



HAL
open science

Towards resolving the transcription factor network controlling myelin gene expression.

Debra L. Fulton, Eric Denarier, Hana C. Friedman, Wyeth W. Wasserman, Alan C.
Peterson

► **To cite this version:**

Debra L. Fulton, Eric Denarier, Hana C. Friedman, Wyeth W. Wasserman, Alan C. Peterson. Towards resolving the transcription factor network controlling myelin gene expression.. *Nucleic Acids Research*, 2011, 39 (18), pp.7974-91. <10.1093/nar/gkr326>. <inserm-00638447>

HAL Id: inserm-00638447

<https://inserm.hal.science/inserm-00638447v1>

Submitted on 4 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Towards resolving the transcription factor network controlling myelin gene expression

Debra L. Fulton^{1,2,3,4}, Eric Denarier^{2,3,4}, Hana C. Friedman^{2,3,4}, Wyeth W. Wasserman^{1,*} and Alan C. Peterson^{2,3,4,*}

¹Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, V5Z 4H4, ²Department of Oncology, ³Department of Human Genetics and ⁴Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec H3A 2B4, Canada

Received February 28, 2011; Revised April 20, 2011; Accepted April 21, 2011

ABSTRACT

In the central nervous system (CNS), myelin is produced from spirally-wrapped oligodendrocyte plasma membrane and, as exemplified by the debilitating effects of inherited or acquired myelin abnormalities in diseases such as multiple sclerosis, it plays a critical role in nervous system function. Myelin sheath production coincides with rapid up-regulation of numerous genes. The complexity of their subsequent expression patterns, along with recently recognized heterogeneity within the oligodendrocyte lineage, suggest that the regulatory networks controlling such genes drive multiple context-specific transcriptional programs. Conferring this nuanced level of control likely involves a large repertoire of interacting transcription factors (TFs). Here, we combined novel strategies of computational sequence analyses with *in vivo* functional analysis to establish a TF network model of coordinate myelin-associated gene transcription. Notably, the network model captures regulatory DNA elements and TFs known to regulate oligodendrocyte myelin gene transcription and/or oligodendrocyte development, thereby validating our approach. Further, it links to numerous TFs with previously unsuspected roles in CNS myelination and suggests collaborative relationships amongst both known and novel TFs, thus providing deeper insight into the myelin gene transcriptional network.

INTRODUCTION

Throughout the mammalian nervous system, large caliber axons are surrounded by myelin sheaths elaborated by glial cells. As demonstrated by the severe clinical consequences of acquired and inherited myelin disruption, such sheaths serve a critical role in the nervous system. Myelin is elaborated in the peripheral nervous system (PNS) by Schwann cells, derived from neural crest, and in the central nervous system (CNS) by oligodendrocytes, derived from neuroectoderm. The lipid-rich myelin sheath is composed of spirally wrapped glial cell plasma membrane that is compacted into closely apposed layers through the action of multiple myelin associated proteins. Despite their distinct embryological origins, cellular architecture and maturation programs, Schwann cells and oligodendrocytes accumulate several myelin proteins in common. However, features unique to CNS and PNS myelin also are well documented extending to differences in myelin proteins [reviewed in (1)] and interlamellar spacing (2). Within each of these glial populations additional molecular heterogeneity and developmentally contextual expression programs are well recognized (3–5). Such diversity, combined with studies that demonstrate multiple enhancer involvement (6) suggest a potentially large and complex transcriptional regulatory architecture.

Maturation of oligodendrocytes involves multiple transitions through progenitor, pre-myelinating, myelinating and mature myelin maintaining states, each distinguished by the expression of specific marker proteins (7). Initiation of myelin synthesis is associated with a rapid and marked increase in the expression of numerous genes encoding myelin proteins (referred to hereafter as ‘myelin genes’). In the mouse, myelin synthesis follows a protracted

*To whom correspondence should be addressed. Tel: +514 934 1934 Ext. 35846; Fax: +514 843 2819; Email: alan.peterson@mcgill.ca
Correspondence may also be addressed to Wyeth W. Wasserman. Tel: +604 875 3812; Fax: +604 875 3819; Email: wyeth@cmmt.ubc.ca
Present address:

Eric Denarier, Institut National de la Santé et de la Recherche Médicale Unité 836, Institut des Neurosciences de Grenoble; Université Joseph Fourier, 38042 Grenoble et Commissariat à l’Énergie Atomique, 38054 Grenoble, Cedex 9, France.

developmental program initiating throughout the PNS and spinal cord (CNS) near birth. More rostral regions of the CNS including optic nerves, brain stem and deep cerebellum initiate myelin synthesis only 5–6 days later with more dorsal regions following in a stereotyped program over the next 2 weeks (8). At the completion of myelin synthesis, myelin gene expression decreases to moderate levels that are maintained throughout maturity. Of relevance to this investigation, enriched populations of pre-, post- and actively myelinating glia emerge in definable domains of the nervous system during specific periods of post-natal development.

Previous expression analyses during oligodendrocyte maturation suggest that both lineage progression and myelin protein production are accompanied by major changes in gene expression (9). The basic helix-loop-helix transcription factor (TF) family, including: OLIG1, OLIG2, ASCL1 and the Id proteins, appears to play significant roles in early oligodendrocyte lineage specification [see reviews in (10,11)]. Other key TF classes appear to operate at both early and late stages of maturation and include homeodomain (Nkx), high-mobility group (Hmg), Pou, Sox, Egr (EGR1) and zinc-finger (ZFP488) family members.

Eukaryote gene transcription is modulated in part by the binding of multiple TFs to ‘promoter’ sequences (12). In this report, we use the term ‘promoter’ to refer to non-coding regulatory sequences proximal to (basal promoters) and distal from (enhancers) the transcription start site (TSS). While computational analyses have been used to predict sets of contributing TFs (13), non-coding sequence is vast making such predictions challenging. To refine the search space, TF binding site (TFBS) detection algorithms typically exploit evolutionarily conserved sequences or phylogenetic footprints (14) to identify candidate regulatory regions under presumed selective pressure. Accordingly, conserved sequences associated with numerous genes demonstrate regulatory function (15–17) and in such active enhancer regions clusters of participating regulatory TFBS, generally referred to as cis-linked regulatory modules (CRMs), are frequently observed (18,19).

A transcriptional regulatory network (TRN) documents both TFs and the genes that they regulate in a biological process. Recent studies have incorporated predicted TRNs to reveal the TFs involved in developing systems, such as neural crest formation (20). While cooperative TF relationships regulating cell fate specification and differentiation in the oligodendrocyte lineage have been documented (21–24), experimental strategies capable of advancing our understanding of the collaborating TFs directing the myelin gene regulatory system remain limited. The aim of this study was to further elucidate the transcriptional machinery responsible for expression of myelin associated genes by establishing a myelin gene TRN model. To achieve this aim, we first used a controlled mouse transgenesis procedure to supplement a known set of confirmed oligodendrocyte myelin gene regulatory sequences (25,26). We next established and validated a CRM detection procedure that achieves selectivity through identification of statistically over-represented

regulatory element combinations. This CRM detection algorithm was applied to the promoter regions of a set of oligodendrocyte genes determined to be co-expressed at or near the initiation of myelin formation. Reconciling the promoter predictions with sequence properties identified in the functionally defined enhancer collection highlighted shared CRM predictions that were amassed to obtain a TRN model of coordinate myelin gene transcription. Finally, we performed TRN model data validations using experimental data and demonstrated its capacity to recognize and expand on previously validated TFs involved in myelin gene transcription. Significantly, the TRN model identifies previously uncharacterized TFs that appear in oligodendrocyte genome-wide expression datasets and predicts collaborative roles amongst both known and novel TFs.

MATERIALS AND METHODS

Overview of methodology

See [Supplementary Figure S1](#) process flow diagram.

Selection of conserved regions for validation in mice

The University of California, Santa Cruz (UCSC) browser human–mouse–dog non-coding sequence alignments (Mouse May 2005 mm7—based on Build 35 assembly by NCBI) (27) for myelin-associated genes were qualitatively reviewed to identify well-conserved putative regulatory regions. Previously tested conserved regions (by other labs) associated with the *Pou3f1* and *Olig1* genes were not considered. We selected non-coding regions for the following genes ([Supplementary Tables S1 and S2](#), and [Supplementary Figure S2](#)): (i) An intergenic sequence 2.5 kilobases (kb) downstream *Cldn11* gene; (ii) a region in the first intron of *Cnp*; (iii) an intergenic region 5.8 kb downstream of the *Mal* gene; (iv) a region in the first intron of *Gjb1*; (v) a region just 5' of the TSS of *Ermn*; (vi) a conserved region 17 kb upstream of *Olig2* gene; (vii) a region 30 kb downstream of the *Olig1* gene and (viii) a sequence 8 kb downstream the *Pou3f1* gene.

Isolation of genomic DNA sequences

The conserved regions were amplified by PCR with Taq DNA polymerase on genomic DNA after selection of the primers in the surrounding sequence using Primer3 (28). Restriction sites *AscI* and *XhoI* were added to the primers for insertion and digestion ([Supplementary Table S3](#)).

Generation of reporter constructs

PCR products were digested and subcloned into *AscI* and *XhoI* sites upstream of the heat shock protein (HSP) promoter in an *HSPeGFPLacZ* Entry vector (25), with the exception of the *Ermn*-associated sequence, which was inserted into the eGFPLacZ Entry vector (a similar vector where the HSP promoter is removed). These reporter constructs were recombined into a ‘Gateway’

Destination vector bearing *Hprt* homology arms using the LR clonase reaction kit (Invitrogen). The final destination vectors were amplified, sequenced across the insert and linearized by restriction enzymes *AgeI* or *Sall* and transfected into hybrid embryonic stem (ES) cells, bearing a deletion of the promoter and exon 1 of the *Hprt* gene, as previously described (15,29). The *Hprt* gene is restored through recombination, which allows the cells to survive after hypoxanthine, aminopterin, thymidine (HAT) selection (30). Positive clones were sequenced at the McGill University and Génome Québec Innovation Center (Montreal, Quebec, Canada) using the forward primer 5-CGCTTGTCTCTGGATGGAAC-3 located in the HSP promoter and the reverse primer 5-AGCCTGGGC AACAGAGAAATATC-3 located in the *Hprt* homology arm. Positive clones were injected into blastocysts to derive chimeras. Subsequent mating to C57Bl/6 females led to recovery of germ-line passage offspring and all subsequent breeding was designed as a back-cross to C57Bl/6. The animal use experimental protocols were approved prior to the study by the Research Institute of the McGill University Health Centre—Facility Animal Care Committee (for Royal Victoria Hospital and Meakins-Cristie Laboratories sites).

Histochemistry, fluorescence microscopy and immunocytochemistry

Wholmount histochemical detection of β -galactosidase activity was performed as described previously (15). Anesthetized mice were perfused with 4% PFA in phosphate buffered saline (PBS). Brains were incubated overnight in fixation buffer at 4°C. Brains were transferred to 30% sucrose the following day, left overnight at 4°C, and then frozen. Thirty micrometer cryosections were evaluated for immunofluorescence and/or direct GFP detection. Primary antibodies: Rabbit anti-GFP antibody (Molecular Probes 1/200), Rat anti-myelin basic protein (Chemicon 1/500), Rat anti PDGFr (BD Pharmingen 1/500) and secondary antibodies: Goat anti-Rabbit Alexa 488 for GFP detection (Molecular Probes 1/1000), and Goat anti-Rat Cy3 for the detection of the other antigens (Jackson ImmunoResearch 1/1000) were incubated at 4°C overnight. Nuclei were labeled with Hoescht.

Gene expression profiling analyses

Dataset 1: P4 versus P10 mouse optic nerve expression data. Optic nerves were dissected and placed into RNAlater™ (Ambion Inc.) and left at 4°C overnight before being stored at -80°C. A total of 75 postnatal (P) Day 4 (P4) C57Bl/6J mice and 70 P10 C57Bl/6J mice were used for microarray Experiment 1, and 55 P4 and 100 P10 mice, the latter divided into two pools, were used in Experiment 2. Each pool of optic nerves was homogenized in Trizol (Life Technologies) using a glass/glass homogenizer. Following chloroform extraction, the lysate was purified using an RNeasy column (Qiagen). RNA quality was assessed using an Agilent Bioanalyzer at the McGill University and Génome Québec Innovation Center (McGill-GQIC). The probe preparation,

hybridization and scanning of microarrays were performed at the McGill-GQIC according to the manufacturer's instructions using Affymetrix Mouse Expression 430 A Array (MOE430) microarray chips. Background correction and normalization were performed in the R environment (31) with the Bioconductor packages (32) using the robust multichip analysis method (RMA) (33,34) of the R Affy package (35). The microarray data was deposited in the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under series number GSE24513. The mouse Affymetrix chip probes were mapped to mouse NCBI Entrez Genes (36) using Bioconductor packages (32). A set of mouse TF genes (37) was mapped to the mouse Affymetrix probes using a developed Perl program. Differential expression analyses for the P4 and P10 datasets were performed using a two-sample *t*-test with a random variance model (38) and the Benjamini and Hochberg method to compute a false discovery rate (FDR—expected proportion of Type I errors within the rejected hypotheses) implemented in the BRB-Array software (<http://linus.nci.nih.gov/~brb/>). Perl scripts were created to convert each of the HTML-formatted expression analysis results to text files and extract and report all significantly ($P \leq 0.001$ and $FDR \leq 0.05$), differentially expressed genes across the pairwise expression profiles.

A GO term enrichment analysis was performed on the differentially expressed set of genes with the AmiGO annotation and ontology toolkit (39) using a hypergeometric test, which incorporated the GO annotations for the Mouse Genome Database (40) and Rat Genome Database (41).

Dataset 11: oligodendrocytes progenitor cells versus early oligodendrocytes and early oligodendrocytes versus mature oligodendrocytes mouse forebrain expression data. Affymetrix Mouse 430 2.0 chip CEL files were downloaded from NCBI GEO (GEO dataset ID: GSE9566) (42). CEL files recording expression for oligodendrocyte progenitor cells (OPCs), early oligodendrocytes (EOLs) and mature oligodendrocytes (MOLs) were analyzed using RMA (as described above) to obtain individual probe set expression values and each pair of experiments was subjected to a two-sample *t*-test with a random variance model (38) implemented in the BRB-Array software (<http://linus.nci.nih.gov/~brb/>). The mouse Affymetrix chip probes were mapped to mouse NCBI Entrez Genes (36) using Bioconductor packages (32). Perl software (described above) was used to extract and report all significantly ($P \leq 0.001$ and $FDR \leq 0.05$), differentially expressed genes across the pairwise expression profiles and TF genes were identified as described above.

Gene expression profiles were compared across the P4 versus P10 (P4-P10d); OPCs versus EOLs (OPC-EOLd) and EOLs versus MOLs (EOL-MOLd) datasets for differential gene expression overlap, enhancer library co-expression and TF-mapped differential gene expression using developed software. The intersection of oligodendrocyte early development expression dataset (IOLEdd) was established using the maximal overlap of

differentially expressed genes between the P4–P10d and OPC–EOLd gene expression datasets.

Evaluation of local sequence conservation for previously validated TFBS

Software was developed to extract human and mouse aligned genomic sequences from the UCSC 28-way multi-species alignments data (43) around annotated TFBS defined in the Annotation Regulatory Binding Site Database (ABS) (44) available within the PAZAR system (45). Sequence identity values for regions surrounding TFBS instances were computed and reported.

Development of the promoter database and CRM analysis algorithm

A database of human–mouse non-coding alignments was developed and populated using a human–mouse subset of the UCSC 28-way alignment data (43,46). All UCSC Hg18 multiz28way chromosome and database files were downloaded from a UCSC FTP site (<ftp://hgdownload.cse.ucsc.edu/>). Kent source utilities were downloaded and compiled on a linux server and Perl-C wrappers were developed for selected C programs. Software was developed to extract 15006 alignments from the chromosome databases based on pre-computed human coordinates in the oPOSSUM database (47). Human-anchored alignment subsets for human and mouse sequences, with aligned gaps removed and exons masked, were saved in MySQL database tables and coordinates for each non-contiguous mouse alignment in a multiple alignment format (MAF) block were recorded. Binding site information was compiled from the literature for a set of glia-related TFs: POU2F1; EGR1; EGR2; EGR3; EGR4; POU3F1; NKX2-2; NKX2-5 to supplement the Jaspar TF binding profile database (Supplementary Figure S3). All Jaspar database TFBS profiles, along with the supplemented profiles, were used to enumerate predicted TFBS motifs in the human–mouse aligned regions requiring only that a predicted motif overlap across the aligned sequences. On the basis of similarity between the binding properties of TFs in the same family, the Combination Site Analysis (CSA) algorithm identifies statistical over-representation of combinations of TFBS class-representatives in a set of co-expressed genes, as compared with a background set (48). For each enriched class-level CRM, the predicted TFBS classes are expanded and each possible TFBS-level CRM is evaluated for over-representation. The CSA⁺ algorithm was developed to incorporate the improved set of regulatory sequence feature predictions and utilize two random background samples for the class combination and TFBS combination evaluation steps.

Validation of promoter analyses method

A reference collection of 25 skeletal muscle genes with defined CRMs was utilized for the validation of the promoter analyses approach (<http://burgundy.cmmt.ubc.ca/tjkwon/> and Supplementary Tables S4 and S5). Using this training collection as input to the CSA⁺ algorithm (described above), parameters were selected for the

boundaries of search regions. The CSA⁺ algorithm predicted combinations of two TFBS (minimal CRMs) for all analyses. A TFBS inter-binding site distance (IBSD) parameter was set to 225 base pairs (bp), consistent with past observations of TF cooperativity up to ~20 helical turns (~210 bp) (49). The analysis was performed using the full database of promoter regions for assessing background properties in addition to a random background promoter sampling procedure. A ‘noise-added’ representation of the training collection was generated for further validation of CSA⁺ by appending an additional set of 25 randomly selected genes (Supplementary Table S6). The same CSA⁺ parameter selections were utilized for all subsequent promoter analyses (Supplementary Table S7).

Jaspar profile clustering and cluster labeling

Vertebrate binding site profiles were extracted from the supplemented Jaspar database and pairwise profile Pearson correlations were computed using the CompareAce program (50). Clustering was performed using the *hclust* function in the R software package using a cut-off of 0.40. TFBS profile clusters were examined manually and each cluster with ≥ 4 members was assigned a class label that, in most cases, reflected the structural class of its members. Clusters with < 4 members were labeled with the corresponding TF name and structural class.

CRM promoter analyses of differentially expressed genes in oligodendrocytes

CSA⁺ analyses were conducted over the IOLEDd oligodendrocyte dataset using parameter values determined in the validation step (Supplementary Table S7). To maintain search regions spanning ≥ 5000 kb we omitted the 2000 bp upstream/downstream parameter in CSA⁺ analyses of the oligodendrocyte data. The mouse IOLEDd gene set was comprised of 203 mouse Entrez genes associated with 202 mouse Ensembl genes. Mapping the mouse ensembl genes to human orthologs resulted in 194 human Ensembl gene identifiers, of which 178 were represented in the promoter database (File 1 in Supplementary Data). A second set of CSA⁺ analyses was performed using a TFBS prediction threshold of 85%, retaining all other parameter values. In total, 16 analyses were performed, each of which provided a ranked list of CRM (TFBS pair) predictions that fell below the 0.05 ranking cut-off.

Enhancer support of CRM predictions

We developed software to extract UCSC MAF 28-way human–mouse alignments (27,43) and predict TFBS in the tested myelin gene enhancer sequences (Supplementary Table S8). Predicted TFBS with aligned positions were stored in a MySQL database. Software was developed to extract and report genomic coordinates for: CRMs predicted by CSA⁺ in non-coding regions of the IOLEDd genes and CRM instances predicted in the myelin gene enhancer collection. Myelin gene enhancer constructs that were expressed in the CNS and linked to genes that demonstrated differential expression in the

IOLEdd dataset were assigned to the 'positive' enhancer group. Two enhancers that failed to direct reporter gene expression in white matter were classified in the negative group: (i) *Cnp*—its associated gene was differentially expressed in the IOLEdd dataset and (ii) *Ermm*—its associated gene (alias A330104H05Rik) was up-regulated in both the OPC_EOLd and EOL_MOLd datasets (the Affymetrix chip used to obtain the P4_P10d dataset did not have a probe for this gene). Software was developed to evaluate the coincidence of CRMs found in the CSA⁺ analyses results and in the enhancer groups. CSA⁺ CRM predictions were included in the enhancer-supported set if they had a co-occurring CRM prediction in a 'positive' enhancer and no corresponding CRM prediction in a 'negative enhancer' (Supplementary Table S9). A Perl program was developed to map the enhancer-supported CRM predictions to CRM class labels using the TF class label mapping described above.

Myelin gene TRN construction

Predicted CRMs in the enhancer-supported CRM set were extracted if the genes linked to the CRMs were identified in a list of abundant CNS myelin proteins [Figure 5a in (51)]. Software was developed to construct a TRN using the enhancer-supported myelin protein CRM predictions and visualized using BioLayout (52).

Mouse optic nerve (P4–P10) and mouse forebrain (P16) expression profile comparisons were reviewed for TFs associated with TFBS classes identified in the TRN. A P7 rat brain oligodendrocyte dataset [GEO dataset: GSE5940 (53)] was also queried for expression of TF genes using the NCBI GEO Profiles viewer (54).

Analysis of overlapping oligodendrocyte CRM predictions

Software was written to identify pairs of enhancer-supported CRM predictions with one overlapping predicted TFBS genomic coordinate position to identify CRM predictions with three TFBS.

Myelin gene TRN database and website tool

A MySQL database was designed and developed to house the oligodendrocyte myelin TRN data. A Perl script was written to populate the database with predicted CRM genomic instances. CRM predictions identified in mouse (mm8) enhancer sequences were also converted to human (hg18) genomic coordinates and inserted into the database using the UCSC Genome Browser LiftOver C-based software (<http://hgdownload.cse.ucsc.edu/admin/execute/>) and developed Perl software. PHP software was developed to create a website tool which enables entry of parameter selections and data extraction from the MySQL database (www.cisreg.ca/MyelinCode). An interface to the UCSC Browser display (27) was developed to facilitate display of selected CRM predictions in the genome browser.

Access to the website tool is available at www.cisreg.ca/MyelinCode. Additional material is available at: www.cisreg.ca/MyelinCode/supporting_material.

RESULTS

Myelin gene associated conserved regions confer reporter gene activity in glial cells

We showed previously that regulatory sequences associated with the myelin genes, myelin basic protein (*Mbp*) and proteolipid protein (myelin) 1 (*Plp1*), often coincided with domains highly conserved between human and mouse (15,25,26,55). To expand on this enhancer set, eight similarly conserved sequences associated with genes known to function in myelinating cells were evaluated. These sequences were located adjacent to the coding regions of mouse genes: claudin 11 (*Cldn11*); 2',3'-cyclic nucleotide 3' phosphodiesterase (*Cnp*); ermin, ERM-like protein (*Ermm*); connexin 32 (*Gjb1*); myelin and lymphocyte protein, T-cell differentiation protein (*Mal*); oligodendrocyte TF 1 (*Olig1*); oligodendrocyte TF 2 (*Olig2*) and pou domain class 3, TF 1 (*Pou3f1*) (Supplementary Figure S2). With one exception, the putative regulatory regions were ligated to an eGFP-lacZ reporter gene with a minimal HSP promoter (Figure 1). In contrast, the *Ermm*-associated reporter construct included its endogenous promoter. These constructs were inserted in the same orientation in single copy at a common site 5' of the *Hprt* locus and their *in vivo* expression phenotypes evaluated in transgenic mice at multiple developmental ages.

Of the candidate non-coding regions tested, three were proximal to genes encoding the tetraspan myelin protein family members: CLDN11, GJB1 and MAL. CLDN11 is an integral membrane protein contributing ~1% of CNS myelin protein (51). A 3' 631 bp sequence directed widespread CNS expression that initiated in pre-myelinating oligodendrocytes at P5 (Figure 2 and Supplementary Figure S4) and persisted in mature myelinating oligodendrocytes at P10 and P90. Reporter gene expression was absent in the PNS (Supplementary Figure S5). GJB1 is a transmembrane gap junction subunit expressed in oligodendrocytes and Schwann cells (56). A 637 bp sequence located in intron 1 corresponds to an orthologous human sequence containing Sox TFBS that, when mutated, are associated with cases of Charcot-Marie-Tooth Disease (57,58). This region directed reporter expression to both the CNS and PNS (Figure 2; Supplementary Figures S4 and S5). In P5 samples, the reporter gene expressed diffusely in both brain and optic nerves while spinal cords labeled intensely from cervical through lumbar levels (Figure 2; Supplementary Figures S4 and S5) consistent with expression in both OPCs and myelinating oligodendrocytes. Reporter expression in white matter continued into maturity. Additionally, reporter expression was robust in myelinating Schwann cells (spinal roots and sciatic nerves) at P5 and P10 but not in cells maintaining mature myelin sheaths (P6 months) (Supplementary Figure S5). MAL is a tetraspan raft-associated proteolipid, which regulates sorting and trafficking of membrane components in myelinating cells (59). A 5' 680 bp sequence directed reporter expression to myelin forming cells in both the CNS and PNS; at P5, spinal cord and peripheral nerves were densely labeled while the chiasmatal end of optic

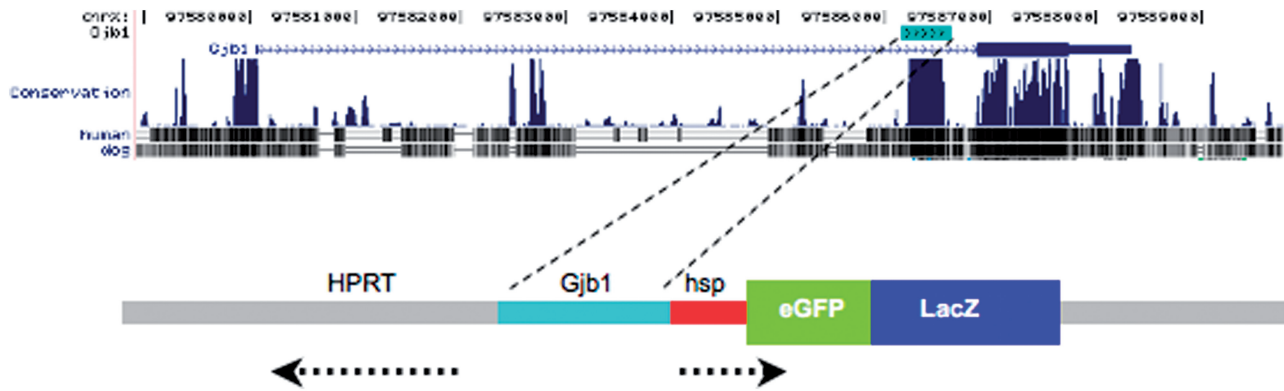


Figure 1. Putative myelin gene-associated regulatory regions were tested for function in mice using reporter constructs inserted in single copy at the *Hprt* locus. (UCSC genome browser displayed using mouse MM8-mapped coordinates <http://genome.ucsc.edu>).

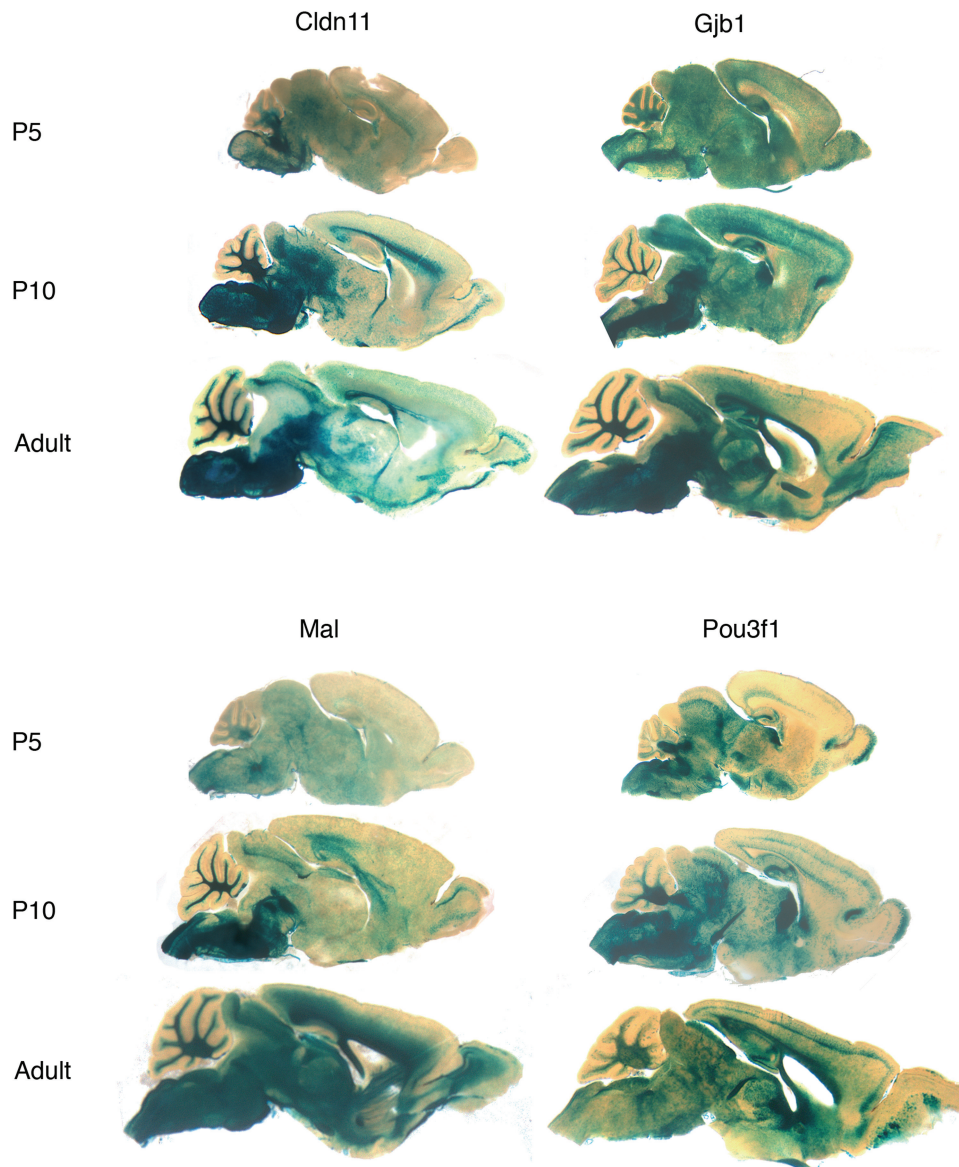


Figure 2. Histochemical detection of β -galactosidase activity in brain sections from P5, P10 and 2–3-month-old mice reveals targeting and developmental expression programming of *Cldn11*, *Gjb1*, *Mal* and *Pou3f1* reporter constructs.

nerves and deep regions of the cerebellum were lightly labeled (Figure 2; Supplementary Figures S4 and S5). Expression continued in white matter and in the PNS at P18 and P5 months (Supplementary Figure S5).

Two of the conserved non-coding regions tested were associated with glial genes encoding the cytoplasmic proteins ERMN and CNP. *Ernm* encodes a cytoskeletal protein that is expressed during late stage myelination (60). At P5 and P10, a 5' 587 bp sequence conferred sparse expression in dorsal spinal cord white matter but not in brain. In mature mice, oligodendrocytes in the cerebellum and brain stem labeled weakly (Supplementary Figures S5 and S6). In the PNS, Schwann cells did not label at any age (Supplementary Figures S5 and S7). CNP accumulates in uncompact regions of the myelin sheath (61). A reporter construct bearing a 742 bp sequence from intron 1 of the CNP gene directed no detectable expression in either the CNS or PNS at any age (data not shown).

Putative enhancer regions associated with three genes encoding TFs with regulatory roles in oligodendrocytes, *Pou3f1*, *Olig1* and *Olig2*, were evaluated. POU3F1, encodes a Pou-homeodomain containing TF expressed in cerebral cortex OPCs, astrocytes (62–64) and Schwann cells (65,66). A 3' 650 bp sequence that encompasses a portion of a previously identified Schwann cell enhancer sequence (67) was evaluated. Reporter expression was observed throughout the CNS at P5, was increased in white matter at P10, and continued at reduced levels throughout maturity (Figure 2; Supplementary Figures S4 and S5). Schwann cells were strongly labeled at P5 and P10 but were weakly labeled in adults (Supplementary Figures S4 and S5). Expression was also observed in several neuronal populations in the hippocampus and the dentate gyrus. OLIG1 is a basic helix-loop-helix TF required for normal oligodendrocyte development (68). A 3' 1101 bp sequence conferred labeling in neurons throughout development and trace levels of expression in the white matter tracks of cerebellum and brain stem at P10 but not at other time points (Supplementary Figures S5–S7). A 5' 894 bp sequence from *Olig2* directed low-level expression in optic nerves of P10 mice. Sparse expression was observed elsewhere in the white matter of the P10 brain and at later time points (Supplementary Figures S5–S7) although robust expression was observed in the lateral septal nucleus and medial cerebellar nucleus at P5 and P10.

To confirm that the *Cldn11*, *Gjb1*, *Mal* and *Pou3f1* reporter constructs expressed in cells of the oligodendrocyte lineage, sections from P10 brain were co-labeled for the OPC marker, platelet derived growth factor receptor alpha (PDGFRA) and the later appearing myelination marker, MBP. In the P10 brain stem, myelination has progressed and oligodendrocytes expressing each of these constructs co-labeled for MBP (Supplementary Figure S8A and B). In contrast, in the P10 cortex, myelination has yet to initiate but cells expressing each of the constructs co-labeled for the oligodendrocyte progenitor marker PDGFRA (Supplementary Figure S9). Thus, as predicted by the reporter gene expression profiles, the *Cldn11*, *Gjb1*, *Mal* and *Pou3f1* enhancer sequences function in the oligodendrocyte lineage.

Coordinate myelin gene expression is detected in mouse forebrain and optic nerves

Transcriptional profiles highlight genes that are temporally co-expressed in a given cell state (69,70) and promoter analyses of such coordinately expressed genes may implicate mediating DNA-binding TFs. In contrast, non-cohesive groupings of expressed genes seldom reveal insights into regulatory mechanisms. Thus, a refined and reliable set of co-expressed genes is key for detection of enriched TFBS in their promoter sequences.

Optic nerves are a structurally uniform CNS domain containing axons projecting from retinal ganglion neurons with few other neural components. In the mouse, oligodendrocytes initiate myelination in optic nerves late on P5 and myelination is advanced by P10. To identify genes induced during the myelination process, we generated gene expression profiles at P4 and P10. Our comparative analyses identified 487 differentially expressed genes at a *P*-value cutoff of 0.001 (File 2 in Supplementary Data).

To illuminate potential functional roles for the differentially expressed genes, we mapped the 487 partially redundant mouse gene identifiers to a set of 504 mouse and orthologous rat genes and performed a Gene Ontology (GO) (71) over-representation analysis using the GO mouse and rat annotation databases and the Molecular Function and Cellular Component branches (Supplementary Tables S10 and S11). The two most significant GO categories highlighted in these analyses were: GO:0005488 'binding' and GO:0019911 'structural constituent of myelin sheath' for the Molecular Function sub-tree (Supplementary Table S10) and GO:0043209 'myelin sheath' and GO:0044421 'extracellular region part' in the Cellular Component sub-tree (Supplementary Table S11), confirming that a portion of the genes differentially expressed at P4 and P10 encode myelin structural proteins.

During the process of lineage progression, oligodendrocytes transition through a developmental process in which four cell states are recognized: progenitors, pre-myelinating, myelinating and myelin maintaining cells, each distinguished by specific oligodendrocyte protein markers. While more mature oligodendrocytes will be present in the P10 optic nerve compared to P4, limited synchronicity of cell lineage progression will result in heterogeneous populations at both time points. In contrast, a recent study designed to establish stage-specific gene expression profiles exploited stage specific markers to recover populations of oligodendrocyte lineage cells at multiple stages of maturation from the P16 mouse forebrain (42). As these marker-enriched populations and the optic nerve samples represent CNS domains divergent in both structure and developmental programming, any myelin genes with differential gene expression concordance may be components of a core set used throughout the lineage. Therefore, we reanalyzed the forebrain data using a method consistent with the optic nerve expression profile analyses and searched for overlap of differentially expressed genes. Using only those genes represented on both microarray chips (Figure 3A and File 2 in

Supplementary Data) the P4 versus P10 optic nerve dataset (P4–P10d) was compared to the oligodendrocyte progenitor cells versus early oligodendrocytes (OPC–EOLd) and early oligodendrocytes versus myelinating oligodendrocytes (EOL–MOLd) datasets. We found that the maximal overlap of differentially expressed genes occurred between the P4–P10d and OPC–EOLd gene expression datasets (Figure 3A and File 3 in **Supplementary Data**), referred to here as the intersection of oligodendrocyte early development expression dataset (IOLEDD). The IOLEDD expression dataset is composed of 203 genes that are either concordantly up- or down-regulated during oligodendrocyte maturation. Notably, five of the myelin genes associated with enhancers in the supplemented myelin enhancer collection were differentially expressed in the IOLEDD expression dataset (Figure 3B). Moreover, comparison of this dataset to a recent proteomic analysis of myelin (51) illuminated that additional myelin genes were captured within the IOLEDD set.

TFs that are co-expressed with tissue-specific genes may be causally associated with the biological process under study (72,73). Therefore, we mapped the optic nerve and forebrain expression data to a curated mouse–human TF catalogue to identify the subset of differentially-expressed

genes that encode transcriptional regulators (37). A subset of 31 TFs in the IOLEDD set was identified and the other sets of differentially expressed genes included 829 candidate TFs (Figure 3C and File 4 in **Supplementary Data**). Notably, TFs previously shown to regulate oligodendrocyte proliferation and differentiation, as well as novel TFs, were captured in these subsets. To illuminate the CRMs responsible for oligodendrocyte development, we used a novel procedure described below to analyze the non-coding regions of the IOLEDD set of differentially expressed genes.

Promoter analyses

Promoter analyses methods typically focus on non-coding regions that satisfy minimum sequence conservation criteria (e.g. $\geq 60\%$ nucleotide identity in sequence alignments of orthologous genes separated by ~ 60 million years of evolution) over a given length of sequence (e.g. 100 bp). However, orthologous non-coding sequences identified by alignment algorithms can exhibit a range of sequence conservation and/or constraint (74) and recent gene regulatory and chromatin immunoprecipitation (ChIP) studies have highlighted binding and activity of

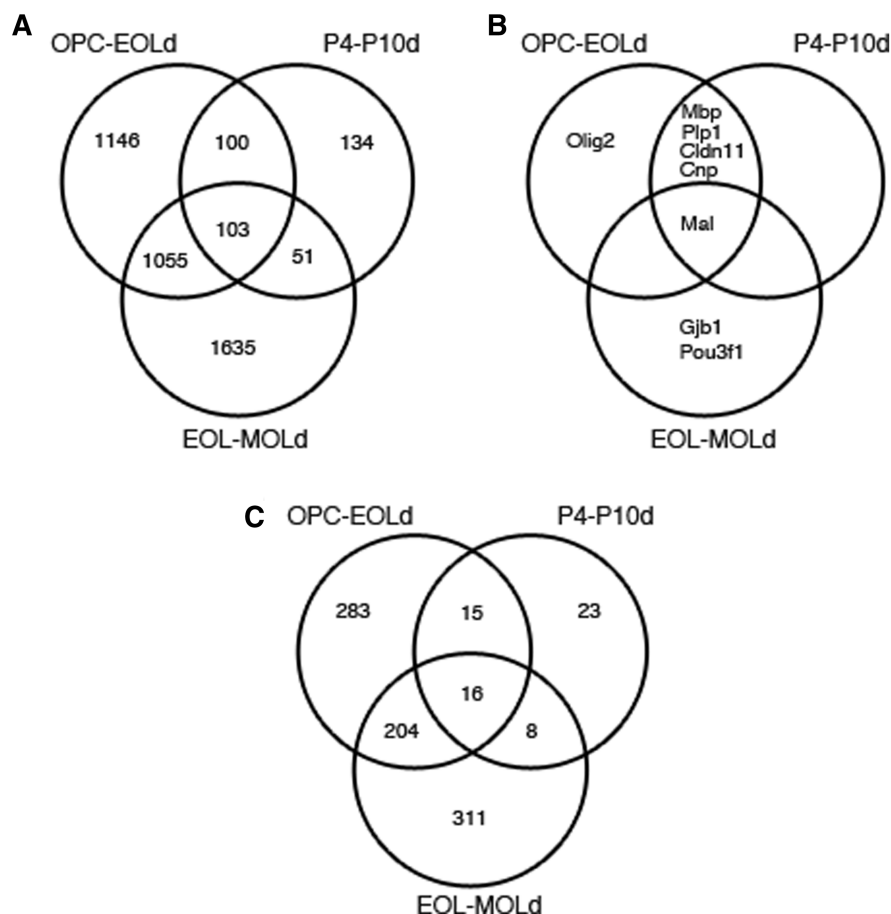


Figure 3. Overlap of differentially expressed genes in the mouse optic nerve and forebrain datasets. (A) Oligodendrocyte gene expression profile comparisons for optic nerve (P4–P10d) and forebrain (OPC–EOLd and EOL–MOLd) were assessed for concordance of differential gene expression. (B) Identification of the overlap concordance for genes linked to the enhancer constructs. (C) Overlap of differentially expressed genes that are TFs.

TFs over sequence regions that possess little or no inter-species sequence identity (75,76). Our evaluation of regions surrounding experimentally validated TFBS in aligned sequences (44) suggested that local sequence conservation flanking TFBS can fall below typically applied threshold cut-offs (Supplementary Figure S10). Accordingly, to facilitate regulatory element detection in sequences with more diverse conservation values, we designed a new CRM detection approach called CSA⁺, that requires only that each *aligned sequence block* (77) be evaluated for aligned binding site predictions. Other binding site conservation approaches have been incorporated in previous promoter sequence analysis studies [for examples see (78,79)].

To validate the capacity of the CSA⁺ algorithm to identify known CRMs, we first applied it to a reference collection of 25 human skeletal muscle-related co-expressed genes (Supplementary Table S4) containing previously characterized CRMs. A procedure for the detection of over-represented TFBS motif combinations was evaluated, which compared motif frequencies in a co-expressed gene set against background sets composed of either random selections of 5000 gene promoters found in our promoter database (see ‘Materials and Methods’ section) or all available gene promoters in the database. The CSA⁺ algorithm predicted minimal CRMs, namely combinations of two TFBS, for all analyses. We applied the reference collection to both approaches and arbitrarily extracted the top 12 ranked TFBS pair results for comparison. The same 4 known muscle CRMs were predicted in the top 12 results by both background methods (Supplementary Tables S12 and S13), while an additional TFBS CRM, was identified using the randomly sampled background method (Supplementary Table S13B.). Importantly, the CSA⁺ method demonstrated refinements in the rankings of known CRMs compared with previous validations using a similar reference collection (47,48). Given the results of the background validations, we hypothesized that a randomized sample of genes may minimize any impact of promoter characteristic properties that might dominate the full database of orthologs. As such, the random background sampling method was incorporated in all subsequent CSA⁺ analyses.

Among the group of differentially expressed genes recognized through gene expression comparisons, some are likely to be co-regulated by a common transcriptional regulatory program while others may be controlled by unrelated regulatory pathways. To evaluate the capacity of the CSA⁺ algorithm to identify over-represented CRM mechanisms in a set of genes that are controlled by different transcriptional programs, an equal number of randomly selected genes was added to the validation reference test set (Supplementary Table S6) and reanalyzed with CSA⁺. The majority of the known CRMs continued to be ranked among the top 12 results (Supplementary Table S14). One TFBS combination, correctly predicted in the ‘specific’ reference collection promoter analyses did not satisfy the score cut-off in the analyses of the ‘noise-added’ test set, while an additional known CRM was raised to rank 11 from rank 14 (Supplementary Tables S13 and S14). Such changes are not surprising as

the frequency of TFBS class pair instances can be obscured and/or altered in larger, less cohesive gene sets, supporting the view that promoter analysis specificity will be maximized when using cohesive sets of co-expressed genes. We retained and incorporated this ‘noise-added’ representation of the muscle reference collection (Supplementary Table S6) in the remainder of our CSA⁺ algorithm validation to approximate the noise that may be present in the differentially expressed oligodendrocyte IOLEdd dataset.

To test the capacity of the promoter analysis method to recover CRMs in varying lengths of search sequence, potentially spanning more than one TSS, sequence range values between 100 and 10000 bp were evaluated (Supplementary Table S7). The results confirmed that discovery of over-represented TFBS combinations was not obfuscated by the length of non-coding regions analyzed.

Promoter analyses of differentially expressed oligodendrocyte genes highlights potential TF cooperativity

The promoter analyses method and search parameters defined above were applied to the IOLEdd oligodendrocyte gene set to elucidate the CRMs that may be responsible for their co-expression. Individual TFBS predictions often implicate the binding of homologous TFs with similar binding properties. Consequently, we clustered the vertebrate TFBS profiles from the JASPAR profile database (80) supplemented with additional TFBS profiles (Supplementary Figure S3) using a binding profile similarity measure. This led to 47 TFBS classes (Supplementary Table S15) of which over half (26/47) were composed of single TFBS profiles while the remainder contained four or more (see ‘Materials and Methods’ section). TFBS in each predicted CRM identified by the oligodendrocyte CSA⁺ analyses were mapped to their respective TFBS class labels. Four recurring class CRMs corresponded to the top 4 of 15 highest ranked hits across all analyses: SOX/HMG GRP - NKX/HOX GRP; SOX/HMG GRP - EGR/ZF GRP; SOX/HMG GRP - ETS GRP and SP1 (ZF) - SPZ1 (HLH). Importantly, most of these highly ranked CRM predictions align with previously established associations between TF families and oligodendrocytes (10,11).

A myelin gene TRN supported by enhancer predicted regulatory elements

Using the set of statistically-supported TFBS signatures enriched in the co-expressed oligodendrocyte genes (IOLEdd), we exploited the demonstrated regulatory capacity of the myelin enhancer sequence collection to improve the specificity of oligodendrocyte-associated CRM predictions. We performed analyses to extract and store all human–mouse aligned TFBS predictions found in the myelin enhancer sequence collection (Supplementary Table S8). Each enhancer included in the myelin gene enhancer collection was classified as ‘positive’ or ‘negative’ or deemed ‘not applicable’, based both on its capacity to activate the reporter gene in oligodendrocytes and co-expression of its associated gene in the IOLEdd

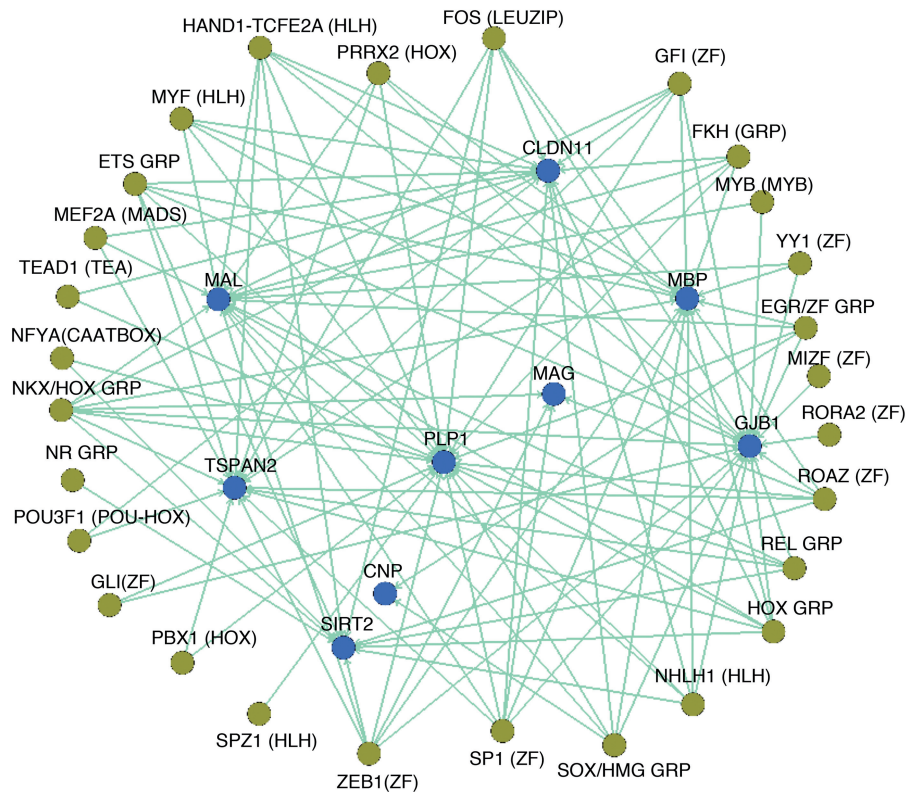


Figure 4. TRN model for myelin genes. Predicted functional linkages between TF classes (gold spheres) and targeted myelin genes (blue spheres) are indicated with connecting edges.

set (Supplementary Table S9). Using the full set of promoter analyses results, we identified and mapped all instances of the ranked TFBS pair predictions found in gene promoters to TFBS classes. This resulted in 143 distinct TFBS class CRM predictions with score values ≤ 0.05 (File 5 in Supplementary Data). To isolate CSA⁺ CRM predictions supported by enhancer sequence data, we identified the overlap with CRMs predicted in the positive enhancers but absent in the negative enhancers. This resulted in 49 unique enhancer-supported CRM predictions targeting the IOLEdd set (File 6 in Supplementary Data). Finally, we identified enhancer-supported CRM pair predictions with overlapping TFBS to distinguish IOLEdd genes with predicted CRMs made up of three TFBS. This analysis identified 300 unique CRMs with 3 TFBS (Files 7 and 8 in Supplementary Data).

A TRN model was established by extracting the associations between the list of enhancer-supported predicted CRMs and oligodendrocyte myelin proteins (51). The TRN model includes 28 unique TF classes that participate in 43 CRMs (Figure 4 and Table 1). Importantly, the network highlights previously identified regulators of oligodendrocyte development: SOX/HMG GRP, POU3F1 (POU-HOX) and NKX/HOX GRP (Table 2). Moreover, a review of gene expression data and related evidence for other TF candidates provided noteworthy support for their involvement in oligodendrocyte maturation (Table 2). For example, a recent CNS myelin proteome list (51) includes proteins (DES; ACTA1;

ACTC1; CKB) that are also components of muscle. The predicted role of Mef2a and the Myf family of TFs in the TRN model and their demonstrated expression with known target genes in developing oligodendrocytes (Table 2) suggest that these TFs also play a key role in CNS myelinogenesis. Notably, >80% of the TF classes are corroborated by experimental support and/or oligodendrocyte TF gene expression evidence. Additionally, oligodendrocyte genome-wide expression analyses support the involvement of >65% of the previously unknown TF class regulators (Table 2). Importantly, the TRN model and associated CRM predictions provide support for cooperative roles between TF classes.

TRN model predictions are supported by experimental data

To facilitate query and extraction of the TRN model data, we populated a relational database with the genomic coordinates of the CRMs predicted in the oligodendrocyte promoter analyses. A website portal was developed (www.cisreg.ca/MyelinCode) to enable data selection and extraction using the following parameters: co-expressed genes, enhancers, CRMs, TFBS and/or a genomic coordinate interval. Extracted CRMs may be downloaded in a text-tabbed file or viewed in the UCSC genome browser (27).

The TRN model data should predict both known and putative regulatory elements and TFs. To validate this

Table 1. Predicted myelin gene TRN

TFBS class I	TFBS class II	Predicted targeted myelin genes
EGR/ZF GRP	GFI (ZF)	<i>Gjb1; Mal</i>
EGR/ZF GRP	HAND1-TCFE2A (HLH)	<i>Gjb1; Mal; Mbp; Plp1</i>
EGR/ZF GRP	HOX GRP	<i>Gjb1; Mal; Mbp; Plp1</i>
EGR/ZF GRP	MYB (MYB)	<i>Gjb1; Mal</i>
EGR/ZF GRP	NHLH1 (HLH)	<i>Gjb1; Mal; Mbp; Plp1</i>
EGR/ZF GRP	NKX/HOX GRP	<i>Gjb1; Mal; Mbp</i>
EGR/ZF GRP	PBX1 (HOX)	<i>Plp1</i>
EGR/ZF GRP	PRRX2 (HOX)	<i>Gjb1; Mal; Mbp</i>
EGR/ZF GRP	SOX/HMG GRP	<i>Cnp; Gjb1; Mal; Mbp; Plp1</i>
EGR/ZF GRP	YY1 (ZF)	<i>Gjb1; Mal; Mbp;</i>
EGR/ZF GRP	ZEB1 (ZF)	<i>Gjb1; Mal; Mbp;</i>
ETS GRP	FOS (LEUZIP)	<i>Cldn11; Gjb1; Plp1; Sirt2</i>
ETS GRP	GFI (ZF)	<i>Cldn11; Gjb1; Plp1; Tspan2</i>
ETS GRP	GLI (ZF)	<i>Gjb1; Plp1</i>
ETS GRP	HAND1-TCFE2A (HLH)	<i>Cldn11; Gjb1; Sirt2; Tspan2</i>
ETS GRP	PBX1 (HOX)	<i>Plp1; Tspan2</i>
ETS GRP	ROAZ (ZF)	<i>Gjb1; Mbp; Plp1; Tspan2</i>
ETS GRP	SOX/HMG GRP	<i>Cldn11; Gjb1</i>
ETS GRP	SPZ1 (HLH)	<i>Plp1</i>
FOS (LEUZIP)	FKH GRP	<i>Cldn11; Mbp; Plp1; Tspan2</i>
FOS (LEUZIP)	NKX/HOX GRP	<i>Cldn11; Mbp</i>
FOS (LEUZIP)	NR GRP	<i>Sirt2</i>
GFI (ZF)	MEF2A (MADS)	<i>Cldn11; Mal</i>
GFI (ZF)	MIZF (ZF)	<i>Gjb1</i>
GFI (ZF)	MYF (HLH)	<i>Cldn11; Gjb1; Mal; Mbp; Plp1</i>
HAND1-TCFE2A (HLH)	HOX GRP	<i>Cldn11; Gjb1; Mbp; Plp1; Tspan2</i>
HAND1-TCFE2A (HLH)	MEF2A (MADS)	<i>Cldn11; Mal; Tspan2</i>
HAND1-TCFE2A (HLH)	NKX/HOX GRP	<i>Cldn11; Mbp; Tspan2</i>
HAND1-TCFE2A (HLH)	RORA2 (NR)	<i>Gjb1</i>
HAND1-TCFE2A (HLH)	TEAD1 (TEA)	<i>Cldn11; Plp1</i>
MEF2A (MADS)	POU3F1 (POU-HOX)	<i>Cldn11; Tspan2</i>
MEF2A (MADS)	SP1 (ZF)	<i>Cldn11; Mal</i>
NHLH1 (HLH)	NFYA (CCAATBOX)	<i>Plp1; Sirt2</i>
NHLH1 (HLH)	TEAD1 (TEA)	<i>Cldn11; Plp1</i>
REL GRP	FKH GRP	<i>Mal</i>
REL GRP	FOS (LEUZIP)	<i>Gjb1; Plp1; Tspan2</i>
REL GRP	HOX GRP	<i>Gjb1; Mal; Plp1; Sirt2; Tspan2</i>
REL GRP	PRRX2 (HOX)	<i>Gjb1; Mal; Plp1; Tspan2</i>
REL GRP	SOX/HMG GRP	<i>Gjb1; Mal</i>
ROAZ (ZF)	NKX/HOX GRP	<i>Gjb1; Mag; Mbp; Plp1; Sirt2</i>
ROAZ (ZF)	SP1 (ZF)	<i>Gjb1; Mag; Mbp; Plp1; Tspan2</i>
ROAZ (ZF)	ZEB1 (ZF)	<i>Gjb1; Mag; Mbp; Plp1; Sirt2; Tspan2</i>
SOX/HMG GRP	TEAD1 (TEA)	<i>Cldn11; Plp1</i>

capacity we reviewed the current literature for experimental data that identified combinations of acting TFs and regulatory elements in myelin gene non-coding regions. To avoid circularity, we required that experimentally evaluated regions did not overlap with and/or were not orthologous to our enhancer sequences. Two criteria-compliant studies were identified for validation of the TRN model data: (i) Li *et al.* (22) and (ii) Gokhan *et al.* (21). Using data from the first study, we extracted TRN model data predictions that fell within the genomic boundaries of the investigated promoter region and compared them with the experimentally-validated regulatory elements determined in the study. Notably, all regulatory elements identified in the study were present in the CRM predictions confirming the ability of the TRN model data to identify bonafide CRMs (Supplementary Figure S11). Remarkably, the remaining set of CRM candidates identified by the TRN model data was composed

of TFBS that bind TFs with previously demonstrated regulatory roles in oligodendrocytes. We then extracted and compared TRN model data with experimental data from the second study and confirmed that predicted CRMs were capable of engaging the TF combinations validated in the study (Supplementary Figure S12). Similarly, the predicted CRMs included putative TFBS associated with TFs known to regulate oligodendrocyte myelination, corroborating their potential involvement. Thus, we confirmed the ability of the TRN model data to identify experimentally validated TFBS and to predict the involvement of regulatory elements and TFs that warrant further investigation.

DISCUSSION

The initiation of CNS myelin elaboration is marked by the up-regulation of numerous myelin-associated genes.

Table 2. Evidence for TFs predicted in the myelin gene TRN

TFBS classes	Gene expression Evidence for associated TFs ↓ - down-regulated; ↑ - up-regulated			Literature support	
	TF	↓ ↑	Dataset	Functional support	Other gene expression support
EGR/ZF GRP	<i>Egr1</i> <i>Egr2</i> <i>Egr3</i>	↓ ↓ ↓	OPC_EOLd OPC_EOLd OPC_EOLd	EGR1 may be involved in the suppression of oligodendrocyte differentiation (89) and its induced expression is associated with events leading to oligodendrocyte apoptosis (87).	Egr TF genes are expressed in rat brain oligodendrocytes at P7 (53).
ETS GRP: The ETS GRP of TF proteins include Ets and Ets variants (ETV), Elk, and Elf TFs.	<i>Elf1</i> <i>Elk3</i> <i>Ets1</i> <i>Ets2</i> <i>Etv1</i> <i>Etv5</i> <i>Etv6</i>	↑ ↓ ↓ ↓ ↓ ↑ ↓ ↑	OPC_EOLd P4_P10d OPC_EOLd P4_P10d OPC_EOLd OPC_EOLd EOL_MOLd OPC_EOLd EOL_MOLd OPC_EOLd EOL_MOLd		Members of the Ets group are expressed in rat brain oligodendrocytes at P7 (53).
FKH GRP	<i>Foxf2</i> <i>Foxg1</i> <i>Foxj1</i> <i>Foxj3</i> <i>Foxn2</i> <i>Foxo3a</i> <i>Foxp1</i>	↓ ↓ ↑ ↑ ↑ ↓ ↑	P4_P10d P4_P10d OPC_EOLd EOL_MOLd EOL_MOLd OPC_EOLd EOL_MOLd OPC_EOLd EOL_MOLd	FOXJ3 may be involved in OPC proliferation (90).	Forkhead TFs are expressed in rat brain oligodendrocytes at P7 (53).
FOS (LEUZIP)—TF Ap1 is composed of Fos and Jun proteins.	<i>Fos</i> <i>Fosb</i>	↓ ↑ ↓	P4_P10d OPC_EOLd EOL_MOLd OPC_EOLd	FOS previously shown to play a role in oligodendrocyte differentiation (91) and oligodendrocyte cell death (87).	Fos TFs are expressed in rat brain oligodendrocytes at P7 (53).
GLI (ZF)	<i>Gli1</i>	↓	P4_P10d	GLI1 (as part of the Sonic hedgehog pathway) promotes the migration and proliferation of optic nerve OPCs (92).	<i>Gli1</i> and <i>Gli2</i> TFs are expressed in rat brain oligodendrocytes at P7 (53).
HOX GRP; PRRX2 (HOX); PBX1 (HOX)	<i>Hoxd1</i> <i>Pbx1</i> <i>Prrx1</i>	↑ ↓ ↓ ↓	EOL_MOLd OPC_EOLd EOL_MOLd EOL_MOLd		Members of the Hox family are expressed in oligodendrocytes (53,93–95).
MEF2A (MADS); MYF (HLH); HAND1-TCFE2A (HLH)	<i>Tcf3</i>	↓ ↑ ↓	P4_P10d OPC_EOLd EOL_MOLd	A number of proteins reported in a CNS myelin proteome list (51) are known to be regulated by MEF2A and the Myf family members in muscle: DES; ACTA1; ACTC1 (96,97) and Ckb (98,99).	<i>Mef2A</i> , <i>Hand1</i> , <i>Myod1</i> , and <i>Myf6</i> TF genes are expressed in rat brain oligodendrocytes at P7 (53);
MIZF (ZF): Mbd2	<i>Mbd2</i>	↑	EOL_MOLd		<i>Mbd2</i> is expressed in rat brain oligodendrocytes at P7 (53).
NFYA (CCAATBOX): NFY binds CCAAT motifs and is a complex composed of three subunits: NFYA, NFYB, and NFYC	<i>Nfyb</i>	↓	OPC_EOLd		<i>Nfya</i> , <i>Nfyb</i> and <i>Nfyc</i> are expressed in rat brain oligodendrocytes at P7 (53).

Table 2. Continued

TFBS classes	Gene expression Evidence for associated TFs ↓ - down-regulated; ↑ - up-regulated			Literature support	
	TF	↓ ↑	Dataset	Functional support	Other gene expression support
NHLH1 (HLH)				NHLH1 is required for normal brain development (100).	<i>Nhlh1</i> is expressed in rat brain oligodendrocytes at P7 (53).
NKX/HOX GRP	<i>Nkx2-2</i> <i>Nkx2-9</i> <i>Nkx6-2</i>	↓ ↑ ↑	EOL_MOLd EOL_MOLd P4_P10d OPC_EOLd EOL_MOLd	Nkx TFs play a role in oligodendrocyte maturation, see reviews in (10,11).	<i>Nkx6-1</i> is expressed in rat brain oligodendrocytes at P7 (53).
NR GRP	<i>Nr2c2</i> <i>Nr2f1</i> <i>Nr2f2</i> <i>Ppargc1a</i> <i>Rxra</i>	↓ ↓ ↓ ↓ ↓ ↑	EOL_MOLd P4_P10d OPC_EOLd EOL_MOLd OPC_EOLd OPC_EOLd EOL_MOLd EOL_MOLd	NR2F1 is required for oligodendrocyte differentiation (101).	Nuclear receptor TFs are expressed in rat brain oligodendrocytes at P7 (53).
Pou3f1 (POU-HOX)	<i>Pou3f1</i>	↑	EOL_MOLd	POU3F1 expression plays a role in the normal assembly of myelin in the CNS (102).	<i>Pou3f1</i> is expressed in rat brain oligodendrocytes at P7 (53).
RELGRP (NF-κ B family)	<i>Nfkb1</i> <i>Nfkb2</i>	↑ ↑	EOL_MOLd OPC_EOLd	Oxidative stress-induced apoptotic death in oligodendrocytes stimulates nuclear translocation and activity of both AP-1 and NF-κ B TFs (103).	REL GRP TF genes are expressed in rat brain oligodendrocytes at P7 (53).
ROAZ (ZF): Zfp423				ZFP423 (Roaz) is a zinc finger TF that is required for proliferation and differentiation of neuronal and glial precursors in the developing brain (104).	<i>Zfp423</i> is expressed in P7 rat brain oligodendrocytes (53).
RORA2 (NR)	<i>Rora2</i> (<i>Rora</i>)	↓ ↓	OPC_EOLd EOL_MOLd		<i>Rora</i> is expressed in P7 rat brain oligodendrocytes (53).
SOX/HMG GRP	<i>Sox2</i> <i>Sox4</i> <i>Sox5</i> <i>Sox6</i> <i>Sox7</i> <i>Sox9</i> <i>Sox11</i> <i>Sox21</i>	↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	EOL_MOLd EOL_MOLd EOL_MOLd OPC_EOLd EOL_MOLd P4_P10d OPC_EOLd OPC_EOLd EOL_MOLd OPC_EOLd EOL_MOLd	Sox TFs are required for oligodendrocyte maturation, see reviews in (10,11).	Sox TFs are expressed in rat brain oligodendrocytes at P7 (53).
SP1 (ZF)				SP1 is a zinc finger protein that regulates <i>Mbp</i> gene expression (105).	<i>Sp1</i> is expressed in P7 rat brain oligodendrocytes (53).
TEAD1 (TEA)	<i>Tead2</i>	↓	P4_P10d	TEAD1 is a known regulator of muscle genes and is associated with notochord maintenance and cell proliferation and survival in mouse development (106). A homologous family member <i>Tead2</i> , in some cases, may perform functionally redundant roles with <i>Tead1</i> (106).	<i>Tead1</i> and <i>Tead2</i> TFs are expressed in rat brain oligodendrocytes at P7 (53).
YY1 (ZF): Ying Yang 1 (<i>Yy1</i>) is a zinc finger TF that can bind DNA.				YY1 is involved in oligodendrocyte progenitor differentiation and myelin gene expression (107).	<i>Yy1</i> is expressed in P7 rat brain oligodendrocytes (53).
ZEB1 (ZF)	<i>Zeb1</i>	↓ ↓	OPC_EOLd EOL_MOLd		<i>Zeb1</i> is expressed in P7 rat brain oligodendrocytes (53).

While gene knock-out and gene-centric transcription studies have highlighted the essential role of some DNA-binding TFs in the myelination process, a comprehensive inventory of collaborating TFs driving myelin-associated gene expression has remained elusive. In this investigation we combined genome-wide oligodendrocyte expression profiling, *in vivo* enhancer analysis and sequence analysis of non-coding regions to establish a TRN model of coordinate myelin-associated gene transcription. Functional characterization of new myelin gene enhancers coupled with novel profiling of the myelination expression program provided the experimental foundation for computational modeling of network components. The derived myelin gene TRN model contains 43 CRMs that link an expanded myelination-relevant TF repertoire of DNA-binding TFs. Furthermore, the TRN model, together with the underlying pro-myelination CRMs, specifies the collaborative relationships among TFs.

To functionally characterize myelin gene enhancers, we used a controlled transgenesis strategy and evaluated reporter-gene expression in oligodendrocytes co-labeled with progenitor or myelinating cell stage markers. This *in vivo* approach accesses the developmental context, provides cellular resolution and controls many of the experimental variables typically associated with analysis of *in vivo* reporter gene expression. The myelin gene enhancer set reported here augments existing resources such as: Vista Enhancer Browser (VEB) (81), which records reporter expression phenotypes in mouse embryos, and the Pleiades Promoter Project (PPP) (82), which identifies targeted transgene activity in the brains of mature mice. Additionally, non-overlapping non-coding regions associated with genes that are common to our enhancer collection have previously been investigated for activity: *Pou3f1* (VEB), *Olig1* (VEB and PPP) and *Olig2* (83). Expansion of such characterized enhancer sequence collections should contribute in a progressive manner to the understanding of gene regulatory mechanisms.

Much like oligodendrocytes in the CNS, myelin production by Schwann cells in the PNS is controlled largely by transcriptional mechanisms (84). Additionally, PNS and CNS myelin share many structural proteins and accordingly, several of the enhancer sequences characterized here direct expression in both oligodendrocytes and Schwann cells. TFs SOX10, EGR2 and POU3F1 act in a combinatorial manner to drive the myelinating transition in Schwann cells and our analyses also predict the cooperative involvement of Sox and Egr (i.e. Krox) TF family members in oligodendrocyte myelination. Recent experimental analyses of the *Mbp* (26,29,55,85) and *Plp1* (25) loci have exposed regulatory architectures composed of multiple enhancers individually directing expression to oligodendrocytes and/or Schwann cells. Future joint investigations of myelin development in the CNS and PNS should offer unique opportunities to elucidate the transcriptional mechanisms responsible for directing commonly expressed genes in these different cell types.

Several groups have performed gene expression profiling to identify genes that participate in oligodendrocyte development and function. We hypothesized that a

core set of differentially expressed genes responsible for oligodendrocyte myelination would be shared by oligodendrocytes maturing on different temporal programs in different CNS domains. Consistent with this expectation, the optic nerve P4–P10 expression comparison dataset shared 65% of its differentially expressed genes with the P16 mouse forebrain stage-specific differential expression profiles and 52% are found in common with the OPC–EOL oligodendrocyte stage transition. Notably, we identified a subset of 31 TFs in the IOLEDD set and 829 genes classified as TFs in the other differentially expressed datasets. While some of these TFs will be responsible for transcriptional regulation of myelin structural proteins, others may control gene expression in related processes such as lipid synthesis and membrane assembly.

The oligodendrocyte myelination TRN model depicts a gene regulatory system that is coordinated by multiple CRMs and includes classes of TFs not previously linked to myelin gene targets. Importantly, the regulatory architecture defined by the TRN accommodates the emerging view of oligodendrocyte combinatorial gene regulation, where specific TFs provide sustained input throughout the myelination process [e.g. Hmg protein *Sox10* (86)] while additional TFs conditionally exert enhancing and/or antagonizing transcriptional effects; for example, in response to environmental influences (87). Further system complexity is embodied in the TRN class nodes, such that the same CRM elements may be engaged by different TF class members resulting in different regulatory outcomes (88). Notably, differences in reporter gene expression programs along with loss of expression at mature developmental stages for some of the enhancers investigated in this study implicate further levels of heterogeneity in transcriptional mechanisms.

In an effort to achieve specificity in the predicted myelin gene TRN, we included only those CRM predictions unique to the positive enhancers and also identified as statistically over-represented by the CSA⁺ analyses. If the enhancer set does not include all myelin gene regulatory sequences, the approach will necessarily exclude contributing CRMs. Nonetheless, the current TRN model provides a significantly refined view of the TFs controlling oligodendrocyte maturation and we expect that progressively enhanced sensitivity will be realized as new myelin gene-associated regulatory sequences are identified and characterized.

Validation of the oligodendrocyte TRN model data using experimental data demonstrated its capacity to identify bonafide regulatory sequence and associated TFs. While a naive CRM prediction approach (i.e. predicting all TFBS combinations in non-coding sequence) would identify these regulatory sequences, the specificity and sensitivity achieved in the TRN model data validation results substantiate our approach. Significantly, the TRN includes TFs with previously recognized roles in oligodendrocyte maturation and further expands on known regulatory mechanisms by predicting CRMs with the capacity to bind factors expressed during oligodendrocyte development. Our study supplements and extends recently developed oligodendrocyte gene expression data resources (9,42,53) and establishes the transition from a gene-centric

to a network-informed systems view of the regulatory mechanisms directing oligodendrocyte myelinogenesis. Further characterization of the TF network controlling myelin gene expression should help refine our understanding of oligodendrocyte development as well as suggest novel therapeutic strategies to potentiate their regenerative capacity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the McGill University and Génome Québec Innovation Centre for their assistance and work on the mouse optic nerve microarray data. They appreciate the excellent mouse transgenesis-related and imaging-related work performed by Irene Tretjakoff, Priscila Valera and Tess Fernandez. They are grateful to Jean-Christophe Deloulme for his assistance with the immunohistochemistry and confocal imaging.

FUNDING

Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award, a Michael Smith Foundation for Health Research (MSFHR) Graduate Doctoral Scholarship award, and a Multiple Sclerosis Society of Canada Fellowship award (to D.L.F.); Commissariat à l'Énergie Atomique, France (to E.D.); Genome Quebec and Genome Canada 'GRID' program, the Multiple Sclerosis Society of Canada and the Multiple Sclerosis Scientific Research Foundation (to A.C.P.); CIHR New Investigator award, MSFHR Scholar and grants from CIHR and National Sciences and Engineering Research Council of Canada (to W.W.W.). Computer hardware resources utilized for this project were supported by the Gene Regulation Bioinformatics Laboratory funded by Canada Foundation for Innovation. Funding for open access charge: MS Society of Canada operating grant (to A.C.P.).

Conflict of interest statement. None declared.

REFERENCES

- Arroyo,E.J. and Scherer,S.S. (2000) On the molecular architecture of myelinated fibers. *Histochem. Cell Biol.*, **113**, 1–18.
- Bischoff,A. and Moor,H. (1967) Ultrastructural differences between the myelin sheaths of peripheral nerve fibres and CNS white matter. *Z. Zellforsch. Mikrosk. Anat.*, **81**, 303–310.
- Wolswijk,G., Riddle,P.N. and Noble,M. (1990) Coexistence of perinatal and adult forms of a glial progenitor cell during development of the rat optic nerve. *Development*, **109**, 691–698.
- Sloane,J.A. and Vartanian,T.K. (2007) WAVE1 and regulation of actin nucleation in myelination. *Neuroscientist*, **13**, 486–491.
- 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. et al. (2005) Region-specific myelin pathology in mice lacking the golli products of the myelin basic protein gene. *J. Neurosci.*, **25**, 7004–7013.
- Forghani,R., Garofalo,L., Foran,D.R., Farhadi,H.F., Lepage,P., Hudson,T.J., Tretjakoff,I., Valera,P. and Peterson,A. (2001) A distal upstream enhancer from the myelin basic protein gene regulates expression in myelin-forming schwann cells. *J. Neurosci.*, **21**, 3780–3787.
- Pfeiffer,S.E., Warrington,A.E. and Bansal,R. (1993) The oligodendrocyte and its many cellular processes. *Trends Cell Biol.*, **3**, 191–197.
- Foran,D.R. and Peterson,A.C. (1992) Myelin acquisition in the central nervous system of the mouse revealed by an MBP-Lac Z transgene. *J. Neurosci.*, **12**, 4890–4897.
- Dugas,J.C., Tai,Y.C., Speed,T.P., Ngai,J. and Barres,B.A. (2006) Functional genomic analysis of oligodendrocyte differentiation. *J. Neurosci.*, **26**, 10967–10983.
- Nicolay,D.J., Doucette,J.R. and Nazarali,A.J. (2007) Transcriptional control of oligodendrogenesis. *Glia*, **55**, 1287–1299.
- Wegner,M. (2008) A matter of identity: transcriptional control in oligodendrocytes. *J. Mol. Neurosci.*, **35**, 3–12.
- Davidson,E.H. (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Academic Press, Burlington, MA, San Diego, USA.
- Wasserman,W.W. and Sandelin,A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.*, **5**, 276–287.
- Sandelin,A. and Wasserman,W.W. (2004) Constrained binding site diversity within families of transcription factors enhances pattern discovery bioinformatics. *J. Mol. Biol.*, **338**, 207–215.
- Farhadi,H.F., Lepage,P., Forghani,R., Friedman,H.C., Orfali,W., Jasmin,L., Miller,W., Hudson,T.J. and Peterson,A.C. (2003) A combinatorial network of evolutionarily conserved myelin basic protein regulatory sequences confers distinct glial-specific phenotypes. *J. Neurosci.*, **23**, 10214–10223.
- Pennacchio,L.A., Ahituv,N., Moses,A.M., Prabhakar,S., Nobrega,M.A., Shoukry,M., Minovitsky,S., Dubchak,I., Holt,A., Lewis,K.D. et al. (2006) *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature*, **444**, 499–502.
- Visel,A., Akiyama,J.A., Shoukry,M., Afzal,V., Rubin,E.M. and Pennacchio,L.A. (2009) Functional autonomy of distant-acting human enhancers. *Genomics*, **93**, 509–513.
- Krivan,W. and Wasserman,W.W. (2001) A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res.*, **11**, 1559–1566.
- Wasserman,W.W. and Fickett,J.W. (1998) Identification of regulatory regions which confer muscle-specific gene expression. *J. Mol. Biol.*, **278**, 167–181.
- Sauka-Spengler,T. and Bronner-Fraser,M. (2008) A gene regulatory network orchestrates neural crest formation. *Nat. Rev. Mol. Cell Biol.*, **9**, 557–568.
- Gokhan,S., Marin-Husstege,M., Yung,S.Y., Fontanez,D., Casaccia-Bonnel,P. and Mehler,M.F. (2005) Combinatorial profiles of oligodendrocyte-selective classes of transcriptional regulators differentially modulate myelin basic protein gene expression. *J. Neurosci.*, **25**, 8311–8321.
- Li,H., Lu,Y., Smith,H.K. and Richardson,W.D. (2007) Olig1 and Sox10 interact synergistically to drive myelin basic protein transcription in oligodendrocytes. *J. Neurosci.*, **27**, 14375–14382.
- Liu,Z., Hu,X., Cai,J., Liu,B., Peng,X., Wegner,M. and Qiu,M. (2007) Induction of oligodendrocyte differentiation by Olig2 and Sox10: evidence for reciprocal interactions and dosage-dependent mechanisms. *Dev. Biol.*, **302**, 683–693.
- Sugimori,M., Nagao,M., Bertrand,N., Parras,C.M., Guillemot,F. and Nakafuku,M. (2007) Combinatorial actions of patterning and HLH transcription factors in the spatiotemporal control of neurogenesis and gliogenesis in the developing spinal cord. *Development*, **134**, 1617–1629.
- Tuason,M.C., Rastikerdar,A., Kuhlmann,T., Goujet-Zalc,C., Zalc,B., Dib,S., Friedman,H. and Peterson,A. (2008) Separate proteolipid protein/DM20 enhancers serve different lineages and stages of development. *J. Neurosci.*, **28**, 6895–6903.
- Dionne,N. (2006) Structure and function of Module 3, a conserved enhancer of the myelin basic protein gene. *Ph.D. Thesis*, McGill University.

27. Kuhn,R.M., Karolchik,D., Zweig,A.S., Wang,T., Smith,K.E., Rosenbloom,K.R., Rhead,B., Raney,B.J., Pohl,A., Pheasant,M. *et al.* (2009) The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res.*, **37**, D755–D761.
28. Rozen,S. and Skaletsky,H.J. (2000) In Krawetz,S. and Misener,S. (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.
29. Farhadi,H.F. and Peterson,A.C. (2006) The myelin basic protein gene: a prototype for combinatorial mammalian transcriptional regulation. *Adv. Neurol.*, **98**, 65–76.
30. Bronson,S.K., Plaehn,E.G., Kluckman,K.D., Hagaman,J.R., Maeda,N. and Smithies,O. (1996) Single-copy transgenic mice with chosen-site integration. *Proc. Natl Acad. Sci. USA*, **93**, 9067–9072.
31. R Development Core Team. (2008) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
32. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Detting,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
33. Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
34. Irizarry,R.A., Hobbs,B., Collin,F., Beazer-Barclay,Y.D., Antonellis,K.J., Scherf,U. and Speed,T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
35. Irizarry,R.A., Bolstad,B.M., Collin,F., Cope,L.M., Hobbs,B. and Speed,T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.
36. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
37. Fulton,D.L., Sundararajan,S., Badis,G., Hughes,T.R., Wasserman,W.W., Roach,J.C. and Sladek,R. (2009) TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol.*, **10**, R29.
38. Wright,G.W. and Simon,R.M. (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, **19**, 2448–2455.
39. Carbon,S., Ireland,A., Mungall,C.J., Shu,S., Marshall,B. and Lewis,S. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.
40. Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Blake,J.A. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
41. Twigger,S.N., Shimoyama,M., Bromberg,S., Kwitek,A.E. and Jacob,H.J. (2007) The Rat Genome Database, update 2007—easing the path from disease to data and back again. *Nucleic Acids Res.*, **35**, D658–D662.
42. Cahoy,J.D., Emery,B., Kaushal,A., Foo,L.C., Zamanian,J.L., Christopherson,K.S., Xing,Y., Lubischer,J.L., Krieg,P.A., Krupenko,S.A. *et al.* (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.*, **28**, 264–278.
43. Miller,W., Rosenbloom,K., Hardison,R.C., Hou,M., Taylor,J., Raney,B., Burhans,R., King,D.C., Baertsch,R., Blankenberg,D. *et al.* (2007) 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res.*, **17**, 1797–1808.
44. Blanco,E., Farre,D., Alba,M.M., Messeguer,X. and Guigo,R. (2006) ABS: a database of Annotated regulatory Binding Sites from orthologous promoters. *Nucleic Acids Res.*, **34**, D63–D67.
45. Portales-Casamar,E., Kirov,S., Lim,J., Lithwick,S., Swanson,M.I., Ticoll,A., Snoddy,J. and Wasserman,W.W. (2007) PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation. *Genome Biol.*, **8**, R207.
46. Kuhn,R.M., Karolchik,D., Zweig,A.S., Trumbower,H., Thomas,D.J., Thakkapallayil,A., Sugnet,C.W., Stanke,M., Smith,K.E., Siepel,A. *et al.* (2007) The UCSC genome browser database: update 2007. *Nucleic Acids Res.*, **35**, D668–D673.
47. Ho Sui,S.J., Fulton,D.L., Arenillas,D.J., Kwon,A.T. and Wasserman,W. (2007) oPOSSUM: integrated tools for analysis of regulatory motif over-representation. *Nucleic Acids Res.*, **35**, W245–W252.
48. Huang,S.S., Fulton,D.L., Arenillas,D.J., Perco,P., Ho Sui,S.J., Mortimer,J.R. and Wasserman,W.W. (2006) In Jiang,T., Yang,U., Chen,Y.P. and Wong,L. (eds), *Series on Advances in Bioinformatics and Computational Biology Volume 3—Proceedings of the 4th Asia-Pacific Bioinformatics Conference*, Vol. 3. Imperial College Press, London UK, Taipei, Taiwan, pp. 247–256.
49. Dunn,T.M., Hahn,S., Ogden,S. and Schleif,R.F. (1984) An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl Acad. Sci. USA*, **81**, 5017–5020.
50. Roth,F.P., Hughes,J.D., Estep,P.W. and Church,G.M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.*, **16**, 939–945.
51. Jahn,O., Tenzer,S. and Werner,H.B. (2009) Myelin proteomics: molecular anatomy of an insulating sheath. *Mol. Neurobiol.*, **40**, 55–72.
52. Goldovsky,L., Cases,I., Enright,A.J. and Ouzounis,C.A. (2005) BioLayout(Java): versatile network visualisation of structural and functional relationships. *Appl. Bioinf.*, **4**, 71–74.
53. Nielsen,J.A., Maric,D., Lau,P., Barker,J.L. and Hudson,L.D. (2006) Identification of a novel oligodendrocyte cell adhesion protein using gene expression profiling. *J. Neurosci.*, **26**, 9881–9891.
54. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Rudnev,D., Evangelista,C., Kim,I.F., Soboleva,A., Tomashevsky,M., Marshall,K.A. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D885–D890.
55. Denarier,E., Forghani,R., Farhadi,H.F., Dib,S., Dionne,N., Friedman,H.C., Lepage,P., Hudson,T.J., Drouin,R. and Peterson,A. (2005) Functional organization of a Schwann cell enhancer. *J. Neurosci.*, **25**, 11210–11217.
56. Scherer,S.S., Deschenes,S.M., Xu,Y.T., Grinspan,J.B., Fischbeck,K.H. and Paul,D.L. (1995) Connexin32 is a myelin-related protein in the PNS and CNS. *J. Neurosci.*, **15**, 8281–8294.
57. Bondurand,N., Girard,M., Pingault,V., Lemort,N., Dubourg,O. and Goossens,M. (2001) Human Connexin 32, a gap junction protein altered in the X-linked form of Charcot-Marie-Tooth disease, is directly regulated by the transcription factor SOX10. *Hum. Mol. Genet.*, **10**, 2783–2795.
58. Houlden,H., Girard,M., Cockerell,C., Ingram,D., Wood,N.W., Goossens,M., Walker,R.W. and Reilly,M.M. (2004) Connexin 32 promoter P2 mutations: a mechanism of peripheral nerve dysfunction. *Ann. Neurol.*, **56**, 730–734.
59. Schaeren-Wiemers,N., Bonnet,A., Erb,M., Erne,B., Bartsch,U., Kern,F., Mantei,N., Sherman,D. and Suter,U. (2004) The raft-associated protein MAL is required for maintenance of proper axon–glia interactions in the central nervous system. *J. Cell Biol.*, **166**, 731–742.
60. Brockschneider,D., Sabanay,H., Riethmacher,D. and Peles,E. (2006) Ermin, a myelinating oligodendrocyte-specific protein that regulates cell morphology. *J. Neurosci.*, **26**, 757–762.
61. Trapp,B.D., Bernier,L., Andrews,S.B. and Colman,D.R. (1988) Cellular and subcellular distribution of 2',3'-cyclic nucleotide 3'-phosphodiesterase and its mRNA in the rat central nervous system. *J. Neurochem.*, **51**, 859–868.
62. Collarini,E.J., Pringle,N., Mudhar,H., Stevens,G., Kuhn,R., Monuki,E.S., Lemke,G. and Richardson,W.D. (1991) Growth factors and transcription factors in oligodendrocyte development. *J. Cell Sci.*, **15**, 117–123.
63. Bogler,O., Entwistle,A., Kuhn,R., Monuki,E., Lemke,G. and Noble,M. (1993) Single cell analysis of the expression of a nuclear protein, SCIP, by fluorescent immunohistochemistry visualized with confocal microscopy. *Histochem. J.*, **25**, 746–761.
64. Collarini,E.J., Kuhn,R., Marshall,C.J., Monuki,E.S., Lemke,G. and Richardson,W.D. (1992) Down-regulation of the POU transcription factor SCIP is an early event in oligodendrocyte differentiation *in vitro*. *Development*, **116**, 193–200.

65. Bermingham, J.R. Jr, Scherer, S.S., O'Connell, S., Arroyo, E., Kalla, K.A., Powell, F.L. and Rosenfeld, M.G. (1996) Tst-1/Oct-6/SCIP regulates a unique step in peripheral myelination and is required for normal respiration. *Genes Dev.*, **10**, 1751–1762.
66. Jaegle, M., Mandemakers, W., Broos, L., Zwart, R., Karis, A., Visser, P., Grosveld, F. and Meijer, D. (1996) The POU factor Oct-6 and Schwann cell differentiation. *Science*, **273**, 507–510.
67. Mandemakers, W., Zwart, R., Jaegle, M., Walbeehm, E., Visser, P., Grosveld, F. and Meijer, D. (2000) A distal Schwann cell-specific enhancer mediates axonal regulation of the Oct-6 transcription factor during peripheral nerve development and regeneration. *EMBO J.*, **19**, 2992–3003.
68. Xin, M., Yue, T., Ma, Z., Wu, F.F., Gow, A. and Lu, Q.R. (2005) Myelinogenesis and axonal recognition by oligodendrocytes in brain are uncoupled in Olig1-null mice. *J. Neurosci.*, **25**, 1354–1365.
69. Marco, A., Konikoff, C., Karr, T.L. and Kumar, S. (2009) Relationship between gene co-expression and sharing of transcription factor binding sites in *Drosophila melanogaster*. *Bioinformatics*.
70. Allocco, D.J., Kohane, I.S. and Butte, A.J. (2004) Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics*, **5**, 18.
71. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C. et al. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.
72. Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.L., Tsung, E.F., Cai, Z. et al. (2004) Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science*, **306**, 2255–2257.
73. Choi, M.Y., Romer, A.I., Hu, M., Lepourcelet, M., Mechoor, A., Yesilaltay, A., Krieger, M., Gray, P.A. and Shivdasani, R.A. (2006) A dynamic expression survey identifies transcription factors relevant in mouse digestive tract development. *Development*, **133**, 4119–4129.
74. King, D.C., Taylor, J., Zhang, Y., Cheng, Y., Lawson, H.A., Martin, J., Analysis, E.g.f.T.R.a.M.S., Chiaromonte, F., Miller, W. and Hardison, R.C. (2007) Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data. *Genome Res.*, **17**, 775–786.
75. Odom, D., Dowell, R., Jacobsen, E., Gordon, W., Danford, T., Macisaac, K., Rolfe, P., Conboy, C., Gifford, D. and Fraenkel, E. (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat. Genet.*, **39**, 730–732.
76. McGaughey, D.M., Vinton, R.M., Huynh, J., Al-Saif, A., Beer, M.A. and McCallion, A.S. (2008) Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at *phox2b*. *Genome Res.*, **18**, 252–260.
77. Blanchette, M. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.*, **14**, 708–715.
78. Blanco, E., Messeguer, X., Smith, T.F. and Guigo, R. (2006) Transcription factor map alignment of promoter regions. *PLoS Comput. Biol.*, **2**, e49.
79. Blanchette, M., Bataille, A.R., Chen, X., Poitras, C., Laganier, J., Lefebvre, C., Deblois, G., Giguere, V., Ferretti, V., Bergeron, D. et al. (2006) Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res.*, **16**, 656–668.
80. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. and Lenhard, B. (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.*, **32**, D91–D94.
81. Visel, A., Minovitsky, S., Dubchak, I. and Pennacchio, L.A. (2007) VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.*, **35**, D88–D92.
82. Portales-Casamar, E., Swanson, D.J., Liu, L., de Leeuw, C.N., Banks, K.G., Ho Sui, S.J., Fulton, D.L., Ali, J., Amirabbasi, M., Arenillas, D.J. et al. (2010) A regulatory toolbox of MiniPromoters to drive selective expression in the brain. *Proc. Natl Acad. Sci. USA*, **107**, 16589–16594.
83. Chen, C.T., Gottlieb, D.I. and Cohen, B.A. (2008) Ultraconserved elements in the Olig2 promoter. *PLoS One*, **3**, e3946.
84. Svaren, J. and Meijer, D. (2008) The molecular machinery of myelin gene transcription in Schwann cells. *Glia*, **56**, 1541–1551.
85. Dib, S., Denarier, E., Dionne, N., Beaudoin, M., Friedman, H.H. and Peterson, A.C. (2011) Regulatory modules function in a non-autonomous manner to control transcription of the *mbp* gene. *Nucleic Acids Res.*, **39**, 2548–2558.
86. Stolt, C.C., Rehberg, S., Ader, M., Lommes, P., Riethmacher, D., Schachner, M., Bartsch, U. and Wegner, M. (2002) Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10. *Genes Dev.*, **16**, 165–170.
87. FitzGerald, U.F., Gilbey, T., Brodie, S. and Barnett, S.C. (2003) Transcription factor expression and cellular redox in immature oligodendrocyte cell death: effect of Bcl-2. *Mol. Cell. Neurosci.*, **22**, 516–529.
88. Stolt, C.C., Schlierf, A., Lommes, P., Hillgartner, S., Werner, T., Kosian, T., Sock, E., Kessar, N., Richardson, W.D., Lefebvre, V. et al. (2006) SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. *Dev. Cell*, **11**, 697–709.
89. Sock, E., Leger, H., Kuhlbrodt, K., Schreiber, J., Enderich, J., Richter-Landsberg, C. and Wegner, M. (1997) Expression of Krox proteins during differentiation of the O-2A progenitor cell line CG-4. *J. Neurochem.*, **68**, 1911–1919.
90. Dugas, J.C., Cuellar, T.L., Scholze, A., Ason, B., Ibrahim, A., Emery, B., Zamanian, J.L., Foo, L.C., McManus, M.T. and Barres, B.A. (2010) Dicer1 and miR-219 Are required for normal oligodendrocyte differentiation and myelination. *Neuron*, **65**, 597–611.
91. Barnett, S.C., Rosario, M., Doyle, A., Kilbey, A., Lovatt, A. and Gillespie, D.A. (1995) Differential regulation of AP-1 and novel TRE-specific DNA-binding complexes during differentiation of oligodendrocyte-type-2-astrocyte (O-2A) progenitor cells. *Development*, **121**, 3969–3977.
92. Merchan, P., Bribian, A., Sanchez-Camacho, C., Lezameta, M., Bovolenta, P. and de Castro, F. (2007) Sonic hedgehog promotes the migration and proliferation of optic nerve oligodendrocyte precursors. *Mol. Cell. Neurosci.*, **36**, 355–368.
93. Booth, J., Nicolay, D.J., Doucette, J.R. and Nazarali, A.J. (2007) Hoxd1 is expressed by oligodendroglial cells and binds to a region of the human myelin oligodendrocyte glycoprotein promoter in vitro. *Cell. Mol. Neurobiol.*, **27**, 641–650.
94. Nicolay, D.J., Doucette, J.R. and Nazarali, A.J. (2004) Hoxb4 in oligodendrogenesis. *Cell. Mol. Neurobiol.*, **24**, 357–366.
95. Nicolay, D.J., Doucette, J.R. and Nazarali, A.J. (2004) Early stages of oligodendrocyte development in the embryonic murine spinal cord proceed normally in the absence of Hoxa2. *Glia*, **48**, 14–26.
96. Di Padova, M., Caretti, G., Zhao, P., Hoffman, E.P. and Sartorelli, V. (2007) MyoD acetylation influences temporal patterns of skeletal muscle gene expression. *J. Biol. Chem.*, **282**, 37650–37659.
97. Hinitz, Y. and Hughes, S.M. (2007) Mef2s are required for thick filament formation in nascent muscle fibres. *Development*, **134**, 2511–2519.
98. Hobson, G.M., Molloy, G.R. and Benfield, P.A. (1990) Identification of cis-acting regulatory elements in the promoter region of the rat brain creatine kinase gene. *Mol. Cell Biol.*, **10**, 6533–6543.
99. Hobson, G.M., Mitchell, M.T., Molloy, G.R., Pearson, M.L. and Benfield, P.A. (1988) Identification of a novel TA-rich DNA binding protein that recognizes a TATA sequence within the brain creatine kinase promoter. *Nucleic Acids Res.*, **16**, 8925–8944.
100. Li, C.M., Yan, R.T. and Wang, S.Z. (1999) Misexpression of a bHLH gene, cNSCL1, results in abnormal brain development. *Dev. Dyn.*, **215**, 238–247.
101. Yamaguchi, H., Zhou, C., Lin, S.C., Durand, B., Tsai, S.Y. and Tsai, M.J. (2004) The nuclear orphan receptor COUP-TFI is important for differentiation of oligodendrocytes. *Dev. Biol.*, **266**, 238–251.
102. Jensen, N.A., Pedersen, K.M., Celis, J.E. and West, M.J. (1998) Neurological disturbances, premature lethality, and central myelination deficiency in transgenic mice overexpressing the homeo domain transcription factor Oct-6. *J. Clin. Invest.*, **101**, 1292–1299.

103. Vollgraf,U., Wegner,M. and Richter-Landsberg,C. (1999) Activation of AP-1 and nuclear factor-kappaB transcription factors is involved in hydrogen peroxide-induced apoptotic cell death of oligodendrocytes. *J. Neurochem.*, **73**, 2501–2509.
104. Alcaraz,W.A., Gold,D.A., Raponi,E., Gent,P.M., Concepcion,D. and Hamilton,B.A. (2006) Zfp423 controls proliferation and differentiation of neural precursors in cerebellar vermis formation. *Proc. Natl Acad. Sci. USA*, **103**, 19424–19429.
105. Wei,Q., Miskimins,W.K. and Miskimins,R. (2003) The Sp1 family of transcription factors is involved in p27(Kip1)-mediated activation of myelin basic protein gene expression. *Mol. Cell Biol.*, **23**, 4035–4045.
106. Sawada,A., Kiyonari,H., Ukita,K., Nishioka,N., Imuta,Y. and Sasaki,H. (2008) Redundant roles of Tead1 and Tead2 in notochord development and the regulation of cell proliferation and survival. *Mol. Cell Biol.*, **28**, 3177–3189.
107. He,Y., Dupree,J., Wang,J., Sandoval,J., Li,J., Liu,H., Shi,Y., Nave,K.A. and Casaccia-Bonnel,P. (2007) The transcription factor Yin Yang 1 is essential for oligodendrocyte progenitor differentiation. *Neuron*, **55**, 217–230.