Supplementary Information

Towards resolving the transcription factor network controlling myelin gene expression

Table of Contents:

Tables

Table S1. Myelin gene enhancer genomic coordinates.

	Mouse Assembly 36 Feb 2006 (mm8)		
Gene Enhancers	Chr	From Coordinate	To Coordinate
Cldn11	3	31357206	31357836
Cnp	11	100391366	100392107
Ermn	2	57868209	57868795
Gjb1	X	97586055	97586691
Mal	2	127318715	127319393
Olig1	16	91189705	91190805
Olig2	16	91097897	91098790
Pou3f1	4	124169971	124170620

Table S2. Myelin gene-associated enhancer sequences.

Gene Enhancers	Construct Sequences
Cldn11	TGTCACCACTCCTCTGAAGCTACCCCAGGCATGATTTGTGTCCTCCACGGCCCTCCAGGAATGGCTTCGACTTCCCACATTTTTTCATTTCTACAGAGGGGCATGAAGCAGTGACTAGTCCCCCTGTCTCCACTCACCCTTCCAGACACGGCCACGGTGCCTGAGCTCCGGGTATACACGAGCTAAAGAAACCGAGATTCCTCAAGAGCTGCTATTCAAGCCACGAGCAGCATGTGCCAACATTCCTTCCACAATCCCGCTTCAGTCCCCATGAAGCTGCACATCTGGCTGTCACACACTCTTTTGTGGGGCCATCTCAAAGGCTGTCCTGATGCTGCCTCTCCCCAGGGCCGGTCCCACAGTGGCCTCCCTGTCCTCCTAGCATGTGATGTCATGAGCATGAATGTCCTGCCAAACATTTAAGATGCGACAGCTCATTGTGTACATTGTATAGAGAAACACTGCTGGGCAGATCAGTTCTAATTAAAACAAGAAAAGAGGCAGGCCAGGACAGAGCCGGACCTTGTTCTATTGTTACCCCAAGCCACCGAGGCAGCACAAGTGTGTAACGTGCACTCTCCAAGCTCACACCCTTCTCCTTTTGTTCCAGTCTTGTCCCCAGCATCTCCAA
Cnp	GCTGGGTTGTAAGGTAGAGGGAATCTTTCTGAAGCTACCTAACCTCTTAGTCTCTGGCTCCAAGGATCACTCTCTGCTCCCCCTCAGAACACCTTGGTGACAGGGCACGCAAGGGCAGCAAGTGCCTGTGGCTCCCTGTCAGCCCTAACTTATATATTCTTGGCACCTCTTAAGGATGTCCTGTTTGGAGGCCTGGGGCTCTATGTATGGGCAAAGCCACAGCTGCCTTGGCTCTCTTGAGTGTGCAGAGGAGGGTGAGGGCAGGGGAGGGTGACCCAAAGACCGCCTCCCTCCACCCAACAATGGCACCAACAGATAATACTCTCTTGTCTGGGTATAGGGGGACAATGGTCCCACAGAGGGCTGGTGGGGGGCAGGGCAGGAGGGCACTGGACAAAGGGCATATTCACAACCACCTCCTCCTCTGCTCTCTCAGTGCGCCCCCTCCCCCAATCTGTAGGATTAGGGCACAGGGCACCTGTATTCAGAGATCTGTGCCCCCTATTCTGCTGTGTGTTAGCAGCACCTCTCTGACAAAAACAAGACATAGCGGTTTCAGTTGTCATCTACAACGGCTTCCGCCCCTGAGAATTTAGCTGGCTTCTGGCTTCTCTATAGCCTTGTGCAGGAACCCACTCCCAGCGCGGGCTCCACAACAGCCCTGCACTCTCTGGTGCTACCCAGACCCATTTGCCAGCTGGTGTCGCCAAGGAGCATCGACTTTCACTTCCCTGCGGCGATA
Ermn	TAATGGGTAGGGTGGGTGTCTGTAGACCTGGGGGGACTGGAAAATCAGCTTCCTGAGAGCACACTATGAAATCAGTTGCCAAAGCCGTTTACAATCAAGTTGTCATGTTTACAGGTCCTTGAAAAGCTGGTTTTAGTCTTTAATAGATGAGAAATGATGCCATTGTCTCCTACAAAATAGCTGAGGTTTTACTCGAACCAAATCTGTTCGGAAATGTTAAGCTGGTTACACAGAACTAATGTGCCTCAAGGTGGTCATTCTCTCTGGATTAAAGCCTGGGAACAATTGTGGCTCCTGTTACAAGGAAAATTACAATGGGCCATTATGGAGAGGGACAAAAATCTCTGTTCCCAGAGGGGGACTGACAAGCGGCAAGTCCCCCTTTGCATGCTAACACAAAAGCCTGATTGCTTCAGATTGCTTGTTTATGCAAATTGAAGGCAACACTAATTCTGAGACTGGAGCTGCGAGGTGGTGGGGATGTTGCTGATTGGCTGGGTTTTGCTCCCGGGCTGGTTACCAAACTGACAAACCCTATCATATTCATTCTGAGCCCTGAAAACCTTTCGGAGGAATCTTGCTATTGC
Gjb1	TCCTCTTGAGTCCCTTTCTTCACCAGGCCTATCTTGTTCTGCTATTAGCCTGTTCTTCGTGCACCACCACCACCACCCCGCCCCAGCTCCTTTACACCTGACTTATGGCAAACCCAGATAGCCCGGTTACCTCGGGGAGCGCCTATCCTTGAGGCCACCCAGACAGCTCCCCTATGGTCTCATGTCTGTGTAGGGGGAGTGGTCCCTGTGGTCGCGCCTGCGTCCCAATTCATCGCCTCCCGGTGCTAGAGAGTCATATGCTGCTTTATACCCAGTGTCTGTACAATGGGACCTTTGTGGGGCGCTGAACAACACCCTTTAGAATTCAGATCAAACGCCCTGACTTCCGCCCACCCTACACACAGAGCACTTGTGTCCCAGCGCCGGACACCCAGCCTGCACCCACCTGTTCTGTGCCCGCCCGGTCCATCGCCGCACCCGCCCCATGGCCCCACCAGAATAATGGCTCTGGGGGAGGGAGATAGAACACAGACAGGTGCAAGTAAGCCAGGACCCCAGGGAATCAGGGTCCTCTGGACTAAGGCCTTCCCCTATATTCTACCAGAAACGGGGTCAGCAGCAGGTTTTCTAGGGGTAGGTCCCTTTCCAGGTATGGATGAGCCAGAAGAGAGGAAAC
Mal	CTGCCTTTGTTTCTCTTTCTGGTGATCTTCCCCCACGAAGGGAGGGTATATGAAAGAGGCTTCTTCATGTATAGCTCTTTTCCCCTGGAAGCCAACTTCTAACACCTCATTCTGCACACAGTAGGTCCAATGAGACCATTCTTTCATAGACCACTTCATAGGAGCTGAGCCTTCCTTATAAAGGAGTCTGTGTGGGCACCCCTCAGGAATCAGCATGTCTTTGCACAGCAGTCCCAACGGTGCATTGAAGGCCTCCCTACACATGGGCCTTAAAGAGACAGTAAAGGGTTCGCACTGCCAGCCGTGGGCTGGAGCAAGCATGCCTTTGGGGCTATACAAGGGCCTGTTGTTCCCAGAATCAATTGGGGCCACTGTGGGGTCACAATGAAGGCCTTTCAAAGAGACAAAGCGCCTCTGGCCTGGCTAAGTGTCCCTGCACACGTGGGCATCTGCCCAGAACATTTAAAATAACACGATTTGAAAAAACAGATTTAAAGTGTACAATCTAGTGGCCTTTGTTCTGTGTCTACCATGTGCAGTGGCCTTTGTTCTGTGTCCACCACATGCAGTGATGGTGGATCAGAGCTGGGTGTGGGGAGGAGGACACTAGAGCTCCCAGCTGGGACCTCTCTTCTCTGTCTTAGGTAGGAAACTGGGTATCTACACAGTTCCAGGCTCAC

Table S2. Myelin gene-associated enhancer sequences (continued).

Gene Enhancers	Construct Sequences
Olig1	CTGCTGAAACACTCCGCTCTTCCTGGAGACTGCAGGAGGCTCGGATGGGGTGGTTGGGGCAGAGCACTGGGCATTAAGCTCCATCCTGGCTTCACATAAAGGAAGAAAACTCAAAACGCAGGGGCGGGGGGTGGGGGGGTGCAACTAAATTTAATTGGCTAGAAAGCAGGATTCCTGGGGTCTCCGCTTTCCCAGCTTCCAGAGACGGCTTTAAGAAGGGATTCTTCTCCGCCTCCCAGCACCAAAAAGAAAGGGTTGAGCTGGGGTGGTAATCCAGGTTTGGTCTGGCCAAGACAAATGACATACAGAACATGCAGTCTCTTTGCACACGGCATGTGCTTTCATGGAAGTGGAAGGGGAGAGAAGAAGGTGGGTAAGGGGGTGTAGAAAGGGGGCACTAAAAACAACTTGCTGTTTGCATATGGTCGGGCGAAATCTGTTTTGTCCCCGTCACTGTGCAGAGTGTCAGTAGTGGGCACCGTGTCGTGACAAGTTTCTTTCTTTCTTTCTTTTCTTTTTTTCTTTTTCTTTTTTTTTTTTAAGTTAAGCTAAGCTCACCCCGAGCCTCGGCCCCCCGAGGTCCTGCAGCTTTAATTGTGTTCAGTATGATCTGGCGAACAATATGTCTTGATAAATGTGAATGGCGATCCCGCGCGGCTCCCCTTGGCATTGCCGCAAAGCCAGGCCTCGCTTTAGTGAACCCGTTCAGGAATGTGGGTTAATTCGGCCGACCCTTTTTACCCCCACAGCTCTGTTTTTGTGAGGTTGGCAGAAACTGACACGTTTTCTTTTGCTTAATTAAGTTGTTCAGTAGAAGGCTGGTCAAACAGTTGCATCTACGTGGGGAGTCTGGGGATGCGCAGAAAGCCGTGGGCGTTTCCAACATAAAAAGGGGGATTCTGTTATTGTTTGGGTGTAGACGGGTCATTTTTTCAAAGTTATTGCTGCCTGAGTGACCACAACCTTGCGCAGAGCCCTGAAAGGACATTAAAGCTGCTTAAACACAACCACCTTGGGCTCCAAAGTGTGTTTCTAAAAAGGAGAAAGAGATGCAGGGCGCAAGAGAAAGAACTAGGGCGATAGGAAGCAGTGGCGTTT
Olig2	GGATTGGGAAGGCATCTTGCCTCCAGCCTGGCATTTTATGAAGAATTAAAAATAACGAAGCCGGCAGAGATACGGAGGGAGGCTAATTTGGAAACCTGGAATGATCGCTTTTAATTTGTTGACAATGTGGTGTTTGCGGGGGGAGGGATTGAGGGAGCCAAGGCACAGGATGGCCGTGCTTGTCTGGGCCTTGTATTAATGGAGTAGCAGCTGAATTACGTGTTCAGAGCGAACACATCAAAGACCATCATCTGAGATCTCATTTCTTCTATTCCACGGTGTCAAAGGTGGCTGCAAATATATCCAGCACTTTCTGGGTGGCATTGTTTAGGATGCTGGCCCCAGCAGATGGCCCCAAAATGTACCTCAGCCTGATGGAGCAGGGAGGGACCACATGGGATGGCATGTCTCCTCATTAGCAAACAGAAAAGAAGGGGGAAAAAATAGAAAGCAGGATGAGAAAGGAGAGAGATGAAAAGAATACACAAAGACATTTACATATTTACATTGTCTCCCTCGCATTTAAAATTCCCCCTTAGCCGAAGTGTCAGTGGGCTTCTTCCTTCTGTTTGTGTGTTTACTTGTGCTAAATAGATTATGGTAATCAGGCACATAAATGCACGCAACTGTTATTATGGCATTTTAATCACGGACTTTTTCACTGGAAGCCTTCATTAGGATTCCCATAATCTTACTTGTTAGTGTAACAAACAGTAATAAATGCCAAAGAGCTTGGATTAAGGCAGGGGCCTATCAGGCTTTAAAAGAGGAATATGGAAATGTTGTTGAATCTCTGCATTACTTATTGAAATGTCACATAAACAGTTCGAGATTCTAAATAGCTTACTCATTATCCAGAGAGGGACGCGGGCAACCTGAGCCCAGTAATGATGT
Pou3f1	CTCGAGCTGCTCAGCCCCCCTCCCCTGACACAAACAATCCTCAGTTACCTCCCCCTCCTGCTCCCCAGAATCTGGGCACAGCTGGAGCCTGCTATGCCCTAGCCACCCCATGAATCACCGCTCTATGGTCCACGGGGGAGTGGTCCAGGGAGCATCCTACGCTTGCTCGGGAGGAGTGAGGGCCAGAACTGCAGCCCTCAAGCAGGCAGTGTCCACAGAAACAATGGGGGCCTGTGGCTAACAGGCGGAATGCAGCTATTGTCCTGCCCTGGCCCCCAACCCCAAGGCCCCAGGTCCCCAGGCCAGGCGGCCTGGCGTGAATCAGTGCGTCAGACTCTCGTGTACCAGGGCTGGGCACACATGACCTGCTGCTTACCTTCTCTGGGTAAGCAGGAGGGTAGGCTGATGGGCCAGACCCCCACCTCTACAGCCAACCTCACAAAGGATCCTCCTGCACAGAATGAGAAGCGAGATGGGACCCACGGGAGTAAAGGCAACCTCCAACCTACCCATCCCTAGGATGCTTGAGACCAGCAGCTCTGAGACCCAAGCTATACATCACAGTGGGGAGGAGACCACATTAGAATAATGCAGGATTAGAGTGGGGTTGCTATAGCGACGTATTAGGGCAATACATCTAGGGAGCCCCA

Table S3. PCR primers used in enhancer sequence amplification (restriction sites are noted in bold).

Gene Enhancers	Primer 1 / Primer 2
Cldn11	ATGC**GGCGCGCC**TGTCACCACTCCTCTGAAGCTACC
	ATGC**CTCGAG**TTGGAGATGCTGGGGACAAGAC
Cnp	ATGC**CTCGAG**GCTGGGTTGTAAGGTAGAGGGAATC
	ATGC**GGCGCGCC**TATCGCCGCAGGGAAGTGAAAGTC
Ermn	ATGC**GGCGCGCC**TAATGGGTAGGGTGGGTGTCTG
	ATGC**CTCGAG**GCAATAGCAAGATTCCTCCGAAAG
Gjb1	ATGC**GGCGCGCC**TCCTCTTGAGTCCCTTTCTTCACC
	ATGC**CTCGAG**GTTTCCTCTCTTCTGGCTCATCC
Mal	ATGC**GGCGCGCC**CTGCCTTTGTTTCTCTTTCTGGTG
	ATGC**CTCGAG**GTGAGCCTGGAACTGTGTAGATACC
Olig1	ATGC**GGCGCGCC**CTGCTGAAACACTCCGCTCTTC
	ATGC**CTCGAG**AAACGCCACTGCTTCCTATCGC
Olig2	ATGC**GGCGCGCC**GGATTGGGAAGGCATCTTGC
	ATGC**CTCGAG**ACATCATTACTGGGCTCAGGT
Pou3f1	ATGC**CTCGAG**TTCTTTGACAATGGGGCTTCTCT
	ATGC**GGCGCGCC**TGGGGCTCCCTAGATGTATTGC

Table S4. List of skeletal muscle reference collection genes.

Ensembl Gene ID	HGNC Symbol	Ensembl Description
ENSG00000143632	ACTA1	Actin, alpha skeletal muscle (Alpha-actin-1) [Source:UniProtKB/Swiss-Prot;Acc:P68133]
ENSG00000159251	ACTC1	Actin, alpha, cardiac muscle 1 [Source:HGNC Symbol;Acc:143]
ENSG00000149925	ALDOA	Fructose-bisphosphate aldolase A (EC 4.1.2.13)(Muscle-type aldolase)(Lung cancer antigen NY-LU-1) [Source:UniProtKB/Swiss-Prot;Acc:P04075]
ENSG00000138435	CHRNA1	Acetylcholine receptor subunit alpha Precursor [Source:UniProtKB/Swiss-Prot;Acc:P02708]
ENSG00000170175	CHRNB1	Acetylcholine receptor subunit beta Precursor [Source:UniProtKB/Swiss-Prot;Acc:P11230]
ENSG00000135902	CHRND	Acetylcholine receptor subunit delta Precursor [Source:UniProtKB/Swiss-Prot;Acc:Q07001]
ENSG00000108556	CHRNE	Acetylcholine receptor subunit epsilon Precursor [Source:UniProtKB/Swiss-Prot;Acc:Q04844]
ENSG00000196811	CHRNG	Acetylcholine receptor subunit gamma Precursor [Source:UniProtKB/Swiss-Prot;Acc:P07510]
ENSG00000104879	CKM	Creatine kinase M-type (EC 2.7.3.2)(Creatine kinase M chain)(M-CK) [Source:UniProtKB/Swiss-Prot;Acc:P06732]
ENSG00000175084	DES	Desmin [Source:UniProtKB/Swiss-Prot;Acc:P17661]
ENSG00000198947	DMD	Dystrophin [Source:UniProtKB/Swiss-Prot;Acc:P11532]
ENSG00000198125	MB	Myoglobin [Source:UniProtKB/Swiss-Prot;Acc:P02144]
ENSG00000081189	MEF2C	Myocyte-specific enhancer factor 2C [Source:UniProtKB/Swiss-Prot;Acc:Q06413]
ENSG00000111046	MYF6	Myogenic factor 6 (Myf-6) [Source:UniProtKB/Swiss-Prot;Acc:P23409]
ENSG00000109063	MYH3	Myosin-3 (Myosin heavy chain 3)(Myosin heavy chain, fast skeletal muscle, embryonic)(Muscle embryonic myosin heavy chain)(SMHCE) [Source:UniProtKB/Swiss-Prot;Acc:P11055]
ENSG00000141048	MYH4	Myosin-4 (Myosin heavy chain 4)(Myosin heavy chain 2b)(MyHC-2b)(Myosin heavy chain IIb)(MyHC-IIb)(Myosin heavy chain, skeletal muscle, fetal) [Source:UniProtKB/Swiss-Prot;Acc:Q9Y623]
ENSG00000197616	MYH6	Myosin-6 (Myosin heavy chain 6)(Myosin heavy chain, cardiac muscle alpha isoform)(MyHC-alpha) [Source:UniProtKB/Swiss-Prot;Acc:P13533]
ENSG00000092054	MYH7	Myosin-7 (Myosin heavy chain 7)(Myosin heavy chain, cardiac muscle beta isoform)(MyHC-beta)(Myosin heavy chain slow isoform)(MyHC-slow) [Source:UniProtKB/Swiss-Prot;Acc:P12883]
ENSG00000168530	MYL1	Myosin light chain 1, skeletal muscle isoform (MLC1F)(A1 catalytic)(Alkali myosin light chain 1) [Source:UniProtKB/Swiss-Prot;Acc:P05976]
ENSG00000198336	MYL4	Myosin light chain 4 (Myosin light chain 1, embryonic muscle/atrial isoform)(Myosin light chain alkali, GT-1 isoform) [Source:UniProtKB/Swiss-Prot;Acc:P12829]
ENSG00000129152	MYOD1	Myoblast determination protein 1 (Myogenic factor 3)(Myf-3) [Source:UniProtKB/Swiss-Prot;Acc:P15172]
ENSG00000122180	MYOG	Myogenin (Myogenic factor 4)(Myf-4) [Source:UniProtKB/Swiss-Prot;Acc:P15173]

Table S4. List of skeletal muscle reference collection genes (continued).

Ensembl Gene ID	HGNC Symbol	Ensembl Description
ENSG00000007314	SCN4A	Sodium channel protein type 4 subunit alpha (Sodium channel protein type IV subunit alpha)(Voltage-gated sodium channel subunit alpha Nav1.4)(Sodium channel protein skeletal muscle subunit alpha)(SkM1) [Source:UniProtKB/Swiss-Prot;Acc:P35499]
ENSG00000181856	SLC2A4	Solute carrier family 2, facilitated glucose transporter member 4 (Glucose transporter type 4, insulin-responsive)(GLUT-4) [Source:UniProtKB/Swiss-Prot;Acc:P14672]
ENSG00000114854	TNNC1	Troponin C, slow skeletal and cardiac muscles (TN-C) [Source:UniProtKB/Swiss-Prot;Acc:P63316]
ENSG00000159173	TNNI1	Troponin I, slow skeletal muscle (Troponin I, slow-twitch isoform) [Source:UniProtKB/Swiss-Prot;Acc:P19237]

Table S5. Skeletal muscle enhancer genomic coordinates.

		Mouse Assembly 36 Feb 2006 (mm8)		
Gene Enhancers	Human Ensembl Gene Id	Chr	From Coordinate	To Coordinate
DMD_MOUSE	ENSG00000198947	X	79208550	79208871
MYG_MOUSE	ENSG00000198125	15	76849839	76850250
MEF2C_MOUSE	ENSG00000081189	13	83981341	83981501
MYH3_MOUSE	ENSG00000125414	11	67053712	7053977
ACHA_MOUSE	ENSG00000138435	2	73381070	73381369
ACHB_MOUSE	ENSG00000170175	11	69612003	69612304
ACHG_MOUSE	ENSG00000196811	1	89036593	89036963
ACHD_MOUSE	ENSG00000135902	1	89021553	89021854
ACHE_MOUSE	ENSG00000108556	11	70435308	70435605
KCRM_MOUSE_1	ENSG00000104879	7	18568267	18568567
KCRM_MOUSE_2	ENSG00000104879	7	18568838	18569622
MYF6_MOUSE	ENSG00000111046	10	106898903	106899202
MYOD_MOUSE	ENSG00000129152	7	46239064	46239781
MYL4_MOUSE	ENSG00000198336	11	104392837	104393509
TNCC_MOUSE	ENSG00000114854	14	30038200	30038502
MYH6_MOUSE	ENSG00000197616	14	53920850	53921149

Table S6. Skeletal muscle reference collection with 25 random genes.

Ensembl Gene ID	HGNC Symbol	Ensembl Description
ENSG00000143632	ACTA1	Actin, alpha skeletal muscle (Alpha-actin-1) [Source:UniProtKB/Swiss-Prot;Acc:P68133]
ENSG00000159251	ACTC1	Actin, alpha, cardiac muscle 1 [Source:HGNC Symbol;Acc:143]
ENSG00000149925	ALDOA	Fructose-bisphosphate aldolase A (EC 4.1.2.13)(Muscle-type aldolase)(Lung cancer antigen NY-LU-1) [Source:UniProtKB/Swiss-Prot;Acc:P04075]
ENSG00000050820	BCAR1	Breast cancer anti-estrogen resistance protein 1 (CRK-associated substrate)(p130cas)(Cas scaffolding protein family member 1) [Source:UniProtKB/Swiss-Prot;Acc:P56945]
ENSG00000108688	CCL7	C-C motif chemokine 7 Precursor (Small-inducible cytokine A7)(Monocyte chemoattractant protein 3)(Monocyte chemotactic protein 3)(MCP-3)(NC28) [Source:UniProtKB/Swiss-Prot;Acc:P80098]
ENSG00000138435	CHRNA1	Acetylcholine receptor subunit alpha Precursor [Source:UniProtKB/Swiss-Prot;Acc:P02708]
ENSG00000170175	CHRNB1	Acetylcholine receptor subunit beta Precursor [Source:UniProtKB/Swiss-Prot;Acc:P11230]
ENSG00000135902	CHRND	Acetylcholine receptor subunit delta Precursor [Source:UniProtKB/Swiss-Prot;Acc:Q07001]
ENSG00000108556	CHRNE	Acetylcholine receptor subunit epsilon Precursor [Source:UniProtKB/Swiss-Prot;Acc:Q04844]
ENSG00000196811	CHRNG	Acetylcholine receptor subunit gamma Precursor [Source:UniProtKB/Swiss-Prot;Acc:P07510]
ENSG00000104879	CKM	Creatine kinase M-type (EC 2.7.3.2)(Creatine kinase M chain)(M-CK) [Source:UniProtKB/Swiss-Prot;Acc:P06732]
ENSG00000166394	CYB5R2	NADH-cytochrome b5 reductase 2 (b5R.2)(EC 1.6.2.2) [Source:UniProtKB/Swiss-Prot;Acc:Q6BCY4]
ENSG00000175084	DES	Desmin [Source:UniProtKB/Swiss-Prot;Acc:P17661]
ENSG00000198947	DMD	Dystrophin [Source:UniProtKB/Swiss-Prot;Acc:P11532]
ENSG00000163435	ELF3	ETS-related transcription factor Elf-3 (E74-like factor 3)(Epithelium-specific Ets transcription factor 1)(ESE-1)(Epithelium-restricted Ets protein ESX)(Epithelial-restricted with serine box) [Source:UniProtKB/Swiss-Prot;Acc:P78545]
ENSG00000187672	ERC2	ERC protein 2 [Source:UniProtKB/Swiss-Prot;Acc:O15083]
ENSG00000138829	FBN2	Fibrillin-2 Precursor [Source:UniProtKB/Swiss-Prot;Acc:P35556]
ENSG00000204007	GLT6D1	Glycosyltransferase 6 domain-containing protein 1 (EC 2.4.1.-)(Galactosyltransferase family 6 domain-containing 1) [Source:UniProtKB/Swiss-Prot;Acc:Q7Z4J2]
ENSG00000100577	GSTZ1	Maleylacetoacetate isomerase (MAAI)(EC 5.2.1.2)(Glutathione S-transferase zeta 1)(EC 2.5.1.18)(GSTZ1-1) [Source:UniProtKB/Swiss-Prot;Acc:O43708]
ENSG00000129636	ITFG1	T-cell immunomodulatory protein Precursor (Protein TIP)(Integrin-alpha FG-GAP repeat-containing protein 1) [Source:UniProtKB/Swiss-Prot;Acc:Q8TB96]
ENSG00000111615	KRR1	KRR1 small subunit processome component homolog (HIV-1 Rev-binding protein 2)(Rev-interacting protein 1)(Rip-1) [Source:UniProtKB/Swiss-Prot;Acc:Q13601]
ENSG00000181541	MAB21L2	Protein mab-21-like 2 [Source:UniProtKB/Swiss-Prot;Acc:Q9Y586]
ENSG00000198125	MB	Myoglobin [Source:UniProtKB/Swiss-Prot;Acc:P02144]
ENSG00000081189	MEF2C	Myocyte-specific enhancer factor 2C [Source:UniProtKB/Swiss-Prot;Acc:Q06413]
ENSG00000139505	MTMR6	Myotubularin-related protein 6 (EC 3.1.3.-) [Source:UniProtKB/Swiss-Prot;Acc:Q9Y217]

Table S6. Skeletal muscle reference collection with 25 random genes (continued).

Ensembl Gene ID	HGNC Symbol	Ensembl Description
ENSG00000111046	MYF6	Myogenic factor 6 (Myf-6) [Source:UniProtKB/Swiss-Prot;Acc:P23409]
ENSG00000109063	MYH3	Myosin-3 (Myosin heavy chain 3)(Myosin heavy chain, fast skeletal muscle, embryonic)(Muscle embryonic myosin heavy chain)(SMHCE) [Source:UniProtKB/Swiss-Prot;Acc:P11055]
ENSG00000141048	MYH4	Myosin-4 (Myosin heavy chain 4)(Myosin heavy chain 2b)(MyHC-2b)(Myosin heavy chain IIb)(MyHC-IIb)(Myosin heavy chain, skeletal muscle, fetal) [Source:UniProtKB/Swiss-Prot;Acc:Q9Y623]
ENSG00000197616	MYH6	Myosin-6 (Myosin heavy chain 6)(Myosin heavy chain, cardiac muscle alpha isoform)(MyHC-alpha) [Source:UniProtKB/Swiss-Prot;Acc:P13533]
ENSG00000092054	MYH7	Myosin-7 (Myosin heavy chain 7)(Myosin heavy chain, cardiac muscle beta isoform)(MyHC-beta)(Myosin heavy chain slow isoform)(MyHC-slow) [Source:UniProtKB/Swiss-Prot;Acc:P12883]
ENSG00000168530	MYL1	Myosin light chain 1, skeletal muscle isoform (MLC1F)(A1 catalytic)(Alkali myosin light chain 1) [Source:UniProtKB/Swiss-Prot;Acc:P05976]
ENSG00000198336	MYL4	Myosin light chain 4 (Myosin light chain 1, embryonic muscle/atrial isoform)(Myosin light chain alkali, GT-1 isoform) [Source:UniProtKB/Swiss-Prot;Acc:P12829]
ENSG00000129152	MYOD1	Myoblast determination protein 1 (Myogenic factor 3)(Myf-3) [Source:UniProtKB/Swiss-Prot;Acc:P15172]
ENSG00000122180	MYOG	Myogenin (Myogenic factor 4)(Myf-4) [Source:UniProtKB/Swiss-Prot;Acc:P15173]
ENSG00000188162	OTOG	Otogelin Precursor [Source:UniProtKB/Swiss-Prot;Acc:Q6ZRI0]
ENSG00000169241	RAG1AP1	RAG1-activating protein 1 (Stromal cell protein) [Source:UniProtKB/Swiss-Prot;Acc:Q9BRV3]
ENSG00000054967	RELT	Tumor necrosis factor receptor superfamily member 19L Precursor (Receptor expressed in lymphoid tissues) [Source:UniProtKB/Swiss-Prot;Acc:Q969Z4]
ENSG00000128482	RNF112	RING finger protein 112 (Zinc finger protein 179)(Brain finger protein) [Source:UniProtKB/Swiss-Prot;Acc:Q9ULX5]
ENSG00000007314	SCN4A	Sodium channel protein type 4 subunit alpha (Sodium channel protein type IV subunit alpha)(Voltage-gated sodium channel subunit alpha Nav1.4)(Sodium channel protein skeletal muscle subunit alpha)(SkM1) [Source:UniProtKB/Swiss-Prot;Acc:P35499]
ENSG00000071537	SEL1L	Protein sel-1 homolog 1 Precursor (Suppressor of lin-12-like protein 1)(Sel-1L) [Source:UniProtKB/Swiss-Prot;Acc:Q9UBV2]
ENSG00000099381	SETD1A	Histone-lysine N-methyltransferase SETD1A (EC 2.1.1.43)(SET domain-containing protein 1A)(hSET1A)(Set1/Ash2 histone methyltransferase complex subunit SET1)(Lysine N-methyltransferase 2F) [Source:UniProtKB/Swiss-Prot;Acc:O15047]
ENSG00000181856	SLC2A4	Solute carrier family 2, facilitated glucose transporter member 4 (Glucose transporter type 4, insulin-responsive)(GLUT-4) [Source:UniProtKB/Swiss-Prot;Acc:P14672]
ENSG00000184402	SS18L1	SS18-like protein 1 (SYT homolog 1) [Source:UniProtKB/Swiss-Prot;Acc:O75177]
ENSG00000099365	STX1B	Syntaxin-1B (Syntaxin-1B1)(Syntaxin-1B2) [Source:UniProtKB/Swiss-Prot;Acc:P61266]
ENSG00000114854	TNNC1	Troponin C, slow skeletal and cardiac muscles (TN-C) [Source:UniProtKB/Swiss-Prot;Acc:P63316]
ENSG00000159173	TNNI1	Troponin I, slow skeletal muscle (Troponin I, slow-twitch isoform) [Source:UniProtKB/Swiss-Prot;Acc:P19237]
ENSG00000198258	UBL5	Ubiquitin-like protein 5 [Source:UniProtKB/Swiss-Prot;Acc:Q9BZL1]

Table S6. Skeletal muscle reference collection with 25 random genes (continued).

Ensembl Gene ID	HGNC Symbol	Ensembl Description
ENSG00000124486	USP9X	Probable ubiquitin carboxyl-terminal hydrolase FAF-X (EC 3.1.2.15)(Ubiquitin thioesterase FAF-X)(Ubiquitin-specific-processing protease FAF-X)(Deubiquitinating enzyme FAF-X)(Fat facets protein-related, X-linked)(Ubiquitin-specific protease 9, X chromosome) [Source:UniProtKB/Swiss-Prot;Acc:Q93008]
ENSG00000157796	WDR19	WD repeat-containing protein 19 [Source:UniProtKB/Swiss-Prot;Acc:Q8NEZ3]
ENSG00000166435	XRRA1	X-ray radiation resistance-associated protein 1 [Source:UniProtKB/Swiss-Prot;Acc:Q6P2D8]
ENSG00000161914	ZNF653	Zinc finger protein 653 (67 kDa zinc finger protein)(Zinc finger protein Zip67) [Source:UniProtKB/Swiss-Prot;Acc:Q96CK0]

Table S7. Parameters used in the validation of the promoter CRM analyses method.

Jaspar TFBS Phylum	Max. TFBS Binding Site Distance (bp)	TFBS Score Threshold (%)	Upstream Sequence From TSS (bp)	Downstream Sequence from TSS (bp)
Vertebrate	225	80	100	5000
Vertebrate	225	80	1000	5000
Vertebrate	225	80	2000	2000
Vertebrate	225	80	2000	3000
Vertebrate	225	80	4000	3000
Vertebrate	225	80	5000	3000
Vertebrate	225	80	6000	3000
Vertebrate	225	80	7000	3000
Vertebrate	225	80	10000	3000

Table S8. Mouse myelin gene-associated enhancer genomic coordinates.

		Mouse Assembly 36 Feb 2006 (mm8)		
Gene Enhancers	Human Ensembl Gene Id	Chr	From Coordinate	To Coordinate
Cldn11	ENSG00000013297	3	31357206	31357836
Cnp	ENSG00000173786	11	100391366	100392107
Ermin	ENSG00000136541	2	57868209	57868795
Gjb1/Cx32	ENSG00000169562	X	97586055	97586691
Mal	ENSG00000172005	2	127318715	127319393
Mbp_M3	ENSG00000197971	18	82683969	82684438
Mbp_M4	ENSG00000197971	18	82680074	82680286
Olig1	ENSG00000184221	16	91189705	91190805
Olig2	ENSG00000205927	16	91097897	91098790
Plp_Pk211	ENSG00000123560	X	132185516	132186725
Plp_WmN1	ENSG00000123560	X	132172873	132174044
Plp_WmN2	ENSG00000123560	X	132174737	132175770
Pou3f1	ENSG00000185668	4	124169971	124170620

Table S9. Positive and negative groupings for oligodendrocyte enhancers.

Enhancer	Enhancer Group
Cldn11	Positive
Mal	Positive
Mbp_M3	Positive
Plp_WmN1	Positive
Cnp	Negative
Ermn	Negative

Table S10. Gene Ontology molecular function term enrichment analysis of optic nerve expression data.

GO molecular function enrichment analysis for genes differentially expressed in murine optic nerve between postnatal day 4 and postnatal day 10 (using p-value cut-off of 0.01 and database filters: MGI – Mouse Genome Informatics and RGD – Rat Genome Database). In some cases, the same mouse and rat gene symbols are indicated.

GO Term	P-value	Sample frequency (MGI + RGD)	Background frequency (MGI + RGD)	Genes
GO:0005488 binding	2.47E-07	309/504 (61.3%)	10628/22524 (47.2%)	*Cdca5 Fkbp10 Emilin1 Kcnj10 S100a1 Ngfr Tpd52 Tgfbr2 S100a1 Elk3 Dock9 Tcf7 Ptp4a3 Idh1 Ndn Nek6 Ywhaq Fos Chfr Bmp6 Eif4ebp2 Cxcl12 Col5a1 Ier3 Ssbp3 Myoc Arap3 Gjb2 Sept8 Jam3 Il1rap Crym Dab2ip Ctse Ywhah Cdt1 Tgm2 Sp5 Gnas Erbb3 Pdlim2 Aplp1 Cacnb4 Ascl1 Klf2 Wdr6 Tcf3 Gas6 Fzd2 Klhl2 Cdc37l1 Rras2 Taldo1 Nid2 Txnip Arpc1a Ywhah Axin1 Tgfbi Kcnj10 Elk3 Arpc1b Col5a1 Sema4d Cebpb Bmp7 Gstm1 Eng Ddx39 Egfl7 S100a4 Sdc1 Eya2 Rrm2 Plp1 Birc2 Rims2 Eps15 Myoc Spred2 Rrm2 Ncapd2 Ckap4 Dab2ip Wnt7b Adamts4 Tcf7 Eps15 Sirt2 Dgkz Mcm6 Emilin1 Mfge8 Cdc37l1 Vldlr Kctd12 Rims2 Ets2 Adipor1 Igfbp2 Plekhb1 Ywhaq Sema4d Mast2 Mknk2 Bgn Ntsr2 Strn Loxl1 Aplp1 Cdca5 Tpd52 Mknk2 Lyve1 Cebpb Gsn F2r Ccnb2 Fgf13 S100a4 Med16 Fli1 Csdc2 Fos Sytl2 Fgfr1 Acta1 Prkcz Nkx6-2 Wnt7b Fmnl3 Ramp2 Cldn11 Fzd2 Wasf2 Igfbp5 Nfic Plekhb1 Bicc1 Id4 Trp53 Itgb5 Nid2 Ehd4 Sept2 Smad1 Sox7 Actn1 Aatk Acat2 Ets2 Dlk1 Ctse Emid2 Cacnb4 Tprkb Klf2 Wasf2 Tgfbr2 Tcfe2a Scap Car2 Otud7b Ets1 Prkcz Lasp1 Acss2 Erbb3 Il1rap Cldn5 Vldlr Tgm2 Gusb Ssbp3 Mcm7 Antxr1 Med9 Rab34 Ascl1 Rasip1 Fgf13 Gli1 Rcn3 Ppic Txnip Dgkz Fcgr2b Scap Cxcr4 Bgn Cyp1b1 Ngfr Tead2 Dlk1 Sept2 Pdgfrb Rap1a Mmp14 Reep5 Gli1 Fkbp10 Mast2 Ski Lsm2 Smarcd3 Uhrf1 Rap1a Cpox Smarcd3 Igfbp2 Piga Dhcr24 Rab34 Mtap7 Ndn Nfic Igfbp5 Tead2 Eng Grb14 Gtf2ird1 Chfr Akt1 Efhd2 Loxl1 Ntsr2 Fcgr2b Akt1 Ets1 Tcf3 Pdk4 Mcm2 Plekha1 Glrb Rras2 Crym Reep5 Creb3l1 Ptx3 Ctsb Fermt3 Itgb5 Surf4 Klf4 Dact1 Birc2 Plp1 Slc2a1 Mfge8 Cdt1 Slc2a1 Wif1 Gnas Lrrc59 Adipor1 Fdft1 Spred2 Id4 Gjb2 Klf4 Arf3 Lasp1 Eif4ebp2 Ddc Gtf2ird1 Racgap1 Smad1 Ptx3 Lyve1 Bmp6 Mcm2 Lrp10 Efhd2 Sh3gl3 Pdgfrb Ehd4 Fzr1 Lrp10 Fnta Foxf2 Mest Dock9 Nkx6-2 Mcm6 Sept9 Gamt Cldn11 Mdk Igfbp4 Timm17a Grb14 Mcm7 Pctk3 Sox7 Pxn Cxcl12 Igfbp4 Glrb Stard4 Tprkb B4galt1 B4galt1 Sh3gl3 Pxn S100a3 Dact1 Axin1 Actn1 Bmp7*

Table S10. Gene Ontology molecular function term enrichment analysis of optic nerve expression data (continued).

GO Term	P-value	Sample frequency (MGI + RGD)	Background frequency (MGI + RGD)	Genes
GO:0019911 structural constituent of myelin sheath	3.56E-07	6/504 (1.2%)	6/22524 (0.0%)	*Mobp Mbp Mobp Plp1 Mal Tspan2*
GO:0019838 growth factor binding	2.14E-05	15/504 (3.0%)	101/22524 (0.4%)	*Ngfr Tgfbr2 Il1rap Erbb3 Col5a1 Eng Igfbp2 Igfbp5 Il1rap Ngfr Igfbp2 Igfbp5 Pdgfrb Igfbp4 Igfbp4*
GO:0005515 protein binding	1.67E-04	264/504 (52.4%)	9167/22524 (40.7%)	*Cdca5 Fkbp10 Emilin1 Kcnj10 S100a1 Ngfr Tpd52 Tgfbr2 S100a1 Dock9 Tcf7 Ptp4a3 Ndn Nek6 Ywhaq Chfr Bmp6 Eif4ebp2 Cxcl12 Col5a1 Ier3 Myoc Arap3 Gjb2 Sept8 Jam3 Il1rap Dab2ip Ctse Ywhah Cdt1 Tgm2 Gnas Erbb3 Pdlim2 Aplp1 Cacnb4 Klf2 Wdr6 Tcf3 Gas6 Fzd2 Klhl2 Cdc37l1 Rras2 Taldo1 Nid2 Txnip Arpc1a Ywhah Axin1 Kcnj10 Elk3 Arpc1b Col5a1 Sema4d Cebpb Bmp7 Gstm1 Eng Ddx39 S100a4 Sdc1 Rrm2 Plp1 Birc2 Rims2 Eps15 Myoc Spred2 Rrm2 Ncapd2 Ckap4 Dab2ip Wnt7b Adamts4 Eps15 Sirt2 Dgkz Mcm6 Emilin1 Mfge8 Cdc37l1 Vldlr Kctd12 Rims2 Ets2 Adipor1 Igfbp2 Plekhb1 Ywhaq Sema4d Mast2 Mknk2 Strn Loxl1 Aplp1 Cdca5 Tpd52 Mknk2 Cebpb Gsn F2r Ccnb2 Fgf13 S100a4 Med16 Fli1 Csdc2 Fos Sytl2 Fgfr1 Acta1 Prkcz Wnt7b Fmnl3 Ramp2 Cldn11 Fzd2 Wasf2 Igfbp5 Plekhb1 Id4 Trp53 Itgb5 Nid2 Ehd4 Sept2 Smad1 Actn1 Aatk Acat2 Dlk1 Ctse Emid2 Cacnb4 Tprkb Klf2 Wasf2 Tgfbr2 Tcfe2a Scap Car2 Otud7b Ets1 Prkcz Lasp1 Acss2 Erbb3 Il1rap Cldn5 Vldlr Mcm7 Antxr1 Med9 Ascl1 Rasip1 Fgf13 Gli1 Rcn3 Txnip Dgkz Fcgr2b Scap Cxcr4 Cyp1b1 Ngfr Dlk1 Sept2 Pdgfrb Rap1a Mmp14 Reep5 Gli1 Fkbp10 Mast2 Ski Lsm2 Smarcd3 Uhrf1 Rap1a Cpox Smarcd3 Igfbp2 Piga Dhcr24 Mtap7 Ndn Nfic Igfbp5 Eng Grb14 Chfr Akt1 Efhd2 Loxl1 Fcgr2b Akt1 Ets1 Tcf3 Mcm2 Glrb Rras2 Crym Reep5 Ctsb Fermt3 Itgb5 Surf4 Dact1 Birc2 Plp1 Slc2a1 Mfge8 Cdt1 Slc2a1 Wif1 Gnas Lrrc59 Adipor1 Fdft1 Spred2 Id4 Gjb2 Lasp1 Eif4ebp2 Ddc Gtf2ird1 Racgap1 Smad1 Bmp6 Mcm2 Lrp10 Efhd2 Sh3gl3 Pdgfrb Ehd4 Fzr1 Lrp10 Fnta Mest Dock9 Mcm6 Sept9 Gamt Cldn11 Igfbp4 Timm17a Grb14 Pctk3 Pxn Cxcl12 Igfbp4 Glrb Tprkb B4galt1 B4galt1 Sh3gl3 Pxn S100a3 Dact1 Axin1 Actn1 Bmp7*
GO:0004450 isocitrate dehydrogenase (NADP+) activity	7.25E-04	4/504 (0.8%)	4/22524 (0.0%)	*Idh1 Idh2 Idh2 Idh1*

Table S10. Gene Ontology molecular function term enrichment analysis of optic nerve expression data (continued).

GO Term	P-value	Sample frequency (MGI + RGD)	Background frequency (MGI + RGD)	Genes
GO:0005520 insulin-like growth factor binding	5.26E-03	6/504 (1.2%)	18/22524 (0.1%)	*Igfbp2 Igfbp5 Igfbp2 Igfbp5 Igfbp4 Igfbp4*
GO:0003700 transcription factor activity	7.61E-03	33/504 (6.5%)	604/22524 (2.7%)	*Elk3 Tcf7 Fos Ascl1 Tcf3 Elk3 Cebpb Tcf7 Cebpb Fos Nkx6-2 Nfic Trp53 Smad1 Sox7 Klf2 Tcfe2a Ets1 Ascl1 Gli1 Gli1 Nfic Gtf2ird1 Ets1 Tcf3 Creb3l1 Klf4 Klf4 Gtf2ird1 Smad1 Foxf2 Nkx6-2 Sox7*

Table S11. Gene Ontology cellular component term enrichment analysis of optic nerve expression data.

GO cellular component enrichment analysis for genes differentially expressed in murine optic nerve between postnatal day 4 and postnatal day 10 (using p-value cut-off of 0.01 and database filters: MGI – Mouse Genome Informatics and RGD – Rat Genome Database). In some cases, the same mouse and rat gene symbols are indicated.

GO Term	P-value	Sample frequency (MGI + RGD)	Background frequency (MGI + RGD)	Genes
GO:0043209 myelin sheath	4.22E-05	8/504 (1.6%)	22/22524 (0.1%)	*Mbp Plp1 Mag Mbp Plp1 Pllp Tubb4 Mag*
GO:0044421 extracellular region part	3.00E-04	45/504 (8.9%)	853/22524 (3.8%)	*Emilin1 Efemp2 Bmp6 Cxcl12 Col5a1 Myoc Erbb3 Nid2 Tgfbi Col5a1 Bmp7 Egfl7 Grn Myoc Wnt7b Adamts4 Emilin1 Mfge8 Vldlr Igfbp2 Bgn Loxl1 Gsn Tgfbi Cnp Nid2 Ltbp3 Emid2 Car2 Vldlr Tgm2 Dkk3 Bgn Col18a1 Igfbp2 Igfbp5 Loxl1 Mfge8 Grn Col18a1 Egfl7 Mdk Igfbp4 Igfbp4 Bmp7*
GO:0005576 extracellular region	8.18E-03	48/504 (9.5%)	1056/22524 (4.7%)	*Emilin1 Efemp2 Bmp6 Cxcl12 Col5a1 Myoc Erbb3 Nid2 Tgfbi Col5a1 Bmp7 Gstm1 Egfl7 Grn Myoc Wnt7b Adamts4 Emilin1 Mfge8 Vldlr Igfbp2 Bgn Loxl1 Gsn Tgfbi Cnp Nid2 Ltbp3 Ltbp3 Emid2 Car2 Vldlr Tgm2 Dkk3 Bgn Col18a1 Igfbp2 Igfbp5 Loxl1 Ctsb Mfge8 Grn Col18a1 Egfl7 Mdk Igfbp4 Igfbp4 Bmp7*

Table S12. Validation of promoter CRM analysis method using the skeletal muscle reference collection and a full ortholog background.

Top ranked CRM predictions from the promoter analysis method validation using the full skeletal muscle reference collection (25 genes), a full ortholog background dataset, and the following analysis parameters: search regions: 2000 base pairs (bp) upstream and downstream of the transcription start site; inter-binding site distance constraint of 225 bp; and all vertebrate TF binding profiles with instances that satisfy an 80% threshold. The algorithm first identifies statistical over-representation of combinations of TFBS class-representatives in a set of co-expressed genes, as compared with the background set. The top six over-represented TFBS class-representative combinations are shown (A). For each enriched TFBS class-representative CRM, the TFBS class-representatives are expanded and each possible TFBS CRM is evaluated for over-representation. The top twelve over-represented CRM predictions are shown (B). Known skeletal muscle CRMs are highlighted in yellow.

Analyses of skeletal muscle-specific genes recovered known cis-regulatory modules composed of MEF2A, SP1, SRF, and TEAD1 TFBS in the top twelve predicted pairs. Another top-ranked TFBS pair prediction was MEF2A/HAND1-TCFE2A(TCF3). HAND1 is a basic helix-loop-helix (bHLH) transcription factor family that, amongst other roles, is required for vascular smooth muscle recruitment (1) and regulation of cardiomyocyte development (2). HAND1 also inhibits MYOD-dependent skeletal muscle cell differentiation and muscle-specific myosin heavy chain protein expression (3). Interestingly, a recent study demonstrated that MEF2A/HAND1 interactions result in synergistic activation of MEF2A-dependent promoters in cardiac muscle (4). As such, this skeletal muscle CRM prediction may warrant further investigation.

Abbreviations used: FG=Foreground; BG=Background

Table S12 A. TFBS class pair over-representation (top six results).

TF Name	Class	TF Name	Class	FG Hits	FG Non Hits	BG Hits	BG Non Hits	Score
MEF2A	MADS	SP1	ZN-FINGER, C2H2	13	12	1214	13907	1.2562e-08
MEF2A	MADS	Hand1-Tcfe2a	bHLH	14	11	1800	13321	1.4654e-07
MEF2A	MADS	ZEB1	ZN-FINGER, C2H2	18	7	3631	11490	5.7498e-07
MEF2A	MADS	Roaz	ZN-FINGER, C2H2	6	19	306	14815	9.2682e-06
MEF2A	MADS	SRF	MADS	4	21	83	15038	1.1715e-05
MEF2A	MADS	TEAD1	TEA	6	19	347	14774	1.8698e-05

Table S12 B. Top twelve over-represented TFBS pairs.

TF Name	Class	TF Name	Class	FG Hits	FG Non Hits	BG Hits	BG Non Hits	Score
MEF2A	MADS	SP1	ZN-FINGER, C2H2	13	37	1214	13907	0.00012414
MEF2A	MADS	TAL1-TCF3	bHLH	9	41	602	14519	0.00014999
MEF2A	MADS	SRF	MADS	4	46	83	15038	0.00018999
MEF2A	MADS	Hand1-Tcfe2a	bHLH	15	35	1800	13321	0.00052526
MEF2A	MADS	Roaz	ZN-FINGER, C2H2	6	44	306	14815	0.00053505
MEF2A	MADS	TEAD1	TEA	6	44	347	14774	0.0010191
MEF2A	MADS	YY1	ZN-FINGER, C2H2	17	33	2294	12827	0.0010286
RORA_1	NUCLEAR RECEPTOR	Hand1-Tcfe2a	bHLH	19	31	2751	12370	0.0013039
Nkx2-5	HOMEO	ESR1	NUCLEAR RECEPTOR	4	46	147	14974	0.001522
YY1	ZN-FINGER, C2H2	NKX2-5_V2	HOMEO	35	15	7109	8012	0.0015308
SP1	ZN-FINGER, C2H2	TEAD1	TEA	11	39	1215	13906	0.0018149
MEF2A	MADS	RORA_1	NUCLEAR RECEPTOR	9	41	876	14245	0.0021661

Table S13. Validation of promoter CRM analysis method using the skeletal muscle reference collection and random background samples.

Top ranked CRM predictions from the promoter analysis method validation using the full skeletal muscle reference collection (25 genes), two randomly sampled background datasets of 5000 ortholog pairs, and the following analysis parameters: search regions: 2000 base pairs (bp) upstream and downstream of the transcription start site; inter-binding site distance constraint of 225 bp; and all vertebrate TF binding profiles with instances that satisfy an 80% threshold. The algorithm first identifies statistical over-representation of combinations of TFBS class-representatives in a set of co-expressed genes, as compared with the background set. The top six over-represented TFBS class-representative combinations are shown (A). For each enriched TFBS class-representative CRM, the TFBS class-representatives are expanded and each possible TFBS CRM is evaluated for over-representation. The top twelve over-represented CRM predictions are shown (B). Known skeletal muscle CRMs are highlighted in yellow.

The initial ranked TFBS pair results for both analyses methods using a full background (see Table S12) versus a random background (see Table S13 below) are consistent. We computed Z ratios for the proportion of TFBS background hits in a full background sample versus the randomized background sample for each matching TFBS pair prediction and found that these differences were not significant at a p-value threshold cut-off of 0.05 (data not shown). Notably, an additional known skeletal muscle CRM: MYF-SP1 was also identified with the random background approach.

Abbreviations used: FG=Foreground; BG=Background

Table S13 A. TFBS class pair over-representation (top six results).

TF Name	Class	TF Name	Class	FG Hits	FG Non Hits	BG Hits	BG Non Hits	Score
MEF2A	MADS	SP1	ZN-FINGER, C2H2	13	12	380	4620	7.3789e-09
MEF2A	MADS	Hand1-Tcfe2a	bHLH	14	11	600	4400	1.7605e-07
MEF2A	MADS	ZEB1	ZN-FINGER, C2H2	18	7	1212	3788	7.0572e-07
MEF2A	MADS	SRF	MADS	4	21	22	4978	6.6204e-06
MEF2A	MADS	Roaz	ZN-FINGER, C2H2	6	19	95	4905	7.3819e-06
MEF2A	MADS	TEAD1	TEA	6	19	115	4885	2.0963e-05

Table S13 B. Top twelve over-represented TFBS pairs.

TF Name	Class	TF Name	Class	FG Hits	FG Non Hits	BG Hits	BG Non Hits	Score
MEF2A	MADS	SP1	ZN-FINGER, C2H2	13	12	412	4580	1.9574e-08
MEF2A	MADS	Hand1-Tcfe2a	bHLH	14	11	587	4405	1.3671e-07
MEF2A	MADS	ZEB1	ZN-FINGER, C2H2	18	7	1180	3812	4.7407e-07
Myf	bHLH	SP1	ZN-FINGER, C2H2	18	7	1215	3777	7.5269e-07
SP1	ZN-FINGER, C2H2	Myb	TRP-CLUSTER	21	4	1755	3237	7.7509e-07
Myf	bHLH	ZEB1	ZN-FINGER, C2H2	23	2	2190	2802	1.0915e-06
MEF2A	MADS	Myb	TRP-CLUSTER	12	13	518	4474	2.4437e-06
MEF2A	MADS	Pdx1	HOMEO	16	9	1008	3984	2.5554e-06
MEF2A	MADS	Roaz	ZN-FINGER, C2H2	6	19	92	4900	6.2484e-06
MEF2A	MADS	SRF	MADS	4	21	29	4963	1.7811e-05
MEF2A	MADS	TEAD1	TEA	6	19	112	4880	1.8315e-05
MEF2A	MADS	Nobox	HOMEO	13	12	764	4228	2.2662e-05

Table S14. Validation of promoter CRM analysis method using the skeletal muscle reference collection with 25 randomly selected genes and random background samples.

Top ranked CRM predictions from the promoter analysis method validation using a dataset of 50 genes (25 skeletal muscle collection genes and 25 additional randomly selected genes), two randomly sampled background datasets of 5000 ortholog pairs, and the following analysis parameters: search regions: 2000 base pairs (bp) upstream and downstream of the transcription start site; inter-binding site distance constraint of 225 bp; and all vertebrate TF binding profiles. The algorithm first identifies statistical over-representation of combinations of TFBS class-representatives in a set of co-expressed genes, as compared with the background set. The top six over-represented TFBS class-representative combinations are shown (A). For each enriched TFBS class-representative CRM, the TFBS class-representatives are expanded and each possible TFBS CRM is evaluated for over-representation. The top twelve over-represented CRM predictions are shown (B). Known skeletal muscle CRMs are highlighted in yellow.

While the TFBS pair results for the noise-added dataset (see Table S14 below) is comparable with that of the analyses performed on the original reference collection (see Table S13), one CRM: MYF-SP1 was not identified. Interestingly, an additional known skeletal muscle CRM: SP1-TEAD1, was raised to rank 11 from rank 14. It is likely that predicted TFBS combinations associated with the random genes elevated the ranking of this known and over-represented CRM.

Table S14 A. TFBS class pair over-representation (top six results).

TF Name	Class	TF Name	Class	FG Hits	FG Non Hits	BG Hits	BG Non Hits	Score
MEF2A	MADS	SP1	ZN-FINGER, C2H2	13	37	392	4608	0.00010643
MEF2A	MADS	SRF	MADS	4	46	30	4970	0.00031695
MEF2A	MADS	Hand1-Tcfe2a	bHLH	15	35	576	4424	0.00039347
MEF2A	MADS	Roaz	ZN-FINGER, C2H2	6	44	97	4903	0.00047758
RORA_1	NUCLEAR RECEPTOR	Hand1-Tcfe2a	bHLH	19	31	878	4122	0.00059197
MEF2A	MADS	TEAD1	TEA	6	44	104	4896	0.00067878

Table S14 B. Top twelve over-represented TFBS pairs.

TF Name (ID)	Class	TF Name (ID)	Class	FG Hits	FG Non Hits	BG Hits	BG Non Hits	Score
MEF2A	MADS	SP1	ZN-FINGER, C2H2	13	37	390	4602	0.0001028
MEF2A	MADS	SRF	MADS	4	46	25	4967	0.00016937
MEF2A	MADS	TAL1-TCF3	bHLH	9	41	207	4785	0.00022234
MEF2A	MADS	Roaz	ZN-FINGER, C2H2	6	44	90	4902	0.00032911
MEF2A	MADS	YY1	ZN-FINGER, C2H2	17	33	723	4269	0.00047675
MEF2A	MADS	Hand1-Tcfe2a	bHLH	15	35	593	4399	0.00054489
MEF2A	MADS	TEAD1	TEA	6	44	110	4882	0.00090635
YY1	ZN-FINGER, C2H2	NKX2-5_V2	HOMEO	35	15	2323	2669	0.00096387
RORA_1	NUCLEAR RECEPTOR	Hand1-Tcfe2a	bHLH	19	31	900	4092	0.001248
NR2F1	NUCLEAR RECEPTOR	MEF2A	MADS	5	45	81	4911	0.0014885
SP1	ZN-FINGER, C2H2	TEAD1	TEA	11	39	390	4602	0.0015315
MEF2A	MADS	RORA_1	NUCLEAR RECEPTOR	9	41	276	4716	0.0016714

Table S15. Vertebrate TF profile clustering and class label assignment.

Cluster #	Jaspar TFBS	Jaspar TFBS ID	TFBS Class Label
1	Arnt	MA0004	HLH GRP
1	MAX	MA0058	HLH GRP
1	MYC-MAX	MA0059	HLH GRP
1	Mycn	MA0104	HLH GRP
1	USF1	MA0093	HLH GRP
2	Arnt-Ahr	MA0006	ARNT-AHR (HLH)
2	Pax6	MA0069	PAX6 (HOX)
3	Ar	MA0007	AR (NR)
3	NR3C1	MA0113	NR3C1 (NR)
4	T	MA0009	BRACHYURY-T (TDOMAIN)
4	ZEB1	MA0103	ZEB1 (ZF)
5	Pax5	MA0014	PAX5 (HOX)
6	HNF4A	MA0114	NR GRP
6	NR1H2-RXRA	MA0115	NR GRP
6	NR2F1	MA0017	NR GRP
6	PPARG-RXRA	MA0065	NR GRP
7	CREB1	MA0018	CREB1 (LEUZIP)
7	RORA_1	MA0071	RORA1 (NR)
7	RORA_2	MA0072	RORA2 (NR)
8	Cebpa	MA0102	CEBPA (LEUZIP)
8	Ddit3-Cebpa	MA0019	DDIT3-CEBPA (LEUZIP)
9	E2F1	MA0024	E2F1 (TFF)
10	HLF	MA0043	HLF (LEUZIP)
10	NFIL3	MA0025	NFIL3 (LEUZIP)
11	ELF5	MA0136	ETS GRP
11	ELK1	MA0028	ETS GRP
11	ELK4	MA0076	ETS GRP
11	GABPA	MA0062	ETS GRP
11	SPIB	MA0081	ETS GRP
12	Evi1	MA0029	EVi1 (ZF)
13	Foxa2	MA0047	FKH GRP
13	FOXD1	MA0031	FKH GRP
13	Foxd3	MA0041	FKH GRP
13	FOXF2	MA0030	FKH GRP
13	FOXI1	MA0042	FKH GRP
13	Foxq1	MA0040	FKH GRP
14	Gfi	MA0038	GFI (ZF)
14	NFYA	MA0060	NFYA (CCAATBOX)
14	PBX1	MA0070	PBX1 (HOX)
15	HNF1A	MA0046	HNF1A (HOX)

Table S15. Vertebrate TF profile clustering and class label assignment (continued).

Cluster #	Jaspar TFBS	Jaspar TFBS ID	TFBS Class Label
16	Myb	MA0100	MYB (MYB)
16	Myf	MA0055	MYF (HLH)
16	NHLH1	MA0048	NHLH1 (HLH)
17	IRF1	MA0050	IRF1 (IRF)
17	IRF2	MA0051	IRF2 (IRF)
17	STAT1	MA0137	STAT1 (STAT)
18	MEF2A	MA0052	MEF2A (MADS)
19	MZF1_1-4	MA0056	MZF1_1-4 (ZF)
19	MZF1_5-13	MA0057	MZF1_5-13 (ZF)
20	NF-kappaB	MA0061	REL GRP
20	NFKB1	MA0105	REL GRP
20	REL	MA0101	REL GRP
20	RELA	MA0107	REL GRP
21	Lhx3	MA0134	HOX GRP
21	Lhx3_V2	MA0135	HOX GRP
21	Nkx2-5	MA0063	HOX GRP
21	Nobox	MA0125	HOX GRP
21	Pdx1	MA0132	HOX GRP
22	ESR1	MA0112	ESR1 (NR)
22	PPARG	MA0066	PPARG (NR)
23	Pax4	MA0068	PAX3 (HOX)
24	RREB1	MA0073	RREB1 (ZF)
25	RXRA-VDR	MA0074	RXRA-VDR (NR)
26	Prrx2	MA0075	PRRX2 (HOX)
27	Sox17	MA0078	SOX/HMG GRP
27	Sox5	MA0087	SOX/HMG GRP
27	SOX9	MA0077	SOX/HMG GRP
27	SRY	MA0084	SOX/HMG GRP
28	SP1	MA0079	SP1 (ZF)
29	SRF	MA0083	SRF (MADS)
30	Staf	MA0088	STAF (ZF)
31	TEAD1	MA0090	TEAD1 (TEA)
32	TAL1-TCF3	MA0091	TAL1 (HLH)
32	YY1	MA0095	YY1 (ZF)
33	Hand1-Tcfe2a	MA0092	HAND1-TCFE2A (HLH)
34	Fos	MA0099	FOS (LEUZIP)
35	TP53	MA0106	TP53 (LSH)
36	Spz1	MA0111	SPZ1 (HLH)
37	Roaz	MA0116	ROAZ (ZF)
38	TLX1-NFIC	MA0119	TLX1-NFIC (HOX-CCAAT)
39	Bapx1	MA0122	NKX/HOX GRP
39	NKX2-2	MA0506	NKX/HOX GRP
39	NKX2-5_V2	MA0507	NKX/HOX GRP
39	NKX3-1	MA0124	NKX/HOX GRP
40	ZNF354C	MA0130	ZNF354C (ZF)

Table S15. Vertebrate TF profile clustering and class label assignment (continued).

Cluster #	Jaspar TFBS	Jaspar TFBS ID	TFBS Class Label
41	MIZF	MA0131	MIZF (ZF)
42	REST	MA0138	REST (ZF)
43	Pou2F1/Oct1	MA0500	POU2F1 (POU-HOX)
44	Egr1/Krox-24/NGFI-A	MA0501	EGR/ZF GRP
44	Egr2/Krox-20	MA0502	EGR/ZF GRP
44	Egr3	MA0503	EGR/ZF GRP
44	Egr4	MA0504	EGR/ZF GRP
45	POU3F1/OCT6	MA0505	POU3F1 (POU-HOX)
46	CTCF_ren	MA0508	CTCF (ZF)
47	Gli	MA0509	GLI (ZF)

Figures

Figure S1. Overview of methodology.

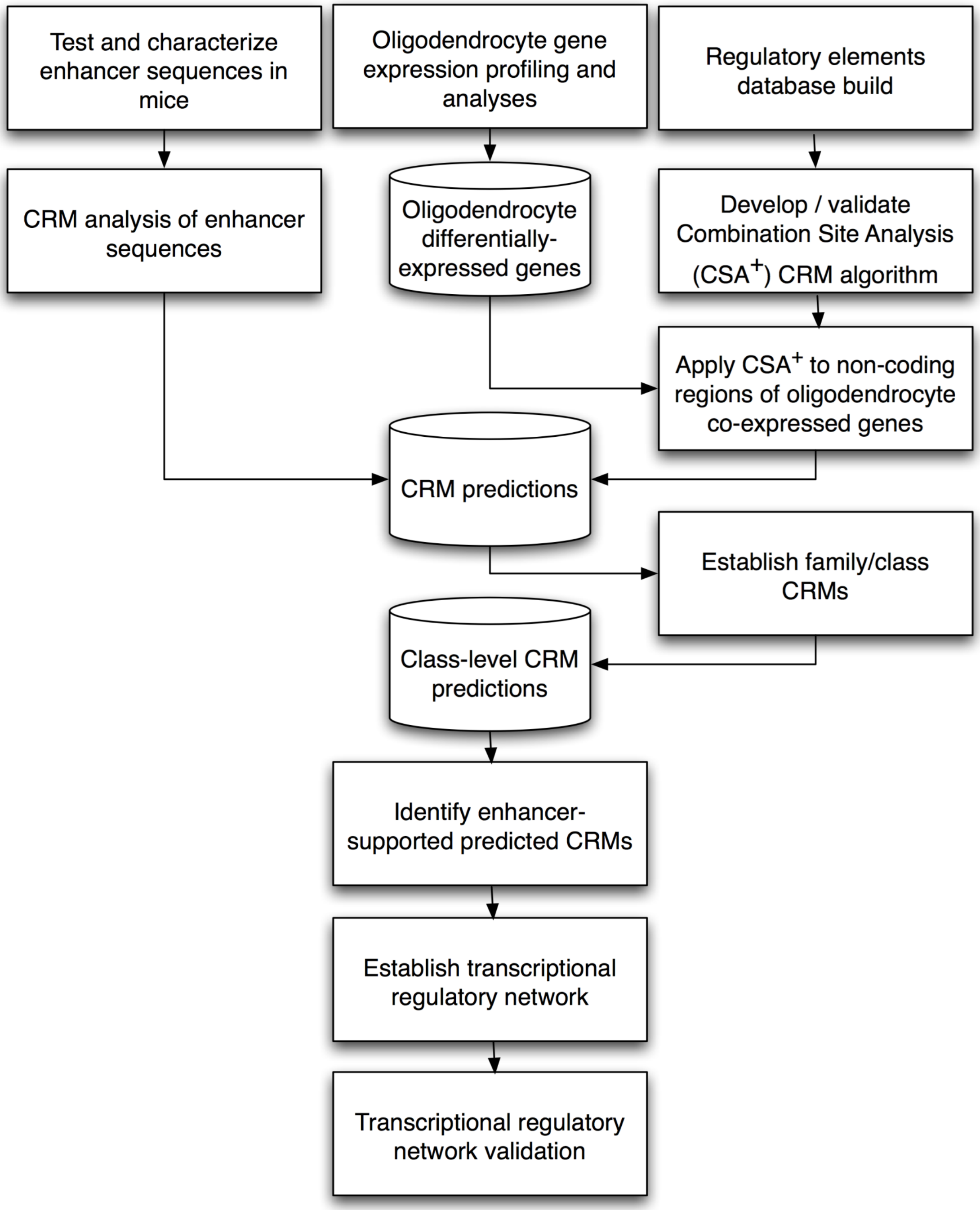
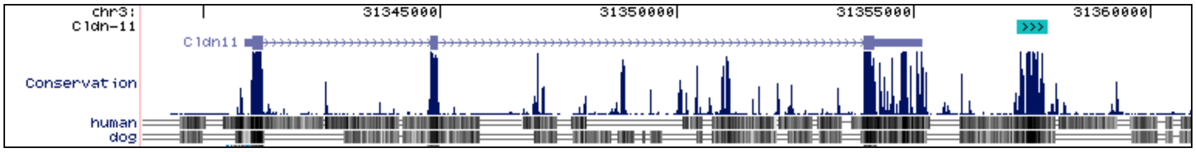
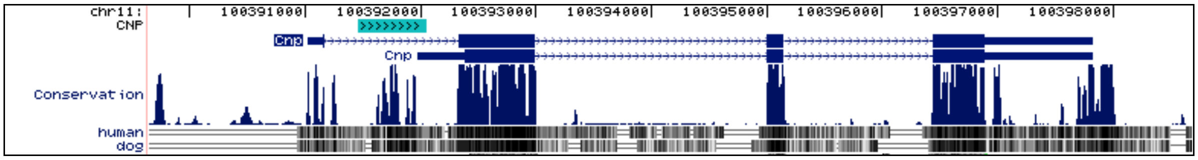


Figure S2. Genomic context for myelin gene-associated enhancer regions. Displayed using mouse MM8-mapped coordinates http://genome.ucsc.edu.

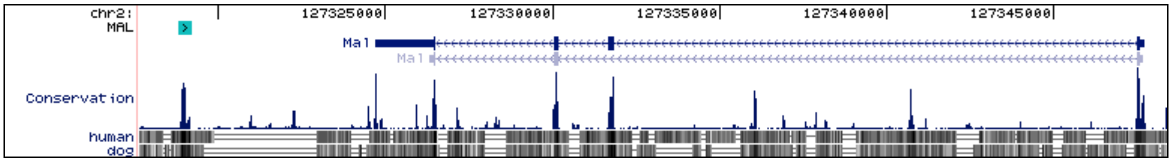
Cldn11



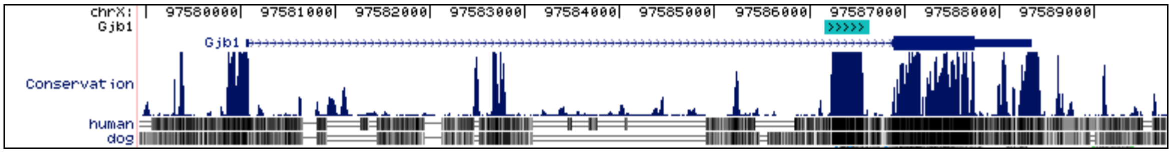
Cnp



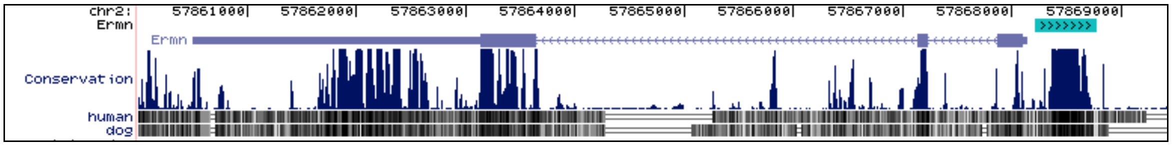
Mal



Gjb1



Ermn



Olig1

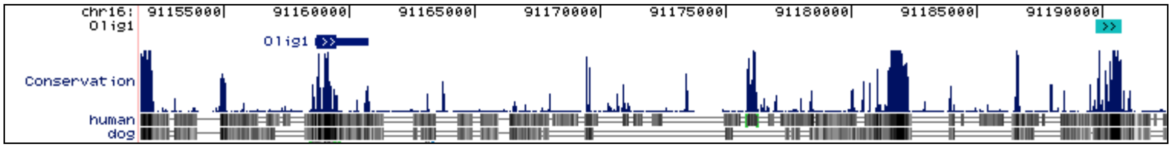
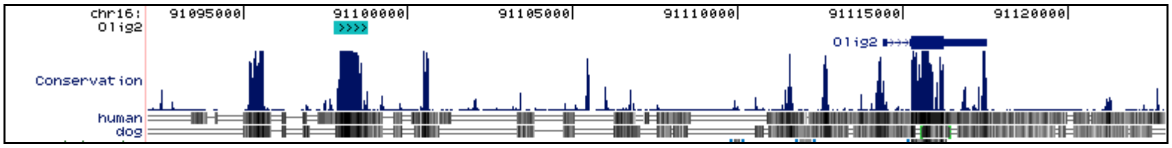


Figure S2. Genomic context for myelin gene-associated enhancer regions (continued).

Olig2



Pou3f1

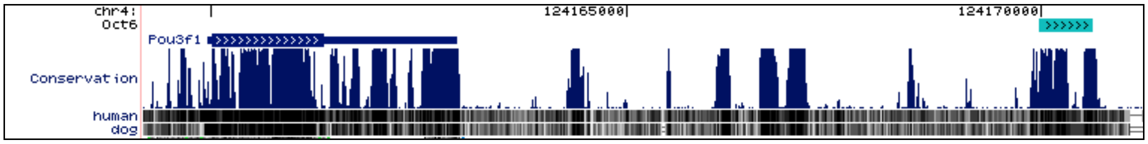
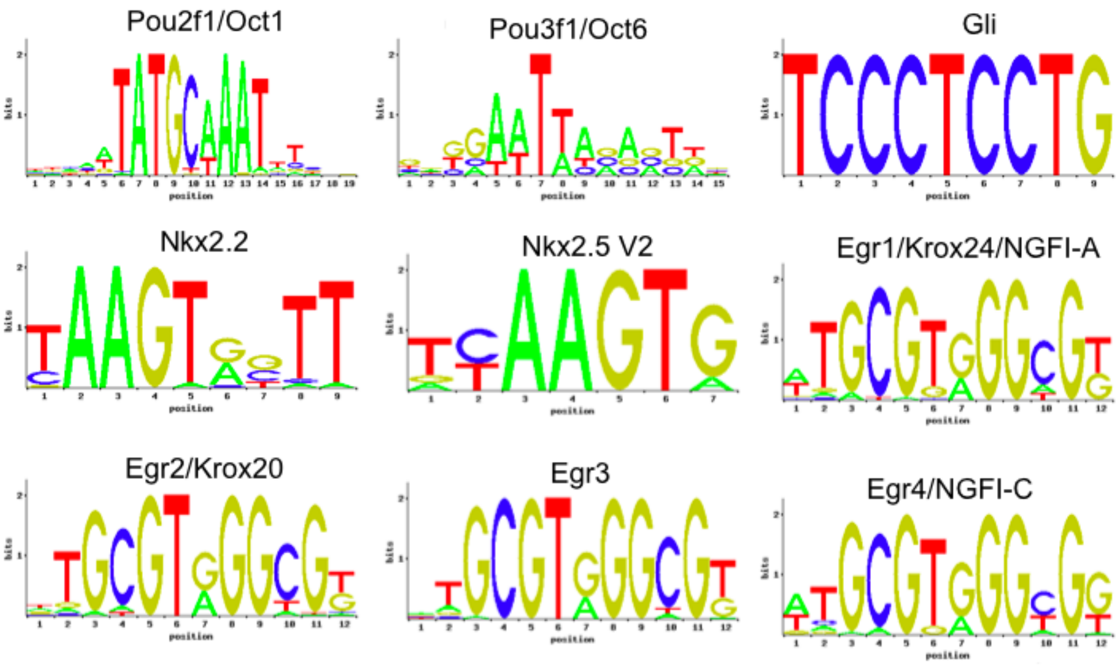


Figure S3. Logos of transcription factor binding site profiles added to the Jaspar database for the promoter analyses.



Pou2f1/Oct1

Pou3f1/Oct6

Gli

Nkx2.2

Nkx2.5 V2

Egr1/Krox24/NGFI-A

Egr2/Krox20

Egr3

Egr4/NGFI-C

Figure S4. Reporter gene activity in mouse spinal cords for *Cldn11*, *Gjb1*, *Pou3f1*, and *Mal* constructs.

Histochemical detection of β-galactosidase activity in mouse spinal cord cross sections at P5, P10, and adult (2 – 3 months old) stages. Increasing ages are placed in successive rows.

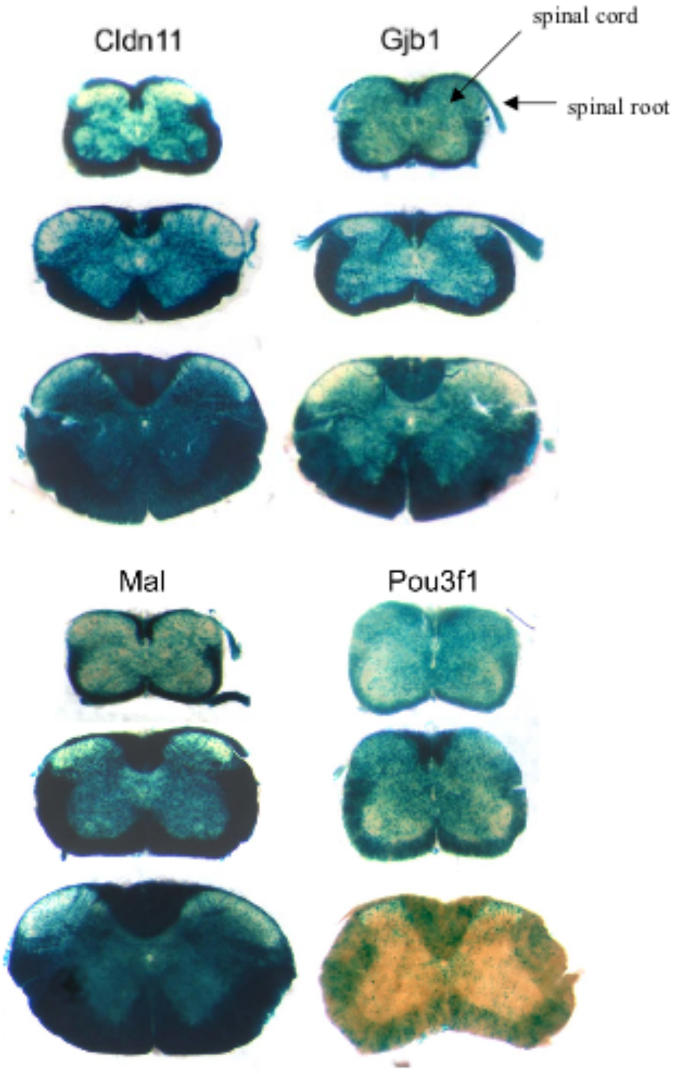


Figure S5. Wholemount staining of mouse spinal cords/spinal roots for pre-weaning and mature developmental stages.

Wholemount staining of mouse spinal cords/spinal roots for P5-P10 and adult (2 – 3 months) developmental stages. In the spinal cord, the *Cldn11* construct is expressed in oligodendrocytes, while the *Gjb1*, *Mal* and *Pou3f1* constructs are expressed in both oligodendrocytes and in Schwann cells of spinal roots.

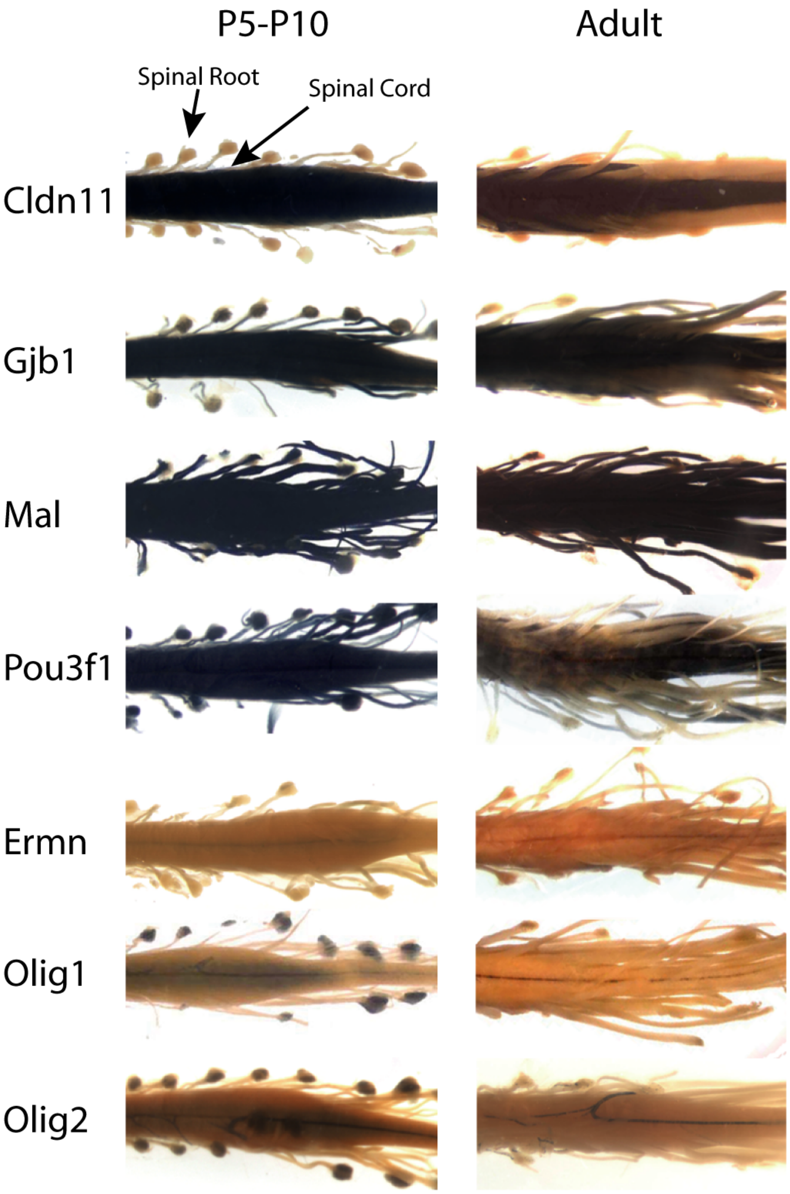


Figure S6. Reporter gene activity in mouse brains for *Ermn*, *Olig1*, and *Olig2* constructs.

Histochemical detection of β-galactosidase activity in mouse brain sagittal sections for reporter constructs that revealed little or no staining in white matter at P5, P10, and adult (2 – 3 months old) stages. Increasing ages are placed in successive rows.

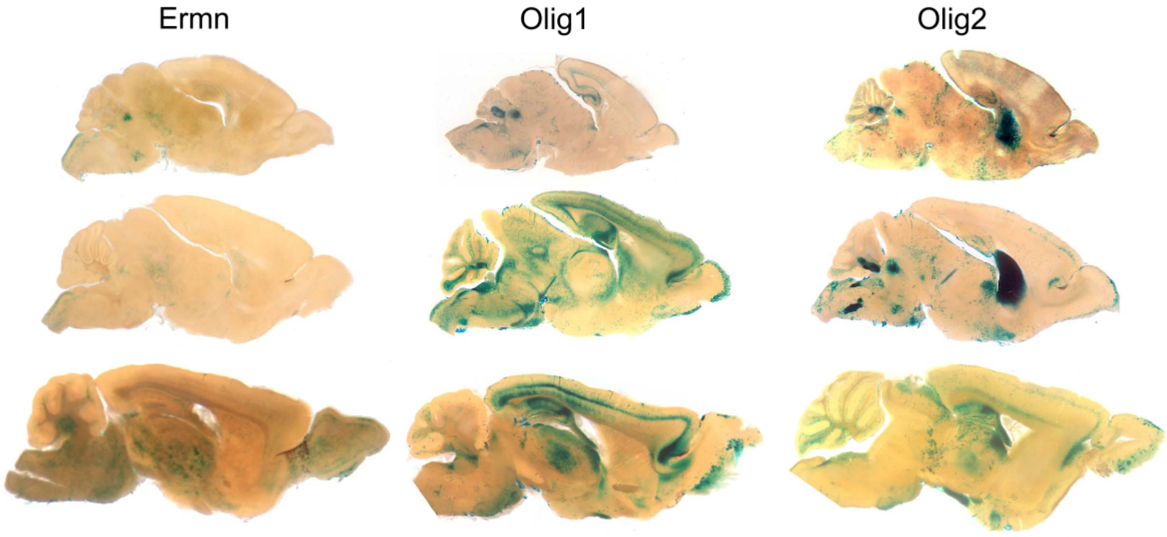


Figure S7. Reporter gene activity in mouse spinal cords for *Ermn*, *Olig1*, and *Olig2* constructs.

Histochemical detection of β-galactosidase activity in mouse spinal cord cross sections for reporter constructs that revealed little or no staining in white matter at P5, P10, and adult (2 – 3 months old) stages. Increasing ages are placed in successive rows.

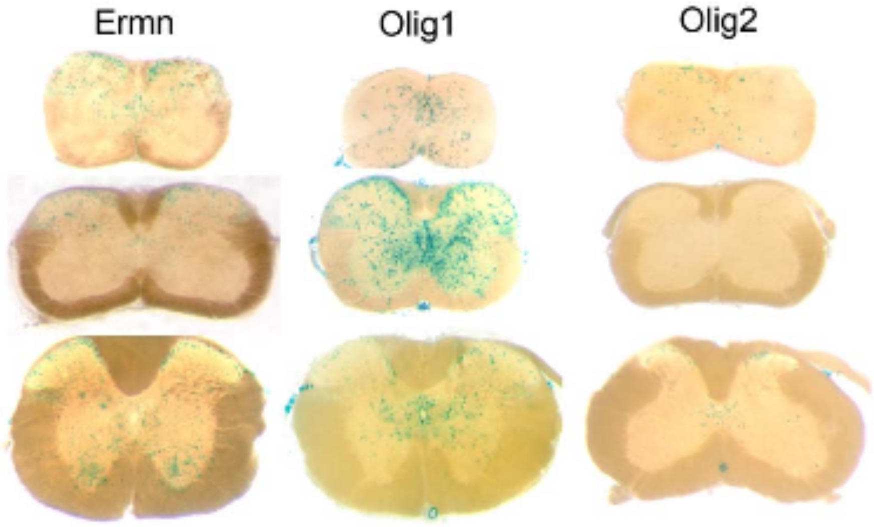
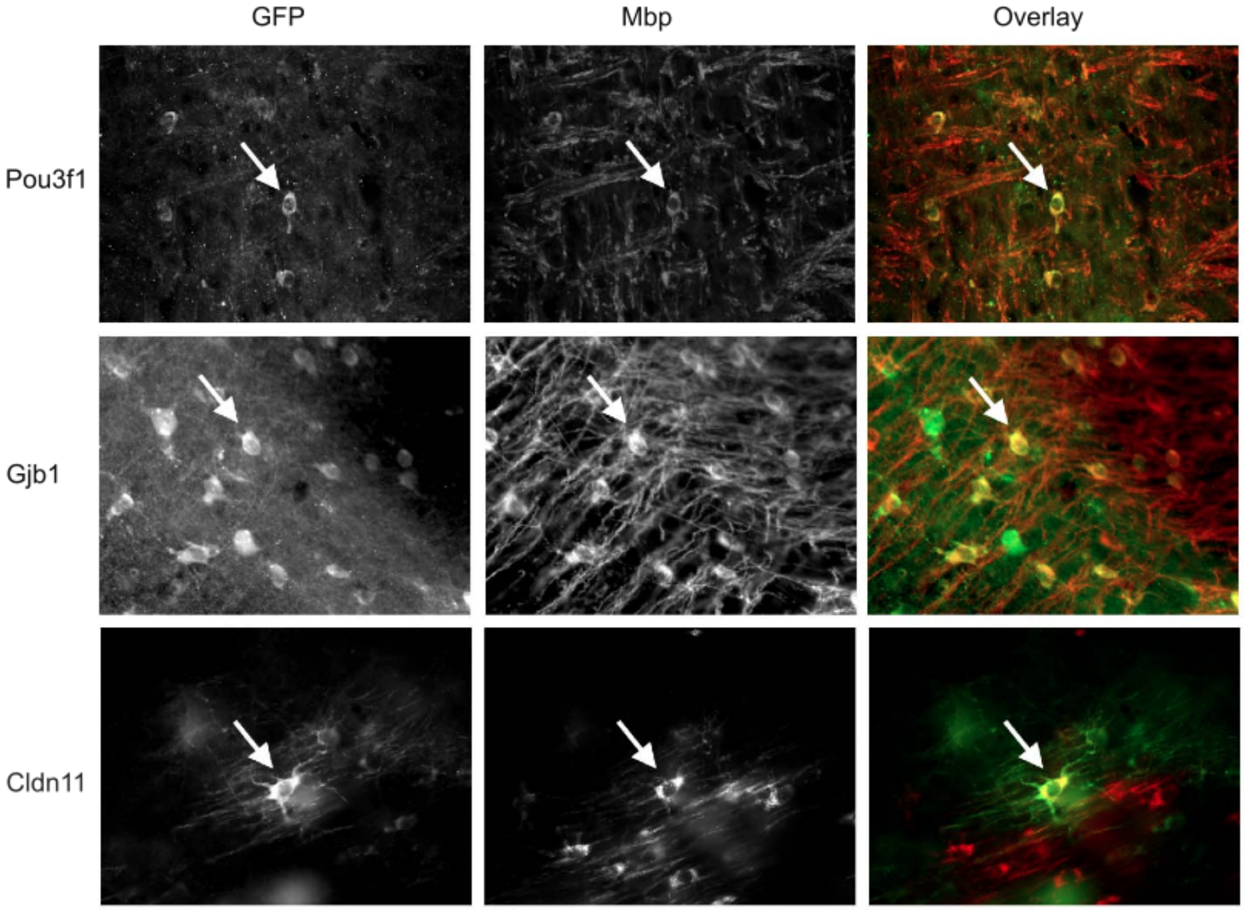


Figure S8. Characterization of cell populations expressing MBP and the *Gjb1*, *Cldn11*, *Pou3f1*, and *Mal* constructs.

Images of brain sections (pons) stained for MBP (red) and GFP (green) by classical (A) and confocal (B) imaging reveals multiple oligodendrocyte lineage cells that co-express MBP and the *Gjb1*, *Cldn11*, and *Pou3f1* transgenes (arrows).

A.



B.

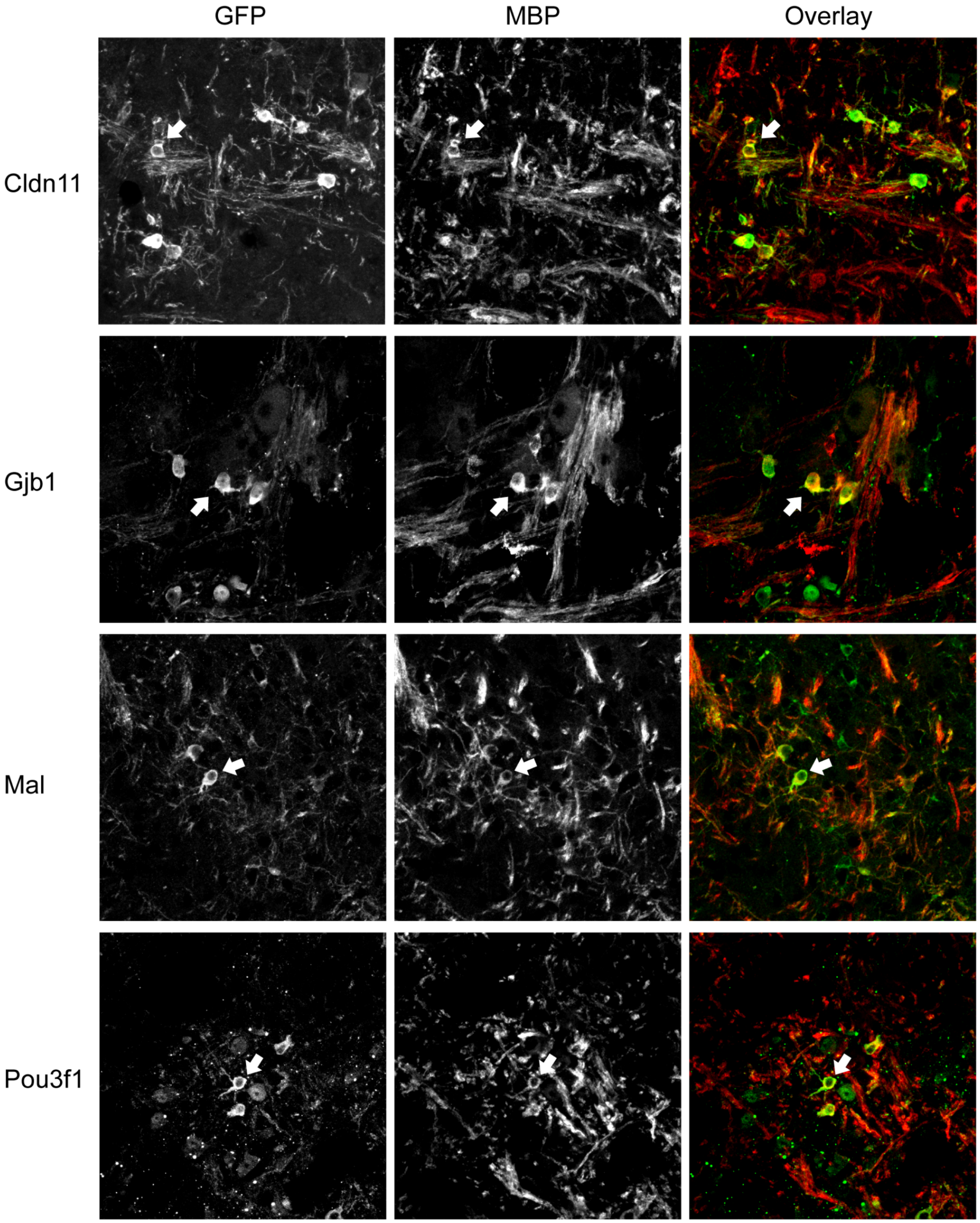


Figure S9. Characterization of cell populations expressing PDGFRA and the *Gjb1*, *Cldn11*, *Pou3f1*, and *Mal* constructs.

Confocal images of P10 brain sections reveal that *Cldn11*, *Gjb1*, *Mal*, and *Pou3f1* constructs are co-expressed in some PDGFRA+ cells.

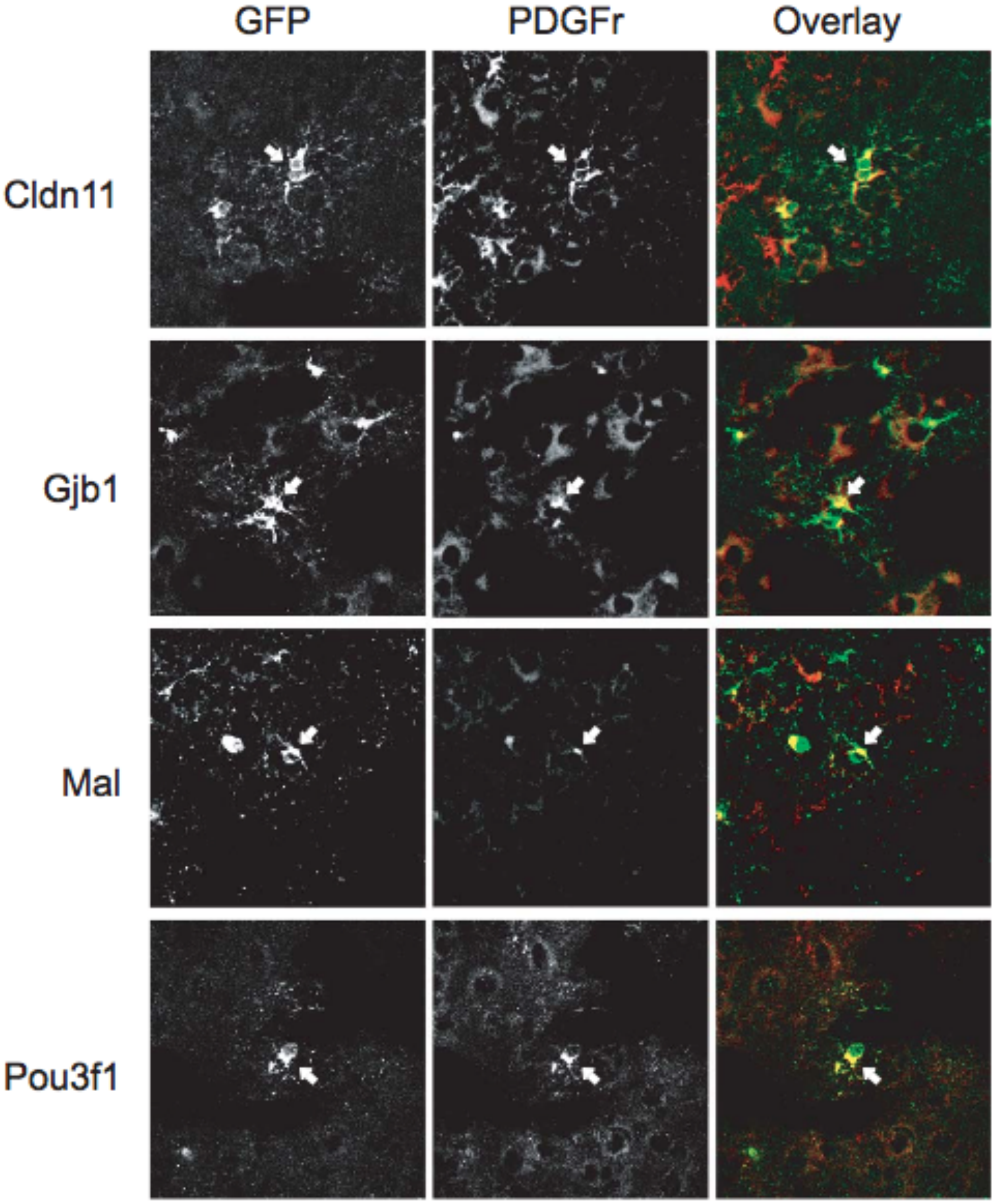


Figure S10. Evaluation of sequence conservation flanking transcription factor binding sites.

Human-mouse sequence conservation flanking an experimentally validated Ap1/Fos transcription factor binding site (using binding site information: AP1_EXTRACTED50 extracted from the Annotated Binding Site database (5)).

Ap1/Fos

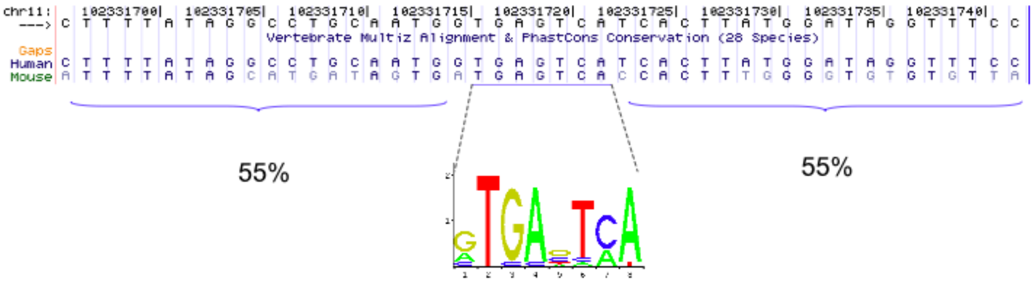


Figure S11. Myelin gene TRN model validation study 1.

Reference: 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.

Summary: The study demonstrated the activity and interaction of Olig (E-box binding) TFs with Sox TFs in the MBP gene non-coding sequence.

Analysis: The myelin TRN data tool was used to extract predicted CRMs for interactions with SOX5 TFBS (the *SOX5* binding site is similar to other Sox family member binding sites) in the *Mbp* upstream non-coding region examined by the study. We used the human orthologous region: chr18:72857989-72858227.

All Sox and E-box binding sites validated by the study were recovered in the CRM predictions (green bars in browser image below are predicted CRM TFBS and red asterisked TFBS were validated in Li *et al*. 2007). Additional TFBSs that may be bound by TFs known to regulate oligodendrocyte transcription (blue asterisks) were also predicted.

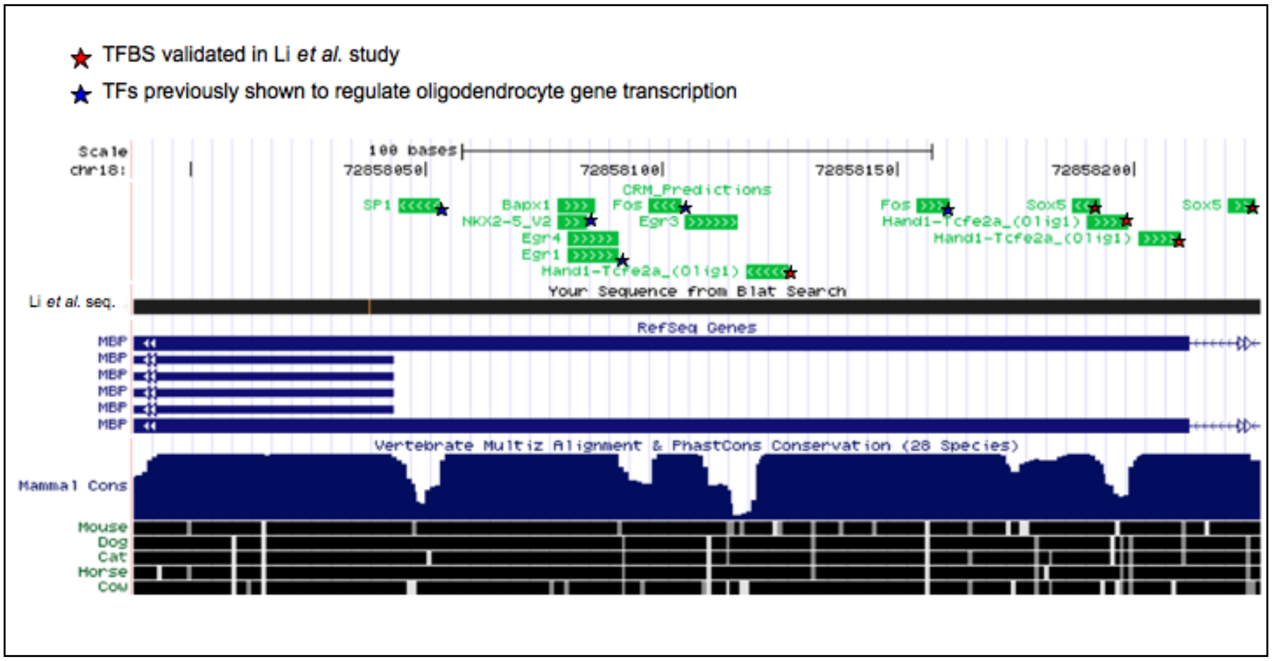


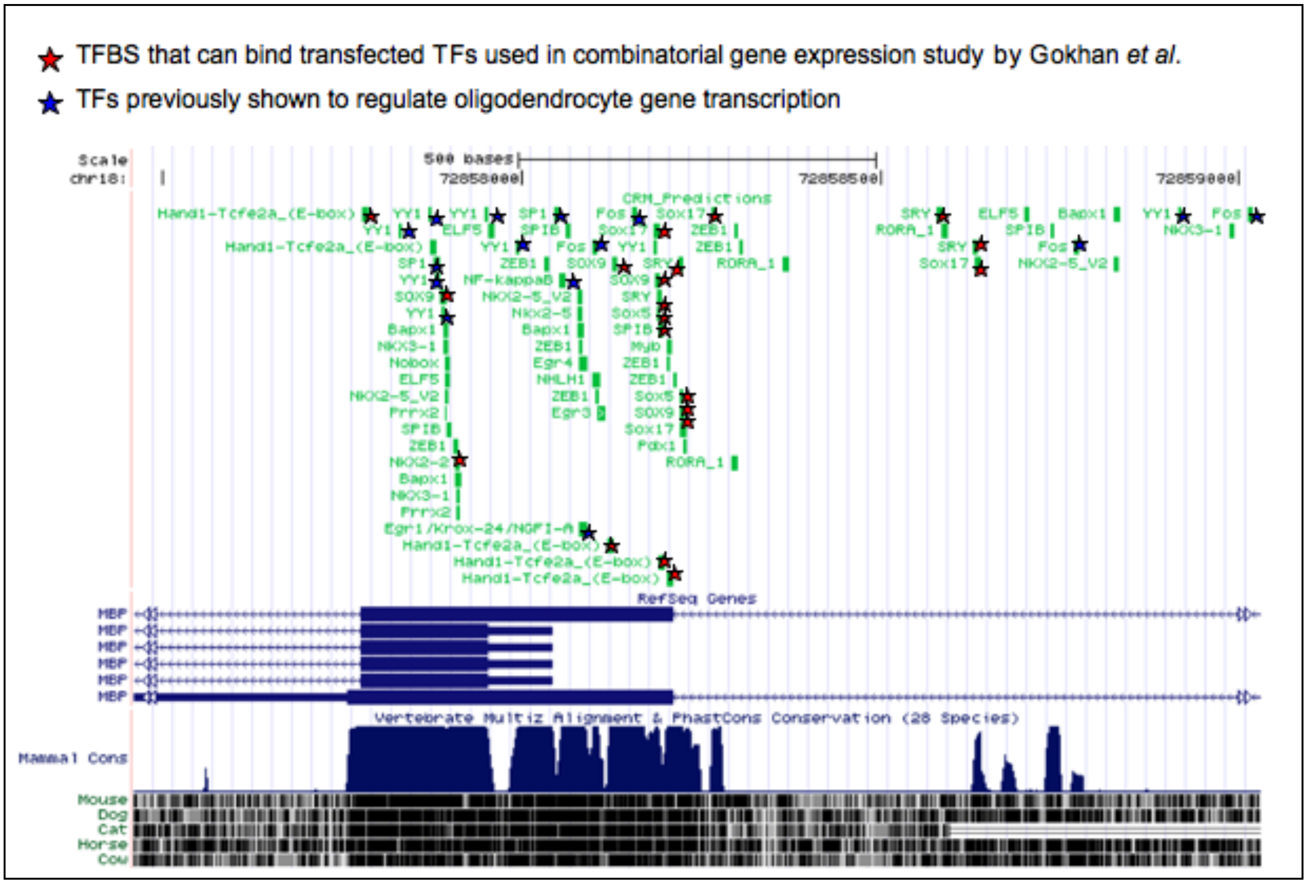
Figure S12. Myelin gene TRN model validation study 2.

Reference: Gokhan, S., Marin-Husstege, M., Yung, S.Y., Fontanez, D., Casaccia-Bonnefil, 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.

Summary: Study explores the combinatorial effects of the canonical E-box binding TFs: OLIG1, OLIG2, ASCL1, SOX10 TFs, inhibitory bHLH TFs: ID2 and ID4, and the NKX2-2 TF in transfection assays using an oligodendrocyte cell line.

Analysis: The myelin TRN data tool was used to extract all predicted CRMs in the *Mbp* upstream non-coding region examined by the study. We used the human orthologous region: chr18:72857970-72859290.

The extracted CRM predictions for this region (green bars in browser image below are predicted CRM TFBS) included TFBS capable of binding all TFs tested in the study (red asterisks). Additional TFBSs that may be bound by TFs known to regulate oligodendrocyte transcription (blue asterisks) were also predicted.



Supplemental references

1. Morikawa, Y. and Cserjesi, P. (2004) Extra-embryonic vasculature development is regulated by the transcription factor HAND1. *Development*, **131**, 2195-2204.
2. Srivastava, D., Cserjesi, P. and Olson, E.N. (1995) A subclass of bHLH proteins required for cardiac morphogenesis. *Science*, **270**, 1995-1999.
3. Bounpheng, M.A., Morrish, T.A., Dodds, S.G. and Christy, B.A. (2000) Negative regulation of selected bHLH proteins by eHAND. *Exp. Cell Res*, **257**, 320-331.
4. Morin, S., Pozzulo, G., Robitaille, L., Cross, J. and Nemer, M. (2005) MEF2-dependent recruitment of the HAND1 transcription factor results in synergistic activation of target promoters. *J. Biol. Chem.*, **280**, 32272-32278.
5. 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**, D63-67.