Supporting Information

Bioorthogonal metabolic labelling of nascent RNA in neurons improves the sensitivity of transcriptome-wide profiling

Esmi L. Zajaczkowski[†], Qiong-Yi Zhao[†], Zong Hong Zhang[†], Xiang Li[†], Wei Wei[†], Paul R. Marshall[†], Laura J. Leighton[†], Sarah Nainar[‡], Chao Feng[‡], Robert C. Spitale[‡ ||]*, Timothy W. Bredy[†]*

[†]Cognitive Neuroepigenetics Laboratory, Queensland Brain Institute, The University of Queensland, Brisbane, QLD, 4072, Australia

[‡]Department of Pharmaceutical Sciences and [||]Department of Chemistry, University of California, Irvine, Irvine, California, 92697, United States

Table of Contents

1. Supplementary Figures

1.1 Supplementary Figure 1

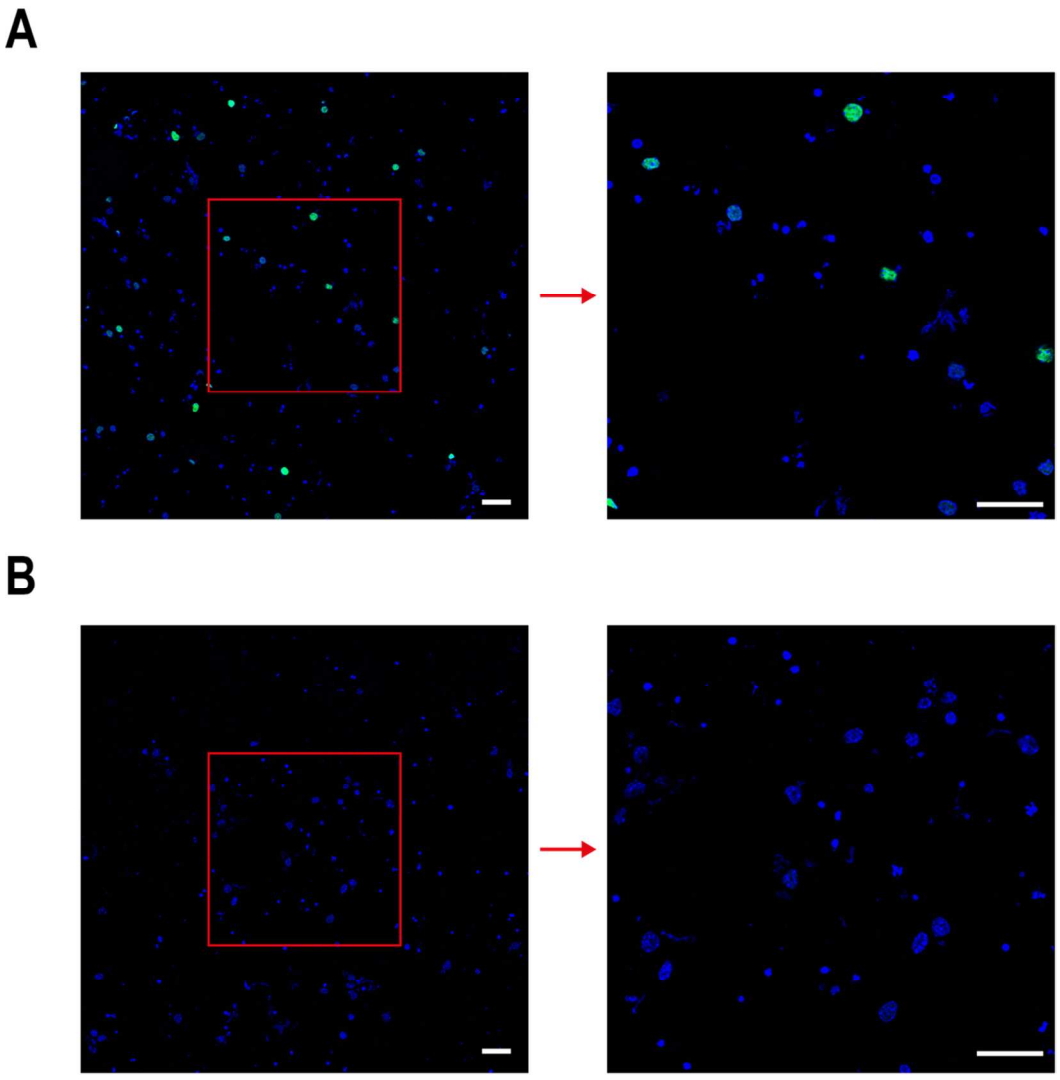

Figure S1: Visualisation of nascent RNA labelled with 5EUracil using an azide-AlexaFluor488 fluorescent dye.
A) Mouse primary cortical neurons (PCNs) treated with a synapsin I-driven UPRT lentivirus 7 days prior were fixed immediately following a 3 hour incubation with 5EUracil and KCl stimulation. Cells were then subjected to a CuAAC reaction (see Appendix 2) using 25 μM azide-AlexaFluor488 (green) and counterstained with 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI, blue). **B)** Same as A) except cells were treated with a control GFP lentivirus 7 days prior instead. All images were taken on a ZIS LSM 510 META confocal microscope using a 63x oil objective, scale bar 50 μm. Red boxes on left-hand side images outline the visual region displayed in the right-hand side image.

1.2 Supplementary Figure 2

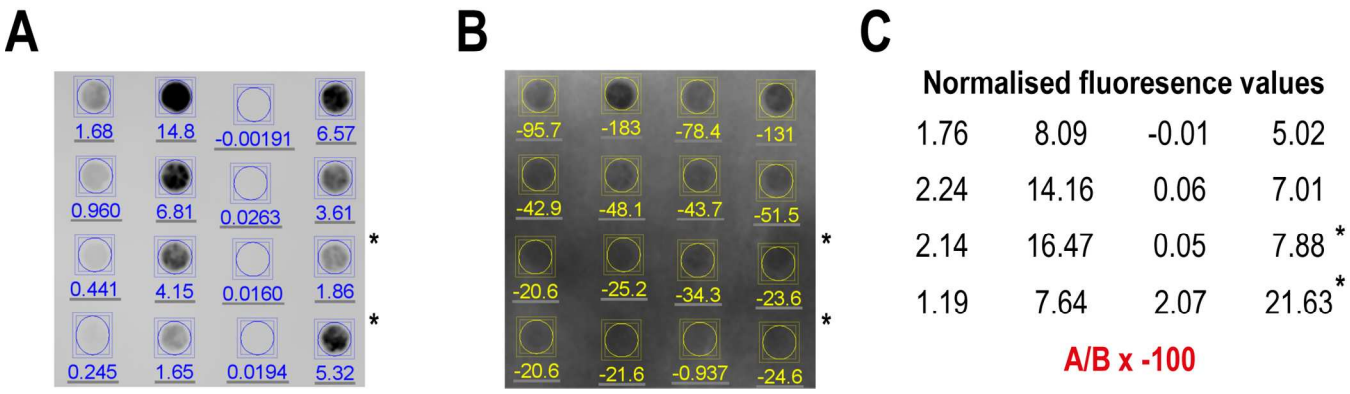

Figure S2: UPRT overexpression is an effective method for capturing nascent RNA labelled with 5EUracil. A) Data shown is a dot blot of RNA biotinylated via CuAAC (see section 1.5.1 and 1.6, serial dilutions starting at 400 ng) and labelled with streptavidin-IR800CW. Total cellular RNA was harvested following 3 hours of KCl stimulation and incubation with nucleic analogs 5EUracil or 5EUridine. The four experimental conditions illustrated are as follows: GFP lentivirus with 5EUracil (virus negative control), UPRT virus with 5EUracil, UPRT virus without 5EUracil (analog negative control), and no virus with 5EUridine (analog positive control). Dot blot image was obtained using the LI-COR Odyssey Fc imaging system. **B)**Same dot blot as shown in A) but instead shows a methylene blue stain, which is used to visualise the total amount of RNA present.**C)** A semi-quantitative measure of the amount of alkyne-labelled RNA was calculated by taking the fluorescent values shown in A) and normalising them to values shown in B) (and the multiplying by -100, formula in red). Note that samples with asterisks (*) were mistakenly blotted in the wrong place and should be swapped around.

1.3 Supplementary Figure 3

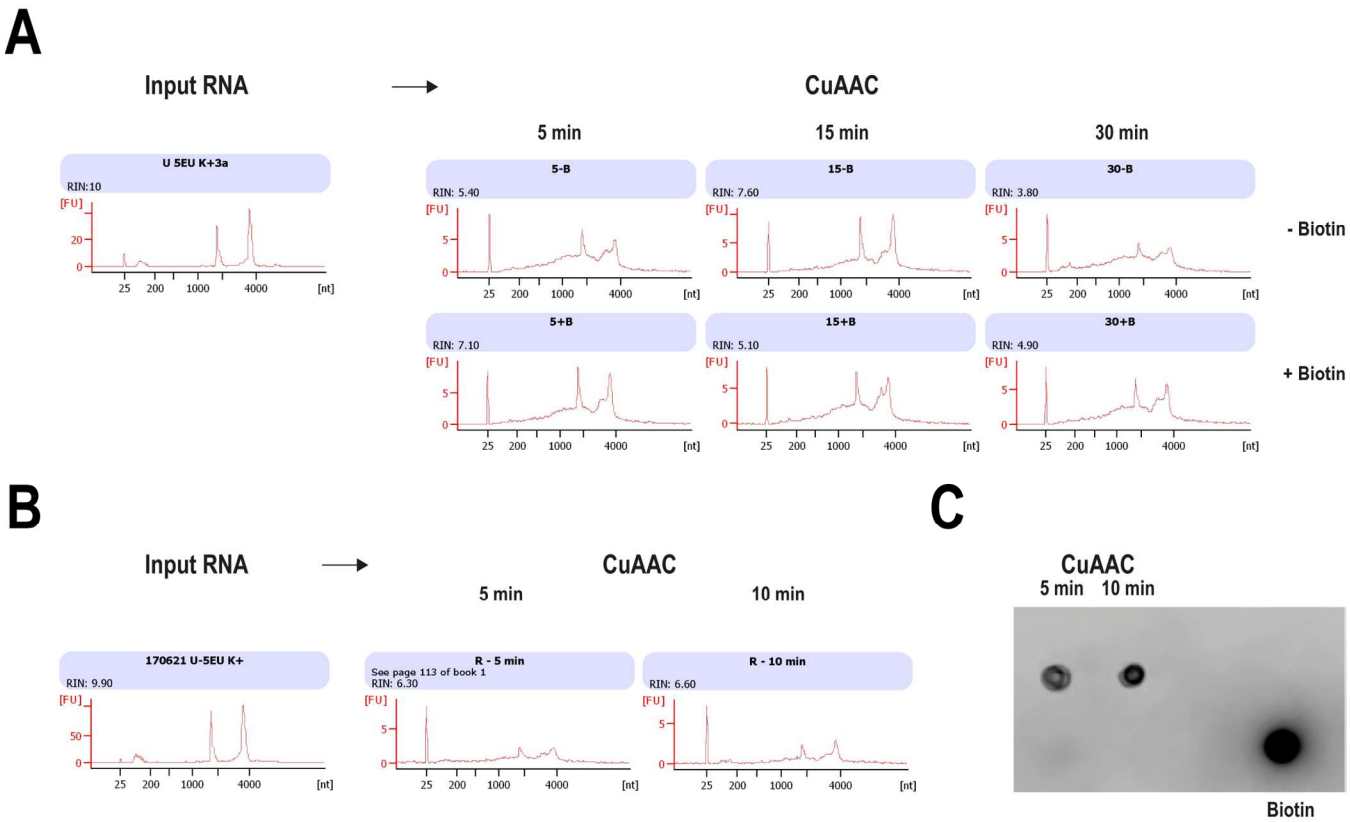

Figure S3: Optimisation of the CuAAC reaction for biotinylation of alkyne-labelled nascent RNA.A) Bioanalyser traces from sixCuAAC reactions that were performed in parallel to investigate three reaction times (5, 15, 30 min) and 2 conditions (with and without azide-biotin) as well as the pre-CuAAC input RNA. RNA quality drops significantly by 30 min.**B)**Bioanalyser traces of pre-CuAAC input RNA and two reaction times for CuAAC (5, 10 min), which have roughly the same RNA quality (RIN 6 – 7). **C)** Dot blot of post-CuAAC RNA from B). 10 min CuAAC reaction time appears to yield a slightly better biotinylation amount as well as being easier to technically replicate with less time variability.Image was obtained using the LI-COR Odyssey Fc imaging system.

1.4 Supplementary Figure 4

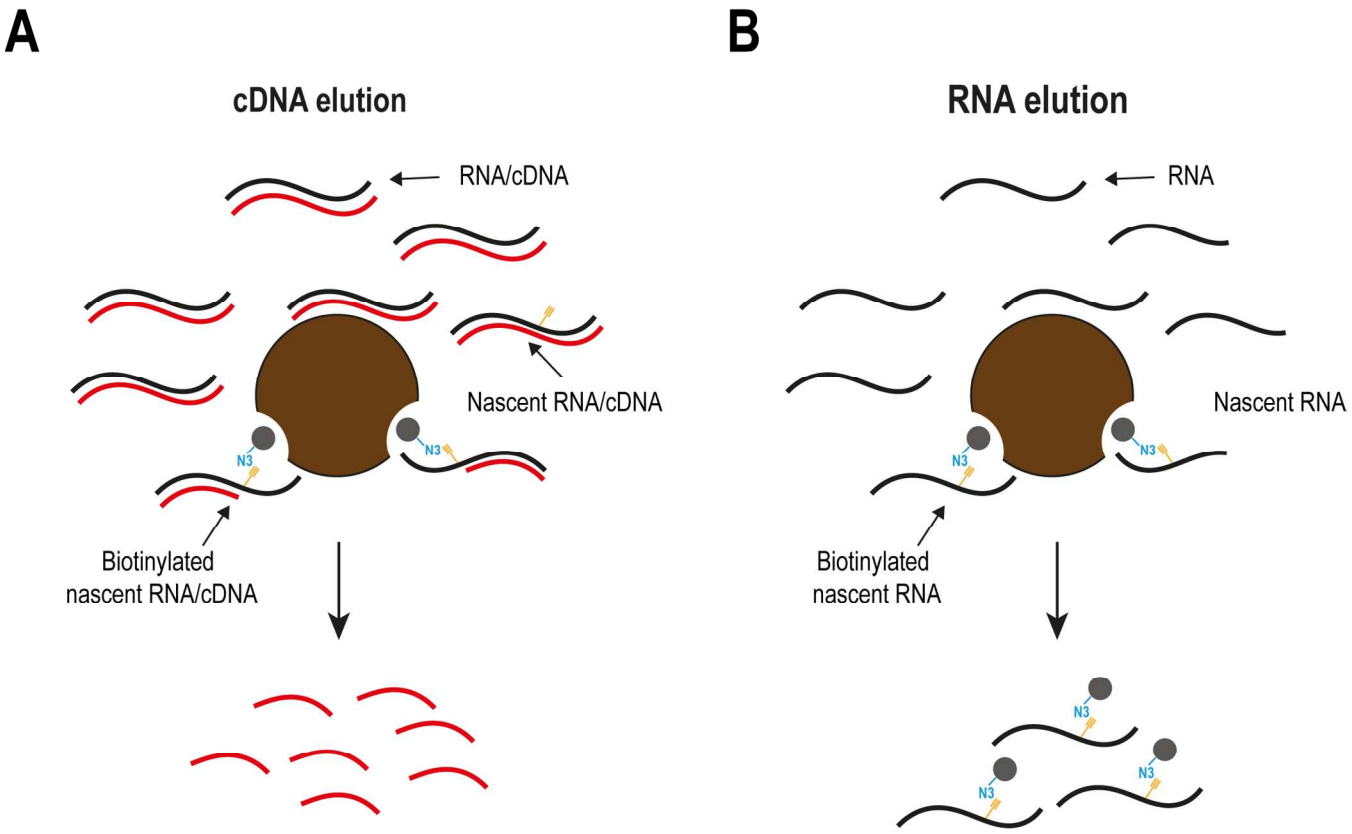

Figure S4: Illustrative diagrams for streptavidin enrichment of nascent RNA/cDNA hybrids and nascent RNA.
A) cDNA elution involves applying post-CuAAC RNA/cDNA hybrids to streptavidin beads and then degrading the RNA to allow the cDNA that was complementary to nascent RNA to be eluted. **B)** RNA elution involves directly binding post-CuAAC RNA to streptavidin beads followed by a heated chemical reaction to disrupt the biotin-streptavidin reaction and elute nascent RNA.

1.5 Supplementary Figure 5

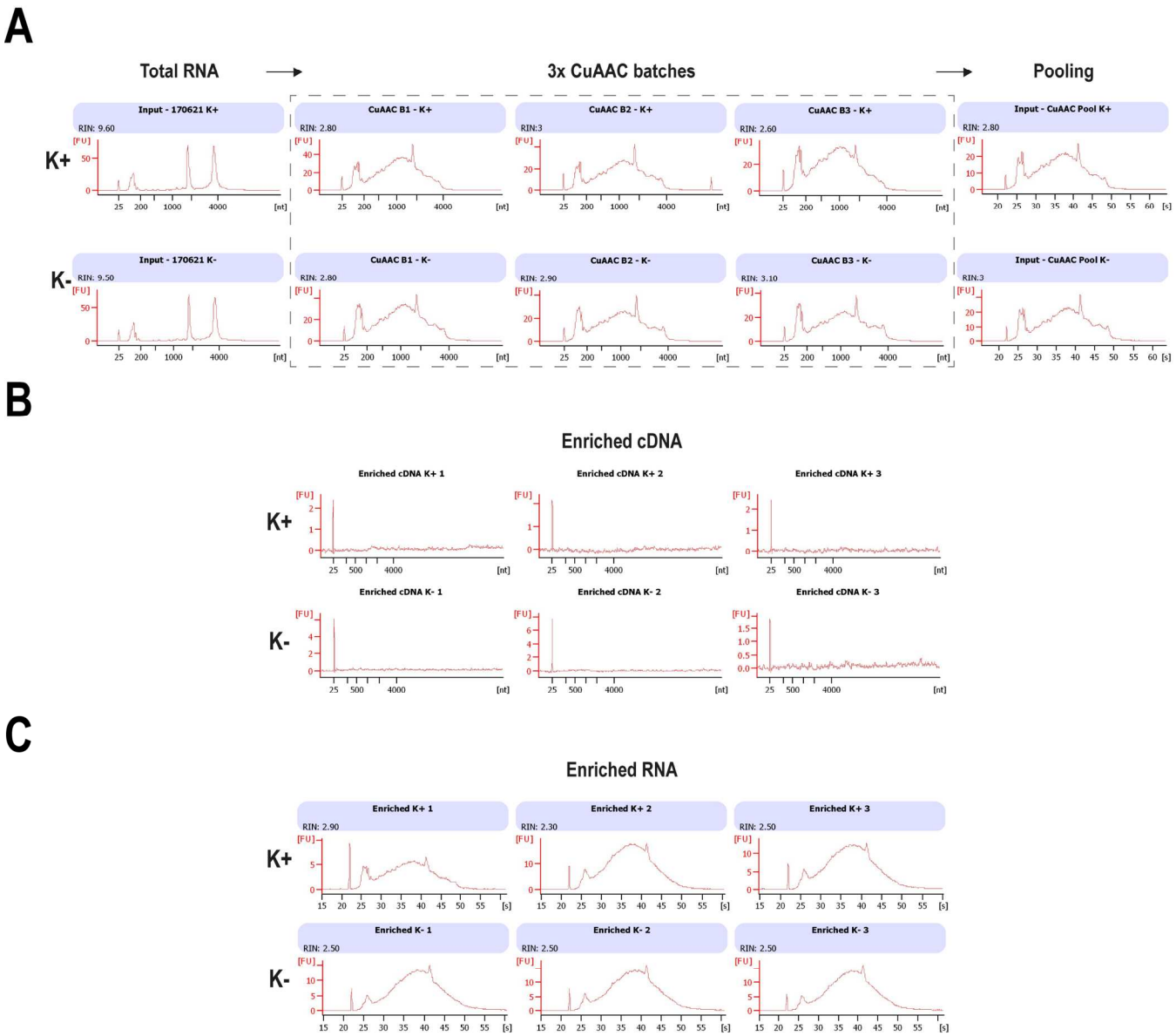

Figure S5: RNA elution is a better method forenriching nascent RNA transcripts.A) Biotinylated post-CuAAC RNA was generated in 3x batches each for non-activated (KCl-) and activated (KCl+) conditions and then pooled together.**B)** Enrichment of cDNA complementary to nascent RNA was performed in triplicate for each condition (KCl- vs KCl+) and run on the Agilent Bioanalyser (traces shown) using a HS DNA chip. Eluted cDNA was unquantifiable using both a Qubit assay (Invitrogen) and the Agilent Bioanalyser.**C)** Enrichment of nascent RNA was performed in triplicate for each condition (KCl- vs KCl+) and run on the Agilent Bioanalyser (traces shown) using a RNA Pico II chip. K: KCl.

1.6 Supplementary Figure 6

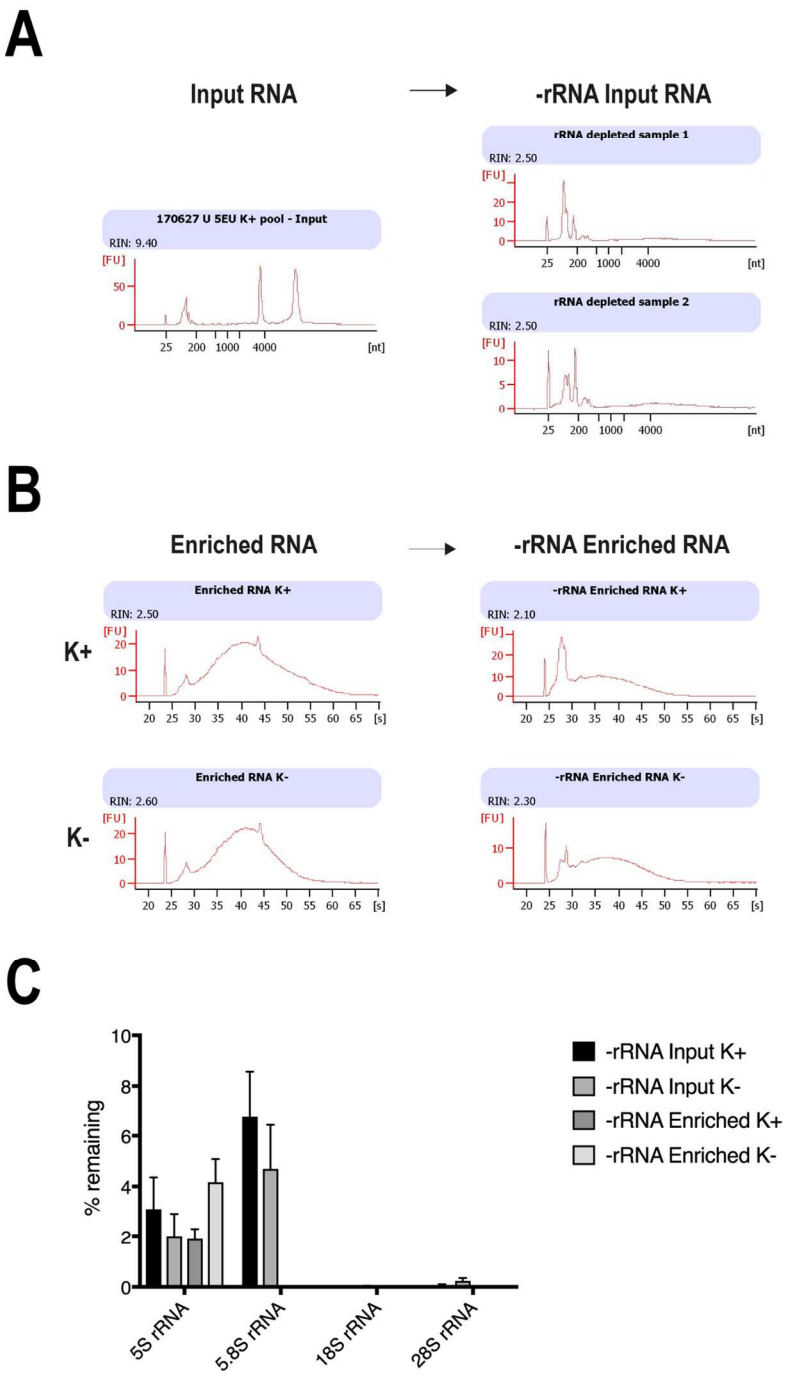

Figure S6: Validating ribosomal RNA(rRNA) removal on input and enriched samples.A) Bioanalyser traces showing the input sample before rRNA removal and after (performed twice), which indicates effective removal of larger 18S and 28S rRNA peaks. **B)** Bioanalyser traces for enriched (KCl- and KCl+) before and after rRNA removal seems to also have a reduced fraction of the larger fragments (presumably 18S and 28S rRNA). **C)** Bar chart showing the %remaining of 5S, 5.8S, 18S, 28S rRNA after rRNA removal. Data values were obtained from qPCR analysis (see **section 1.3.2**) by first normalising to housekeeping gene *Pgk1* and then calculating the fold change relative to the input sample before rRNA removal (and then multiplied by 100 to get %remaining).*n=4* per group.K: KCl.

1.7 Supplementary Figure 7

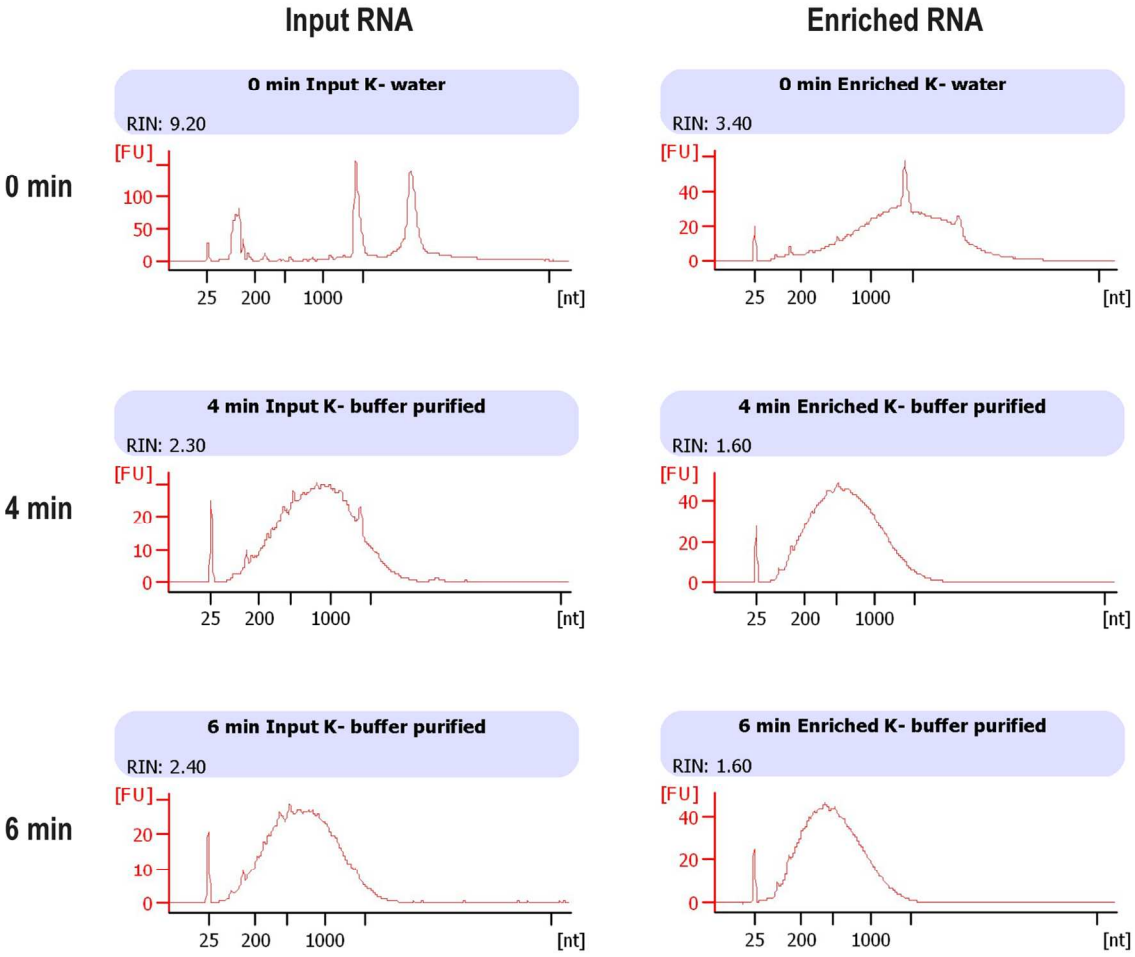

Figure S7: Optimisation of RNA fragmentation prior to double-stranded cDNA (dscDNA) synthesis. Input and enriched RNA samples were fragmented at 94°C in buffer and primer mix from the NEBNext® Ultra™ II RNA First Strand Synthesis Module (#E7771) for 4 and 6 min and then purified using the Zymo Research RNA Clean & Concentrator-5 kit (#R1014) before running on the Agilent Bioanalyser using a RNA Pico II chip.

1.8 Supplementary Figure 8

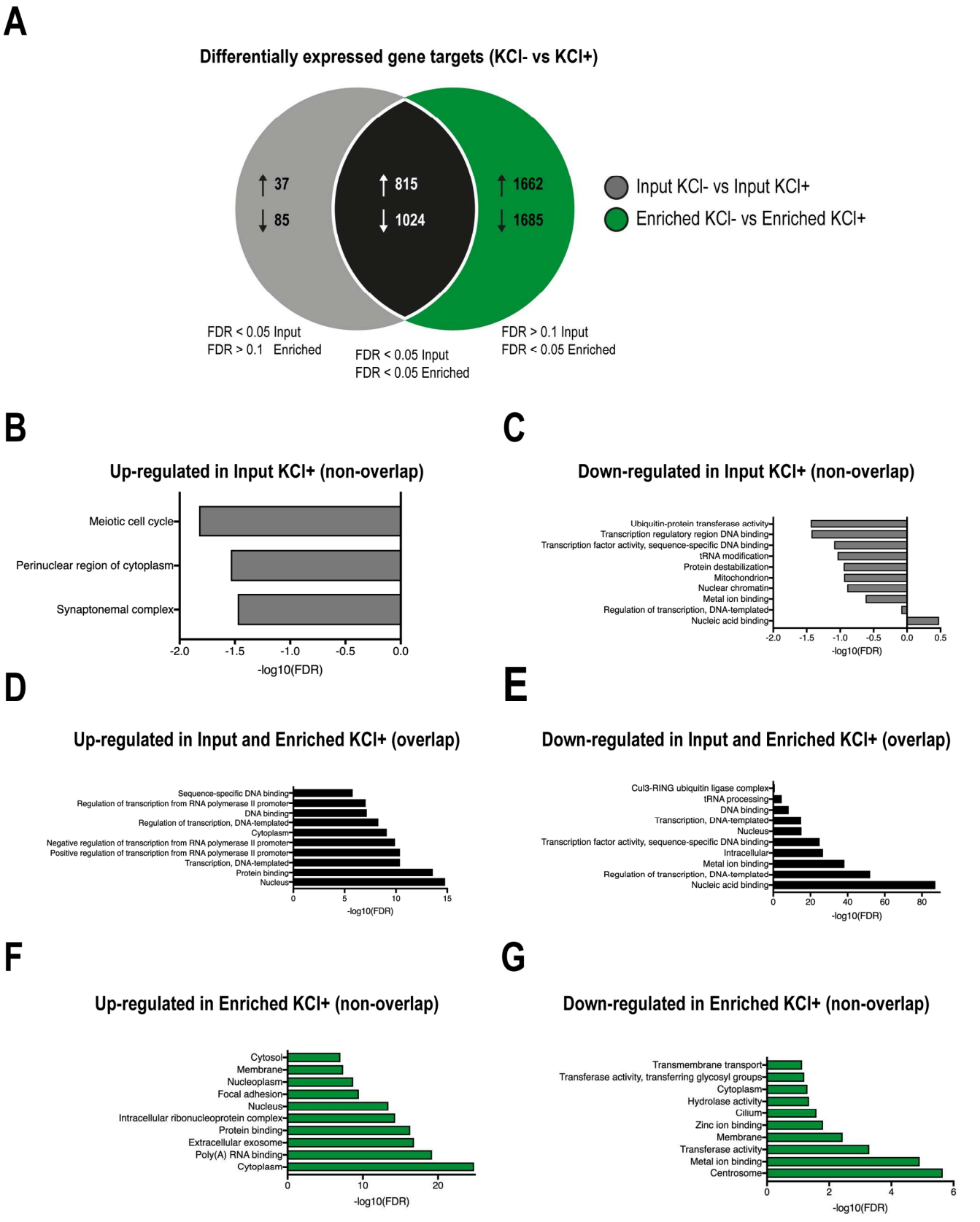

Figure S1: DAVID functional annotation of differentially expressed target genes.A) Venn diagram illustrating the overlap in target genes (up-regulated and down-regulated) between the comparisons “Input KCl+ vs Input KCl-”and “Enriched KCl+ vs

Enriched KCl-". **B) – G)** Each graph shows the top ten DAVID functional annotation Gene Ontology terms (smallest FDRs) with their respective –log10(FDR) values (smaller FDR → bigger and more positive value) for each set of genes illustrated inA).

1.9 Supplementary Figure 9

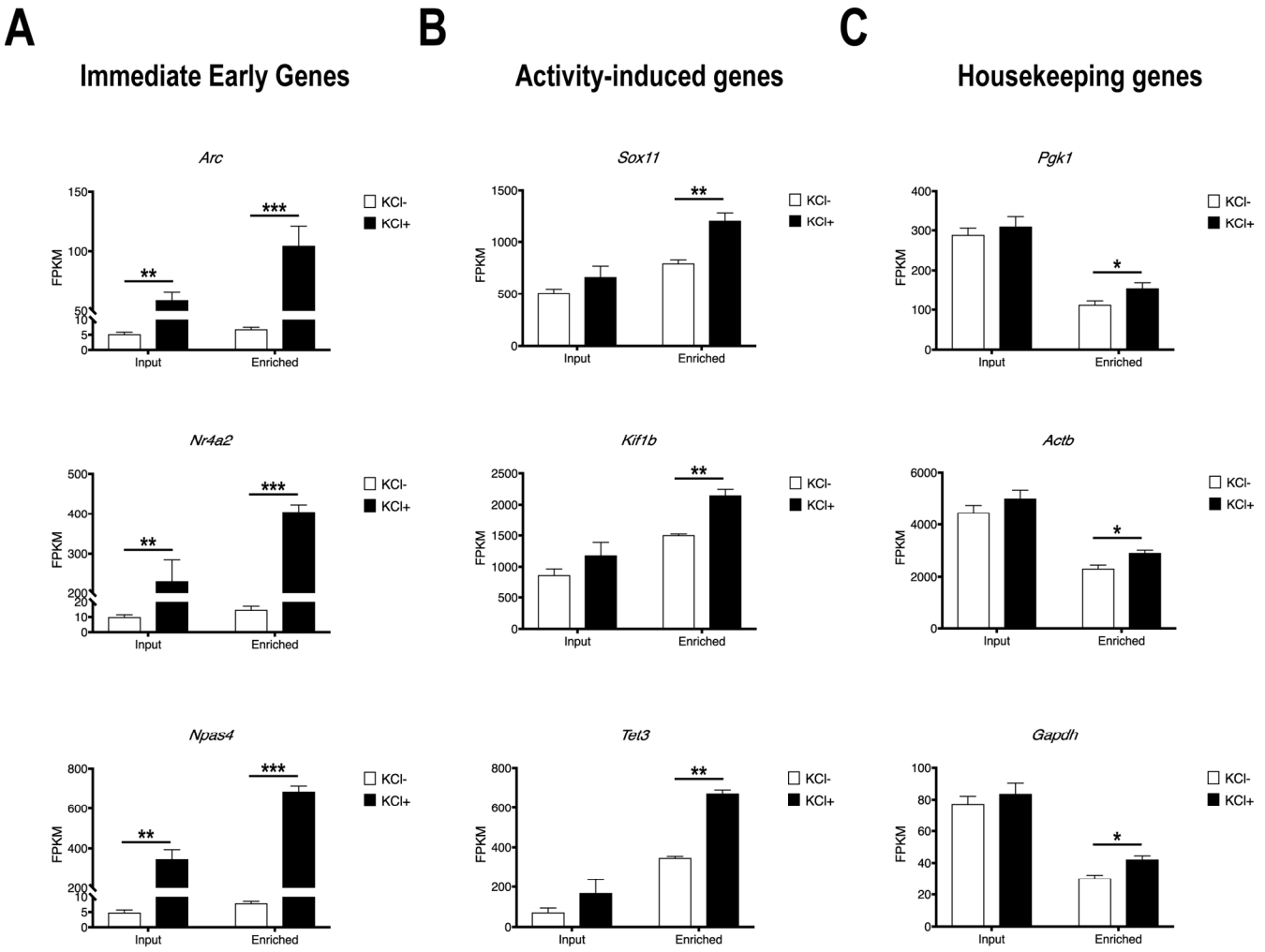

Figure S9: Nascent RNA sequencing improves sensitivity for detecting differences in activity-dependent gene expression. A) Nascent RNA sequencing is able to better detect Immediate Early Gene expression. **B)** Nascent RNA sequencing is able to detect activity-dependent differences in genes that are missed by standard total RNA-seq. **C)** Conventional housekeeping genes show a lower abundance of fragments after enrichment of nascent RNA (enriched) compared to total RNA-seq (input) but also exhibit activity-dependent differences in gene expression. *FDR<0.05, **FDR<0.001, ***FDR<0.000000001. FPKM: Fragments per Kilobase Million; FDR: False Discovery Rate; KCl: potassium chloride.

1.10 Supplementary Figure 10

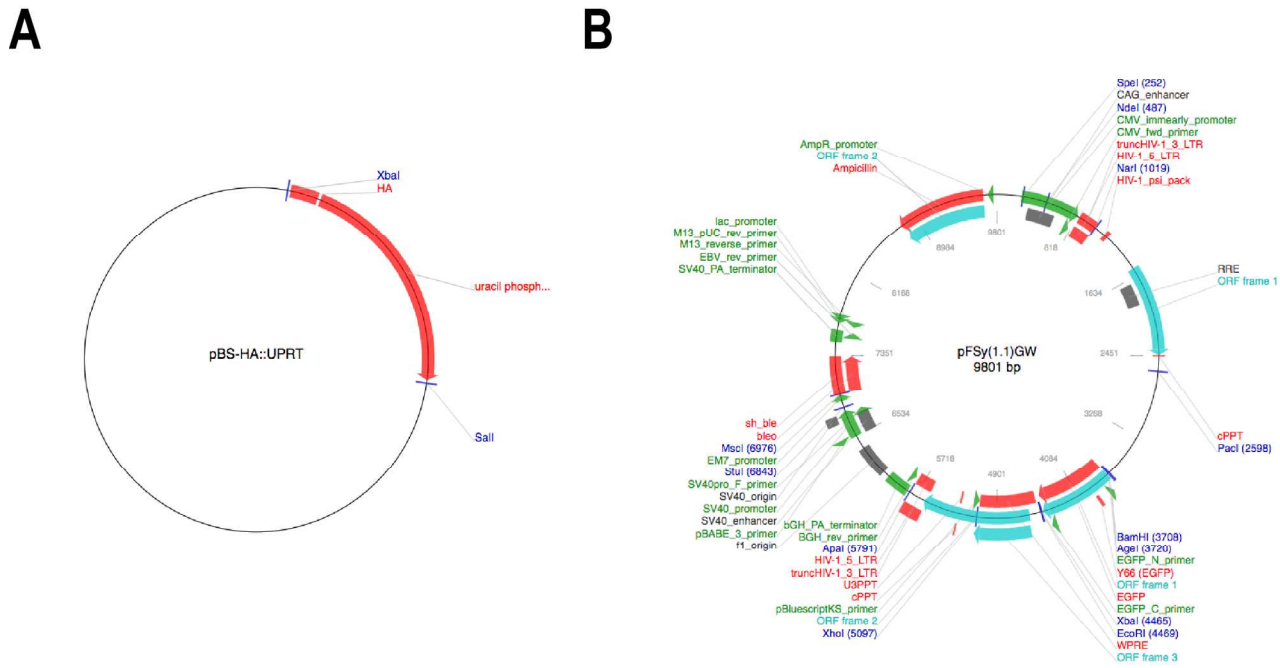

Figure S10: Components used to generate the transfer vector pFSy(1.1)GW containing the HA::UPRT insert.A) The pBS-HA::UPRT plasmid (kindly deposited by Mike Cleary & Chris Doe[2] into Addgene, plasmid #47110) was used as a template for PCR amplification of the HA::UPRT sequence. **B)** The original pFSy(1.1)GW vector (kindly deposited by Pavel Osten[1] into Addgene, plasmid #27232) before the HA::UPRT sequence replaced the EGFP sequence located in-between the BamHI and XBaI cutting sites. The original unmodified vector was packaged as a control. Images displayed are sourced from the Addgene website (https://www.addgene.org/).

2. Supplementary Tables

2.1 Supplementary Table 1

Table 1: Single-indexed i7 adapter sequences for each sample (*n = 4* biological replicates per group) for a 2x lanes on a HiSeq 4000 2x150 run.

	Sample code	Sample name	Single-indexed i7 adapter sequence	Kapa HyperPrep Kit Adapters Set A
Lane 1	A1	-rRNA Input K+ 1	ACAGTG	5
	A2	-rRNA Input K+ 2	GCCAAT	6
	B1	-rRNA Input K- 1	CTTGTA	12
	B2	-rRNA Input K- 2	GTGAAA	19
	C1	-rRNA Enriched K+ 1	CGATGT	2
	C2	-rRNA Enriched K+ 2	TGACCA	4
	D1	-rRNA Enriched K- 1	CAGATC	7
	D2	-rRNA Enriched K- 2	CCGTCC	16
Lane 2	A3	-rRNA Input K+ 3	CGATGT	2
	A4	-rRNA Input K+ 4	TGACCA	4
	B3	-rRNA Input K- 3	CAGATC	7
	B4	-rRNA Input K- 4	CCGTCC	16
	C3	-rRNA Enriched K+ 3	ACAGTG	5
	C4	-rRNA Enriched K+ 4	GCCAAT	6
	D3	-rRNA Enriched K- 3	CTTGTA	12
	D4	-rRNA Enriched K- 4	GTGAAA	19
Naming scheme	**A**	-rRNA Input K+	4x biological replicates	A1, A2, A3, A4
	B	-rRNA Input K-	4x biological replicates	B1, B2, B3, B4
	C	-rRNA Enriched K+	4x biological replicates	C1, C2, C3, C4
	D	-rRNA Enriched K-	4x biological replicates	D1, D2, D3, D4

2.2 Supplementary Table 2

Table 2: Primers for generating the synapsin I-driven UPRT overexpression lentivirus. Two rounds of PCR amplification were used to amplify and add XbaI and BamHI restriction sites to the HA::UPRT fragment obtained from the pBS-HA::UPRT plasmid (Addgene #47110) prior to digestion and cloning into thepFsy(1.1)GW vector backbone (Addgene #27232), replacing the original EGFP sequence.

	Primers	Notes
Round 1: + XbaI	Forward: 5'- ATGTACCCCTACGATGT-3' Reverse: 5'-AAA TCTAGACTACATGGTTCCAAAGT-3'	XbaI restriction site: TCTAGA
Round 2: + Kozak + BamHI	Forward: 5'-AAA GGATCC**GCCGCCACC**ATGTACCCCTACGATGT-3' Reverse: 5'-AAA TCTAGACTACATGGTTCCAAAGT-3'	BamHI restriction site: GGATCC Kozak sequence: **GCCGCCACC** XbaI restriction site: TCTAGA

HA::UPRT coding sequence

ATGTACCCCTACGATGTGCCCGATTACGCCACTAGTGCTAGCAGATCTGCGGCCAACATGGCGCA
GGTCCCAGCGAGCGGAAAGCTCCTTGTCGATCCCCGATATTCGACAAACGACCAGGAAGAAAGC
ATTCTCCAGGACATCATCACGAGGTTTCCCAATGTGGTGCTCATGAAGCAGACGGCTCAGCTTCG
AGCGATGATGACCATCATTCGTGATAAAGAAACACCGAAGGAAGAATTCGTCTTCTACGCCGACC
GCCTGATTCGCCTCCTCATCGAAGAAGCTTTGAACGAACTGCCGTTCGAAAAGAAGGAAGTGACA
ACCCCTCTGGATGTGTCATACCATGGAGTTTCCTTCTATTCCAAGATCTGTGGCGTCTCGATTGTG
AGAGCTGGCGAGTCGATGGAAAGCGGCTTGCGGGCAGTTTGCCGCGGCTGCCGCATCGGGAAA
ATCCTCATCCAGAGAGACGAAACAACTGCGGAGCCTAAGCTGATCTACGAGAAGCTGCCTGCCGA
CATTAGAGATCGCTGGGTGATGCTGCTAGATCCGATGTGCGCGACGGCGGGCAGTGTGTGCAAA
GCGATCGAGGTCCTCCTGAGGCTCGGCGTGAAGGAAGAGAGAATCATTTTCGTCAACATTTTGGC
TGCTCCCCAAGGCATTGAACGTGTTTTCAAGGAATACCCGAAAGTCCGCATGGTCACTGCTGCTG
TTGACATCTGCCTGAACTCGAGGTACTACATCGTCCCCGGCATTGGTGATTTCGGTGACCGGTAC
TTTGGAACCATGTAG

2.3 Supplementary Table 3

Table 3: qPCR primers used during the study to investigate the effect of synapsin I-driven UPRT lentivirus on UPRT expression within mouse primary cortical neurons.

Target	Primer	Primer sequence (5'-3')
Pgk1	qPGK F	TGCACGCTTCAAAAGCGCACG
	qPGK R	AAGTCCACCCTCATCACGACCC
Uprt	qUPRT F	GATTGTGAGAGCTGGCGAGT
	qUPRT R	GCAGGCAGCTTCTCGTAGAT

2.4 Supplementary Table 4

Table 4: qPCR primers used to validate ribosomal RNA removal.

Target	Primer	Primer sequence (5'-3')
Pgk1	qPGK F	TGCACGCTTCAAAAGCGCACG
	qPGK R	AAGTCCACCCTCATCACGACCC
*5S rRNA**	q5SrRNA F	CATACCACCCTGAACGCG
	q5SrRNA R	CTACAGCACCCGGTATTCCC
5.8S rRNA	q5.8SrRNA F	GACTCTTAGCGGTGGATCAC
	q5.8SrRNA R	GCAAGTGCGTTCGAAGTGT
18S rRNA	q18SrRNA F	CTGGATACCGCAGCTAGGAA
	q18SrRNA R	GAATTTCACCTCTAGCGGCG
28S rRNA	q28SrRNA F	AAGCGTTGGATTGTTCACCC
	q28SrRNA R	TCCTCAGCCAAGCACATACA

* 5S rRNA primer pair is non-specific; it also picks up expression of *Tmem38b*, which encodes a mouse transmembrane protein.

2.5 Supplementary Table 5

Table 5: Total and mappable reads for samples sequenced using a 2x150 run on a HiSeq 4000.

Sample ID	Sample name	Total reads	Mappable reads	Ratio (%)
A1	-rRNA Input K+ 1	99,268,860	86,422,585	87.06
A2	-rRNA Input K+ 2	116,477,742	100,426,648	86.22
A3	-rRNA Input K+ 3	101,906,268	87,531,669	85.89
A4	-rRNA Input K+ 4	91,429,842	67,865,072	74.23
B1	-rRNA Input K- 1	104,976,440	91,819,904	87.47
B2	-rRNA Input K- 2	115,971,574	94,061,984	81.11
B3	-rRNA Input K- 3	80,656,104	69,732,454	86.46
B4	-rRNA Input K- 4	110,339,894	90,770,763	82.26
C1	-rRNA Enriched K+ 1	116,782,840	96,929,752	83.00
C2	-rRNA Enriched K+ 2	102,594,202	82,353,665	80.27
C3	-rRNA Enriched K+ 3	119,252,686	99,048,592	83.06
C4	-rRNA Enriched K+ 4	139,448,294	119,270,005	85.53
D1	-rRNA Enriched K- 1	97,335,200	82,714,249	84.98
D2	-rRNA Enriched K- 2	106,035,272	85,824,224	80.94
D3	-rRNA Enriched K- 3	92,890,624	78,885,906	84.92
D4	-rRNA Enriched K- 4	102,807,580	80,488,372	78.29

Naming scheme	**A**	-rRNA Input K+	4x biological replicates	A1, A2, A3, A4
	B	-rRNA Input K-	4x biological replicates	B1, B2, B3, B4
	C	-rRNA Enriched K+	4x biological replicates	C1, C2, C3, C4
	D	-rRNA Enriched K-	4x biological replicates	D1, D2, D3, D4