Plasma profiling reveals three proteins associated to amyotrophic lateral sclerosis

Anna Häggmark¹,a, Maria Mikus¹,a, Atefeh Mhosenchian¹, Mun-Gwan Hong¹, Björn Forström¹, Beata Gajewska²,³, Anna Baraniczky-Kuźma²,³, Mathias Uhlen¹, Jochen M. Schwenk¹, Magdalena Kuźma-Kozakiewicz³,⁴,b & Peter Nilsson¹,b

¹SciLifeLab, School of Biotechnology, KTH-Royal Institute of Technology, Stockholm, Sweden
²Department of Biochemistry, Medical University of Warsaw, Warsaw, Poland
³Neurodegenerative Diseases Research Group, Medical University of Warsaw, Warsaw, Poland
⁴Department of Neurology, Medical University of Warsaw, Warsaw, Poland

Abstract

Objective: Amyotrophic lateral sclerosis (ALS) is the most common adult motor neuron disease leading to muscular paralysis and death within 3–5 years from onset. Currently, there are no reliable and sensitive markers able to substantially shorten the diagnosis delay. The objective of the study was to analyze a large number of proteins in plasma from patients with various clinical phenotypes of ALS in search for novel proteins or protein profiles that could serve as potential indicators of disease. Methods: Affinity proteomics in the form of antibody suspension bead arrays were applied to profile plasma samples from 367 ALS patients and 101 controls. The plasma protein content was directly labeled and protein profiles obtained using 352 antibodies from the Human Protein Atlas targeting 278 proteins. A focused bead array was then built to further profile eight selected protein targets in all available samples. Results: Disease-associated significant differences were observed and replicated for profiles from antibodies targeting the proteins: neurofilament medium polypeptide (NEFM), solute carrier family 25 (SLC25A20), and regulator of G-protein signaling 18 (RGS18). Interpretation: Upon further validation in several independent cohorts with inclusion of a broad range of other neurological disorders as controls, the alterations of these three protein profiles in plasma could potentially provide new molecular markers of disease that contribute to the quest of understanding ALS pathology.

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder characterized by degeneration of motor neurons in primary motor cortex, brain stem, and spinal cord. It leads to progressive weakness of limb and bulbar muscles, spasticity and muscle wasting, followed by death due to respiratory insufficiency. ALS is the most common adult onset motor neuron disease, affecting worldwide over 100,000 people yearly. Due to the lack of effective treatment, the prognosis remains poor with median survival ranging from 3 to 5 years from the symptoms onset. However, the disease course and prognosis may be affected by genetic factors, age at onset, localization of onset, and clinical phenotype. Approximately 80–85% of ALS cases are sporadic (SALS), the rest are inherited (familial ALS, FALS) mainly as an autosomal dominant trait. Although the disease starts earlier in FALS patients, their survival may not
differ from that of SALS. Overall, ALS patients with respiratory phenotype survive the least, and the bulbar phenotype is characterized by the second shortest survival. On the other hand, individuals suffering from pure upper motor neuron phenotype (PUMN) – a very rare ALS phenotype – may survive over 15 years, so may some of the patients with pure lower motor neuron phenotype (PLMN) if the disease does not spread into the upper motor neuron (UMN) evolving into classic ALS. The factors determining clinical phenotypes remain elusive; there are also discussions if the nonclassic cases should be considered as ALS or – more generally – as motor neuron disease.

Several mechanisms involved in motor neuron degeneration in ALS have been described to date. Molecular indicators of glutamate excitotoxicity, impaired axonal transport, oxidative stress, mitochondrial dysfunction, growth factor deficiency, protein aggregation, abnormal RNA metabolism, and apoptosis have been observed in human brain, spinal cord, and transgenic disease models long before symptom onset. Studies on cerebrospinal fluid (CSF) allowed to identify phosphorylated neurofilament heavy chain and complement C3 as well as cytokine panels of granulocyte–monocyte colony stimulating factors as potential disease biomarkers. Extensive studies on plasma proteins are, however, still lacking.

The aim of this study was to analyze the protein profiles in plasma of a large group of clinically well-characterized patients with various clinical phenotypes of ALS in search for profiles that could contribute to improved understanding of ALS pathology. We have here employed an affinity proteomics approach applying antibody suspension bead arrays with the potential to screen hundreds of proteins in hundreds of patient samples simultaneously in a parallel analysis. This method was utilized with 352 antibodies generated within the Human Protein Atlas (HPA) targeting 278 proteins and profiled in plasma from 367 ALS patients and 101 controls.

Materials and Methods

Patients

In total, 367 ALS patients (350 SALS and 17 FALS) were diagnosed and followed at the Department of Neurology, Medical University of Warsaw, Poland. The ALS diagnosis was made according to revised El Escorial criteria. Patients were further characterized according to age, gender, clinical phenotype, site of onset, diagnosis delay, disease duration, and functional status (ALS-functional rating scale). Classic ALS was defined as concomitant UMN and lower motor neuron (LMN) involvement of either limb or bulbar onset with rapid symptoms generalization. Cases were classified as bulbar phenotype if isolated dysarthria and/or dysphagia were present for at least 6 months from the first symptom onset. PUMN phenotype was characterized by isolated UMN involvement without electrophysiological signs of the LMN damage; it was considered a clinically possible ALS. Since the PLMN phenotype was not classified in the revised El Escorial criteria, we used the previously used term of “suspected ALS” elaborated in the primary El Escorial criteria. PLMN was considered if the symptoms were restricted to the LMN and if genetic factors (CAG expansion or SMN1 deletion) and electrophysiological conduction block were excluded. Neuroimaging was used to exclude organic changes of the brain, brainstem, and spinal cord in all cases. The control group included 101 healthy individuals or patients diagnosed at the same department with benign non-neurodegenerative diseases such as headache or radiculopathy. The study was approved by the Institutional Bioethical Committees according to Helsinki declaration. All subjects signed an informed consent prior to inclusion in the study. For sample demographics, see Table 1.

Samples

Ethylenediaminetetraacetic acid plasma samples were collected and separated from primary mononuclear cells by

| Table 1. Sample demographics. (A) Number of patients and distribution of age and gender in the ALS and control groups. (B) Number of patients and disease duration in months for the clinical phenotypes, site of onset, and El Escorial criteria. |
| --- | --- | --- |
| (A) | N | Age Mean (SD) | Female (%) |
| ALS | 367 | 58 (13) | 49 |
| SALS | 350 | 58 (13) | 49 |
| FALS | 17 | 56 (13) | 53 |
| Control | 101 | 45 (19) | 60 |
| (B) | N | Disease duration, months Mean (SD) |
| Clinical phenotype | | |
| Classic | 202 | 25 (24) |
| Bulbar | 36 | 16 (13) |
| PLMN | 51 | 50 (51) |
| PUMN | 10 | 45 (37) |
| Site of onset | | |
| Bulbar | 69 | 18 (16) |
| Limb | 230 | 32 (34) |
| El Escorial criteria | | |
| Definite | 66 | 26 (24) |
| Probable | 117 | 22 (18) |
| Possible | 66 | 27 (30) |
| Suspected | 51 | 50 (51) |

Information on phenotype, onset, and El Escorial criteria was incomplete for 80 patients. ALS, amyotrophic lateral sclerosis; SALS, sporadic ALS; FALS, familial ALS; PLMN, pure lower motor neuron; PUMN, pure upper motor neuron.
centrifugation with GRADISOL L (Aqua-Med Łódź, Poland). Samples were then immediately frozen and stored at −80°C until further use.

**Antibody selection**

Proteins proposed for this study were selected by a generous and inclusive literature search, as well as previous internal protein profiling studies in the context of neuroscience. The used antibody set was based on availability within the HPA project and included 352 antigen-purified and protein microarray validated antibodies targeting 278 unique proteins.

**Suspension bead array**

The procedure for suspension bead arrays was performed as described previously. In short, samples were distributed in 96-well microtiter plates, diluted 1:10 in Phosphate buffered saline, and the protein content directly labeled with biotin. For the bead array, antibodies were immobilized onto magnetic color-coded beads with one bead identity corresponding to a certain antibody. Samples were then further diluted 1:50 in an assay buffer, heat treated at 56°C for 30 min, combined into a 384-well microtiter plate, and incubated with the bead array at RT on a shaker overnight. Unbound proteins were removed by washing and proteins on the beads were detected through a streptavidin-conjugated fluorophore (Invitrogen). Results from the FlexMap3D instrument (Luminex Corp.) were reported per bead identity as median fluorescence intensities (MFI). For a more extensive description, see Data S1.

**Antibody validation**

Three antibodies were validated with Western blot and epitope mapping using peptide arrays. For Western blot, plasma samples were diluted 1:40 in MilliQ and proteins separated on a gel before blotted on a membrane (all Invitrogen). Detection antibody was applied at 1 µg/mL and binding allowed at 4°C overnight followed by readout through a chemiluminescent substrate (BioRad). The three polyclonal antibodies were epitope mapped using high-density peptide arrays (Roche NimbleGen) containing synthetic 12-mer peptides with 11 amino acids overlap covering the whole sequence used for antibody production. See Data S1 for a more detailed protocol.

**Statistical analysis**

Data were processed and visualized in R. Samples with less than 20 counted beads per identity as well as outliers identified by robust principal component analysis (rPCA, R package “rcov”) were excluded from further analysis. For biological interpretation, MFI values were normalized within each 96-well plate by probabilistic quotient normalization as described earlier. In addition, to diminish any plate effect, multidimensional normalization was applied with the assumption that the mean of each plate for each antibody should be close. Technical variation was assessed by calculating the coefficient of variation for each antibody based on samples measured in triplicates. The association between groups was tested using normalized data by Wilcoxon rank sum or Kruskal-Wallis tests and unadjusted P-values lower than 0.01 were regarded as significant. Spearman correlation coefficients were calculated to evaluate associations and Euclidian distance was applied for hierarchical clustering. Receiver operating characteristic (ROC) analysis was performed to evaluate the classification power of selected proteins and normalized data were log transformed for logistic regression applied for either the antibodies individually or combined. The model was trained with data from 324 individuals and to assess the classification power, the remaining 144 were used for testing.

**Results**

With the aim to identify protein profiles associated to ALS, we have performed an extensive profiling of plasma samples from 367 ALS patients and 101 controls. The selection of protein targets was based on thorough literature mining with an aim to include all published proteins with potential relation to ALS and the disease pathogenesis. Furthermore, we added a selection of targets from previous in-house efforts of antibody-based protein profiling of plasma and CSF within neuroscience. In the end, availability of validated antibodies within the HPA directed the final set to a unique panel of 352 antibodies targeting 278 unique proteins.

The samples were randomly divided into two sample plates and screening using the panel of 352 antibodies was performed for both plates. After analysis of all 468 individuals, eight antibodies were found to reveal significant or close to significant differences (P < 0.05) in both plates. To further verify the findings, additional antibodies were included toward these targets and all samples were analyzed once more with a direct labeling assay. Of these eight, three proteins were replicated with a significant difference between ALS patients and controls. See Figure 1 for an overview of the study design.

The three antibodies showing significant differences between ALS patients and controls were generated toward neurofilament medium polypeptide (NEFM, HPA022845), regulator of G-protein signaling 18 (RGS18, HPA028727), and solute carrier family 25 (carnitine/acylcarnitine.
translocase) member 20 (SLC25A20, HPA029863). All three showed higher levels of the proteins in ALS patients (Fig. 2) with a technical variation of 14%, 13% and 13%, respectively. The NEFM antibody was analyzed with Western blot and gave a distinct band of expected molecular weight (150 kDa) and one additional band at higher molecular weight (~450 kDa), which could indicate multimeric forms. A band of predicted molecular weight (33 kDa) was also observed for the SLC25A20 antibody. The antibody toward RGS18 showed no band in this analysis. The high-density peptide microarray-based epitope mapping revealed clear binding to several adjacent peptides and epitopes for all three antibodies. The results can be summarized as the shortest common amino acid sequences for adjacent peptides and the minimal epitopes identified were VVTI for NEFM (corresponding to amino acids 746–749, ENSP00000221166), VITNSIT for RGS18 (amino acids 160–166, ENSP00000356430), and KYTG-TLDC, APPGKY, and YKGFNAV for SLC25A20 (amino acids 148–155, 240–245, and 266–272, respectively, ENSP00000326305).

Relating protein profiles to clinical parameters

The generated protein profiles were also related to available clinical information and the result from each antibody in the screening on the bead array visualized according to El Escorial criteria, site of onset, and clinical phenotype. Aside from comparisons including the control group, a significant difference was also observed for phenotypes, where the group of patients with definite ALS had higher NEFM levels compared to suspected ALS (Fig. 3). Signal intensities were correlated with disease duration, diagnosis delay, and functional status revealing neither positive nor negative correlations, that is, beyond ±0.5. Due to the difference in age between ALS patients and controls included in this study, age-related effects were investigated for the three proteins to exclude that the levels of these proteins were related to age and not disease. For this purpose, correlations of age to the obtained signal intensities for the control group and ALS patients were performed separately and neither a positive nor a negative correlation was observed (correlation coefficients between 0.02 and 0.11).

Figure 1. Overview of the study design. During screening, two plates with plasma samples, including 324 and 144 individuals respectively, were profiled with 352 antibodies targeting 278 proteins using bead arrays and a direct labeling approach. Eight antibodies were found to detect altered levels of their targets in the amyotrophic lateral sclerosis patients compared to the control group and these were selected for an additional replication round including both sample sets. Of the eight, three antibodies were in the end confirmed to show consistent trends.

Figure 2. Differences in protein levels between amyotrophic lateral sclerosis (ALS) patients and controls. Three antibodies targeting NEFM, RGS18, and SLC25A20 were found to reproducibly show significant differences between ALS patients ($n = 367$) and controls ($n = 101$) ($P$-values are shown below respective plot). NEFM, neurofilament medium polypeptide; RGS18, regulator of G-protein signaling 18; SLC25A20, solute carrier family 25 (carnitine/acylcarnitine translocase).
Multivariate analysis

We performed hierarchical clustering using data for NEFM, RGS18, and SLC25A20 and observed two main clusters, see Figure 4A. In the smaller cluster, individuals with generally low levels of all three proteins in both ALS patients and controls were found while in the larger, further subclustering revealed a group of ALS patients with generally high levels of these proteins. When visualizing according to the clinical annotation only including ALS patients, none of the available clinical parameters could clearly be related to the main or subclusters (Fig. 4B).

Profiles generated in the direct labeling assay were analyzed both separately and in combination through logistic regression. The models were trained with data from one of the assay plates \((n = 324)\) and tested using the other \((n = 144)\). We obtained areas under curve (AUC) of 0.80, 0.86, and 0.78 for antibodies targeting NEFM, RGS18, and SLC25A20, respectively. When combining the three antibodies in one model, an AUC of 0.86 was achieved indicating that the three profiles did not generate a model with better classification power than the RGS18 antibody alone, see Figure 5.

Discussion

In this study of plasma profiling within ALS, the suspension bead array technology was utilized to profile large number of samples and analytes in parallel. A screening of 468 samples using 352 antibodies resulted in the
finding of three interesting proteins with potential to discriminate ALS patients from controls where NEFM as well as both RGS18 and SLC25A20 were found at higher levels in ALS patients. While NEFM is highly enriched in the nervous system, the other two have a more widespread expression and have been detected in various cells and tissues at both RNA and protein level. NEFM, the protein with the highest statistical significance in this study, is one of the subunits of neurofilaments. According to both gene expression based on RNAseq data and Immunohistochemistry staining in the HPA (www.proteinatlas.org), as well as the GeneSapiens database, expression of this protein is highly restricted to neurons in the central and peripheral nervous system. The neurofilaments have been mentioned in the context of several neurodegenerative disorders, in which for instance mutations and abnormal accumulation of filaments have been identified as markers. Increased levels of neurofilament light and phosphorylated neurofilament heavy polypeptide have been reported in CSF and plasma of ALS patients. Light and heavy subunits have also been associated with disease progression and survival, where higher concentrations in CSF and/or plasma were associated with a faster progression and shorter survival. Neurofilaments are heavily phosphorylated and accumulation of the hyperphosphorylated forms have been identified in ALS. Also glycosylations of these proteins have been described as a modification related to neurode-

Figure 4. Sample clustering based on NEFM, RGS18, and SLC25A20. (A) Cluster analysis based on intensities from the NEFM, RGS18, and SLC25A20 antibodies revealed two main clusters. The left cluster included individuals, both ALS patients and controls, with generally low levels of the three proteins. In the right cluster, further subclusters were observed with a tendency of separating the ALS patients with high levels from controls and patients with intermediate levels. (B) Based on ALS patients with complete available clinical information, a similar cluster was generated, and also here, two main clusters were obtained but not explained by any of the included parameters. The intensity scale is visualized from low (white) to high (black). NEFM, neurofilament medium polypeptide; RGS18, regulator of G-protein signaling 18; SLC25A20, solute carrier family 25 (carnitine/acylcarnitine translocase); ALS, amyotrophic lateral sclerosis.
generation and in a transgenic rat model of ALS, a
decrease in the O-GlcNAcylation of NEFM was
observed.23–25 Since NEFM is used as an indicator of neu-
ronal damage, its increased plasma concentration might
be a consequence of the axonal death in ALS. Interest-
ingly, it has also been suggested as an autoimmune target
in the context of this disease.26,27

RGS18 belongs to a large family of regulators of G-
protein signaling proteins, all sharing a highly con-
served RGS domain of ~120 amino acids. RGS proteins
are known inhibitors of signal transduction and
increase GTPase activity of the alpha subunits of G
proteins.28 RGS18 expression has been reported in
human platelets, leukocytes, megakaryocytes, and lymphocytes29–31 and it was found to decrease in abun-
dance in degenerating muscle of rainbow trout.32

Muscle wasting secondary to the loss of motor neurons
is a key symptom of ALS. The increase in RGS18 level
in plasma may reflect its leakage from the degenerated
muscles in patients with ALS.

SLC25A20, a member of the large mitochondrial carrier
family, catalyzes transport across the inner mitochondrial
membrane. Cytosolic acylcarnitines are transferred into
the mitochondrial matrix for oxidation by the β-oxida-
tion pathway in exchange for carnitines.33 Disturbances of
ATP production, calcium homeostasis, and regulation of
apoptosis, all related to mitochondria, have all been
reported in ALS. Abnormalities of this organelle’s mor-
phology and biochemistry in brain cortex, spinal cord,
and muscles have been seen both in ALS patients and in
SOD1 transgenic mice.2,3,34,35

In this study, a number of proteins showed trends that
could be of potential interest but was either not signifi-
cant in the screening at set threshold, only significant in
one of the plates or not significant in the replication
assay. These proteins include ERRFI1, LRRCC9, C9,
SLC9A3R1, PSMD7, TFPT, CRYAA, MMP9, GSK3B,
MAP3K1, TJP2, S100A8, and ZFP36L1 for which levels
were higher in the ALS patients, whereas CST3,
DYNCHI1, KIFC3, PIK3CA, PRPH, and PSMB1 were

---

**Figure 5.** The performance of classification models of individual and combined proteins. The classification power of the three antibodies separately and combined was evaluated through ROC analysis, where 324 samples were used for the purpose of training the models and the additional 144 for testing. The area under the curve (AUC) for the combined model was not improved compared to what was obtained by RGS18 alone. ROC, receiver operating characteristic; RGS18, regulator of G-protein signaling 18.
found at lower levels in these patients. It is noticeable that we did not find any significant changes in the plasma level of several proteins directly or indirectly associated to the pathogenesis of ALS (data not shown). The reasons that the antibodies utilized did not show differences could though be many, for example, that the proteins are not present in plasma at detectable levels (currently approximately in the high pg/mL to low ng/mL range).

For target verification using Western blot, the NEFM and SLC25A20 antibody detected a target of the predicted molecular mass. The absence of bands for RGS18 could be due to the difference in representation of protein in a denatured state on a membrane compared to solution, potentially affecting the antibodies’ recognition to find their epitopes. However, for all three antibodies, epitopes were successfully identified using high-density peptide arrays.

In the replication stage of this study, several antibodies were included toward the selected protein targets. For all additional antibodies toward NEFM, RGS18, and SLC25A20 signal intensities were low in the direct labeling assay, indicating that these antibodies might not capture their targets in the described setup. They were also generated toward other parts of the respective proteins and their epitopes might therefore differ in accessibility and detectability as a response to heat treatment.

The three proteins were found at higher levels in plasma of ALS patients compared to controls, however, they were not able to distinguish between FALS and SALS. Interestingly, there were also no differences in levels for subclassifications, site of onset, or phenotypes, aside from the levels of NEFM for definite and suspected ALS. This is especially interesting for PUMN, the most rare clinical phenotype with comparably very long survival, which could potentially be due to a different pathogenic background. There has also been a vivid discussion concerning the inclusion of PLMN phenotype in the diagnosis of ALS. Although primarily classified as suspected ALS the phenotype was excluded from the ALS diagnosis according to the revised El Escorial criteria. However, data published recently reported the existence of common genetic, risk and survival determining factors in progressive muscular atrophy (PLMN of a long duration; >48 months) and classic ALS. Our results sustain the common pathogenesis of these disorders by showing the comparably increased levels of the three identified proteins in plasma of all ALS phenotypes, including PLMN, as compared to controls.

The present study has enabled a representation and assessment of the individual variation in plasma protein levels not only in ALS patients as a whole but also in clinically relevant subgroups, comparisons that would have been limited if it were not for the large sample size. However, to increase the understanding of the function and role of these proteins, further characterizations are required in a broad range of independent sample materials, not at least it important to investigate the potential association and relation to other neurological and muscle degenerative disorders.

To our current knowledge, this is the most extensive plasma profiling study conducted within ALS and we have here identified three interesting proteins as candidate markers to be further explored, to elucidate their potential role in the pathogenesis of ALS, as well as their usefulness in monitoring the response to therapy.

Acknowledgments

This study was funded by grants from Science for Life Laboratory Stockholm, ProNova VINN Excellence Centre for Protein Technology (VINNOVA, Swedish Governmental Agency for Innovation Systems), Knut and Alice Wallenberg Foundation, and grant 5/JPND/2011 (M. K.-K., A. B.-K.) from Joint Program for Neurodegenerative Disorders. We thank the whole group of Affinity Proteomics at SciLifeLab Stockholm for their contributions and we also acknowledge the entire staff of the Human Protein Atlas for their efforts in generating the antibodies. We would also especially like to thank all the patients who donated samples for this study.

Conflict of Interest

None declared.

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Extended information on materials and methods for the suspension bead array and antibody validation.