THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A retroviral link to vertebrate myelination through retrotransposon RNA-mediated control of myelin gene expression

Link:
[Link to publication record in Edinburgh Research Explorer](#)

Document Version:
Peer reviewed version

Published In:
Cell

A retroviral link to vertebrate myelination through retrotransposon RNA-mediated control of myelin gene expression

Tanay Ghosh[1,2,3]*, Rafael G. Almeida[4], Chao Zhao[1,2,3], Abdelkrim Mannioui[5], Elodie Martin[6], Alex Fleet[2,3], Civia Z. Chen[1,2,3], Peggy Assinck[1,2,3], Sophie Ellams[1], Ginez A. Gonzalez[2,3], Stephen C. Graham[7], David H Rowitch[2,8], Katherine Stott[9], Ian Adams[10], Bernard Zalc[6], Nick Goldman[11], David A. Lyons[4], Robin JM Franklin[1,2,3,12]*

[1]AltosLabs-Cambridge Institute of Science, Cambridge CB21 6GP, United Kingdom

[2] Wellcome – MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0AW, United Kingdom

[3] Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0AW, United Kingdom

[4] Centre for Discovery Brain Sciences, Chancellor's Building, GU 587, University of Edinburgh, 49 Little France Crescent, Edinburgh, EH16 4SB, United Kingdom

[5] Sorbonne Université, Institut de Biologie Paris-Seine (IBPS), Aquatic Facility, 75005 Paris, France

[6] Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, 75013 Paris, France

[7]Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

[8] Department of Paediatrics, University of Cambridge, United Kingdom

[9]Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

[10] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom

[11] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

[12] Lead contact

*Corresponding authors

Email: Tanay Ghosh: writetotanay@gmail.com , Robin JM Franklin: rfranklin@altoslabs.com

SUMMARY

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 RNA binds to the transcription factor SOX10 to regulate transcription of myelin basic protein (Mbp, the major constituent of myelin) in rodents. RNLTR12-int like sequences (which we name *RetroMyelin*) are found in all jawed-vertebrates and we further demonstrate their function in regulating myelination in two different vertebrate classes (zebrafish and frogs). Our study therefore suggests that retroviral endogenization played a prominent role in the emergence of vertebrate myelin.

INTRODUCTION

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 hundreds of million 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.

Here we analyzed gene-networks associated with oligodendrocytes and their derivation from progenitors, and found RNA level expression of RNLTR12-int, a retrotransposon of retroviral origin in rats, that is crucial for the expression of myelin basic protein (Mbp). We demonstrate that by interacting with the transcription factor SOX10, RNLTR12-int encoded RNA regulates the transcription of Mbp. We identified RNLTR12-int like sequences (which we name *RetroMyelin*) in all jawed-vertebrates. Our phylogenetic analysis revealed that acquisition of *RetroMyelin* in species genome occurred likely through a convergent evolution. We further demonstrate the function of *RetroMyelin* in regulating myelination/Mbp in two other vertebrate classes (zebrafish and frogs). Our study therefore reveals an important role for retroviral endogenization in the emergence of vertebrate myelin.

RESULTS

RNLTR12-int is a potential regulator of Myelin/Mbp

We identified retrotransposon-specific probes using the rat Affymetrix Chip and then performed meta-analysis of microarray-based gene expression data (GSE245032) in oligodendrocyte progenitor cells (OPCs) and OLs, isolated from postnatal day 7 (P7) rat brains (**Figure S1A**). Analysis revealed a set of protein-coding genes and retrotransposons with differential expression during OL differentiation (**Figure S1B**, **C**). Differentially expressed retrotransposons included long terminal repeats (LTR), short interspersed elements (SINE) and long interspersed elements (LINE) (**Figure S1B-D**).

To identify potential functional relationships among retrotransposon 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 (**Figure 1 A, B**). The module with the highest significant association with OLs (turquoise, **Figure 1B**) was overrepresented for the gene ontology (GO) term 'myelination' (**Figure 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 (**Figure 1D**), and correlated to a myelin gene co-expression network (**Figure 1E**, **Figure S1E**). We undertook further experimental analysis of the relationship of RNLTR12-int with Mbp, a key component of myelin[3].

Experimental validation of RNLTR12-int expression

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 (**Figure S2A**). Expression level of Cytoplasmic beta actin (Actb) remained unchanged in OL and OPC, and were therefore used as an endogenous gene for normalization control (**Figure S2B**). 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) (**Figure 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 (**Figure 1F, Figure S2C**).

RNLTR12-int copies in the rat genome

The RNLTR12-int consensus sequence is an internal sequence of endogenous retrovirus 1 (ERV1) that has remnants of classic Gag-Pol ORFs (required for provirus replication cycle), and likely serves as a long non-coding RNA (**Figure S2D**). RNA-seq reads (GSM3967160) from OPCs were aligned to the entire consensus sequence of RNLTR12-int, confirming its full-length expression at RNA level (**Figure S2E**). In the rat genome, we identified 118 sequences of RNLTR12-int which were flanked by long terminal repeat (LTR) elements and 80 sequences which were not flanked by the LTR (**Figure 1G**, **Figure S3A**; see STAR methods for details).

Among these 80 sequences, we found 2 genes (Pygl, Ncbp2) whose intronic regions (in opposite strand) contain fragments of RNLTR12-int. However, unlike RNLTR12-int, Pygl and Ncbp2 expression levels are not different in OPC and OL (**Figure S3B**). RNLTR12-int copies in the rat genome are less than 1% divergent from its consensus sequence. It is not clear at present whether any specific locus or multiple loci direct RNLTR12-int expression in OPC and OL.

Mbp expression is RNLTR12-int dependent

We next examined whether the 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 (**Figure 2A**), RNLTR12-int siRNA administration inhibited expression of MBP after 5 days of differentiation (**Figure 2B**) and prevented development of complex oligodendrocyte morphologies as determined by O4 immuno-staining (**Figure S4A**). We observed 98% reduction of MBP+ OLs due to inhibition of RNLTR12-int compared to controls (**Figure 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 (**Figure S4B**). RT-qPCR analyses revealed that the RNA level 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 (**Figure 2C**), suggesting a role for mRNLTR12-int in differentiated cells undergoing maturation. Consistent with this, Mbp mRNA levels were drastically reduced (**Figure 2C**), implying Mbp transcription is affected due to inhibition of mRNLTR12-int.

In the peripheral nervous system (PNS), myelin sheaths are generated by Schwann cells (SC)[3], where myelin protein zero (MPZ, alias P0) and MBP are both important for myelin compaction[14]. RT-qPCR analyses confirm that knock-down of RNLTR12-int siRNA in differentiating SC decreases the expression of both P0 and Mbp, suggesting that not only in the CNS, but also in the PNS, the mRNA expression of P0 and Mbp are RNLTR12-int dependent (**Figure S4C**).

We tested the effect of RNLTR12-int inhibition on global gene expression in differentiating OPCs by sequencing OPC samples treated with siRNA against RNLTR12-int or control siRNA. Gene ontology analyses revealed that the most differentially expressed genes related to myelination and were downregulated in the knock-down samples (**Figure 2D**). Since RNLTR12-int fragments might exist inside a gene, we tested whether our results might have arisen due to off-target effects. A 125 bp region, flanked by qPCR primers to amplify RNLTR12-int including our siRNA target, was used as an input in BLAT (using rat genome) and annotations extracted for all BLAT hits. No BLAT hits fell within the exonic region of any genes. Some hits fell within intronic regions in 19 genes: however, none of these were differentially expressed by inhibition of RNLTR12-int (**Figure S4D**). Restoring expression of RNLTR12-int following RNLTR12-int siRNA silencing significantly elevated MBP expression in differentiating OPCs (**Figure S5A**). Together, these data supported our conclusion that the effects of RNLTR12-int inhibition were indicative of a direct role in myelination.

RNLTR12-int regulates Mbp expression in developmental myelination

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 (**Figure 3A**). We used adeno associated virus (AAV) carrying SOX10-EGFP-shmiR (**Figure 3A**) *in vitro* to confirm it reliably infected and inhibited RNLTR12-int expression in differentiating OPCs (**Figure 3B**). We then injected AAV SOX10-EGFP-shmiR into the deep cortex of postnatal day (P) 1 rat (**Figure 3A**) and subsequently harvested brains at P14. We found that almost all EGFP immuno-positive cells were also OLIG2 positive (99.7% for shmiR Scrambled, total GFP+ DAPI+ cells counted: 959; 99.5% for shmir-RNLTR12-int, total GFP+ DAPI+ cells counted: 777) (**Figure 3C**), indicating oligodendrocyte lineage fidelity. Our protocol infected approximately 43% of OLIG2+ cells (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) (**Figure 3C**). We found an equal proportion of EGFP+ cells were immunoreactive for CC1 (an early differentiation marker of OLs) in both shmiR-RNLTR12-int and shmiR-Scrambled cases (425 CC1+GFP+ out of 574 GFP+ cells in shmiR-scrambled; 409 CC1+GFP+ out of 561 GFP+ cells in shmir-RNLTR12-int) (**Figure 3D**). Additionally, the expression of another

early differentiation marker, CNPase, showed a modest decrease in shmiR-RNLTR12-int cases (**Figure S5B**). These results indicate that the differentiation of OPCs into oligodendrocytes was not impaired. However, we observed a significant reduction of MBP expression (both in protein and RNA level) (**Figure 3E, Figure 3F**) in shmiR-RNLTR12-int infected brains. To test the effect of RNLTR12-int inhibition on myelination we injected AAV carrying construct (**Figure 3A**) into the corpus callosum of postnatal day (P)6 rats and harvested brains at postnatal Day (P) 21. Electron microscopy (EM) analyses revealed a higher G-ratio (calculated as the diameter of the axon divided by the diameter of the entire nerve fiber including myelin sheath) and a greater number of unmyelinated axons in shmiR-RNLTR12-int infected corpus callosum (**Figure 3G, Figure S5C**), indicating less efficient myelination and hypomyelination. No significant difference in periodicity (measured as the distance between major dense lines) was observed (**Figure S5D**). These results collectively indicate a dependence on RNLTR12-int for MBP expression and *in vivo* myelination.

RNLTR12-int transcript is essential for SOX10 binding to Mbp promoter

Because non-coding RNA also binds to certain transcription factors to affect target gene expression[15-20], we tested if SOX10 mediated transcription of Mbp[21-22] is regulated by RNLTR12-int. To examine if SOX10 bound 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* (**Figure 4A**). Furthermore, electrophoretic mobility shift assay (EMSA) and surface plasmon resonance (SPR) analyses suggest the potential of direct binding between SOX10 and mRNLTR12-int (**Figure 4B, C**). We next asked whether the binding of SOX10 to the Mbp promoter is RNLTR12-int-dependent. As illustrated in **Figure 4D**, two conserved sequence elements (S2, S1)[22] located 153 bp apart on the Mbp promoter are important for SOX10 binding. 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 (**Figure 4D**). 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 (**Figure 4D, E**).

Other than Mbp, transcription of many other genes in the oligodendrocytes requires SOX10[23]. Myelin regulatory factor (Myrf) is one such gene which carries an evolutionary conserved SOX10 binding region (called ECR9) in the intron 1 that was experimentally validated earlier[24]. Myrf deletion was reported not to affect OPC differentiation to OL but affect the later stages of myelination[25]. Using ChIP-qPCR, we showed that RNLTR12-int is also required for SOX10 binding to the ECR9 region of Myrf (**Figure S5E**). Furthermore, we compared the list of downregulated genes due to RNLTR12-int inhibition (RNAseq data: GSE237744) with the potential SOX10 target genes in the OL (ChIP-seq data, GSE64703). We found that only 0.8% genes (31 genes out of 3920) are downregulated due to inhibition of RNLTR12-int (**Figure 4F**), suggesting RNLTR12-int is potentially needed for the expression of a small subset of SOX10 targets. Future studies are required to identify any common molecular characteristic among these SOX10 target regions.

RNLTR12-int like sequence in jawed vertebrates

Distribution of transposons in a specie's genome is not random[26]: 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 these the most ancient living animals are the cartilaginous fishes[1-3, 6]. After establishing a direct regulatory relationship between Mbp and mRNLTR12-int in mammalian CNS, we searched for RNLTR12-int-like 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 classes analysed, including the ancient cartilaginous fishes (elephant sharks, whale sharks) (**Figure 5A, Table S1**). 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) (**Figure 5A**). *RetroMyelin* in jawed vertebrates all belonged to the ERV1 family, every occurrence being annotated as an internal sequence of ERV1. 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 full coverage, suggesting that they are indeed expressed at RNA-level (**Figure 5B**).

Acquisition of *RetroMyelin* in host genome

RetroMyelin sequences in vertebrate genomes could have been acquired either once before their speciation from the most recent common ancestor, or through multiple retroviral invasions after speciation had begun. In the former case, sequence divergence would have started at the ancestor stage and copies randomly sampled from a species would not be clustered together in a phylogenetic tree (**Figure 6A**). However, if *RetroMyelin* were acquired after speciation then the sequence copies of each species would form separate cluster as divergence would only happen within a species (**Figure 6B**). To test this, we sampled multiple *RetroMyelin* copies from 22 different species. In the reconstructed phylogenetic tree, we found that the copies from each species are clustered together, suggesting the occurrence of separate invasion events after final speciation followed by diversification of sequence copies within that host species (**Figure 6C**). Thus, the acquisition of *RetroMyelin* sequences into host specie's genomes and their co-option for Mbp expression likely happened through convergent evolution.

Selective constraints during evolution generally limit the rapid divergence of essential genomic sequences. We employed a mutational rate heterogeneity model with the assumption that the variation of evolutionary rates of nucleotide substitution over sites followed a gamma distribution[27]. Presence of mutational rate heterogeneity was confirmed by a likelihood ratio test ($2\delta = 1528.5$; $p < 0.005$), 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 (52%) that are relatively conserved or were slowly evolving (relative rate<1) (**Figure 6D**). Together, these data are suggestive of selective constraints acting on some parts of these sequences during evolution.

***RetroMyelin* regulates myelination in other non-mammalian classes of vertebrate**

To test whether *RetroMyelin* played a similar role in the regulation of Mbp in other vertebrate classes, we employed zebrafish (*Danio rerio*) and frog (*Xenopus laevis*). 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 (**Figure S6**). No difference of morphological development (eye area and the body length) was observed at 5 days post-fertilization (dpf) in zebrafish larva that had been injected with a *RetroMyelin* guide crRNA compared to Cas9-only control animals (**Figure 7A, B**). However, a significant reduction of EGFP fluorescence intensity was detected along the dorsal and ventral tracts of the spinal cord (**Figure 7A, C**), suggesting the involvement of *RetroMyelin* in regulating Mbp transcription in fish. Furthermore, *RetroMyelin* disruption in frog resulted in a significant reduction of myelin/MBP expression in tadpoles without altering oligodendrocytes number in the brain stem (**Figure S7**). Together our experiments suggest a conserved function of *RetroMyelin* in regulating Mbp transcription between fish, amphibians and mammals.

DISCUSSION

Generation of myelin by oligodendrocytes is not only important in development, normal physiology[28, 29] and in regeneration following demyelination[30, 31] but has also played a central role in evolution[1-3]. Compacted myelin accelerates nerve conduction in vertebrates and thus facilitates both predatory and escape behaviors[2]. Myelin basic protein (MBP) is essential for the membrane growth and compaction of CNS myelin[2-5]. The myelin sheath of the CNS and the peripheral nervous system (PNS) comprises of many other proteins such as MAL (myelin and lymphocyte protein), MAG (myelin-associated glycoprotein), CNP (Cyclic nucleotide phosphodiesterase), PLP (proteolipid protein, a tetraspan-transmembrane proteolipid), TSPAN2 (tetraspanin 2), PMP22 (peripheral myelin protein 22, a fatty acid-binding protein) and MPZ (Myelin protein zero, P0) (an Ig-like cell adhesion protein). PMP22 and MPZ are produced by the Schwann cells (SC) which myelinate PNS. Additionally, though not a component of myelin sheath *per se*, F-actin promote myelin wrapping by controlling the growth and spreading of the leading edge of myelinating oligodendrocytes[32]. Orthologs of these myelin-associated proteins also exist in non-myelinated jawless vertebrates and

invertebrates and are engaged to perform other functions in those species[3]; however, MBP is only found in jawed vertebrates. Notably, invertebrates and jawless vertebrates also harbour a vertebrate myelin equivalent structure called multilayered ensheathment where the axons are wrapped by glia and this improves the speed and precision of impulse propagation by preventing unwanted electrical crosstalk between adjacent axons[3, 33]. However, the fundamental difference between the glial ensheathment and the jawed vertebrates myelin is its compaction. In comparison with the multilayered glial sheaths in invertebrates, the compact myelin sheaths in jawed vertebrates provide added advantages of rapid conduction velocity while maintaining relatively smaller axonal diameter[3]. MBP is strongly basic in nature at physiological pH[34, 3]. This assists MBP to interact with the negatively charged membrane phospholipids and pull together the adjacent plasma membranes like a 'zipper'[3]. MBP is also an intrinsically disordered polypeptide and has undergone phase transition to form self-assembled polymeric 'sieve' that restricts the diffusion of membrane bound proteins into compact myelin[34-36]. Together, MBP is essential for the myelin compaction and its control[34, 35]. The open reading frame (ORF) of Mbp exists as a part in the transcriptional unit that consist of an ancient gene Golli (gene-of-the-oligodendrocyte-lineages)[3, 37]. Golli also exists in lamprey and its function is unknown[37]. GOLLI is neither the homolog of MBP, nor the element of the myelin sheath or needed for myelination[3, 37, 38]. Thus, the de novo appearance of the ORF of Mbp was an essential addition to the jawed vertebrate genome in the evolution of compact myelin.

In our analyses, we found that retroviral originated RNLTR12-int-like sequence, that we called *RetroMyelin,* exists in all jawed vertebrates where myelination/MBP is evident. Interestingly, *RetroMyelin* sequence copies display high sequence similarity within (but not between) individual species and form distinct clusters specific to each species in a sequence phylogenetic tree. This suggests that their acquisition most likely resulted from independent invasions of similar retroviruses after speciation rather than from a shared ancestral root. Alternatively, what would we expect to see if these sequence copies originated from a common ancestor? If an invasion event occurred at the common ancestral stage, these sequences would have undergone divergence and fixation during that period; otherwise, they risk being lost from the population. Following speciation, each species would have inherited copies from this common ancestor and potentially acquiring further specific mutations.

Observing a sequence phylogenetic tree then reveals that orthologous copies from different species tend to cluster closer together, seemingly intermixing rather than maintaining unique species-specific clusters. While myelination requires numerous crucial genes inherited by vertebrate species from a common ancestor, *RetroMyelin* was likely independently captured post-speciation and adapted potentially due to more effectively regulating these genes, thus offering selective advantages.

We investigated whether any molecular relationships exist between RNLTR12-int and MBP. We demonstrated that RNLTR12-int is critically required for the transcriptional regulation of the Mbp gene, therefore myelin basic protein expression is dependent on *RetroMyelin*. RNLTR12-int encoded RNA binds to the transcription factor SOX10 and this binding is vital for the SOX10 mediated transcription of Mbp. Although the detailed mechanism of transcriptional regulation mediated by *RetroMyelin* and SOX10 is out of the scope of the present study, our data suggests that RNA can act as a transcriptional activator and further support the emerging view of the functional diversity of RNA[20].

Did myelin first emerge in the PNS or CNS? It is proposed that myelination have commenced simultaneously in both systems during evolution[39]. However, the initial form of myelination in the CNS may have been similar to that of the PNS since Schwann cells (SC) can be derived from CNS progenitors and exhibit the capacity to myelinate CNS axons[40] and zebrafish CNS proteome analysis shows many myelin proteins akin to those in the mouse PNS[41]. Further work will be required to resolve this question. Myelin protein zero (P0) is the major constituent of PNS myelin[3] and together with MBP contribute to the formation of major dense lines in the PNS myelin[14]. The intracellular domain of P0 contains basic amino acids that likely perform a similar function to MBP, given that the PNS myelin of Mbp-deficient "shiverer" mice is essentially normal[42, 14]. Here we show that both P0 and Mbp expression in Schwann cells (SC) are dependent on RNLTR12-int. It is possible that initiation of the myelination program through *RetroMyelin* occurred at the same time in the PNS and CNS. Future studies are needed for detail mechanistic understanding of the molecular function of RNLTR12-int in SC.

Upon infection of host target cells, the retroviral RNA genome is reverse transcribed and becomes integrated into the host genomic DNA. When this occurs in the germline, there is vertical transmission of retroviral sequence into the host gene pool and the newly integrated

viral genome is called endogenous retrovirus (ERV). The founder ERV can generate further copies. However, their fate, be it fixed in the genome or eliminated, is determined by selection forces (such as natural selection, random genetic drift) acting at the level of host population[7]. In the present study, we propose that in vertebrate species, *RetroMyelin* sequences originate due to germline invasion of ancient retroviruses carrying an RNLTR12-int-like sequence. Our phylogenetic analyses suggest separate invasion events might have taken place after final speciation. In the case of jawless vertebrates and invertebrates, these sequences might have been purged due to lack of fixation and thereby lost from those species. In jawed vertebrates, the diversification of the introduced sequence copies within each host occurred and was eventually exapted to regulate Mbp through convergent evolution. This is reminiscent of the multilayered ensheathment and myelination of axons that originated independently in at least five different clades of bilaterian phylogeny, an example of convergent evolution[3, 43]. It is also noteworthy that syncytin genes that are required for placental development came from ERV via independent event of domestication across diverse mammalian species as well as one species of lizard, another related example of convergent evolution[7, 44-46]. In the present study, we explored three different vertebrate classes (mammalia, amphibia and fish) and demonstrated examples of the function of *RetroMyelin* in regulation of MBP in all three. *RetroMyelin* function in reptiles and birds, the other vertebrate classes, remains to be addressed.

In summary, 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, a critical step in vertebrate evolution. The vertebrate genome contains many extinct ERVs[7], and our study highlights the impact of viral infection on the evolution of myelin within this important sub-phylum which includes humans.

Limitations of the study

RetroMyelin sequences are repetitive elements that exist in multiple copies within a genome, often sharing high sequence similarity or even identical sequences. When sequencing reads align to a *RetroMyelin*, they can also align or map to other copies of the same *RetroMyelin* located elsewhere in the genome. This issue of multiple mapping creates ambiguity, making it challenging to precisely determine the specific genomic location to which a sequencing read

belongs. Consequently, downstream analyses, such as identifying the specific locus expressing a *RetroMyelin*, become complicated. Even with long-read sequencing technologies, the problem of multiple mapping persists, especially when dealing with evolutionarily younger sequences with minimal divergence. This ongoing challenge makes it difficult to pinpoint any specific locus exclusively expressing *RetroMyelin*. Moreover, it is important to consider that *RetroMyelin* might potentially be expressed from multiple loci. As *RetroMyelin* plays a crucial role in myelination, maintaining its expression is essential for a cellular 'fail-safe' mechanism in adverse events. In cases where one locus is compromised, such as due to somatic mutations or stress, other loci may act as buffers, ensuring sufficient levels of transcript expression. Additionally, we acknowledge the limitations in conducting a gain-of-function (GOF) experiment to determine the sufficiency of *RetroMyelin* in producing compact myelin in agnathostomes such as the lamprey. Lamprey do not have Mbp in their genome, and other myelination genes are vested for different functions. Unlike a GOF experiment under controlled laboratory settings, the evolution of a complex trait like myelin is a gradual, intricate process driven by the interplay between genetic, environmental, and selective pressures, necessitating genetic 'fixation' of *RetroMyelin* and the shaping of a new network for myelin production. Therefore, the creation of compact myelin in lamprey through a GOF experiment extends beyond merely introducing *RetroMyelin* and Mbp sequences in a short-term setting. Another limitation is the challenge of experimentally testing *RetroMyelin* function in all jawed vertebrate species. Future studies are needed to gain a detailed understanding of RNLTR12-int's role in Schwann cell biology. Further studies will also elucidate the precise mechanism of the binding interaction between SOX10 and *RetroMyelin*.

AUTHOR CONTRIBUTIONS

Conceptualization: T.G., R.J.M.F.; Formal analysis: T.G., N.G.; Investigation: T.G., C.Z., R.G.A., A.M., E.M., A.F., C.Z.C., P.A., S.E., G.A.G., K.S., SCG; Methodology: T.G., R.G.A., R.J.M.F., D.A.L., B.Z., 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.A.L., N.G., R.G.A., D.H.R, C.Z., I.A., B.Z., K.S., P.A., C.Z.C., A.F., G.A.G., A.M., E.M., S.E.

DECLARATION OF INTERESTS

Authors declare that they have no competing interests.

Main Figure titles and legends

Figure 1. Network analyses of differentially expressed retrotransposons and protein coding genes in oligodendrocytes

(A) Cluster dendrogram and derived modules (color labeled) after WGCNA. Modules are clusters of genes which are highly connected based on co-expression relations of their expression values across OPC and OL (see STAR methods for detail). Probe level meta-analysis of Affymetrix microarray data (GSE245032) was performed and normalized intensity values were used as input in WGCNA (see STAR methods for detail).

(B) Association of modules with oligodendrocytes is determined by WGCNA. * $P<0.0007$ (Bonferroni threshold). P value in parentheses, correlation is written above 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). Up and down: indicate the modular gene expression changes (elevated or repressed 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. Protein coding genes in the network are represented by round shaped nodes and RNLTR12-int is represented by a brown square. Top 5 hubs (genes which have topmost connectivity in the network) are represented by red nodes and the rest nodes are colored blue, Edges (lines) represent connection (based on expression correlation) between two nodes. Red edges: top 3 edges (topmost strong connections with RNLTR12-int).

(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$\pm$SEM, *$P<0.05$, Student's t test (unpaired, two tailed).

(G) Total number of RNLTR12-int (either flanked or not flanked by LTR) per chromosome in rats which were identified by our criteria (see STAR methods) is plotted as a bar diagram. See also Figures S1, S2 and S3.

Figure 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 beta-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$\pm$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$\pm$SEM, ** P<0.0001, Two-way ANOVA.

(C) Effect of the inhibition of RNLTR12-int on RNA level expression of Pdgfra, Cnp and Mbp was determined by RT-qPCR. Data were normalised to Actb. N=3 independent experiments, mean$\pm$SEM, * P<0.05, Student's t test (unpaired, two tailed) with Welch's correction.

(D) Global changes of gene expression due to inhibition of RNLTR12-int in differentiating OPC as determined by RNA-seq (GSE237744). Left: Volcano plot representing differential expression, Up (red dot) and down (green dot) regulated genes. Right: GO-term overrepresented in differentially expressed genes. Padj: adjusted P value (Benjamini-Hochberg FDR after Wald test). Genes involved in myelination (green bars) are downregulated. N=4 (control siRNA), 3 (siRNLTR12-int) independent experiments.

See also Figure S4 and S5.

Figure 3. Inhibition of RNLTR12-int expression affects Mbp expression during developmental myelination in rats

(A) SOX10 driven EGFP (Emerald GFP) construct that carry shmiR where shRNA is embedded into a microRNA (miR-30) cassette. bGHp(A): Bovine growth hormone poly A signal. This construct is SOX10 driven and expresses EGFP. This also produces shRNA that function through RNAi pathway. AAV carrying SOX10-EGFP-shmiR-RNLTR12-int/Scrambled construct were injected into the deep cortex of rat at P1 and brains were harvested at P14.

(B) AAV carrying the above construct were infected into the cultured OPC and allowed to differentiate for 4 days, then RNA was isolated. RT-qPCR analysis of RNLTR12-int revealed its reduced expression in shmiR-RNLTR12-int infected cells. N=3 independent experiments, mean±SEM, * P<0.05, Student's t test (unpaired, two tailed).

(C-F) AAV carrying the above construct (shown in A) were injected into newborn rat brain (at P1). Brains were harvested at P14 for immunofluorescence **(C-E)** or *in situ* (RNAscope) **(F)**.

(C) Top: immunofluorescence analysis of OLIG2 and GFP immunostaining. Bottom left: quantification of OLIG2+GFP+ cells were plotted as a percentage of GFP+DAPI+ cells. Bottom right: quantification of OLIG2+GFP+ cells were plotted as a percentage to OLIG2+ cells.

(D) Left: immunofluorescence analysis of CC1 and GFP immunostaining. Right: quantification of CC1+GFP+ cells among GFP+ cells and represented as a percentage.

(E) Left: Immunofluorescence analysis using antibody to MBP, GFP, OLIG2. Right: MBP intensities were quantified and plotted relative to shmiR-Scrambled infected samples.

(C-E) N=3 different rats, mean±SEM, ** P<0.01, Student's t test (unpaired, two tailed). Scale bar: 22 μm.

(F) Top: RNA *In situ* (using RNAscope technique) analyses of Mbp, Gfp puncta. Scale bar: 11 μm. Bottom: Average number of cells, Gfp and Mbp puncta counts per mm^2 were plotted. N=3 different rats, mean±SEM, * P<0.05, Student's t test (unpaired, two tailed).

(G) AAV carrying the above construct (shown in A) were injected into the corpus callosum of the newborn rat brain (at P6). Left: TEM micrographs from rat brains (P21) injected with shmiR-Scrambled or shmiR-RNLTR12-int. Scale bar: 1μm. Right (Top): scatter plots representing G-ratio in relation to axon diameter. N= 183 (for shmiR-Scrambled) to 203 (shmiR-RNLTR12-int) axons from 3 different rats. The fitted linear regression lines were shown, and the statistical significance (ANCOVA) of their slope and intercepts difference were determined. For slope difference: p>0.05, indicating that the relation between G-ratio and

the axon diameter does not change due to inhibition of RNLTR12-int. For intercept difference: *** $P < 2 \times 10^{-16}$, indicating that axons of a given diameter have a significantly higher G-ratio in shmiR-RNLTR12-int compared to shmiR-Scrambled cases. Right (Bottom): myelinated and unmyelinated axons were counted and plotted as a percentage. N=3 different rats. Mean$\pm$SEM **adjusted p< 0.01, Two-way ANOVA (Bonferroni post-test).

See also Figure S5.

Figure 4. SOX10 binds to RNLTR12-int encoded transcript and this is required for SOX10 binding to Mbp promoter

(A) 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). U1: U1 spliceosomal small nuclear RNA.

(B) SOX10 binds RNLTR12-int as determined by EMSA. Binding reactions contained RNLTR12-int transcript (3.4 nM) and increasing concentrations (0.125-4μM) of SOX10. Complexes were separated on a 0.6 % native agarose gel and detected by GelRed.

(C) Binding kinetics of RNLTR12-int transcript and SOX10, analysed by surface plasmon resonance (SPR). SOX10 protein was sequentially injected at five different concentrations (0.1-1.6μM) over the streptavidin-coated flow cell sensor surface immobilized with 3' end biotin-labelled RNLTR12-int or kept blank. Sensorgram of the single cycle kinetics (SCK) represents the data corrected for non-specific binding to the flow cell surface (grey). The curve fit with 1:1 binding is shown (dark pink).

(D) 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[22]: S2 ('AACAAT') and S1 ('TTCAAA'). Black right arrow after S1: Transcription start site (TSS). (Bottom) PCR amplified immunoprecipitated DNA, 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).

(E) 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.

(F) In OL, only a fraction (0.8%) of SOX10 targets is downregulated upon RNLTR12-int inhibition. Venn diagram represents the overlap between SOX10 targets (ChIP-seq data:

GSE64703) and the genes downregulated due to RNLTR12-int inhibition (RNAseq data: GSE237744). ** P<0.0001, Hypergeometric test. See also Figure S5.

Figure 5. Phylogenetic relationship of RNLTR12-int and Mbp

(A) Common Tree display represents a hierarchical view of evolutionary 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 classes wherever *RetroMyelin* is identified in our study. All identified *RetroMyelin* belong to the ERV1 family, carrying internal sequence of ERV1.

(B) Expressed sequence tags (ESTs) were aligned to the *RetroMyelin* sequence of human, mouse, rats, zebrafish, *Xenopus laevis* and elephant sharks. BLASTN was used to search ESTs. For the displayed ESTs: 0<E-value<2.9e-18 and 100<alignment score<2298. For human, mouse, rat and zebrafish: consensus sequence was used. For elephant shark and *Xenopus laevis*: the best hit obtained after nhmmer search on RNLTR12-int (Table S1, see STAR methods) was used. magenta: alignment score$\geq$100, red: alignment score$\geq$200.

See also Table S1.

Figure 6. Acquisition of *RetroMyelin* in host genome occurred independently after each host's final speciation, testifying convergent evolution.

(A) With a hypothetical example, here we show that how a tree would look like if retroviral element acquisition occurred once before their speciation from the most recent common ancestor. A representation of a hypothetical phylogenetic tree reconstructed from copies of a specific retrotransposon sequence existing in different species. Let us assume that a viral invasion had happened before speciation. The newly invaded proviral genome (X0) generates copies 1 and 2 in a host genome. Subsequently retrotransposition of copy 1 propagated as copies 3 & 4, while copies 5 & 6 were generated from copy 2. 3, 4, 5 and 6 completely lost their capacity to further retrotranspose due to mutations associated with each recombination event. Selective forces acting at the level of host population subsequently purge copies 1 and 2 from the population pool. Speciation then produces species A and B (speciation 1), and from these ancestors C & D and E & F, respectively, are generated (speciation 2). Assume that each species received the versions of the original copy of 3, 4, 5, 6 (represented by a, b, c, d, e, f). A scientist collects samples of two copies (randomly) from each extant species (C, D, E, F) for reconstruction of a phylogenetic tree. In the resulting tree, we can see that the phylogeny of the retrotransposon copies (c3, e3, f3, c4, d4, etc.) does not recapitulate the species tree and the copies sampled from each species are not clustered together.

(B) A hypothetical example [alternative to (A)] showing what a tree would look like if the acquisition of retroviral element occurred after each host's final speciation. Similar to (A), speciation 1 followed by 2 generated species C, D, E and F. Infection (dotted arrows) of retroviruses (V1 to V3) to the host animals (C, D, E, F) then happened. Subsequently, some ERV copies were fixed in the species genome and underwent sequence divergence within the host species over the course of time. We can view the consequences of this incidence in the form of a sequence phylogeny with a common evolutionary origin (red star) somewhere back in time. This means that the ultimately derived sequences (coloured dots) in species (C to F) are of common origin, so they are homologous though not orthologous. Black dotted lines demarcate sequences determined from the three species. A scientist collects two random copies [similar colour coded like (A)] from extant species (C to F) to reconstruct a phylogenetic

tree with these sequences. We can observe that the copies sampled from each species are clustered together in the resulting tree.

(C) Data from our study suggests *RetroMyelin* acquisition occurred after each host's final speciation. Phylogenetic tree (inference based on maximum likelihood and bootstrapping) was reconstructed from the copies of *RetroMyelin* sequences from each species (for detail see STAR methods). Copies of *RetroMyelin* in each species were clustered together, with high bootstrap support values throughout (coloured circles), and particularly at host-species-defining branches.

(D) *RetroMyelin* sequences have common evolutionary origin [like the example shown in (B)], therefore, suitable for evolutionary rate heterogeneity analysis that we present here (see method for detail). Frequency of site-wise relative rates is represented as a histogram. Expected values of relative rate per bin were calculated from the cumulative distribution function of the inferred gamma distribution (the inferred shape parameter (α) = 2.224, standard error = 0.156) and plotted (red dots).

See also STAR methods.

Figure 7. *RetroMyelin* regulates myelination in zebrafish

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).

(A) Automated imaging of control and *RetroMyelin* gRNA injected zebrafish larva at 5 days postfertilization (5dpf). red boxes: locations of zoomed-in example regions in left panel.

(B) no difference is found in larval morphological development, analysed by eye area and body length. Adjusted P value after Two way ANOVA (Benjamini-Hochberg FDR post test), mean±SEM, N=52-60 (control), 19-23 (*RetroMyelin* gRNA) animals;

(C) GFP fluorescence intensity along the dorsal and ventral tracts of the spinal cord was measured and plotted. * adjusted $P < 0.05$, ** adjusted $P < 0.01$, Two way ANOVA (Benjamini-Hochberg FDR post test), mean±SEM, N=60 (control), 23 (*RetroMyelin* gRNA) animals. See also Figure S6, S7.

Supplemental Figure titles and legends

Figure S1. **Genome wide expression of retrotransposons in OPC and OL, related to Figure 1**

(A) Dendrogram representation of OPC and OL samples, showing OPC and OL samples were formed two separate clusters. Average linkage hierarchical clustering (using Euclidean distance matrix) was performed with respect to the intensity levels of probes in Affymetrix GeneChip.

(B) Probe level analysis of differential expression. Differentially expressed retrotransposons probes (red and green dots) and probes for protein coding genes (grey dots). Probes for which the expression difference between OL and OPC is 2-fold up or down and the adjusted P value<0.05 were considered differentially expressed. FC: Fold change. Padj: adjusted p value (Benjamini Hochberg FDR).

(C) Left: Heatmap plot and dendrogram representation of all differentially expressed probes in OL and OPC samples. Right: Differential gene expression (log_2 value) of retrotransposon probes were plotted. UP: induced probes, DOWN: repressed probes. Proportion of different retrotransposon type in induced and repressed probes were plotted in a pie chart.

(D) Probability density plots showing the behavior of retrotransposon probes, non-retrotransposon probes, all probes and differentially expressed retrotransposon probes. Vertical dotted lines: 2-fold change of expression. * P<0.05, Wilcoxon rank sum test.

(E) Expression levels of RNLTR12-int and myelination hubs were represented as heat map and a dendrogram derived after two-way hierarchical clustering.

Figure S2. Validation of RNLTR12-int expression in MACS isolated OPC (A2B5+) and OL (Mog+), related to Figure 1 and STAR methods.

(A) Top: (left) RT-PCR analysis of cytoplasmic beta-actin (Actb) using intron flanking primers. Example shown detection of gDNA contamination in RNA that resulted an upper band. Only single lower band obtained in OPC and OL samples, confirms no gDNA contamination. (right) RT-PCR analysis of RNLTR12-int expression in OPC and OL. PCR product was run on a 2% agarose gel. Bottom: qPCR melting curve analysis resulted only single peak for Actb (left) and RNLTR12-int (right).

(B) Actb expressions remained unchanged in OPC and OL as determined by Affymetrix GeneChip. Normalized intensity of all Actb specific probes (17 probes) were used and plotted relative to OPC. mean$\pm$SEM; N=9 (OPC), 8 (OL), P>0.05, Student's t-test (unpaired, two-tailed).

(C) Expression levels of Mag, Tspan2 and PB1D9 is elevated in OL as compared to OPC determined by RT-qPCR. mean$\pm$SEM; N=3-4, *P<0.05, **P<0.01, Student's t-test (unpaired, two-tailed).

(D) RNLTR12-int encoded transcript is likely to be a non-coding RNA as predicted by CNIT algorithm[58]. Score of a known protein coding gene, TATA-box binding protein (Tbp, Transcript ID: ENSRNOT00000002038.4) is shown.

(E) RNA sequencing reads (GSM3967160) were aligned to RNLTR12-int consensus sequence and the depth was plotted.

Figure S3. **Some examples of RNLTR12-int copies in rat genome, related to Figure 1 and STAR methods.**

Schematic view of some examples of RNLTR12-int in rat genome (rn6) which are flanked (**A**, left) or not flanked (**A**, right) by long terminal repeat (LTR). length in base pair (bp) is written in the middle of the double headed arrow. Genomic location is mentioned below of a vertical dotted line. On each horizontal line the blue colored symbol '>' indicates forward strand, '<' indicates reverse strand. chr: Chromosome. **(B)** Expression levels of Pygl and Ncbp2 in OPC and OL (determined by Affymetrix GeneChip) were plotted relative to OPC samples. N=9 (OPC), 8 (OL), Mean+SEM, P>0.05, Student's t-test (unpaired, two-tailed).

Figure S4. Functional Impact of RNLTR12-int Inhibition on oligodendrocyte Morphology, cell fate, gene expression in Schwann cells and elimination of any off-target effect, related to Figure 2 and STAR methods.

(A) Absence of complex oligodendrocyte morphology due to inhibition of RNLTR12-int. Left: Immunofluorescence analysis of O4 immunostaining 5 days after transfection of siRNA. Scale bar: 26 μm. siRNLTR12-int: siRNA against RNLTR12-int, control siRNA: siGENOME non-targeting siRNA pool (Dharmacon).

Right: Quantification of oligodendrocyte average outgrowth area (O4 immuno-stained) and the total number of cells (O4 immuno-positive) counted were plotted. mean+SEM, N= 3 independent experiments. ** p<0.01, Student's t test with Welch's correction (unpaired, two tailed).

(B) Inhibition of RNLTR12-int did not initiate astrocytic fate. Immunofluorescence analysis of GFAP immunostaining 5 days after transfection of siRNA. Scale bar: 60μm. Percentage of GFAP+ cell in total DAPI+ cells were plotted. N=3 independent experiments (each time 3 replicates), mean+SEM, p>0.05, Two-way ANOVA. siRNLTR12-int: siRNA against RNLTR12-int, control siRNA: siGENOME non-targeting siRNA pool (Dharmacon).

(C) Effect of inhibition of RNLTR12-int on RNA level expression of P0 and Mbp in Schwann Cells (SC). RNA level expressions were determined by RT-qPCR. Upon transfection [siRNLTR12-int: siRNA against RNLTR12-int or control siRNA: siGENOME non-targeting siRNA pool (Dharmacon)], SC were maintained in differentiation media for 4 days and RNA was harvested

for RT-qPCR. N=3 independent experiments (each time 2-3 replicates), mean$\pm$SEM, ** P<0.01, * P<0.05, Two-way ANOVA.

(D) Exclusion of any off-target effect. siRNA-mediated inhibition of RNLTR12-int in differentiating OPC does not alter expression of genes whose intronic region (either in the same or the opposite strand) contains fragment of RNLTR12-int. Normalized values from RNA-seq data plotted, mean$\pm$SEM, ns: P>0.05, Benjamini-Hochberg FDR after Wald test.

Figure S5. Effects of RNLTR12-int modulation on oligodendrocyte differentiation, myelination and gene regulation, related to Figure 2, Figure 3 and Figure 4

(A) Overexpression of RNLTR12-int restores MBP expression in RNLTR12-int siRNA silenced differentiating OPCs. pcDNA 3.1 (+) bearing a full length RNLTR12-int consensus sequence or null vector was transfected 3 days after siRNA administration (schema above). Representative immune-stained image (left) and quantification of MBP+ oligodendrocytes (right) shown. * adjusted P < 0.05, Brown-Forsythe and Welch's ANOVA (Benjamini-Hochberg FDR post test), mean+SEM, N=3 independent experiments. Scale bar: 45µm.

(B) CNPase expression is affected due to inhibition of RNLTR12-int during developmental myelination in rats. AAV carrying the construct (shown in **Figure 3A**) were injected into newborn rat brain (at P1). Brains were harvested at P14 for immunofluorescence.

Left: Immunofluorescence analysis using antibody to CNPase, GFP, OLIG2. Right: CNPase intensities were quantified and plotted relative to shmiR-Scrambled infected samples.

N=3 different rats, mean$\pm$SEM, * P<0.05, Student's t test (unpaired, two tailed). Scale bar: 30 µm.

(**C-D**) AAV carrying the construct (shown in **Figure 3A**) were injected to the corpus callosum at P6 rats and brains were harvested at P21 for electron microscopy (TEM).

(**C**) G-ratios but not the axonal diameters differ due to inhibition of RNLTR12-int. G-ratios (Top) and axonal diameters (Bottom) were plotted. N= 183 (for shmiR-Scrambled) to 203 (shmiR-RNLTR12-int) axons from 3 different rats. mean$\pm$SEM, ** P<0.0001, Student's t test (unpaired, two tailed) with Welch's correction.

(**D**) Periodicity remained similar due to RNLTR12-int inhibition. TEM images shown (Top). Scale bar: 20 nm. Bottom: Average periodicity values were plotted. N= 35-36 myelin sheath

images were used from 3 different rats. For each image an average of 2-5 different measurements were taken. mean$\pm$SEM, P>0.05, Student's t test (unpaired, two tailed).

(E) SOX10 occupancy to Myrf promoter is affected due to inhibition of RNLTR12-int as determined by ChIP analysis 4 days after siRNA transfection. Left: SOX10 binding conserved region (ECR9)[24] in the intron 1 of Myrf gene was PCR amplified [amplified region: chr1: 233229291-233229458 (rn5), strand: (-)] from the immunoprecipitated DNA and run on 2% agarose gel. Right: qPCR analysis after ChIP. N=3 independent experiments. mean$\pm$SEM, * P<0.05, Ratio paired t-test (two-sided).

Figure S6. **Assessing CRISPR/cas9 mediated knockout in zebrafish injected with *RetroMyelin* gRNA, related to Figure 7 and STAR methods.**

(A-B) Example chromatograms from Sanger sequencing of a PCR product amplified from genomic DNA extracted from a control-injected larva (A) or *RetroMyelin* gRNA-injected larva (B) Predicted genomic sequence indicated on top, and gRNA target and PAM sequence indicated by red box and arrow. Below the traces are predicted alleles identified from the chromatograms using CRISP-ID[88], showing only few SNPs in control, but multiple indels and variations in the gRNA-injected larva. **(C-D)** Alignments of the sequencing results for PCR products amplified from genomic DNA extracted from several control (C) or *RetroMyelin* gRNA-injected (D) larvae to the predicted genomic sequence (top, gRNA sequence in blue and predicted cut site in red). Note widespread mismatches in *RetroMyelin* gRNA-injected traces, indicating gRNA activity.

Figure S7. *RetroMyelin* regulates myelination in frog, related to Figure 7

(A) Analysis of overlapping peaks arising from control (left) or *RetroMyelin* gRNAs-injected Xenopus (right) using CRISP-ID[88]. Shown is representative chromatogram from Sanger sequencing of PCR amplicons from genomic DNA extracted from control (left) or *RetroMyelin* gRNA-injected *Xenopus* tadpole (right). Alignments of sequences were generated from mixture sequences present in the same animal and compared to the reference sequence (Ref) for the target region. The direction of sgRNA binding site shown in red arrow. CCC (surround by a red rectangle) is the PAM sequence. Three sequences are detected in the injected animal, a: wt sequence, b: sequence with nucleotides mutations around the sgRNA target sequence, c: sequence with nucleotides mutations around the sgRNA and 5 bp nucleotides deletions).

(B) Longitudinal Z-series of wholemount across the brain stem of Tg(mbp:gfp-ntr) control (upper panel) and *RetroMyelin* gRNAs-injected (RNLTR12-/-) (lower panel) Xenopus tadpoles triply labeled for GFP (green), MBP (red) and XMM (white) [XMM (*Xenopus* myelin marker): mAb recognizes specifically an unknown *Xenopus* myelin protein[89]]. scale bar= 23μm. Quantification of oligodendrocytes (GFP+ cells) was plotted. P>0.05, Mann Whitney test (Two-tailed), N=4 (control), 5 (*RetroMyelin* gRNA) different sections, mean±SEM. Intensity of XMM and MBP were quantified and plotted relative to control. * P<0.05, Student's t test (unpaired, two tailed), N=4 (control), 5 (*RetroMyelin* gRNA) different sections.

STAR*METHODS

Detailed methods include the following:

- KEY RESOURCES TABLE

- RESOURCE AVAILABILITY

  o Lead contact

  o Materials availability

  o Data and code availability

- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

  o Rats

  o Isolation of OPC and OL

  o Culture and differentiation of OPC

  o Schwann cells isolation and culture

  o Zebrafish

  o Frog

- METHOD DETAILS

  o Retrotransposable element annotation from Affymetrix microarray

  o Probe level meta-analyses of Affymetrix microarray data

  o Weighted gene co-expression network analysis (WGCNA)

  o Overrepresentation analysis

  o Coding potential determination

  o Aligning RNA-seq reads to RNLTR12-int consensus sequence

  o RNA-seq with siRNLTR12-int

  o RNLTR12-int copies in rat genome

  o Identification of RNLTR12-int like sequence (*RetroMyelin*)

  o Common Tree display

  o Phylogenetic analysis

  o Rate heterogeneity analysis

  o EST mapping

  o Transfection

  o shmiR cloning and AAV PHP.eB packaging

  o In vitro and in vivo infection

- o Immunofluorescence staining

- o RNA In situ (RNAscope)

- o qRT-PCR analysis

- o RNA immunoprecipitation (RIP)

- o Chromatin-immuno-precipitation (ChIP)

- o Comparison of gene list from ChIP-seq data with RNAseq data

- o Cloning, in vitro transcription (IVT) and labelling of RNA

- o Electrophoretic mobility shift assay (EMSA)

- o Surface plasmon resonance (SPR) experiment

- o Electron microscopy

- o Zebrafish experiment

- o Frog (*Xenopus laevis*) experiment

- QUANTIFICATIION AND STATISTICAL ANALYSIS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Robin JM Franklin (rfranklin@altoslabs.com).

Materials availability

Any stable/unique materials used in this study are available by request from the lead contact.

Data and code availability

Meta-analysis of microarray data and RNA-seq data has been deposited to Gene expression omnibus (GEO) under the accession number GSE245032 and GSE237744 respectively and are publicly available as of the date of publication. ChIP-seq data (GSE64703) and RNA-seq data (GSM3967160) used in this study were obtained from GEO.

All codes are deposited in GitHub: https://github.com/Tanay-math/RetroMyelin and are publicly available as of the date of publication.

Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Rats

Sprague–Dawley rats were received from C. River (Margate, UK). Rats were maintained in individually vented cages at temperature 22°C ± 1°C and humidity 60% ± 5%, in a 12 hr light:dark cycle; food and water were supplied ad libitum in a standard facility for rodents at the University of Cambridge, UK. All animal studies were conducted under the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB).

Isolation of OPC and OL

Postnatal day 7 (P7) rats were sacrificed using an overdose of pentobarbital and whole brain was dissected immediately in ice cold hibernate A low fluorescence (HALF) medium[47, 48]. Subsequently OPC was isolated as described earlier[47]. Per 10 million cell suspension, we used 2.5 μg mouse-anti-rat-A2B5-IgM antibody (Millipore; MAB312) and 20 μl of rat-anti-mouse-IgM antibody magnetic beads (Miltenyi, 130-047-302) for OPC isolation by magnetic-activated cell sorting (MACS). OLs were isolation from A2B5 negative fraction by using 2.5 μg of goat-anti-mouse-MOG-Biotinylated (RD systems, BAF2439) and 20μl mouse-anti-biotin magnetic micro beads (Miltenyi, 130-105-637).

Isolated OPC/OL were immediately placed in cell lysis buffer from mirVana (Ambion) RNA isolation kit. Otherwise, OPCs were seeded in 24-well-plate or 6-well-plate with poly-d-lysine (PDL) coated wells.

Culture and differentiation of OPC

OPCs were maintained in OPC medium[47, 48] [supplemented with 30 ng ml^{-1} basic fibroblast growth factor (bFGF) (Peprotech, 100-18B) and 30 ng ml^{-1} platelet-derived growth factor (PDGF) (Peprotech, 100-13A)] and 60μg ml^{-1} N-Acetyl cysteine (Sigma, A7250) in an incubator at 37°C and 5% CO_2 and 5% O_2. Medium was replaced in alternative days. OPC medium (100 ml) composition: A total volume of 100 ml of Dulbecco's modified Eagle's medium (DMEM)/F12 (Thermo Fisher, 11039-021) is supplemented with 1mM sodium pyruvate (Thermo Fisher, 11360-070), 5 mg apo-transferrin (Sigma, T2036), 1 ml SATO stock solution [described earlier (51)], 10μg ml^{-1} human recombinant insulin (GIBCO, 12585014).

For differentiation of OPC to OL: OPC medium was supplemented with 40 ng ml^{-1} 3,3',5-Triiodo-L-thyronine (T3) (Sigma; T6397)[47, 48]. Medium was replaced on alternate days for 5 days.

Schwann cell isolation and culture

Freshly plated Schwann cells were obtained from the rat sciatic nerves at postnatal day 2-4 (P2-4) as described earlier[49, 50]. Briefly, nerves were digested with an enzyme mixture containing trypsin (1.25 mg ml^{-1}) and collagenase (2 mg ml^{-1}) for 45 min at 37°C. Subsequently, the nerves were triturated with a P1000 pipette, then P200 pipette. The cells were pelleted by centrifugation at 300 g for 5 min at room temperature, then plated on Poly-L-

lysine/laminin coated dishes. Immediately following isolation, the Schwann cells were grown in DMEM + 5% horse serum, supplemented with the antimitotic agent AraC (1 mM) for 3 days to eliminate fibroblasts. After 3 days, Schwann cells were replated, then maintained in Schwann cell defined medium as previously described[49, 50].

Zebrafish

Zebrafish (transgenic myelin reporter line: Tg(mbp:EGFP-CAAX)ue2[51]) were maintained in the University of Edinburgh BVS Aquatics facility under project license PP5258250.

Frog

Xenopus laevia were raised in the animal facility of Institut du Cerveau – ICM, Paris, France (agreement # A75-13-19) under the project license approved by the ethical committee of the French Ministry of Higher Education and Research (APAFIS#5842-2016101312021965).

METHOD DETAILS

Retrotransposable element annotation from Affymetrix microarray

Retrotransposon probes in Affymetrix Rat Genome 230 2.0 were identified and annotated as described previously[52]. Affymetrix Rat Genome 230 2.0 contains 31,000 probesets. Probe sequence were used for BLAT search in the rat genome (rn4, Nov. 2004, version 3.4) and multimapping probes were identified. Genomic co-ordinates of these probes were then compared with Repeatmasked region (rn4 release rat genome) by querying repeat masker database[53] and annotation (class, family and element) were tabulated. Annotation file is available in GitHub: https://github.com/Tanay-math/RetroMyelin.

Probe level analyses of Affymetrix microarray data

Raw data with accession number GSE11218 and GSE5940 (using Affymetrix expression GeneChip Rat Genome 230 2.0 and 230A, respectively) were extracted from Gene expression omnibus (GEO). Raw data from both the platform comprise 9 OPC and 8 OL samples. Probe level information was extracted using R library rat2302probe (for Rat genome 230 2.0) and rae230aprobe (for Rat genome 230A). Unique probe sequence in 230 2.0 and 230A platform were retained and the intersect probes (90507 in number) between two platforms were used

for analysis. Background subtracted raw data were quantile normalized. Probes for which intensity were greater than sample median level in at least half the arrays for one experimental condition in a dataset were considered present[52]; absent probes were removed. Further differentially expressed probes between OL and OPC were identified by using R Bioconductor package limma[54]. P-values were adjusted to correct for multiple testing using FDR (Benjamini-Hochberg). Probes for which the difference between groups were $\geq$ 2 fold and the adjusted P-value < 0.05 were considered differentially expressed. This meta-analysis has been deposited to GEO with accession number: GSE245032.

Weighted gene co-expression network analysis (WGCNA)

WGCNA was performed by using functions[55] available as a R package (1.20 library) installer. Because each gene is represented by multiple probes in microarray, aggregate[56] function was used to collapse probe with median expression. Pearson correlation matrix (preserving the sign of the correlation coefficient) of pairwise comparison of gene expression levels was generated and transformed to an adjacency matrix by using soft threshold power=16 for fitting with approximate scale free topology; thereafter a topological overlap matrix (TOM) was generated. Topological overlap (TO) based dissimilarity (1-TO) measure was used for generating average linkage hierarchical clustering and cutreeDynamic function was used to identify modules. Module eigengene (ME) (i.e. 1st principal component obtained after singular value decomposition) is a summary of the gene expression levels of all genes in a module. Binary numeric variables called 'oligodendrocytes' was generated to define OPC (=0) and OL (=1). Correlation co-efficients (Pearson) between MEs with oligodendrocytes and p-values (Student asymptotic) of corresponding correlations were obtained by using WGCNA functions. Bonferroni threshold level of significance (0.01/14 = 0.00071) was set based on the number of modules (=14).

Overrepresentation analysis

DAVID version 6.8 (the Database for Annotation, Visualization and Integrated Discovery)[57] functional annotation tool was used for analyzing GO (Gene ontology) terms overrepresented in WGCNA modules. Benjamini-Hochberg adjusted p value <0.05 was considered significant.

Coding potential determination

Coding and non-coding potential was evaluated using CNIT (Coding-Non-Coding Identifying Tool) program[58].

Aligning RNA-seq reads to RNLTR12-int consensus sequence

RNAseq raw data from OPC (GSM3967160) were pre-processed through sortMeRNA[59] to filter out reads from ribosomal RNA, adapter was trimmed, and quality paired reads were extracted using Trimmomatics[60]. RNLTR12-int consensus sequence (from Repbase-GIRI) was indexed and reads alignment to RNLTR12-int was performed by STAR_2.5.2b[61]. Subsequently sorted bam file was generated, and depth was calculated using samtools[62].

RNA-seq with siRNLTR12-int

OPC were isolated from P7 rats and maintained in culture media (see the section: Culture and differentiation of OPC). Approx. 40,000 cells were plated in a 12 well plate. Subsequently transfection was performed (see the section: Transfection) and cells were kept at differentiating media for 4-5 days. 4 (control) to 3 (siRNLTR12-int) independent biological replicates were prepared (each time 4-5 P7 rat brains were pooled for OPC isolation). RNA was isolated using mirVana RNA isolation kit (Ambion). Subsequently, DNAase treatment and purification was performed and RNA quality was verified in a bioanalyzer before used for RNAseq library preparation. RIN values were: 9-10. Qiaseq Fast Select – rRNA HMR (Cat. No. / ID: 334385) was used for ribodepletion and paired end directional library was prepared using NEB NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (E7760L), and run on Novaseq SP.

As mentioned above, data were pre-processed through sortMeRNA, trimmomatics. Rat transcriptome (Rnor_6.0.cdna+Rnor_6.0.ncrna) index was build and quantification was performed using Salmon[63] (Quasi mode). Differential expression was analysed using DEseq2 package[64]. Criteria for differential expression: adjusted P value (Wald test followed by Benjamini-Hochberg FDR)<0.05 and |fold change|>2.
Accession number of this data set is GSE237744. A list of differentially expressed genes and their normalized values can be found in GSE237744.

Off-target identification: A 125 bp region which is flanked by the qPCR primers to amplify RNLTR12-int (that region also includes our siRNA target) was used as an input in BLAT (using rat genome). Subsequently annotation was extracted for all the BLAT hits using HOMMER tools. A total of 19 genes were found whose intronic region contains a fragment of RNLTR12-int (either in the same or in the opposite strand), none were found in the exonic region. Differential expression (siRNLTR12-int vs control) of these 19 genes were extracted from the analysis above.

RNLTR12-int copies in rat genome

Repeatmasker annotation of repeats in the rat genome (rn6) were parsed through home written perl scripts. The following categories were used to identify a) RNLTR12-int sequence that is flanked by LTR: LTR to LTR total distance < 9 kb. At least 75% of the sequence flanked by the LTR should be repeatmasked. b) RNLTR12-int is not flanked by LTR: if RNLTR12-int fragmented sequences were separated by small insertions or other repeats up to 1 kb, they were assembled; and the whole contig size would be < 9 kb. In both (a) and (b): all features should be in the same chromosome and in same orientation.

Identification of RNLTR12-int like sequence (*RetroMyelin*)

We employed the remote homology search algorithm nhmmer[65], a probabilistic inference method based on a hidden Markov models, and used the RNLTR12-int consensus sequence (Repbase-GIRI) as an input to search for *RetroMyelin* in other genomes. The E-value threshold for inclusion was <0.01. Co-ordinates of the top hits were then queried using RepeatMasker[51] to extract repeat annotation. Also, RepeatMasker[53] was run (search engine: rmblast/hmmer/cross_match and selecting model organism as a DNA source) independently using the sequence of the top hits as an input to identify as well as to verify repeat type and family. Identified repeats which are just a simple low-complexity repeat (e.g. $(TC)_n$) are not considered as *RetroMyelin* (**Table S1**).

Common Tree display

Using the taxonomic IDs of the subset of organisms as an input and by employing the Common Tree tool from NCBI, a hierarchical representation of the relationships among the taxa and their lineages was generated. The Phylip-format tree file was then visualized using iTOL[66].

Phylogenetic analysis

Sample collection: Top hit sequence of the identified *RetroMyelin* in different animals were used as an input for BLAT. For each BLAT result, the top 5 five highly similar sequences as well as 5 randomly selected sequences were collected. However, we do not consider sequences showing the following features in BLAT: a) the sequence alignment is such that only a small fragment of the full-length input sequence (e.g. 153 nt small fragment out of 1270 nt full length input) is aligned and/or b) alignment is such that a small region is aligned to the genome followed by a long stretch of unaligned region, which is then followed by another small region that is aligned. Such sequences are so fragmented that they would not carry any functional copy. Furthermore, if sequences were not identified as ERV1-int by RepeatMasker[53] run (search engine: rmblast/hmmer/cross_match and selecting model organism as a DNA source) then they were not considered further. In some animals we can not find 10 functional hits and therefore we collected samples from as many of them as were available.

Alignment and phylogenetic tree: MAFFT version 7 (a fast Fourier transform based multiple sequence alignment program)[67] (keeping mafft flavour=auto) is employed to align *RetroMyelin* sequences. The alignment was curated using the BMGE (Block mapping and gathering with entropy) algorithm[68]. RAxML (RAxML-HPC BlackBox version 8.2.12)[69] was employed for phylogenetic tree inference with maximum-likelihood and bootstrapping. Following the MRE-based Bootstopping criterion[70], RAxML performed a rapid bootstrap search and collected 500 replicates. Trees were visualized in iTOL[66].

Rate heterogeneity analysis

Rate heterogeneity analysis was performed using BASEML (in PAML version 4.8a)[71]. This analysis is suitable for any sequences that have a common evolutionary origin[72] (see **Figure 6B**), and uses the phylogeny that describes their relationships in order to detect if there is variation in rates over sites where each site is assumed to have a constant rate over the tree, but potentially different from other sites. We used the REV+G (GTR+G) model with the assumption of the variation of evolutionary rates over sites following a gamma distribution. 4 gamma distributed rate categories were used. The parameters in the rate matrix (REV) were

as described[73, 74]. The shape parameter (α) of the gamma distribution of rates over sites was estimated. H0 (no rate heterogeneity i.e., α→∞) was compared with H1 (gamma distributed rate heterogeneity) using a likelihood ratio test and the 2δ test statistics as described[27]. Following rejection of H0, site wise empirical Bayesian (posterior mean) relative rate estimates were computed as described[75].

EST mapping

Input nucleotide sequences were individually queried to expressed sequence tags (ESTs) database by using NCBI blastn suit. ESTs for which $0 \leq$ E-value $\leq$ 6e-26 and $119 \leq$ alignment score $\leq$ 2298, were mapped to the RNLTR12-int like sequences from the following animals: consensus sequence of Human (Dfam ID: DF0000205.5), mouse (Dfam ID: DF0001916.1), rat (RNLTR12-int from Repbase), zebrafish (Dfam ID: DF0002922.2). Consensus sequence was not available for elephant sharks, so we mapped to the genomic location obtained after nhmmer i.e. KI636671.1: 5058-5765 (Callorhinchus_milii-6.1.3).

Transfection

Lipofectamine™ RNAiMAX transfection reagent (ThermoFisher scientific) was used and followed the manufacturer protocol. 48 hours before transfection MACS isolated OPC (approx. 25,000) were seeded per well of a 24-well-plate (PDL coated) and cells were maintained in proliferation media. Before transfection media was replaced by differentiating media. 1 µl Lipofectamine™ RNAiMAX and 30pmol siRNA was used for transfection. siRNA against RNLTR12-int (Dharmacon) and control siRNA (siGENOME Non-Targeting siRNA pool #2) were procured from Dharmacon. 48 hours after transfection cells were harvested for RNA isolation. 5 days after transfection cells were fixed for immunofluorescence analyses.
siRNA sequence against RNLTR12-int: 5' AAGUGAGGGCCUUCUAUGCUU 3'.

Transfection of siRNA against RNLTR12-int and control siRNA in Schwann cells was performed as described above. Before transfection Schwann cell defined media was supplemented with 2mM dibutyryl cyclic adenosine monophosphate (dbcAMP) and 20 ng ml^{-1} neuregulin 1 (Ngr1) to induce differentiation. Media was replaced on alternate days and cells were harvested 4 days after transfection.

shmiR cloning and AAV PHP.eB packaging

SOX10 Multiple Species Conserved enhancer element 5 conjugated with cfos basal promoter (https://www.addgene.org/115783/)[76] sequence was used for oligodendrocyte lineage specific expression of EGFP-shmiR with a bGHp(A) poly A signal (**Figure 3A**). This SOX10 driven entire sequence (**Figure 3A**) was cloned between the inverted terminal repeat of adeno associated virus 2 (AAV2ITR) vector. Cloning and large scale AAV PHP.eB packaging of the construct and purification was performed from AMS Biotechnology (Europe) Limited, UK. The viral titre was 10^{13} GCml^{-1}. The following sequences were embedded into the miR-30 backbone of the shmiR vector:

shmiR-RNLTR12-int:

5' ATAGAAGGCCCTCACTTTTA[GTGAAGCCACAGATG]TAAAAGTGAGGGCCTTCTATGC 3'

shmiR-Scrambled:

5' TGGACGCTCGATTGATCATA[GTGAAGCCACAGATG]TATGATCAATCGAGCGTCCAAT 3'

The loop sequence is indicated by parenthesis.

***In vitro* and *in vivo* infection**

For *in vitro* study, OPC was isolated from rat pups using MACS (above) and approximately 30,000 cells were seeded to each well of a 24-well-plate, 48 hours before infection. Media was replaced to differentiating media (above) before infection. Approximately 5×10^{10} GC of virus was used per well. 4 days after infection total RNA was isolated using the mirVana kit (Ambion), DNase I (ThermoFisher scientific) treated before reverse transcription.

For *in vivo* experiment, 1µl of virus suspension (approximately 1×10^{10} GC) was injected stereotaxically into the cerebral cortex at 0.1mm anterior, 1.2mm lateral and 1.2mm deep with reference to Bregma under isoflurane anaesthesia, in neonatal rats at P1.

Immunofluorescence staining

Cells were fixed with 4% PFA (10 min, room temperature), washed with PBS and proceeded for immunofluorescence staining as described[48]. For *in vivo* experiment, animals (at P14) were perfused using 4% PFA in PBS, brains were postfixed overnight with 4% PFA, cryoprotected with 20% sucrose, and embedded in OCT-medium (TissueTek). Immunofluorescence was

performed on 12 µm cryostat section and closely followed the protocol as described earlier[77]. For CNPase immunostaining antigen retrieval was performed at 98°C for 5-10 min in a citrate buffer (10mM citrate buffer pH 6.0, 0.05% Tween).

The following antibodies were used in this study:

Primary antibody	Class	Host species	Dilution	Source	Catalogue number
Anti-MBP	IgG	Rat	1 to 500 or 1 to 100	Serotec	MCA409S
Anti-OLIG2	IgG	Rabbit	1 to 1000 or 1 to 100	Millipore	AB9610
Anti-GFAP	IgY	Chicken	1 to 1000	Abcam	ab4674
Anti-O4	IgM	Mouse	1 to 1000	R&D Systems	MAB1326
Anti-CC1	IgG2b	Mouse	1 to 100	Merck/milipore	OP80
Anti-CNPase	IgG2b	Mouse	1 to 100	Atlas	AMAb91072
Anti-GFP	IgY	Chicken	1 to 100	Abcam	ab13970
Alexa Fluor 488 Anti-Mouse	IgM	Goat	1 to 1000	Invitrogen	A21042
Alexa Fluor 488 Anti-Chicken	IgY (IgG) (H+L)	Donkey	1:200	Jackson Immuno Research	703-546-155
Cy™3 AffiniPure Anti-Rat	IgG (H+L)	Donkey	1 to 1000 or 1 to 500	Jackson Immuno Research	712-165-150
Alexa Fluor 647 Anti-Rabbit	IgG	Donkey	1 to 1000 or 1 to 500 or 1 to 200	Invitrogen	A31573
Alexa Fluor 488 Anti-Chicken	IgG (H+L)	Goat	1 to 1000 or 1 to 500	Invitrogen	A11039
Alexa Fluor-568-Anti-mouse	IgG2b	Goat	1 to 500	Invitrogen	A21144
Alexa Fluor-555-Anti-Mouse	IgG	Donkey	1 to 200	Invitrogen	A31570

Images were acquired with an SP5 (Leica TCS) or SP8 (Leica TCS) confocal microscope. All acquisition settings were kept constant for the whole experiment. Images were processed

using the ImageJ version 2.1.0 software[78]. For *in vivo* immunoassayed sections, integrated density of MBP/CNP immunostaining was measured using ImageJ 2.1.0. Per animal, 4-7 images from the adjacent brain sections were used to calculate the average intensity and 3 animals per group were analysed. Cell counting was performed using Cell Counter tool in ImageJ 2.1.0. O4 immuno-positive outgrowth area were measured using ImageJ[78] analysed particle function (Circularity 0.0-1.0) and the total O4+ area was divided by the total number of O4+ cells to obtain the average outgrowth area[79]. Per experiment 4-6 images from the 3 different wells were used and 3 independent experiments per group were analysed. Immunofluorescence experiments were performed and analysed blind to treatment.

RNA In situ (RNAscope)

In situ hybridization was performed on 12 μm cryostat section using RNAscope® fluorescent multiplex Assay [as described by the manufacturer's protocol: Cat#320850, ACD, Bio-Techne] on the automated BOND RX system (Leica). The following predesigned probes were used: Gfp: RNAscope® 2.5 LS Probe - EGFP-O4 (Cat#538858, ACD, Bio-Techne Ltd), Mbp: RNAscope® 2.5 LS Probe - Rn-Mbp-O1-C2 (Cat#1106048-C2, ACD, Bio-Techne Ltd), RNAscope 2.5 LS Negative Control Probe_dapB (Cat#312038). Slides were stained with DAPI, and probes of interest were stained with fluorophores Opal 540 (for Gfp) and Opal 650 (for Mbp). Images were taken in a stellaris 8 confocal microscope (Leica), and images were analysed using ImageJ version 2.1.0 software[78]. Per animal 3-5 adjacent brain sections were imaged and the average cell numbers, number of puncta (Gfp, Mbp) were calculated. 3 animals per group were analysed.

qRT-PCR analysis

Total RNA was isolated using the mirVana kit (Ambion). Total RNA was DNase I (ThermoFisher scientific) treated before reverse transcription. cDNA was generated by using random hexamer (Invitrogen) and SuperScript™ IV Reverse Transcriptase (Invitrogen™) following manufacturer protocol. Real time PCR was performed using SYBR green master mix (Applied Biosystems, ABI) in QuantStudio 7 Flex (Applied Biosystems, ABI) machine. The real time PCR program was 50°C/ 2 min, 95°C / 10 min., 40 cycles 95°C/ 15 s, 60°C/ 1min. Dissociation protocol was run.

Rat specific primer sequences used in qPCR:

Primer	Sequence (5' to 3')	Amplicon (bp)
RNLTR12-int	F: CCACTGAGGAAAGACGGAAT	125
	R: ACCTGGAGCACTCCCTACCT	
Actb	F: CACCATGTACCCAGGCATTG	111
	R: CACACAGAGTACTTGCGCTC	
Mbp	F: ATCGGCTCACAAGGGATTCA	108
	R: CGTCTTGCCATGGGAGATCC	
Tspan2	F: TTAAGCTCCAGCTCATTGGAA	101
	R: TGAGTTCCGTATTGCACAGC	
Mag	F: CCTGGCAGAGAATGCCTATG	115
	R: GACTGTCTCCCCCTCTACCG	
PB1D9	F: CAAGAGGCAGAGGCAAACAC	81
	R:TTTGGTTTTTAGATGTATAAGGTCTCA	
Pdgfra	F: AAGATCTGTGACTTCGGGCT	106
	R: AAATGCTCTCAGGTGCCATC	
CNP1	F: GTGCTGCACTGTACAACCAA	111
	R: GGACAGTTTGAAGGCCTTGC	
Mpz (P0)	F: CGTCTCAGGTCACGCTCTAT	114
	R: AGAAGAGCAACAGCAGCAAC	
U1 snRNA	F: TTTTCCCAGGGCGAGGCTTA	100
	R: CCCCACTACCACAAATTATGCA	

RNLTR12-int sequence is located in multiple chromosomes which can be identified using Affymetrix probe set (1379497_at). We did BLAT (from UCSC) and designed the primers from

the top hit (chr7:82148213-82148189, assembly: rn5), which was one among many that showed 100% identity. Since RNLTR12-int is intron-less, we excluded the possibility of trace genomic DNA (gDNA) in our RNA samples by treating with DNase I and designing intron flanking primers targeting cytoplasmic beta actin (Actb) gene to detect gDNA contamination. We observed only cDNA specific expected size amplicon band (approx. 125 bp) in our OPC and OL samples (**Figure S2A**). Further melting curve analysis by quantitative PCR (qPCR) yielded only single peak, suggesting a single amplicon (**Figure S2A**), confirming that there was no gDNA contamination.

Relative quantification of qPCR data was analyzed by Paffl method[80]. PCR efficiency of the target gene and the endogenous control was determined from the slope of the respective standard curve.

RNA immunoprecipitation (RIP)

Crosslinking RNA immunoprecipitation was performed by following standard Abcam protocol. 5-6 Sprague–Dawley rat brains (P7) were used. Meninges were removed and cellular dissociation was performed up to 33% Percoll step to remove debris and fat. Approximately 4-6 X 10^7 cells were obtained, re-suspended in Hank's balanced salt solution (HBSS) and immediately cross-linked with formaldehyde (with a final concentration 1%) for 10 min at room temperature. Crosslinking was stopped with glycine (final concentration 0.125M) and subsequently washed with PBS. Nuclei was isolated in nuclei isolation buffer and nuclear pellet was re-suspended in RIP buffer. Sonication (30 s on, 30 s off, 10 min) was performed using Bioruptor (Diagenode). Pre-cleaning of lysate was carried out using blocked Protein G sepharose (ab193259, Abcam). Blocking of Protein G sepharose was carried out by using 500ng/ul yeast tRNA and 1mg/ml RNase-free BSA. A total of 45% (450 μl) of pre-cleaned lysate was used for anti-SOX10 (sc-365692, Santa Cruz) or anti-immunoglobulin G1 (IgG1) (sc-3877, Santa Cruz) samples, and 10% was used for input sample. A total of 4 ug antibody (SOX10 or immunoglobulin G1) was used for immunoprecipitation, and blocked Protein G sepharose was used for the pull-down step after immunoprecipitation. Eluted solution was de-crosslinked and proteinase K treated, DNAase treatment (TurboDNAse) was performed, and RNA was isolated. RNA samples were analyzed by qPCR.

Chromatin-immuno-precipitation (ChIP)

Four P7 rat brains were pooled each time for OPC isolation and approximately 2×10^5 cells were seeded per well of a six-well-plate (PDL coated). Transfection were performed as mentioned above. 4 days after transfection, cells were cross linked with 16% HCHO added per well onto the media (final concentration 1%) and kept for 10 min and subsequently followed the standard ChIP protocol[81]. After fixation and quenching by glycine (final concentration: 0.125 M), cells were dissociated from PDL-coated wells using TrypLE Express (Thermo Fisher; 12604013).

Sonication (30 s on, 30 s off, 30 min)[82] was performed using a bioruptor (Diagenode) and the fragment size 200–500 bp was verified. Protein G Sepharose beads were blocked using 1 mg/ml BSA (stock 10 mg/ml). As described in RIP method (above), lysate was pre-cleaned and immunoprecipitation using antibody (anti SOX10 or IgG1) and blocked Protein G sepharose was used for the pull-down step after immunoprecipitation. Washing and preparation of DNA was performed as described[81]. Precipitated DNA samples were analysed by PCR and qPCR. The following primers were used for PCR after ChIP:

Mbp promoter:

F: 5' CATTGTTGTTGCAGGGGAGG 3'

R: 5' GCTCGTCGGACTCTGAGG 3'

Myrf ECR9:

F: 5' CACGTGGCTGACGGGATTTC 3'

R: 5' CCACAGCTGTGGCTGCTGGC 3'

Comparison of gene list from ChIP-seq data (GSE64703) with RNAseq data (GSE237744)

SOX10 targeted genomic co-ordinates in the CNS (oligodendrocytes) were obtained from GSE64703[23] and the annotation were acquired using HOMER annotatePeaks function[83] and using assembly: rn5. Non-redundant gene symbols were then compared with the gene symbols of downregulated genes due to RNLTR12-int inhibition (GSE237744).

Cloning, in vitro transcription (IVT) and labelling of RNA

Consensus   sequence   of   RNLTR12-int   (obtained   from   Repbase-Giri: https://www.girinst.org/repbase/) was cloned in pcDNA3.1 (+) using the restriction site NheI/XbaI.

Purified linearized plasmid (from above) was used as a template and transcribed *in vitro* using HiScribe™ T7 ARCA mRNA Kit (New England Biolabs) following manufacturer protocol (E2060), polyA tail was not incorporated. mRNA was purified using Direct-zol™ RNA Miniprep kit (Zymo Research).

RNA was 3' labelled with biotin using Pierce™ RNA 3' End Biotinylation Kit (ThermoFisher) and purified using Monarch® RNA Cleanup Kit (New England Biolabs).

Electrophoretic mobility shift assay (EMSA)

In vitro transcribed RNLTR12-int transcript was heated at 65°C for 5 min and then allowed to cool down at room temperature for 10 mins before starting the binding reactions. Binding reactions were 8μl and contained 3.4nM RNA and increasing concentrations (0.125μM, 0.25μM, 0.3μM, 0.5μM, 1μM, 2μM, 3μM, 4μM) of recombinant Human SOX10 protein (EP8458340, Euprotein). After incubation at room temperature for 20 mins, then mixed with loading dye (R1161, Thermofisher). complexes were separated on a 0.6% native agarose gel [containing GelRed (Millipore, SCT123)] using 1X TBE buffer at 4°C.

Surface plasmon resonance (SPR) experiment

The SPR experiment was performed on a Biacore T200 (GE Healthcare/Cytiva) using a Series S Sensor Chip SA (Cytiva) following the manufacturer's instructions. Tris Buffered Saline (TBS) pH7.4, supplemented with 0.002% Tween20, 100 mM glycine and 5% glycerol, was used as the running buffer throughout. Immobilization was achieved by flowing approximately 5 μgml^{-1} biotinylated RNA for 300 seconds at a flow rate of 2 μlmin^{-1}. For single cycle kinetics (SCK), recombinant Human SOX10 protein (Euprotein) was flowed (for 60 seconds at a flow rate of 30 μlmin^{-1}) as analyte at sequentially increasing concentrations (0.1 μM, 0.2 μM, 0.4 μM, 0.8 μM, 1.6 μM). Data were analysed using the inbuilt BIAEval software (Cytiva). Data was fitted with 1:1 binding model described in the user guide (https://cdn.cytivalifesciences.com/api/public/content/digi-33042-pdf).

Electron microscopy

The rats were injected with AAV (carrying shmiR constructs) into the corpus callosum (CC) in the left hemisphere at P6. At P21, the rats were perfused through the left ventricle with 20ml

of cold PBS, followed by perfusion with 20ml of 1% PFA and 1.25% glutaraldehyde prepared in 0.1M Sodium cacodylate buffer (pH 7.3), and finally with 50ml of 4% PFA and 5% glutaraldehyde in Sodium cacodylate (pH 7.3). Following post-fixation for 48 hours, the cortices containing AAV-injected CC were sliced in sagittal orientation at ~1 mm thickness and treated with 1% osmium tetroxide (ThermoFisher) overnight before going through epoxy resin embedding process described previously[84]. The injected site was located on semithin (800nm) sections stained with 0.5% toluidine blue (Merck). Ultrathin sections were cut and collected onto formvar coated copper grids (Agar Scientific) and stained with UranyLess and lead citrate (Electron Microscopy Sciences) before image acquisition using Hitachi HT7800 Transmission EM. G-ratio was quantified on images with 6,000X magnification, periodicity was measured using images at 100,000X magnification.

For G-ratio calculation the perimeter of the of the axon (P1) and the perimeter of the entire nerve fiber including myelin sheath (P2) were measures using measure function in Fiji[75]. All images were in similar scale. Considering axonal section as a circle in a two-dimensional plane, the perimeter=πD (D=diameter, π=22/7). G-ratio (i.e. the diameter of the axon divided by the diameter of the entire nerve fiber including myelin sheath) was then calculated as P1 divided by P2. G-ratio values (Y-axis) and the corresponding axonal diameters (X-axis) were plotted. Regression analysis and hypothesis testing was performed in 'R'. Best fit linear regression lines (G-ratio~diameter) were obtained by least square method. The deviation from linearity was insignificant ($p>0.05$, Runs test, two sided). Residuals were normally distributed ($p>0.05$, Shapiro-Wilk normality test). Heteroskedasticity between two linear models (residuals vs fitted for each data set) remained insignificant ($p>0.05$, F-test, two sided). The difference between the slopes and the intercepts of two regression lines were tested using analysis of covariance (ANCOVA). shmiR-Scrambled and shmiR-RNLTR12-int were defined as a categorical variable 'condition'. A joint model (G-ratio~diameter*condition) and an additive model (G-ratio~diameter+condition) was fitted for hypothesis testing of the slope and intercepts, respectively. Coefficients were estimated, and p-values were determined (using t-distribution). In our study we found no significant difference in slopes ($p>0.05$), while intercepts were significantly ($p< 2\times10^{-16}$) different.

Zebrafish experiment

To induce mutations in *RetroMyelin* loci in the zebrafish genome, we designed a guide crRNA targeting the sequence TAATGAAGCAATCAAACAAG*TGG* (PAM sequence italicized), which is conserved in the top 10 hit loci most similar to the RNLTR12-int. Ribonucleoprotein complex (RNP) were prepared by annealing 20 µM crRNA with 20 µM tracrRNA (IDT DNA) at 95°C, and incubating 1.6 µM of annealed gRNA with 2 µM Engen Cas9 enzyme (New England Biolabs) at 37°C for 10min. 1nL RNP or 1nL of Cas9-only control solution were microinjected into fertilized one-cell stage eggs of the transgenic myelin reporter line Tg(mbp:EGFP-CAAX)ue2[49]. Embryos were kept at 28.5°C in 10mM HEPES-buffered E3 Embryo medium.

At 5dpf, larvae were anesthetized in 0.16mg/mL tricaine (ethyl 3-aminobenzoate methanesulfonate salt, Sigma-Aldrich) and loaded and oriented for imaging using an LPSampler and VAST BioImager fitted with a 600 µm capillary (Union Biometrica) as previously described[85]. We imaged the whole length of the embryo with a Zeiss Axio Examiner D1 equipped with a CSU-X1 spinning confocal scanner, a Zeiss AxioCam 506 m CCD camera and a Zeiss C-Plan-Apochromat 10x 0.5NA objective, by tiling 5 z-stacks per embryo with a z-step of 3 µm. This system automatically collects individual larva from a multi-well plate, positions every larva consistently in the same angle, acquires a brightfield image for morphological analyses, and a high-resolution optically-sectioned GFP image for myelination analyses. Subsequently, individual larva was dispensed into a multi-well plate, from which we collected genomic DNA for each larva using the HOTSHOT method[86].

For analyses of RNP activity, we amplified a 607 bp PCR products around the target gRNA sequences from genomic DNA of each imaged larva using primers 5'-TTC GGC TGC GGT AGC AAA-3' and 5'-CGC TGT CTC AGT GGT CAG-3', and Onetaq DNA polymerase (New England Biolabs) according to the manufacturer's instructions (briefly, 94°C-30s, 35X [94°C-30s, 54°C-30s, 68°C-50s], 68°C-2min). PCR reactions were run in a 1% agarose gel, and a single band of the expected size was purified for each larva using Monarch Gel Extraction Kit (New England Biolabs). Purified amplicons from 6 Cas9-only control injected larva and from 24 gRNA-injected (and imaged) larva were sent for Sanger sequencing using the PCR primers (Source Bioscience).

For morphology analyses of brightfield images, we used Fiji/ImageJ to mark the rostral and caudal ends of each larva (mouth and tail fin) to measure body length, and the 'wand' tool (with a constant threshold set manually with the first 2-3 samples and used throughout the

dataset) to select the eye area while blinded to experimental condition. For myelination image processing and analyses, maximum intensity projections were created for all samples and a semi-automated Fiji/ImageJ script defined regions of interest in spinal cord (whole, dorsal and ventral as well as bins along the anterior-posterior length) in a manner blinded to experimental group[85]. The mean gray value was obtained for each region of interest. In the figures, all zebrafish images represent a lateral view, anterior to the left and dorsal on top. For the myelination figure panel, control and gRNA-injected images were prepared in parallel using the same brightness and contrast settings, and the 'fire' look-up table was used to convey fluorescence intensity/gray values as colour changes, indicated in the calibration bar.

Frog (*Xenopus laevis*) experiment

Maintenance of tadpoles, Crispr/Cas9 injection, albino selection and immunolabeling procedures were performed as described[87]. Preparation of sgRNAs, generation of S1P5 null mutant: CRISPR sgRNA were as described[87]. We design three guide RNAs (gRNA) using CHOPCHOP (https://chopchop.cbu.uib.no/) with 1 to 3 mismatches capable of covering a target region in multiple sequences of *Xenopus RetroMyelin*:

(The red color indicates the mismatch in gRNA, the brown color indicates the PAM sequence.)

xl.seq1_2gRNA3	5' TGGAACGGACCGTCAAATCG GGG 3'
xl.seq3gRNA3	5' TGAAAAGGCCCATCAAACCG GGG 3'
xl.seq4_5gRNA	5' TGGAAAGGCCCGTCAAACCT GGG 3'

The region including the guide RNA target site were amplified by PCR using primers designed with Primer3, and Sanger sequencing with the forward primers was performed. The sequences data were analyzed with CRISP-ID[88].

FP: 5' GTGCATGGTCAGGTGTTCTC 3'

RP: 5' GGATCCACGCATCCCACTGCAA 3'

Images were acquired using a Nikon A1R-HD25-confocal microscope. Z-series of wholemount tadpole were performed at 0.375 µm step. Maximum orthogonal projection of images was carried out using Fiji software (NIH, Bethesda, Maryland). To analyze of Olig (this is not OLIG1 or OLIG2, for simplicity reason we wrote XMM)[89] and MBP expression, images were quantified for labeled area fraction (%) by automated counting using ImageJ software. Briefly, images were normalized by subtracting background and a threshold (MaxEntropy) was

applied. Mean values of area fraction (%) were obtained from images of tadpole injected with gRNA vs control.

QUANTIFICATIION AND STATISTICAL ANALYSIS

R (version 4.0.2) (http://www.R-project.org) or GraphPad Prism (version 8.4.3) was used for statistical analysis. Shapiro-Wilk test[90] was performed to test normality. Welch's t test was performed whenever heteroscedasticity existed in the data set.

Supplemental Table

Table S1. *RetroMyelin* sequences in different species, related to Figure 5.

This table list the top hit of the identified repeat type, after searching remote homology using nhmmer[65]. Specific repeat annotation, wherever available, is written under the parathesis in column 5. NA: Not applicable. Not found: wherever no hit is being obtained above the inclusion threshold (E-value<0.01).

REFERENCES

1. Zalc, B., Colman, D.R. (2000). Origins of vertebrate success. Science *288*, 271-272. https://doi.org/10.1126/science.288.5464.271c.

2. Salzer, J.L., Zalc, B. (2016). Myelination. Curr Biol. *26*, R971-R975. https://doi.org/10.1016/j.cub.2016.07.074.

3. Nave, K.A., Werner, H.B. (2021). Ensheathment and Myelination of Axons: Evolution of Glial Functions. Annu. Rev. Neurosci. *44*, 197-219. https://doi.org/10.1146/annurev-neuro-100120-122621.

4. Lemke, G. (1988). Unwrapping the genes of myelin. Neuron *1*, 535-543. https://doi.org/10.1016/0896-6273(88)90103-1.

5. Fields, R.D., Stevens-Graham, B. (2002). New insights into neuron-glia communication. Science *298*, 556-562. https://doi.org/10.1126/science.298.5593.556.

6. Nawaz, S., Schweitzer, J., Jahn, O., Werner, H.B. (2013). Molecular evolution of myelin basic protein, an abundant structural myelin component. Glia *61*, 1364-1377. https://doi.org/10.1002/glia.22520.

7. Johnson, W.E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses. Nat. Rev. Microbiol. *17*, 355-370. https://doi.org/10.1038/s41579-019-0189-2.

8. Thompson, P.J., Macfarlan, T.S., Lorincz, M.C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Mol Cell *62*, 766-776. https://doi.org/10.1016/j.molcel.2016.03.029.

9. Kazazian, H.H. Jr. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632. https://doi.org/10.1126/science.1089670.

10. Kazazian, H.H. Jr., Moran, J.V. (2017). Mobile DNA in Health and Disease. N. Engl. J. Med. *377*, 361-370. https://doi.org/10.1056/NEJMra1510092.

11. Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership regulates early development and ESC identity. Cell *174*, 391-405. https://doi.org/10.1016/j.cell.2018.05.043.

12. Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific

populations of early human embryonic cells. Cell Stem Cell *16*, 135-141. https://doi.org/10.1016/j.stem.2015.01.005.

13. Robb, G.B., Brown, K.M., Khurana, J., Rana, T.M. (2005). Specific and potent RNAi in the nucleus of human cells. Nat. Struct. Mol. Biol. *12*, 133-137. https://doi.org/10.1038/nsmb886.

14. Martini, R., Mohajeri, M.H., Kasper, S., Giese, K.P., Schachner, M. (1995). Mice doubly deficient in the genes for P0 and myelin basic protein show that both proteins contribute to the formation of the major dense line in peripheral nerve myelin. J Neurosci. *15*, 4488-95. https://doi.org/10.1523/JNEUROSCI.15-06-04488.1995.

15. Holmes, Z.E., Hamilton, D.J., Hwang, T., Parsonnet, N.V., Rinn, J.L., Wuttke, D.S., Batey, R.T. (2020). The Sox2 transcription factor binds RNA. Nat. Commun. *11*, 1805. https://doi.org/10.1038/s41467-020-15571-8.

16. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978-981. https://doi.org/10.1126/science.aad3346.

17. Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat. Genet. *43*, 621-629. https://doi.org/10.1038/ng.848.

18. Cassiday, L.A., Maher, LJ 3rd. (2002). Having it both ways: transcription factors that bind DNA and RNA. Nucleic Acids Res. *30*, 4118-4126. https://doi.org/10.1093/nar/gkf512.

19. Jeon Y, Lee J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119-133. https://doi.org/10.1016/j.cell.2011.06.026.

20. Gottesfeld, J.M., Barbas, C.F. 3rd. (2003). RNA as a transcriptional activator. Chem. Biol. *10*, 584-585. https://doi.org/10.1016/s1074-5521(03)00153-4.

21. Stolt, C.C., Rehberg, S., Ader, M., Lommes, P., Riethmacher, D., Schachner, M., Bartsch, U., Wegner, M. (2002). Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10. Genes Dev. *16*, 165-70. https://doi.org/10.1101/gad.215802.

22. Li, H., Lu, Y., Smith, H.K., Richardson, W.D. (2007). Olig1 and Sox10 interact synergistically to drive myelin basic protein transcription in oligodendrocytes. J Neurosci. *27*, 14375-14382. https://doi.org/10.1523/JNEUROSCI.4456-07.2007.

23. Lopez-Anido, C., Sun, G., Koenning, M., Srinivasan, R., Hung, H.A., Emery, B., Keles, S., Svaren, J. (2015). Differential Sox10 genomic occupancy in myelinating glia. Glia *63*, 1897-1914. https://doi.org/10.1002/glia.22855.

24. Hornig, J., Fröb, F., Vogl, M.R., Hermans-Borgmeyer, I., Tamm, E.R., Wegner, M. (2013). The transcription factors Sox10 and Myrf define an essential regulatory network module in differentiating oligodendrocytes. PLoS Genet. *9*, e1003907. https://doi.org/10.1371/journal.pgen.1003907.

25. Duncan, G.J., Plemel, J.R., Assinck, P., Manesh, S.B., Muir, F.G.W., Hirata, R., Berson, M., Liu, J., Wegner, M., Emery, B., Moore, G.R.W., Tetzlaff, W. (2017). Acta Neuropathol. *134*, 403-422. https://doi.org/10.1007/s00401-017-1741-7.

26. Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L., Feschotte, C. (2018). Ten things you should know about transposable elements. Genome Biol. *19*,199. https://doi.org/10.1186/s13059-018-1577-z.

27. Goldman, N., Whelan, S. (2000). Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Mol. Biol. Evol. *17*, 975-978. https://doi.org/10.1093/oxfordjournals.molbev.a026378.

28. McKenzie, I.A., Ohayon, D., Li, H., de Faria, J.P., Emery, B., Tohyama, K., Richardson, W.D. (2014). Motor skill learning requires active central myelination. Science *346*, 318-322. https://doi.org/10.1126/science.1254960.

29. Xin, W., Chan, J.R. (2020). Myelin plasticity: sculpting circuits in learning and memory. Nat. Rev. Neurosci. *21*, 682-694. https://doi.org/10.1038/s41583-020-00379-8.

30. Franklin, R.J.M., ffrench-Constant, C. (2017). Regenerating CNS myelin - from mechanisms to experimental medicines. Nat. Rev. Neurosci. *18*, 753-769. https://doi.org/10.1038/nrn.2017.136.

31. Fancy, S.P., Chan, J.R., Baranzini, S.E., Franklin, R.J.M., Rowitch, D.H. (2011). Myelin regeneration: a recapitulation of development? Annu. Rev. Neurosci. *34*, 21-43. https://doi.org/10.1146/annurev-neuro-061010-113629.

32. Nawaz, S., Sánchez, P., Schmitt, S., Snaidero, N., Mitkovski, M., Velte, C., Brückner, B.R., Alexopoulos, I., Czopka, T., Jung, S.Y., Rhee, J.S., Janshoff, A., Witke, W., Schaap, I.A.T., Lyons, D.A., Simons, M. (2015). Actin filament turnover drives leading edge growth during myelin sheath formation in the central nervous system. Dev. Cell *34*, 139-151. https://doi.org/10.1016/j.devcel.2015.05.013.

33. Kottmeier, R., Bittern, J., Schoofs, A., Scheiwe, F., Matzat, T., Pankratz, M., Klämbt, C. (2020). Wrapping glia regulates neuronal signaling speed and precision in the peripheral nervous system of Drosophila. Nat. Commun. *11*, 4491. https://doi.org/10.1038/s41467-020-18291-1.

34. Aggarwal, S., Yurlova, L., Snaidero, N., Reetz, C., Frey, S., Zimmermann, J., Pähler, G., Janshoff, A., Friedrichs, J., Müller, D.J., Goebel, C., Simons, M. (2011). A size barrier limits protein diffusion at the cell surface to generate lipid-rich myelin-membrane sheets. Dev. Cell *21*, 445-56. https://doi.org/10.1016/j.devcel.2011.08.001.

35. Aggarwal, S., Snaidero, N., Pähler, G., Frey, S., Sánchez, P., Zweckstetter, M., Janshoff, A., Schneider, A., Weil, M.T., Schaap, I.A., Görlich, D., Simons, M. (2013). Myelin membrane assembly is driven by a phase transition of myelin basic proteins into a cohesive protein meshwork. PLoS Biol. *11*, e1001577. https://doi.org/10.1371/journal.pbio.1001577.

36. Zuchero, J.B., Barres, B.A. (2011). Between the sheets: a molecular sieve makes myelin membranes. Dev. Cell *21*, 385-6. https://doi.org/10.1016/j.devcel.2011.08.023.

37. Werner, H.B. (2013). Do we have to reconsider the evolutionary emergence of myelin? Front. Cell Neurosci. *7*, 217. https://doi.org/10.3389/fncel.2013.00217.

38. Jacobs, E.C., Pribyl, T.M., Feng, J.M., Kampf, K., Spreur, V., Campagnoni, C., Colwell, C.S., Reyes, S.D., Martin, M., Handley, V., Hiltner, T.D., Readhead, C., Jacobs, R.E., Messing, A., Fisher, R.S., Campagnoni, A.T. (2005). Region-specific myelin pathology in mice lacking the golli products of the myelin basic protein gene. J Neurosci. *25*, 7004-13. https://doi.org/10.1523/JNEUROSCI.0288-05.2005.

39. Zalc, B. (2016). The acquisition of myelin: An evolutionary perspective. Brain Res. *1641*, 4-10. https://doi.org/10.1016/j.brainres.2015.09.005.

40. Zawadzka, M., Rivers, L.E., Fancy, S.P., Zhao, C., Tripathi, R., Jamen, F., Young, K., Goncharevich, A., Pohl, H., Rizzi, M., Rowitch, D.H., Kessaris, N., Suter, U., Richardson, W.D., Franklin, R.J.M. (2010). CNS-resident glial progenitor/stem cells produce

Schwann cells as well as oligodendrocytes during repair of CNS demyelination. Cell Stem Cell. *6*, 578-90. https://doi.org/10.1016/j.stem.2010.04.002.

41. Siems, S.B., Jahn, O., Hoodless, L.J., Jung, R.B., Hesse, D., Möbius, W., Czopka, T., Werner, H.B. (2021). Proteome Profile of Myelin in the Zebrafish Brain. Front. Cell Dev. Biol. *9*, 640169. https://doi.org/10.3389/fcell.2021.640169.

42. Hogan, E.L., Greenfield, S. (1984). Animal models of genetic disorders of myelin. In: Myelin (Morell. P. ed), 489-534. New York: Plenum.

43. Hartline, D.K., Colman, D.R. (2007). Rapid conduction and the evolution of giant axons and myelinated fibers. Curr. Biol. *17*, R29-35. https://doi.org/10.1016/j.cub.2006.11.042.

44. Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., Heidmann, T. (2013). Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *368*, 20120507. https://doi.org/10.1098/rstb.2012.0507.

45. Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., Nilsson, M.A., Menzies, B.R., Renfree, M.B., Pierron, G., Zeller, U., Heidmann, O., Dupressoir, A., Heidmann, T. (2015). Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. Proc. Natl. Acad. Sci. USA *112*, E487-96. https://doi.org/10.1073/pnas.1417000112.

46. Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O.A., Meurice, G, Heidmann, O., Dupressoir, A., Miralles, A., Ramirez-Pinilla, M.P., Heidmann, T. (2017). An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. Proc. Natl. Acad. Sci. USA *114*, E10991-E11000. https://doi.org/10.1073/pnas.1714590114.

47. Segel, M., Neumann, B., Hill, M.F.E., Weber, I.P., Viscomi, C., Zhao, C., Young, A., Agley, C.C., Thompson, A.J., Gonzalez, G.A., Sharma, A., Holmqvist, S., Rowitch, D.H., Franze, K., Franklin, R.J.M., Chalut, K.J. (2019). Niche stiffness underlies the ageing of central nervous system progenitor cells. Nature *573*, 130-134. https://doi.org/10.1038/s41586-019-1484-9.

48. Neumann, B., Baror, R., Zhao, C., Segel, M., Dietmann, S., Rawji, K.S., Foerster, S., McClain, C.R., Chalut, K., van Wijngaarden, P., Franklin, R.J.M. (2019). Metformin

restores CNS remyelination capacity by rejuvenating aged stem cells. Cell Stem Cell *25*, 473-485. https://doi.org/10.1016/j.stem.2019.08.015.

49. Arthur-Farraj, P., Wanek, K., Hantke, J., Davis, C. M., Jayakar, A., Parkinson, D. B., Mirsky, R., Jessen, K.R. (2011). Mouse schwann cells need both NRG1 and cyclic AMP to myelinate. Glia *59*, 720–733. https://doi.org/10.1002/glia.21144.

50. Jessen, K.R., Brennan, A., Morgan, L., Mirsky, R., Kent, A., Hashimoto, Y., Gavrilovic, J. (1994). The Schwann cell precursor and its fate: a study of cell death and differentiation during gliogenesis in rat embryonic nerves. Neuron *12*, 509-27. https://doi.org/10.1016/0896-6273(94)90209-7.

51. Almeida, R.G., Czopka, T., Ffrench-Constant, C., Lyons, D.A. (2011). Individual axons regulate the myelinating potential of single oligodendrocytes in vivo. Development *138*, 4443-4450. https://doi.org/10.1242/dev.071001.

52. Reichmann, J., Crichton, J.H., Madej, M.J., Taggart, M., Gautier, P., Garcia-Perez, J.L., Meehan, R.R., Adams, I.R. (2012). Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. PLoS Comput. Biol. *8*, e1002486. https://doi.org/10.1371/journal.pcbi.1002486.

53. AFA Smit, Hubley, R., Green, P. (2013-2015). RepeatMasker Open-4.0. http: //www.repeatmasker.org.

54. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

55. Langfelder, P., Horvath, S. (2008). WGCNA, an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

56. Becker, R.A., Chambers, J. M. , Wilks A. R. (1988). The New S Language. Wadsworth & Brooks/Cole.

57. Huang, da W., Sherman, B.T., Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44-57. https://doi.org/10.1038/nprot.2008.211.

58. Guo, J.C., Fang, S.S., Wu, Y., Zhang, J.H., Chen, Y., Liu, J., Wu, B., Wu, J.R., Li, E.M., Xu, L.Y., Sun L., Zhao, Y. (2019). CNIT: a fast and accurate web tool for identifying protein-

coding and long non-coding transcripts based on intrinsic sequence composition. Nucleic Acids Res. *47*, W516-W522. https://doi.org/10.1093/nar/gkz400.

59. Kopylova, E., Noé, L., Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics *28*, 3211-7. https://doi.org/10.1093/bioinformatics/bts611.

60. Bolger, A.M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

61. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. https://doi.org/10.1093/bioinformatics/bts635.

62. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-9. https://doi.org/10.1093/bioinformatics/btp352.

63. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417-419. https://doi.org/10.1038/nmeth.4197.

64. Love, M.I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

65. Wheeler, T.J., Eddy, S.R. (2013). nhmmer: DNA homology search with profile HMMs. Bioinformatics *29*, 2487-9. https://doi.org/10.1093/bioinformatics/btt403.

66. Letunic, I., Bork, P. (2019). Interactive Tree Of Life (iTOL) v4, recent updates and new developments. Nucleic Acids Res. *47*, W256-W259. https://doi.org/10.1093/nar/gkz239.

67. Katoh, K., Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772-80. https://doi.org/10.1093/molbev/mst010.

68. Criscuolo, A., Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple

sequence alignments. BMC Evol. Biol. *10*, 210. https://doi.org/10.1186/1471-2148-10-210.

69. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312-3. https://doi.org/10.1093/bioinformatics/btu033.

70. Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M., Stamatakis, A. (2010). How many bootstrap replicates are necessary? J. Comput. Biol. *17*, 337-54. https://doi.org/10.1089/cmb.2009.0179.

71. Yang, Z. (1997). PAML, a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555-6. https://doi.org/10.1093/bioinformatics/13.5.555.

72. Liberles, D.A., Dittmar, K. (2008). Characterizing gene family evolution. Biol. Proced. Online *10*, 66-73. https://doi.org/10.1251/bpo144.

73. Yang, Z. (1994). Estimating the pattern of nucleotide substitution. J. Mol. Evol. *39*, 105-11. https://doi.org/10.1007/BF00178256.

74. Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites, approximate methods. J. Mol. Evol. *39*, 306-14. https://doi.org/10.1007/BF00160154.

75. Yang, Z., Wang, T. (1995). Mixed model analysis of DNA sequence evolution. Biometrics *51*, 552-61.

76. Pol, S.U., Lang, J.K., O'Bara, M.A., Cimato, T.R., McCallion, A.S., Sim, F.J. (2013). Sox10-MCS5 enhancer dynamically tracks human oligodendrocyte progenitor fate. Exp. Neurol. *247*, 694-702. https://doi.org/10.1016/j.expneurol.2013.03.010.

77. Rawji, K.S., Kappen, J., Tang, W., Teo, W., Plemel, J.R., Stys, P.K., Yong, V.W. (2018). Deficient surveillance and phagocytic activity of myeloid cells within demyelinated lesions in aging mice visualized by ex vivo live multiphoton imaging. J. Neurosci. *38*, 1973-1988. https://doi.org/10.1523/JNEUROSCI.2341-17.2018.

78. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676-82. https://doi.org/10.1038/nmeth.2019.

79. Rawji, K.S., Young, A.M.H., Ghosh, T., Michaels, N.J., Mirzaei, R., Kappen. J., Kolehmainen, K.L., Alaeiilkhchi, N., Lozinski, B., Mishra, M.K., Pu, A., Tang, W., Zein, S., Kaushik, D.K., Keough, M.B., Plemel, J.R., Calvert, F., Knights, A.J., Gaffney, D.J., Tetzlaff, W., Franklin, R.J.M., Yong, V.W. (2020). Niacin-mediated rejuvenation of macrophage/microglia enhances remyelination of the aging central nervous system. Acta Neuropathol. *139*, 893-909. https://doi.org/10.1007/s00401-020-02129-7.

80. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45. https://doi.org/10.1093/nar/29.9.e45.

81. Weinmann, A.S., Farnham, P.J. (2002). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods *26*, 37-47. https://doi.org/10.1016/S1046-2023(02)00006-3.

82. Ghosh, T., Aprea, J., Nardelli, J., Engel, H., Selinger, C., Mombereau, C., Lemonnier, T., Moutkine, I., Schwendimann, L., Dori, M., Irinopoulou, T., Henrion-Caude, A., Benecke, A.G., Arnold, S.J., Gressens, P., Calegari, F., Groszer, M. (2014). MicroRNAs establish robustness and adaptability of a critical gene network to regulate progenitor fate decisions during cortical neurogenesis. Cell Rep. *7*, 1779-88. https://doi.org/10.1016/j.celrep.2014.05.029.

83. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576-89. https://doi.org/10.1016/j.molcel.2010.05.004.

84. Zhao, C., Fancy, S.P., ffrench-Constant, C., Franklin, R.J. (2008) Osteopontin is extensively expressed by macrophages following CNS demyelination but has a redundant role in remyelination. Neurobiol. Dis. *31*, 209-17. https://doi.org/10.1016/j.nbd.2008.04.007.

85. Early, J.J., Cole, K.L., Williamson, J.M., Swire, M., Kamadurai, H., Muskavitch, M., Lyons, D.A. (2018). An automated high-resolution in vivo screen in zebrafish to identify chemical regulators of myelination. Elife *7*, e35136. https://doi.org/10.7554/eLife.35136.

86. Meeker, N.D., Hutchinson, S.A., Ho, L., Trede, N.S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. Biotechniques *43*, 610-614. https://doi.org/10.2144/000112619.

87. Mannioui, A., Vauzanges, Q., Fini, J.B., Henriet, E., Sekizar, S., Azoyan, L., Thomas, J.L., Pasquier, D.D., Giovannangeli, C., Demeneix, B., Lubetzki, C., Zalc, B. (2018). The Xenopus tadpole: An in vivo model to screen drugs favoring remyelination. Mult. Scler. *24*, 1421-1432. https://doi.org/10.1177/1352458517721355.

88. Dehairs, J., Talebi, A., Cherifi, Y., Swinnen, J.V. (2016). CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Sci. Rep. *6*, 28973. https://doi.org/10.1038/srep28973

89. Steen, P., Kalghatgi, L., Constantine-Paton, M. (1989). Monoclonal antibody markers for amphibian oligodendrocytes and neurons. J. Comp. Neurol. *289*, 467-80. https://doi.org/10.1002/cne.902890311.

90. Royston, J.P. (1982). Algorithm AS 181: The W Test for Normality. Appl. Stat. *181*, 176–180.

A

Dissimilarity

Modules

Scale
100 genes

B

Oligodendrocytes

greenyellow	-0.78 (2e-04)	*
pink	-0.96 (1e-09)	*
brown	-0.84 (3e-05)	*
cyan	-0.86 (8e-06)	*
tan	-0.74 (6e-04)	*
purple	-0.72 (0.001)	
salmon	-0.76 (4e-04)	*
blue	-0.74 (7e-04)	*
green	-0.82 (6e-05)	*
black	0.79 (2e-04)	*
red	0.79 (2e-04)	*
yellow	0.83 (4e-05)	*
magenta	0.79 (2e-04)	*
turquoise	0.96 (6e-10)	**

Correlation scale

-1 DOWN

0

1 UP

C

Threshold (Padj=0.05)

mitotic cell cycle process — green
neurogenesis — blue
angiogenesis — salmon
epithelium development — purple
biological adhesion — tan
regulation: DNA damage checkpoint — cyan
axonogenesis — brown
mitotic nuclear division — pink
cell division — greenyellow
regulation: G1/S transition in mitotic — black red
steroid biosynthetic process — red
sterol biosynthetic process — yellow
ion homeostasis — magenta
MHC class II protein complex — turquoise
myelination — turquoise

DOWN

UP

-log₁₀ (Padj)

D

Relative expression

RNLTR12-int, B3A, B4, MT2A, PB1D9

E

Tspan2, Pllp, Sh3tc2, Mbp, Mal, Mag, Lpar1, Plp1, Ndrg1, Prx, Rab40c, Adgrg6, Ugt8, Sirt2, Myrf, Tf, Aspa, Atrn, Pmp22, Lama2, Nfasc, RNLTR12-int

F

RNLTR12-int

Relative expression

OPC, OL

Mbp

OPC, OL

G

\# RNLTR12-int

RNLTR12-int flanked by LTR
RNLTR12-int not flanked by LTR

Chromosome

Figure 1

64

A

Control siRNA   siRNLTR12-int

RNLTR12-int

Actb

1   2   3   1   2   3

Biological replicates

RNLTR12-int

Relative expression

Control siRNA   siRNLTR12-int

*

B

DAPI   OLIG2   MBP/OLIG2/DAPI

Control siRNA

siRNLTR12-int

#MBP+ / #OLIG2+

Control siRNA   siRNLTR12-int

**

C

Pdgfra

Relative expression

P>0.05

Control siRNA   siRNLTR12-int

Cnp

Relative expression

Control siRNA   siRNLTR12-int

*

Mbp

Relative expression

Control siRNA   siRNLTR12-int

*

D

FC=2   FC=2

$-\log_{10}$ (Padj)

Threshold
(Padj=0.05)

$\log_2$(siRNLTR12-int/Control siRNA)

G2/M transition of mitotic cell cycle
regulation of oligodendrocyte differentiation
Glycoprotein
plasma membrane
Disulfide bond
Lipoprotein
Cell membrane
cell adhesion
paranode region of axon
cns myelination
myelination
myelin sheath

Threshold
(Padj=0.05)

Nfasc
Nkx6-2
Plp1
Kel
Gjc2
Gjc3
Ermn
Mag
Mobp
Myrf
Mbp

$-\log_{10}$ (Padj)

Figure 2

A

SOX10 → EGFP — shmiR siRNLTR12-int/siScrambled bGHp(A)

AAV carrying cassette

AAV injection (P1)

↓

Immunofluorescence (P14)
in situ (RNAscope) (P14)

B

RNLTR12-int

Relative expression

shmiR-Scrambled   shmiR-RNLTR12-int

*

C

DAPI/OLIG2   OLIG2   GFP/OLIG2

shmiR-Scrambled
shmiR-RNLTR12-int

● shmiR-Scrambled   ● shmiR-RNLTR12-int

#OLIG2+GFP+/#GFP+DAPI+   P>0.05

#GFP+OLIG2+/ #OLIG2+   P>0.05

D

GFP   CC1   GFP/CC1

shmiR-Scrambled
shmiR-RNLTR12-int

#CC1+GFP+/ #GFP+   P>0.05

● shmiR-Scrambled
● shmiR-RNLTR12-int

E

DAPI   GFP/OLIG2   MBP   MBP/GFP

shmiR-Scrambled
shmiR-RNLTR12-int

MBP

Relative intensity   **

● shmiR-Scrambled
● shmiR-RNLTR12-int

F

RNA In situ (RNAscope)

DAPI/GFP   MBP   DAPI/GFP/MBP

shmiR-Scrambled
shmiR-RNLTR12-int

Av. cell/mm²   P>0.05

Mbp
Av. Mbp puncta/mm²   *

Gfp
Av. Gfp puncta/mm²   P>0.05

● shmiR-Scrambled   ● shmiR-RNLTR12-int

G

AAV injection (P6) → Transmission electron microscopy (P21)

shmiR-Scrambled
shmiR-RNLTR12-int

G-ratio vs Axon diameter (µm)
***
● shmiR-Scrambled
● shmiR-RNLTR12-int

Cell counts (%)
** ** P>0.05 **
☐ Myelinated
☐ Unmyelinated
shmiR-Scrambled   shmiR-RNLTR12-int

Figure 3

66

A RIP-qPCR

B [SOX10] (μM), [BSA] (μM)

C

$K_{on} (M^{-1}s^{-1}) : (188\pm1.66) \times 10^2$
$K_{off} (s^{-1}) : (4.80\pm0.029) \times 10^{-4}$
$KD (nM) \sim 26\ nM$

D Mbp promoter

ChIP-qPCR

E

F

ChIP seq : SOX10 targets in OL

RNA seq : Down regulated in siRNLTR12-int

Figure 4

Figure 5

A

Invasion before speciation:

X0

Invasion & propagation

Retrotransposition

Purged out from the population pool

Retrotransposition capacity lost

Speciation

1

2

Copies of retrotransposon

A
a3 a4 a5 a6

B
b3 b4 b5 b6

Copies of retrotransposon

C
c3 c4 c5 c6

D
d3 d4 d5 d6

E
e3 e4 e5 e6

F
f3 f4 f5 f6

Sampling of any 2 copies:

c3 c4   d4 d6   e3 e5   f3 f6

Tree generated:

X0

B

Invasion after speciation:

Animal phylogeny   Infection   Viral phylogeny

1. Retrotransposition & subsequently lost this capacity
2. Retroelement purged from the population pool OR fixed & diverge within the host species

Sequences of common evolutionary origin

Sampling of any 2 copies:

Tree generated:

C

Cow
Rat
Mouse
Gorilla
Chimpanzee
Human
Turtle
Horse
Zebra finch
Chicken

bootstrap
- 80% – 90%
- 90% – 95%
- 95% – 97.5%
- 97.5% – 100%

Xenopus laevis

Zebrafish
Stickleback
Cichlid
Turbot
Atlantic cod
Medaka
Spiny-chromis
Takifugu
Whale shark
Elephant shark
Coelacanth

10
Scale

D

Figure 6

A Zebrafish 5dpf Tg(mbp:EGFP-CAAX)

Control

Retromyelin gRNA

0.5mm

Control

Retromyelin gRNA

Gray Value
700
350
0

] dorsal
] ventral

50μm

B

Eye area

P>0.05

Body length

P>0.05

● Control   ● *Retromyelin* gRNA

C

- ○ Control ventral
- ■ *RetroMyelin* gRNA ventral
- ● Control dorsal
- ■ *RetroMyelin* gRNA dorsal

rostral   caudal
0μm   1200μm   2400μm

length of the spinal cord

Figure7

A

B

C

D

E

Figure S1

A

Actb

RNLTR12-int

bp
250
150
100

bp
150
100

-(dF/dT)

Actb

RNLTR12-int

Temperature (°C)

B

Actb

Relative level

P>0.05

OPC   OL

C

Mag

Relative expression

OPC   OL

Tspan2

OPC   OL

PB1D9

OPC   OL

D

CINT score

Tbp

RNLTR12-int

Coding
Threshold
Non-coding

E

Depth

Nucleotide position
Consensus sequence (RNLTR12-int)

Figure S2

A

Examples of RNLTR12-int flanked by LTR

chr18

chr15

chrX

chr19

Examples of RNLTR12-int not flanked by LTR

chr1

chr12

chr6

chr9

Scale

2000 bp

B

Figure S3

Figure S4

A

OPCs → Control siRNA/ siRNA RNLTR12-int | Differentiation media → 3 days → Null vector/vector carrying RNLTR12-int | Differentiation media → 4 days → Immunofluorescence

OLIG2 | MBP | MBP/OLIG2

Null vector: Control siRNA

siRNA RNLTR12-int

Vector carrying RNLTR12-int: siRNA RNLTR12-int

#MBP+/#OLIG2+

B

DAPI | GFP/OLIG2 | CNPase | CNPase/GFP

shmiR-Scrambled

shmiR-RNLTR12-int

CNPase

Relative intensity (AU)

● shmiR-Scrambled
● shmiR-RNLTR12-int

C

G-ratio

Axonal diameter (μm)

P>0.05

● shmiR-Scrambled
● shmiR-RNLTR12-int

D

shmiR-Scrambled

shmiR-RNLTR12-int

Periodicity (nm)

P>0.05

● shmiR-Scrambled
● shmiR-RNLTR12-int

E

Myrf ECR9   IP:SOX10 | IP:IgG | Input
ChIP

Control siRNA

siRNLTR12-int

PCR

ChIP qPCR: Myrf ECR9

Fold enrichment

P>0.05

IP: SOX10 | IP: IgG | IP: SOX10 | IP: IgG
Control | siRNLTR12-int

Figure S5

75

Figure S6

A

Xenopus laevis   Control

Ref AAATCCTCTCTCGATCCCCGGTTTGACGGGCCCTTTCAAGTCCTGCTGACAACTGCCACCT
a   AAATCCCCTCTCGATCCCCGATTTGACGGGCCCTTTCAAGTCCTGCTCACAACTGCCACCT
b   AAATCTTCACTCGAGCCCCGATTTGACGGGTCCGTTCACTGACTACAGCACCT

Retromyelin gRNA

Ref AAATCCTCTCTCGATCCCCG---GTTTGACGGGCCCTTTCAAGTCTGCTGACAACTGCCACCTCT
a   AAATCCTCTCTCGATCCCCG---GTTTGACGGGCCCTTTCAAGTCTGCTGCTCACAACTGCCACCTCT
b   AAATCCCCTCTCGATCCCCG-GCTTGC--CGGGGCCAATTCCAATCCAGCTCACCGCTGTCACCTCT
c   AAATCTTCACTCGAGGCGCGGAGGGTGCGTTGTGCGACGTCCTGCCTACTGTCACCGCT
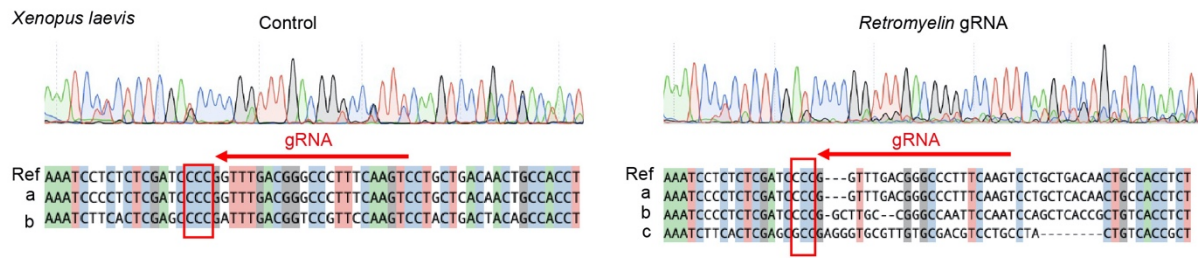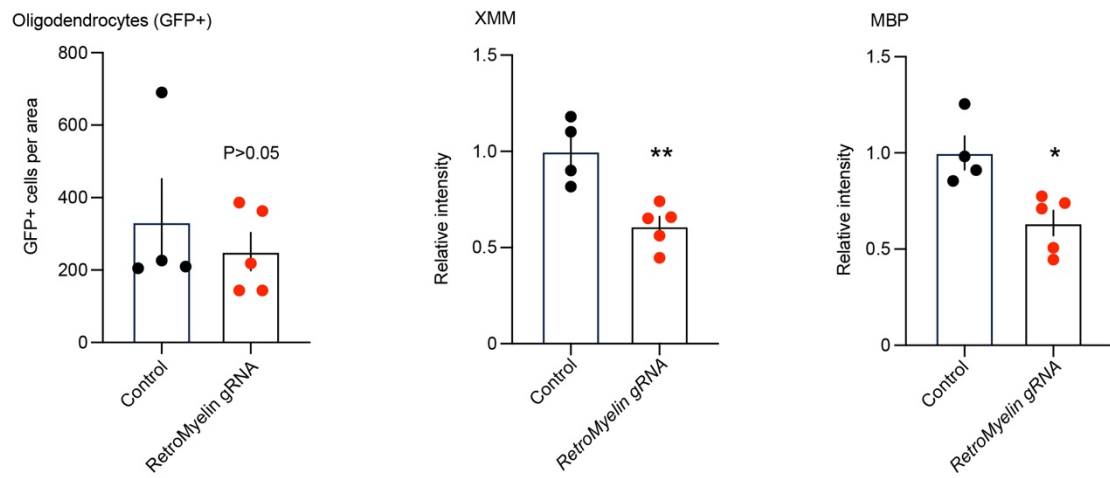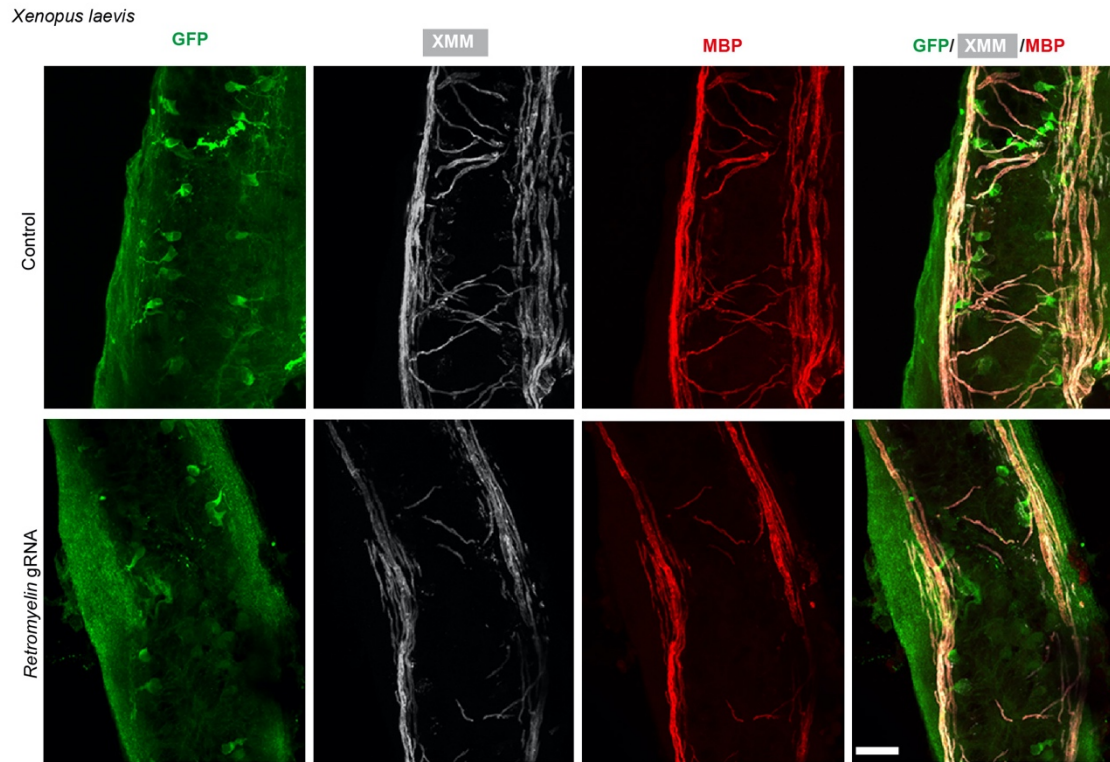
B

Xenopus laevis

Figure S7

Table S1. *RetroMyelin* sequences in different species, related to Figure 5 and STAR methods.

Vertebrates	Genome assembly	Top hit	E-value	Matching repeat	Family
Human	hg38	chr16:35270573-35269845	4.6E-36	ERV1-int (HERVS71-int)	LTR/ERV1
Chimpanzee	panTro4	chr13:18277202-18274245	7.4E-101	ERV1-int (PTERV1c-int)	LTR/ERV1
Gorilla	gorGor3	CABD02423608:217-1465	3.6E-87	ERV1-int (HERVS71-int)	LTR/ERV1
Mouse	mm10	chr7:30112940-30114182	1.5e-142	ERV1-int (MERV1_I-int)	LTR/ERV1
Rat	rn5	chr15:24404804-24399534	0	ERV1-int (RNLTR12-int)	LTR/ERV1
Cow	bosTau7	chr29:42154010-42153024	1.8e-29	ERV1-int (BtERVF2_I-int)	LTR/ERV1
Horse	equCab2	chr1:27399148-27399594	2.9e-18	ERV1-int (ERV1-3-EC_I-int)	LTR/ERV1
Zebrafinch	taeGut1	chr27:4207784-4209022	1.1e-19	ERV1-int (TguERV2_I-int)	LTR/ERV1
Chicken	galgal4	chrW_JH375236_random:9151-8296	4.1E-14	ERV1-int (GGLTR7-int)	LTR/ERV1
Turtle	chrPic1	AHGY01435164:4348-5025	1.3e-15	ERV1-int (ERV1-1B_CPB-I)	LTR/ERV1
Xenopus	xenLae2	chr4S: 78506971-78506302	1.5e-14	ERV1-int	LTR/ERV1
Coelacanth	latCha1	JH128184:158800-157936	2.5e-17	ERV1-int	LTR/ERV1
Stickleback	gasAcu1	chrUn:56084321-56084993	1.7E-07	ERV1-int (Gypsy-30_GA-I)	LTR/ERV1
Takifugu	fr3	HE592075:741-39	4E-14	ERV1-int (FERV-R2_I-int)	LTR/ERV1
Spiny chromis	ASM210954v1	MVNR01000444.1:145194-144494	1.60E-08	ERV1-int	LTR/ERV1

Cichlid	NeoBri1.0	JH422275.1:16925254-16925968	2.20E-10	ERV1-int	LTR/ERV1
Medaka	Om_v0.7.RACA	NVQA01006789.1:101671-102370	5.70E-12	ERV1-int	LTR/ERV1
Turbot	ASM318616v1	chr4:26831079-26830372	4.50E-10	ERV1-int	LTR/ERV1
Atlantic cod	gadMor3.0	chr4:29532684-29533397	2.1e-12	ERV1-int	LTR/ERV1
Zebrafish	danRer7	chr15:43196012-43195219	1.9e-14	ERV1-int (ERV1-3-I_DR)	LTR/ERV1
Whale shark	GCF_001642345.1_ASM164234v2	NW_018046349.1:22177-21460	2.8e-19	ERV1-int	LTR/ERV1
Elephant shark	Callorhinchus_milii-6.1.3	KI636671.1:5058-5765	1.7E-14	ERV1-int	LTR/ERV1
Lamprey	Pmarinus_7.0/petMar2	Not found	NA	NA	NA
Lancelet	braflo2	Bf_V2_65:4822224-4821722	3.6E-07	(TC)n	Simple_repeat
Sea Urchin	strPur2	Scaffold69942:23181-22682	1.70E-09	(TC)n	Simple_repeat
Sea anemone	nemVec1	Not found	NA	NA	NA
C. elegans	ce10	Not found	NA	NA	NA
Drosophila	dm6	Not found	NA	NA	NA

This table list the top hit of the identified repeat type, after searching remote homology using nhmmer[65]. Specific repeat annotation, wherever available, is written under the parathesis in column 5. NA: Not applicable. Not found: wherever no hit is being obtained above the inclusion threshold (E-value<0.01).