



Proteomics profiling and machine learning in nusinersen-treated patients with spinal muscular atrophy

Chiara Panicucci¹ · Eray Sahin² · Martina Bartolucci³ · Sara Casalini¹ · Noemi Brolatti¹ · Marina Pedemonte⁴ · Serena Baratto¹ · Sara Pintus¹ · Elisa Principi¹ · Adele D'Amico⁵ · Marika Pane⁶ · Marina Sframeli⁷ · Sonia Messina⁷ · Emilio Albamonte⁸ · Valeria A. Sansone⁸ · Eugenio Mercuri⁶ · Enrico Bertini⁵ · Ugur Sezerman⁹ · Andrea Petretto³ · Claudio Bruno^{1,10}

Received: 25 February 2024 / Revised: 11 August 2024 / Accepted: 25 August 2024
© The Author(s) 2024

Abstract

Aim The availability of disease-modifying therapies and newborn screening programs for spinal muscular atrophy (SMA) has generated an urgent need for reliable prognostic biomarkers to classify patients according to disease severity. We aim to identify cerebrospinal fluid (CSF) prognostic protein biomarkers in CSF samples of SMA patients collected at baseline (T0), and to describe proteomic profile changes and biological pathways influenced by nusinersen before the sixth nusinersen infusion (T302).

Methods In this multicenter retrospective longitudinal study, we employed an untargeted liquid chromatography mass spectrometry (LC-MS)-based proteomic approach on CSF samples collected from 61 SMA patients treated with nusinersen (SMA1 n=19, SMA2 n=19, SMA3 n=23) at T0 at T302. The Random Forest (RF) machine learning algorithm and pathway enrichment analysis were applied for analysis.

Results The RF algorithm, applied to the protein expression profile of naïve patients, revealed several proteins that could classify the different types of SMA according to their differential abundance at T0. Analysis of changes in proteomic profiles identified a total of 147 differentially expressed proteins after nusinersen treatment in SMA1, 135 in SMA2, and 289 in SMA3.

Overall, nusinersen-induced changes on proteomic profile were consistent with i) common effects observed in all SMA types (i.e. regulation of axonogenesis), and ii) disease severity-specific changes, namely regulation of glucose metabolism in SMA1, of coagulation processes in SMA2, and of complement cascade in SMA3.

Conclusions This untargeted LC-MS proteomic profiling in the CSF of SMA patients revealed differences in protein expression in naïve patients and showed nusinersen-related modulation in several biological processes after 10 months of treatment. Further confirmatory studies are needed to validate these results in larger number of patients and over a broader timeframe.

Keywords Spinal muscular atrophy · Proteomics · Cerebrospinal fluid · Biomarkers · Machine learning · Artificial intelligence

Introduction

Spinal muscular atrophy (SMA) is an autosomal recessive neurodegenerative disorder, due to defects in the survival motor neuron (SMN1) gene on chromosome 5, leading to degeneration of motor neurons in the spinal cord, and

inducing a progressive muscular hypotonia and weakness [1].

According to the age at onset and the best motor function achieved, SMA has been classified into five main types with pediatric onset (types 1 to 3, from the most severe to the mildest), and two more recently classified phenotypes,

Extended author information available on the last page of the article

type 0 with antenatal onset, and type 5 with adult onset and a mild phenotype [1]. Disease severity inversely correlates with the number of copies of the SMN2 gene [2, 3], the SMN1 paralogue gene, which produces predominantly an alternatively spliced mRNA transcript lacking the exon 7 and encoding for an unstable SMN protein (SMN Δ 7) [4].

In the last decade, significant improvements have been made in SMA treatment, resulting in three available therapies, namely the SMN2 splicing-modifiers nusinersen (antisense oligonucleotide administered intrathecally) and risdiplam (small molecule with oral route), which upregulate SMN protein levels [5, 6], and onasemnogene abeparvovec (an adeno-associated virus vector-based gene therapy administered once i.v.) [7]. These treatments have improved life expectancy and quality of life in SMA patients, particularly when started in the pre-symptomatic or early stages of the disease [8, 9].

In this scenario, neonatal screening programs are mandatory to ensure rapid diagnosis and guarantee a prompt therapeutic intervention, making the discovery of novel biomarkers urgently required.

To date, the only available and clinically relevant biomarkers are represented by the SMN2 copy number, alongside with two alternative splicing-modulating variants (rs121909192 and rs1454173648), although not uniquely predictive of disease severity [1]. In addition to genetic biomarkers, neurofilaments (NfL) have shown potentiality as serum and cerebrospinal fluid (CSF) biomarker, but inconsistent results prevent their wide use in the clinical setting, especially in adult patients [10].

In the last few years, omics-based techniques have been applied for biomarkers discovery in CSF samples of nusinersen-treated SMA patients [11–18], providing valuable insights on nusinersen-related modification of CSF proteome and metabolome, and identified putative treatment response biomarkers. However, no prognostic biomarkers in naïve patients have been explored so far.

In the context of omics data generation and analysis, machine learning algorithms demonstrated exceptional potential for managing extensive and multidimensional datasets, serving as fundamental tools to facilitate biomarker discovery across various fields, including neuroscience, but their application in SMA remains unexplored [19–22].

We conducted a multi-center retrospective longitudinal study, using an untargeted liquid-chromatography mass spectroscopy-based proteomic approach on CSF samples collected from a large cohort ($n=61$) of SMA1, SMA2 and SMA3 patients at baseline (before starting nusinersen treatment, T0), and after 10 months of therapy (before the sixth nusinersen infusion, T302). The primary objective of the study was to identify CSF prognostic biomarkers at baseline, by combining proteomics data and a machine learning

algorithm. The secondary objective was to describe changes in the proteomic profile at T302 and identify which biological pathways are influenced by nusinersen, through the application of classical bioinformatics approaches.

Results

Study design and population

The study design is summarized in Fig. 1a.

Sixty-one genetically confirmed SMA patients treated with nusinersen, including 19 SMA1, 19 SMA2, and 23 SMA3 subjects, were enrolled in the study (Fig. 1a). Demographic characteristics and clinical data of patients are reported in Table 1.

The median age at T0 was 2.1 years (IQR: 0.7–4.9) in SMA1, 7.7 years (IQR: 2.3–11.8) in SMA2, and 15.3 years (IQR: 7–29.3) in SMA3 (Table 1). Differences in age at T0 were significantly found between the SMA types (SMA1 vs. SMA2, $p < 0.005$; SMA1 vs. SMA3, $p < 0.0001$; SMA2 vs. SMA3, $p < 0.005$; Kruskal-Wallis test, multiple comparison), while gender did not differ among groups (Pearson's Chi squared test p -value = 0.927).

At baseline, median CHOP values in SMA1 was 13/64 (IQR: 3.7–24), while median HFMSE scores were 9/66 (IQR: 2.7–13) in SMA2 and 47/66 (IQR: 29–58) in SMA3 (Fig. 1b).

Identification of putative prognostic biomarkers through machine learning

Proteome analysis at T0 identified 1625 unique proteins. Principal component analysis (PCA) shows separation of proteome profiles in the different SMA subtypes (Fig. 2a). PERMANOVA test carried out on Euclidian distance of proteomics data at T0 revealed that only SMA subtype ($F = 5.1894$, $p = 0.001$) and age ($F = 3.4401$, $p = 0.001$) can significantly explain the variance in the data, while the rest of the factors, gender, BMI group, SMN2 copy number, and response status, were not significant ($p > 0.05$). Therefore, for the differential abundance analysis, age information was used as a confounding factor.

A machine learning approach, the Random forest (RF) classification algorithm, was then used to test whether the protein expression at T0 could be used to predict the SMA subtypes. All the trained RF models obtained gave accuracy of 1, which is an indicator of over-fitting, and the same accuracy value was obtained from prediction on test datasets, too. To eliminate the possible effects of post-processing of the data and hyper tuning of the RF parameters in model trainings, that might cause data leaking and over-fitting,

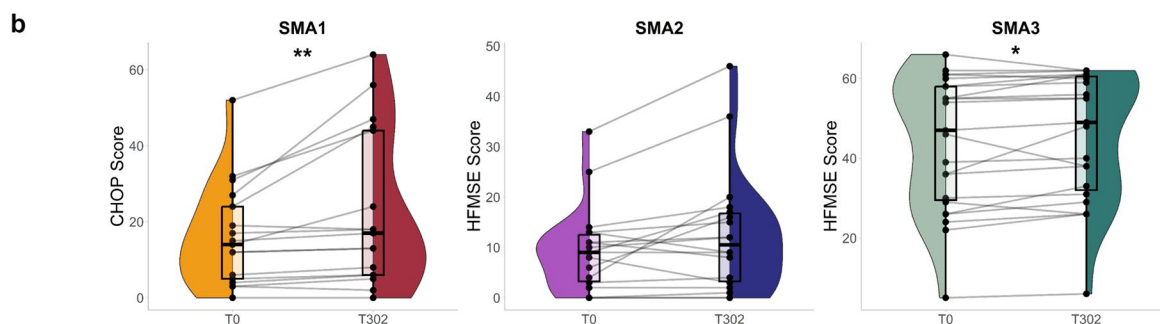
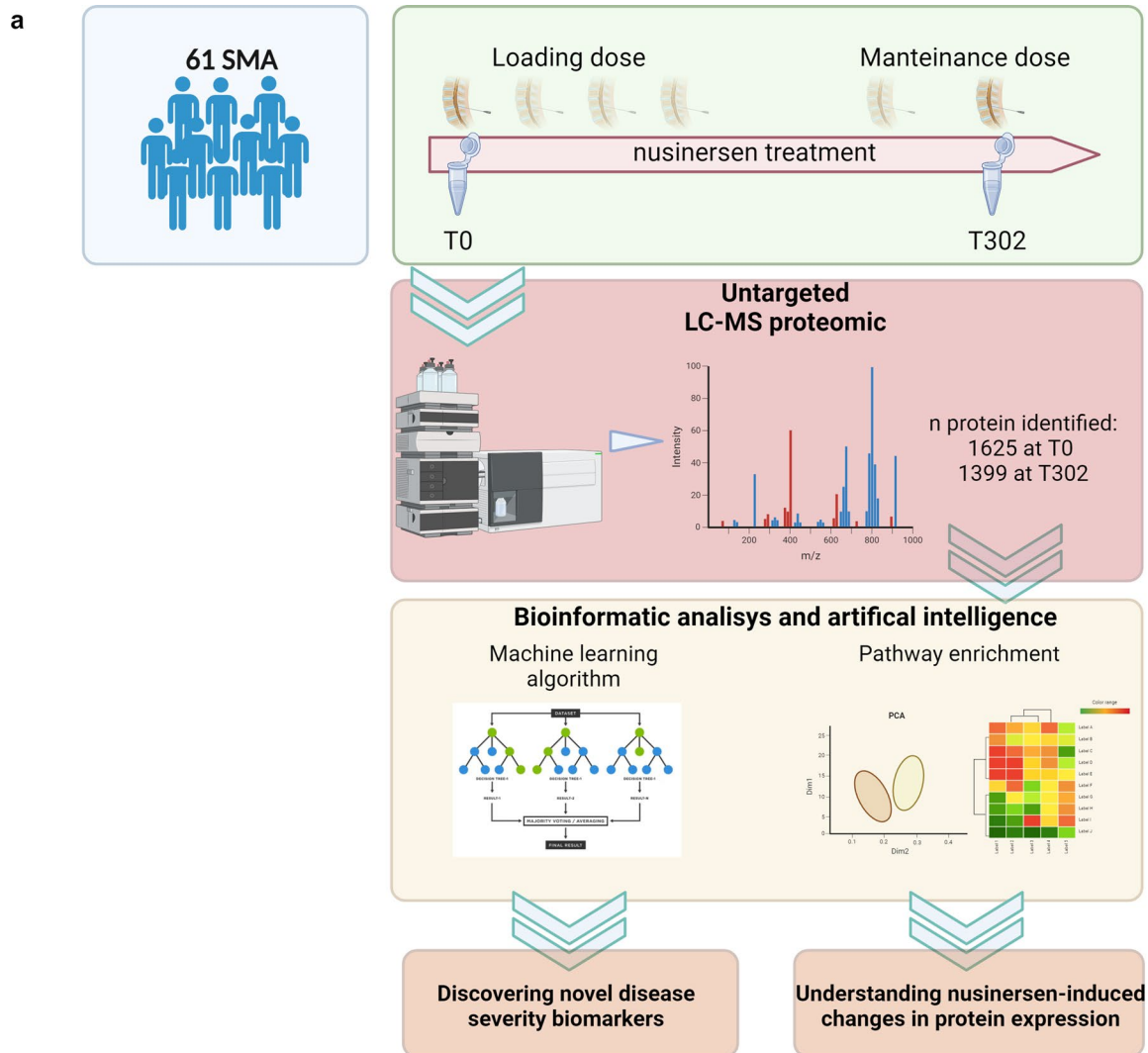


Fig. 1 Study design and population. **(a)** A graphical description of the study is shown. **(b)** Changes in CHOP scores for SMA1 ($n=17$), and HFMSE scores in SMA2 ($n=18$) or SMA3 ($n=23$) patients before (T0) and after (T302) nusinersen treatment. Plots show points for each sample score and paired sample data were connected by lines. Violin

plot shows the distribution of the scores via density, and box plots showing interquartile range; box has sides belonging to lower and upper quartiles, and median depicted by horizontal line. Significance was tested via paired Wilcoxon rank-sum test. Significance test results were shown by “**” for p-value < 0.05; “***” for p-value < .01

Table 1 Study population description. Age at baseline, gender, number of SMN2 copies, response to therapy, respiratory support and nutrition information are listed in the table

	SMA1 (n=19)	SMA2 (n=19)	SMA3 (n=23)	Total (n=61)
Age at T0 (median, Q1-Q3)	2.1 (0.7–4.9)	7.7 (2.3–11.8)	15.3 (7–29.3)	6.9 (2.3–13)
Gender (F/M)	11/8	12/7	12/11	30/26
SMN2 copies (n,%)				
2 SMN2	17 (89%)	2 (10%)	3 (13%)	22 (36%)
3 SMN2	2 (10%)	17 (89%)	10 (43%)	29 (48%)
> 3 SMN2	0	0	9 (39%)	9 (15%)
Responders (n,%)	11 (52%)	6 (31%)	6 (23%)	23 (37%)
Respiratory support (n,%)				
Use of non-invasive ventilation	7 (37%)	8 (42%)	1 (4%)	16 (26%)
Tracheostomy	7 (37%)	1 (5%)	0	8 (13%)
Nutrition (n,%)				
Oral	7 (37%)	19 (100%)	23 (100%)	49 (80%)
Nasogastric tube	8 (42%)	0	0	8 (13%)
Gastrostomy	4 (21%)	0	0	4 (6%)

different strategies were followed, including splitting raw dataset into train and test, and applying normalization procedures independently, changing the ratio of train and test datasets, and varying size of grids for hyper parameter tuning

(the detailed strategies employed during optimization and prevention of overfitting and data leakage are described in the Supplementary Methods). All models consistently gave accuracy of 1 on both trained and test datasets, signifying the strength of the proteomics data in SMA subtype classification. Lastly, to further decrease the potential bias of the model on feature selection for the classification arising from the relatively small sample size, 100 different RF models were trained by changing the samples in the train and test datasets, and common proteins across top 30 features, sorted based on mean decrease in Gini coefficient of all 100 RF models, were determined (Supplementary Table 1). Accordingly, 9 proteins were obtained as common features from the 100 models: heat shock protein 90 beta family member 1 (HSP90B1), serpin family F member 2 (SERPINF2), golgi membrane protein 1 (GOLM1), contactin 1 (CNTN1), PVR cell adhesion molecule (PVR), beta-1,4-glucuronyltransferase 1 (B4GAT1), cathepsin F (CTSF), neurexin 3 (NRXN3), and chromosome 16 open reading frame 89 (C16orf89) (Fig. 2b). The maximum number of modes for model training was set to 3, and we observed that the classification of the three SMA subtypes can be achieved using combinations of two proteins selected based on their differential abundance profiles in SMA subtypes. Out of the 9 selected proteins, CNTN1, B4GAT1, NRXN3, and C16orf89 were

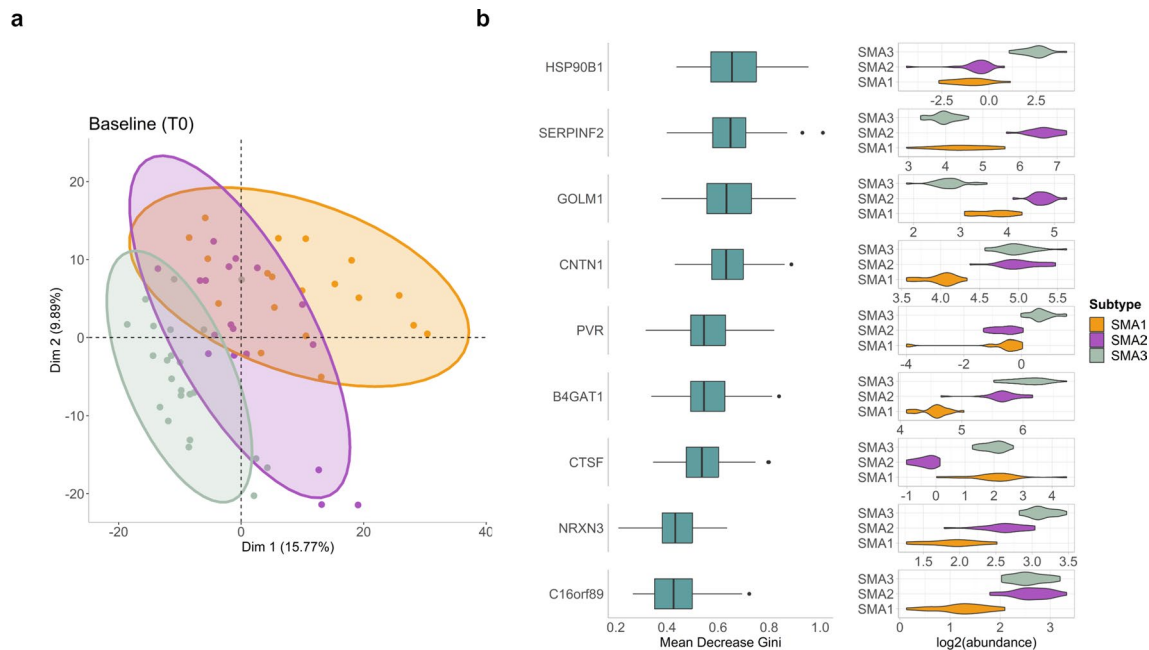


Fig. 2 Random Forest identified new putative biomarkers for SMA severity. **(a)** PCA plot showing the separate clustering patterns of SMA1 (orange), SMA2 (purple) and SMA3 (green) proteomic profiles at T0 based on dimensions 1 and 2. On PCA plots, each sample was shown by points, and ellipses correspond to 95% confidence intervals for each of the subtypes. **(b)** Nine proteins common among top 30 features obtained from 100 different Random Forest models for clas-

sification of SMA types at T0. On the left side of figure, the importance score (based on mean decrease in Gini index) distributions of each protein from 100 models were illustrated by box plots. Log₂ abundance of each protein in each SMA type at T0 were shown by violin plots on the right side. Box plots showing interquartile range, and outliers were shown by black dots

identified as key proteins that distinctly separate the most severe form SMA1 from the milder ones (Fig. 2b).

We checked the correlations between age and the 9 proteins above (Supplementary Table 1). Except SERPINF2 and CTSF, 7 of them show a significant moderate correlation with age ($p_{\text{adj}} < 0.05$). Among those, only GOLM1 ($\rho = -0.45$) had negative correlation, while the other 6 proteins, HSP90B1 ($\rho = 0.61$), CNTN1 ($\rho = 0.48$), B4GAT1 ($\rho = 0.77$), PVR ($\rho = 0.56$), NRXN3 ($\rho = 0.47$), and C16orf89 ($\rho = 0.55$) revealed positive associations.

Nusinersen-driven changes in proteomic profile at T302

Proteome analysis at T302 identified 1399 unique proteins. Initially, all samples from each group, independent from their response status, were incorporated into the analysis of

protein changes between T0 (before treatment) and T302 (after 302 days of treatment). When PCA plots were generated using different pairs of axes, more distinct separation was observed when examining dimensions other than first two, as PCA results can be influenced by biases originating from any of the preceding steps, starting from data collection to the end of data processing [23] (Fig. 3a-c).

When the significant changes in proteomic profiles in response to nusinersen-treatment were examined between T0 and T302 for each SMA subtype individually, 147, 135, and 289 proteins were founded, changing significantly in SMA1, SMA2, and SMA3, respectively. Volcano plots highlight downregulated and upregulated proteins at T302 compared to the baseline (Fig. 3d-f), and the heatmaps show hierarchical clustering of significantly changing proteins (Suppl. Figure 1a-c).

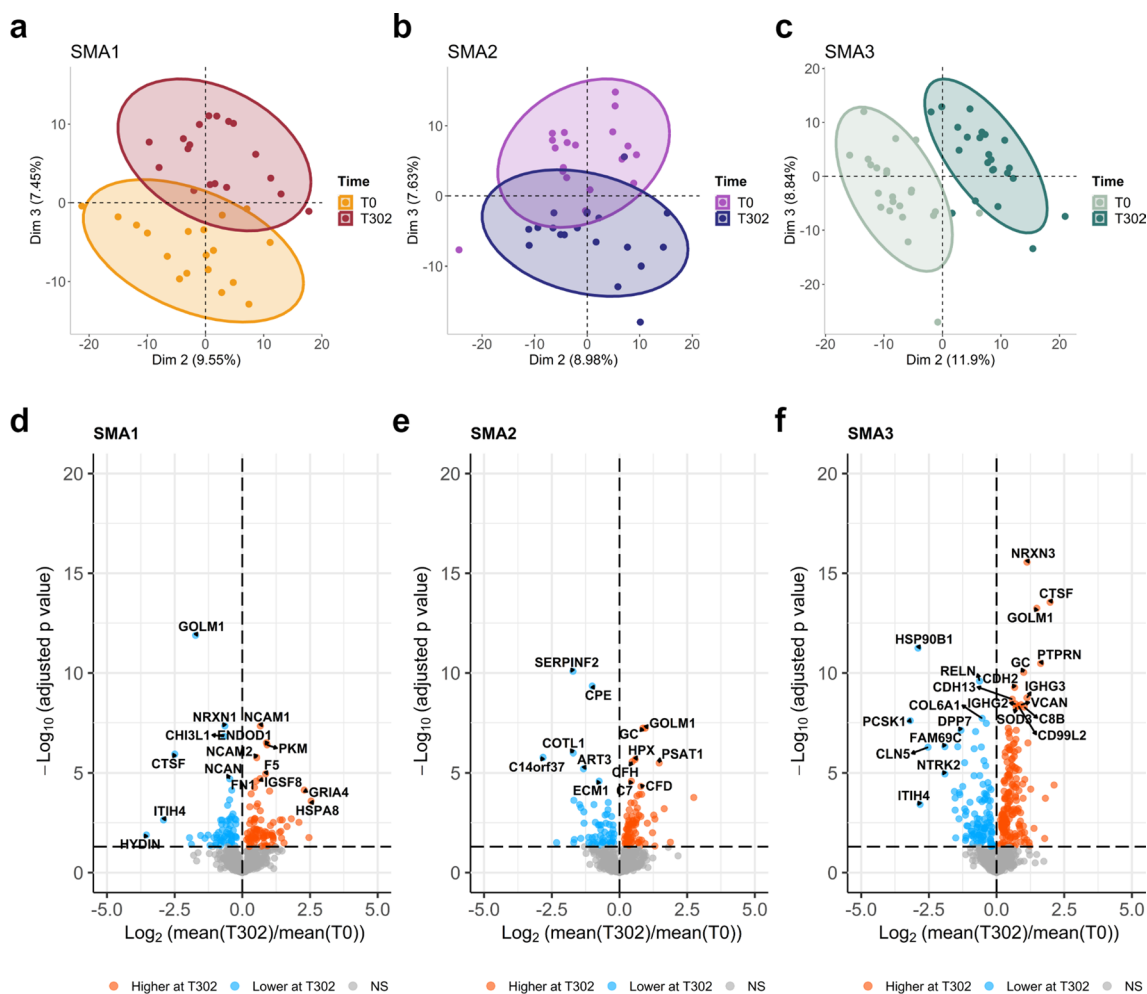


Fig. 3 Nusinersen modulates CSF proteome at T302. The change in clustering based on proteomics between T0 and T302 is shown for (a) SMA1, (b) SMA2, and (c) SMA3 patients, based on second- and third-dimension combinations. PCA plots showing distribution and clustering of samples, and ellipses correspond to 95% confidence intervals

for each time point. Volcano plots are used to show the down- or up-regulated proteins after nusinersen treatment for (d) SMA1, (e) SMA2, and (f) SMA3 samples. Significantly changed proteins ($p_{\text{adj}} < 0.05$) were colored on volcano plots based on their higher (orange) or lower (blue) abundances at T302 compared to T0

Gene ontology (GO) enrichment analysis of biological processes (BP) was performed to investigate enriched pathways of differentially expressed protein (DEPs) after nusinersen treatment. Overall, a common response across all SMA subtypes was observed for proteins down-regulated by the treatment, with the main top enriched pathways related to axonogenesis (GO terms: regulation of axonogenesis, axon guidance, axon development (Fig. 4a-c). Axonogenesis related proteins included components of the semaphorin family (SEMA6A, SEMA7A, SEMA4B), leucine rich repeat and Ig domain containing 1 (LINGO1), plexin-B2 (PLXNB2), L1 cell adhesion molecule (L1CAM), multiple EGF like domains 8 (MEGF8), dystroglycan 1 (DAG1), and reelin (RELN).

Other than the axonogenesis, SMA1 subjects showed platelet degranulation and aminoglycan metabolic processes within the top modulated BPs for down-regulated DEPs.

Pathway enrichment analysis of proteins up-regulated after nusinersen treatment, revealed disease-specific results for SMA1 and a similar signature for the milder phenotypes SMA2 and SMA3 (Fig. 4d-f).

In SMA1 patient's proteome after treatment, up-regulated DEPs were related to glucose metabolism (GO terms: canonical glycolysis, NADH regeneration, glucose catabolic processes to pyruvate), including proteins as pyruvate kinase (PKM), enolase 1 (ENO1), aldolase, fructose-bisphosphate C (ALDOC), phosphoglyceratemutase 1 (PGAM1).

In SMA2 and SMA3 proteomes, up-regulated DEPs mainly belonged to two highly intertwined cascades,

namely the coagulation and the complement pathways. In details, proteome of SMA2 subjects after treatment revealed increased coagulation processes (GO terms: protein coagulation cascade, blood coagulation and fibrinolysis), identified by proteins as fibulin1 (FBLN1), coagulation factor II, IX and XII (F1, F9, F12), kininogen1 (KNG1), fibronectin 1 (FN1), serpin family G member 1 (SERPING1), fibrinogen A (FGA), and apolipoprotein E (APOE).

SMA3 proteome after treatment showed enhanced complement activation (GO terms: regulation of complement activation and humoral immune response), including proteins as complement C1q B and C chain (C1QB, C1QC), complement C8 beta chain (C8B), complement molecules (C2, C3, C4A, C6, C7), complement C1r (C1R), complement C1s (C1S), complement Factor I (CFI), clusterin (CLU), and serpin family G member 1 (SERPING1). Complement activation (GO terms: complement activation, classical pathway) also appeared in SMA1 pathway enrichment but was not as significant for this subtype. Proteins included were C1S, CFI, C2, and C4A.

Proteome changes in responder versus non-responder patients

In addition to assessing the temporal changes in protein levels for each SMA subtype using all patient samples, we conducted a comparison of protein fold changes between baseline (T0) and after 302 days of treatment (T302) in responder (R) and non-responder (non-R) patients.

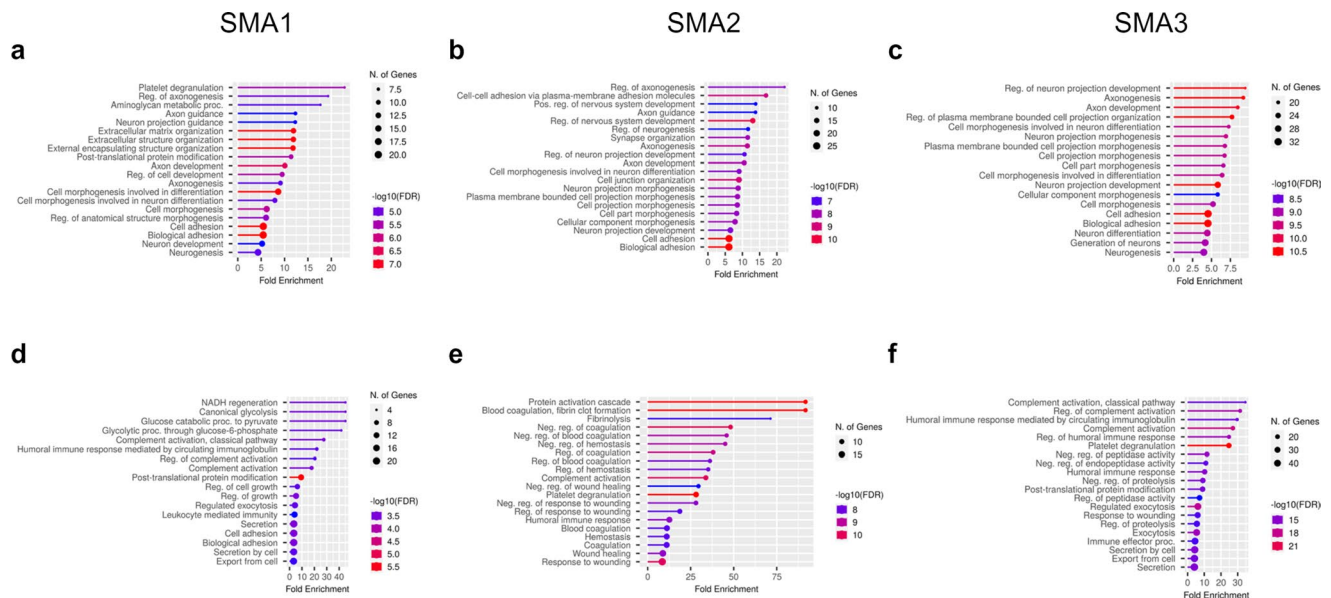


Fig. 4 Biological processes modulated by nusinersen at T302. Top 20 GO pathways for biological processes (BP) enriched by mainly down-regulated (a-c) and up regulated DEPs (d-f), obtained from two clusters from each heatmap, are shown by lollipop plots for each SMA type. Fold enrichment represents the ratio of the percentage of

matched proteins in the query list with pathway-associated proteins to the percentage of query with the background. Bar colors are corresponding to the false discovery rate (FDR) corrected p values for each enriched pathway, and the size of dots are proportional to the number of proteins associated with the respective pathway

Based on outcome measures improvement at T302, we classified as responders 52.6% of SMA1, 33.3% of SMA2, and 26.1% of SMA3. When we check the importance of response status in treatment-based protein level changes by PERMANOVA test for each SMA subtype, only SMA1 had a significance close to threshold ($F = 1.2544$, $p = 0.085$), while the protein changes were determined to be similar between R and non-R in SMA2 ($F = 0.8361$, $p = 0.739$) and SMA3 ($F = 1.0221$, $p = 0.416$) groups. Accordingly, principal components analysis showed better separation of treatment-based CSF proteome change between R and non-R patients in the SMA1 group compared to the SMA2 and SMA3 groups (Fig. 5a-c-d).

The comparison of changes in proteome expression between R and Non-R patients identified 45 DEPs in the SMA1 group, 26 in SMA2, and 26 in SMA3. Volcano plot highlighting downregulated and upregulated proteins at T302 in R vs. non-R was therefore depicted for SMA1 only (Fig. 5b). Among proteins upregulated in SMA1 Responders after treatment, we highlight the cellular communication network factor 3 (NOV/CCN3, $p = 0.0005$), involved

in immunomodulatory processes via T regulatory cells [24], the SPOCK1 ($p = 0.00094$), an extracellular proteoglycan that belongs to the secreted protein acidic and rich in cysteine (SPARC) family and is involved in the regulation of blood brain barrier (BBB) permeability [25], and the heat shock proteins HSPA1A/HSPA1B ($p = 0.002$), which are highly conserved cellular response to internal and external stress [26]. Fibroblast growth factor receptor 2 (FGFR2, $p = 0.002$), involved in lymphocyte and macrophage/microglia infiltration as well as myelin and axon degeneration [27], was instead in the pool of downregulated proteins.

Discussion

Advancements in omics technologies have significantly expanded our understanding and exploration of biological mechanisms, facilitating biomarker discovery in neuroscience [28–31]. Machine learning algorithms, a subset of artificial intelligence, prove particularly effective in handling omics datasets, capturing intricate patterns that might be

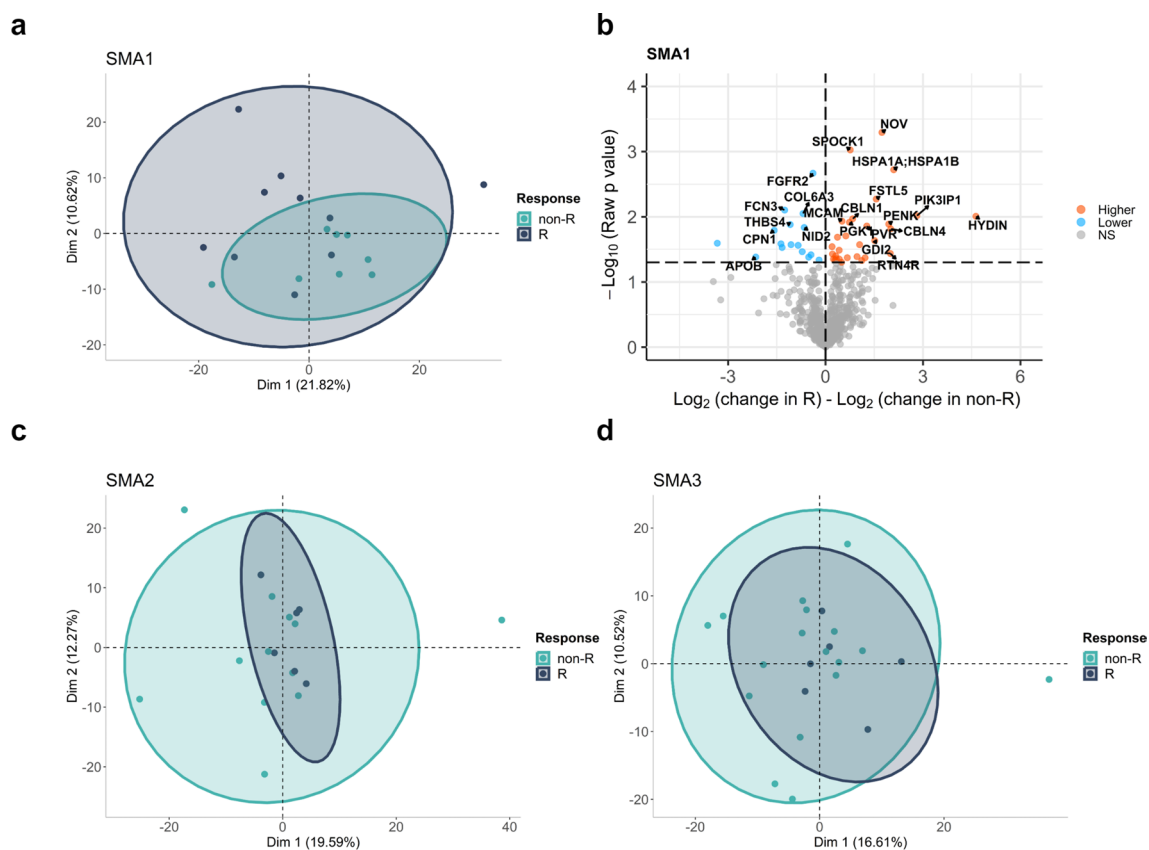


Fig. 5 Responders vs. non-Responders analysis. The clustering patterns of the responders (green) and non-responders (dark grey) based on changes in protein levels after nusinersen treatment were shown by PCA plots based on first two dimensions for (a) SMA1, (c) SMA2, and (d) SMA3. Each dot represents one sample, and ellipses correspond to 95% confidence intervals for each of the subtypes. (b) Volcano

plot for comparison of time-wise changes in protein levels in each responder and non-responder samples. Fold change differences were calculated by subtracting log_2 fold change of proteins between T302 and T0 in non-responders from log_2 fold change between T302 and T0 in responders, and higher and lower changes in responders compared to non-responders were shown by orange and blue colors, respectively

overlooked by traditional statistical methods [19–22, 32, 33].

In the first part of this study, we exploited machine learning to discover novel prognostic biomarkers for SMA. The Random Forest algorithm, applied to the protein expression profile of naïve patients, identified different pivotal proteins capable of classifying the different SMA types based on their differential abundance. Among them, CNTN1 and NRXN3 clearly differentiated the severe form from the milder ones.

CNTN1, one of the axo-glial adhesion molecules, showed the lowest expression levels in SMA1 and higher expression in SMA2 and SMA3 patients. CNTN1 is expressed both in the central and peripheral nervous system [34] and is primarily located at the paranodes, where it forms a tripartite complex contributing to the stabilization of the connection between the axon and the myelin loops [35]. Abnormal intercellular communication between axons and Schwann cells have been recently associated with immature, dysfunctional and vulnerable motor axons in severe SMA patients' human tissues and in a severe SMA mouse model (SMA Δ 7 mice) [36].

Moreover, genetic ablation of *Cntn1* in a zebrafish model results in infertility, reduced animal size, ataxic swimming behavior, and a curved spine, with hypomyelinated neurons compared to wild-type [37]. Similarly, the *Cntn1*^{-/-} mice, display failure to thrive and severe locomotor abnormalities, along with a decreased number of myelinated axons in both the spinal cord and the optic nerve [37]. In the *Smn1*^{-/-}*Smn2*^{+/-} mouse model, a transcriptomic analysis identified the *tifa-il6-cntn1* pathway as one of the top upregulated pathways in the central nervous system, where *Cntn1* might serve as a protective factor for neurons [38]. Together, these findings point to a crucial role of CNTN1 in the developing nervous system, with possible implications in the pathogenesis of SMA.

NRXN3, a member of the neurexin family (NRXN1, 2 and 3), presynaptic cell adhesion molecules necessary for synapse formation and maintenance [39–41], exhibited a progressive increase in expression levels, with the lowest levels observed in SMA1 patients and the highest in SMA3 patients at baseline. Interestingly, a reduction or altered *Nrxn2* splicing have been observed in both transgenic SMA mice and zebrafish [42], while in vitro studies in motor neurons derived from SMA patients' iPSCs showed a reduction of NRXN2 expression compared to controls, while an over-expression of NRXN2 improved the SMA motor neuron survival, and increase neurite length, suggesting it might act as a gene modifier in SMA [43].

Therefore, considering the role that CNTN1 and NRXN3 play in the nervous system, we foresee their potential as prognostic biomarkers to distinguish severe SMA1 from milder phenotypes in the context of newborn screening

programs, especially in the management of asymptomatic patients. A moderate correlation was found between these markers and the age at baseline, so that future studies to ensure the robustness of our results should focus on larger cohorts with a more uniform age distributions.

To translate our findings into clinical applications, targeted assays such as Enzyme-Linked Immuno Assay (ELISA) for the identified proteins, validation in independent cohorts, and evaluation in more accessible samples like serum are underway.

In the second part of the study, we showed that, at T302, nusinersen significantly modulates the CSF proteome. Our aim was to provide insights about nusinersen-dependent protein changes at the biological processes level, rather than focusing on individual differentially expressed proteins. Through Gene ontology (GO) analysis of biological processes, we found that nusinersen causes both a common effect in all types of SMA, regardless of the severity, and specific effects in the different forms.

Specifically, nusinersen modulated biological processes related to axonogenesis and axon development in all SMA patients. SMN1 defects have been shown to significantly affect motor neuron axon development, maturation, and function in type 1 SMA patients and in a mouse model of severe SMA [36]. A marked delay of SMA motor neuron axon radial growth, which prevents a correct interaction with Schwann cells, have been demonstrated and associated with axon degeneration and loss after birth [36]. As a result, indirect markers of axon degeneration are released by suffering motor neuron and detectable in CSF and bloodstream, such as phosphorylated neurofilament heavy chain (pNfH) and neurofilament light chain (NfL), which are increased both in pre-symptomatic SMA patients with 2 SMN2 copies and symptomatic SMA patients [10, 44–46].

Our findings indicate that nusinersen modulates axonogenesis processes, in line with recent electrophysiological studies demonstrating partial enhancements in axonal maturation in SMA patients undergoing nusinersen treatment [47, 48]. Currently, no biochemical biomarkers capturing a reduction in axon deterioration upon nusinersen treatment are available, except for neurofilaments, which dosage, however, has produced discordant results [10, 45, 46, 49–52].

Among the differentially expressed proteins belonging to the axonogenesis GO term, we detected components of the semaphorine family, in particular SEMA4B, SEMA6A, SEMA7A which were downregulated by nusinersen treatment in all patients. These results are consistent with recent untargeted proteomic studies that identified a reduction in SEMA6A and SEMA7A levels in the CSF of SMA3 patients after 22 months of nusinersen treatment [15], as well as the downregulation of SEMA7A in responders SMA3 subjects

after 10 months of nusinersen [11]. Our data, combined with the previous ones, suggest semaforins, especially SEMA6A and SEMA7A, as a new class of biomarkers for monitoring nusinersen treatment in all SMA patients, regardless of their severity, reflecting axon maintenance and remodeling, as already emerged in other condition [53].

Beyond axonogenesis, nusinersen showed modulation of bioenergetic pathways in SMA1 and inflammatory pathways in SMA2 and SMA3. Previously, Errico et al. reported nusinersen affecting the glucose metabolism selectively in the CSF of SMA1 patients through NMR-metabolomic analysis [13]. Here, we confirm and strengthen these results at a protein level, showing a nusinersen-mediated boost in the protein expression of glycolytic enzymes such as PKM, ENO1, ALDOC, and PGAM1. SMA1 patient's reduced tolerance to fasting is well known [54–56], and an imbalance in glucose metabolism has been reported in SMA2 [57]. Our results suggest a correction of glucose metabolism by nusinersen, although additional studies on blood samples will be necessary to understand whether this effect is specific to the CSF or systemic.

Moreover, with regard to the effects of nusinersen in SMA2 and SMA3 proteome, we report an increase in proteins belonging to the coagulation and the complement cascades, highly interconnected processes [58, 59], sharing common ancestral pathways [60], and contributing to a complex inflammatory network [61]. Furthermore, mediators of the coagulation cascade act as inhibitors/activators of the complement cascade, and vice versa [59]. In our results, SERPING1, an inhibitor of the classical complement cascade, was upregulated in both SMA2 and SMA3 at T302. In SMA2, an upregulation of F12, a potent activator of the C1q molecule in the classical pathway, was noted, while SMA3 showed an increase of several complement cascade components such as C1q, C1s, C2, C3, C4A, C6 and C7. Recently, Faravelli et al. reported that members of the innate immune complement pathways, including SERPING1, were among the most abundant proteins in the CSF of SMA3 patients as detected by untargeted LC-MS [15]. These findings align with recent data suggesting that the complement cascade activation participates to neuroinflammatory processes and contributes to diseases pathogenesis in various neurodegenerative disorders, including Parkinson's, Alzheimer's and Huntington's diseases, Amyotrophic Lateral Sclerosis and SMA [62–65]. As such, different clinical trials investigating complement therapeutic targets are ongoing in neurologic disorders [65]. Moreover, in a pre-clinical SMA model, the aberrant upregulation of the classical complement proteins C1q and C3, was associated with dysfunction and selective elimination of synapses, and an *in vivo* treatment with an anti-C1q antibody induced a rescue of synapses [64].

In addition to complement activation, increased inflammatory mediators have been reported in both CSF and serum in preclinical models and in SMA patients, indicating that immune dysfunction and neuroinflammation processes are involved in the pathogenesis of SMA [38, 66–69]. In this context, we noted a decreased expression of inflammatory markers, including communication network factor 3 (NOV/CCN3), heat shock proteins HSPA1A/HSPA1B, and fibroblast growth factor receptor 2 (FGFR2), at T302 in SMA1 Responders, suggesting that treatment with nusinersen promotes better control of pro-inflammatory processes.

In summary, this untargeted LC-MS proteomic profiling in the CSF of SMA patients revealed differences in protein expression in naïve patients and showed nusinersen-related modulation in several biological processes after 10 months of treatment.

We acknowledge some limitations of the study, including the absence of CSF samples from healthy controls and untreated SMA patients, the relatively small sample size within each SMA subtype, the lack of an assay of widely used biomarkers such as NfL to parallel the results obtained, and the limited follow-up period of 10 months, that may constrain the robustness of the conclusions. Further validation in larger, independent cohorts, with uniform age distributions and long-term monitoring, is crucial to confirm the prognostic value of the identified biomarkers and better understand the duration of the observed treatment effects.

Nevertheless, despite these limitations, the integration of proteomics and machine learning in a large cohort of patients has enabled the identification of novel prognostic CSF protein biomarkers for disease severity stratification, providing a new strategy to help guiding precision therapeutic treatment.

Additional confirmatory studies on CSF and serum samples from other SMA patients and healthy controls, even after a longer treatment period, are needed to validate these results.

Methods

Patients characteristics

Genetically confirmed SMA patients treated with nusinersen, have been enrolled in the study from 4 tertiary Italian neuromuscular centers (IRCCS Gaslini Institute-Genoa, Nemo Center-Milan, Ospedale Bambino Gesù-Rome, and Policlinico G. Martino-Messina). SMN2 copy numbers were also available for all but one patient.

All SMA1 patients participated in the compassionate Expanded Access Program (EAP) [70], while SMA2 and SMA3 patients received treatment in a real-world

setting. Nusinersen was administered following the standard protocol.

On the day of nusinersen administration, anthropometric measures, information on ventilatory support (need for non-invasive ventilation or tracheostomy), and nutritional route (oral, nasogastric tube, or gastrostomy) were noted, and functional outcome measures were performed by trained physical therapists at each site (Table 1). SMA1 patients were scored by the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND) scale [71], while SMA2 and SMA3 patients were evaluated through the Hammersmith Functional Motor Scale – Expanded (HFMSE) [72]. Patients were categorized into Responders (R) and non-Responders (non-R) based on changes in motor function scores from T0 to T302, following previously described criteria. Specifically, SMA1 patients showing an increase in the CHOP score by 4 points, and SMA2-SMA3 subjects achieving a 3-point improvement on the HFMSE scale at T302 were defined Responders [73, 74].

The study was conducted in accordance with the Declaration of Helsinki and ICH GCP guidelines. All subjects and guardians provided written informed consent for the analysis of biological samples, following the approval of local ethics committees (Prot. SMALiQ_2022, 503/2022 - id 12319).

Sample preparation and proteomic analysis

On the day of administration, CSF samples were collected after a fasting period of 4–6 h and stored at -80°C at each site. For the purposes of this study, only CSF collected at baseline (T0; before the first nusinersen intrathecal administration) and at day 302 (T302; before the sixth infusion), were centralized at IRCCS Gaslini, and subsequently subjected to proteomic analysis.

One hundred μl of CSF were denatured, reduced and alkylated in 100 μl of iST-LYSE buffer (PreOmics) for 10 min at 95°C , 1000 rpm. Proteins were isolated by PAC method [75]. Briefly, protein aggregation was induced by addition of 70% ACN and 200 μg of magnetic beads were added to capture aggregated proteins. Magnetic beads were retained by magnet and the supernatant was removed. Beads were washed one time with 1 ml acetonitrile, followed by two wash with 1 ml 70% ethanol. Washed beads were resuspended in 100 μl TRIS 25 mM pH 8 and captured proteins were digested O.N. at 37°C with 0.7 μg Trypsin and 0.3 μg LysC. Obtained peptides were desalted in Stage-Tips [76] and analyzed by a nano-Ultra High Performance Liquid Chromatography-Tandem Mass Spectrometry system (nano-UHPLC-MS/MS) using an Ultimate 3000 RSLC coupled to an Orbitrap Fusion Tribrid mass spectrometer.

Elution was performed with an EASY spray column ($75\ \mu\text{m} \times 50\ \text{cm}$, $2\ \mu\text{m}$ particle size, Thermo Scientific) at a flow rate of 250 nL/min with a 100 min non-linear gradient consisting of an increase from 7 to 27% solution B (80% ACN and 20% H_2O , 5% DMSO, 0.1% FA) in 57 min, with a further increase to 45% B in 15 min, followed by 1-minute wash at 80% B and a 20 min re-equilibration at 2% B.

MS scans were acquired at a resolution of 120,000 between 375 and 1,500 m/z and an AGC target of 4.0×10^5 , 50 ms maximum injection time. Advanced Peak Determination was enabled for MS1 measurements. MS/MS spectra were acquired in the linear ion trap (rapid scan mode) with an AGC target of 3.0×10^4 and a 30 ms maximum injection time. For precursor selection, we prioritized the least abundant signals in the three ranges 375–575 m/z , 574–775 m/z and 774–1500 m/z . Quadrupole isolation with a 0.7 m/z isolation window was used, and Dynamic Exclusion was set at 25 s. HCD was performed using 30% collision energy.

MaxQuant software version 1.6.17.0 was used to process data [77]. A false discovery rate was set at 0.01 for the identification of proteins, peptides and peptide-spectrum match (PSM). A minimum of 6 amino acids was required for peptide identification. Andromeda engine, incorporated into MaxQuant software, was used to search MS/MS spectra against Uniprot human database (release UP000005640_9606 December 2020). In the processing the variable modifications are Acetyl (Protein N-Term), Oxidation (M), Deamidation (NQ), on the contrary, the Carbamidomethyl (C) was selected as fixed modification. Quantification intensities were calculated by the default fast MaxLFQ algorithm with the activated option 'match between runs'.

Machine learning algorithm

The Random Forest (RF) machine learning algorithm, previously applied to metabolomics and proteomics data for classification and regression analysis [78, 79], was adopted to classify SMA types based on protein expression levels at baseline. Random Forest models were trained for classification of SMA subtypes using the proteomics data obtained at T0. All steps were carried in "R" free software using "caret" (ver. 6.0–94) and "randomForest" (ver. 4.7–1.1) packages. The detailed strategies employed during optimization and prevention of overfit and data leakage are described in the Supplementary Methods. All the steps were carried out individually, with proteomics data only. To investigate the effects of different hyperparameter tuning strategies on obtained classification accuracies, data partitioning ratio, mtry (number of features to be randomly selected at each split), number of trees, and maximum number of nodes were assigned to different sets of values. Besides varying sets of parameters, addition of cross-validation and post-processing (normalization, transformation, scaling) of

training and test data after partitioning were included in optimizations to avoid potential data loss. At the end of optimizations, normalized and scaled T0 proteomics data were used, and partitioned into train and test datasets in 80% and 20% ratios, respectively. Model training was carried out using 1000 trees and maximum number of nodes, set to 3. Classification accuracy of the trained model on test data was calculated by comparing the real and predicted SMA classes. List of important variables were recorded for the trained model by extracting top 30 features sorted by mean decrease in Gini coefficient. Each step, ranging from data partitioning to the selection of important variables, was repeated 100 times, with the seed set to numbers between 1 and 100. Finally, a set of features common in all lists of top 30 important variables from 100 trained models was obtained.

Statistical analysis

Log₂-transformed and median-centered intensity values were used for downstream statistical analyses. PERMANOVA with Euclidean distance method was used to test significance of subtype, sex, age, smn2 gene copy number, and BMI. For comparison of protein compositions of SMA1, SMA2, and SMA3 subtype samples before treatment, analysis of covariance (ANCOVA) was used by adjusting for age. Features with raw p value less than 0.05 were evaluated as significant and used in clustering for each pairwise comparison. Before and after treatment comparisons of SMA1, SMA2, and SMA3 patients were carried out individually, using paired two-sided t-test, and obtained p values were adjusted for multiple hypothesis testing using Benjamini-Hochberg method. Log₂ fold changes in responders and non-responders were calculated by subtracting the values at T302 from T0 for each SMA subtype, and unpaired two-sided t-test was used for the comparison. Raw p values were used for significance. For each pairwise comparison results, volcano plots were produced using EnhancedVolcano (ver. 1.16.0). Principal component analysis (PCA) was carried out using FactoMineR (ver. 2.9), and visualizations with different pairs of dimensions were obtained using factoextra package (ver. 1.0.7).

For correlation analysis between age and 9 biomarker candidates, Spearman correlation was performed and obtained p values were adjusted using Benjamini-Hochberg method.

Hierarchical clustering

Intensity data of the filtered features were scaled to have a mean value of zero with standard deviation of one. The Euclidean distances were calculated between samples and between proteins using dist function, and then, Ward-D2 clustering was performed for samples and for proteins by hclust function. Heatmaps were plotted via heatmap.2 function in gplots

package (ver. 3.1.3). In order to divide clustered into two clusters, mainly corresponding to increasing and decreasing proteins for each pairwise comparison, cutree function with k=2 was used.

Pathway enrichment

Gene Ontology (GO) enrichment analysis was performed for proteins in each cluster, separately. For that purpose, respective gene symbols of the proteins in each cluster were converted into Entrez ID's by mapIds function in AnnotationDbi package org. Hs.eg.db (ver. 3.14.0). If Entrez ID could not be obtained for a protein, and additional manual curation step was performed using protein ID as query on UniProt website (www.uniprot.org/uniprotkb/<protein.id>/entry, <https://doi.org/10.1093/nar/gkh131>). If the protein IDs for a single protein entry correspond to different proteins, it was excluded from enrichment analysis. For each clusters, obtained list of Entrez IDs were submitted to ShinyGO (ver 0.76, <http://bioinformatics.sdstate.edu/go/>, <https://doi.org/10.1093/bioinformatics/btz931>) after selecting "Human" as the organism with other default settings. Enrichment was performed using three different GO pathway databases; "GO Biological Process", "GO Cellular Compartment", and "GO Molecular Function". Tables for all enriched pathways and lollipop plots were exported.

In addition to the GO enrichment analysis, the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of biological processes (BP) was performed. The results are provided as Supplementary Fig. 2a-f.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00018-024-05426-6>.

Acknowledgements ES was scholar of TUBITAK 2244 Industrial PhD between 2019 and 2023 (project No: 118C039). CP, MP, AD, EB, CB are members of the European Reference Network for Neuromuscular Diseases (ERN-NMD).

Author contributions CP, SC, NB, MP, AD, MP, MS, SM, EA, VAS, EM, EB, and CB were involved in patients' follow-up, CSF collection and drug administration. SB, AP, EP, and MB performed laboratory assessment with AP supervision. CP, CB, MB, AP, ES and US analyzed the data and wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding This work was partly supported by Ricerca Corrente 2021–2022 (Italian Ministry of Health).

Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement.

Data availability The mass spectrometry data have been deposited to the ProteomeX-change Consortium via the PRIDE partner repository with the dataset identifiers: PXD047529 (Reviewer account details: Username: reviewer_pxd047529@ebi.ac.uk Password: oRxG3oEc).

Declarations

Ethics approval and consent to participate The study was conducted in accordance with the Declaration of Helsinki and ICH GCP guidelines. All subjects and guardians provided written informed consent for the analysis of biological samples, following the approval of local ethics committees (Prot. SMALiQ_2022, 503/2022 - id 12319).

Consent for publication Authors gave consent for the publication of the manuscript.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Mercuri E, Sumner CJ, Muntoni F, Darras BT, Finkel RS (2022) Spinal muscular atrophy. *Nat Rev Dis Primers* 8:52. <https://doi.org/10.1038/s41572-022-00380-8>
- Feldkötter M, Schwarzer V, Wirth R, Wienker TF, Wirth B (2002) Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am J Hum Genet* 70:358–368. <https://doi.org/10.1086/338627>
- Calucho M, Bernal S, Alías L, March F, Venceslá A, Rodríguez-Álvarez FJ, Aller E, Fernández RM, Borrego S, Millán JM, Hernández-Chico C, Cuscó I, Fuentes-Prior P, Tizzano EF (2018) Correlation between SMA type and SMN2 copy number revisited: an analysis of 625 unrelated Spanish patients and a compilation of 2834 reported cases. *Neuromuscul Disord* 28:208–215. <https://doi.org/10.1016/j.nmd.2018.01.003>
- Monani UR, Lorson CL, Parsons DW, Prior TW, Androphy EJ, Burghes AH, McPherson JD (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum Mol Genet* 8:1177–1183. <https://doi.org/10.1093/hmg/8.7.1177>
- Finkel RS, Chiriboga CA, Vajsar J, Day JW, Montes J, De Vivo DC, Yamashita M, Rigo F, Hung G, Schneider E, Norris DA, Xia S, Bennett CF, Bishop KM (2016) Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet* 388:3017–3026. [https://doi.org/10.1016/S0140-6736\(16\)31408-8](https://doi.org/10.1016/S0140-6736(16)31408-8)
- Mercuri E, Baranello G, Boespflug-Tanguy O, De Waele L, Goe-mans N, Kirschner J, Masson R, Mazzone ES, Pechmann A, Pera MC, Vuillerot C, Bader-Weder S, Gerber M, Gorni K, Hoffart J, Kletzl H, Martin C, McIver T, Scalco RS, Yeung WY, Servais L, SUNFISH Working Group (2023) Risdiplam in types 2 and 3 spinal muscular atrophy: a randomised, placebo-controlled, dose-finding trial followed by 24 months of treatment. *Eur J Neurol* 30:1945–1956. <https://doi.org/10.1111/ene.15499>
- Mercuri E, Muntoni F, Baranello G, Masson R, Boespflug-Tanguy O, Bruno C, Corti S, Daron A, Deconinck N, Servais L, Straub V, Ouyang H, Chand D, Tauscher-Wisniewski S, Mendonca N, Lavrov A, STRIVE-EU study group (2021) Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy type 1 (STRIVE-EU): an open-label, single-arm, multicentre, phase 3 trial. *Lancet Neurol* 20:832–841. [https://doi.org/10.1016/S1474-4422\(21\)00251-9](https://doi.org/10.1016/S1474-4422(21)00251-9)
- De Vivo DC, Bertini E, Swoboda KJ, Hwu WL, Crawford TO, Finkel RS, Kirschner J, Kuntz NL, Parsons JA, Ryan MM, Butterfield RJ, Topaloglu H, Ben-Omran T, Sansone VA, Jong YJ, Shu F, Staropoli JF, Kerr D, Sandrock AW, Stebbins C, Petrillo M, Braley G, Johnson K, Foster R, Gheuens S, Bhan I, Reyna SP, Fradette S, Farwell W, NURTURE Study Group (2019) Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: interim efficacy and safety results from the phase 2 NURTURE study. *Neuromuscul Disord* 29:842–856. <https://doi.org/10.1016/j.nmd.2019.09.007>
- Strauss KA, Farrar MA, Muntoni F, Saito K, Mendell JR, Servais L, McMillan HJ, Finkel RS, Swoboda KJ, Kwon JM, Zaidman CM, Chiriboga CA, Iannaccone ST, Krueger JM, Parsons JA, Shieh PB, Kavanagh S, Wigderson M, Tauscher-Wisniewski S, McGill BE, Macek TA (2022) Onasemnogene Abeparvovec for presymptomatic infants with three copies of SMN2 at risk for spinal muscular atrophy: the phase III SPRINT trial. *Nat Med* 28:1390–1397. <https://doi.org/10.1038/s41591-022-01867-3>
- Rich KA, Fox A, Yalvac M, Heintzman S, Tellez M, Bartlett A, Severyn S, Linsenmayer M, Kelly K, Reynolds J, Sterling GB, Weaver T, Rajneesh K, Pino MG, Arnold WD, Elsheikh B, Kolb SJ (2022) Neurofilament levels in CSF and serum in an adult SMA cohort treated with Nusinersen. *J Neuromuscul Dis* 9:111–119. <https://doi.org/10.3233/JND-210735>
- Kessler T, Latzer P, Schmid D, Warnken U, Saffari A, Ziegler A, Kollmer J, Möhlenbruch M, Ulfert C, Herweh C, Wildemann B, Wick W, Weiler M (2020) Cerebrospinal fluid proteomic profiling in nusinersen-treated patients with spinal muscular atrophy. *J Neurochem* 153:650–661. <https://doi.org/10.1111/jnc.14953>
- Bianchi L, Sframeli M, Vantaggiato L, Vita GL, Ciranni A, Polito F, Oteri R, Gitto E, Di Giuseppe F, Angelucci S, Versaci A, Messina S, Vita G, Bini L, Aguenouz M (2021) Nusinersen Modulates Proteomics Profiles of Cerebrospinal Fluid in spinal muscular atrophy type 1 patients. *Int J Mol Sci* 22:4329. <https://doi.org/10.3390/ijms22094329>
- Errico F, Marino C, Grimaldi M, Nuzzo T, Bassareo V, Valsecchi V, Panicucci C, Di Schiavi E, Mazza T, Bruno C, D'Amico A, Carta M, D'Ursi AM, Bertini E, Pellizzoni L, Usiello A (2022) Nusinersen induces Disease-Severity-Specific Neurometabolic effects in spinal muscular atrophy. *Biomolecules* 12:1431. <https://doi.org/10.3390/biom12101431>
- Schorling DC, Kölbl H, Hentschel A, Pechmann A, Meyer N, Wirth B, Rombo R, SMARtCARE consortium, Sickmann A, Kirschner J, Schara-Schmidt U, Lochmüller H, Roos A (2022) Cathepsin D as biomarker in cerebrospinal fluid of nusinersen-treated patients with spinal muscular atrophy. *Eur J Neurol* 29:2084–2096. <https://doi.org/10.1111/ene.15331>
- Faravelli I, Gagliardi D, Abati E, Meneri M, Ongaro J, Magri F, Parente V, Petrozzi L, Ricci G, Farè F, Garrone G, Fontana M, Caruso D, Siciliano G, Comi GP, Govoni A, Corti S, Ottoboni L (2023) Multi-omics profiling of CSF from spinal muscular atrophy type 3 patients after nusinersen treatment: a 2-year follow-up multicenter retrospective study. *Cell Mol Life Sci* 80:241. <https://doi.org/10.1007/s00018-023-04885-7>
- Beaudin M, Kamali T, Tang W, Hagerman KA, Dunaway Young S, Ghiglieri L, Parker DM, Lehallier B, Tesi-Rocha C, Sampson



- JB, Duong T, Day JW (2023) Cerebrospinal fluid proteomic changes after Nusinersen in patients with spinal muscular atrophy. *J Clin Med* 12:6696. <https://doi.org/10.3390/jcm12206696>
17. Babić M, Banovi M, Berečić I, Banić T, BabićLeko M, Ulašević M, Junakovi A, Kopic J, Sertić J, Barišić N, Šimić G (2023) Molecular biomarkers for the diagnosis, prognosis, and pharmacodynamics of spinal muscular atrophy. *Journal of clinical medicine*. *J Clin Med* 12:5060. <https://doi.org/10.3390/jcm12155060>
 18. Meneri M, Abati E, Gagliardi D, Faravelli I, Parente V, Ratti A, Verde F, Ticozzi N, Comi GP, Ottoboni L, Corti S (2023) Identification of novel biomarkers of spinal muscular atrophy and therapeutic response by Proteomic and Metabolomic Profiling of Human Biological Fluid samples. *Biomedicines* 11:1254. <https://doi.org/10.3390/biomedicines11051254>
 19. Chang CH, Lin CH, Lane HY (2021) Machine learning and novel biomarkers for the diagnosis of Alzheimer's Disease. *Int J Mol Sci* 22:2761. <https://doi.org/10.3390/ijms22052761>
 20. Mégret L, Mendoza C, Arrieta Lobo M, Brouillet E, Nguyen TT, Bouaziz O, Chambaz A, Néri C (2022) Precision machine learning to understand micro-RNA regulation in neurodegenerative diseases. *Front Mol Neurosci* 15:914830. <https://doi.org/10.3389/fnmol.2022.914830>
 21. Xu J, Mao C, Hou Y, Luo Y, Binder JL, Zhou Y, Bekris LM, Shin J, Hu M, Wang F, Eng C, Oprea TI, Flanagan ME, Pieper AA, Cummings J, Leverenz JB, Cheng F (2022) Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. *Cell Rep* 41:111717. <https://doi.org/10.1016/j.celrep.2022.111717>
 22. Catanese A, Rajkumar S, Sommer D, Masrori P, Hersmus N, Van Damme P, Witzel S, Ludolph A, Ho R, Boeckers TM, Mulaw M (2023) Multiomics and machine-learning identify novel transcriptional and mutational signatures in amyotrophic lateral sclerosis. *Brain* 146:3770–3782. <https://doi.org/10.1093/brain/awad075>
 23. Elhaik E (2022) Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Sci Rep* 12:14683. <https://doi.org/10.1038/s41598-022-14395-4>
 24. Naughton M, Moffat J, Eleftheriadis G, de la Vega Gallardo N, Young A, Falconer J, Hawkins K, Pearson B, Perbal B, Hogan A, Moynagh P, Loveless S, Robertson NP, Gran B, Kee R, Hughes S, McDonnell G, Howell O, Fitzgerald DC (2020) CCN3 is dynamically regulated by treatment and disease state in multiple sclerosis. *J Neuroinflammation* 17:349. <https://doi.org/10.1186/s12974-020-02025-7>
 25. O'Brown NM, Patel NB, Hartmann U, Klein AM, Gu C, Megason SG (2023) The secreted neuronal signal Spock1 promotes blood-brain barrier development. *Dev Cell* 58:1534–1547e6. <https://doi.org/10.1016/j.devcel.2023.06.005>
 26. Lyon MS, Milligan C (2019) Extracellular heat shock proteins in neurodegenerative diseases: new perspectives. *Neurosci Lett* 711:134462. <https://doi.org/10.1016/j.neulet.2019.134462>
 27. Rajendran R, Böttiger G, Stadelmann C, Karnati S, Berghoff M (2021) FGF/FGFR pathways in multiple sclerosis and in its Disease models. *Cells* 10:884. <https://doi.org/10.3390/cells10040884>
 28. Montaner J, Ramiro L, Simats A, Tiedt S, Makris K, Jickling GC, Dobbie S, Sanchez JC, Bustamante A (2020) Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat Rev Neurol* 16:247–264. <https://doi.org/10.1038/s41582-020-0350-6>
 29. Aerqin Q, Wang ZT, Wu KM, He XY, Dong Q, Yu JT (2022) Omics-based biomarkers discovery for Alzheimer's disease. *Cell Mol Life Sci* 79:585. <https://doi.org/10.1007/s00018-022-04614-6>
 30. Raghunathan R, Turajane K, Wong LC (2022) Biomarkers in neurodegenerative diseases: Proteomics spotlight on ALS and Parkinson's Disease. *Int J Mol Sci* 23:9299. <https://doi.org/10.3390/ijms23169299>
 31. Schumacher-Schuh A, Bieger A, Borelli WV, Portley MK, Awad PS, Bandres-Ciga S (2022) Advances in Proteomic and metabolomic profiling of neurodegenerative diseases. *Front Neurol* 12:792227. <https://doi.org/10.3389/fneur.2021.792227>
 32. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos G (2020) A random forest based biomarker discovery and power analysis framework for diagnostics research. *BMC Med Genomics* 13:178. <https://doi.org/10.1186/s12920-020-00826-6>
 33. Bettencourt C, Skene N, Bandres-Ciga S, Anderson E, Winchester LM, Foote IF, Schwartzentruber J, Botia JA, Nalls M, Singleton A, Schilder BM, Humphrey J, Marzi SJ, Toomey CE, Kleiflat AA, Harshfield EL, Garfield V, Sandor C, Keat S, Tamburin S, Frigério CS, Lourida I; Deep Dementia Phenotyping (DEMON) Network; Ranson JM, Llewellyn DJ (2023) Artificial intelligence for dementia genetics and omics. *Alzheimers Dement* 19:5905–5921. <https://doi.org/10.1002/alz.13427>
 34. Kalafatakis I, Savvaki M, Velona T, Karagogeos D (2021) Implication of contactins in demyelinating pathologies. *Life (Basel)* 11:51. <https://doi.org/10.3390/life11010051>
 35. Rios JC, Melendez-Vasquez CV, Einheber S, Lustig M, Grumet M, Hemperly J, Peles E, Salzer JL (2000) Contactin-associated protein (Caspr) and contactin form a complex that is targeted to the paranodal junctions during myelination. *J Neurosci* 20:8354–8364. <https://doi.org/10.1523/JNEUROSCI.20-22-08354.2000>
 36. Kong L, Valdivia DO, Simon CM, Hassinan CW, Delestrée N, Ramos DM, Park JH, Pilato CM, Xu X, Crowder M, Grzyb CC, King ZA, Petrillo M, Swoboda KJ, Davis C, Lutz CM, Stephan AH, Zhao X, Weetall M, Naryshkin NA, Crawford TO, Mentis GZ, Sumner CJ (2021) Impaired prenatal motor axon development necessitates early therapeutic intervention in severe SMA. *Sci Transl Med* 13:eabb6871. <https://doi.org/10.1126/scitranslmed.abb6871>
 37. Djannatian M, Timmler S, Arends M, Luckner M, Weil MT, Alexopoulos I, Snaidero N, Schmid B, Misgeld T, Möbius W, Schifferer M, Peles E, Simons M (2019) Two adhesive systems cooperatively regulate axon ensheathment and myelin growth in the CNS. *Nat Commun* 10:4794. <https://doi.org/10.1038/s41467-019-12789-z>
 38. Yang CW, Chen CL, Chou WC, Lin HC, Jong YJ, Tsai LK, Chuang CY (2016) An Integrative Transcriptomic Analysis for identifying Novel Target genes corresponding to Severity Spectrum in spinal muscular atrophy. *PLoS ONE* 11:e0157426. <https://doi.org/10.1371/journal.pone.0157426>
 39. Kamal N, Jafari Khamirani H, Dara M, Dianatpour M (2023) NRXN3 mutations cause developmental delay, movement disorder, and behavioral problems: CRISPR edited cells based WES results. *Gene* 2023 867:147347. <https://doi.org/10.1016/j.gene.2023.147347>
 40. Lloyd BA, Han Y, Roth R, Zhang B, Aoto J (2023) Neurexin-3 subsynaptic densities are spatially distinct from Neurexin-1 and essential for excitatory synapse nanoscale organization in the hippocampus. *Nat Commun* 14:4706. <https://doi.org/10.1038/s41467-023-40419-2>
 41. Trotter JH, Wang CY, Zhou P, Nakahara G, Südhof TC (2023) A combinatorial code of neurexin-3 alternative splicing controls inhibitory synapses via a trans-synaptic dystroglycan signaling loop. *Nat Commun* 14:1771. <https://doi.org/10.1038/s41467-023-36872-8>
 42. See K, Yadav P, Giegerich M, Cheong PS, Graf M, Vyas H, Lee SG, Mathavan S, Fischer U, Sendtner M, Winkler C (2014) SMN deficiency alters Nrnx2 expression and splicing in zebrafish and mouse models of spinal muscular atrophy. *Hum Mol Genet* 23:1754–1770. <https://doi.org/10.1093/hmg/ddt567>
 43. Rizzo F, Nizzardo M, Vashisht S, Molteni E, Melzi V, Taiana M, Salani S, Santonicola P, Di Schiavi E, Bucchia M, Bordoni A, Faravelli I, Bresolin N, Comi GP, Pozzoli U, Corti S (2019) Key

- role of SMN/SYNERIP and RNA-Motif 7 in spinal muscular atrophy: RNA-Seq and motif analysis of human motor neurons. *Brain* 142:276–294. <https://doi.org/10.1093/brain/awy330>
44. Blaschek A, Kölbl H, Schwartz O, Köhler C, Gläser D, Eggermann K, Hannibal I, Schara-Schmidt U, Müller-Felber W, Vill K (2022) Newborn screening for SMA - can a wait-and-see strategy be responsibly justified in patients with four SMN2 copies? *J Neuromuscul Dis* 9:597–605. <https://doi.org/10.3233/JND-221510>
 45. Nitz E, Smitka M, Schallner J, Akgün K, Ziemssen T, von der Hagen M, Tüngler V (2021) Serum neurofilament light chain in pediatric spinal muscular atrophy patients and healthy children. *Ann Clin Transl Neurol* 8:2013–2024. <https://doi.org/10.1002/acn3.51449>
 46. Olsson B, Alberg L, Cullen NC, Michael E, Wahlgren L, Kroksmark AK, Rostasy K, Blennow K, Zetterberg H, Tulinus M (2019) NFL is a marker of treatment response in children with SMA treated with nusinersen. *J Neurol* 266:2129–2136. <https://doi.org/10.1007/s00415-019-09547-y>
 47. Kariyawasam D, D'Silva A, Howells J, Herbert K, Geelan-Small P, Lin CS, Farrar MA (2020) Motor unit changes in children with symptomatic spinal muscular atrophy treated with nusinersen. *J Neurol Neurosurg Psychiatry* 92:78–85. <https://doi.org/10.1136/jnnp-2020-324254>
 48. Kariyawasam DST, D'Silva AM, Herbert K, Howells J, Carey K, Kandula T, Farrar MA, Lin CS (2022) Axonal excitability changes in children with spinal muscular atrophy treated with nusinersen. *J Physiol* 600:95–109. <https://doi.org/10.1113/JP282249>
 49. Faravelli I, Meneri M, Saccomanno D, Velardo D, Abati E, Gagliardi D, Parente V, Petrozzi L, Ronchi D, Stocchetti N, Calderini E, D'Angelo G, Chidini G, Prandi E, Ricci G, Siciliano G, Bresolin N, Comi GP, Corti S, Magri F, Govoni A (2020) Nusinersen treatment and cerebrospinal fluid neurofilaments: an explorative study on spinal muscular atrophy type 3 patients. *J Cell Mol Med* 24:3034–3039. <https://doi.org/10.1111/jcmm.14939>
 50. Darras BT, Crawford TO, Finkel RS, Mercuri E, De Vivo DC, Oskoui M, Tizzano EF, Ryan MM, Muntoni F, Zhao G, Staropoli J, McCampbell A, Petrillo M, Stebbins C, Fradette S, Farwell W, Sumner CJ (2019) Neurofilament as a potential biomarker for spinal muscular atrophy. *Ann Clin Transl Neurol* 6:932–944. <https://doi.org/10.1002/acn3.779>
 51. Milella G, Introna A, D'Errico E, Fraddosio A, Scaglione G, Morea A, Ucci M, Ruggieri M, Mastrapasqua M, Megna M, Puntillo F, Simone IL (2021) Cerebrospinal fluid and clinical profiles in adult type 2–3 spinal muscular atrophy patients treated with Nusinersen: an 18-Month single-centre experience. *Clin Drug Investig* 41:775–784. <https://doi.org/10.1007/s40261-021-01071-0>
 52. Wurster CD, Steinacker P, Günther R, Koch JC, Lingor P, Uzelac Z, Witzel S, Wollinsky K, Winter B, Osmanovic A, Schreiber-Katz O, Al Shweiki R, Ludolph AC, Petri S, Hermann A, Otto M, MND-Net (2020) Neurofilament light chain in serum of adolescent and adult SMA patients under treatment with nusinersen. *J Neurol* 267:36–44. <https://doi.org/10.1007/s00415-019-09547-y>
 53. Carulli D, de Winter F, Verhaagen J (2021) Semaphorins in adult nervous system plasticity and disease. *Front Synaptic Neurosci* 13:672891. <https://doi.org/10.3389/fnsyn.2021.672891>
 54. Ørngreen MC, Zacho M, Hebert A, Laub M, Vissing J (2003) Patients with severe muscle wasting are prone to develop hypoglycemia during fasting. *Neurology* 61:997–1000. <https://doi.org/10.1212/01.wnl.0000086813.59722.72>
 55. Berti B, Onesimo R, Leone D, Palermo C, Giorgio V, Buon-senso D, Pane M, Mercuri E (2020) Hypoglycaemia in patients with type 1 SMA: an underdiagnosed problem? *Arch Dis Child* 105:707. <https://doi.org/10.1136/archdischild-2019-318120>
 56. Deguise MO, Chehade L, Kothary R (2021) Metabolic dysfunction in spinal muscular atrophy. *Int J Mol Sci* 22:5913. <https://doi.org/10.3390/ijms22115913>
 57. Ørngreen MC, Andersen AG, Eisum AS, Hald EJ, Raaschou-Pedersen DE, Løkken N, Høi-Hansen CE, Vissing J, Born AP, van Hall G (2021) Prolonged fasting-induced hyperketosis, hypoglycaemia and impaired fat oxidation in child and adult patients with spinal muscular atrophy type II. *Acta Paediatr* 110:3367–3375. <https://doi.org/10.1111/apa.16074>
 58. Amara U, Rittirsch D, Flierl M, Bruckner U, Klos A, Gebhard F, Lambris JD, Huber-Lang M (2008) Interaction between the coagulation and complement system. *Adv Exp Med Biol* 632:71–79. https://doi.org/10.1007/978-0-387-78952-1_6
 59. Amara U, Flierl MA, Rittirsch D, Klos A, Chen H, Acker B, Brückner UB, Nilsson B, Gebhard F, Lambris JD, Huber-Lang M (2010) Molecular intercommunication between the complement and coagulation systems. *J Immunol* 185:5628–5636. <https://doi.org/10.4049/jimmunol.0903678>
 60. Krem MM, Di Cera E (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem Sci* 27:67–74. [https://doi.org/10.1016/s0968-0004\(01\)02007-2](https://doi.org/10.1016/s0968-0004(01)02007-2)
 61. Rittirsch D, Flierl MA, Ward PA (2008) Harmful molecular mechanisms in sepsis. *Nat Rev Immunol* 8:776–787. <https://doi.org/10.1038/nri2402>
 62. Magdalon J, Mansur F, Teles E, Silva AL, de Goes VA, Reiner O, Sertié AL (2020) Complement system in Brain Architecture and Neurodevelopmental disorders. *Front Neurosci* 14:23. <https://doi.org/10.3389/fnins.2020.00023>
 63. Stevens B, Allen NJ, Vazquez LE, Howell GR, Christopherson KS, Nouri N, Micheva KD, Mehalow AK, Huberman AD, Stafford B, Sher A, Litke AM, Lambris JD, Smith SJ, John SW, Barres BA (2007) The classical complement cascade mediates CNS synapse elimination. *Cell* 131:1164–1178. <https://doi.org/10.1016/j.cell.2007.10.036>
 64. Vukojicic A, Delestrée N, Fletcher EV, Pagiazitis JG, Sankaranarayanan S, Yednock TA, Barres BA, Mentis GZ (2019) The classical complement pathway mediates Microglia-Dependent remodeling of spinal motor circuits during development and in SMA. *Cell Rep* 29:3087–3100e7. <https://doi.org/10.1016/j.celrep.2019.11.013>
 65. Gavriilaki M, Kimiskidis VK, Gavriilaki E (2020) Precision Medicine in Neurology: the inspirational paradigm of complement therapeutics. *Pharmaceuticals (Basel)* 13:341. <https://doi.org/10.3390/ph13110341>
 66. Deguise MO, Kothary R (2017) New insights into SMA pathogenesis: immune dysfunction and neuroinflammation. *Ann Clin Transl Neurol* 4:522–530. <https://doi.org/10.1002/acn3.423>
 67. Wan B, Feng P, Guan Z, Sheng L, Liu Z, Hua Y (2018) A severe mouse model of spinal muscular atrophy develops early systemic inflammation. *Hum Mol Genet* 27:4061–4076. <https://doi.org/10.1093/hmg/ddy300>
 68. Bonanno S, Cavalcante P, Salvi E, Giagnorio E, Malacarne C, Cattaneo M, Andreatta F, Venerando A, Pensato V, Gellera C, Zanin R, Arnoldi MT, Dosi C, Mantegazza R, Masson R, Maggi L, Marcuzzo S (2022) Identification of a cytokine profile in serum and cerebrospinal fluid of pediatric and adult spinal muscular atrophy patients and its modulation upon nusinersen treatment. *Front Cell Neurosci* 16:982760. <https://doi.org/10.3389/fncel.2022.982760>
 69. Nuzzo T, Russo R, Errico F, D'Amico A, Tewelde AG, Valletta M, Hassan A, Tosi M, Panicucci C, Bruno C, Bertini E, Chaberry A, Pellizzoni L, Usiello A (2023) Nusinersen mitigates neuroinflammation in severe spinal muscular atrophy patients. *Commun Med (Lond)* 3:28. <https://doi.org/10.1038/s43856-023-00256-2>
 70. Pane M, Palermo C, Messina S, Sansone VA, Bruno C, Catteruccia M, Sframeli M, Albamonte E, Pedemonte M, D'Amico A, Brigati G, de Sanctis R, Coratti G, Lucibello S, Bertini E, Vita G, Danilo Tiziano F, Mercuri E, Italian EAP, Working Group (2018) An observational study of functional abilities in infants, children,

- and adults with type 1 SMA. *Neurol* 2018 91:e696–e703. <https://doi.org/10.1212/WNL.0000000000006050>
71. Glanzman AM, Mazzone E, Main M, Pelliccioni M, Wood J, Swoboda KJ, Scott C, Pane M, Messina S, Bertini E, Mercuri E, Finkel RS (2010) The children's hospital of Philadelphia Infant Test of Neuromuscular disorders (CHOP INTEND): test development and reliability. *Neuromuscul Disord* 20:155–161. <https://doi.org/10.1016/j.nmd.2009.11.014>
 72. O'Hagen JM, Glanzman AM, McDermott MP, Ryan PA, Flickinger J, Quigley J, Riley S, Sanborn E, Irvine C, Martens WB, Annis C, Tawil R, Oskoui M, Darras BT, Finkel RS, De Vivo DC (2007) An expanded version of the Hammersmith Functional Motor Scale for SMA II and III patients. *Neuromuscul Disord* 17:693–697. <https://doi.org/10.1016/j.nmd.2007.05.009>
 73. Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, Chiriboga CA, Saito K, Servais L, Tizzano E, Topaloglu H, Tulinius M, Montes J, Glanzman AM, Bishop K, Zhong ZJ, Gheuens S, Bennett CF, Schneider E, Farwell W, De Vivo DC, ENDEAR Study Group (2017) Nusinersen versus Sham Control in infantile-onset spinal muscular atrophy. *N Engl J Med* 377:1723–1732. <https://doi.org/10.1056/NEJMoa1702752>
 74. Pechmann A, Behrens M, Dörnbrack K, Tassoni A, Wenzel F, Stein S, Vogt S, Zöllner D, Bernert G, Hagenacker T, Schara-Schmidt U, Walter MC, Bertsche A, Vill K, Baumann M, Baumgartner M, Cordts I, Eisenkölbl A, Flotats-Bastardas M, Friese J, Günther R, Hahn A, Horber V, Husain RA, Illsinger S, Jahnel J, Johannsen J, Köhler C, Kölbl H, Müller M, von Moers A, Schwerin-Nagel A, Reihle C, Schlachter K, Schreiber G, Schwartz O, Smitka M, Steiner E, Trollmann R, Weiler M, Weiß C, Wiegand G, Wilichowski E, Ziegler A, Lochmüller H, Kirschner J, SMARtCARE study group (2022) Improved upper limb function in non-ambulant children with SMA type 2 and 3 during nusinersen treatment: a prospective 3-years SMARtCARE registry study. *Orphanet J Rare Dis* 17:384. <https://doi.org/10.1186/s13023-022-02547-8>
 75. Bath TS, Tollenaere MX, Rüter P, Gonzalez-Franquesa A, Prabhakar BS, Bekker-Jensen S, Deshmukh AS, Olsen JV (2019) Protein aggregation capture on Microparticles enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteom* 18:1027–1035. <https://doi.org/10.1074/mcp.TIR118.001270>
 76. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* 11:319–324. <https://doi.org/10.1038/nmeth.2834>
 77. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372. <https://doi.org/10.1038/nbt.1511>
 78. Alakwaa FM, Chaudhary K, Garmire LX (2018) Deep learning accurately predicts estrogen receptor status in breast Cancer Metabolomics Data. *J Proteome Res* 17:337–347. <https://doi.org/10.1021/acs.jproteome.7b00595>
 79. Bahado-Singh RO, Sonek J, McKenna D, Cool D, Aydas B, Turkoglu O, Bjorndahl T, Mandal R, Wishart D, Friedman P, Graham SF, Yilmaz A (2019) Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix. *Ultrasound Obstet Gynecol* 54:110–118. <https://doi.org/10.1002/uog.20168>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Chiara Panicucci¹  · Eray Sahin² · Martina Bartolucci³ · Sara Casalini¹ · Noemi Brolatti¹ · Marina Pedemonte⁴ · Serena Baratto¹ · Sara Pintus¹ · Elisa Principi¹ · Adele D'Amico⁵ · Marika Pane⁶ · Marina Sframeli⁷ · Sonia Messina⁷ · Emilio Albamonte⁸ · Valeria A. Sansone⁸ · Eugenio Mercuri⁶ · Enrico Bertini⁵ · Ugur Sezerman⁹ · Andrea Petretto³ · Claudio Bruno^{1,10} 

✉ Claudio Bruno
claudiobruno@gaslini.org

¹ Center of Translational and Experimental Myology, IRCCS Istituto Giannina Gaslini, Via G. Gaslini, 5, I-16147 Genova, Italy

² Department of Biostatistics and Bioinformatics, Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

³ Core Facilities-Clinical Proteomics and Metabolomics, IRCCS Istituto Giannina Gaslini, Genova, Italy

⁴ Pediatric Neurology Unit, IRCCS Istituto Giannina Gaslini, Genova, Italy

⁵ Unit of Neuromuscular and Neurodegenerative Disorders, IRCCS Bambino Gesù Children's Hospital, Rome, Italy

⁶ Centro Clinico Nemo, IRCCS Fondazione Policlinico Universitario Agostino Gemelli, Rome, Italy

⁷ Department of Neurosciences, University of Messina, Messina, Italy

⁸ Neurorehabilitation Unit, Centro Clinico NeMO, University of Milan, Milan, Italy

⁹ Department of Biostatistics and Medical Informatics, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

¹⁰ Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health- DINO GMI, University of Genova, Genova, Italy