# A retroviral origin of vertebrate myelin

**Authors:** Tanay Ghosh[1,2]*, Rafael G. Almeida[3], Chao Zhao[1,2], Ginez Gonzalez M[1,2], Katherine Stott[4], Ian Adams[5], David H Rowitch[1,6], Nick Goldman[7], David A. Lyons[3], Robin JM Franklin[1,2]*

**Affiliations:**

[1] Wellcome – MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW, United Kingdom
[2] Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0AW, United Kingdom
[3] Centre for Discovery Brain Sciences, Chancellor's Building, GU 587, 49 Little France Crescent, Edinburgh, EH16 4SB, United Kingdom
[4] Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom
[5] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom
[6] Department of Paediatrics, University of Cambridge, United Kingdom
[7] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

*Corresponding authors.

Email: Tanay Ghosh: tg369@cam.ac.uk, Robin Franklin: rjf1000@cam.ac.uk

**Abstract:**

Myelin, the insulating sheath that surrounds neuronal axons, is produced by oligodendrocytes in the central nervous system (CNS). This evolutionary innovation, which first appears in jawed vertebrates, enabled rapid transmission of nerve impulses, more complex brains and greater morphological diversity. Here we report that RNA level expression of RNLTR12-int, a retrotransposon of retroviral origin, is essential for myelination. We show RNLTR12-int-encoded non-coding RNA binds to the transcription factor SOX10 to regulate transcription of myelin basic protein (Mbp, the major constituent of myelin). RNLTR12-int like sequences (which we name *RetroMyelin*) are found in all jawed-vertebrates and also regulate myelination in zebrafish, suggesting retroviral endogenization as a key step in the emergence of vertebrate myelin.

Myelination, the process by which axons are invested with a myelin sheath, had a profound impact on vertebrate evolution (1-3). By conferring the ability to transmit by rapid saltatory conduction and assisting neuronal viability by providing local metabolic support, the myelin sheath allowed axons to function over much greater lengths and hence vertebrates to attain a larger size and diversity than would have occurred in the absence of myelination. Myelination also allowed rapid conduction without needing to increase axonal diameter, enabling the packing of larger numbers of axons necessary for the evolution of complex central nervous systems (CNS). Phylogenetically, compacted myelin and genes critical to myelination such as myelin basic protein (Mbp) (2-5) likely appeared concurrently with the emergence of jaws in vertebrates, with myelin found in the most ancient living vertebrate, the Chondrichthyes (cartilaginous fish) but not in the Agnatha (jawless fish) (1-3, 6). Despite the significant functional advantages associated with myelination, a molecular explanation of what triggered this critical event in vertebrate evolution remains elusive.

Endogenous retrovirus (ERV) type retrotransposons are remnants of ancient retroviral sequence that persist in the genome and provide evidence of endogenization of diversified retrovirus to their vertebrate host that began several millions of years ago (7). Retrotransposons are repeated regions in the eukaryotic genome with the potential to mobilize within the genome via an RNA intermediate. The majority have lost the ability to undergo transposition due to accumulation of mutations: however, through evolution many have become 'domesticated' to exert gene regulatory functions by contributing to regulatory elements or being the source of functionally relevant mRNAs that can modify transcriptional networks (8-10). Although functional retrotransposons have been implicated in the maintenance of stem cell identity (11, 12), their role in regulating gene networks during maturation of oligodendrocytes (OLs) to form myelinating OLs is unknown.

We identified retrotransposon-specific probes using the rat Affymetrix Chip and then performed meta-analysis of microarray-based gene expression data in oligodendrocyte progenitor cells (OPCs) and OLs, isolated from postnatal day 7 (P7) rat brains (**Fig. S1 A**). Analysis revealed a set of protein-coding genes and retrotransposons with differential expression during OL differentiation (**Fig. S1 B, C**). Differentially expressed retrotransposons included long terminal repeats (LTR), short interspersed elements (SINE) and long interspersed elements (LINE) (**Fig. S1 B-D**).

To identify potential functional relationships among retrotransposons RNAs and protein-coding mRNAs in OLs we employed weighted gene co-expression network analysis (WGCNA). This revealed 13 different modules of highly co-regulated genes, which were significantly either up or down regulated in OLs (**Fig. 1 A, B**). The module with the highest significant association with OLs (turquoise, **Fig. 1B**) was overrepresented for the gene ontology (GO) term 'myelination' (**Fig. 1C**), suggesting that retrotransposon members of this module were potentially involved in regulating expression of myelination genes. Within this module, RNLTR12-int showed relatively high levels of expression in OLs as compared to OPCs (**Fig. 1D**), and correlated to a myelin gene co-expression network (**Fig. 1E, Fig. S1E**). We undertook further experimental analysis of the relationship of RNLTR12-int with Mbp, a key component of myelin (3).

To validate RNLTR12-int expression in OPCs and OLs, we extracted RNAs from Magnetic activated cell sorting (MACS)-purified OPCs (A2B5+) and OLs (MOG+) from postnatal day 7 (P7) rat brains and confirmed the expression of RNLTR12-int in OPCs and OLs by RT-PCR

(**Fig. S2 A**). We found 2.5-fold higher levels of RNLTR12-int expression in OLs compared to OPCs by qPCR (Reverse transcription followed by qPCR, RT-qPCR) (**Fig. 1F**). Similarly, expression of Mbp, myelin associated glycoprotein (Mag), tetraspanin 2 (Tspan2) and PB1D9, a SINE-retrotransposon within the same WGCNA module as RNLTR12-int, were higher in OLs versus OPCs (**Fig. 1F, S2 C**).

The RNLTR12-int consensus sequence is an internal sequence of endogenous retrovirus 1 (ERV1) that has only remnants of classic Gag-Pol ORFs (required for provirus replication cycle), lacks long-terminal repeat (LTR), and likely serves as a long non-coding RNA (**Fig. S2D**). RNA-seq reads from OPCs were aligned to the entire consensus sequence of RNLTR12-int, confirming its full-length expression at RNA level (**Fig. S2E**).

We next examined whether RNLTR12-int transcript (mRNLTR12-int) is required for regulation of Mbp expression in differentiating OPCs using small interfering RNA (siRNA) mediated RNA interference (RNAi) (13) under differentiation conditions. As shown (**Fig. 2A**), RNLTR12-int siRNA administration inhibited expression of MBP after 5 days of differentiation (**Fig. 2B**) and prevented development of complex oligodendrocyte morphologies as determined by O4 immuno-staining (**Fig. S3**). We observed 98% reduction of MBP+ OLs due to inhibition of RNLTR12-int compared to controls (**Fig. 2B**). In contrast, the proportion (12-13%) of cells immunostained with the astrocyte marker, glial fibrillary acid protein (GFAP), were unchanged in siRNLTR12-int and control siRNA transfected samples (**Fig. S4A**). RT-qPCR analyses revealed that the RNA expression of the early differentiation marker 2',3'-Cyclic-nucleotide 3'-phosphodiesterase (CNP) was decreased while the OPC marker platelet derived growth factor receptor alpha (Pdgfra) remained unaltered (**Fig. S4B**),

suggesting a role for mRNLTR12-int in differentiated cells undergoing maturation. Consistent with this, Mbp mRNA levels were drastically reduced (95%) (**Fig. 2C**), implying Mbp transcription is affected due to inhibition of mRNLTR12-int.

We next examined the effect of RNLTR12-int inhibition in the oligodendrocyte lineage during developmental myelination. We used a SOX10-driven shmiR construct where shRNA against RNLTR12-int is embedded into a microRNA cassette with a lineage-specific Emerald GFP (EGFP) reporter (**Fig. S5 A**). We used adeno associated virus (AAV) carrying SOX10-EGFP-shmiR (**Fig. S5 A**) *in vitro* to confirm it reliably infected and inhibited *RNLTR12-int* expression in differentiating OPCs (**Fig. S5 B**). We then injected AAV SOX10-EGFP-shmiR into the deep cortex of postnatal day (P) 1 rat (**Fig. S5 A**) and subsequently harvested brains at P14. We found that all EGFP immuno-positive cells (total GFP+ cells counted: 956 for shmiR-scrambled; 773 for shmir-RNLTR12-int) were also OLIG2 positive (**Fig. S5 C**), indicating oligodendrocyte lineage fidelity. Our protocol infected approximately 43% (956 GFP+OLIG+ out of 2214 OLIG+ cells counted in shmiR-scrambled; 773 GFP+OLIG2+ out of 1758 OLIG2+ cells counted for shmir-RNLTR12-int) of OLIG2+ cells (**Fig. S5 C**). We found an equal proportion (425 CC1+GFP+ out of 574 GFP+ cells in shmir-scrambled; 409 CC1+GFP+ out of 561 GFP+ cells in shmir-RNLTR12-int) of EGFP+ cells were immunoreactive for CC1 (an early differentiation marker of OLs) in both shmiR-RNLTR12-int and shmiR-Scrambled cases (**Fig. S5 D**), indicating that differentiation of OPCs into oligodendrocytes was not impaired. However, we observed a significant reduction of MBP expression in shmiR-RNLTR12-int infected brains (**Fig. 2D**) suggesting that MBP expression *in vivo* is RNLTR12-int dependent.

Because non-coding RNA also binds to certain transcription factors to affect target gene expression (14-19), we tested if SOX10 mediated transcription of Mbp (20,21) is regulated by RNLTR12-int. We found 15 potential SOX10 binding sites in RNLTR12-int consensus sequence (**Fig. 3A**). To examine if SOX10 can binds to mRNLTR12-int *in vivo*, we used a SOX10 antibody and performed RNA immunoprecipitation (RIP) on post-natal rat brains. RT-qPCR analysis on RIP samples indicated that SOX10 bound mRNLTR12-int *in vivo* (**Fig. 3A**), and we confirmed direct binding between SOX10 and mRNLTR12-int using surface plasmon resonance analysis (**Fig. 3B**). We next asked whether the binding of SOX10 to the Mbp promoter is RNLTR12-int-dependent. As illustrated in **Figure 3C**, two conserved sequence elements (S2, S1) (21) located 153 bp apart on the Mbp promoter are important for SOX10 binding (**Fig. 3C**). siRNLTR12-int transfected OPCs were allowed to differentiate for 4 days in culture. Subsequently chromatin immunoprecipitation (ChIP) with SOX10 antibody and qPCR analysis was performed. We identified a significant enrichment of the Mbp promoter region in control siRNA transfected samples (containing SOX10 immunoprecipitant), confirming SOX10 localization to Mbp promoter region comprising S2 and S1 (**Fig. 3C**). However, there was no enrichment of the Mbp promoter region in siRNLTR12-int transfected samples, suggesting mRNLTR12-int is essential for SOX10 binding to the Mbp promoter (**Fig. 3C, Fig. 3D**).

Distribution of transposons in a species genome is not random (22): insertions which confer survival fitness to their host are preferentially retained through natural selection. Compacted myelin or MBP expression is evident only in jawed vertebrates and among them the most ancient living animals are the cartilaginous fishes (e.g. sharks) (1-3, 6). After establishing a direct regulatory relationship between Mbp and mRNLTR12-int in mammalian CNS, we searched for RNLTR12-int-similar sequences within jawed vertebrates, jawless vertebrates and

selected invertebrates based on a profile hidden Markov model, followed by repeat annotation and repeat family identification. We were able to detect RNLTR12-int-like sequences (which we have named *RetroMyelin* (Retrotransposon sequences associated with Myelin evolution) in all the jawed vertebrate phyla analysed, including the ancient cartilaginous fishes (elephant sharks, whale sharks) (**Fig. 4A**). In contrast, we found no such sequence in jawless vertebrates (e.g., lamprey), fish-like jawless chordates (lancelet) and invertebrates (*Drosophila, C. elegans,* Sea anemone, Sea urchin) (**Fig. 4A**). *RetroMyelin* in jawed vertebrates all belonged to the ERV1 family, annotated as an internal sequence of ERV1. We performed phylogenetic analysis of *RetroMyelin* and built a phylogenetic tree based on a probabilistic model. *RetroMyelin* sequences were clustered in accordance with the current view of the relationships among the phyla (**Fig. 4B**). Furthermore, we took *RetroMyelin* sequences from human, mouse, rat, zebrafish and elephant shark and employed the ncbi BLASTN search engine to query expressed sequence tags (ESTs) databases. We found ESTs representing these sequences with a full coverage, suggesting that they are indeed expressed at RNA-level (**Fig. S6**).

Selective constraints during evolution generally limits the rapid divergence of essential genomic sequences. Mutational rate heterogeneity analysis is a proxy to understand such determinants. We employed a rate heterogeneity model with the assumption of the variation of evolutionary rates of nucleotide substitution over sites following a gamma distribution (23). Presence of mutational rate heterogeneity was confirmed by a likelihood ratio test ($2\delta = 967.9$; $p < 0.005$), providing strong evidence of rate variation over sites in *RetroMyelin*. Evolutionary relative rate estimates among the sites in *RetroMyelin* suggested that there are many sites (58%) that are relatively conserved or were slowly evolving (relative rate<1) (**Fig. 4C**). Together this is suggestive of selective constraint acting on some parts of these sequences during evolution.

To test whether *RetroMyelin* played a similar role in the regulation of Mbp in other vertebrate classes we used a zebrafish reporter line Tg(mbp:EGFP-CAAX), where EGFP fluorescent intensity is a proxy measure of Mbp promoter activation. Using CRISPR/Cas9 technology, we acutely disrupted a small region of *RetroMyelin* sequences in the zebrafish genome by injecting fertilized eggs with a *RetroMyelin* guide crRNA and Cas9 enzyme (**Fig. 4D**). A significant reduction of EGFP fluorescence intensity was detected at 5 days post-fertilization (dpf) in zebrafish larva that had been injected with a *RetroMyelin* guide crRNA compared to Cas9-only control animals (**Fig. 4D**), suggesting a conserved function of *RetroMyelin* in regulating Mbp transcription between fish and mammals.

Generation of myelin by oligodendrocytes is not only important in development, normal physiology (24, 25) and in regeneration following demyelination (26, 27) but also played a central role in evolution (1-3). We provide evidence that transcriptional regulation of Mbp is critically governed by *RetroMyelin* encoded transcripts. MBP is essential for myelination of the vertebrate CNS (2-5), unlike the other myelin proteins MAL, MAG, CNP, PLP, PMP22, TSPAN2 that are evolutionarily older and exist in non-myelinated jawless vertebrates and invertebrates (3). In our phylogenetic analysis, we found *RetroMyelin* exists in all jawed vertebrates where myelination is evident. We propose that in vertebrate lineages, ERV1 type of retrotransposons most probably originate due to germline invasion and genomic integration of retroviruses carrying an RNLTR12-int like sequence. The original ERV was either purged from the population pool and lost, or became fixed and exapted to regulate Mbp and myelination. The role for ERV1 class retrotransposons in Schwann Cells, which myelinate the peripheral nervous system, remains to be addressed in future work. Our study suggests that *RetroMyelin* was co-opted to regulate transcription of Mbp and thus endogenization of ERV1 into the vertebrate genome is coupled to the evolutionary emergence of myelination.

**References**


1. B Zalc, DR Colman, Science 288, 271-272 (2000).

2. JL Salzer, B Zalc, Curr Biol. 26, R971-R975 (2016).

3. KA Nave, HB Werner, Annu Rev Neurosci. 44, 197-219 (2021).

4. G Lemke, Neuron 1, 535-543 (1988).

5. RD Fields, B Stevens-Graham, Science 298, 556-562 (2002).

6. S Nawaz et al., Glia 61, 1364-1377 (2013).

7. WE Johnson, Nat Rev Microbiol. 17, 355-370 (2019).

8. PJ Thompson et al., Cell 62, 766-776 (2016).

9. HH Jr Kazazian, Science 303,1626-1632 (2004).

10. HH Jr Kazazian et al., N Engl J Med. 377, 361-370 (2017).

11. M Percharde et al., Cell 174, 391-405 (2018).

12. J Göke et al., Cell Stem Cell. 16,135-141 (2015).

13. GB Robb et al., Nat Struct Mol Biol. 12, 133-137 (2005).

14. ZE Holmes et al., Nat Commun. 11, 1805 (2020).

15. AA Sigova et al., Science 350, 978-981(2015).

16. T Hung et al., Nat Genet. 43, 621-629 (2011).

17. LA Cassiday et al., Nucleic Acids Res. 30, 4118-4126 (2002).

18. Y Jeon et al., Cell 146, 119-133 (2011).

19. JM Gottesfeld et al., Chem Biol. 10, 584-585 (2003).

20. CC Stolt et al., Genes Dev. 16, 165-170 (2002).

21. H Li et al., J Neurosci. 27, 14375-14382 (2007).

22. G Bourque et al., Genome Biol. 19, 199 (2018).

23. Goldman N, Whelan S, Mol Biol Evol. 17, 975-978 (2000).

24. IA McKenzie et al., Science 346, 318-322 (2014).

25. W Xin, JR Chan, Nat Rev Neurosci. 21, 682-694 (2020).

26. RJM Franklin, C ffrench-Constant, Nat Rev Neurosci. 18, 753-769 (2017).

27. Fancy SP et al., Annu Rev Neurosci. 34, 21-43 (2011).

**Acknowledgments:**

We thank Peter Humphreys and Darran Clements for the technical assistance with confocal imaging, Prof. Travis J Wheeler (University of Montana) for fruitful discussion regarding remote homology search with repeat sequence.

**Funding:**

Adelson Medical Research Foundation (AMRF) (R.J.M.F., D.H.R.).

Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute core support grant 203151/Z/16/Z (R.J.M.F., D.H.R.).

Wellcome Trust Investigator Award and European Research Council Advanced Grant (no 789054; to D.H.R.).

UK Multiple Sclerosis Society grant MS50 (R.J.M.F.).

This study was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Author contributions:**

Conceptualization: T.G., D.H.R., R.J.M.F.; Formal analysis: T.G., N.G.; Investigation: T.G., C.Z., R.G.A., G.G.M., K.S.; Methodology: T.G., R.G.A., R.J.M.F., D.A.L., N.G., I.A.; Project administration: T.G., R.J.M.F.; Supervision: R.J.M.F., T.G.; Writing – original draft: T.G., R.J.M.F.; Writing – review & editing: T.G., R.J.M.F., D.H.R., D.A.L., N.G., C.Z., I.A., R.G.A., K.S., G.G.M.

**Competing interests:**

Authors declare that they have no competing interests.

**Data and materials availability:**

All data are available in the main text or the supplementary materials.

**Supplementary Materials**

Materials and Methods

Figs. S1 to S6

Tables S1

References (28-61)

**Figure 1**

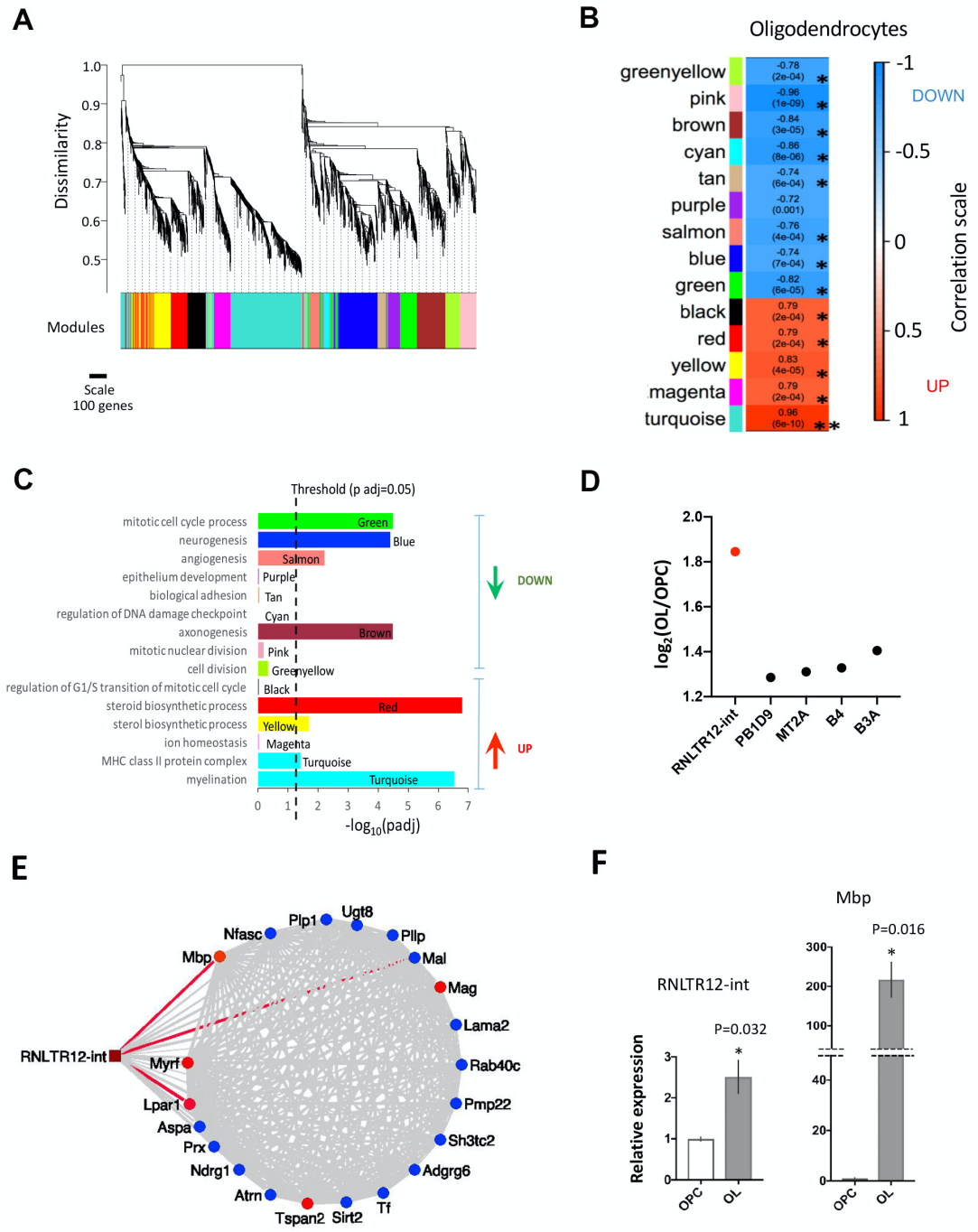


**Fig. 1. Network analyses of differentially expressed retrotransposons and protein coding genes in oligodendrocytes.** (**A**) Cluster dendrogram and derived modules (color labeled) after WGCNA. (**B**) Association of modules with oligodendrocytes is determined by WGCNA. * p <0.0007 (Bonferroni threshold). p value in parentheses, correlation is written on top and the

box is color coded according to the color scale on the left. (**C**) GO-terms which are significantly overrepresented in modules. padj: adjusted p value (Benjamini-Hochberg FDR). Red and green arrow indicate the modular gene expression changes (up or down respectively) in OL as compared to OPC. (**D**) Expression changes (OL vs OPC) of retrotransposons in turquoise module were plotted, as determined by microarray. P adj (Benjamini-Hochberg FDR)<0.01. (**E**) Potential interaction of RNLTR12-int with myelination gene-network. Nodes: round shape, Edges: line, top 5 hubs: red round, top 3 edges: red line. Edges represent connection (expression correlation) between two nodes (**F**) Expressions of RNLTR12-int transcript and Mbp in OL and OPC were determined using RT-qPCR and relative expressions as compared to OPC were plotted. Data normalized to Actb. N=4 (P7 rats), mean±SEM, *p<0.05, Student's t test (unpaired, two tailed).

## Figure 2

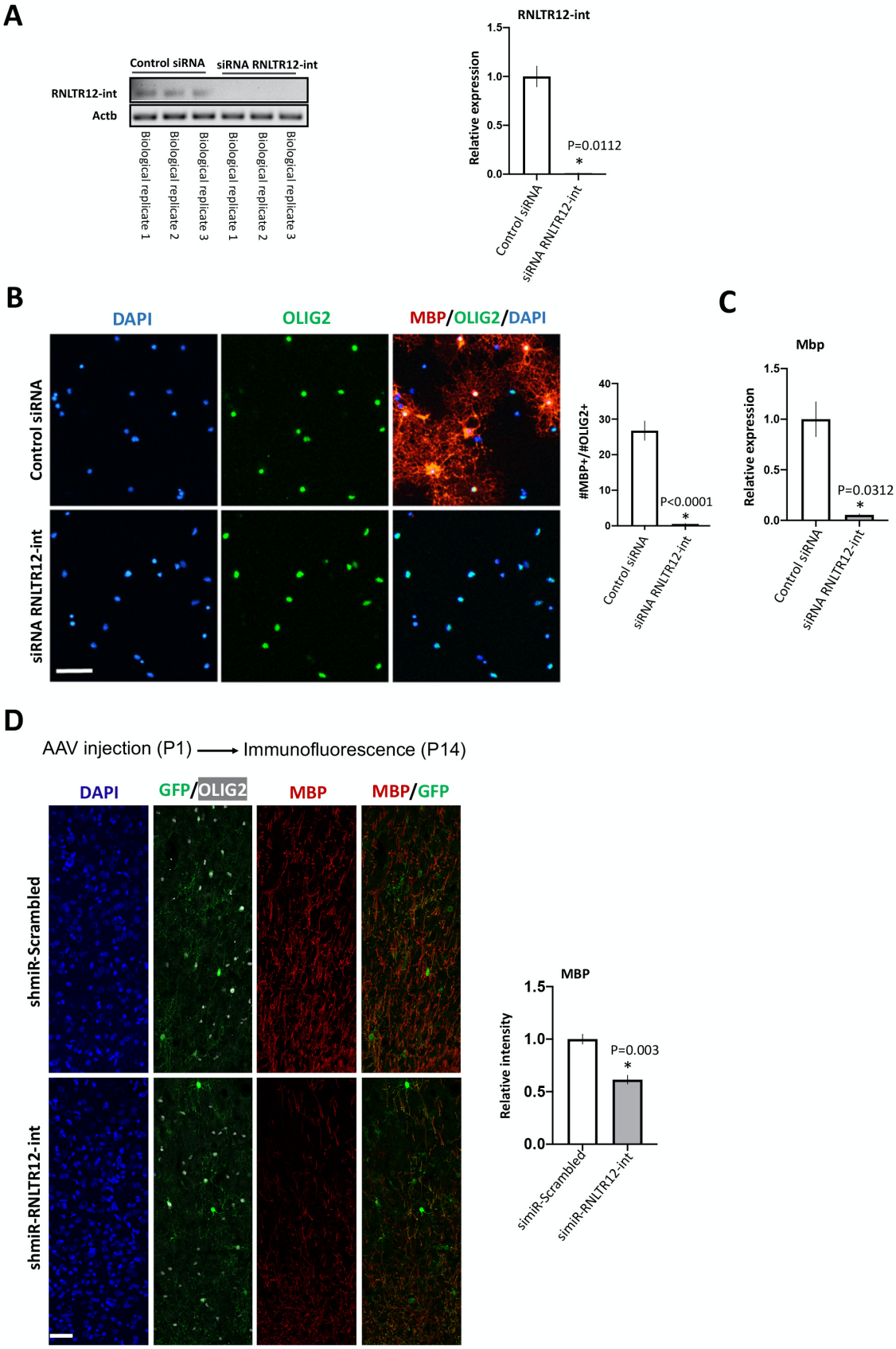
A
Control siRNA
siRNA RNLTR12-int
RNLTR12-int
Actb
Biological replicate 1
Biological replicate 2
Biological replicate 3
Biological replicate 1
Biological replicate 2
Biological replicate 3
RNLTR12-int
Relative expression
1.5
1.0
0.5
0.0
P=0.0112
*
Control siRNA
siRNA RNLTR12-int
B
DAPI
OLIG2
MBP/OLIG2/DAPI
Control siRNA
siRNA RNLTR12-int
#MBP+/#OLIG2+
40
30
20
10
0
P<0.0001
*
Control siRNA
siRNA RNLTR12-int
C
Mbp
Relative expression
1.5
1.0
0.5
0.0
P=0.0312
*
Control siRNA
siRNA RNLTR12-int
D
AAV injection (P1) → Immunofluorescence (P14)
DAPI
GFP/OLIG2
MBP
MBP/GFP
shmiR-Scrambled
shmiR-RNLTR12-int
MBP
Relative intensity
1.5
1.0
0.5
0.0
P=0.003
*
shmiR-Scrambled
shmiR-RNLTR12-int

**Fig. 2. Inhibition of RNLTR12-int expression affects Mbp expression.** (**A**) RNLTR12-int expression was inhibited upon transfection of siRNA against RNLTR12-int (siRNLTR12-int) as determined by RT-PCR. PCR products were run in a 2% agarose gel (left). RT-qPCR analysis (right). Data were normalized to cytoplasmic b-actin (Actb). Control siRNA: siGENOME non-targeting siRNA pool (Dharmacon). N=3 independent experiments (each time 4 P7 brains were pooled for OPC isolation), mean±SEM, *p<0.05, Student's t test (unpaired, two tailed) with Welch's correction. (**B**) Left: Immunofluorescence analysis using antibody to OLIG2 (Oligodendrocyte lineage specific marker) and to MBP (differentiation/maturation marker). Upon transfection, OPCs were maintained in differentiation media for 5 days. Scale bar: 60μm. Right: Quantification of matured OLs (MBP+, OLIG2+), presented as a percentage of total OLIG2+ cells. N=3 independent experiments (each time 3 replicates), mean±SEM, *p<0.0001, Two-way ANOVA. (**C**) Effect of the inhibition of RNLTR12-int on RNA level expression of Mbp was determined by RT-qPCR. Data were normalised to Actb. N=3 independent experiments, mean±SEM, * p<0.05, Student's t test (unpaired, two tailed) with Welch's correction. (**D**) AAV carrying SOX10-EGFP-shmiR-RNLTR12-int/Scrambled construct were injected into the deep cortex of rat at P1 (also see **Fig. S5**) and brains were harvested at P14. Left: Immunofluorescence analysis using antibody to MBP, GFP, OLIG2. Scale bar: 22 um. Right: MBP intensity were quantified and plotted relative to shmiR-Scrambled infected samples. N=3 different rats, mean±SEM, *p<0.01, Student's t test (unpaired, two tailed).

## Figure 3

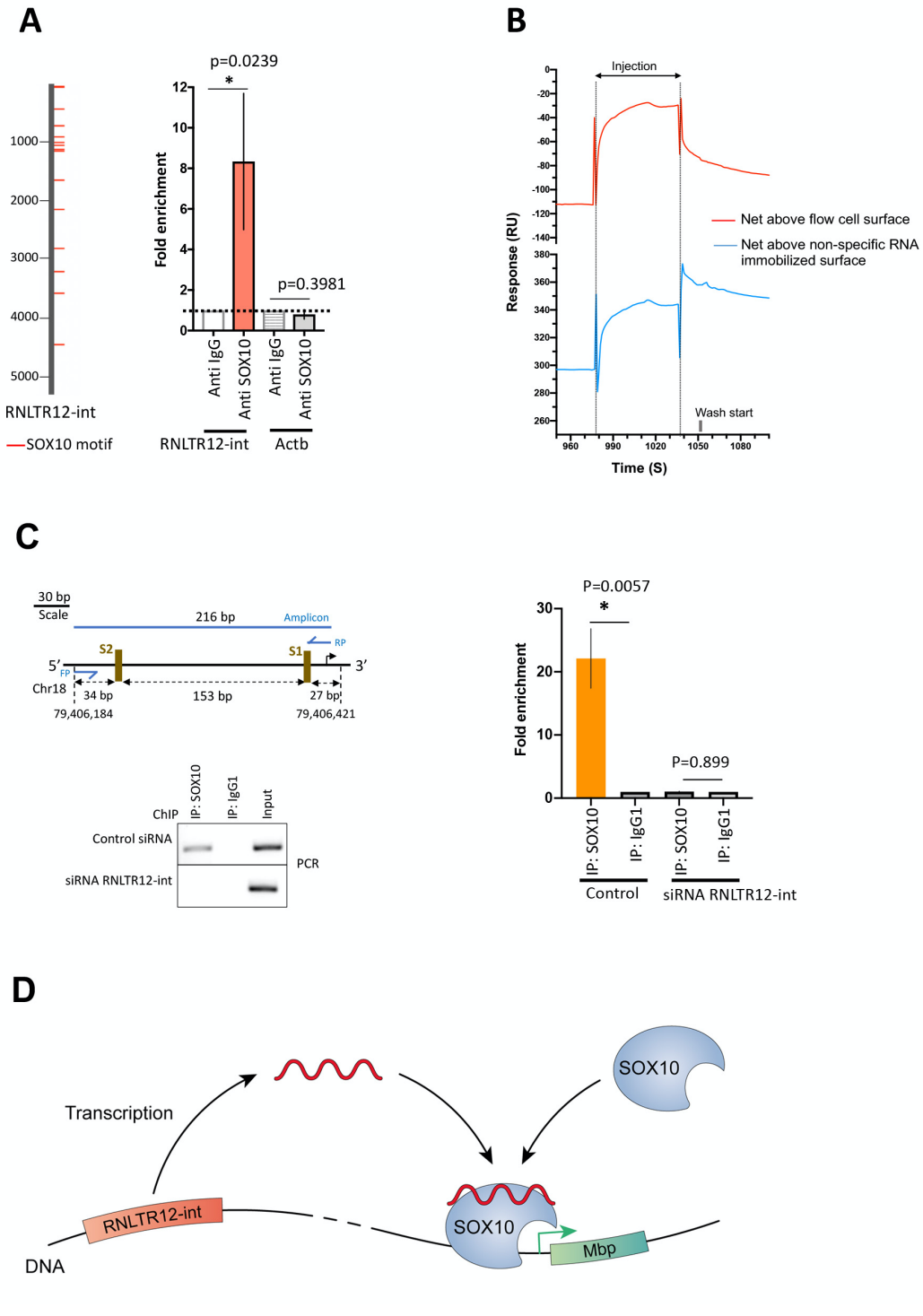


**Fig. 3. SOX10 binds to RNLTR12-int encoded transcript and this is required for SOX10 binding to Mbp promoter.** (**A**) Left: Putative SOX10 binding motif was mapped onto

RNLTR12-int consensus sequence (grey vertical bar). Horizontal lines: SOX10 binding sites (red) and nucleotide position (black). JASPAR (Supplementary materials and methods) relative score threshold= 90%, TFM p value threshold<0.01. Right: RT-qPCR after crosslinking RIP (from P7 rat brains) with SOX10 or control immunoglobulin G1, showing SOX10 association with RNLTR12-int transcript. N=4 independent experiments (each time 5-6 brains were pooled). mean±SEM, * p<0.05, Ratio paired t-test (two-sided). (**B**) Direct binding of RNLTR12-int transcript and SOX10, determined by surface plasmon resonance. SOX10 protein was injected over the streptavidin-coated flow cell sensor surface immobilized with 3' end biotin-labelled RNLTR12-int/non-specific RNA or kept blank. Sensorgram represents the data corrected for non-specific binding to the flow cell surface (red) or for non-specific binding to the flow cell surface immobilized with non-specific RNA (Blue). Net positive response above these backgrounds indicates direct binding. (**C**) ChIP analyses with SOX10 or control immunoglobulin G1, 4 days after siRNA transfection, showing SOX10 occupancy to Mbp promoter is affected in absence of RNLTR12-int transcript. Left: (Top) schematic representation of Mbp promoter. Rat genome assembly: rn6, FP: forward primer, RP: reverse primer for PCR amplification after ChIP. SOX10 binding conserved elements: S2 ('AACAAT') and S1 ('TTCAAA'). Black right arrow after S1: Transcription start site (TSS). (Bottom) PCR amplified immunoprecipitated samples, run on 2% agarose gel. Right: qPCR analysis after ChIP. N=3 independent experiments. mean±SEM, * p<0.01, Ratio paired t-test (two-sided). (**D**) Illustration represents SOX10 mediated transcription of Mbp is RNLTR12-int transcript dependent. Copies of RNLTR12-int exists in the rat genome. For simplicity reason only one is drawn. RNLTR12-int is transcribed. Association of the transcription factor SOX10 to Mbp promoter requires a direct binding interaction between SOX10 and RNLTR12-int transcript. Green arrow: TSS.

**Figure 4**

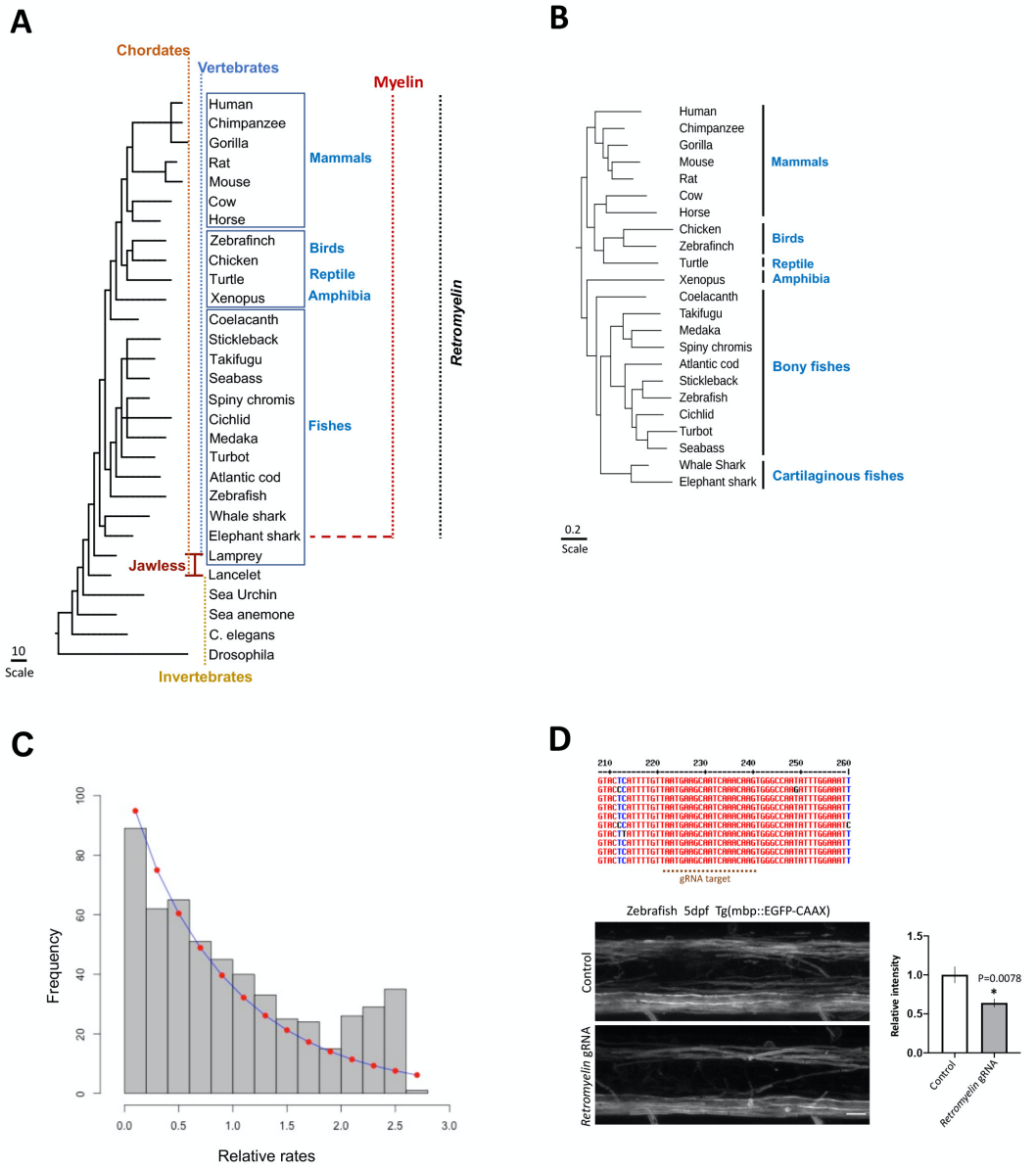


**Fig. 4. Phylogenetic relationship of RNLTR12-int and Mbp** (**A**) Taxonomy common tree display, represents a hierarchical view of relationships among taxa and their lineages. Appearance of Myelin/MBP is associated with the gain of hinge-jaw in vertebrates (1-3, 6). Presence of myelin/MBP (1-3, 6) in most ancient living vertebrate (the cartilaginous fishes) (red dotted horizontal line) to higher vertebrates (red dotted vertical line) is marked. Black dotted vertical line marks the vertebrate phyla wherever *RetroMyelin* is identified in our study. All identified *RetroMyelin* belong to ERV1 family and carrying internal sequence of ERV1.

(**B**) Phylogenetic tree was reconstructed based on identified *RetroMyelin* by using a maximum likelihood method. *RetroMyelin* were clustered in line with the relationship among phyla. (**A-B**) trees are unrooted but a root was placed in the midpoint for clarity. (**C**) Evolutionary rate heterogeneity analysis of *RetroMyelin*. Relative rate per site was represented as histogram (grey bars). Expected values of relative rate per bin were calculated from the cumulative distribution function of the inferred gamma distribution (the inferred shape parameter ($\alpha$) = 0.978, standard error = 0.067) and plotted (red dots). (**D**) Reduction of Mbp-promoter driven EGFP expression upon genome editing of a small region of *RetroMyelin* sequences in a transgenic zebrafish reporter line (Tg(mbp:EGFP-CAAX). Top: top 10 hits of *RetroMyelin* loci in zebrafish were aligned and gRNA target region was marked (dotted line). Bottom left: fluorescence image of anterior-posterior position of larva (5dpf). Bottom right: EGFP intensity were quantified and plotted relative to control (Cas9-only). N=8 different larvae, mean$\pm$SEM, *$p<0.01$, Student's t test (unpaired, two tailed).