptor) to enable removal of the STOP cassette upon Tamoxifen treatment. Induced expression of ΔNIC after the onset of symptoms in *STOP CreER*^{T} mice (Fig. 4a) resulted in high levels of Cre-mediated recombination (Extended Data Fig. 9a) and protein levels similar to those of *ΔNIC* mice (Extended Data Fig. 9b). *ΔNIC* activation had a dramatic effect on phenotypic progression, relieving neurological symptoms and restoring normal survival (Fig. 4b-c). Separation of the phenotypic scores into the six tested components showed clear reversal of tremor, hypoactivity and gait abnormalities (Extended Data Fig. 9c). In contrast, control *STOP* mice lacking the *CreER*^{T} transgene developed severe symptoms and failed to survive beyond 26 weeks. Thus, despite its radically reduced length and relatively low abundance, *ΔNIC* was able to effectively rescue MeCP2-deficient mice from RTT-like phenotypes.

This finding prompted us to explore whether ΔNIC could be used for gene therapy in *Mecp2*-null mice. A human version of the *ΔNIC* gene (Fig. 4d), driven by a minimal *Mecp2* promoter[25], was tagged with a short Myc epitope (in place of EGFP) and packaged into a self-complementary adeno-associated viral vector (scAAV). Neonatal mice (P1-2) injected intra-cranially with this virus (Fig. 4d) expressed hΔNIC protein (Extended Data Fig. 10a). Treated *Mecp2*-null mice showed reduced symptom severity and greatly extended survival compared with controls receiving vehicle alone (Fig. 4e-g). Despite the lack of fine control over infection rate, we did not observe deleterious effects due to overexpression, even in WT animals (Extended Data Fig. 10b-d). It is possible that the moderate instability of hΔNIC protein mitigates toxic effects associated with overexpression, widening the dosage window. The results also demonstrate that hΔNIC protein is functional without the large EGFP tag. Minimal MeCP2 may therefore be therapeutically advantageous, as the shortening the

157 coding sequence creates room for additional regulatory sequences within the limited capacity of
158 scAAV vectors, which may enable more precise control of expression.

159 Our findings support a simple model whereby the predominant function of MeCP2 is to recruit the
160 NCoR/SMRT co-repressor complex to methylated sites on chromatin. This scenario agrees with
161 recent evidence that inhibition of gene transcription is proportional to MeCP2 occupancy within
162 gene bodies[26,27]. Importantly, minimal MeCP2 protein (ΔNIC) is missing all or part of several domains
163 that have been highlighted as potentially important, including the AT-hooks[28], several activity-
164 dependent phosphorylation sites[29,30] , an RNA binding motif[7] and interaction sites for proteins
165 implicated in micro-RNA processing[10], splicing[11] and chromatin remodelling[9]. While these parts of
166 the protein may have biological relevance, their presence is evidently not required for prevention of
167 the RTT-like phenotype. Importantly, the discovery that the MBD and NID are sufficient to partially
168 restore neuronal function to MeCP2-deficient mice allowed us to explore the therapeutic potential
169 of the minimal protein, with encouraging results. These results potentially set a precedent for
170 reducing the length of other gene therapy constructs by identifying dispensable regions that cannot
171 be predicted by evolutionary conservation.

172

173

References

1. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–8 (1999).

2. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69,** 905–14 (1992).

3. Skene, P. J. *et al.* Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* **37,** 457–468 (2010).

4. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16,** 261–274 (2015).

5. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16,** 898–902 (2013).

6. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science (80-.).* **320,** 1224–9 (2008).

7. Jeffery, L. & Nakielny, S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.* **279,** 49479–49487 (2004).

8. Nan, X. *et al.* Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2709–14 (2007).

9. Agarwal, N. *et al.* MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res.* **35,** 5402–8 (2007).

10. Cheng, T.-L. *et al.* MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. *Dev. Cell* **28,** 547–60 (2014).

11. Young, J. I. *et al.* Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17551–8 (2005).

12. Ragione, F. Della, Vacca, M., Fioriniello, S., Pepe, G. & Esposito, M. D. MECP2 , a multi-talented modulator of chromatin architecture. *Brief. Funct. Genomics* **15,** 1–12 (2016).

13. Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **21,** 4886–4892 (1993).

14. Kriaucionis, S. & Bird, A. The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res.* **32,** 1818–23 (2004).

15. Brown, K. *et al.* The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* **25,** 558–570 (2016).

16. Nan, X., Tate, P., Li, E. & Bird, A. DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.* **16,** 414–21 (1996).

17. Kudo, S. *et al.* Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J. Med. Genet.* **40,** 487–93 (2003).

18. Kruusvee, V. *et al.* Structure of the MeCP2–TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. *Proc. Natl. Acad. Sci.* 17007311114 (2017). doi:10.1073/pnas.1700731114

19. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science (80-.).* **315,** 1143–7 (2007).

20. Cheval, H. *et al.* Postnatal inactivation reveals enhanced requirement for MeCP2 at distinct age windows. *Hum. Mol. Genet.* **21,** 3806–3814 (2012).

21. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null mutation causes

neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27,** 322–6 (2001).

22. Goffin, D. *et al.* Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat. Neurosci.* **15,** 274–83 (2012).

23. Shahbazian, M. *et al.* Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35,** 243–54 (2002).

24. Samaco, R. C. *et al.* A partial loss of function allele of Methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. *Hum. Mol. Genet.* **17,** 1718–1727 (2008).

25. Gadalla, K. K. E. *et al.* 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 (2017).

26. Lagger, S. *et al.* MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLOS Genet.* **13,** e1006793 (2017).

27. Kinde, B., Wu, D. Y., Greenberg, M. E. & Gabel, H. W. DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc. Natl. Acad. Sci.* **113,** 15114–15119 (2016).

28. Baker, S. A. *et al.* An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell* **152,** 984–96 (2013).

29. Zhou, Z. *et al.* Brain-Specific Phosphorylation of MeCP2 Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation. *Neuron* **52,** 255–269 (2006).

30. Li, H., Zhong, X., Chau, K. F., Williams, E. C. & Chang, Q. Loss of activity-induced phosphorylation of MeCP2 enhances synaptogenesis, LTP and spatial memory. *Nat. Neurosci.* **14,** 1001–8 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements

This work was supported by the Sylvia Aiken Charitable Trust, the Rett Syndrome Research Trust and Wellcome. R.T. was funded by a BBSRC Doctoral Training Partnership studentship. We thank the following people for assistance: Atlanta Cook (advice on designing the truncated proteins), Alan McClure (animal husbandry), David Kelly (microscopy), Martin Waterfall (flow cytometry) and Alastair Kerr (statistics). We also thank members of the Bird, Cobb, M. E. Greenberg and G. Mandel labs for helpful discussions. A.B. and S.R.C. are members of the Simons Initiative for the Developing Brain at the University of Edinburgh.

Author contributions

R.T., A.B. and S.R.C. designed research. R.T., J.S., M.V.K., K.K.E.G., J.G., D.D.S and R.D.H performed the experiments. R.T. and S.R.C. analysed the data. R.T. and A.B. wrote the manuscript. All authors reviewed the manuscript.

Author information

 Conflict of interest statement: A.B. is a member of the Board of ArRETT, a company based in the United States with the goal of developing therapies for Rett syndrome. Correspondence and requests for materials should be addressed to A.B. (a.bird@ed.ac.uk).

Figure 1: Stepwise truncation of MeCP2 protein to retain only the MBD and NID

a, Diagram of human MeCP2 protein sequence with the Methyl-CpG Binding Domain (MBD) and the NCoR/SMRT Interaction Domain (NID); annotated to show single nucleotide polymorphisms (SNPs) in males in the general population (black lines) and RTT-causing missense mutations (red lines). Sequence identity between human and other vertebrate MeCP2 proteins is shown by purple bars and insertions by dark lines. **b,** Schematic diagram of the deletion series based on the mouse e2 isoform that were generated in this study, compared with WT-EGFP[15].

Figure 2: Deletion of the MeCP2 N- and C-terminal regions has minimal phenotypic consequence

a, b, Phenotypic severity scores of hemizygous male (**a**) *ΔN* mice (*n*=10) and (**b**) *ΔNC* mice (*n*=10), compared to their *WT* littermates (*n*=10) over one year. Graphs show mean scores ± S.E.M. Published *Mecp2*-null data (*n*=12)[15] is shown as a comparator. **c, d,** Kaplan-Meier plots showing survival of the cohorts shown in panels **a** and **b**. *Mecp2*-null data (*n*=24)[15] is shown as a comparator. **e,f,g,** Behavioural analysis of separate cohorts performed at 20 weeks of age: *ΔN* (*n*=10) and *ΔNC* mice (*n*=10 for Open Field/Rotarod; 11 for Elevated Plus Maze), each compared to *WT* littermates (*n*=10). Graphs show individual values and medians, and statistical significance as follows : not significant ('n.s.') $P>0.05$, * $P<0.05$. **e,** Time spent in the closed and open arms of the Elevated Plus Maze during a 15 min trial. Genotypes were compared using KS tests: *ΔN* closed arms $P=0.988$ and open arms $P=0.759$; *ΔNC* closed arms $P=0.950$ and open arms $P=0.932$. **f,** Time spent in the central region of the Open Field test was measured during a 20 minute trial. Genotypes were compared using t-tests: *ΔN* $P=0.822$; *ΔNC* * $P=0.020$. **g,** Average latency to fall from the Accelerating Rotarod in four trials was calculated for each of the three days of the experiment. Genotypes were compared using KS tests: *ΔN* day 1 $P=0.759$, day 2 $P=0.401$ and day 3 $P=0.055$; *ΔNC* day 1 $P=0.988$, day 2 $P=0.401$ and day 3 $P=0.759$.

Figure 3: Additional deletion of the intervening region leads to protein instability and mild RTT-like symptoms

a, Western blot analysis of whole brain extract showing protein sizes and abundance of MeCP2 in *ΔNIC* mice and *WT-EGFP* controls, detected using a GFP antibody. Histone H3 was used as a loading control. *denotes a non-specific band detected by the GFP antibody. For gel source data, see Supplementary Information. **b,** Flow cytometry analysis of protein levels in nuclei from whole brain ('All') and the high-NeuN subpopulation ('Neurons') in *ΔNIC* mice (*n*=3) and *WT-EGFP* controls (*n*=3), detected using EGFP fluorescence. Graph shows mean ± S.E.M. and genotypes were compared by t-test: 'All' *** *P*=0.0002 and 'Neurons' *** *P*=0.0001. 'au' = arbitrary units. **c,** Quantitative PCR analysis of mRNA prepared from whole brain of *ΔNIC* mice (*n*=3) and *WT-EGFP* controls (*n*=3). *Mecp2* transcript levels were normalised to *Cyclophilin A* mRNA. Graph shows mean ± S.E.M. (relative to *WT-EGFP*) and genotypes were compared by t-test: *P*=0.110. **d,** Phenotypic severity scores of *ΔNIC* mice (*n*=10) compared to *WT* littermates (*n*=10) over one year. Graph shows mean scores ± S.E.M. *Mecp2*-null data (*n*=12)[15] is shown as a comparator. **e,** Kaplan-Meier plot showing survival of the cohort shown in panel **d**. One *ΔNIC* animal died at 43 weeks, after receiving phenotypic scores of ≤2.5. *Mecp2*-null data (*n*=24)[15] is shown as a comparator. **f, g, h,** Behavioural analysis of a separate cohort performed at 20 weeks of age: *ΔNIC* (*n*=10) compared to *WT* littermates (*n*=10). Graphs show individual values and medians, and statistical significance as follows: not significant ('n.s.') *P*>0.05, * *P*<0.05, ** *P*<0.01. **f,** Time spent in the closed and open arms and centre of the Elevated Plus Maze during a 15 minute trial. Genotypes were compared using KS tests: closed arms ** *P*=0.003, open arms *P*=0.055 and centre * *P*=0.015. **g,** Time spent in the central region of the Open Field measured during a 20 minute trial. Genotypes were compared using a t-test: *P*=0.402. **h,** Average latency to fall from the Accelerating Rotarod in four trials was calculated for each of the three days of the experiment. Genotypes were compared using KS tests: day 1 *P*=0.164, day 2 *P*=0.055 and day 3 **

316 *P*=0.003. Changed performance (learning/worsening) over the three day period was determined
317 using Friedman tests: wild-type animals *P*=0.601, *ΔNIC* animals ** *P*=0.003.
318
319 **Figure 4: Activation or viral transduction of ΔNIC ameliorates neurological phenotypes in MeCP2-**
320 **deficient mice**
321
322 **a,** Timeline of Cre-mediated activation of *ΔNIC* induced by Tamoxifen injections. **b,** Phenotypic
323 severity scores (mean ± SEM) of mice injected with Tamoxifen (arrows) from 4-28 weeks: *WT* (*n*=4),
324 *WT CreER*^{T} (*n*=4), *STOP* (*n*=9) and *STOP CreER* (*n*=9). **c,** Kaplan-Meier plot showing survival of the
325 cohort shown in panel **b**. **d,** Diagram of the DNA sequence inserted into an scAAV viral vector,
326 comprising a 426 nt *Mecp2* promoter driving the human *ΔNIC* coding sequence plus a C-terminal
327 Myc tag and 3' UTR. A vector containing full-length human *MECP2*[25] is shown for comparison. **e,**
328 Timeline of the scAAV-mediated gene therapy experiment. **f,** Phenotypic severity scores (mean ±
329 SEM) of scAAV-injected mice from 5-30 weeks: *WT* + vehicle (*n*=15), *Mecp2*-null + vehicle (*n*=20) and
330 *Mecp2*-null + *hΔNIC* (*n*=17). **g,** Kaplan-Meier plot showing survival of the cohort shown in panel **f**.
331 Four *Mecp2*-null + *hΔNIC* animals reached their humane end-point. Five *Mecp2*-null + *ΔNIC* animals
332 were culled due to injuries unrelated to RTT-like phenotypes at 16, 23, 25, 26 and 29 weeks of age
333 (data shown as ticks). Survival of *Mecp2*-null + *ΔNIC* animals was compared to *Mecp2*-null + vehicle
334 controls using the Mantel-Cox test: *P*=<0.0001.
335
336

Extended Data Figure 1: Design of the MeCP2 deletion series

a, Diagram of the genomic DNA sequences encoding WT and ΔNIC MeCP2, showing the retention of the extreme N-terminal amino acids encoded in exons 1 and 2 and the first 10 bp of exon 3, the deletion of the N- and C-terminal regions, the replacement of the intervening region with a linker and SV40 NLS, and the addition of the C-terminal EGFP tag. Colour key: 5'UTR=white, MBD=blue, NID=pink, other MeCP2 coding regions=grey, SV40 NLS=orange, linkers=dark grey and EGFP=green. **b,** The N-terminal ends of the sequences of all three truncated proteins (e1 and e2 isoforms) showing the fusion of the extreme N-terminal amino acids to the MBD (starting with P72). **c, d,** Protein sequence alignment of the MBD (**c**) and NID (**d**) regions using ClustalWS, shaded according to BLOSUM62 score. Both alignments are annotated with RTT-causing missense mutations[31] (red), activity-dependent phosphorylation sites[29,32,33] (orange), sequence conservation, interaction domains and known[34]/predicted[35] structure. Interaction sites: methyl-DNA binding (residues 78-162[13]), AT hook 1 (residues 183-195[36]), AT hook 2 (residues 257-272[28]), NCoR/SMRT binding (residues 285-309[5]). The bipartite nuclear localisation signal (NLS) is also shown (residues 253-256 and 266-271). The regions retained in ΔNIC are: MBD resides 72-173 (highlighted by the blue shading in panel **c**) and NID resides 272-312 (highlighted by the pink shading panel d). Residue numbers correspond to that of mammalian e2 isoforms.

Extended Data Figure 2: Truncated MeCP2 proteins retain the ability to bind methylated DNA and the NCoR/SMRT complex

a, EGFP-tagged truncated proteins immunoprecipitate components of the NCoR/SMRT co-repressor complex: NCoR, HDAC3 and TBL1XR1. WT and R306C were used as positive and negative controls for binding, respectively. 'In' = input, 'IP' = immunoprecipiate. For gel source data, see Supplementary Information. **b,** EGFP-tagged truncated MeCP2 proteins localise to mCpG-rich heterochromatic foci when overexpressed in mouse fibroblasts (NIH-3T3 cells). WT and R111G were used as controls to show focal and diffuse localisation, respectively. Scale bars indicate 10 μm. **c,** EGFP-tagged truncated

proteins recruit TBL1X-mCherry to heterochromatin when co-overexpressed in NIH-3T3 cells. WT and R306C were used as positive and negative controls for TBL1X-mCherry recruitment, respectively. scale bars indicate 10 µm. Quantification (right) shows the percentage of cells with focal TBL1X-mCherry localisation, evaluated relative to WT using Fisher's exact tests: R306C **** $P<0.0001$, ΔN $P=0.071$, ΔNC $P=0.604$, ΔNIC $P=0.460$. Total numbers of cells counted: WT $n=117$, R306C $n=119$, ΔN $n=113$, ΔNC $n=119$, ΔNIC $n=125$.

Extended Data Figure 3: Generation of ΔN and ΔNC mice

Diagrammatic representation of *ΔN* (**a**) and *ΔNC* (**b**) knock-in mouse line generation. The endogenous *Mecp2* allele was targeted in male ES cells. The site of Cas9 cleavage in the WT sequence is shown by the scissors symbol (used for production of *ΔN* knock-in ES cells). The selection cassette was removed *in vivo* by crossing chimaeras with deleter (*CMV-Cre*) transgenic mice. Southern blot analysis shows correct targeting of ES cells and successful cassette deletion in the knock-in mice. The solid black line represents the sequence encoded in the targeted vector and the dotted lines indicate the flanking regions of mouse genomic DNA. For gel source data, see Supplementary Information.

Extended Data Figure 4: *ΔN* and *ΔNC* knock-in mice express truncated proteins at approximately *WT* levels and display minimal phenotypes

a, Western blot analysis of whole brain extract showing protein sizes and abundance of MeCP2 in *ΔN* and *ΔNC* mice and *WT-EGFP* controls, detected using a GFP antibody. Histone H3 was used as a loading control. *denotes a non-specific band detected by the GFP antibody. For gel source data, see Supplementary Information. **b,** Flow cytometry analysis of protein levels in nuclei from whole brain ('All') and the high-NeuN subpopulations ('Neurons') in *WT-EGFP* ($n=3$), *ΔN* ($n=3$) and *ΔNC* ($n=3$) mice, detected using EGFP fluorescence. Graph shows mean ± S.E.M. and genotypes were compared to *WT-EGFP* controls by t-test: 'All' *ΔN* $P=0.338$, *ΔNC* ** $P=0.003$; and 'Neurons' *ΔN* $P=0.672$, *ΔNC* * $P=0.014$. **c,** Flow cytometry analysis of protein levels in *WT* ($n=3$) and *WT-EGFP* ($n=3$) mice, detected

using an MeCP2 antibody. Graph shows mean ± S.E.M. and genotypes were compared by t-test: 'All' *P*=0.214; and 'Neurons' *P*=0.085. **d,** Example histogram (of one *WT-EGFP* sample) showing how the 'Neuronal' subpopulation was defined according to NeuN-AF647 staining. 'au' = arbitrary units. **e, f, g,** Growth curves of the backcrossed scoring cohorts (**e** and **f**; see Fig. 2a-d) and an outbred (**g**; 75% C57BL/6J) cohort of *ΔNC* mice (*n*=7) and WT littermates (*n*=9). Graphs show mean values ± S.E.M. Genotypes were compared using repeated measures ANOVA: *ΔN P*=0.362, *ΔNC* **** *P*<0.0001, ΔNC (outbred) *P*=0.739. *Mecp2*-null data (*n*=20)[15] is shown as a comparator for the backcrossed cohorts. **h,** Behavioural analysis of *ΔN* (*n*=10) and *ΔNC* mice (*n*=10) compared to their WT littermates (*n*=10) at 20 weeks of age (see Fig. 2e-g). Total distance travelled in the Open Field test was measured during a 20 minute trial. Graphs show individual values and medians. Genotypes were compared using t-tests: *ΔN P*=0.691; *ΔNC P*=0.791. 'n.s.' = not significant.

Extended Data Figure 5: Generation of *ΔNIC* and *STOP* mice

Diagrammatic representation of *ΔNIC* and *STOP* mouse line generation. The endogenous *Mecp2* allele was targeted in male ES cells. The site of Cas9 cleavage in the WT sequence is shown by the scissors symbol. The selection cassette was removed *in vivo* by crossing chimaeras with deleter (*CMV-Cre*) transgenic mice to produce constitutively expressing *ΔNIC* mice, or retained to produce *STOP* mice. Southern blot analysis shows correct targeting of ES cells and successful cassette deletion in the *ΔNIC* knock-in mice. The solid black line represents the sequence encoded in the targeted vector and the dotted lines indicate the flanking regions of mouse genomic DNA. For gel source data, see Supplementary Information.

Extended Data Figure 6: *ΔNIC* mice have a normal lifespan and no activity phenotype but decreased body weight

a, Kaplan-Meier plot showing survival of an outbred (75% C57BL/6J) cohort of *ΔNIC* mice (*n*=10) and their *WT* littermate (*n*=1). **b,** Growth curve of the backcrossed cohort used for phenotypic scoring (see Fig. 3d-e). Graph shows mean ± S.E.M. Genotypes were compared using repeated measures ANOVA **** *P*<0.0001. *Mecp2*-null data (*n*=20)[15] is shown as a comparator. **c,** Behavioural analysis of *ΔNIC* mice (*n*=10) compared to their WT littermates (*n*=10) at 20 weeks of age (see Fig. 3f-h). Total distance travelled the Open Field test was measured during a 20 minute trial. Graph shows individual values and medians. Genotypes were compared using a t-test *P*=0.333. 'n.s.' = not significant.

Extended Data Figure 7: *ΔNIC* mice have a less severe phenotype than the mildest mouse model of RTT, R133C

a, b, c, Repeat presentation of phenotypic analysis of *ΔNIC* mice and *WT* littermates in Fig. 3d-e and Extended Data Fig. 6b, this time including EGFP-tagged R133C mice (*n*=10)[15] as a comparator.

Extended Data Figure 8: '*STOP*' mice with transcriptionally silenced *ΔNIC* resemble *Mecp2*-nulls

a, Western blot analysis of whole brain extract showing protein sizes and abundance of MeCP2 in *STOP* mice and WT-EGFP and *ΔNIC* controls, detected using a GFP antibody. Histone H3 was used as a loading control. *denotes a non-specific band detected by the GFP antibody. For gel source data, see Supplementary Information. **b,** Flow cytometry analysis of protein levels in nuclei from whole brain ('All') and the high-NeuN subpopulation ('Neurons') in *WT-EGFP* (*n*=3), *ΔNIC* (*n*=3) and *STOP* (*n*=3) mice, detected using EGFP fluorescence. Graph shows mean ± S.E.M. and genotypes were compared using t-tests: **** denotes a *P* value <0.0001. 'au' = arbitrary units. **c,** Phenotypic scoring of STOP mice (*n*=22) compared to published *Mecp2*-null data (*n*=12)[15]. Graph shows mean scores ± S.E.M. **d,** Kaplan-Meier plot showing survival of STOP mice (*n*=14) compared to *Mecp2*-null data (*n*=24)[15].

Extended Data Figure 9: Successful activation of *ΔNIC* in Tamoxifen-injected *STOP CreER^T* mice led to symptom reversal

a, Southern blot analysis of genomic DNA to determine the level of recombination mediated by CreER^T in Tamoxifen-injected (+Tmx) *STOP CreER^T* animals. *WT*, *WT CreER^T*, *ΔNIC* and *STOP* samples, with or without Tamoxifen injection, were included as controls. (Bsu36I digestion, see restriction map in Extended Data Fig. 5.) **b,** Protein levels in Tamoxifen-injected *STOP CreER^T* animals were determined using western blotting (upper, *n*=5) and flow cytometry (lower, *n*=3). Constitutively expressing *ΔNIC* mice (*n*=3) were used as a comparator. Graphs show mean values ± S.E.M. (quantification by western blotting is shown normalised to ΔNIC). Genotypes were compared using t-tests: western blotting *P*=0.434; flow cytometry 'All' nuclei *P*=0.128 and 'Neuronal' nuclei * *P*=0.016. 'au' = arbitrary units. For gel source data, see Supplementary Information. **c,** Heatmap of the phenotypic scores of the Tamoxifen-injected *STOP CreER^T* (upper; *n*=9) and *STOP* (lower; *n*=9 until 8

weeks of age, see survival plot in Fig. 4c) animals (see Fig. 4b), divided into the six categories. The plot is shaded according to the mean score for each category.

Extended Data Figure 10: Virus-encoded ΔNIC is expressed in brain and does not have adverse consequences in WT mice

a, b, Representative confocal images from thalamus and brainstem of scAAV-injected *Mecp2*-null (**a**) and *WT* (**b**) mice; scale bars indicate 20 µm visualised using an antibody against the Myc epitope (red) and the neuronal marker NeuN (green). Nuclei are stained with DAPI (blue). Graphs show transduction efficiency (mean ± SEM) in different brain regions (*n*=3 mice per genotype, 27 fields from each brain region). **c,** Phenotypic scoring (mean ± SEM) of scAAV-injected mice from 5-30 weeks: *WT* + vehicle (*n*=15), *Mecp2*-null + vehicle (*n*=20) and *WT* + *hΔNIC* (*n*=14). **d,** Kaplan-Meier plot showing survival of the cohort shown in panel **c**. One *WT* + *hΔNIC* animal was culled due to injuries at 28 weeks of age (shown by a tick). An arrow indicates the timing of the viral injection.

Full Methods

Nomenclature

According to convention, all amino acid numbers given refer to the e2 isoform. Numbers refer to homologous amino acids in human (NCBI accession P51608) and mouse (NCBI accession Q9Z2D6) until residue 385 where there is a two amino acid insertion in the human protein.

Mutation analysis

Mutational data was collected as described previously[5]: RTT-causing missense mutations were extracted from the RettBASE dataset[31]; and polymorphisms identified in males in the general population were extracted from the Exome Aggregation Consortium (ExAC) database[37].

Design of the truncated MeCP2 proteins

The MBD and NID were defined as residues 72-173 and 272-312, respectively. All three constructs retain the extreme N-terminal sequences encoded by exons 1 and 2 - present in isoforms e1 and e2, respectively. They also include the first three amino acids of exons 3 (EEK) to preserve the splice acceptor site. The intervening region (I) was replaced in ΔNIC by the NLS of SV40 preceded by a flexible linker. The sequence of the NLS is PKKKRKV (DNA sequence: CCCAAGAAAAAGCGGAAGGTG) and of the linker is GSSGSSG (DNA sequence: GGATCCAGTGGCAGCTCTGGG). All three proteins were C-terminally tagged with EGFP connected by a linker. To be consistent with a previous study tagging full-length MeCP2[15], the linker sequence CKDPPVAT (DNA sequence: TGTAAGGATCCACCGGTCGCCACC) was used to connect the C-terminus of ΔN to EGFP. To connect the NID to the EGFP tag in ΔNC and ΔNIC, the flexible GSSGSSG linker was used instead (DNA

sequence: GGGAGCTCCGGCAGTTCTGGA). For expression in cultured cells, cDNA sequences encoding e2 isoforms of the MeCP2 deletion series were synthesised (GeneArt, Thermo Fisher Scientific) and cloned into the pEGFPN1 vector (Clontech) using XhoI and NotI restriction sites (NEB). Point mutations (R111G and R306C) were inserted into the WT-EGFP plasmid using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Primer sequences for R111G: Forward TGGACACGAAAGCTTAAACAAGGGAAGTCTGGCC and Reverse GGCCAGACTTCCCTTGTTTAAGCTTTCGTGTCCA; and R306C: Forward CTCCCGGGTCTTGCACTTCTTGATGGGGA and Reverse TCCCCATCAAGAAGTGCAAGACCCGGGAG. For ES cell targeting, genomic sequences encoding exons 3 and 4 of the EGFP-tagged truncated proteins were synthesised (GeneArt, Thermo Fisher Scientific) and cloned into a previously used[19] targeting vector using MfeI restriction sites (NEB). This vector contains a Neomycin resistance gene followed by a transcriptional 'STOP' cassette flanked by *loxP* sites ('floxed') in intron 2.

Cell culture

HeLa and NIH-3T3 cells were grown in DMEM (Gibco) supplemented with 10% foetal bovine serum (FBS; Gibco) and 1% Penicillin-Streptomycin (Gibco). ES cells were grown in Glasgow MEM (Gibco) supplemented with 10% FBS (Gibco - batch tested), 1% Non-essential amino acids (Gibco), 1% Sodium Pyruvate (Gibco), 0.1% β-mercaptoethanol (Gibco) and 1000 units/ml LIF (ESGRO).

Immunoprecipitation

HeLa cells were transfected with pEGFPN1-MeCP2 plasmids using JetPEI (PolyPlus Transfection) and harvested after 24-48 hours. Nuclear extracts were prepared using Benzonase (Sigma E1014-25KU) and 150 mM NaCl, and MeCP2-EGFP complexes were captured using GFP-Trap_A beads (Chromotek) as described previously[5]. Proteins were analysed by western blotting using antibodies against GFP

(NEB #2956), NCoR (Bethyl A301-146A), HDAC3 (Sigma 3E11) and TBL1XR1 (Bethyl A300-408A), all at a dilution of 1:1000; followed by LI-COR secondary antibodies: IRDye® 800CW Donkey anti-Mouse (926-32212) and IRDye® 800CW Donkey anti-Rabbit (926-32213) or IRDye® 680LT Donkey anti-Rabbit (926-68023) at a dilution of 1:10,000.

MeCP2 localisation and TBL1X-mCherry recruitment assay

NIH-3T3 cells were seeded on coverslips in 6-well plates (25,000 cells per well) and transfected with 2 µg plasmid DNA (pEGFPN1-MeCP2 alone or pEGFPN1-MeCP2 and pmCherry-TBL1X[5]) using JetPEI (PolyPlus Transfection). After 48 hours, cells were fixed with 4% (w/v) paraformaldehyde, stained with DAPI (Sigma) and then mounted using ProLong Diamond (Life Technologies). Fixed cells were photographed on a confocal microscope (Leica SP5) using LAS AF software (Leica). The number of co-transfected cells with TBL1X-mCherry recruitment to heterochromatic foci was determined for each MeCP2 construct. In total, 113-125 cells per construct were counted (from three independent transfection experiments). This analysis was performed blind. The total proportion of cells with TBL1X-mCherry recruitment by each mutant MeCP2 protein was compared to WT using Fisher's exact tests.

Generation of knock-in mice

Targeting vectors were introduced into 129/Ola E14 TG2a ES cells by electroporation, and G418-resistant clones with correct targeting at the *Mecp2* locus were identified by PCR and Southern blot screening. CRISPR/Cas9 technology was used to increase the targeting efficiency of *ΔN* and *ΔNIC* lines: the guide RNA sequence (GGTTGTGACCCGCCATGGAT) was cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (a gift from Feng Zhang; Addgene plasmid #42230[38]), which was introduced into the ES cells with the targeting vectors. This introduced a double-strand cut in intron 2 of the wild-type gene

(at the site of the NeoSTOP cassette in the targeting vector). Mice were generated from ES cells as previously described[21]. The 'floxed' NeoSTOP cassette was removed *in vivo* by crossing chimaeras with homozygous females from the transgenic *CMV-Cre* deleter strain (JAX Stock #006054) on a C57BL/6J background. The *CMV-Cre* transgene was subsequently bred out. All mice used in this study were bred and maintained at the University of Edinburgh or Glasgow animal facilities under standard conditions and procedures were carried out by staff licensed by the UK Home Office and according with the Animal and Scientific Procedures Act 1986. Knock-in mice were caged with their wild-type littermates.

Biochemical characterisation of knock-in mice

For biochemical analysis, brains were harvested by snap-freezing in liquid nitrogen at 6-13 weeks of age, unless otherwise stated. Brains of hemizygous male mice were used for all analysis, unless otherwise stated. For Southern blot analysis, half brains were homogenised in 50 mM Tris HCl pH7.5, 100 mM NaCl, 5mM EDTA and treated with 0.4 mg/ml Proteinase K in 1% SDS at 55°C overnight. Samples were treated with 0.1 mg/ml RNAseA for 1-2 hours at 37°C, before phenol:chloroform extraction of genomic DNA. Genomic DNA was purified from ES cells using Puregene Core Kit A (Qiagen) according to manufacturer's instructions for cultured cells. Genomic DNA was digested with restriction enzymes (NEB), separated by agarose gel electrophoresis and transferred onto ZetaProbe membranes (BioRad). DNA probes homologous to either exon 4 or the end of the 3' homology arm were radioactively labelled with [α32]dCTP (Perkin Elmer) using the Prime-a-Gene Labeling System (Promega). Blots were probed overnight, washed, and exposed in Phosphorimager cassettes (GE Healthcare) before scanning on a Typhoon FLA 7000. Bands were quantified using ImageQuant software.

Protein levels in whole brain crude extracts were quantified using western blotting. Extracts were prepared as described previously[15], and blots were probed with antibodies against GFP (NEB #2956)

567 at a dilution of 1:1,000, followed by LI-COR secondary antibodies (listed above). Histone H3 (Abcam
568 ab1791) was used as a loading control (dilution 1:10,000). Levels were quantified using Image Studio
569 Lite Ver 4.0 software and compared using t-tests. WT-EGFP mice[15] were used as controls.
570
571 For flow cytometry analysis, fresh brains were harvested from 12 week-old animals and Dounce-
572 homogenised in 5 ml homogenisation buffer (320 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 10 mM
573 Tris HCl pH.7.8, 0.1 mM EDTA, 0.1% NP40, 0.1 mM PMSF, 14.3mM β-mercaptoethanol, protease
574 inhibitors (Roche)), and 5 ml of 50% OptiPrep gradient centrifugation medium (50% Optiprep (Sigma
575 D1556-250ML), 5 mM $CaCl_2$, 3mM $Mg(Ac)_2$, 10 mM Tris HCl pH7.8, 0.1M PMSF, 14.3mM β-
576 mercaptoethanol) was added. This was layered on top of 10 ml of 29% OptiPrep solution (v/v in H_2O,
577 diluted from 60% stock) in Ultra clear Beckman Coulter centrifuge tubes, and samples were
578 centrifuged at 10,100 xg for 30 mins, 4°C. Pelleted nuclei were resuspended in Resuspension buffer
579 (20% glycerol in DPBS (Gibco) with protease inhibitors (Roche)). For flow cytometry analysis, nuclei
580 were pelleted at 600 xg (5 mins, 4°C), washed in 1 ml PBTB (5% (w/v) BSA, 0.1% Triton X-100 in DPBS
581 with protease inhibitors (Roche)), and then resuspended in 250 µl PBTB. To stain for NeuN, NeuN
582 antibody (Millipore MAB377) was conjugated to Alexa Fluor 647 (APEX Antibody Labelling Kit,
583 Invitrogen A10475), added at a dilution of 1:125 and incubated under rotation for 45 mins at 4°C.
584 Flow cytometry (BD LSRFortessa SORP using FACSDIVA v8.0.1 software) was used to obtain the mean
585 EGFP fluorescence for the total nuclei (n=50,000 per sample) and the high NeuN (neuronal)
586 subpopulation (n>8,000 per sample). The protein levels of the novel mouse lines were compared to
587 *WT-EGFP* controls using t-tests. To compare protein levels in *WT-EGFP* mice to wild-type littermates,
588 nuclei were also stained with an MeCP2 antibody (Sigma M7443) conjugated to Alexa Fluor 568
589 (APEX Antibody Labelling Kit, Invitrogen A10494) at a dilution of 1:125.
590
591 To determine mRNA levels, RNA was purified and reverse transcribed from half brains (harvested at
592 11 weeks of age); and *Mecp2* and *Cyclophilin A* transcripts were analysed by qPCR using LightCycler

480 SW 1.5 software as previously described[15]. mRNA levels in *ΔNIC* mice were compared to *WT-EGFP* controls using a t-test.

Phenotypic characterisation of knock-in mice

Consistent with a previous study[15], mice were backcrossed for four generations to reach ~94% C57BL/6J before undergoing phenotypic characterisation. Two separate cohorts, each consisting of 10 mutant animals (11 for *ΔNC* Elevated Plus Maze) and 10 wild-type littermates, were produced for each novel knock-in line. One cohort was scored and weighed regularly from 4-52 weeks of age as previously described[19,20]. Survival was graphed using Kaplan-Meier plots. (A preliminary outbred [75% C57BL/6J] cohort of 7 *ΔNC* mice and 9 wild-type littermates was also analysed.) Previously published[15] data for *Mecp2*-null and R133C-EGFP (both backcrossed onto C57BL/6J) were included as comparators. The second backcrossed cohorts underwent behavioural analysis at 20 weeks of age (see [20] and [15] for detailed protocols). Tests were performed over a two-week period: Elevated Plus Maze on day 1, Open Field test on day 2, and Accelerating Rotarod test on days 6-9 (one day of training followed by three days of trials). All analysis was performed blind to genotype.

Statistical analysis

Growth curves were compared using repeated measures ANOVA (the animals that died within the experimental period – one wild-type in each *ΔNC* cohort and one *ΔNIC* in their cohort – were excluded from this analysis to enable a balanced design). Survival curves were compared using the Mantel-Cox test. For behavioural analysis, when all data fitted a normal distribution (Open Field centre time and distance travelled), genotypes were compared using t-tests (unpaired, two-tailed). If not (Elevated Plus Maze time in arms/centre and Accelerating Rotarod latency to fall), genotypes were compared using Kolmogorov-Smirnov tests. Change in performance over time in the Accelerating Rotarod test was determined using Friedman tests. All analysis was performed using GraphPad Prism 7 software.

Genetic activation of minimal MeCP2 (ΔNIC)

Transcriptionally silent minimal MeCP2 (ΔNIC) was activated in symptomatic null-like *'STOP'* mice following the procedure used in [19]. In short, the *ΔNIC Mecp2* allele was inactivated by the retention of the NeoSTOP cassette in intron 2 by mating chimaeras with wild-type females instead of *CMV-Cre* deleter mice. Resulting STOP/+ females were crossed with heterozygous *CreER^T* transgenic males (JAX Stock #004682) to produce males of four genotypes (87.5% C57BL/6J). A cohort consisting of all four genotypes *WT* (*n*=4), *WT CreER^T* (*n*=4), *STOP* (*n*=9) and *STOP CreER^T* (*n*=9), was scored and weighed weekly from 4 weeks of age. From 6 weeks (when *STOP* and *STOP CreER^T* mice displayed RTT-like symptoms), all individuals were given a series of Tamoxifen injections: two weekly followed by five daily, each at a dose of 100 μg/g body weight. Brain tissue from Tamoxifen-treated *STOP CreER^T* (n=8), WT (n=1) and *WT CreER^T* (n=1) animals was harvested at 28 weeks of age (after successful symptom reversal in *STOP CreER^T* mice) for biochemical analysis. Brain tissue from one Tamoxifen-treated *STOP* mouse (harvested at its humane end-point) was also included in the biochemical analysis (methods described above).

Vector delivery of minimal MeCP2 (ΔNIC)

The AAV vector expressing minimal MeCP2 (ΔNIC) was tested in *Mecp2*-null and *WT* mice maintained on a C57BL/6J background. Self-complementary AAV (scAAV) particles, comprising AAV2 ITR-flanked genomes packaged into AAV9 capsids, were generated at the UNC Gene Therapy Center Vector Core facility. Particles were produced as previously described[39] by transfection of HEK293 cells with helper plasmids (pXX6-80, pGSK2/9) and a plasmid containing the ITR-flanked construct in the presence of polyethyleneimine (Polysciences, Warrington, PA). For translational relevance, the ΔNIC-expressing construct utilised the equivalent human *MECP2_e1* coding sequence tagged with a small C-terminal Myc epitope to replace the EGFP tag used in knock-in experiments. The transgene was under the control of an endogenous *Mecp2* promoter fragment as previously described[25].

Vector was formulated in high-salt PBS (containing 350 mM total NaCl) supplemented with 5% sorbitol. Virus (3 µl per site; dose = 1 x 10^{11} viral genome per mouse) was injected bilaterally into the neuropil of unanaesthetised P1/2 males, as described previously[40]. Control injections used the same diluent lacking vector ('vehicle control'). The injected pups were returned to the home cage and assessed weekly from 5 weeks of age, as described above (performed blind to genotype). Cohorts were as follows: *WT* + vehicle (*n*=15); *Mecp2*-null + vehicle (*n*=20; 19 of which were scored as one reached its humane end-point early); *WT* + *hΔNIC* (*n*=14); and *Mecp2*-null + *hΔNIC* (*n*=17).

To validate the expression of virally-delivered *hΔNIC*, mice were deeply anesthetized with pentobarbitone (50 mg, intraperitoneally) and transcardially perfused with 4% paraformaldehyde (0.1 M PBS). A vibrating microtome (Leica VT1200) was used to obtain 70 µm sections of the brain. Sections were washed three times in 0.3 M PBS followed by blocking using 5% normal goat serum in 0.3 M PBS with 0.3% Triton X-100 (PBST) for 1 hour at room temperature. Samples then were incubated for 48 hours on a shaker at 4°C with the following primary antibodies against: Myc (Abcam ab9106, 1:500 dilution) and NeuN (Abcam 104224; 1:500). Samples were washed three times with 0.3 M PBST and incubated in secondary antibodies (Alexa Fluor 594 goat anti- rabbit (Abcam 150080; 1:500) and Alexa Fluor 647 goat anti-mouse (Stratech scientific LTD, 115-605-003JIR; 1:500) at 4°C overnight. Finally, sections were incubated with DAPI (Sigma; 1:1,000) for 30 minutes at room temperature before mounting with Fluoroshield with DAPI (Sigma, F6057). Z-series at 0.6–1.3 µm intervals were captured using a Zeiss LSM710 or Zeiss Axiovert LSM510 laser confocal microscope (40x objective). To estimate transduction efficiency, the ratio of Myc-positive nuclei to DAPI-stained nuclei was calculated form random sections of hippocampus (CA1), layer 5 of primary motor cortex, thalamus, hypothalamus, and brainstem (*n* = 3 mice per genotype, 27 fields from each brain region). *Mecp2*-null + *hΔNIC* were analysed after reaching their humane endpoints (aged 33, 35 and 36 weeks). *WT* + *hΔNIC* mice were harvested for analysis at 4 weeks of age.

All data are available from the authors on reasonable request. Source data underlying all graphs and full scans of all western and Southern blots are included.

Supplemental References

31. RettBase: Rett Syndrome Variation Database. at <http://mecp2.chw.edu.au/>

32. Tao, J. *et al.* Phosphorylation of MeCP2 at Serine 80 regulates its chromatin association and neurological function. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4882–7 (2009).

33. Ebert, D. H. *et al.* Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* **499,** 341–5 (2013).

34. Ho, K. L. *et al.* MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* **29,** 525–31 (2008).

35. PHD Secondary structure prediction method. at <https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html>

36. Lyst, M. J., Connelly, J., Merusi, C. & Bird, A. Sequence-specific DNA binding by AT-hook motifs in MeCP2. *FEBS Lett.* **590,** 2927–2933 (2016).

37. Exome Aggregation Consortium (ExAC), Cambridge, MA. at <http://exac.broadinstitute.org>

38. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/VCas Systems. *Science (80-.).* **339,** 819–823 (2013).

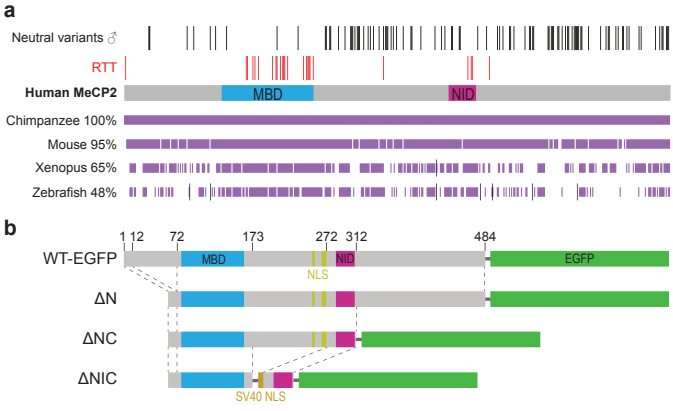
39. Clément, N. & Grieger, J. C. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol. Ther. Methods Clin. Dev.* **3,** 16002 (2016).

40. Gadalla, K. K. E. *et al.* Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice. *Mol. Ther.* **21,** 18–30 (2013).

698

699

700


a
Neutral variants ♂
RTT
Human MeCP2
MBD
NID
Chimpanzee 100%
Mouse 95%
Xenopus 65%
Zebrafish 48%

b
1 12
72
173
272 312
484
WT-EGFP
MBD
NID
NLS
EGFP
ΔN
ΔNC
ΔNIC
SV40 NLS

Figure 1

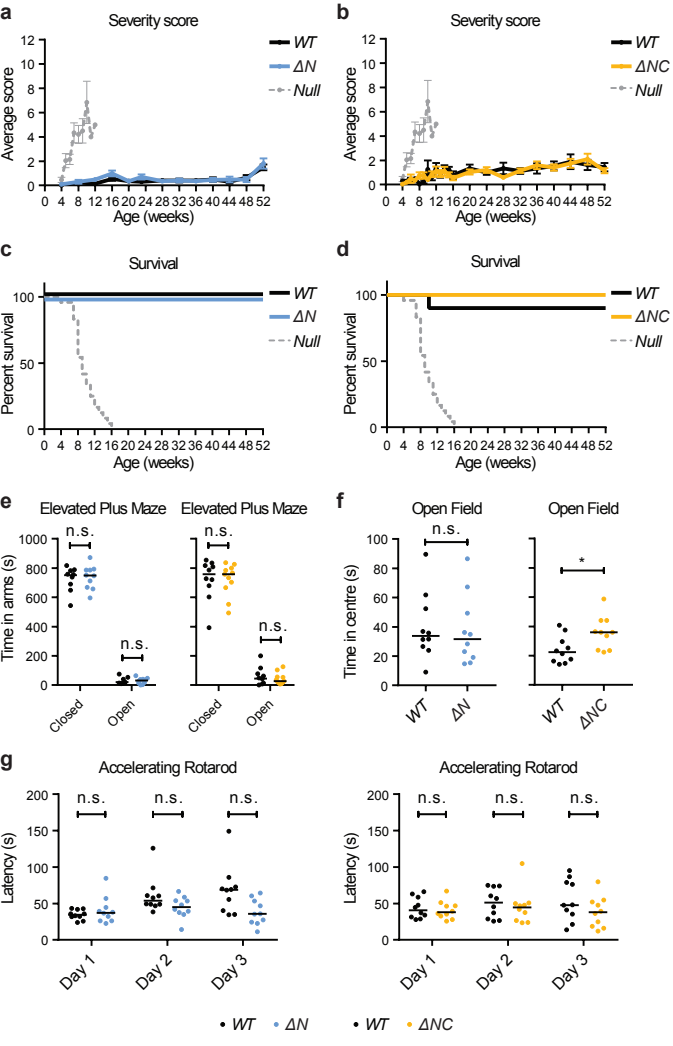

a
Severity score
Average score
Age (weeks)
WT
ΔN
Null
b
Severity score
Average score
Age (weeks)
WT
ΔNC
Null
c
Survival
Percent survival
Age (weeks)
WT
ΔN
Null
d
Survival
Percent survival
Age (weeks)
WT
ΔNC
Null
e
Elevated Plus Maze
Elevated Plus Maze
Time in arms (s)
n.s.
n.s.
n.s.
n.s.
Closed
Open
Closed
Open
f
Open Field
Open Field
Time in centre (s)
n.s.
*
WT
ΔN
WT
ΔNC
g
Accelerating Rotarod
Accelerating Rotarod
Latency (s)
n.s.
n.s.
n.s.
n.s.
n.s.
n.s.
Day 1
Day 2
Day 3
Day 1
Day 2
Day 3
WT
ΔN
WT
ΔNC

Figure 2

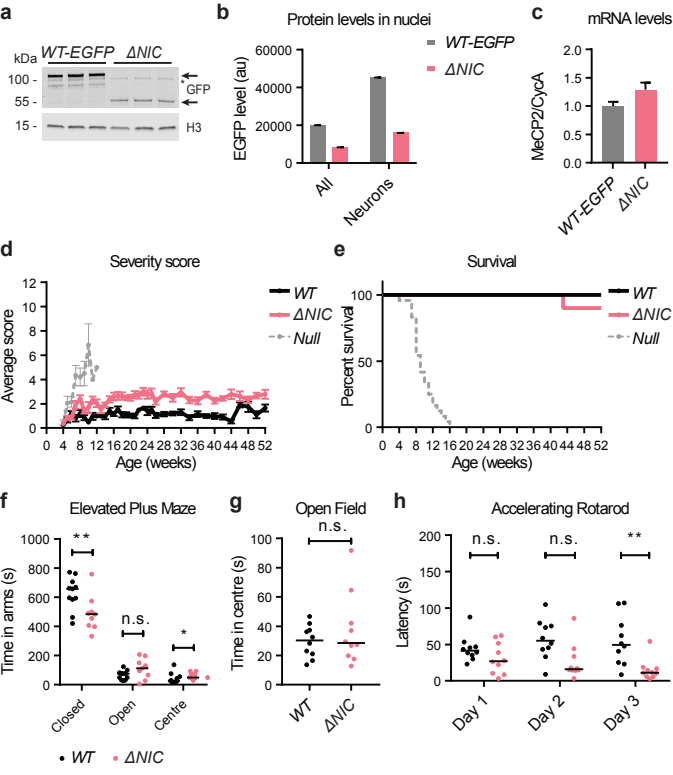

a
kDa
WT-EGFP
ΔNIC
100 -
55 -
15 -
GFP
H3
b
Protein levels in nuclei
WT-EGFP
ΔNIC
EGFP level (au)
60000
40000
20000
0
All
Neurons
c
mRNA levels
MeCP2/CycA
2.0
1.5
1.0
0.5
0.0
WT-EGFP
ΔNIC
d
Severity score
Average score
WT
ΔNIC
Null
Age (weeks)
e
Survival
Percent survival
100
50
0
WT
ΔNIC
Null
Age (weeks)
f
Elevated Plus Maze
Time in arms (s)
**
n.s.
*
Closed
Open
Centre
WT
ΔNIC
g
Open Field
Time in centre (s)
n.s.
WT
ΔNIC
h
Accelerating Rotarod
Latency (s)
n.s.
n.s.
**
Day 1
Day 2
Day 3

Figure 3

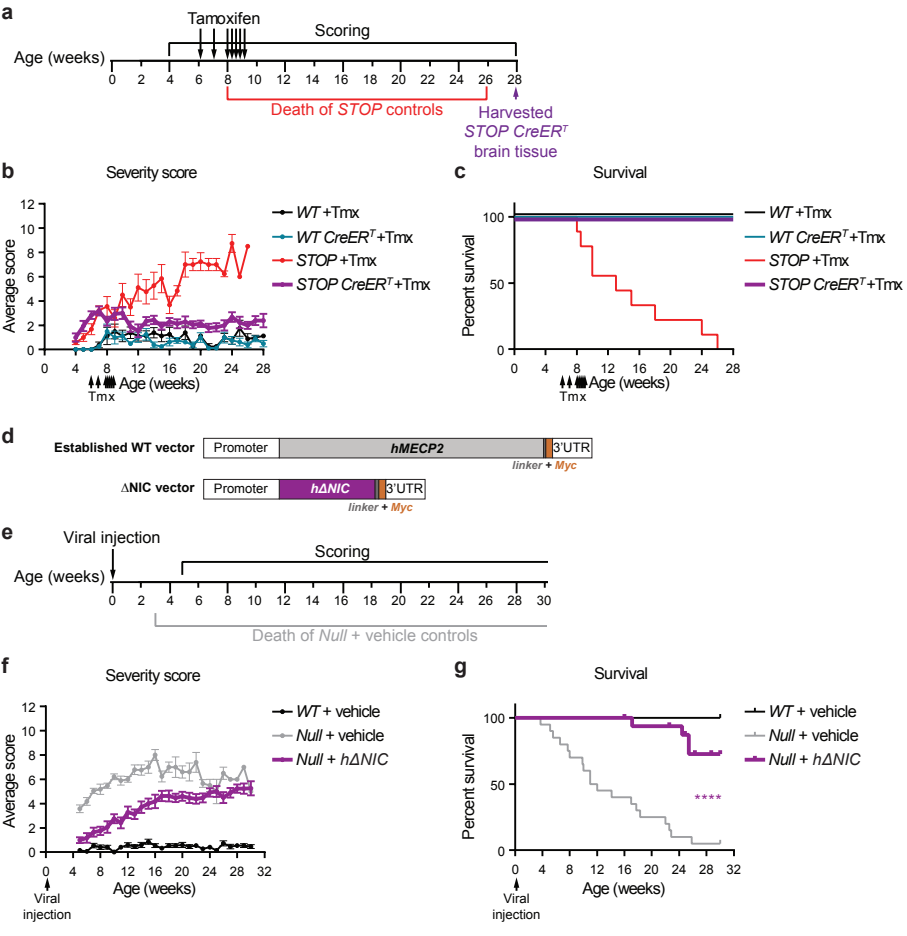

a
Tamoxifen
Scoring
Age (weeks)
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28
Death of STOP controls
Harvested
STOP CreER^T
brain tissue
b
Severity score
Average score
Age (weeks)
Tmx
WT +Tmx
WT CreER^T+Tmx
STOP +Tmx
STOP CreER^T+Tmx
c
Survival
Percent survival
Age (weeks)
Tmx
WT +Tmx
WT CreER^T+Tmx
STOP +Tmx
STOP CreER^T+Tmx
d
Established WT vector
Promoter
hMECP2
3'UTR
linker + Myc
ΔNIC vector
Promoter
hΔNIC
3'UTR
linker + Myc
e
Viral injection
Scoring
Age (weeks)
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Death of Null + vehicle controls
f
Severity score
Average score
Age (weeks)
Viral injection
WT + vehicle
Null + vehicle
Null + hΔNIC
g
Survival
Percent survival
****
Age (weeks)
Viral injection
WT + vehicle
Null + vehicle
Null + hΔNIC

Figure 4