Deciphering peripheral nerve myelination by using Schwann cell expression profiling

Rakesh Nagarajan*, Nam Le*, Heather Mahoney, Toshiyuki Araki, and Jeffrey Milbrandt†

Departments of Pathology and Internal Medicine, Washington University School of Medicine, 660 South Euclid Avenue, Box 8118, St. Louis, MO 63110

Edited by Eric M. Shooter, Stanford University School of Medicine, Stanford, CA, and approved May 6, 2002 (received for review February 8, 2002)

Although mutations in multiple genes are associated with inherited demyelinating neuropathies, the molecular components and pathways crucial for myelination remain largely unknown. To approach this question, we performed genome-wide expression analysis in several paradigms where the status of peripheral nerve myelination is dynamically changing. Anchor gene correlation analysis, a form of microarray analysis that integrates functional information, using correlation-based clustering, with a statistically rigorous test, the Westfall and Young step-down algorithm, was applied to this data set. Biological pathways active in myelination, genes encoding proteins involved in myelin synthesis, and genes whose mutation results in myelination defects were identified. Many known genes and previously uncharacterized ESTs not heretofore associated with myelination were also identified. One of these ESTs, *MASR* (myelin-associated SUR4 protein), encodes a member of the SUR4 family of fatty acid desaturases, enzymes involved in elongation of very long chain fatty acids. Its specific localization in myelinating Schwann cells indicates a crucial role for *MASR* in normal myelin lipid synthesis.

Rapid saltatory conduction of electrical signals along the distance of an axon is facilitated by myelin, a lipid-rich ionic insulator that wraps around the axon. In the central nervous system, myelination is accomplished by oligodendrocytes, whereas the analogous role is fulfilled by Schwann cells in the peripheral nervous system. Disruption of myelination in diseases, such as hereditary motor and sensory neuropathies in the peripheral nervous system and multiple sclerosis in the central nervous system, are highly prevalent and result in significant morbidity and mortality (1).

Alterations in several genes including myelin protein zero (*MPZ*), peripheral myelin protein (*PMP22*), connexin32 (*Cx32*), and *EGR2* have been associated with inherited demyelinating neuropathies (for review, see ref. 1). However, the molecular components of myelination remain largely unknown, and a significant proportion of patients with inherited myelinopathies do not have mutations in any of the above genes, underscoring the need to identify additional candidate genes that may be involved.

With the advent of microarray technology, genome-wide expression analyses are now possible. The predominant types of microarray experiments include comparisons of predefined sample groups (e.g., benign vs. tumor) and assessments of expression over a continuous variable (e.g., a time course). Many of these studies have used clustering techniques, such as hierarchical clustering, correlation clustering, and self-organizing maps (SOMs), to distinguish tumor subtypes via “expression fingerprints” and to identify functionally related gene clusters (2–5). The underlying concept derived from these studies is that genes important in a common process share similar expression profiles. However, the complexity of mammalian systems makes it difficult to identify functionally similar genes by cluster analysis. Furthermore, although several methods have been proposed to add statistical rigor to analyses of microarray experiments dealing with predefined sample groups (6–8), it is currently unclear how to assess the statistical significance of techniques such as SOMs or k-means clustering, which are used to analyze contin-

uous variable experiments. In addition, the analysis of microarray data invariably requires performing multiple comparisons, which results in a high occurrence rate of false positives (type I error). A generalized method that can be applied to control this type I error rate, and that also takes into account the dependence structure between variables, is the Westfall and Young step-down algorithm for calculating adjusted *P* values by permutation (9).

In this study, we combined the idea of identifying functionally related genes by their coregulated expression profiles with a statistical test that accounts for multiple hypotheses. We integrated Westfall and Young step-down algorithm with correlation-based clustering into an algorithm called anchor gene correlation analysis (AGCA). AGCA was applied to a small compendium of peripheral nerve expression profiles and enabled the identification of genes that are important during myelination and that may be involved in causing inherited neuropathies. Additionally, through this approach we discovered a gene belonging to the SUR4 family, whose members are known to be involved in very long chain fatty acid (VLCFA) synthesis, a necessary step in generating sphingomyelin and galactocerebroside (Gal-C). Thus, we present here a generalized global approach to studying peripheral nerve biology that has facilitated identification of molecular components of biological processes and has permitted process-directed gene discovery.

Methods

Microarray Analysis. Total RNA was prepared from sciatic nerves from at least ten C57BL/6 mice at each time point for nerve crush (days 0, 1, 4, 7, 10, 14, and 28), nerve development (P0 and P56), or nerve transection (days 0, 14, and 28), as well as for *Egr2*^{lo/lo} mice and WT littermates. Replicates of the uninjured P14 and P56 time points were prepared entirely independently from two separate pools of ten animals each. From the total RNA, biotinylated cRNA probes were generated, fragmented, and applied as described (10) to Mouse MU74A (Version 1) GENECHIP arrays (Affymetrix, Santa Clara, CA). Affymetrix software was used to filter inaccurately represented probe sets. The raw chip data are provided as supporting information, which is published on the PNAS web site, www.pnas.org.

For analysis of the proliferation, cytokine response, and macrophage infiltration processes, we preprocessed (filtered) the data for probe sets that are called “present” at 4 (5,931 probe sets), 1 (5,189 probe sets), or 7 days (5,945 probe sets) post-crush, respectively. We further selected for probe sets that have a ≥ 2 -fold increase in expression compared with uninjured nerve. Similarly, absolute call filtering, fold change filtering, and deviation/mean filter of 0–0.35 for replicate chips, were performed

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: AGCA, anchor gene correlation analysis; VLCFA, very long chain fatty acid; MASR, myelin-associated SUR4 protein; ER, endoplasmic reticulum; MGCC, myelin gene coregulated cluster; RT, reverse transcription; MGAs, myelin gene anchors.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database [accession no. AF480860 (*MASR* mRNA)].

*R.N. and N.L. contributed equally to this work.

†To whom reprint requests should be addressed. E-mail: jeff@milbrandt.wustl.edu.

for the myelination process (see Fig. 2), using DCHIP ANALYZER software (11) and FUNCTION EXPRESS (software written in our lab). Genes that were “called” absent on all chips by the DCHIP software and genes whose expression varied by >0.35 in two independent microarray experiments conducted on adult nerve were eliminated. AGCA was performed to find statistically significant coregulated genes. The correlation coefficient of the expression profile of an anchor gene and the expression profile of each of the other probe sets was determined using the Pearson product moment correlation coefficient (PPMCC) function. To measure the significance of the correlations, we applied the Westfall and Young step-down algorithm essentially as described (9). Briefly, the t statistic was replaced by the PPMCC and a one-tailed analysis was performed. One thousand random permutations were performed singly for each anchor or for all anchors. Adjusted P values were calculated under asymptotic normality enforcing monotonicity constraints. For analysis of P0 vs. uninjured adult P56 nerve and *Egr2^{lo/lo}* nerves vs. WT (P14) nerve, probe sets were selected that were called present in the baseline chips and called “decreased” by ≥ 2 fold change based on the difference call metric by the Affymetrix analysis software.

Sciatic Nerve Injury and Postnatal Development. All surgical procedures followed National Institutes of Health guidelines for the care and use of laboratory animals at Washington University. For each injury time point, ten 8-week-old (P56) male C57BL/6 mice were anesthetized, and the right sciatic nerve was injured at the hip level by compressing (crush) for 30 sec or cutting (transection) the nerve and ensuring the ends were not in contact. The contralateral nerve was exposed, but left uninjured (control). After the indicated length of time, the animals were killed for immediate collection of tissues for RNA isolation. For the developmental paradigm, 60 P0 and 10 P56 animals were anesthetized, collecting both sciatic nerves for RNA isolation.

Cell Lines, Transfection, and Subcellular Localization. Preparation of a purified population of cultured Schwann cells has been described (12). Transient transfections of CV-1 cells with *MASR*-GFP fusion proteins (generated using plasmid EPEHisCsNeo, a gift from Jane Wu, Department of Pediatrics, Washington University School of Medicine) were performed in 24-well plates, using SuperFect (Qiagen) as described by the manufacturer. Twenty-four hours after transfection, cells were stained with 1 mM MitoTracker (Molecular Probes) or 1 mM LysoTracker (Molecular Probes) for 30 min, washed, and re-incubated in media. For immunocytochemistry, transfected CV-1 cells were fixed in ice-cold methanol for 5 min, followed by ice-cold acetone for 30 s. Anti-calnexin polyclonal antibody (StressGen Biotechnologies, Victoria, BC, Canada) was used at 1:100 for 1 h, followed by Cy3-conjugated anti-rabbit secondary antibody (The Jackson Laboratory) at 1:100.

Quantitative Reverse Transcription (RT)-PCR (TaqMan) Analysis of Gene Expression and Sequence Analysis. Expression levels of genes were measured by quantitative RT-PCR analysis with 18S rRNA normalization of samples similarly as described (13). Primer sequences used for quantitative analysis of each gene are available on request. Sequence of the mouse *MASR* cDNA was obtained by sequencing EST IMAGE clones 2182209 and 2352559, using standard methods and sequence information available in National Center for Biotechnology Information Unigene cluster Mm.26171. Human myelin-associated SUR4 protein (*MASR*) cDNA was found by homology, using sequence available under GenBank accession no. AK027031.

In Situ Hybridization Analysis. Sense and antisense digoxigenin-labeled RNA probes for *in situ* hybridization of *MASR* were transcribed from EST IMAGE clone 1400770 after digesting

with *NotI* or *XhoI*, respectively. For *MPZ* probes, nucleotides 181–700 of *MPZ* mRNA (accession no. NM.008623) were cloned into Bluescript II KS ($\pm$). *In situ* hybridization analysis was performed on fresh-frozen nerve sections as described (14).

Generation of *Egr2^{lo/lo}* Mice. Briefly, the same recombination arms as those used to disrupt *Egr2* by Schneider-Maunoury *et al.* (15) was used to generate the *Egr2^{lo/lo}* mice. A PGK (phosphoglycerate kinase promoter) neomycin-resistance cassette flanked by loxP sites was inserted into *EGR2*’s only intron. Introduction of this cassette results in severely decreased *Egr2* expression and concomitant hypomyelination of peripheral nerves as determined by light and electron microscopy.

Results

Anchor Gene Correlation Analysis Identifies Coregulated Genes Involved in Multiple Processes. To study the process of peripheral nerve myelination after injury, microarray time courses were generated following nerve crush and transection, paradigms in which the state of myelination is dynamically changing. The sciatic nerve was chosen for its accessibility and relative homogeneity and predominance of Schwann cells. An appropriate time course following sciatic nerve crush injury was selected, such that the processes of demyelination, Wallerian degeneration, Schwann cell proliferation, axonal regeneration, and subsequent remyelination could all be monitored. The segments of nerve distal to the site of injury were collected for microarray processing at days 0 (uninjured adult P56 nerve), 1, 4, 7, 10, 14, and 28 post-injury. Unlike the crush paradigm, no axonal regeneration, and therefore no remyelination, occurs in the distal stump following transection of a peripheral nerve, resulting in a sustained decrease of myelination-specific gene expression. Thus, distal nerve segments from time points of 14 and 28 days post-transection were also collected.

Because the uninjured adult nerve (P56) serves as a baseline for comparison, entirely independent replicate samples and microarray chips were performed at this time point. When these replicates were compared, 98% of 12,654 probe sets were called unchanged when using the AFFYMETRIX EXPRESSION SUITE software (see Fig. 6*a*, which is published as supporting information on the PNAS web site). However, when comparing the normal nerve chip to that obtained from injured nerve (4 days post-crush), more than 15% of 12,654 probe sets exhibited changes in expression from baseline (see Fig. 6*b*).

Using these expression profiles, we asked whether AGCA could identify genes involved in biological processes active after nerve crush. Following nerve injury, Schwann cells dissociate from the degenerating axon and begin proliferating (16, 17). To examine this process by using AGCA, Ki-67 antigen, a marker intimately associated with cell proliferation, was chosen as the anchor gene. Ki-67 antigen expression rises sharply following injury, peaks at 4 days post-injury, and then returns to baseline levels, consistent with prior BrdUrd labeling and Northern blot studies (17). After a filtering step to select the most informative probe sets (see *Methods*) and performing AGCA, we found 19 known genes whose expression profiles correlated significantly with that of Ki-67 (adjusted $P \leq 0.05$), and 15 of these genes were highly associated with proliferation (Fig. 1*a*, and see Table 1, which is published as supporting information on the PNAS web site). However, the gene for proliferating cell nuclear antigen (*PCNA*), also a widely accepted marker for proliferation, was not identified. We reasoned that because proliferation is monitored by several widely accepted markers, the employment of multiple anchor genes would provide a more comprehensive identification of genes involved in this process. Indeed, by querying for genes correlating significantly with Ki-67 or *PCNA* individually or combined, an additional nine genes were identified, seven of which are associated with proliferation. Genes in this list in-

cluded cyclins, enzymes essential for nucleotide metabolism, histones, and mitotic motor proteins. Thus, whereas AGCA can be performed with one anchor gene, the use of multiple anchors identifies a greater number of genes associated with the process of interest.

Following peripheral nerve injury, the inflammatory response is crucial for Wallerian degeneration and subsequent regeneration of the nerve (18). Circulating macrophages must be recruited into the endoneurium, which requires the release of inflammatory cytokines. In monitoring the cytokine response, monocyte chemoattractant protein 1 (*MCP-1*) and *IL-6* were chosen as anchor genes because they are potent cytokines for macrophage recruitment and inflammatory initiation, respectively, following nerve injury (19, 20). Consistent with previous data, *MCP-1* and *IL-6* are up-regulated immediately after nerve crush with levels peaking 1 day after injury. To further validate this profile as representative of the inflammatory cytokine response, genes whose temporal expression profiles correlated (adjusted $P \leq 0.05$) with *MCP-1* and/or *IL-6* were identified using AGCA. We found that several cytokines and inflammatory-related genes clustered with these anchors (Fig. 1b, and see Table 1). Seven of nine known genes in this group encode cytokines or inflammatory proteins.

With the release of inflammatory cytokines, macrophages are subsequently recruited to phagocytose cellular and myelin debris, allowing regrowth of axons through the nerve tubules. The genes encoding the common macrophage marker CD68 and the antigen recognized by the macrophage-specific antibody F4/80 (*Emr1*) were selected as anchors to monitor macrophage infiltration. The expression level of these two genes gradually increases following injury, peaks after 7 days, then declines toward basal levels, consistent with previous observations of macrophage infiltration into injured nerve (16, 21). Indeed, 11 of the 14 known genes identified were related to macrophages or the inflammatory process, including *CSF-1 receptor*, a well characterized macrophage marker (Fig. 1b, and see Table 1).

Identification of Genes Important for Myelination by Using AGCA. The examination of the proliferative and inflammatory responses by using AGCA confirmed that functionally related genes very often share similar expression patterns. To explore the molecular components of the remyelination program, we applied AGCA to the peripheral nerve crush and transection injury expression profiles. Because *MPZ*, *PMP22*, *MBP*, *MAG*, *periaxin*, proteolipid protein (*PLP*), and the DM20 variant of *PLP* have all been shown to be critical for myelination, these genes, which we designate myelin gene anchors (MGAs), were chosen as anchors to identify other genes involved in the myelination process (1, 22, 23).

In our algorithm, the probe sets selected by the filtering step (see *Methods*) were further stratified by taking into account the magnitude of change in expression level following injury. Using this magnitude filter, we selected 293 probe sets whose decrease in expression after nerve crush (day 4) and after nerve transection (day 14 or day 28) was greater than or equal to the average decrease of expression of the MGAs after these injuries (Fig. 2). Finally, application of AGCA identified 98 probe sets, which we call the myelin gene coregulated cluster (MGCC), whose pattern of expression following injury was correlated significantly (adjusted $P \leq 0.05$) with that of one or all of the MGAs (Fig. 2, and see Table 2, which is published as supporting information on the PNAS web site).

To further assess the association of these genes with myelination, their expression levels were examined in peripheral nerve from a developmental perspective. Because myelination commences shortly after birth in rodents, we would expect myelination-associated genes, similar to *MPZ* and *MAG* for example, to be expressed at a lower level at age P0, increasing in expression as myelination ensues postnatally. We found that

Fig. 1. Genes with similar expression profiles participate in a common process. The normalized expression of genes identified by AGCA, whose expression profiles correlate significantly with that of *Ki-67* and/or *PCNA* (a), *MCP-1* and/or *IL-6* (b, blue lines), or *Emr1* and/or *CD68* (b, red lines), is plotted as a function of days after nerve crush.

approximately 50% of the genes in the MGCC are expressed ≥ 2 -fold higher in adult nerve compared with P0 nerve (Fig. 2). We also examined the expression of MGCC genes in Schwann cells derived from *Egr2^{lo/lo}* mice whose nerves are hypomyelinated because of severely decreased expression of the key myelination transcriptional regulator, *Egr2* (N.L., R.N., T.A., and J.M., unpublished data). This insufficiency of *Egr2* activity diminishes expression of its target genes and the subsequent induction of myelination. We reasoned that MGCC genes involved in myelination would also be expressed at lower levels in *Egr2^{lo/lo}* hypomyelinated nerve. Indeed, we found that approximately 50% of genes in the MGCC were decreased ≥ 2 -fold in hypomyelinated nerve, indicating that it includes many genes important for myelination by Schwann cells (Fig. 2).

Inspection of the genes in the MGCC revealed that three genes

Fig. 2. Stratification algorithm using AGCA identifies genes involved in myelination. Gene expression profiles following peripheral nerve crush and transection were established. Using AGCA, 98 probe sets, termed the MGCC, were identified that correlated significantly (adjusted $P \leq 0.05$) with the MGAs, genes crucial for myelination (*MPZ*, *MBP*, *PMP22*, *periaxin*, *MAG*, *PLP*, or *DM20*). At each step there was an enrichment for genes whose expression was decreased in sciatic nerves of age P0 mice and in sciatic nerves of a mouse model of congenital hypomyelination (*Egr2^{lo/lo}*). The average expression plot of all probe sets in the MGCC resembles profiles characteristic of genes encoding critical myelin proteins during development and following nerve crush and transection injuries.

have been implicated in myelination and/or are involved in neuropathies: *NDRG1*, myelin and lymphocyte protein (*MAL*), and protein tyrosine phosphatase-epsilon (*PTP-ε*). Truncation mutations in *NDRG1* have recently been shown to cause hereditary motor and sensory neuropathy-Lom, an autosomal recessive form of Charcot-Marie-Tooth (CMT) disease that results in hypomyelination and demyelination (24). *MAL*, a proteolipid protein expressed by both oligodendrocytes and myelinating Schwann cells, is localized to compact myelin and associated with glycosphingolipids. Similar to *PMP22*, overexpression of *MAL* results in aberrant myelin formation (25). Finally, *PTP-ε* is a signaling molecule involved in cell cycle exit and cellular differentiation, and *PTP-ε*-deficient animals exhibit peripheral nerve hypomyelination at an early postnatal age (26).

Interestingly, the MGCC also contains genes that map to many of the chromosomal loci linked to inherited neuropathies for which no genetic defect has been determined. For example, although the genetic mutations for CMT1A, CMT1B, and CMT1D have been identified in *PMP22*, *MPZ*, and *Egr2*, respectively, the genetic mutation for CMT1C has not been assigned. Recent linkage analysis, however, maps CMT1C to chromosome 16p13.1-p12.3 (27). Intriguingly, one of the MGCC genes, epithelial membrane protein 2 (*EMP2*), localizes to this chromosomal region. *Emp2* is a member of the *PMP22* family and has demonstrated similar functional properties to *PMP22*, such as regulation of apoptosis (28). Thus, complemented with disease linkage information, our analysis has identified *Emp2* as a positional candidate for CMT1C.

In inspecting the MGCC, we attempted to identify genes pertaining to specific biological pathways. Among the genes in the MGCC were many that encode proteins critical for lipid metabolism. This is consistent with the fact that cholesterol, as well as sphingomyelin and Gal-C, which both contain esterified very long chain fatty acids, are essential constituents of myelin. Despite this fact, with the exception of HMG CoA reductase, which is the rate-limiting enzyme in cholesterol biosynthesis and is coregulated with *MPZ* during development, very little is known about the regulation of other lipid synthetic enzymes in Schwann cells. In our analysis, we found that of the thirteen cholesterol synthetic enzymes represented on the microarray, nine were included in the MGCC (Fig. 3). Additionally, ATP-binding cassette A2 (*ABCA2*), a member of the ABCA family of transporters known to traffic cholesterol and phospholipids (29), as well as oxysterol-binding protein, which traffics cholesterol and sphingomyelin from the Golgi to the plasma membrane (30), were members of the MGCC. Thus, we have discovered that many proteins necessary for lipid metabolism are strictly coregulated during myelination, including a majority of the enzymes involved in cholesterol synthesis.

As a means of rapidly assessing the association of other MGCC genes with myelination, we examined their expression under conditions [i.e., adult (P56) vs. newborn (P0) sciatic nerve and adult sciatic nerve vs. cultured Schwann cells] in which myelin protein levels vary. Unlike myelinating Schwann cells of adult peripheral nerve, cultured Schwann cells and newborn Schwann cells minimally express myelination-associated genes such as *MPZ*, *PMP22*, *MAG*, and *EGR2*. Accordingly, *MPZ* and *NDRG1* expression, in contrast to the expression of nonmyelinating markers, *L1* and *NCAM*, was substantially greater in adult sciatic nerve relative to cultured Schwann cells (Fig. 4a) or P0 sciatic nerve (Fig. 4b). Inspection of other genes and ESTs in the MGCC, including *osteoglycin* and *IL-16*, revealed that they were also expressed at significantly higher levels in adult sciatic nerve relative to cultured Schwann cells (Fig. 4a) or P0 sciatic nerve (Fig. 4b), suggesting that they are expressed in myelinating Schwann cells.

Fig. 3. Multiple genes encoding enzymes required for cholesterol synthesis are identified as genes associated with the myelination process. The cholesterol synthesis pathway, including seventeen enzymes and reaction substrates and products, is shown. Enzymes represented on the chip are denoted with blue font, and those present in the MGCC are designated by an asterisk. A pseudocolor representation of these genes after nerve crush, transection (Txn), or during nerve development (Dev) is shown. Full names and respective probe set numbers for the cholesterol biosynthesis enzymes are listed in Table 3, which is published as supporting information on the PNAS web site.

Identification of a Fatty Acid Desaturase Involved in Myelination. The MGCC included many ESTs whose expression was coregulated with genes encoding critical myelin proteins, suggesting they might also encode proteins important for myelination. To ascertain a molecular function via homology, several of these ESTs were sequenced. Analysis of the full-length sequence of one of the ESTs belonging to Unigene Mm.26171 revealed that it was a fatty acid desaturase (FAD) similar to the yeast VLCFA-synthesizing enzymes, *SUR4* and *FEN1*. Mammalian homologues of *FEN1* and *SUR4* have recently been identified and designated *CIG30*, *SSC1*, and *SSC2* (31). Protein alignment of the full-length mouse and human Mm.26171 protein with *Cig30*, *Ssc1*, and *Ssc2* reveals extensive homology (see Fig. 7, which is published as supporting information on the PNAS web site). For

Fig. 4. Multiple genes identified by AGCA are expressed at higher levels in adult sciatic nerve vs. cultured Schwann cells (a) and adult vs. newborn sciatic nerve (b). mRNA levels of *NDRG1*, *osteoglycin*, *IL-16*, and four different ESTs (denoted by their Unigene ID), identified by AGCA using the MGAs, were examined in sciatic nerve and cultured Schwann cells by quantitative RT-PCR. The fold change represents the ratio of expression in sciatic nerve to cultured Schwann cells.

Fig. 5. The expression pattern of *MASR* indicates that it functions in myelinating Schwann cells and is located within mitochondria. (a) *In situ* hybridization of *MPZ* and *MASR* in sciatic nerve reveals that *MASR* is expressed similarly to *MPZ*, indicating that *MASR* is specifically expressed in myelinating Schwann cells. (b) CV-1 cells transiently transfected with an N-terminal EGFP-fusion of *MASR* reveals that *MASR* colocalizes with MitoTracker staining and is distinct from staining with anti-calnexin, an endoplasmic reticulum marker. The composite overlay of *MASR* and MitoTracker or anti-calnexin shows that *MASR* is localized in mitochondria (yellow), but not in ER.

example, they all contain 5–6 transmembrane domains and share highly conserved motifs, such as KXXEXXDT, FXHXXHH, HXXMYXYY, TXXQXXQ, and the consensus sequence HXXHH, which is frequently found in FADs. Thus, we designate the FAD gene identified in this analysis, *MASR*.

MASR was highly expressed in peripheral nerve, brain, cerebellum, spinal cord, and fetal brain when using RT-PCR analysis, suggesting that it has a critical function in myelin-rich tissues (see Fig. 8a, which is published as supporting information on the PNAS web site). It is preferentially expressed, however, in the peripheral nervous system, because levels of *MASR* mRNA are 70-fold greater in adult sciatic nerve compared with adult brain. *MASR* expression is also 17-fold higher in sciatic nerve compared with cultured Schwann cells (Fig. 4a), and its expression in developing nerve increases 9-fold from P0–P14 (data not shown), indicating that its expression correlates with peripheral nervous system myelination. Furthermore, expression analysis of all four family members by RT-PCR demonstrated that *MASR* is the member of this family preferentially expressed in peripheral nerve (see Fig. 8b). Because VLCFAs are basic components of essential myelin lipids, it is likely that *MASR* is expressed in myelinating Schwann cells. Indeed, *in situ* hybridization for *MASR* in adult mouse sciatic nerve revealed a characteristic speckled expression pattern similar to that observed for *MPZ* (Fig. 5a) and *Egr2* (data not shown), markers of myelinating Schwann cells, further supporting *MASR*'s role in peripheral nerve myelination.

Synthesis of fatty acids beyond C₁₈, which are called VLCFAs,

is catalyzed by FADs in the endoplasmic reticulum (ER) and mitochondria. The COOH-terminal sequence of all four SUR4 family members contains a putative ER-retention signal, leading to the prediction that they are localized to the ER. However, because this signal is weak in *MASR*, we examined its subcellular localization directly. *MASR* was N- or C-terminally tagged with enhanced green fluorescent protein (EGFP), and the resulting fusion protein was expressed in CV-1 cells. EGFP fluorescence colocalized with MitoTracker, a mitochondrial-specific dye, but was distinct from either LysoTracker, an endosome/lysosome-specific dye (data not shown), or immunofluorescence performed against calnexin, an ER-retained protein (Fig. 5b). *MASR*'s localization to the mitochondria is consistent with previous findings that VLCFA synthesis occurs in both mitochondria and the ER. In further support of *MASR*'s activity in fatty acid elongation, Moon *et al.* (32) concurrently identified the *MASR* gene and named it *LCE* (long chain fatty acyl elongase). Through biochemical analyses, they demonstrate the fatty acid elongation activity of *MASR* by observing the incorporation of malonyl-CoA into various fatty acids. Indeed, its expression in sciatic nerve, homology to fatty acid desaturases, biochemical analyses, and subcellular localization indicate that *MASR* has a critical function in the production of myelin. Supporting this hypothesis is the observation that, just as *MASR* expression is lower in the hypomyelinating nerve of *Egr2*^{lo/lo} mice, the expression of *SSCI* is decreased in the central nervous system of myelin-deficient mouse mutants, *quaking* and *jimpy* (31). As such, it is possible that mutations in *MASR* may result in inherited peripheral neuropathy.

Discussion

To describe the molecular components involved in myelination and to identify candidate genes that may be mutated in patients with inherited neuropathy, we have studied the dynamics of peripheral nerve myelination by establishing a compendium of expression profiles. As an alternative to using clustering algorithms, which do not take advantage of the existing wealth of biological information and lack a statistical basis, we have used a biologically based and statistically rigorous approach, AGCA, to identify such candidate genes.

These genes function in diverse pathways, such as lipid metabolism, membrane trafficking, cell adhesion, cell signaling, and cytoskeletal reorganization, which together orchestrate the myelination process by Schwann cells. Concordantly, genes whose mutation is associated with peripheral neuropathies also fall within these various functional pathways. Of particular note, we have been able to identify multiple genes specific not only to myelin synthesis *per se*, but to other biological pathways important for myelination in general. For example, in addition to discovering that numerous genes encoding enzymes for lipid synthesis were strictly coregulated, genes encoding proteins that function in cytoskeletal reorganization were also identified by this algorithm. Cytoskeletal reorganization is indeed critical for myelination because Schwann cells are required to migrate and wrap around axons. We found that a number of cytoskeletal structural components, such as myosin I, myosin VIIa, spectrin, smoothelin, and intermediate filament protein, as well as signaling molecules involved in cytoskeletal reorganization, such as JIP-1, were coregulated with myelination anchor genes. Intriguingly, we also identified utrophin, a cytoskeletal-related protein that binds to F-actin and is part of a Schwann cell-specific dystrophin–dystroglycan complex that includes periaxin (33). Mutations in periaxin disrupt this membrane complex and result in Charcot–Marie–Tooth (CMT) myelinopathies, suggesting that utrophin mutations could also cause this disease.

The complementation of our expression analysis with chromosomal linkage information should be useful in identifying candidate genes for inherited neuropathies as we have demon-

strated with *Emp2*. Another positional candidate for causing neuropathy based on this idea of complementation is *Sirt2*, which is a member of the Sir2 family of NAD-dependent histone deacetylase enzymes thought to regulate gene silencing, aging, and the cell cycle (34), is located at 19q13, and overlaps with the mapping of CMT2B2 at 19q13.3 (35), as well as a severe form of CMT4 at 19q13.1-q13.3 (36). Finally, AF1q, a transmembrane protein found to cause acute leukemia when fused with the protein MLL (mixed lineage leukemia) (37), also overlaps with the mapping of CMT2B1 at 1q21 (38). As disease linkage analysis and genomic sequencing information becomes more refined, the role of global expression profiling complemented with mapping information will become increasingly useful in understanding the molecular pathogenesis of disease.

Discovering genes involved in a specific process through expression analysis is one of the great promises of microarray technology. Toward this end, we applied a statistically rigorous correlation clustering algorithm, AGCA, to our microarray data set of multiple Schwann cell paradigms and discovered MASR, a VLCFA elongation enzyme involved in myelin lipid synthesis. We believe that similar analyses of expression profiles will be increasingly used to associate gene products with particular functional processes.

We thank Mark Watson, William Shannon, and Aditya Phatak for advice on bioinformatics. This work was supported by National Institutes of Health Grant NS4074.

- Warner, L. E., Garcia, C. A. & Lupski, J. R. (1999) *Annu. Rev. Med.* **50**, 263–275.
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868.
- Coller, H. A., Grandori, C., Tamayo, P., Colbert, T., Lander, E. S., Eisenman, R. N. & Golub, T. R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3260–3265.
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitarawan, S., Dmitrovsky, E., Lander, E. S. & Golub, T. R. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2907–2912.
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., *et al.* (1999) *Science* **286**, 531–537.
- Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13790–13795.
- Thomas, J. G., Olson, J. M., Tapscott, S. J. & Zhao, L. P. (2001) *Genome Res.* **11**, 1227–1236.
- Chow, M. L., Moler, E. J. & Mian, I. S. (2001) *Physiol. Genomics* **5**, 99–111.
- Dudoit, S., Yang, Y. H. Y., Callow, M. & Speed, T. (2000) in *University of California, Berkeley, Department of Statistics Technical Report #578* (Stanford Univ. School of Medicine, Univ. of California, Berkeley), pp. 1–38.
- Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. (1999) *Nat. Genet.* **21**, 20–24.
- Li, C. & Wong, W. H. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 31–36.
- Chandross, K. J., Cohen, R. I., Paras, P., Jr., Gravel, M., Braun, P. E. & Hudson, L. D. (1999) *J. Neurosci.* **19**, 759–774.
- Nagarajan, R., Svaren, J., Le, N., Araki, T., Watson, M. & Milbrandt, J. (2001) *Neuron* **30**, 355–368.
- Araki, T., Nagarajan, R. & Milbrandt, J. (2001) *J. Biol. Chem.* **276**, 34131–34141.
- Schneider-Maunoury, S., Topilko, P., Seitanidou, T., Levi, G., Cohen-Tannoudji, M., Pournin, S., Babinet, C. & Charnay, P. (1993) *Cell* **75**, 1199–1214.
- Fu, S. Y. & Gordon, T. (1997) *Mol. Neurobiol.* **14**, 67–116.
- Carroll, S. L., Miller, M. L., Frohnert, P. W., Kim, S. S. & Corbett, J. A. (1997) *J. Neurosci.* **17**, 1642–1659.
- Bruck, W. (1997) *Brain Pathol.* **7**, 741–752.
- Toews, A. D., Barrett, C. & Morell, P. (1998) *J. Neurosci. Res.* **53**, 260–267.
- Wells, M. R., Racis, S. P., Jr., & Vaidya, U. (1992) *J. Neuroimmunol.* **39**, 261–268.
- Dailey, A. T., Avellino, A. M., Benthem, L., Silver, J. & Kliot, M. (1998) *J. Neurosci.* **18**, 6713–6722.
- Boerkoel, C. F., Takashima, H., Stankiewicz, P., Garcia, C. A., Leber, S. M., Rhee-Morris, L. & Lupski, J. R. (2001) *Am. J. Hum. Genet.* **68**, 325–333.
- Garbern, J. Y., Cambi, F., Tang, X. M., Sima, A. A., Vallat, J. M., Bosch, E. P., Lewis, R., Shy, M., Sohi, J., Kraft, G., *et al.* (1997) *Neuron* **19**, 205–218.
- Kalaydjieva, L., Gresham, D., Gooding, R., Heather, L., Baas, F., de Jonge, R., Blechschmidt, K., Angelicheva, D., Chandler, D., Worsley, P., *et al.* (2000) *Am. J. Hum. Genet.* **67**, 47–58.
- Frank, M., Atanasoski, S., Sancho, J. P., Magyar, J. P., Rulicic, T., Schwab, M. E. & Suter, U. (2000) *J. Neurochem.* **75**, 1927–1939.
- Peretz, A., Gil-Henn, H., Sobko, A., Shinder, V., Attali, B. & Elson, A. (2000) *EMBO J.* **19**, 4036–4045.
- Street, V. A., Goldy, J. D., Golden, A. S., Tempel, B. L., Bird, T. D. & Chance, P. F. (2001) *Am. J. Hum. Genet.* **70**, 244–250.
- Wang, C. X., Wadehra, M., Fisk, B. C., Goodglick, L. & Braun, J. (2001) *Blood* **97**, 3890–3895.
- Schmitz, G., Kaminski, W. E. & Orso, E. (2000) *Curr. Opin. Lipidol.* **11**, 493–501.
- Ridgway, N. D. (2000) *Biochim. Biophys. Acta* **1484**, 129–141.
- Tvrdek, P., Westerberg, R., Silve, S., Asadi, A., Jakobsson, A., Cannon, B., Loison, G. & Jacobsson, A. (2000) *J. Cell Biol.* **149**, 707–718.
- Moon, Y. A., Shah, N. A., Mohapatra, S., Warrington, J. A. & Horton, J. D. (2001) *J. Biol. Chem.* **276**, 45358–45366.
- Sherman, D. L., Fabrizio, C., Gillespie, C. S. & Brophy, P. J. (2001) *Neuron* **30**, 677–687.
- Yang, Y. H., Chen, Y. H., Zhang, C. Y., Nimmakayalu, M. A., Ward, D. C. & Weissman, S. (2000) *Genomics* **69**, 355–369.
- Leal, A., Morera, B., Del Valle, G., Heuss, D., Kayser, C., Berghoff, M., Villegas, R., Hernandez, E., Mendez, M., Hennies, H. C., *et al.* (2001) *Am. J. Hum. Genet.* **68**, 269–274.
- Delague, V., Barcil, C., Tuffery, S., Bouvagnet, P., Chouery, E., Koussa, S., Maisonnobe, T., Loiselet, J., Megarbane, A. & Claustres, M. (2000) *Am. J. Hum. Genet.* **67**, 236–243.
- Tse, W., Zhu, W., Chen, H. S. & Cohen, A. (1995) *Blood* **85**, 650–656.
- Bouhouche, A., Benomar, A., Birouk, N., Mularoni, A., Meggouh, F., Tassin, J., Grid, D., Vandenbergh, A., Yahyaoui, M., Chkili, T., *et al.* (1999) *Am. J. Hum. Genet.* **65**, 722–727.