Large-scale informatic analysis to algorithmically identify blood biomarkers of neurological damage

Grant C. O'Connell[a,1], Megan L. Alder[a], Christine G. Smothers[a], and Julia H. C. Chang[a]

[a]School of Nursing, Case Western Reserve University, Cleveland, OH 44106

Edited by Vincent T. Marchesi, Yale University School of Medicine, New Haven, CT, and approved July 9, 2020 (received for review April 23, 2020)

The identification of precision blood biomarkers which can accurately indicate damage to brain tissue could yield molecular diagnostics with the potential to improve how we detect and treat neurological pathologies. However, a majority of candidate blood biomarkers for neurological damage that are studied today are proteins which were arbitrarily proposed several decades before the advent of high-throughput omic techniques, and it is unclear whether they represent the best possible targets relative to the remainder of the human proteome. Here, we leveraged mRNA expression data generated from nearly 12,000 human specimens to algorithmically evaluate over 17,000 protein-coding genes in terms of their potential to produce blood biomarkers for neurological damage based on their expression profiles both across the body and within the brain. The circulating levels of proteins associated with the top-ranked genes were then measured in blood sampled from a diverse cohort of patients diagnosed with a variety of acute and chronic neurological disorders, including ischemic stroke, hemorrhagic stroke, traumatic brain injury, Alzheimer's disease, and multiple sclerosis, and evaluated for their diagnostic performance. Our analysis identifies several previously unexplored candidate blood biomarkers of neurological damage with possible clinical utility, many of which whose presence in blood is likely linked to specific cell-level pathologic processes. Furthermore, our findings also suggest that many frequently cited previously proposed blood biomarkers exhibit expression profiles which could limit their diagnostic efficacy.

molecular diagnostics | stroke | multiple sclerosis | traumatic brain injury | Alzheimer's disease

Collectively, neurological disorders are the leading cause of disability and second leading cause of death worldwide (1). The identification and development of precision blood biomarkers of neurological damage could dramatically improve how we diagnose and treat these debilitating conditions, and ultimately reduce their burden. For example, it is well established that rapid and accurate diagnosis of acute neurological injuries such as stroke and traumatic brain injury during the early stages of care significantly reduces mortality and morbidity (2, 3). However, the symptom-based assessments that are currently used for recognition of such injuries during triage have limited accuracy, and up to 35% of patients are misdiagnosed at initial clinician contact (4–8). In these acute conditions, the development of biomarker-based screening tools with the ability to accurately detect neurological damage could substantially reduce rates of mistriage, enable earlier access to intervention, and improve patient outcomes (9). With respect to chronic neurodegenerative diseases such as Alzheimer's disease and multiple sclerosis, developing accurate blood biomarkers of neurological damage could allow for more confident early diagnosis, noninvasive tracking of disease progression, and real-time monitoring of response to therapy (10, 11).

Due to its specialized function, the proteomic composition of the brain is highly unique relative to other organs. Cellular disruption of neural tissue results in the release of brain-specific proteins into the extracellular environment, and ultimately into peripheral circulation. Thus, the detection of these proteins in the blood can serve as a surrogate marker of neurological damage. From a logical perspective, candidate proteins most ideally suited to serve as such biomarkers are those which display three predominant properties. First, they should exhibit highly enriched expression in brain tissue relative to other tissues, ensuring specificity. Second, they should be highly abundant within brain tissue, as lowly expressed proteins may not be released into circulation at high enough levels to enable detection. Third, they should exhibit ubiquitous expression across all brain regions, reducing the risk of false negative diagnosis in the case of focal damage.

A large number of existing candidate blood biomarkers of neurological damage studied today are proteins which were arbitrarily labeled as being brain specific decades ago (12–19); however, in many cases, their degree of enrichment in brain tissue has been poorly validated, especially in humans. Furthermore, a majority of these proteins were proposed as biomarkers without consideration for brain abundance or expressional variability across brain regions. Additionally, because many of them were suggested before the widespread availability and use of high-throughput omic techniques, they have predominantly been studied in low-throughput investigations only considering a handful of targets. Due to the unsystematic manner in which these existing candidates have been proposed and investigated, it is currently unclear whether they represent the best possible biomarkers relative to the remainder of the human proteome.

Thus, our goal was to systematically search the protein-coding genome to identify genes with the highest potential to produce blood biomarkers of neurological damage. To do this, we leveraged mRNA expression data generated from nearly 12,000 human specimens to algorithmically evaluate over 17,000 protein-coding genes in terms of a novel biomarker suitability score accounting

Significance

The discovery and development of precision blood biomarkers which can accurately detect damage to brain tissue could transform how we diagnose and treat neurological pathologies. In this study, we used mRNA expression data generated from thousands of tissue samples to algorithmically evaluate nearly every protein-coding gene in the human genome in terms of potential to produce blood biomarkers for neurological damage based on expression profiles both across the body and within the brain. This unprecedented analysis identifies a plethora of previously unexplored candidate blood biomarkers which could have clinical utility for noninvasive diagnosis and monitoring of various common neurological conditions, including traumatic brain injury, stroke, and multiple sclerosis.

Author contributions: G.C.O. designed research; G.C.O. secured funding; G.C.O., M.L.A., and C.G.S. performed research; G.C.O. analyzed data; and G.C.O., M.L.A., C.G.S., and J.H.C.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: grant.oconnell@case.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2007719117/-/DCSupplemental.

First published August 6, 2020.

for brain enrichment, brain abundance, and brain regional variability. Then, to determine whether the top-ranked genes identified in our algorithmic analysis could code for proteins with the potential to provide more detailed diagnostic information regarding the specific cellular nature of pathology, we leveraged single-cell sequencing data generated from human brain tissue to determine which cell populations the top-ranked genes are expressed within. Finally, in order to directly evaluate their diagnostic potential, the circulating levels of proteins associated with the top-ranked genes were measured in blood sampled from a diverse cohort of patients diagnosed with a variety of acute and chronic neurological disorders, including ischemic stroke, hemorrhagic stroke, traumatic brain injury, Alzheimer's disease, and multiple sclerosis.

Our collective analysis identifies several previously unexplored candidate blood biomarkers of neurological damage with potential clinical utility, many of which whose presence in blood is likely linked to specific cell-level pathologic processes. Furthermore, our findings also suggest that several of the most frequently cited previously proposed blood biomarkers exhibit expression profiles which could limit their diagnostic performance in many clinical-use scenarios.

Results

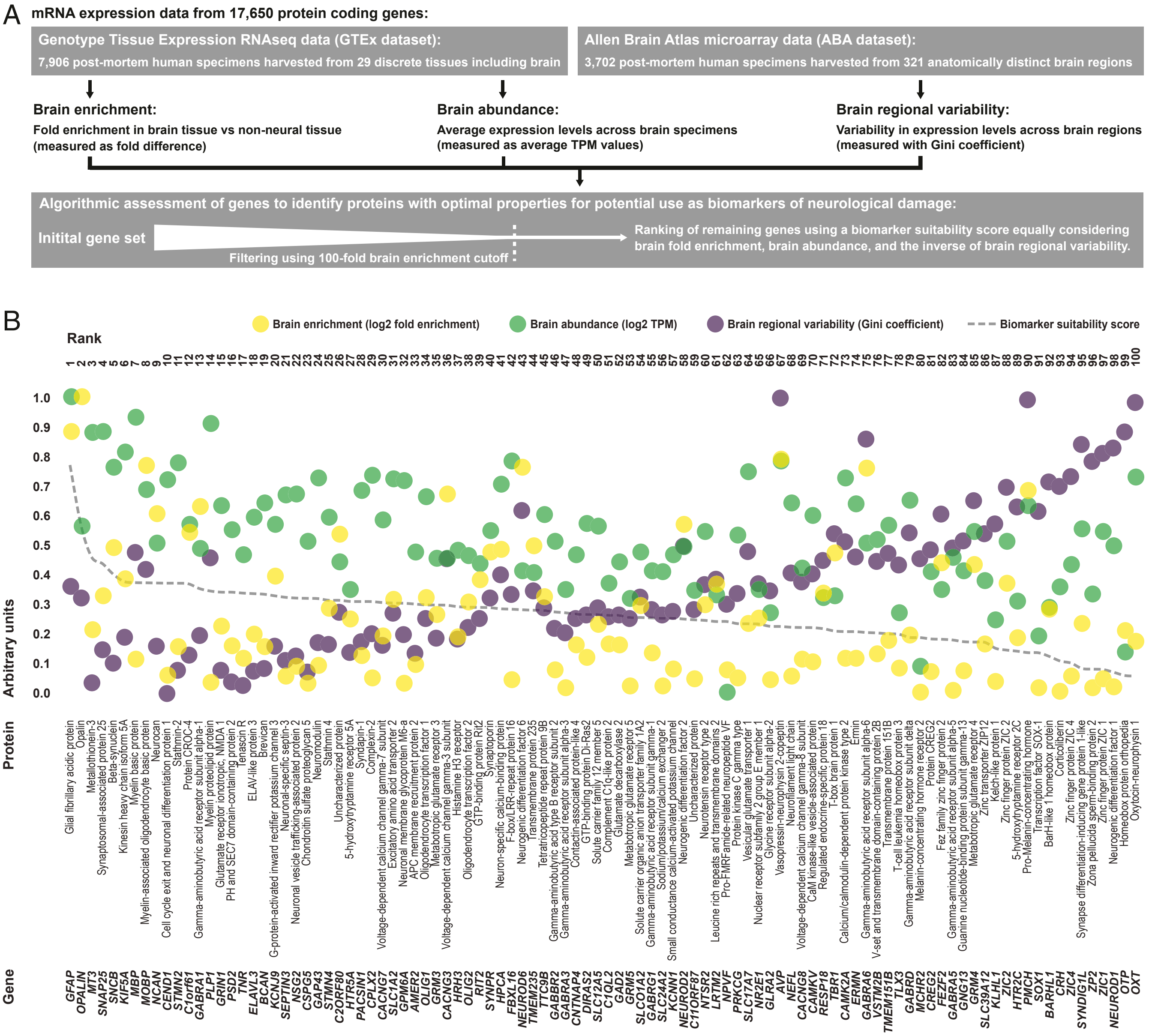
Algorithmic Ranking of Biomarker Suitability. In order to discover genes with optimal expression profiles to produce blood biomarkers of neurological damage, we systematically searched the protein-coding genome to identify genes which exhibit high levels of expressional enrichment in brain tissue relative to nonneural tissues, abundant expression within the brain, and low variability in expression levels across brain regions. To do this, genomewide mRNA expression data were obtained via two publicly available datasets. The first dataset originated from the Genotype-Tissue Expression (GTEx) project (20), and was generated via RNA sequencing of 7,906 postmortem normal human specimens harvested from 28 different nonneural tissues, as well as eight anatomically distinct regions of brain (*SI Appendix*, Tables S1 and S2). The second dataset originated from the Allen Brain Atlas (ABA) project (21), and was generated via microarray analysis of 3,702 normal postmortem normal human brain specimens harvested from 232 anatomically distinct brain regions (*SI Appendix*, Tables S3 and S4). Data were filtered to retain 17,650 genes which were detected in both datasets and informatically annotated as coding for protein products based on the presence of an open reading frame. To determine how specific the expression of each protein-coding gene is to brain tissue, we calculated the level of fold enrichment in brain samples relative to samples from nonneural tissues within the GTEx dataset. In order to determine how abundantly expressed each gene is within the brain, we further calculated the average expression levels across brain samples contained in the GTEx dataset. In order to determine how ubiquitously expressed each gene is across the brain, we used the Gini coefficient, a statistical measure of inequality (22), to assess variability in expression levels across samples from different anatomical brain regions within the ABA dataset. Genes were then filtered to only retain those whose expression levels were enriched at least 100-fold within the brain, and the remaining genes were subsequently ranked for suitability to produce blood biomarkers of neurological damage based on a biomarker suitability score calculated by taking the mean of unity-normalized brain fold enrichment, brain abundance, and inverse brain regional variability values (Fig. 1*A*).

Clustering of samples in both datasets based on the expression levels of all 17,650 protein-coding genes using t-distributed stochastic neighborhood embedding (t-SNE) produced distinct clusters consistent with tissue type and brain region, indicating that the processed data were of high fidelity and properly annotated (*SI Appendix*, Fig. S1). In terms of algorithmic ranking, only 100 genes remained after filtering based on brain fold enrichment cutoff. The relationships between biomarker suitability score, brain fold enrichment, brain abundance, and brain regional variability for these remaining genes are indicated in Fig. 1*B*. The highest ranked genes according to biomarker suitability score generally exhibited a combination of high brain enrichment, high brain abundance, and low regional variability, while lower ranked genes tended to exhibit lower levels of brain enrichment, lower brain abundance, and higher regional variability.

The genes associated with two well-studied candidate biomarkers of neurological damage, glial fibrillary acidic protein (GFAP) (23), and myelin basic protein (MBP) (24–26), ranked in the top of the analysis at first and seventh, respectively. The gene coding for neurofilament light chain (NfL), another previously proposed and increasingly studied neurological damage biomarker (27), ranked 68th in the analysis. However, the remaining 97 of the 100 top-ranked genes all coded for proteins which have been largely unexplored as blood biomarkers to date, suggesting that there is a plethora of proteins with diagnostic potential which have yet to be investigated. For example, the remaining 6 of the top 8 ranked genes, which code for oligodendrocytic myelin paranodal and inner loop protein (OPALIN), metallothionein-3 (MT-3), synaptosomal-associated protein 25 (SNAP-25), beta-synuclein (β-synuclein), kinesin heavy chain isoform 5A (KIF5A), and myelin-associated oligodendrocyte basic protein (MOBP), all displayed optimal expression profiles, but their products have yet to be widely evaluated for use in blood-based diagnostics. Surprisingly, genes associated with some of the most high-profile and frequently cited previously proposed candidate biomarkers of neurological damage (10, 28–30), such as S100 calcium binding protein B (S100B), neuron-specific enolase (NSE), ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCH-L1), alpha-II spectrin (spectrin-αII), Tau, neurofilament heavy chain (NfH), prion protein (PrP), and amyloid beta (Aβ), all failed to meet the fold enrichment cutoff.

Enrichment of the Top-Ranked Genes in Brain Tissue. Fig. 2 depicts the transcriptional expression levels of the top 50 algorithmically ranked genes, along with those of genes associated with several notable previously proposed candidate biomarkers, in each of the 29 tissue types interrogated in the GTEx dataset. Genes associated with three previously proposed biomarkers, GFAP, MBP, and NfL, demonstrated modest to high levels of brain enrichment, exhibiting 1,670-fold, 143-fold, and 119-fold higher expression levels in brain specimens relative to specimens from nonneural tissues. However, genes associated with the remaining previously proposed biomarkers which we examined displayed relatively low levels of enrichment; transcripts associated with S100B, NSE, UCH-L1, spectrin αII, Tau, NfH, PrP, and Aβ were only enriched 3- to 23-fold in brain specimens, suggesting they are not as brain specific as previously thought, and thus may be limited in their diagnostic specificity. Interestingly, some of the most brain-enriched genes were those coding for proteins which have yet to be widely investigated as blood biomarkers. For example, the genes which code for OPALIN, β-synuclein, and MOBP were transcriptionally enriched 2,430-fold, 475-fold, and 1,130-fold, respectively. This suggests that there are several proteins which have yet to be studied as blood biomarkers which may offer equivalent or higher levels of diagnostic specificity relative to those which are currently being investigated.

Abundance of the Top-Ranked Genes in Brain Tissue. Fig. 3 depicts the mean transcriptional expression levels of the top 50 algorithmically ranked genes, along with those of genes associated with several notable previously proposed candidate biomarkers of neurological damage, in brain specimens of the GTEx dataset. Genes associated with all of the previously proposed biomarkers which we examined exhibited expression levels which fell above


A mRNA expression data from 17,650 protein coding genes:

Genotype Tissue Expression RNAseq data (GTEx dataset):
7,906 post-mortem human specimens harvested from 29 discrete tissues including brain

Allen Brain Atlas microarray data (ABA dataset):
3,702 post-mortem human specimens harvested from 321 anatomically distinct brain regions

Brain enrichment:
Fold enrichment in brain tissue vs non-neural tissue
(measured as fold difference)

Brain abundance:
Average expression levels across brain specimens
(measured as average TPM values)

Brain regional variability:
Variability in expression levels across brain regions
(measured with Gini coefficient)

Algorithmic assessment of genes to identify proteins with optimal properties for potential use as biomarkers of neurological damage:

Initital gene set

Filtering using 100-fold brain enrichment cutoff

Ranking of remaining genes using a biomarker suitability score equally considering brain fold enrichment, brain abundance, and the inverse of brain regional variability.

B
● Brain enrichment (log2 fold enrichment)
● Brain abundance (log2 TPM)
● Brain regional variability (Gini coefficient)
Biomarker suitability score
Rank
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
Arbitrary units
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Protein
Glial fibrillary acidic protein
Opalin
Metallothionein-3
Synaptosomal-associated protein 25
Beta-synuclein
Kinesin heavy chain isoform 5A
Myelin basic protein
Myelin-associated oligodendrocyte basic protein
Neurocan
Cell cycle exit and neuronal differentiation protein 1
Stathmin-2
Protein CROC-4
Gamma-aminobutyric acid receptor subunit alpha-1
Myelin proteolipid protein
Glutamate receptor ionotropic, NMDA 1
PH and SEC7 domain-containing protein 2
Tenascin R
ELAV-like protein 3
Brevican
G-protein-activated inward rectifier potassium channel 3
Neuronal-specific septin-3
Neuronal vesicle trafficking-associated protein 2
Chondroitin sulfate proteoglycan 5
Neuromodulin
Stathmin 4
Uncharacterized protein
5-hydroxytryptamine receptor 5A
Syndapin-1
Complexin-2
Voltage-dependent calcium channel gamma-7 subunit
Excitatory amino acid transporter 2
Neuronal membrane glycoprotein M6-a
APC membrane recruitment protein 2
Oligodendrocyte transcription factor 1
Metabotropic glutamate receptor 3
Voltage-dependent calcium channel gamma-3 subunit
Histamine H3 receptor
Oligodendrocyte transcription factor 2
GTP-binding protein Rit2
Synaptoporin
Neuron-specific calcium-binding protein
F-box/LRR-repeat protein 16
Neurogenic differentiation factor 6
Transmembrane protein 235
Tetratricopeptide repeat protein 9B
Gamma-aminobutyric acid type B receptor subunit 2
Gamma-aminobutyric acid receptor subunit alpha-3
Contactin-associated protein-like 4
GTP-binding protein Di-Ras2
Solute carrier family 12 member 5
Complement C1q-like protein 2
Glutamate decarboxylase 2
Metabotropic glutamate receptor 5
Solute carrier organic anion transporter family 1A2
Gamma-aminobutyric acid receptor subunit gamma-1
Sodium/potassium/calcium exchanger 2
Small conductance calcium-activated potassium channel
Neurogenic differentiation factor 2
Uncharacterized protein
Neurotensin receptor type 2
Leucine rich repeats and transmembrane domains 2
Pro-FMRFamide-related neuropeptide VF
Protein kinase C gamma type
Vesicular glutamate transporter 1
Nuclear receptor subfamily 2 group E member 1
Glycine receptor subunit alpha-2
Vasopressin-neurophysin 2-copeptin
Neurofilament light chain
Voltage-dependent calcium channel gamma-8 subunit
CaM kinase-like vesicle-associated protein
Regulated endocrine-specific protein 18
T-box brain protein 1
Calcium/calmodulin-dependent protein kinase type 2
Ermin
Gamma-aminobutyric acid receptor subunit alpha-6
V-set and transmembrane domain-containing protein 2B
Transmembrane protein 151B
T-cell leukemia homeobox protein 3
Gamma-aminobutyric acid receptor subunit delta
Melanin-concentrating hormone receptor 2
Protein CREG2
Fez family zinc finger protein 2
Gamma-aminobutyric acid receptor subunit alpha-5
Guanine nucleotide-binding protein subunit gamma-13
Metabotropic glutamate receptor 4
Zinc transporter ZIP12
Kelch-like protein 1
Zinc finger protein ZIC 2
5-hydroxytryptamine receptor 2C
Pro-Melanin-concentrating hormone
Transcription factor SOX-1
BarH-like 1 homeobox protein
Corticoliberin
Zinc finger protein ZIC 4
Synapse differentiation-inducing gene protein 1-like
Zona pellucida sperm-binding protein 2
Zinc finger protein ZIC 1
Neurogenic differentiation factor 1
Homeobox protein orthopedia
Oxytocin-neurophysin 1
Gene
GFAP
OPALIN
MT3
SNAP25
SNCB
KIF5A
MBP
MOBP
NCAN
CEND1
STMN2
C1orf61
GABRA1
PLP1
GRIN1
PSD2
TNR
ELAVL3
BCAN
KCNJ9
SEPTIN3
NSG2
CSPG5
GAP43
STMN4
C2ORF80
HTR5A
PACSIN1
CPLX2
CACNG7
SLC1A2
GPM6A
AMER2
OLIG1
GRM3
CACNG3
HRH3
OLIG2
RIT2
SYNPR
HPCA
FBXL16
NEUROD6
TMEM235
TTC9B
GABBR2
GABRA3
CNTNAP4
DIRAS2
SLC12A5
C1QL2
GAD2
GRM5
SLCO1A2
GABRG1
SLC24A2
KCNN1
NEUROD2
C11ORF87
NTSR2
LRTM2
NPVF
PRKCG
SLC17A7
NR2E1
GLRA2
AVP
NEFL
CACNG8
CAMKV
RESP18
TBR1
CAMK2A
ERMN
GABRA6
VSTM2B
TMEM151B
TLX3
GABRD
MCHR2
CREG2
FEZF2
GABRA5
GNG13
GRM4
SLC39A12
KLHL1
ZIC2
HTR2C
PMCH
SOX1
BARHL1
CRH
ZIC4
SYNDIG1L
ZP2
ZIC1
NEUROD1
OTP
OXT

Fig. 1. Algorithmic search strategy and resultant top-ranked candidate genes according to biomarker suitability score. (*A*) Experimental workflow used to algorithmically assess the protein-coding genome for candidate biomarkers of neurological damage. (*B*) Relationship between biomarker suitability score, brain fold enrichment, brain abundance, and brain regional variability for the top-100 ranked genes. Brain fold enrichment, brain abundance, and brain regional variability values are presented scaled between 0 and 1 using unity normalization.

the 90th percentile relative to all 17,650 protein-coding genes included in our analysis, suggesting high levels of abundance in brain tissue. Comparatively, many highly ranked genes coding for proteins that have yet to be widely considered as blood biomarkers exhibited similar or higher expression levels. For example, the genes which code for MT-3, SNAP-25, β-synuclein, and KIF5A all displayed expression levels which fell above the 99th percentile. From an analytical perspective, the fact that these genes exhibit such high transcriptional abundance in brain tissue is encouraging in terms of the possibility that their protein products could be detected in blood in the case of neurological damage.

Variability in Expression Levels of the Top-Ranked Genes across Brain Regions. Fig. 4 depicts the transcriptional expression levels of the top 50 algorithmically ranked genes, along with those of genes associated with other notable previously proposed candidate biomarkers, across specimens from each of the 232 brain regions examined in the ABA dataset, with the degree of regional variability indicated by Gini coefficient. Gini coefficient values ranged from 0.06 to 1.0 across all 17,650 protein-coding genes examined in the analysis, with values closer to 0 indicating nearly homogeneous expression across brain regions, and values approaching 1 indicating extreme inequality in expression across brain regions. Many genes associated with previously proposed biomarkers exhibited relatively low expressional variability, indicated by Gini coefficient values ranging from 0.11 to 0.36. However, a handful, including GFAP, MBP, NfL, and NfH, were more variable, with respective measured Gini values of 0.43, 0.53, 0.47, and 0.60. Comparatively, many top ranked genes associated with proteins that have yet to be widely investigated as blood biomarkers exhibited an equivalent or lower degree of expressional variability. For example, the genes which code for OPALIN, MT-3, SNAP-25, β-synuclein, and KIF5A

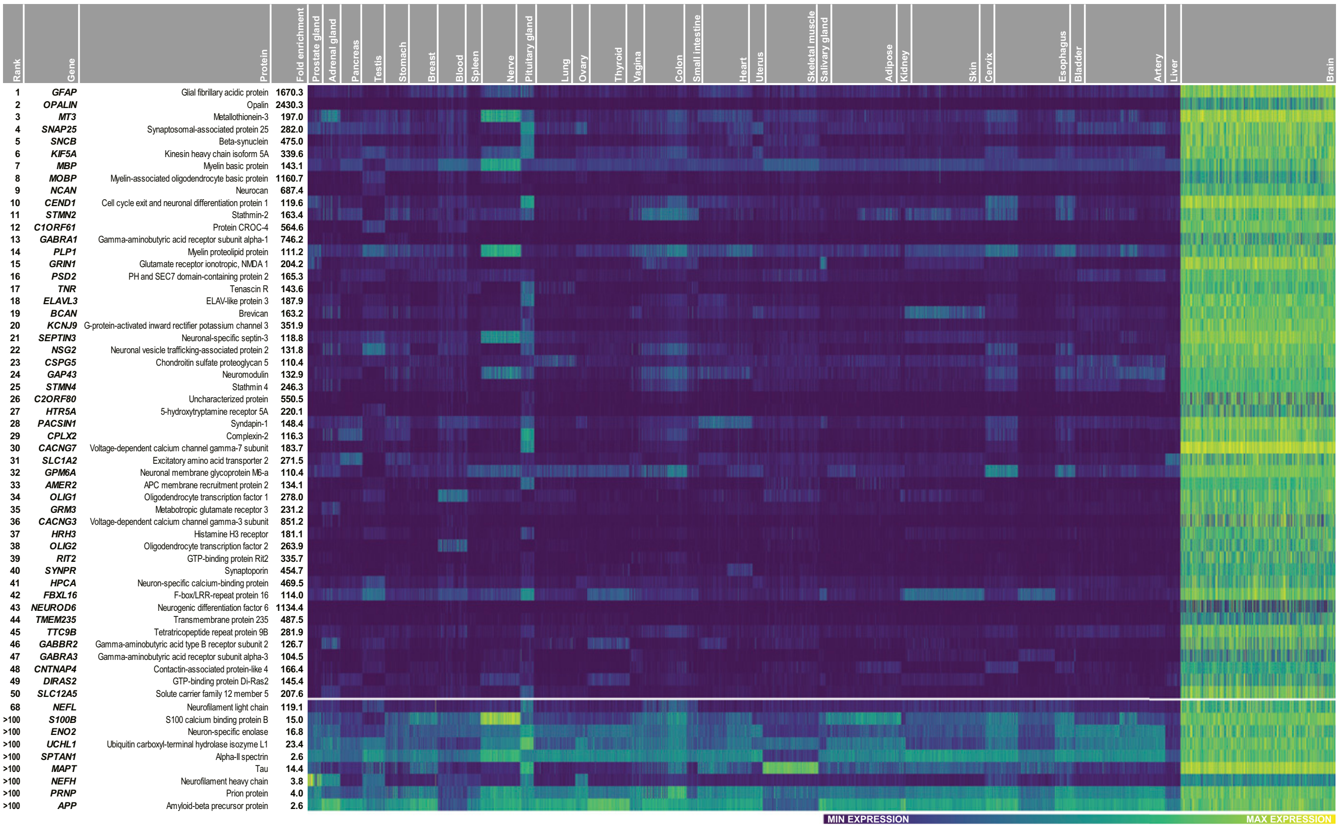


Fig. 2. Transcriptional enrichment of the top-ranked candidate genes in brain tissue. Transcriptional expression levels of the top-50 algorithmically ranked genes, along with those of genes associated with several notable previously proposed candidate biomarkers of neurological damage, in each of the 29 tissue types interrogated in the GTEx dataset. Levels of fold enrichment in brain tissue relative to nonneural tissues are indicated.

displayed respective Gini coefficient values of 0.40, 0.16, 0.25, 0.21, and 0.29. This suggests that there are several yet to be explored candidate blood biomarkers which may exhibit enough homogeneity in brain expression to allow for detection of neurological damage across a variety of brain regions with relatively equivalent levels of diagnostic sensitivity.

Cellular and Subcellular Expression Profiles of Top-Ranked Genes. In an effort to provide biological context relevant to the use of their protein products as biomarkers, we examined the cellular and subcellular expression profiles of the top-50 ranked genes, as well as genes associated with other notable previously proposed candidate markers of neurological damage.

In order to determine which cell populations the candidate genes are expressed by within the brain, we leveraged a publicly available single-cell RNA-sequencing dataset originally generated by Darmanis et al. (31) to examine their transcriptional expression levels in five distinct types of cells isolated from surgically resected brain tissue harvested from a mix of living adult human donors (*SI Appendix*, Tables S5 and S6). A majority of the top-ranked genes exhibited relatively cell-specific expression profiles (Fig. 5*A*). For example, the genes coding for GFAP and MT-3 were predominantly expressed by astrocytes, while the genes coding for SNAP-25, β-synuclein, and KIF5A were predominantly expressed by neurons. Likewise, the genes coding for OPALIN, MBP, and MOBP were predominantly expressed by oligodendrocytes. This suggests that it is possible that the presence of any of these given proteins in the blood could be diagnostically attributable to damage to a specific brain cell population, which could provide pathophysiological context that adds to their clinical value as biomarkers. Genes associated with multiple notable previously proposed biomarkers including S100B, PrP, and spectrin-αII exhibited less isolated cellular expression profiles, suggesting they may not be able to provide similar contextual value.

In order to determine the subcellular distribution of proteins associated with the candidate genes, we utilized the Compartments database developed by Binder et al. (32) to retrieve subcellular localization confidence scores generated from an aggregate of sequence-based prediction, literature text-mining, and high-throughput microscopy-based screening. A majority of the top-ranked genes generated in our analysis code for proteins associated with the plasma membrane, cytosol, or cytoskeleton, and few coded for nuclear or secreted proteins (Fig. 5*B*). This is encouraging in terms of their use as biomarkers, as secreted proteins are more likely to be found in the blood in the absence of cellular damage, limiting diagnostic specificity, and proteins with predominantly nuclear localization may not easily diffuse from damaged cells into the extracellular environment, limiting their ability to be detected. Furthermore, in the case of some proteins associated with top-ranked genes, their highly specified subcellular localization, when considered along with their cellular expression profiles, could allow them to provide even more detailed pathophysiological context as biomarkers. For example, KIF5A is a cytoskeletal microtubule motor protein which we observed as being primarily expressed by neurons; thus it is possible that the presence of KIF5A in the blood may be associated with axonal damage specifically.

Circulating Levels of Top-Ranked Candidate Biomarkers in Patients with Neurological Disorders. In order to directly evaluate their potential for use as blood biomarkers, ELISA was used to measure the serum levels of proteins associated with the top-eight ranked

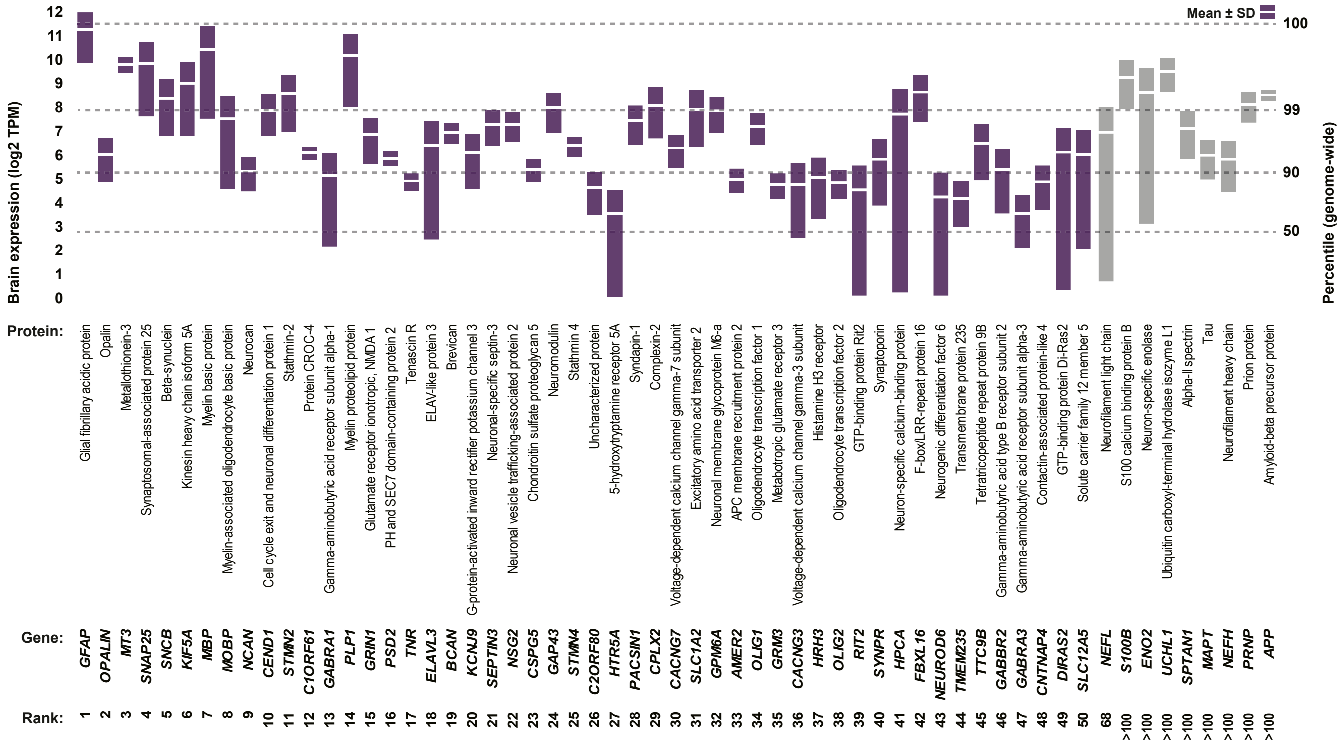


Fig. 3. Transcriptional abundance of top-ranked candidate genes in brain tissue. Average transcriptional expression levels of the top-50 algorithmically ranked genes, along with those of genes associated with several notable previously proposed candidate biomarkers of neurological damage, in brain specimens of the GTEx dataset.

candidate genes in a diverse cohort of patients with acute and chronic neurological damage, as well as in a group of neurologically normal controls. Subjects with neurological damage included patients with definitive clinical diagnoses of traumatic brain injury (TBI, $n = 13$), ischemic stroke (IS, $n = 43$), hemorrhagic stroke (HS, $n = 5$), Alzheimer's disease (AD, $n = 20$), and multiple sclerosis (MS, $n = 20$). In cases of acute neurological damage such as traumatic brain injury and stroke, blood was collected immediately upon hospital admission. Neurologically normal controls ($n = 85$) included patients coenrolled in various other investigations of chronic disease who were deemed neurologically normal by a trained clinician. Patients with neurological damage and control subjects were relatively well matched in terms of gender and the prevalence of common comorbidities such as diabetes, dyslipidemia, and hypertension. The collective group of patients with neurological damage were also of similar age as control subjects, although ischemic stroke and Alzheimer's disease patients tended to be older, while multiple sclerosis patients tended to be younger (*SI Appendix*, Table S7). However, we observed no significant correlations between the serum levels of any of the eight proteins and age after controlling for clinical diagnosis (*SI Appendix*, Table S8); this suggests that age was a relatively small contributor to the overall variance in the serum levels of these proteins in our study population, and that any modest intergroup age differences were unlikely to meaningfully confound downstream analyses.

The levels of all eight proteins were significantly elevated in the serum of patients with neurological damage relative to neurologically normal controls, providing evidence that they are in fact released into circulation as a result of damage to brain tissue. However, some elevations were condition specific, while others were more ubiquitous. For example, circulating levels of GFAP, KIF5A, and SNAP-25, were significantly elevated across nearly all neurological conditions (Fig. 6 *A*, *D*, and *F*), suggesting their presence in blood may serve as a general marker of neurological damage. Conversely, circulating levels of MT-3 were only significantly elevated in patients with acute ischemic stroke (Fig. 6*C*); this observation, taken with prior reports that MT-3 is robustly up-regulated in astrocytes and neurons in response to ischemic and hypoxic conditions (33–35), suggests MT-3 may have utility for detecting ischemia specifically. Likewise, elevations in circulating levels of MBP, MOBP, and OPALIN were most dramatic in patients with multiple sclerosis compared to patients with other neurological conditions (Fig. 6 *B*, *G*, and *H*), suggesting that they may have specific utility for detecting demyelinating pathologies.

Correlational analysis further supported the notion that the presence of these proteins in the blood could be linked to specific cellular pathology. The serum levels of the eight proteins all exhibited varying degrees of positive correlations with each other across all subjects after controlling for clinical diagnosis; however, hierarchal clustering of the proteins based on covariance produced perfect clustering according to the specific cell populations their transcripts were predominantly expressed by in our earlier single-cell analysis. For example, MBP, OPALIN, and MOBP, which displayed predominant transcriptional expression by oligodendrocytes, distinctly clustered with each other based on the correlation between their circulating levels. Similarly, SNAP-25, β-synuclein, and KIF5A, which displayed predominant transcriptional expression by neurons, formed a second distinct cluster, while MT-3 and GFAP, which displayed predominant transcriptional expression by astrocytes, formed a third distinct cluster (Fig. 6*I*). This overall pattern of correlation, taken with patterns of blood elevations we observed across different neurological conditions, provides evidence that presence of these proteins in the blood may be attributable to specific pathophysiological processes affecting specific populations of cells. In particular, this may suggest that increased blood levels of MBP, MOBP, and OPALIN are a result of myelin damage, that increased blood levels

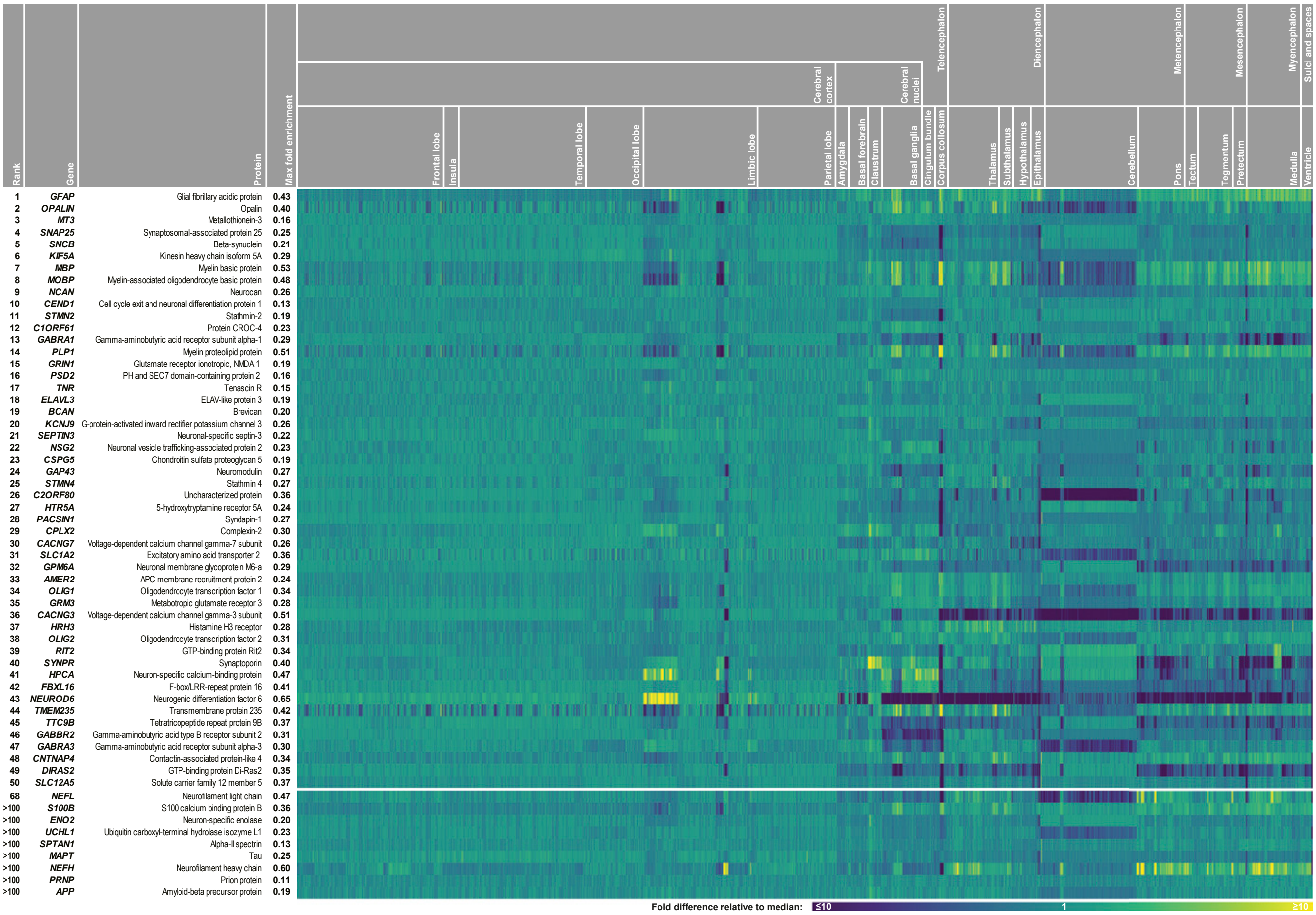


Fig. 4. Variability in expression levels of the top-ranked candidate genes across brain regions. Transcriptional expression levels of the top-50 algorithmically ranked genes, along with those of genes associated with other notable previously proposed candidate biomarkers of neurological damage, across samples from each of the 232 brain regions examined in the ABA dataset. The degree of variability in expression levels across brain regions is indicated by the Gini coefficient.

of SNAP-25, β-synuclein, and KIF5A may be attributable to general neuron or axonal damage, and that increased blood levels of MT-3 and GFAP may be attributable to astrocyte damage or activation.

Diagnostic Performance. Receiver operating characteristic (ROC) analysis was subsequently used to determine how well the circulating levels of proteins associated with the top-eight ranked genes could diagnostically discriminate between patients with neurological damage and control subjects. In terms of differentiating the total pool of patients with neurological damage from control subjects, all eight proteins demonstrated modest diagnostic ability, producing area under curve (AUC) values ranging from 0.68 to 0.80, with SNAP-25, GFAP, and KIF5A offering the highest overall diagnostic performance, and MBP, β-synuclein, and MT-3 offering the lowest. When used in combination, the coordinate blood levels of all eight proteins yielded greater overall diagnostic performance than any individual marker alone, producing an AUC of 0.94, and could discriminate between groups with 84% sensitivity and 89% specificity (Fig. 7*A*). While this suggests that all eight proteins have utility for detection of general neurological damage, the levels of diagnostic performance we observed when considering the total subject pool are likely not reflective of their true diagnostic potential. Each of the neurological conditions we examined are a result of different pathophysiological processes; accordingly, some conditions were better diagnosed than others, and several proteins demonstrated condition-specific discriminatory ability consistent with their patterns of blood elevations.

For example, separate comparisons of each individual neurological condition to the control group revealed that MOBP, OPALIN, and MBP exhibited relatively poor performance for diagnosis of acute pathology. However, consistent with their potential link to oligodendrocyte damage, they were the highest three performing individual markers in diagnosis of multiple sclerosis; their use in combination produced an AUC of 0.92, and allowed for discrimination between groups with 90% sensitivity and 87% specificity (Fig. 7*F*). While they performed fairly well across all conditions, three markers potentially linked to astrocyte activation and neuron damage, SNAP-25, GFAP, and KIF5A, exhibited higher levels of performance in diagnosis of patients with acute pathology as opposed to patients with chronic pathology. They were the highest three performing individual markers for diagnosis of traumatic brain injury, where their use in combination produced an AUC of 0.98, and allowed for discrimination between groups with 100% sensitivity and 92% specificity (Fig. 7*B*). They were also the highest three individual performing markers with respect to diagnosis of hemorrhagic stroke, where their use in combination yielded an AUC of 0.99, and allowed for discrimination between groups with 100% sensitivity and 99% specificity (Fig. 7*D*). Unsurprisingly given its potential link to ischemic injury, MT-3 offered the highest individual level of diagnostic performance for detection of ischemic stroke; when used in

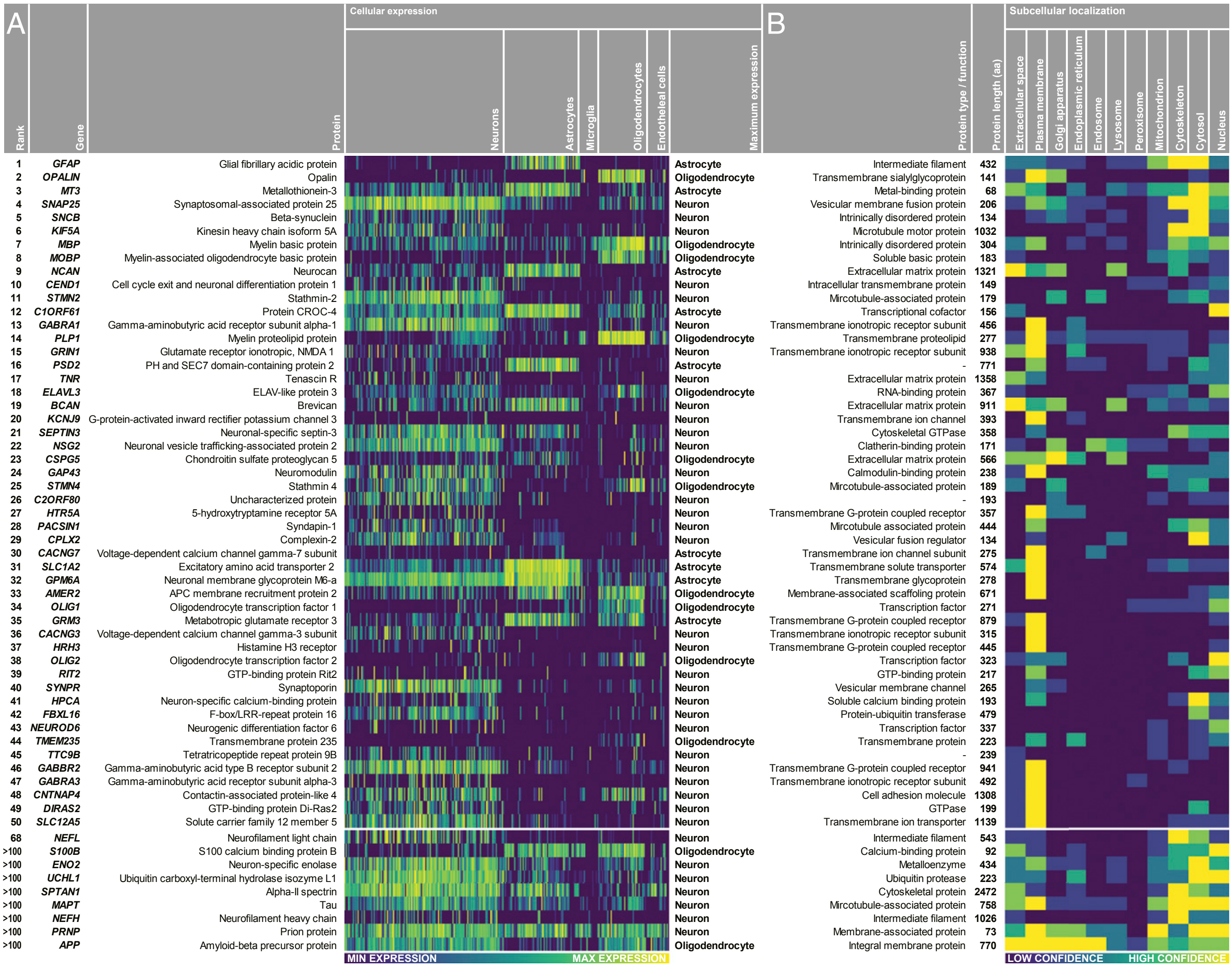


Fig. 5. Cellular and subcellular expression profiles of the top-ranked candidate genes. (*A*) Transcriptional expression levels of the top-50 algorithmically ranked genes, along with those of genes associated with other notable previously proposed candidate biomarkers of neurological damage, across various distinct human brain cell populations profiled by single-cell sequencing. The cell population with the highest average expression levels for each gene is indicated. (*B*) Descriptions of proteins coded for by the top-50 algorithmically ranked genes, as well as genes associated with other notable previously proposed candidate biomarkers of neurological damage, along with their predicted subcellular localizations. Predicated protein subcellular localizations are indicated as confidence scores; higher values indicate a greater degree of confidence a given protein exhibits a given localization.

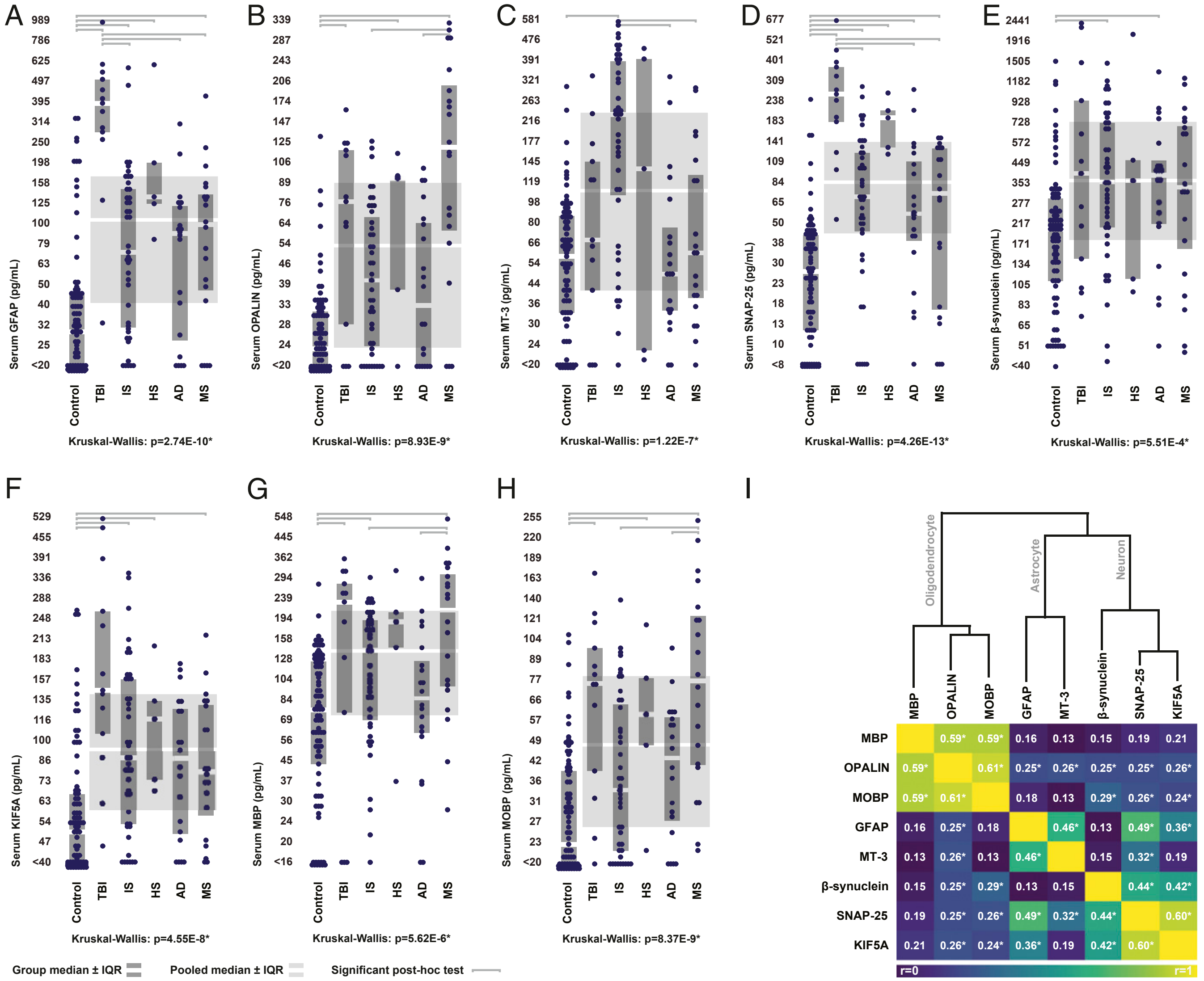
combination with the next two highest performing markers, SNAP-25 and KIF5A, the three markers produced an AUC of 0.90, and could discriminate between groups with 72.0% sensitivity and 93.0% specificity (Fig. 7*C*).

Further supportive of the idea that the presence of these proteins in the blood may be attributed to different pathophysiological mechanisms, in many cases, they displayed the ability to diagnostically discriminate between different neurological conditions with fairly high levels of accuracy, especially when used in combination (*SI Appendix*, Fig. S2). Taken together, the collective levels of diagnostic performance observed across these analyses suggest that the top-ranked candidate biomarkers identified using our algorithmic search strategy could have true clinical utility for detection of neurological damage across a multitude of common pathologies.

Discussion

In the work described here, we employed an algorithmic approach to systematically evaluate the protein-coding genome to identify genes with the highest potential to produce blood biomarkers of neurological damage. The results produced by our analysis, which is unprecedented in both its scale and scope, provide valuable insights into the diagnostic suitability of numerous previously proposed candidate biomarkers, and identify several previously unexplored candidate biomarkers with strong potential for future clinical use.

Three previously proposed candidate biomarkers, GFAP, MBP, and NfL, fell near the top of our algorithmic analysis, ranking at 1st, 7th, and 68th, respectively. These results strongly support those of numerous prior investigations which have reported elevations in their circulating levels in various states of neurological damage and provide further evidence that they may have true diagnostic utility (23–27). Unlike GFAP, MBP, and NfL, the remaining previously proposed biomarkers which were interrogated in our algorithmic analysis, including S100B, NSE, UCH-L1, spectrin-αII, Tau, NfH, PrP, and Aβ, all ranked outside of the top 100. This phenomenon was largely attributable to the fact that they displayed a shockingly low degree of brain enrichment at the transcriptional level, which was unexpected given that many of them are frequently cited as being brain-



	MBP	OPALIN	MOBP	GFAP	MT-3	β-synuclein	SNAP-25	KIF5A
MBP		0.59*	0.59*	0.16	0.13	0.15	0.19	0.21
OPALIN	0.59*		0.61*	0.25*	0.26*	0.25*	0.25*	0.26*
MOBP	0.59*	0.61*		0.18	0.13	0.29*	0.26*	0.24*
GFAP	0.16	0.25*	0.18		0.46*	0.13	0.49*	0.36*
MT-3	0.13	0.26*	0.13	0.46*		0.15	0.32*	0.19
β-synuclein	0.15	0.25*	0.29*	0.13	0.15		0.44*	0.42*
SNAP-25	0.19	0.25*	0.26*	0.49*	0.32*	0.44*		0.60*
KIF5A	0.21	0.26*	0.24*	0.36*	0.19	0.42*	0.60*	

Fig. 6. Circulating levels of proteins associated with the top-eight ranked candidate genes in patients with neurological disease. (*A*–*H*) Circulating concentrations of proteins associated with each of the top-eight ranked candidate genes in neurologically normal controls and patients diagnosed with traumatic brain injury (TBI), ischemic stroke (IS), hemorrhagic stroke (HS), Alzheimer's disease (AD), and multiple sclerosis (MS). Data are presented on a log2 scale, with the minimum axis value corresponding to the lower limit of quantification. Protein concentrations were compared between groups with the Kruskal–Wallis test. Post hoc comparisons were made using the Mann–Whitney *u* test, with *P* values adjusted for multiple comparisons using the Holm–Bonferroni method. (*I*) Correlation matrix depicting the relationships between the circulating levels of proteins associated with the top-eight ranked genes across all subjects after controlling for diagnosis. Strength of partial correlations were assessed via Spearman's rho, and *P* values were adjusted for multiple comparisons using the Holm–Bonferroni method. Hierarchical clustering indicates similarity in expression based on correlation; the predominant brain cell population of expression associated with each major cluster of proteins is indicated in the cluster dendrogram. *Statistically significant.

specific proteins (29, 30, 36). Because our findings suggest that their expression is poorly restricted to brain, they may be limited in their ability to accurately detect damage to brain tissue in the presence of concurrent damage to peripheral tissues and organs, which is often a feature in many neurological conditions. For example, in stroke, common comorbidities such as diabetes and atherosclerosis cause cellular damage to the vasculature and kidneys (37). In Alzheimer's disease, a multitude of common age-related maladies and general atrophy cause similar cellular damage across the body (38). Perhaps most obviously, traumatic brain injury is often a result of collision sports, vehicular accidents, falls, and other types of events which commonly result in numerous peripheral injuries (39). Thus, generally, it is doubtful that blood measures of any of these proteins would be able to offer high levels of diagnostic performance in many true clinical-use scenarios.

This is likely why S100B, NSE, and UCH-L1 have performed poorly as blood biomarkers in several conditions, particularly ischemic stroke (40, 41) and multiple sclerosis (42, 43), when trialed under clinically applicable study designs. From this perspective, it is particularly interesting that the US Food and Drug Administration (FDA) recently approved a dual-analyte assay measuring circulating levels of UCH-L1 in combination with GFAP for clinical detection of traumatic brain injury (44). Based on our observations, it is possible that a majority of the discriminatory power of this assay is attributable to GFAP alone. If this is true, it would mean that a single-analyte assay targeting only GFAP may be able to achieve similar or equivalent performance, while being more cost effective and easier to implement at the point of care.

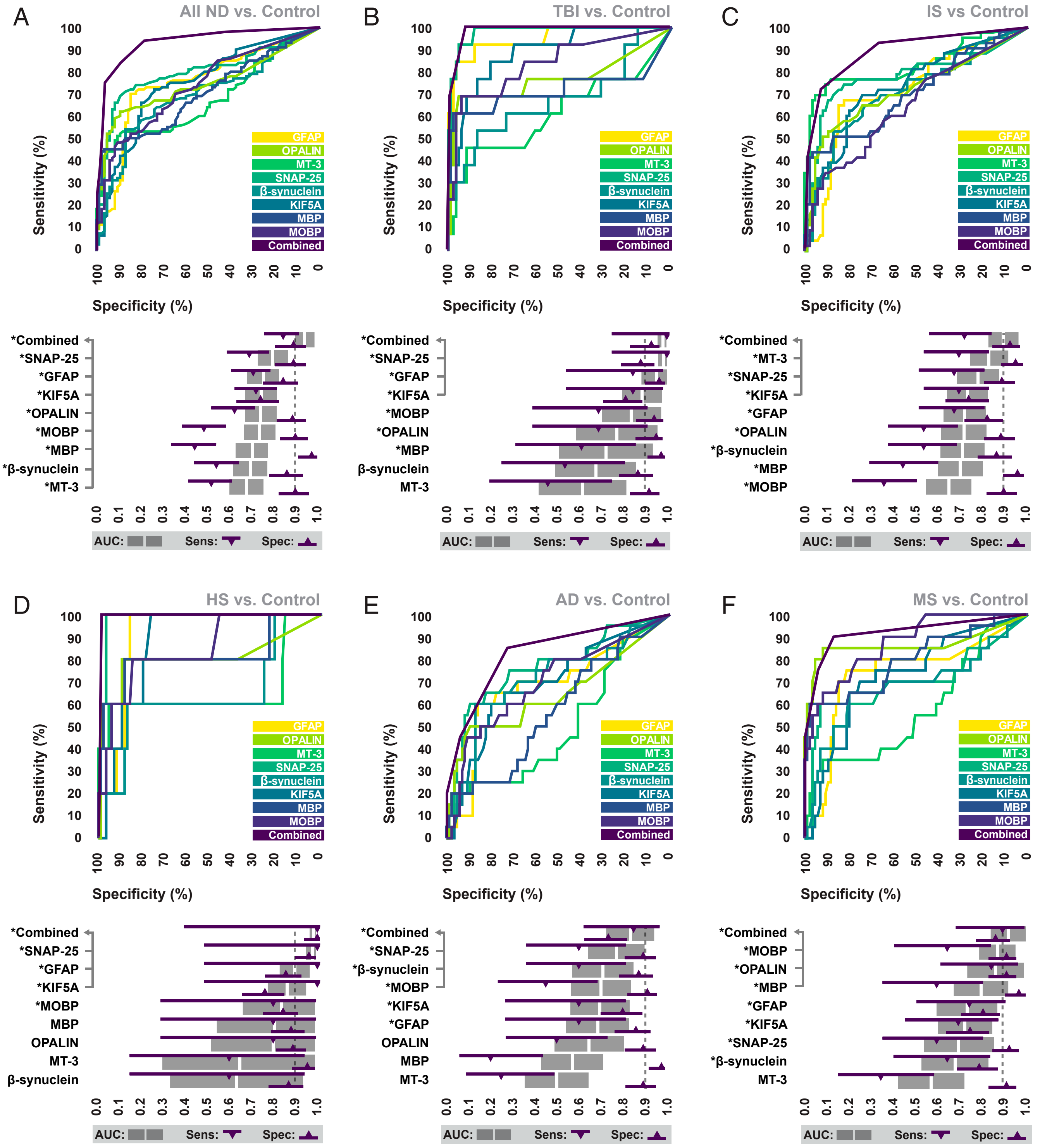


Fig. 7. Diagnostic performance of proteins associated with the top-eight ranked candidate genes. (*A*–*F*) ROC curves indicating the individual and combined abilities of circulating levels of proteins associated with the top-eight ranked genes to discriminate between neurologically normal controls and the total pool of patients diagnosed with neurological damage (All ND), patients diagnosed with traumatic brain injury (TBI), ischemic stroke (IS), hemorrhagic stroke (HS), Alzheimer's disease (AD), and multiple sclerosis (MS). AUC, sensitivity, and specificity values with 95% confidence intervals are indicated. Sensitivity and specificity values are associated with the cutoff yielding the highest Youden value. For the comparison between control subjects and the total pool of patients with neurological damage, the combined ROC curve was generated using the circulating levels of all eight proteins. For comparisons between control subjects and specific neurological diagnoses, combined ROC curves were generated using the circulating levels of only the top three individually performing proteins, as determined by AUC. The statistical significance of AUC values was tested using the DeLong method, and *P* values were adjusted for multiple comparisons using the Holm–Bonferroni method. *Statistically significant AUC. A complete summary of all diagnostic statistics with optimal cutoff values can be found in *SI Appendix*, Table S9.

It is important to note that with respect to some of the previously proposed biomarkers assessed in our analysis, the lack of brain specificity we observed may not necessarily preclude them from being informative in certain neurological conditions, especially if they are directly involved in pathogenesis. For example, because Alzheimer's disease is pathophysiologically associated with an accumulation of neurofibrillary tangles and plaques comprised of tau and Aβ, measures of these proteins in the blood, particularly disease-related isoforms, may still offer diagnostic utility. However, even in this case, a lack of brain specificity may still limit the degree of accuracy these markers can truly achieve when measured in peripheral circulation; this could explain the fact that measures of tau and Aβ in cerebrospinal fluid have historically performed better than measures in blood for detection of Alzheimer's pathology (45).

In addition to providing insights regarding the aforementioned previously proposed biomarkers, our algorithmic analysis also identified several previously unexplored candidates, as proteins associated with 97 of the top-100 ranked genes have yet to be widely investigated for blood-based diagnostic use. While we only focused on the highest six ranked of these markers in our confirmatory blood analysis, many others displayed body- and brain-wide expression profiles which suggest that they may have similar diagnostic utility. If they can be validated, this large pool of newly identified candidate markers could be extremely valuable, given that it is becoming increasingly evident that single blood markers are inadequate to provide high levels of accuracy for many neuro applications, and that the development of true precision diagnostics will likely require the use of multiple markers as part of multianalyte algorithmic assays (46–49).

Many of the top-ranked previously uninvestigated markers which we assessed in our blood analysis displayed patterns of elevations and levels of diagnostic performance which were encouraging in terms of future clinical use. For example, levels of OPALIN and MOBP were robustly elevated in patients with multiple sclerosis and appeared to indicate cellular damage to oligodendrocytes specifically. When combined with MBP, which is an often studied cerebral spinal fluid biomarker of multiple sclerosis-related neurological damage (50–52), they were able to discriminate between multiple sclerosis patients and controls with 90% sensitivity and 87% specificity, albeit in a limited sample size. Thus, diagnostics targeting circulating levels of these proteins could be useful in the initial diagnosis of multiple sclerosis, which is often subjective and ambiguous (53). However, these proteins may have the most future value for noninvasive longitudinal tracking of disease progression. Most molecular biomarkers which are clinically used in multiple sclerosis, such as oligoclonal band screening and autoantibody tests, detect immune system activity and are therefore indirect indicators of neurodegeneration (11). Because the presence of OPALIN, MOBP, and MBP in the blood appear to be directly linked to brain tissue damage, they could be useful for evaluating efficacy of therapeutic interventions, which is essential in developing and maintaining long-term individualized treatment strategies.

Two previously unexplored blood markers identified in our analysis, SNAP-25 and KIF5A, appeared to be linked to cellular damage to neurons specifically, and displayed a strong ability to detect acute neurological conditions, especially traumatic brain injury. In fact, when combined with GFAP, they were able to discriminate between traumatic brain injury patients and controls with 100% sensitivity and 92% specificity. Thus, these markers could be extremely valuable in screening patients for traumatic brain injury during triage, where timely recognition can avoid debilitating complications by ensuring patients are referred to an appropriate care team which is trained to manage neurological injury (3). Given that the symptom-based assessments predominantly used by clinicians for recognition of traumatic brain injury in the prehospital and early in-hospital setting have been reported to be as low as 30% sensitive (4, 5), and the only biomarker-based screening tool with FDA approval, which targets GFAP and UCH-L1, has been reported to be only about 35% specific (44), KIF5A and SNAP-25 could give much needed additional diagnostic power to current clinical tools if our findings can be validated in a larger cohort of patients.

Another relatively unexplored blood marker identified in our analysis, MT-3, displayed properties which suggest it could be similarly valuable for detection of ischemic stroke during triage. Much like traumatic brain injury, early recognition is essential for positive outcome, as it is estimated that 1.9 million neurons are permanently lost every minute without intervention (54). However, the ability to confidently rule out a stroke diagnosis in patients presenting with neurological symptoms is also important, as mistriaged patients with nonstroke disorders can put significant resource strain on stroke centers (55, 56). We observed robust elevation in circulating level of MT-3 in ischemic stroke patients at hospital admission, and MT-3 was able to discriminate between ischemic stroke patients and controls with 70% sensitivity and 95% specificity. Given that the most commonly used symptom-based stroke recognition tools available to clinicians in the prehospital and early in-hospital setting are often unreliable, with levels of sensitivity ranging from 44 to 95%, and levels of specificity ranging from 21 to 78% (57), MT-3 could be a strong candidate for future use in biomarker-based stroke screening if our findings can be confirmed in a larger-scale follow-up investigation.

While our results are exciting, it is important to note that this study was not without limitations. Perhaps most notable is that our algorithmic analysis was performed at the transcriptional level, even though our end goal was to identify protein biomarkers. While the correlation between mRNA and protein expression is far from perfect, several studies have concluded that cellular mRNA levels are the primary determinant of protein levels, and that the former is a strong predictor of the latter (58, 59). Furthermore, the fact that a well-studied and established protein biomarker ranked first in our analysis, and that all of the top-ranked markers were validated at the protein level in our serum assays, suggests that our approach was robust despite this potential limitation. Another potential limitation lies in that we used conventional ELISA techniques to assay the concentrations of the top-eight ranked candidate markers in our serum analysis; due to the limited analytical sensitivity associated with conventional ELISA, in many instances, analyte levels in several samples fell below lower limits of quantitation or were undetected. While this did not hinder our ability to detect the presence of intergroup differences, it could have negatively impacted diagnostic performance. Thus, analysis of these markers in future work using more sensitive immunoassay techniques such as digital ELISA may reveal even higher levels of diagnostic performance than those which we have reported here (60, 61). Furthermore, because the capacity of digital ELISA to support multiplexing is increasing (62), it possible that the future development of a multiplexed digital ELISA simultaneously targeting some or all of these proteins as a neuro panel could offer both high levels of analytical performance and convenience.

It is also worthwhile to note that the general biomarker discovery strategy employed in our analysis could be modified to identify additional neurological disease biomarkers. The primary goal of our analysis was to identify candidate markers of general neurological damage; thus, we searched for brain-enriched proteins which exhibit ubiquitous expression across brain regions. However, our analysis could be modified to search for brain-enriched proteins with region-specific expression profiles whose presence in the blood could indicate the precise location of pathology. Furthermore, our analysis could also be modified to identify other brain-enriched molecular species which could serve similar diagnostic function as the candidate proteins described here. For example, similar algorithmic analysis of publicly available micro-RNA (miRNA) datasets could be used to identify brain-enriched miRNAs (63), which may be beneficial in that miRNAs originating from the brain may be more detectable in the blood than proteins, given that they can be measured via PCR-based methods. Likewise, the gene expression datasets used in our analysis could be used to informatically predict brain-enriched metabolites (64); if this information were combined with other existing disease-specific gene expression datasets, it may be possible to identify blood-borne brain-originating metabolites with a high degree of disease specificity.

Our collective findings provide an unprecedented depth of insight into numerous previously proposed candidate blood biomarkers of neurological damage, and suggest that several may have limited diagnostic utility in many clinical-use scenarios due to a low degree of brain specificity. Just as importantly, we have also identified a plethora of previously unexplored biomarkers which have strong potential for clinical use in several common neurological conditions, particularly traumatic brain injury, stroke, and multiple sclerosis. Further clinical validation of these markers could lead to the development of blood-based precision molecular diagnostics with the potential to transform how we detect and monitor these debilitating pathologies.

Materials and Methods

Detailed methodology for all informatic analyses, recruitment of subjects, and serum protein measurements can be found in *SI Appendix*, *Materials and Methods*.

Protection of Human Subjects. All procedures performed in this study involving human participants were in accordance with the ethical standards set forth in the 1964 Helsinki declaration and its later amendments, and were approved by the institutional review boards of University Hospitals (Cleveland, OH) and Ruby Memorial Hospital (Morgantown, WV). Written informed consent was obtained from all subjects or their authorized representatives prior to all study procedures.

Data Availability. Genotype Tissue Expression RNA sequencing data are available from https://gtexportal.org/home/datasets. Allen Brain Atlas microarray data are available from https://human.brain-map.org/static/download. Brain tissue single-cell RNA sequencing data are available from the NCBI GEO database via accession number GSE67835 and can be found at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67835. Compartments subcellular localization data are available from https://compartments.jensenlab.org/Downloads. Blood biomarker data are available via the Open Science Framework and can be found at https://osf.io/rab7q.

ACKNOWLEDGMENTS. We thank Dr. Paul D. Chantler of West Virginia University, and Dr. Taura L. Barr, formerly of West Virginia University, for providing access to stroke and traumatic brain injury blood samples. We also thank Dr. Mary E. Kerr and Ali L. Clark for critical review of the manuscript. Work was financially supported by Case Western Reserve University Frances Payne Bolton School of Nursing start-up funds issued to G.C.O.

1. V. L. Feigin *et al.*; GBD 2016 Neurology Collaborators, Global, regional, and national burden of neurological disorders, 1990-2016: A systematic analysis for the global burden of disease study 2016. *Lancet Neurol.* **18**, 459–480 (2019).
2. K. R. Lees *et al.*; ECASS, ATLANTIS, NINDS and EPITHET rt-PA Study Group, Time to treatment with intravenous alteplase and outcome in stroke: An updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *Lancet* **375**, 1695–1703 (2010).
3. I. K. Moppett, Traumatic brain injury: Assessment, resuscitation and early management. *Br. J. Anaesth.* **99**, 18–31 (2007).
4. G. Fuller, G. McClelland, T. Lawrence, W. Russell, F. Lecky, The diagnostic accuracy of the HITSNS prehospital triage rule for identifying patients with significant traumatic brain injury: A cohort study. *Eur. J. Emerg. Med.* **23**, 61–64 (2016).
5. G. Fuller, T. Lawrence, M. Woodford, F. Lecky, The accuracy of alternative triage rules for identification of significant traumatic brain injury: A diagnostic cohort study. *Emerg. Med. J.* **31**, 914–919 (2014).
6. A. E. Arch *et al.*, Missed ischemic stroke diagnosis in the emergency department by emergency medicine and neurology services. *Stroke* **47**, 668–673 (2016).
7. N. M. Lever *et al.*, Missed opportunities for recognition of ischemic stroke in the emergency department. *J. Emerg. Nurs.* **39**, 434–439 (2013).
8. B. Jiang *et al.*, Pre-hospital delay and its associated factors in first-ever stroke registered in communities from three cities in China. *Sci. Rep.* **6**, 29795 (2016).
9. D. N. Kernagis, D. T. Laskowitz, Evolving role of biomarkers in acute cerebrovascular disease. *Ann. Neurol.* **71**, 289–303 (2012).
10. H. Hampel *et al.*, Blood-based biomarkers for Alzheimer disease: Mapping the road to the clinic. *Nat. Rev. Neurol.* **14**, 639–652 (2018).
11. M. Comabella, X. Montalban, Body fluid biomarkers in multiple sclerosis. *Lancet Neurol.* **13**, 113–126 (2014).
12. M. D. Weingarten, A. H. Lockwood, S. Y. Hwo, M. W. Kirschner, A protein factor essential for microtubule assembly. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1858–1862 (1975).
13. L. F. Eng, J. J. Vanderhaeghen, A. Bignami, B. Gerstl, An acidic protein isolated from fibrous astrocytes. *Brain Res.* **28**, 351–354 (1971).
14. B. W. Moore, D. McGregor, Chromatographic and electrophoretic fractionation of soluble proteins of brain and liver. *J. Biol. Chem.* **240**, 1647–1653 (1965).
15. B. W. Moore, A soluble protein characteristic of the nervous system. *Biochem. Biophys. Res. Commun.* **19**, 739–744 (1965).
16. R. H. Laatsch, M. W. Kies, S. Gordon, E. C. Alvord Jr., The encephalomyelitic activity of myelin isolated by ultracentrifugation. *J. Exp. Med.* **115**, 777–788 (1962).
17. J. F. Doran, P. Jackson, P. A. Kynoch, R. J. Thompson, Isolation of PGP 9.5, a new human neurone-specific protein detected by high-resolution two-dimensional electrophoresis. *J. Neurochem.* **40**, 1542–1547 (1983).
18. G. G. Glenner, C. W. Wong, Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **120**, 885–890 (1984).
19. P. N. Hoffman, R. J. Lasek, The slow component of axonal transport. Identification of major structural polypeptides of the axon and their generality among mammalian neurons. *J. Cell Biol.* **66**, 351–366 (1975).
20. J. Lonsdale *et al.*; GTEx Consortium, The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
21. E. H. Shen, C. C. Overly, A. R. Jones, The allen human brain Atlas: Comprehensive gene expression mapping of the human brain. *Trends Neurosci.* **35**, 711–714 (2012).
22. S. O'Hagan, M. Wright Muelas, P. J. Day, E. Lundberg, D. B. Kell, GeneGini: Assessment via the Gini coefficient of reference "housekeeping" genes and diverse human transporter expression profiles. *Cell Syst.* **6**, 230–244.e1 (2018).
23. Z. Yang, K. K. W. Wang, Glial fibrillary acidic protein: From intermediate filament assembly and gliosis to neurobiomarker. *Trends Neurosci.* **38**, 364–374 (2015).
24. D. G. Thomas, N. R. Hoyle, P. Seeldrayers, Myelin basic protein immunoreactivity in serum of neurosurgical patients. *J. Neurol. Neurosurg. Psychiatry* **47**, 173–175 (1984).
25. D. G. T. Thomas, J. W. Palfreyman, J. G. Ratcliffe, Serum-myelin-basic-protein assay in diagnosis and prognosis of patients with head injury. *Lancet* **1**, 113–115 (1978).
26. N. Wąsik *et al.*, Serum myelin basic protein as a marker of brain injury in aneurysmal subarachnoid haemorrhage. *Acta Neurochir. (Wien)* **162**, 545–552 (2020).
27. M. Khalil *et al.*, Neurofilaments as biomarkers in neurological disorders. *Nat. Rev. Neurol.* **14**, 577–589 (2018).
28. H. Zetterberg, D. H. Smith, K. Blennow, Biomarkers of mild traumatic brain injury in cerebrospinal fluid and blood. *Nat. Rev. Neurol.* **9**, 201–210 (2013).
29. H. J. Kim, J. W. Tsao, A. G. Stanfill, The current state of biomarkers of mild traumatic brain injury. *JCI Insight* **3**, e97105 (2018).
30. O. Y. Glushakova, A. V. Glushakov, E. R. Miller, A. B. Valadka, R. L. Hayes, Biomarkers for acute diagnosis and management of stroke in neurointensive care units. *Brain Circ.* **2**, 28–47 (2016).
31. S. Darmanis *et al.*, A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015).
32. J. X. Binder *et al.*, COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database (Oxford)* **2014**, bau012 (2014).
33. S. Yanagitani *et al.*, Ischemia induces metallothionein III expression in neurons of rat brain. *Life Sci.* **64**, 707–715 (1999).
34. K. Tanji *et al.*, Expression of metallothionein-III induced by hypoxia attenuates hypoxia-induced cell death in vitro. *Brain Res.* **976**, 125–129 (2003).
35. T. Yuguchi *et al.*, Expression of growth inhibitory factor mRNA after focal ischemia in rat brain. *J. Cereb. Blood Flow Metab.* **17**, 745–752 (1997).
36. J. Kamtchum-Tatuene, G. C. Jickling, Blood biomarkers for stroke diagnosis and management. *Neuromolecular Med.* **21**, 344–368 (2019).
37. K. I. Gallacher, B. D. Jani, P. Hanlon, B. I. Nicholl, F. S. Mair, Multimorbidity in stroke. *Stroke* **50**, 1919–1926 (2019).
38. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
39. E. Rickels, K. von Wild, P. Wenzlaff, Head injury in Germany: A population-based prospective study on epidemiology, causes, treatment and outcome of all degrees of head-injury severity in two distinct areas. *Brain Inj.* **24**, 1491–1504 (2010).
40. C. Ren *et al.*, Assessment of serum UCH-L1 and GFAP in acute stroke patients. *Sci. Rep.* **6**, 24588 (2016).
41. A. Bustamante *et al.*, Blood biomarkers for the early diagnosis of stroke: The stroke-chip study. *Stroke* **48**, 2419–2425 (2017).
42. M. W. Koch, S. George, W. Wall, V. Wee Yong, L. M. Metz, Serum NSE level and disability progression in multiple sclerosis. *J. Neurol. Sci.* **350**, 46–50 (2015).
43. E. T. Lim *et al.*, Serum S100B in primary progressive multiple sclerosis patients treated with interferon-beta-1a. *J. Negat. Results Biomed.* **3**, 4 (2004).
44. J. J. Bazarian *et al.*, Serum GFAP and UCH-L1 for prediction of absence of intracranial injuries on head CT (ALERT-TBI): A multicentre observational study. *Lancet Neurol.* **17**, 782–789 (2018).

45. B. Olsson *et al*., CSF and blood biomarkers for the diagnosis of Alzheimer's disease: A systematic review and meta-analysis. *Lancet Neurol.* **15**, 673–684 (2016).

46. G. C. Jickling, F. R. Sharp, Biomarker panels in ischemic stroke. *Stroke* **46**, 915–920 (2015).

47. M. I. Hiskens, A. G. Schneiders, M. Angoa-Pérez, R. K. Vella, A. S. Fenning, Blood biomarkers for assessment of mild traumatic brain injury and chronic traumatic encephalopathy. *Biomarkers* **25**, 213–227 (2020).

48. S. Ray *et al*., Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat. Med.* **13**, 1359–1362 (2007).

49. G. C. O'Connell, P. Stafford, K. B. Walsh, O. Adeoye, T. L. Barr, High-throughput profiling of circulating antibody signatures for stroke diagnosis using small volumes of whole blood. *Neurotherapeutics* **16**, 868–877 (2019).

50. F. Barkhof *et al*., A correlative triad of gadolinium-DTPA MRI, EDSS, and CSF-MBP in relapsing multiple sclerosis patients treated with high-dose intravenous methylprednisolone. *Neurology* **42**, 63–67 (1992).

51. J. N. Whitaker, Myelin encephalitogenic protein fragments in cerebrospinal fluid of persons with multiple sclerosis. *Neurology* **27**, 911–920 (1977).

52. K. J. Lamers, H. P. de Reus, P. J. Jongen, Myelin basic protein in CSF as indicator of disease activity in multiple sclerosis. *Mult. Scler.* **4**, 124–126 (1998).

53. A. J. Solomon, J. R. Corboy, The tension between early diagnosis and misdiagnosis of multiple sclerosis. *Nat. Rev. Neurol.* **13**, 567–572 (2017).

54. J. L. Saver, Time is brain–quantified. *Stroke* **37**, 263–266 (2006).

55. J. Neves Briard *et al*., Stroke mimics transported by emergency medical services to a comprehensive stroke center: The magnitude of the problem. *J. Stroke Cerebrovasc. Dis.* **27**, 2738–2745 (2018).

56. N. Goyal, S. Male, A. Al Wafai, S. Bellamkonda, R. Zand, Cost burden of stroke mimics and transient ischemic attack after intravenous tissue plasminogen activator treatment. *J. Stroke Cerebrovasc. Dis.* **24**, 828–833 (2015).

57. Z. Zhelev, G. Walker, N. Henschke, J. Fridhandler, S. Yip, Prehospital stroke scales as screening tools for early identification of stroke and transient ischemic attack. *Cochrane Database Syst. Rev.* **4**, CD011427 (2019).

58. J. J. Li, P. J. Bickel, M. D. Biggin, System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270 (2014).

59. M. Jovanovic *et al*., Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015).

60. G. C. O'Connell *et al*., Use of high-sensitivity digital ELISA improves the diagnostic performance of circulating brain-specific proteins for detection of traumatic brain injury during triage. *Neurol. Res.* **42**, 346–353 (2020).

61. G. C. O'Connell *et al*., Diagnosis of ischemic stroke using circulating levels of brain-specific proteins measured via high-sensitivity digital ELISA. *Brain Res.* **1739**, 146861 (2020).

62. J. Lambert *et al*., Development of a high sensitivity 10-plex human cytokine assay using Simoa Planar Array technology. *J. Immunol.* **202**, 52.21 (2019).

63. G. C. O'Connell, C. G. Smothers, C. Winkelman, Bioinformatic analysis of brain-specific miRNAs for identification of candidate traumatic brain injury blood biomarkers. *Brain Inj.* **34**, 965–974 (2020).

64. A. Karnovsky *et al*., Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* **28**, 373–380 (2012).