Computational analysis of SOD1-G93A mouse muscle biomarkers for comprehensive assessment of ALS progression

Pedro Gómez-Gálvez[1,2,3,4,*], Victoria Navarro[4,5], Ana M. Castro[3], Carmen Paradas[4,5,6,*], Luis M. Escudero[3,4,*]

[1] MRC Laboratory of Molecular Biology, Cambridge, United Kingdom

[2] Physiology, Development and Neurobiology department, University of Cambridge, Cambridge, United Kingdom.

[3] Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla and Departamento de Biología Celular, Facultad de Biología, Universidad de Sevilla, Seville, Spain

[4] CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain

[5] Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocio CSIC/Universidad de Sevilla, Seville, Spain

[6] Neuromuscular Disease Unit, Neurology Department, Hospital Universitario Virgen del Rocío, Seville, Spain

* Corresponding authors

KEYWORDS: Amyotrophic lateral sclerosis, motor neuron degeneration, SOD1-G93A, image biomarker, image analysis, graph-theory

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is, to date, a fatal neurodegenerative disease that currently presents significant challenges in both diagnosis and treatment. The rapid progression and critical prognosis of the disease make accurate and prompt diagnosis crucial for enhancing ALS patient outcomes and quality of life. In this regard, the identification of ALS biomarkers is imperative for evaluating potential treatments, being SOD1^{G93A} murine models widely employed for their validation in preclinical studies. Here, we introduce and apply an upgraded version of NDICIA^{TM} (Neuromuscular DIseases Computerized Image Analysis), an image analysis tool to test and quantify the presence of neuromuscular disease image biomarkers. We applied this method to examine morphological and network features in histological images of skeletal muscle biopsies acquired from SOD1^{G93A} and wild type (WT) mice at different stages of the disease progression: first we quantified the level of difference among mutant and WT groups of images, second we filtered out those features being the primary factors in every separation and finally we analysed how those differences and features change as the disease progresses through its different stages. This analysis revealed the way the pathology of mutant muscles evolves, differentiating from WT in muscle fibres arrangement (graph-theory properties) at presymptomatic stages, and presenting incremental atrophic-like morphological properties at later stages. In consequence, our assay pointed out muscle organization features as an important defining factor in ALS, which could be an excellent outcome measure to early detect a beneficial effect of therapeutic interventions in preclinical trials.

INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS) is the most common neurodegenerative disorder of the motor neuron system (Gwinn et al., 2007; Hilton et al., 2016). The absence of an effective treatment for this devastating disease (Czarzasta et al., 2017; Gwinn et al., 2007) makes urgent finding out new therapies to prevent, palliate, and, as ultimate ambition, cure this fatal disease. While most of the cases are registered as sporadic forms (Gwinn et al., 2007; Pramatarova et al., 1995), about 10% of ALS cases are familial, caused by mutations in different genes like *C9ORF72* (Renton et al., 2011), *Cu/Zn* superoxide dismutase (SOD1) (Rosen et al., 1993), FUS (Kwiatkowski et al., 2009), TDP-43 (Sreedharan et al., 2008), as the most frequent among many others (Vajda et al., 2017).

Mouse models are the most used model organism to investigate human diseases (Rosenthal & Brown, 2007). Some important studies ratified them as a suitable model to research about neuronal diseases in humans (Gurney, 1997; Guy et al., 2007; Wegorzewska et al., 2009; Wong et al., 2002). In ALS, different mouse models such as SOD1, TDP-43, FUS and VCP (Scott et al., 2008), have been widely used to study ALS pathogenesis, being the SOD1 mouse models the most popular (Weydt et al., 2003). Among the transgenic SOD1 murine models, the one expressing a G93A mutant form of human SOD1 ($SOD1^{G93A}$) (Gurney et al., 1994), is the most extensively studied and broadly used for developing preclinical therapeutic trials for ALS (Gurney, 1997). This genetic alteration promotes an overexpression of the mutant SOD1 enzyme in mice leading to a rapidly progressive motor neuron disease, in a clinical and pathological form with a remarkable resemblance to ALS observed in patients. The onset of clinical manifestation of ALS in $SOD1^{G93A}$ mice has been described approximately at 90 days of age (De Giorgio et al., 2019), with rapid progression leading to death around 120-130 days old (Bruijn et al., 1997; Gurney et al., 1994). However, recent bibliographic research, including $SOD1^{G93A}$ mice raised under diverse conditions and with different genetic backgrounds, indicates a broader range of onset times, with an average onset reported at 16 weeks of age following an analysis of incidences in $SOD1^{G93A}$ mice aged 2 to 6 months (Ciuro et al., 2023).

In comparison to other transgenic SOD1 mouse models, the one that overexpresses the wild-type human protein ($SOD1^{WT}$) has been suggested to exhibit pathogenic mechanisms similar to those of the $SOD1^{G93A}$ mutant, which may promote the manifestation of ALS (Bosco et al., 2010; Graffmo et al., 2013; Rotunno & Bosco, 2013). Nevertheless, there is ongoing scientific controversy regarding the role of $SOD1^{WT}$ mouse model in ALS (Audet et al., 2010; Rotunno & Bosco, 2013), and its significance is yet to be established.

Examining the tissue architecture in muscle biopsies presents a potential avenue for elucidating the underlying mechanisms of ALS pathogenesis and to uncover an imaging biomarker. The cellular arrangement is intricately regulated and maintained throughout morphogenesis, thus exhibiting a strong association with tissue function. Numerous investigations have recently employed the organizational characteristics of densely packed tissues to gain insights into organ functionality (Gibson et al., 2006; Hayashi & Carthew, 2004; Sánchez-Gutiérrez et al., 2016; Sánchez-Gutiérrez et al., 2017; Savadjiev et al., 2012). Focusing on the skeletal muscles, they constitute a distinctive variety of compact tissue, which comprises of bundled fascicles partitioned by a fine connective tissue known as endomysium (Helliwell, 1999). In accordance with metabolic properties, muscle fibres can be classified into two main types, i.e. type I (slow fibres) and type II (fast fibres), which can be distinguished by the differential expression of the myosin protein isoforms (Pette & Staron, 2000). This tissue organization, easy to observe in a muscle biopsy, emerges as a tessellation where fibres behave similar as polygons and the endomysium as the connector without leaving empty space between them. In fact, it was shown that natural packed tissues (i.e. muscles) are subject to physiological constraints and the cellular mosaics follow some organizational patterns (Sánchez-Gutiérrez et al., 2016).

In this sense, there have been some efforts to automatically extract geometric features from normal muscle biopsies (Behan et al., 2002; Garton et al., 2010; Karen et al., 2009; Kim et al., 2007; Klemencic et al., 1998; Meunier et al., 2010), but those approaches lack the ability to interpret the results accurately. In response, our study builds upon some previous works (Sáez, Acha, et al., 2013; Sáez, Rivas, et al., 2013; Sánchez-Gutiérrez et al., 2017) and introduces a

method that allows the categorization of packed tissues in a detailed, reproducible and objective manner. In particular, we adapted the Neuromuscular DIseases Computerized Image Analysis (NDICIA) (Sáez, Acha, et al., 2013) to extract and analyse morphological, structural and network information in a robust way. NDICIA considers the biopsy as a connected graph, in which nodes are the fibres, and the links are demarcated by the vicinity of each one. To identify the differences between healthy and affected muscles, evaluation of widely known morphological pathogenic features, such as fibre shape or quantity of extracellular connective tissue is critical (Fontelonga et al., 2019; Rooney, 2006). Furthermore, the changes in the pattern of slow and fast fibres uncovered in studies about neurodegenerative muscle diseases, underscore the importance of utilizing graph theory for tissue analysis (Dubowitz & Sewry, 2007; Sáez, Rivas, et al., 2013). From its foundation, NDICIA has used Principal Component Analysis (PCA) to reduce dimensionality of images features, enabling an easier analysis about how different two groups of images are, and a filtering of the most relevant features associated with the differentiation. However, some years ago, other algorithms to reduce dimensionality have emerged as promising to enhance or complement NDICIA outcomes. This is the case of the widely used, Uniform Manifold Approximation and Projection (UMAP) method (McInnes et al., 2018). UMAP (and t-SNE (Van Der Maaten & Hinton, 2008) among others) incorporates non-lineal transformations after fine-tuning of some hyperparameters, operating as a stochastic system, unlike PCA, that relies on lineal transformations based on covariance, being a deterministic system without the need to adjust any parameter.

In this study, we used NDICIA pipelines with PCA and UMAP methods to analyse the skeletal muscle organization in SOD1^{G93A} mice. Our analysis identified significant differences in the organizational pattern between pre-symptomatic and late stages of the disease, providing a signature of the mutant condition.

RESULTS

Topological analysis of WT and SOD1^{G93A} mice muscles.

We analysed the soleus muscle from SOD1^{G93A} and WT mice in terms of organization and morphology of their fibres to investigate the disease progression in this animal model for ALS. We used immunohistochemical staining in thin biopsies of these animals (**methods**). In particular, anti-collagen VI antibody tagging the collagen located in the extracellular matrix and the antibody anti-myosin slow (type I) providing the identification of the type of fibre (fast fibre or slow fibre) (**Fig. 1, methods**). Thus, we compared images coming from SOD1^{G93A} and WT mice at different ages: postnatal day 60 (p60), 100 (p100) and 120 (p120). A set of 103 images of SOD1^{G93A} soleus muscles (23 different mice), and 208 from WT animals (30 different mice) were classified into the aforementioned three groups of age (**Fig. 1**, **methods**). These subsets of images were the source for the NDICIA pipeline.

First, we applied the segmentation module of NDICIA, which allowed us to distinguish between muscle fibres and the collagen network present in the images (as shown in **Fig. 2A-E**). Furthermore, the fibres were classified as slow or fast fibres based on the slow myosin heavy chain staining (**Fig. 2B**, **2D**, **methods**). Subsequently, we constructed a network over the segmented tissue considering the fibres as nodes connected with their neighbouring fibres (**Fig. 2F**).

Second, we extracted 67 features (**Table 1**) from every image, similarly to the approach followed in a previous study (Sánchez-Gutiérrez et al., 2017): the first 14 features considered the shape and size of cells; while features 15-66 were designed to capture the organization of cells and their network connectivity, using properties derived from graph theory. The 67th feature reflected the proportion of slow and fast fibres.

In the last step of the NDICIA pipeline, we executed the Principal Component Analysis (PCA) module to identify the most relevant combination of features that will separate two groups of images, scored by a PCA descriptor (**methods**). The value of the descriptor is an indicator of the difference between the groups of images, with a larger descriptor value indicating a higher level of difference between the groups ((Costa & Cesar, 2009; Sánchez-Gutiérrez et al., 2013a), **methods**). Based on our experience and previous comparisons between groups containing more than 15 images, a PCA descriptor greater than "1" was considered indicative of a good separation between samples.

In order to assess whether we were able to discriminate between biopsies from ALS mouse models and WT mice, and to determine which features were most relevant for differentiating between the two groups, we carried out three types of comparisons in the following order: i) using the 10 purely geometric characteristics without considering the presence of fast or slow fibres (10 characteristics (ccs), features [1-2, 7-14], **Table 1**) ii) applying 14 geometric features and the proportion of slow/fast fibres, that is, expanding the subset of 10 ccs after adding the geometric features considering the presence of slow/fast fibres (15 ccs, features [1-14, 67], **Table 1**) iii) using the entire set of 67 features including all geometric and network features (67 ccs, features [1-67] , **Table 1**).

Therefore, we launched the NDICIA pipeline to obtain morphometric differences between WT and SOD1^{G93A} muscles, that is, using the subset of 10 ccs, and evaluated if they were enough to differentiate mutant from healthy muscles:

First, we studied the stages where the phenotypic differences between SOD1^{G93A} and WT mice were more evident, specifically at 120 days of age. After applying the PCA module to compare the muscles of p120 SOD1^{G93A} and WT mice, we obtained a high PCA value of 2.6, indicating a significant differentiation between the two groups (**Fig. 3A**). The selected characteristics that contributed significantly to this differentiation were standard deviation cell area (cc. 2); average cell convex hull (cc. 11) and standard deviation cell convex hull (cc. 12) (**Fig. 3A, Table 2**). Note that each feature holds an absolute weight in the PCA separation, and the selected features were arranged decreasingly from more to less weight from left to right (we will be consistent by showing the features in this manner along the manuscript) (**Table 2**). These results pointed out that the mutant muscles contained cells with larger average area variability (cc. 2) than the WT muscles, along with less convex and more heterogeneous cell shapes (ccs. 11 and 12). These features contribute to differentiate accurately mutant from WT samples.

In the second comparison, we investigated the morphometric differences between p100 SOD1^{G93A} and WT mice. At this age, the clinical phenotype of the disease is present but not as severe as in p120. In this case the analysis released a lower PCA descriptor (0.74). The selected features for this comparison were

average cell minor axis (cc. 8); standard deviation cell area/(cell area + collagen area) (cc. 14) and standard deviation (major/minor axes) of cells (cc. 10) (**Fig. 3B, Table 2**). The results showed that the SOD1^{G93A} muscles had a slightly lower cell minor axis (cc. 8), a more heterogeneous collagen distribution (cc. 14), and a higher variability of major/minor axes (cc. 10) compared to WT muscles. However, these features were not sufficient to achieve a clear separation between the two groups.

Third, we analysed the morphological differences in tissues from p60 mice, before the emergence of ALS symptoms in SOD1^{G93A}. The resulting PCA value (0.33) showed that purely geometric features were not sufficient to differentiate between mutant and WT at this age stage (**Fig. 3C, Table 2**).

After evaluating the discriminative capability of the 10 ccs subset of features, we proceeded to evaluate the 15 ccs subset, which incorporated geometric properties while considering the differentiation of muscle fibres into slow or fast.

In the p120 comparison, the differentiation slightly improved (from PCA value: 2.6 to PCA value 2.94) after filtering the features below: proportion slow cells/fast cells (cc. 67); average cell convex hull (cc. 11); standard deviation cell convex hull (cc. 12); standard deviation cell area/(cell area + collagen area) (cc. 14); standard deviation cell area (cc. 2) and standard deviation slow cells area (cc. 4) (**Fig. 3D, Table 2**). The same features emerged as in the 10 ccs comparison, with the addition of the proportion of slow/fast fibres (cc. 67) being larger in the number of fast fibres in the ALS model than in WT, and a higher variation of slow fibres area (cc. 4) in the mutant compared to the healthy mice (**Table 2**).

In the p100 analysis, we observed an enhancement in the PCA descriptor (from 0.74 to 0.95) after selecting the following characteristics: proportion slow cells/fast cells (cc. 67); standard deviation (major/minor axes) of cells (cc. 10); standard deviation cell area/(cell area + collagen area) (cc. 14); average major/minor axes of cells (cc. 9); standard deviation cell convex hull (cc. 12) and standard deviation fast cells area (cc. 6) (**Fig. 3E, Table 2**). Once again, the higher number of fast cells in SOD1^{G93A} than in WT (cc. 67) and the higher variability in fast fibres area were found to be crucial in achieving a separation

close to the experimental threshold, with a PCA value of approximately 1 (**Table 2**).

In p60 observation, there was an improvement in the differentiation between mice genotypes in terms of PCA (from 0.33 to 0.88), but it was not sufficient to clearly distinguish the two groups, selecting the properties: proportion slow cells/fast cells (cc. 67); standard deviation cell convex hull (cc. 12); average major/minor axes) of cells (cc. 9); standard deviation cell area/(cell area + collagen area) (cc. 14) and standard deviation fast cells area (cc. 6) (**Fig. 3F, Table 2**). Thus, the inclusion of fast/slow fibres discrimination, specifically the larger number of fast fibres (cc. 67) and the more variable area of fast cells (cc. 6) considerably improved the separation, together with the previously selected features using 10 ccs. Interestingly, these features were also chosen in the p100 comparison, where the power of separation was greater than here (**Fig. 3E, Table 2**).

Finally, we conducted the NDICIA protocol using the full set of 67 ccs, which included the subset of 15 ccs as well as purely graph theory features and network features that compared the geometric characteristics of the fibres neighbourhood (**Table 1**)**:**

In the p120 analysis, the PCA descriptor was further enhanced (from 2.94 to 3.49), with the following features being highlighted: proportion slow cells/fast cells (cc. 67); average cell convex hull (cc. 11); standard deviation cell convex hull (cc. 12); standard deviation (cell/neighbouring cells) area (cc. 24); average (cell/neighbouring cells) minor axis (cc. 27); standard deviation cell area (cc. 2); standard deviation slow cells area (cc. 4) (**Fig. 3G, Table 2**). Taking as reference the previous 15 ccs analysis (**Fig. 3D**), only two new features were included from the 67 ccs set (ccs. 24 and 27). $SOD1^{G93A}$ was the one with larger variability in neighbouring cell area (cc. 24) and a smaller average minor axis of neighbouring cells (cc. 27) compared to WT (**Table 2**).

In the analysis of p100 mice, the PCA value was found to be improved (from 0.95 to 1.12) with the selection of specific features, including standard deviation (cell/neighbouring cells) area (cc. 24); proportion slow cells/fast cells (cc. 67); standard deviation cell area/(cell area + collagen area) (cc. 14); standard

deviation (cell/neighbouring cells) major axis (cc. 26); standard deviation cell convex hull (cc. 12); standard deviation fast cells area (cc. 6) and standard deviation (cell/neighbouring cells) (major/minor axes) (cc. 30) (**Fig. 3H, Table 2**). Here, as a novelty compared to the previous analysis using 15 ccs (**Fig. 3E**), the bigger standard deviation in neighbourhood cell area (cc. 24) and cell axes length (ccs. 26 and 30) in mutant images than in WT, resulted in a more noticeable distinction among genotypic groups (**Fig. 3H)**.

Ultimately and interestingly, our pipeline using the complete set of 67 ccs was able to distinguish p60 SOD1^{G93A} and WT images on the PCA graph (PCA descriptor: 1.18), when the clinical phenotype has not appeared in mutant mice yet (presymptomatic mice) (**Fig. 1A-B, Fig. 3I**, **Table 2**). The features extracted were: fast cells neighbours of slow cells (cc. 20); standard deviation (major/minor axes) of cells (cc. 10); standard deviation cell convex hull (cc. 12); average (major/minor axes) of cells (cc. 9); standard deviation neighbouring cells of fast cells (cc. 18); standard deviation shortest paths length from fast cells to fast cells (cc. 62) and average clustering coefficient of slow cells (cc. 45) (**Fig. 3I**, **Table 2**). In addition to the geometric features identified in prior analyses, the presence of some network features related with the type of fibres (ccs. 20, 18, 62 and 45) played a crucial role in achieving this separation. The most significant feature in this regard was the average number of fast fibres that are neighbours of slow cells, being this number minor in SOD1^{G93A} than in WT tissues. In SOD1^{G93A}, the number of neighbours of fast cells was more diverse than in WT (cc.18). Besides, the standard deviation of the shortest paths between fast cells was larger in mutants (cc. 62), suggesting that the fast cells were more scattered and heterogeneously distributed throughout the tissue. Additionally, the average clustering coefficient of slow cells was higher in SOD1^{G93A} than in WT muscles, indicating that slow cells in mutants tended to form larger clusters (cc. 45).

Importantly, a univariate statistical analysis was performed on the selected features from the 67 ccs set of features comparisons differentiating mutant and WT at different stages. The results showed a significant difference when comparing SOD1^{G93A} and WT features values (p-value<0.05 *, **Table 2**, **methods**), highlighting the importance of these features in distinguishing between the two groups. However, none of the significant features in the

independent comparisons was shared by the p60, p100 and p120 comparisons (**Table 2**). This further emphasizes the importance of NDICIA analysis, which integrates contributions from multiple features to achieve an accurate differentiation between groups, enabling a comprehensive and objective analysis of disease progression in ALS.

Uniform Manifold Approximation and Projection (UMAP) method to upgrade the dimensionality reduction module.

PCA is a classic method to reduce the dimensionality throughout a linear transformation of data to be projected, usually, in a two-dimensional plane. Aiming to ensure a methodology contemplating a non-linear transformation reduction, we upgraded NDICIA by using the Uniform Manifold Approximation and Projection (UMAP) algorithm ((McInnes et al., 2018)**, methods**). With this change, we could reach new (non-linear) separations exploring different avenues to decipher the contribution of the same sets of features. The NDICIA workflow remained the same, with the only difference being the use of UMAP instead of the traditional PCA method. The UMAP algorithm was run 100 times for each comparison to reduce the effect of stochasticity, and the most frequently repeated features were captured, along with the average UMAP descriptor (**methods**). Then, we carried out experiments similar to those in the PCA analysis, checking the power of geometric properties (10 ccs), geometric features discriminating the type of fibre (15 ccs), and the whole set of characteristics including geometric, fibre type and network feature (67 ccs) (**Table 1**):

In the geometric evaluation (10 ccs), the UMAP descriptors were very similar to those observed using the PCA method (**Table 2**). In p120, a high average UMAP descriptor (average UMAP descriptor: 2.44±0.87) well-differentiated $SOD1^{G93A}$ from WT images (**Fig. 4A, Table 2**). The more irregular and variable cell shape (ccs. 2, 8, 10, 11 and 12) and the larger heterogeneity of collagen distribution (cc. 14) in mutants than in WT were the features that drove this comparison. However, in p100 (average UMAP descriptor: 0.63±0.03) and p60 (average UMAP descriptor: 0.37±0.04) comparisons was not possible to obtain a clear dissociation (**Fig. 4B-C, Table 2**).

The inclusion of slow and fast fibre discrimination to geometric features (15 ccs) provided an increase in all age-stages separations. For p120 analysis, the larger standard deviation of slow and fast fibres area (ccs. 4 and 6) in SOD1^{G93A} than in healthy mice enhanced the dissociation (average UMAP descriptor: 3.25 ±0.15) (**Fig. 4D, Table 2**). In p100 evaluation, the greater variance in fast fibre area (cc. 6) and the higher proportion of slow/fast fibres (cc. 67) in mutants than in WT, as in PCA, raised the UMAP value close to 1 (average UMAP descriptor: 0.94±0.04) (**Fig. 4E, Table 2**). As in p100, in the presymptomatic comparison at p60, the major number of slow fibres in relation to fast cells (cc. 67) and the larger standard deviation of fast fibres area (cc. 6) in ALS mouse model improved the separation close to 1 (average UMAP descriptor: 0.95±0.08) (**Fig. 4F, Table 2**).

Finally, using the whole set of 67 features to find the hallmark of ALS in mouse model, all SOD1^{G93A} and WT ages groups enhanced using UMAP instead of PCA method. This was not always the case when working with smaller subsets of characteristics (subsets 10 ccs and 15 ccs) (**Fig. 4, Table 2**). In p120 comparison (average UMAP descriptor: 3.95±0.21), the most frequent selected features were (ordered by frequency from left to right): standard deviation cell area (cc. 2); average (cell/neighbouring cells) minor axis (cc. 27); standard deviation slow cells area (cc. 4); average cell convex hull (cc. 11); standard deviation (cell/neighbouring cells) minor axis (cc. 28); standard deviation cell convex hull (cc. 12); average (cell/neighbouring cells) area (cc. 23)(**Fig. 4G, Table 2**). All these features were related to more heterogeneous fibres shape and size in mutant than in WT mice, a similar result than using PCA algorithm (**Table 2**). In p100 comparison (average UMAP descriptor: 1.26±0.03), the emerging features with larger number of presences along the 100 realizations were: standard deviation cell area/(cell area + collagen area) (cc. 14); standard deviation fast cells area (cc. 6); average (cell/neighbouring cells) convex hull (cc. 31); fast cells neighbours of slow cells (cc. 20); standard deviation betweenness centrality of slow cells (cc. 58); standard deviation (cell/neighbouring cells) (major/minor axes) (cc. 30); average clustering coefficient of slow cells (cc. 45) (**Fig. 4H, Table 2**). Here, three of the features previously observed in PCA analysis showed up (ccs. 14, 6, 30). Besides it came out some other features such as the larger

heterogeneity in the convex shape of fibres (cc. 31) in mutant. Or more important, some network features related with the type of fibre (ccs. 20, 58, 45). These features could be influenced by a larger clustering of slow cells in SOD1^{G93A} than in WT. In p60 comparison (average UMAP descriptor: 1.30±0.10) the selected features were: fast cells neighbours of slow cells (cc. 20); standard deviation fast cells area (cc. 6); standard deviation neighbouring cells of fast cells (cc. 18); fast cells neighbours of fast cells (cc. 22); average (cell/neighbouring cells) convex hull (cc. 31); standard deviation shortest paths length from fast cells to fast cells (cc. 62); average cell minor axis (cc. 8) (**Fig. 4l, Table 2**). Comparing these characteristics with PCA, it matched three features (ccs. 20, 18, 60). Anyhow, as in 100 days analysis, it emerged ccs. 6 and 31 probably due to the larger diversity of fast cells size (cc. 6), and more different convexity of neighbouring cells (cc. 31) in ALS mouse model than in WT, together with another geometric feature (cc. 8) due to the larger cell minor axis of mutant fibres. Additionally, cc. 22 was selected probably because fast fibres arrangement is more scattered in SOD1^{G93A} tissue (less clusters of fast cells) than in WT, which are more grouped.

Importantly, the univariate statistical analysis of the features appearing differentiating mutant and WT at different age stages showed mostly significant difference when comparing SOD1^{G93A} and WT features values (p-value<0.05 *, **Table 2**, **methods**). Moreover, as in PCA, none of them was both, selected and significant in the three age comparisons (**Table 2**). However, in general terms, the number of individual significant differences after using UMAP was equal or better than using PCA.

NDICIA to question the role of SOD1^{WT} mice as ALS model.

The previous results have shown that NDICIA is able to find differences between samples from WT and affected mice. One possibility could be that the computational model would be involved into an overfitting, forcing the separation of any two tissues, even being tissues with similar characteristics. To confirm the biological relevance of the differences found, we decided to add a third animal model to the comparisons, the overexpression of the SOD1^{WT} construct. We focused on the p120 stage where the separation between WT, and SOD1^{G93A} samples were clearest and used the whole set of features to compare the

previous images with images from $SOD1^{WT}$ (**Fig. 1E-F**, **Fig. 5A**). First, using the PCA strategy, in the case of the comparison between WT and $SOD1^{WT}$ samples, the PCA value was very small indicating that both set of images were quite similar (PCA descriptor: 0.26, **Fig. 5B**). However, when we compared $SOD1^{WT}$ and $SOD1^{G93A}$, found a clear separation between the samples displaying a high PCA descriptor (PCA descriptor: 2.57, **Fig. 5C**). Interestingly, 5 out of the 7 features (ccs. 2, 4, 12, 24 and 27) were also present in the PCA comparison between p120 $SOD1^{G93A}$ and WT samples (**Fig. 3A**). The results were equivalent when we applied the UMAP approach to discard possible non-linear relationship among features enabling a good separation for $SOD1^{G93A}$ and $SOD1^{WT}$ groups. In the case of WT vs $SOD1^{WT}$ was not possible to find a separation between the groups (average UMAP descriptor: 0.36±0.02, **Fig. 5D**), while $SOD1^{G93A}$ and $SOD1^{WT}$ were clearly differentiated (average UMAP descriptor: 3.14±0.09, **Fig. 5E**). Both sets of results suggested that $SOD1^{WT}$ samples are very similar to WT in terms of muscle fibres shape and size, and drastically different from $SOD1^{G93A}$. Additionally, these results shed light on the controversy surrounding the potential pathogenic role of SOD1 protein overexpression in ALS disease. The results refute this theory and characterize $SOD1^{WT}$ mouse model as healthy mouse in terms of cell shape and organization.

Discussion

As of now, a conclusive ALS diagnosis test or biomarker remains elusive. Currently, ALS diagnosis relies mainly on physician observation of patient signs, along with electromyography (de Carvalho et al., 2005), and a battery of tests to rule out other diseases (Brooks, 1994; Brooks et al., 2000). These clinical data allow physicians to determine different evidence levels for suffering from ALS (suspected, possible, probable and definitive diagnosis). The importance of identifying biomarkers to validate new treatments is evident. In this study, we study the $SOD1^{G93A}$ mouse, a well-known murine model in ALS preclinical research (Picher-Martel et al., 2016). By using a computational approach, we quantify and evaluate differences between mutant and healthy mouse muscles, with the ultimate goal of identifying potential ALS image biomarkers. The muscles samples analysed across this study range from 60 to 120 days of age (classified

in p60, p100 and p120 groups), where at around 3-4 months of age, the ALS onset is noticeable in the SOD1^{G93A} mice (Ciuro et al., 2023; De Giorgio et al., 2019). In this work, we have tried to characterize the illness properties after ALS phenotype appears and to unveil any possible indicator of ALS presence in presymptomatic stages, when morphology appears normal under optic microscopy examination.

Our robust multivariate analysis by using the NDICIA software protocol (Sáez, Acha, et al., 2013), which has been successfully applied to study muscle organization (Sáez, Acha, et al., 2013; Sánchez-Gutiérrez et al., 2017), reports how different two groups of images are, and identifies the features that most effectively distinguish between them. Applying NDICIA to our p100 and p120 muscle samples, after ALS symptoms presumably came out, we found clear separations of WT and mutant phenotypes in terms of PCA value, with the separation becoming more pronounced as the illness progress. Therefore, NDICIA objectively perceived the differences between the WT and mutant muscles. We found that the main properties to separate both groups were related to the fibres geometry, being more homogeneous in size and shape in WT than in SOD1^{G93A}, noticing the pattern of myofibre neurogenic atrophy in affected tissue. Additionally, it was also important in the analysis the higher quantity and heterogeneous distribution of collagen along the SOD1^{G93A} tissue. These differences increased as the mice get older. These results are obvious simply comparing WT and SOD1^{G93A} biopsies by naked eye. The atrophy of muscle fibres in SOD1^{G93A} after symptoms is reflected in the irregularity and heterogeneity of the fibre shape, while the WT muscle fibres geometry is more homogeneous. The network features associated with the type of fibre (slow/fast) were not useful in differentiating WT and SOD1^{G93A} in p100 and p120.

On the other hand, the results obtained in p60 mice were unexpected. Despite of similar morphology properties in presymptomatic SOD1^{G93A} and WT samples, NDICIA identified significant subgroup separation primarily based on network properties related to fibre type discrimination. The results suggest the presence of larger clusters of slow fibres in SOD1^{G93A} muscle tissues than in WT, where fast fibres are scattered throughout the tissue. Although the proportion of slow/fast fibres is always larger in mutant at any stage (p60, p100 or p120), and

this could bias the network results, it is only at p60 that these kinds of properties enhance the separation. These findings could be associated with the early changes on neuromuscular transmission detected in (Rocha et al., 2013), long before the onset of motor symptoms. Additionally, the higher number of slow fibres in SOD1^{G93A} could be related to the well-known preferential denervation of fast fibres that in a second step, is partially compensated with fast-to-slow fibre type transitions (Hegedus et al., 2008). We acknowledge that our graph theory approach has some limitations, such as the few number of cells embedded in an image. However, our results suggest an interesting way to test if a more specific graph theory analysis of larger muscle samples (larger networks) could unveil organizational properties (i.e. finding a specific and presymptomatic progressive pattern of fast-to-slow fibre type transitions, resulting in grouping slow cells in a restricted way). If so, this computational image analysis of muscle biopsies could be used as a diagnostic tool to support the diagnosis of ALS in early stages, highlighting the importance of network features for early diagnosis. To this end, our analytical tool could benefit from the advanced capabilities of deep learning-based techniques for image processing and segmentation. In particular, we propose the use of convolutional neuronal networks (CNNs) in future works, such as the widely used U-Net (Falk et al., 2019), to accurately segment and identify muscle fibres and extracellular matrix in biopsy images. To achieve this, the CNNs learn from a comprehensive training dataset, enabling them to automatically carry out the segmentation process with high precision. In that sense, the processed images from our study could serve as a valuable resource to build that training dataset to feed the U-Net or other similar CNNs, for automatic prediction on new samples. Overall, it could alleviate the bottleneck of image processing, boosting the data analysis.

We understand how crucial it is to accurately represent and analyse high-dimensional data in order to diagnose effectively using our tool. Therefore, we are continuously seeking better methods to reduce the dimensionality of our data and enhance the performance of NDICIA. Likewise, we integrated UMAP into our pipeline, a successful and more sensitive tool than previous methods to visualize complex expression profiling data (Cao et al., 2019; Dorrity et al., 2020; Packer et al., 2019). Thus, in order to get possible better separations and different

combinations of selected features taking advantage of the non-lineal relationships considered by UMAP, we tested the same comparisons WT-SOD1^{G93A} as we studied with PCA. The weakest point of this new algorithm is its stochastic nature, and for this reason, we develop a hundred of repeated simulations in order to get a consistent average result. Anyhow, when the analyses were developed considering the whole set of features (67 ccs), all the results slightly improved. In p60 and p120, the results with UMAP analysis were very similar to those obtained with PCA method; that is, at 60 days network ccs related with the type of fibre are main precursors of the separation, while in p120 the geometric properties take the main role. The major difference is observed in p100, where some network properties related with the fibre type in combination stand out (together with other geometric features previously described in PCA method), suggesting again a larger clustering of slow fibres in ALS model. As expected, the information provided by UMAP corroborated that p100 mice are in an intermediate stage between p60 and p120. It can be inferred a gradual transition from network ccs at 60 days, to geometric ones at 120 days as main performers separating WT and SOD1^{G93A}, probably due to the clear presence of muscle atrophy. Therefore, besides the support carried out by UMAP tool in this study, its performance has been validated to be used by NDICIA in future works.

Following the classic (PCA) and the updated (UMAP) NDICIA protocols, we attempted to analyse the discrepancy in the scientific community: could SOD1 overexpression in mice (SOD1^{WT}) be considered an ALS model? Some reviews have suggested that misfolded SOD1 protein aggregates present in post-mortem ALS tissues could play an important causative role in SALS due to its oxidative and toxic capacity (Bosco et al., 2011; Furukawa, 2012). Anyway, it is still unclear if SOD1 aggregates may “cause” ALS or if it is a harmless downstream effect of the disease. Trying to elucidate this question, we compared SOD1^{WT} with WT and SOD1^{G93A} mice at p120. WT and SOD1^{WT} groups were almost impossible to differentiate, suggesting they display very similar properties. When we compared SOD1^{WT} and SOD1^{G93A}, the results showed a clear separation between both groups, similarly to what happened between WT and SOD1^{G93A}, suggesting that these two genotypes were very different. Moreover, when we compared SOD1^{WT} vs SOD1^{G93A} and WT vs SOD1^{G93A}, most of the relevant ccs that emerged in the

image differentiation analysis were the same in both comparisons. These ccs were related to a higher homogeneity between the myofibers in $SOD1^{WT}$ and WT than in $SOD1^{G93A}$ muscle. Our results support that $SOD1^{WT}$ mice exhibit similar features to healthy mice and therefore could not be considered an ALS mice model so far.

In summary, our study has demonstrated the remarkable potential of NDICIA for clinical applications in identifying the signature of ALS in muscle as a potential biomarker, even when the disease is clinically imperceptible. This early pathological detection could drive the specific clinical and preclinical trials, bringing forward the onset of potential treatment, before muscle degeneration becomes consequential, thereby would allow to intervene not only in the progression but also in the onset of the disease.

METHODS

Transgenic mice

The transgenic mice colony was obtained by crossing [B6SJL-TgN(SOD1-G93A) 1Gur/J] ($SOD1^{G93A}$) males with wild type B6SJL (WT) females to maintain the genetic background. All animals were provided by Jacksons Laboratories. Non-transgenic littermates ($SOD1^{WT}$) were used as age matched controls. All animal experiments were performed in accordance with the Spanish and the European Union regulations (RD53/2013 and 2010/63/UE) and approved by the Animal Research Committee from the University of Seville (Spain). Animals were housed under controlled temperature and humidity conditions, alternating 12-hour light cycles. Experiments and procedures with animals were designed to minimize the animal suffering and reduce the number of animals used.

Tissue sampling

Samples were obtained from $SOD1^{G93A}$, $SOD1^{WT}$ mouse models and WT at three different ages stablished in order to track the progression of ALS disease. A presymptomatic group, at postnatal day 60 (p60), where the mice are mostly non-symptomatic for ALS. An intermediate group, at 100 days (p100), represents a progressing disease stage. And finally, we stablished the last group at 120 days of age (p120), where most of the mice are severely affected by motor symptoms.

$SOD1^{WT}$ mice muscle were only imaged at p120 to be compared against WT and $SOD1^{G93A}$ mice.

Tissue samples were obtained from each mouse line at the corresponding time point. For the muscle extraction, animals were euthanized by cervical dislocation and soleus muscle was isolated from hind limbs. Muscle tissue was preserved by snap-freezing. To avoid damage to the cells during the process, the whole muscle was first covered completely using talc powder and then immersed in N_2 during 1 minute. All samples were placed in cryogenic safe tubes and stored at -80ºC.

Our total database of imaged mice muscles consisted of 386 images collected from 65 different mice. For p60 comparisons, we collected 27 images from 5 WT mice and 19 images from 4 $SOD1^{G93A}$ mice. For p100, we acquired 78 images from 14 WT mice and 37 images from 5 $SOD1^{G93A}$ mice. Finally, for p120, we obtained 103 images from 11 WT mice, 47 images from 14 $SOD1^{G93A}$ mice, and 75 images from 12 $SOD1^{WT}$ mice.

Immunofluorescence

For immunofluorescence, soleus was cut transversally into 10 µm sections, using a cryostat (Leica), and placed sequentially onto Superfrost™ Plus microscope slides, in order to obtain serial slices of each muscle sample. Tissue slices were fixated using 4% paraformaldehyde for 20 minutes at room temperature and permeabilized by incubating 10 minutes with PBS-0.1% Triton X-100. Continuedly, tissue was blocked using 1% BSA, during 1 hour at room temperature, and incubated with primary antibodies over night at 4ºC. The following antibodies were used: to detect muscle fibre types we used mouse anti-myosin heavy chain (slow) (Leica, clone WB-MHCs; 1:200); to delimit the fibres we used rabbit anti-collagen type VI (Abcam, ab6588, 1:300). Finally, slices were incubated 1 hour at room temperature with the following secondary antibodies: Cy3 donkey anti-mouse (#715-165-151, Jackson Immunoresearch, 1:500), Alexa488 donkey anti-rabbit (#715-545-150, Jackson Immunoresearch, 1:500). After staining, the preparations were mounted using Fluoromount-G (#00-4958-02, Invitrogen). All images were captured with a 40X objective, under identical conditions and using the same settings in a fluorescence microscope (Nikon

Eclipse Ti-E), with a XY resolution of 0.104 µm per pixel, and an image size of 4080 x 3072 pixels.

NDICIA. Image segmentation

Segmentation is the first step in the image processing protocol. In this stage, the collagen network and the fibres must be clearly differentiated among themselves (**Fig. 2E**). This is a semi-automatic process carried out from morphological operations in Matlab R2020b (Mathworks) inspired by (Sáez, Acha, et al., 2013). This procedure was developed as follows: (1) The starting point was disjointing the colour channels of the (RGB) images got from microscope. The green channel represented the extracellular matrix, mainly composed of collagen fibres (**Fig. 2C**). This one was our mainstay in the segmentation of muscle fibres, demarcating their contours. The red channel denoted the type of muscle fibres (slow or fast fibres). (2) After green channel selection, the intensity map was auto-adjusted to enhance the contrast for better differentiation between the fibres and extracellular matrix. (3) We used a method of segmentation based on h-minima and h-maxima transformation (Sáez, Acha, et al., 2013). This type of segmentation has been successfully applied in several medical enforcements (Ali et al., 2011; Chanho Jung & Changick Kim, 2010; Sáez, Acha, et al., 2013). We transformed the green channel of the image using different h-values. For each h-value, we got a range of intensities in which the pixels representing the extracellular matrix were connected. Overlapping the pixels obtained in the individual experiences, we accomplished a complete mesh of connective tissue, joining cells. So, the “holes” in the mesh corresponded with the muscle fibres. (4) We removed the noise and the artefacts of the images by comparative of the individual cell areas. (5) Next, some morphological operations were carried out for filling little holes, compacting, and splitting joined fibres. (6) Finally, manual curation by expert was necessary. The software used was Adobe Photoshop CS6.

NDICIA. Fibre type classification

The detection and classification of muscle fibre types into fast or slow fibres was carried out after the initial segmentation process. Considering that red channel (R) specifically stained the slow fibres, we classified a muscle fibre as a

slow fibre if one if the following conditions were fulfilled: 1) The average intensity value in the fibre should be higher than a quarter of the maximum intensity value in the red channel image. 2) A minimum of fifteen per cent of all the intensity values in the fibre area must be greater than half of the maximum intensity value of the red channel image. This helped to avoid detecting fast fibres with isolated intensity peaks due to image artefacts.

Otherwise, the muscle fibre was considered as fast fibre. After this automatic process we ensure the classification with a supervised review of the whole set of images by using a custom Matlab script.

NDICIA. Features extraction

Once the contours of the muscle fibres were well segmented and the types of fibre classified, we extracted and quantified some properties from the tissues (**Fig. 2F, Table 1**):

Geometric features. A set of characteristics were extracted measuring geometric properties of the image regions. These features were related to the area, the minor and major axes length and the geometric shape of the fibres (ccs. 1-12), besides the amount of collagen surrounding the muscle fibres (ccs. 13-14). The region occupied by collagen that “belongs” to a specific cell is calculated using the Voronoi algorithm (Voronoi, 1908) with the segmented fibres as source (Voronoi seeds). This procedure estimated the equidistant "frontiers" between the muscle fibres. Thus, the region of collagen fibres enclosed within the frontiers of a muscle fibre “belongs” to that fibre.

Network features. These types of features were elicited from the connectivity between muscle fibres. The cell connections were defined by leveraging the sharing fibres frontiers generated after applying the previously cited Voronoi algorithm. Thus, two cells sharing a frontier were defined as neighbours. In this way, some relational features were extracted from each neighbourhood: the number of neighbour’s fibres of each fibre and the ratio between a fibre geometric property (area, major axis, minor axis, convexity, or quantity of surrounding collagen) in comparison with its neighbouring cells (ccs. 15-34, **Table 1**). These features were used to measure properties of organization. Moreover, after considering the tissue as a network, we calculate some graph theory

characteristics such as eccentricity, clustering coefficient, strength, betweenness centrality, shortest path length (ccs. 35-66, **Table 1**). Lastly, the proportion of slow and fast fibres along the muscle sample was calculated (cc. 67, **Table 1**).

NDICIA. Features selection

In the final step of NDICIA we executed the features selection module. To this aim, we run a protocol based on dimensionality reduction that choose a subset of features that allow the best separation between two collections of images. In this work, we have used two different dimensionality reduction methods, the Principal Component Analysis (PCA) and the Uniform Manifold Approximation and Projection (UMAP). With the use of these methods, we transformed and projected the full set of features (67 dimensions) into two dimensions. At this point, we could analyse the position of the "points" (images) along the two dimensional graphic and estimate how much separate the two groups of images are. To evaluate this degree of separation, we used a variant of the Calinski-Harabasz descriptor (Costa & Cesar, 2009), carefully explained in (Sánchez-Gutiérrez et al., 2013b). This value is what we have called along the manuscript as PCA or UMAP descriptor (depending on the previously used algorithm).

The features selection module makes use of the dimensionality reduction method and the separation descriptor in an iterative way to capture the most relevant features distinguishing better the two groups of images. The pipeline was established as follows: 1. We run the dimensionality reduction method for each pair of features and store the 10 pairs of features with a higher separation descriptor. 2. From each of the 10 selected pairs of features, we test all the possible trios by adding a feature. Thus, we run again the dimensionality reduction method and store the 5 trios with larger separation descriptor per each initial pair of features. At this moment, we account with 50 trios of characteristics. 3. The third iteration is similar to the second step but keeping only the two best combinations of quartets from each initial trio, getting a total of 100 quartets. 4. Similar than before but adding only one feature to the previous 100 combinations of features, that is, the feature that more enhance the separation descriptor. 5. The step 4 is repeated until reaching a maximum of 7 features per combination or until the separation descriptor is lower than the step before. Finally, we have reached a total of 100 combinations of features with a specific separation

descriptor each. The result of this protocol will be the combination of features being the one with the highest separation descriptor value.

The weight of each feature in the separation is given by the larger absolute value of him corresponding eigenvector, enabling the transformation of the features values in the final 2D projections.

NDICIA. Dimensionality reduction methods

PCA algorithm is the classical method used for dimensionality reduction in NDICIA. This technique assumes data follow a lineal combination and transform the data values in two (usually) principal components, that is, two axes in which the variability of data is larger.

On the other hand, we have upgraded the dimension reduction of NDICIA after switching PCA with UMAP. UMAP is a novel manifold learning technique for dimension reduction that works with non-linear relationships (McInnes et al., 2018). This is a stochastic method that needs to set several hyperparameters, being the key parameters:

n_neighbours: size of local neighbourhood (in terms of number of neighbouring sample points) used for manifold approximation. Larger values result in more global views of the manifold, while smaller values result in more local data being preserved. Range: 0 - 100

min_dist: the effective minimum distance between embedded points. Smaller values will result in a more clustered/clumped embedding where nearby points on the manifold are drawn closer together. Larger values will result on a more even dispersal of points. Range: 0 - 0.99.

‘n_components’, the dimension of the space to embed into.

‘metric’: kind of metric to use to compute distances in high dimensional space.

For this adjustment we have used the UMAP matlab implementation (v1.3.4) (Meehan Connor, Ebrahimian Jonathan, Moore Wayne, 2020), applying the default values for hyperparameters. ‘n_neighbours’ = 30, ‘min_dist’ = 0.3, ‘n_components’ = 2 and ‘metric’ = euclidean. With the selection of these values,

we tried to get an embedding output that captures a balance between local and global differences.

In order to use the UMAP in combination with our features selection method for feasible diagnosis (despite it lacks reproducibility), we have repeated 100 times our features selection pipeline. With each of these 100 realizations we got a combination of features that provides the best separation between two groups of images. We stored all these combinations of features and chose, by order of frequency, the more repeated features along the 100 realizations. If the more usual length of combinations of features was six features, then we chose only the six more repeated features. Moreover, we got the mean (and standard deviation of the mean) UMAP descriptor considering the average of all the UMAP descriptors for each realization. Finally, to get a visual representation of the protocol result, we plotted only the graph of the randomization with the more similar UMAP descriptor to the average UMAP value (**Fig. 4 and Fig. 5D-E**).

Statistics

After applying NDICIA, we obtained the best set of features separating groups of images at different age stage. So, to inquiry in the individual power of these features across all age stages, we applied a univariate analysis comparing the distribution of values between the two groups of images (WT and $SOD1^{G93A}$) (**Table 2**). We evaluated whether features values of these two kinds of data presented normal distribution and similar variance using Kolmogorov-Smirnov test with the correction of Lilliefors (if $n<50$ we used Shapiro-wilk test) and two-sample F-test respectively. If data followed a normal distribution and had similar variance, we employed the two-tailed Student's t-test. In case data presented normal distribution but not equal variance we employed the two-tailed Welch test to compare means from both groups. If data did not present normal distribution, we used the Wilcoxon test to compare medians from both groups.

Code and data availability

The full code used along this project was developed using Matlab R2020b (Mathworks). The code is freely accessible under the GPLv3 license on https://github.com/ComplexOrganizationOfLivingMatter/MouseMuscleSOD1. Images and all the data necessary to reproduce this study will available after publication.

FIGURES LEGEND

Figure 1. Images of WT and SOD1^{G93A} soleus muscles biopsies. Fluorescence images showing slow (sMHC, red stained) and fast fibres (non-stained, observable in dark green as counterpart of slow fibres), connected by collagen VI content including the endomysium and perimysium (light green). Collagen is demarcating the fibres outlines, and slow fibres are stained due to the presence of slow myosin heavy chain staining. (**A**, **C**, **E**) Images from WT samples of soleus muscle. (**B**, **D**, **F**) Images from SOD1^{G93A} mutant mice. The images are also classified by age groups: p60 (**A**, **B**), p100 (**C**, **D**) and p120 (**E**, **F**). sMHC tags for myosin heavy chain in slow fibres, while col VI indicates the collagen VI staining. Scale bar represents 50 µm.

Figure 2. NDICIA features extraction pipeline. NDICIA processes the images from the microscope (**A**) leveraging the information provided by the red channel (**B**), staining the slow muscle fibres, and the green channel (**C**), staining the collagen VI that delimits the cell compartments. Green channel allows for segmentation of muscle fibres (**C**, **E**), while red channel for fibre type detection (**B**, **D**). A connectivity network is built by linking neighbouring muscle fibres (**F**) and a set of 67 characteristics (geometric, network and fibre type proportion) are extracted. In (**D**), red segmented cells represent to slow fibres, meanwhile the dark green ones to fast fibres. In (**F**), same colour code applies for the nodes used to construct the tissue network that discriminate the type of fibre, where the neighbouring cells are linked by a blue line.

Figure 3. WT vs SOD1^{G93A}, NDICIA comparison (PCA). PCA graphs generated as outputs by NDICIA after feature selection. The selected characteristics are indicated with its numerical index atop each graph. Each illustrates the reduction of dimensionality from the selected features to the two principal components (PCA 1 on the X axis and PCA 2 on the Y axis). These graphs are categorized in three groups based on the subset used as source for NDICIA: Geometric (10 ccs, **A, B, C**), Geometric including fibre type discrimination (15 ccs, **D, E, F**) and full set of features (67 ccs, **G, H, I**). Furthermore, the PCA graphs are further subdivided by age: p120 (**A, D, G**), p100 (**B, E, H**) and p60 (**C, F, I**). Red dots

represent each SOD1^{G93A} sample, while blue dots the WT ones. The PCA values plotted on the graphs show how distinct the SOD1^{G93A} and WT groups are: the closer to the value is to 0, the more they overlap, while values greater than 1 indicate a clear separation between the groups (**methods**).

Figure 4. WT vs SOD1^{G93A}, NDICIA comparison (UMAP). Similar panels as in Figure 3 but using UMAP as reduction of dimensionality method instead of PCA. Here, the selected features are the most frequent selected along 100 repetitions of NDICIA pipeline using UMAP. The number of features, with a maximum of 7 features as in PCA, is determined by calculating the mode value after evaluating the number of selected features across 100 randomizations (**methods**). The UMAP value displays the average and standard deviation considering the 100 repetitions (**methods**). Additionally, each graph displayed in the different panels embodies to the randomization with the more similar UMAP value to the average UMAP value. The colour code and panels order are akin to Figure 3.

Figure 5. Analysis of pathological condition of SOD1^{WT} mice. In panel **A**, it can be observed an example of p120 SOD1^{WT} soleus muscle biopsy. As in **Fig. 1**, sMHC tags for myosin heavy chain in slow fibres, while col VI indicates the collagen VI staining. Scale bar represents 50 µm. Next, it is plotted the NDICIA results comparing SOD1^{WT} vs WT (**B, D**) and SOD1^{WT} vs SOD1^{G93A} (**C, E**) at 120 days using the full set of 67 ccs. First, we handled with PCA method (**B, C**) and second with UMAP (**D, E**), Colour codes and panels organization follow the same rules than in Figure 3 and 4. Green dots represents SOD1^{WT} samples.

Table 1. Set of features analysed using NDICIA. Table shows the name and description of 67 characteristics analysed along this article. The characteristics can be classified as geometric (1-14), network (15-66) and the proportion of slow cells (67). The bold names mark the properties related with slow or fast cell type discrimination.

Table 2. Quantitative and statistical results after NDICIA application. This table is organized in three sheets, 67 ccs, 15 ccs and 10 ccs. Each one collects all NDICIA results for each comparison developed along the manuscripts using as source the subset of 67, 15 or 10 ccs: PCA descriptor; selected features; absolute weights of these features in the PCA graph discrimination; UMAP

descriptor; average and standard deviation of the features values in each compared group, and an individualized statistical analysis to explore if the selected features present significant difference when comparing the different groups of images. The statistical tests were selected using a protocol depending on the data characteristics (**methods**), and the goodness of the statistical difference was evaluated by three levels of *p*-value: <0.05*, <0.01**, 0.005***.

Bibliography

Ali, K., Jalil, A., Gull, M. U., & Fiaz, M. (2011). Medical Image Segmentation Using H-minima Transform and Region Merging Technique. *2011 Frontiers of Information Technology*, 127–132. https://doi.org/10.1109/FIT.2011.31

Audet, J.-N., Gowing, G., & Julien, J.-P. (2010). Wild-type human SOD1 overexpression does not accelerate motor neuron disease in mice expressing murine Sod1G86R. *Neurobiology of Disease*, *40*(1), 245–250. https://doi.org/10.1016/j.nbd.2010.05.031

Behan, W. M., Cossar, D. W., Madden, H. A., & McKay, I. C. (2002). Validation of a simple, rapid, and economical technique for distinguishing type 1 and 2 fibres in fixed and frozen skeletal muscle. *J Clin Pathol*, *55*(5), 375–380.

Bosco, D. A., LaVoie, M. J., Petsko, G. A., & Ringe, D. (2011). Proteostasis and Movement Disorders: Parkinson's Disease and Amyotrophic Lateral Sclerosis. *Cold Spring Harbor Perspectives in Biology*, *3*(10), a007500–a007500. https://doi.org/10.1101/cshperspect.a007500

Bosco, D. A., Morfini, G., Karabacak, N. M., Song, Y., Gros-Louis, F., Pasinelli, P., Goolsby, H., Fontaine, B. A., Lemay, N., McKenna-Yasek, D., Frosch, M. P., Agar, J. N., Julien, J.-P., Brady, S. T., Brown, R. H., & Jr. (2010). Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. *Nature Neuroscience*, *13*(11), 1396–1403. https://doi.org/10.1038/nn.2660

Brooks, B. R. (1994). El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and th. *J Neurol Sci*, *124 Suppl*, 96–107.

Brooks, B. R., Miller, R. G., Swash, M., Munsat, T. L., & World Federation of Neurology Research Group on Motor Neuron Diseases. (2000). El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders : Official Publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases*, *1*(5), 293–299.

Bruijn, L. I., Becher, M. W., Lee, M. K., Anderson, K. L., Jenkins, N. A., Copeland, N. G., Sisodia, S. S., Rothstein, J. D., Borchelt, D. R., Price, D. L., & Cleveland, D. W. (1997). ALS-Linked SOD1 Mutant G85R Mediates Damage to Astrocytes and

Promotes Rapidly Progressive Disease with SOD1-Containing Inclusions. *Neuron*, *18*(2), 327–338. https://doi.org/10.1016/S0896-6273(00)80272-X

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x

Chanho Jung, C., & Changick Kim, C. (2010). Segmenting Clustered Nuclei Using H-minima Transform-Based Marker Extraction and Contour Parameterization. *IEEE Transactions on Biomedical Engineering*, *57*(10), 2600–2604. https://doi.org/10.1109/TBME.2010.2060336

Ciuro, M., Sangiorgio, M., Leanza, G., & Gulino, R. (2023). A Meta-Analysis Study of SOD1-Mutant Mouse Models of ALS to Analyse the Determinants of Disease Onset and Progression. *International Journal of Molecular Sciences*, *24*(1), 216. https://doi.org/10.3390/IJMS24010216/S1

Costa, L. da F., & Cesar, R. M. (2009). Shape classification and analysis: Theory and practice, second edition. In *Shape Classification and Analysis: Theory and Practice* (2th ed.). CRC Press.

Czarzasta, J., Habich, A., Siwek, T., Czapliński, A., Maksymowicz, W., & Wojtkiewicz, J. (2017). Stem cells for ALS: An overview of possible therapeutic approaches. *International Journal of Developmental Neuroscience*, *57*, 46–55. https://doi.org/10.1016/j.ijdevneu.2017.01.003

de Carvalho, M., Chio, A., Dengler, R., Hecht, M., Weber, M., & Swash, M. (2005). Neurophysiological measures in amyotrophic lateral sclerosis: Markers of progression in clinical trials. *Amyotrophic Lateral Sclerosis*, *6*(1), 17–28. https://doi.org/10.1080/14660820410020600

De Giorgio, F., Maduro, C., Fisher, E. M. C., & Acevedo-Arozena, A. (2019). Transgenic and physiological mouse models give insights into different aspects of amyotrophic lateral sclerosis. *Disease Models & Mechanisms*, *12*(1). https://doi.org/10.1242/dmm.037424

Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S., & Trapnell, C. (2020). Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nature Communications*, *11*(1), 1–6. https://doi.org/10.1038/s41467-020-15351-4

Dubowitz, V., & Sewry, C. A. (2007). Muscular dystrophies and allied disorders V: Facioscapulohumeral, myotonic and oculopharyngeal muscular dystrophy. In *Muscle Biopsy* (pp. 395–405). Elsevier. https://doi.org/10.1016/B978-1-4160-2593-1.50019-X

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., … Ronneberger, O. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*. https://doi.org/10.1038/s41592-018-0261-2

Fontelonga, T. M., Jordan, B., Nunes, A. M., Barraza-Flores, P., Bolden, N., Wuebbles, R. D., Griner, L. M., Hu, X., Ferrer, M., Marugan, J., Southall, N., & Burkin, D. J. (2019). Sunitinib promotes myogenic regeneration and mitigates disease

progression in the mdx mouse model of Duchenne muscular dystrophy. *Human Molecular Genetics*, *28*(13), 2120–2132. https://doi.org/10.1093/hmg/ddz044

Furukawa, Y. (2012). Pathological Roles of Wild-Type Cu, Zn-Superoxide Dismutase in Amyotrophic Lateral Sclerosis. *Neurology Research International*, *2012*, 1–6. https://doi.org/10.1155/2012/323261

Garton, F., Seto, J. T., North, K. N., & Yang, N. (2010). Validation of an automated computational method for skeletal muscle fibre morphometry analysis. *Neuromuscul Disord*, *20*(8), 540–547. https://doi.org/S0960-8966(10)00266-X [pii] 10.1016/j.nmd.2010.06.012

Gibson, M. C., Patel, A. B., Nagpal, R., & Perrimon, N. (2006). The emergence of geometric order in proliferating metazoan epithelia [Journal Article]. *Nature*, *442*(7106), 1038–1041.

Graffmo, K. S., Forsberg, K., Bergh, J., Birve, A., Zetterstrom, P., Andersen, P. M., Marklund, S. L., & Brannstrom, T. (2013). Expression of wild-type human superoxide dismutase-1 in mice causes amyotrophic lateral sclerosis. *Human Molecular Genetics*, *22*(1), 51–60. https://doi.org/10.1093/hmg/dds399

Gurney, M. E. (1997). The use of transgenic mouse models of amyotrophic lateral sclerosis in preclinical drug studies. *Journal of the Neurological Sciences*, *152*, s67–s73. https://doi.org/10.1016/S0022-510X(97)00247-5

Gurney, M. E., Pu, H., Chiu, A. Y., Dal Canto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., & Deng, H. X. (1994). Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science (New York, N.Y.)*, *264*(5166), 1772–1775. http://www.ncbi.nlm.nih.gov/pubmed/8209258

Guy, J., Gan, J., Selfridge, J., Cobb, S., & Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science (New York, N.Y.)*, *315*(5815), 1143–1147. https://doi.org/10.1126/science.1138389

Gwinn, K., Corriveau, R. A., Mitsumoto, H., Bednarz, K., Brown, R. H., Cudkowicz, M., Gordon, P. H., Hardy, J., Kasarskis, E. J., Kaufmann, P., Miller, R., Sorenson, E., Tandan, R., Traynor, B. J., Nash, J., Sherman, A., Mailman, M. D., Ostell, J., Bruijn, L., … Group, for T. A. R. (2007). Amyotrophic Lateral Sclerosis: An Emerging Era of Collaborative Gene Discovery. *PLoS ONE*, *2*(12), e1254. https://doi.org/10.1371/journal.pone.0001254

Hayashi, T., & Carthew, R. W. (2004). Surface mechanics mediate pattern formation in the developing retina. *Nature*, *431*(7009), 647–652. https://doi.org/10.1038/nature02952

Hegedus, J., Putman, C. T., Tyreman, N., & Gordon, T. (2008). Preferential motor unit loss in the SOD1 G93A transgenic mouse model of amyotrophic lateral sclerosis. *The Journal of Physiology*, *586*(14), 3337–3351. https://doi.org/10.1113/jphysiol.2007.149286

Helliwell, T. R. (1999). Muscle: Part 1 - Normal structure and function. *Current Orthopaedics*, *13*, 33 41.

Hilton, J. B., White, A. R., Crouch, P. J., Basak, A. N., Wilhelm, C., Gobbi, C., Birve, A., Biskup, S., Zecca, C., Fernandez-Santiago, R., Kaugesaar, T., Schols, L.,

Marklund, S. L., Andersen, P. M., Finkelstein, D. I., Culvenor, J. G., Masters, C. L., Duce, J., White, A. R., … Crouch, P. J. (2016). Endogenous Cu in the central nervous system fails to satiate the elevated requirement for Cu in a mutant SOD1 mouse model of ALS. *Metallomics*, *8*(9), 1002–1011. https://doi.org/10.1039/C6MT00099A

Karen, P., Stevanec, M., Smerdu, V., Cvetko, E., Kubinova, L., & Erzen, I. (2009). Software for muscle fibre type classification and analysis. *Eur J Histochem*, *53*(2), 87–95.

Kim, Y. J., Brox, T., Feiden, W., & Weickert, J. (2007). Fully automated segmentation and morphometrical analysis of muscle fiber images. *Cytometry A*, *71*(1), 8–15. https://doi.org/10.1002/cyto.a.20334

Klemencic, A., Kovacic, S., & Pernus, F. (1998). Automated segmentation of muscle fiber images using active contour models. *Cytometry*, *32*(4), 317–326. https://doi.org/10.1002/(SICI)1097-0320(19980801)32:4<317::AID-CYTO9>3.0.CO;2-E [pii]

Kwiatkowski, T. J., Bosco, D. A., LeClerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T., Valdmanis, P., Rouleau, G. A., Hosler, B. A., Cortelli, P., de Jong, P. J., Yoshinaga, Y., Haines, J. L., Pericak-Vance, M. A., Yan, J., … Brown, R. H. (2009). Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *Science*, *323*(5918), 1205–1208. https://doi.org/10.1126/science.1166066

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*. http://arxiv.org/abs/1802.03426

Meehan Connor, Ebrahimian Jonathan, Moore Wayne, M. S. (2020). *Uniform Manifold Approximation and Projection (UMAP)* (1.3.4). MATLAB Central File Exchange.

Meunier, B., Picard, B., Astruc, T., & Labas, R. (2010). Development of image analysis tool for the classification of muscle fibre type using immunohistochemical staining. *Histochem Cell Biol*, *134*(3), 307–317. https://doi.org/10.1007/s00418-010-0733-7

Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., Waterston, R. H., & Murray, J. I. (2019). A lineage-resolved molecular atlas of C. Elegans embryogenesis at single-cell resolution. *Science*, *365*(6459). https://doi.org/10.1126/science.aax1971

Pette, D., & Staron, R. S. (2000). Myosin isoforms, muscle fiber types, and transitions. *Microsc Res Tech*, *50*(6), 500–509. https://doi.org/10.1002/1097-0029(20000915)50:6<500::AID-JEMT7>3.0.CO;2-7 [pii] 10.1002/1097-0029(20000915)50:6<500::AID-JEMT7>3.0.CO;2-7

Picher-Martel, V., Valdmanis, P. N., Gould, P. V, Julien, J.-P., & Dupré, N. (2016). From animal models to human disease: a genetic approach for personalized medicine in ALS. *Acta Neuropathologica Communications*, *4*(1), 70. https://doi.org/10.1186/s40478-016-0340-5

Pramatarova, A., Figlewicz, D. A., Krizus, A., Han, F. Y., Ceballos-Picot, I., Nicole, A., Dib, M., Meininger, V., Brown, R. H., & Rouleau, G. A. (1995). Identification of new mutations in the Cu/Zn superoxide dismutase gene of patients with familial

amyotrophic lateral sclerosis. *American Journal of Human Genetics*, *56*(3), 592–596.

Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., … Traynor, B. J. (2011). A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, *72*(2), 257–268. https://doi.org/10.1016/j.neuron.2011.09.010

Rocha, M. C., Pousinha, P. A., Correia, A. M., Sebastião, A. M., & Ribeiro, J. A. (2013). Early Changes of Neuromuscular Transmission in the SOD1(G93A) Mice Model of ALS Start Long before Motor Symptoms Onset. *PLoS ONE*, *8*(9), e73846. https://doi.org/10.1371/journal.pone.0073846

Rooney, J. E. (2006). Severe muscular dystrophy in mice that lack dystrophin and 7 integrin. *Journal of Cell Science*, *119*(11), 2185–2195. https://doi.org/10.1242/jcs.02952

Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J. P., Deng, H.-X., Rahmani, Z., Krizus, A., McKenna-Yasek, D., Cayabyab, A., Gaston, S. M., Berger, R., Tanzi, R. E., Halperin, J. J., Herzfeldt, B., … Brown, R. H. (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*, *362*(6415), 59–62. https://doi.org/10.1038/362059a0

Rosenthal, N., & Brown, S. (2007). The mouse ascending: perspectives for human-disease models. *Nature Cell Biology*, *9*(9), 993–999. https://doi.org/10.1038/ncb437

Rotunno, M. S., & Bosco, D. A. (2013). An emerging role for misfolded wild-type SOD1 in sporadic ALS pathogenesis. *Frontiers in Cellular Neuroscience*, *7*, 253. https://doi.org/10.3389/fncel.2013.00253

Sáez, A., Acha, B., Montero-Sánchez, A., Rivas, E., Escudero, L. M., & Serrano, C. (2013). Neuromuscular disease classification system. *Journal of Biomedical Optics*, *18*(6), 066017. https://doi.org/10.1117/1.JBO.18.6.066017

Sáez, A., Rivas, E., Montero-Sánchez, A., Paradas, C., Acha, B., Pascual, A., Serrano, C., & Escudero, L. M. (2013). Quantifiable diagnosis of muscular dystrophies and neurogenic atrophies through network analysis. *BMC Medicine*, *11*(1), 77. https://doi.org/10.1186/1741-7015-11-77

Sánchez-Gutiérrez, D., Sáez, A., Gómez-Gálvez, P., Paradas, C., & Escudero, L. M. (2017). Rules of tissue packing involving different cell types: human muscle organization. *Scientific Reports*, *7*(1), 40444. https://doi.org/10.1038/srep40444

Sánchez-Gutiérrez, D., Sáez, A., Pascual, A., & Escudero, L. M. (2013a). Topological Progression in Proliferating Epithelia Is Driven by a Unique Variation in Polygon Distribution. *PLoS ONE*, *8*(11), e79227. https://doi.org/10.1371/journal.pone.0079227

Sánchez-Gutiérrez, D., Sáez, A., Pascual, A., & Escudero, L. M. (2013b). Topological Progression in Proliferating Epithelia Is Driven by a Unique Variation in Polygon

Distribution [Journal Article]. *PLoS ONE*, *8*(11), e79227. https://doi.org/10.1371/journal.pone.0079227

Sánchez-Gutiérrez, D., Tozluoglu, M., Barry, J. D., Pascual, A., Mao, Y., & Escudero, L. M. (2016). Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO Journal*, *35*(1), 77–88. https://doi.org/10.15252/embj.201592374

Savadjiev, P., Strijkers, G. J., Bakermans, A. J., Piuze, E., Zucker, S. W., & Siddiqi, K. (2012). Heart wall myofibers are arranged in minimal surfaces to optimize organ function. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9248–9253. https://doi.org/10.1073/pnas.1120785109

Scott, S., Kranz, J. E., Cole, J., Lincecum, J. M., Thompson, K., Kelly, N., Bostrom, A., Theodoss, J., Al-Nakhala, B. M., Vieira, F. G., Ramasubbu, J., & Heywood, J. A. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph Lateral Scler*, *9*(1), 4–15. https://doi.org/789666722 [pii] 10.1080/17482960701856300

Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J. C., Williams, K. L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J. D., Leigh, P. N., Al-Chalabi, A., Miller, C. C., Nicholson, G., & Shaw, C. E. (2008). TDP-43 Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis. *Science*, *319*(5870), 1668–1672. https://doi.org/10.1126/science.1154584

Vajda, A., McLaughlin, R. L., Heverin, M., Thorpe, O., Abrahams, S., Al-Chalabi, A., & Hardiman, O. (2017). Genetic testing in ALS. *Neurology*, *88*(10), 991–999. https://doi.org/10.1212/WNL.0000000000003686

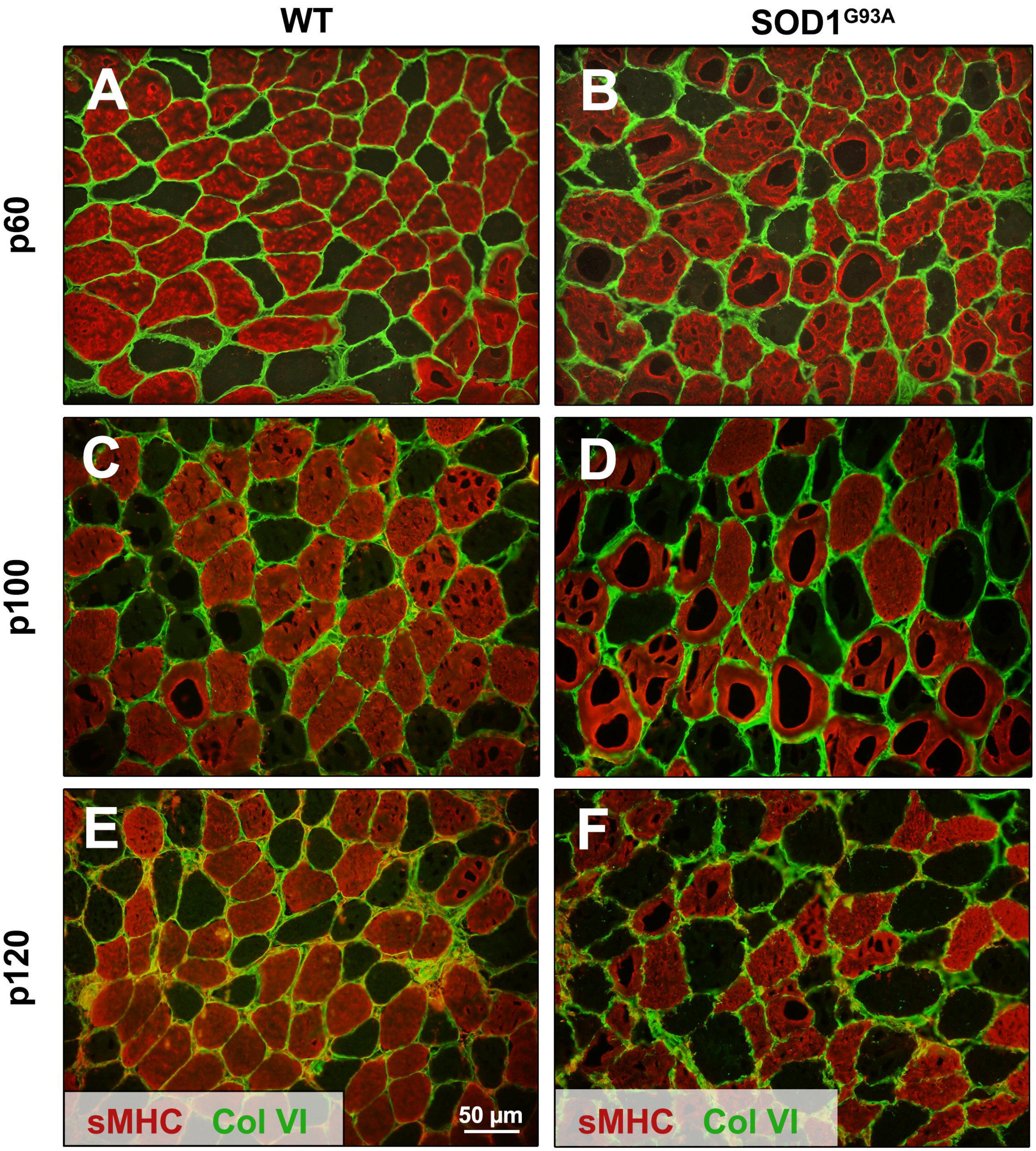
Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. In *Journal of Machine Learning Research* (Vol. 9).

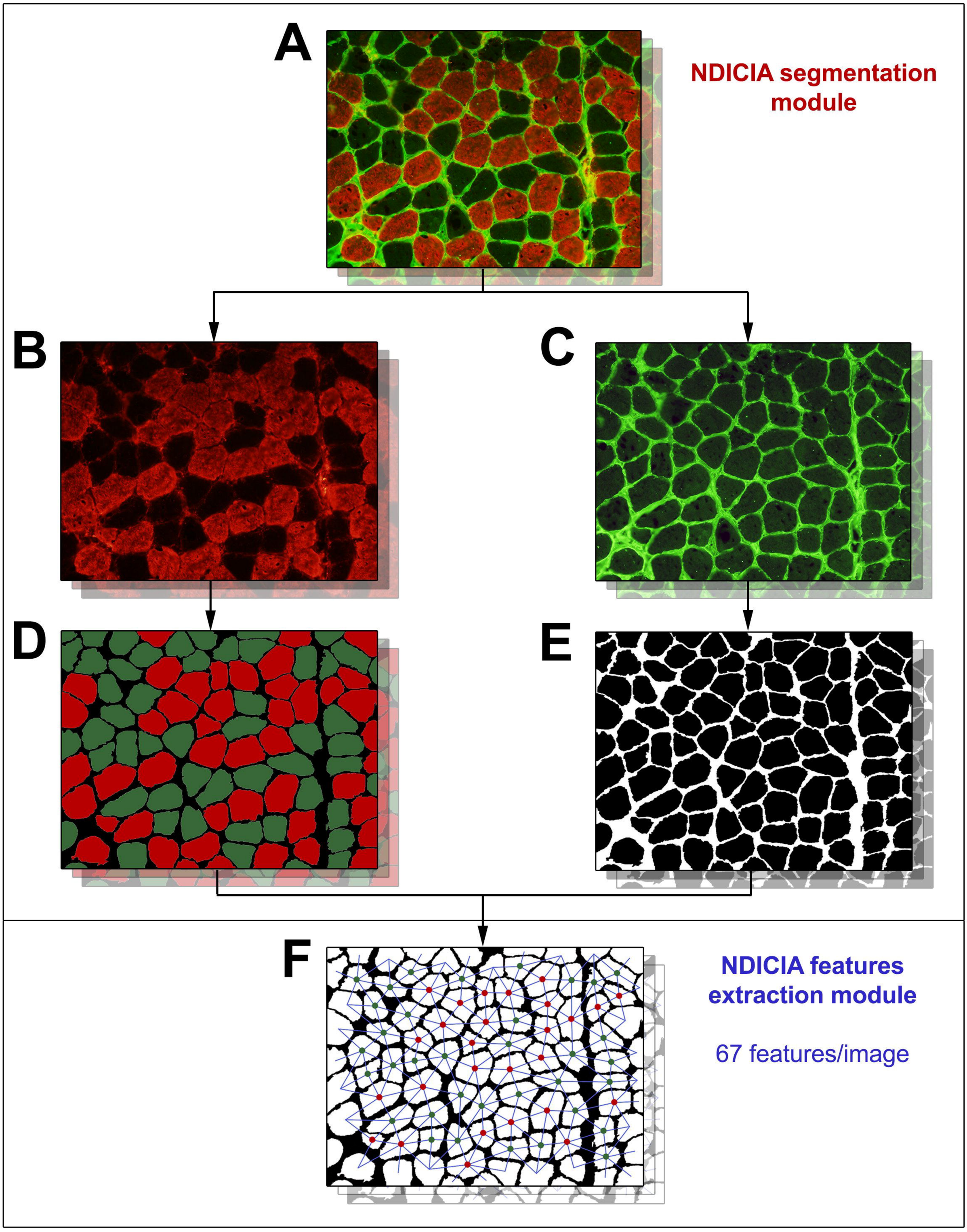
Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal Für Die Reine Und Angewandte Mathematik*, *134*, 198–287.

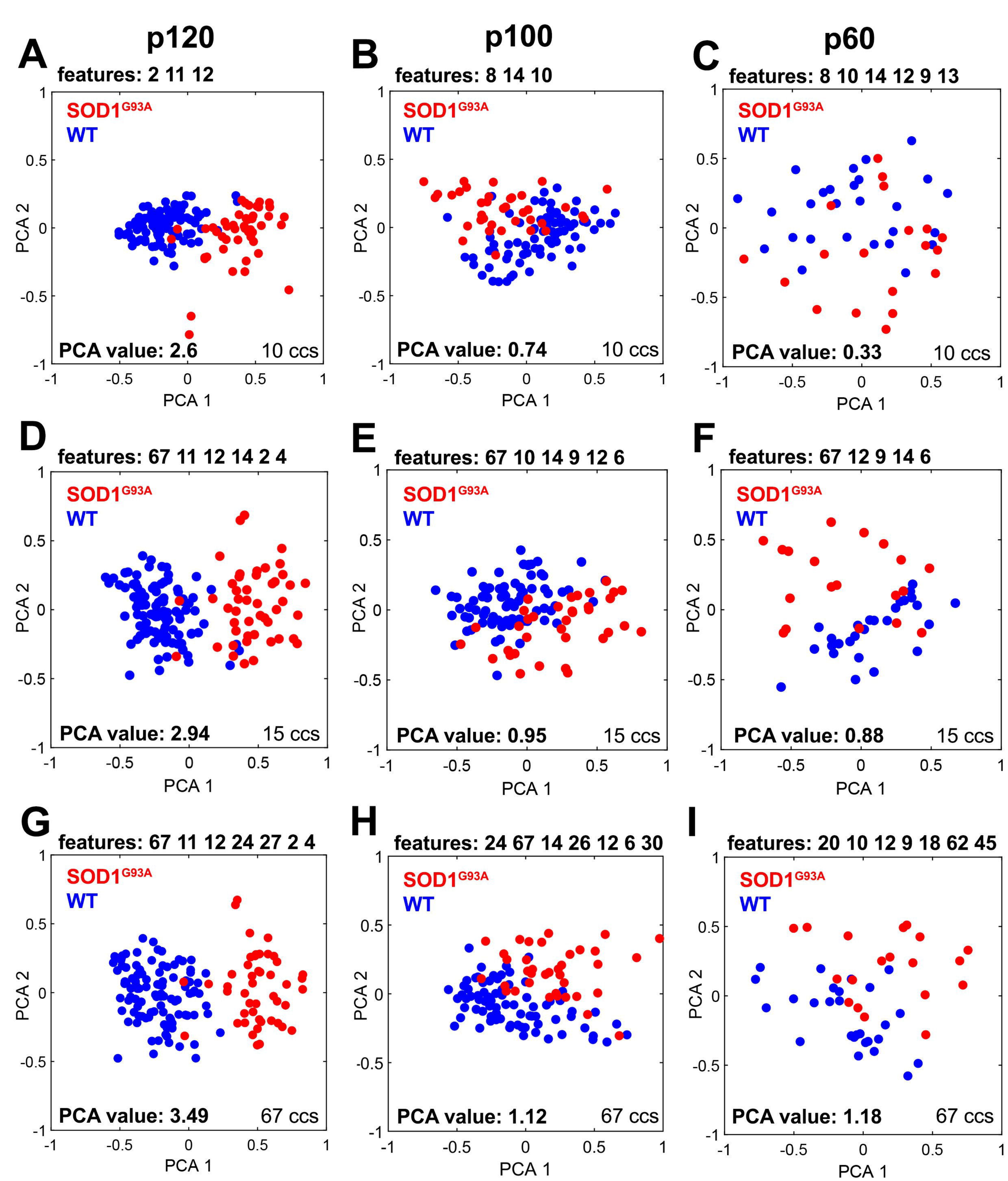
Wegorzewska, I., Bell, S., Cairns, N. J., Miller, T. M., & Baloh, R. H. (2009). TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(44), 18809–18814. https://doi.org/10.1073/pnas.0908767106

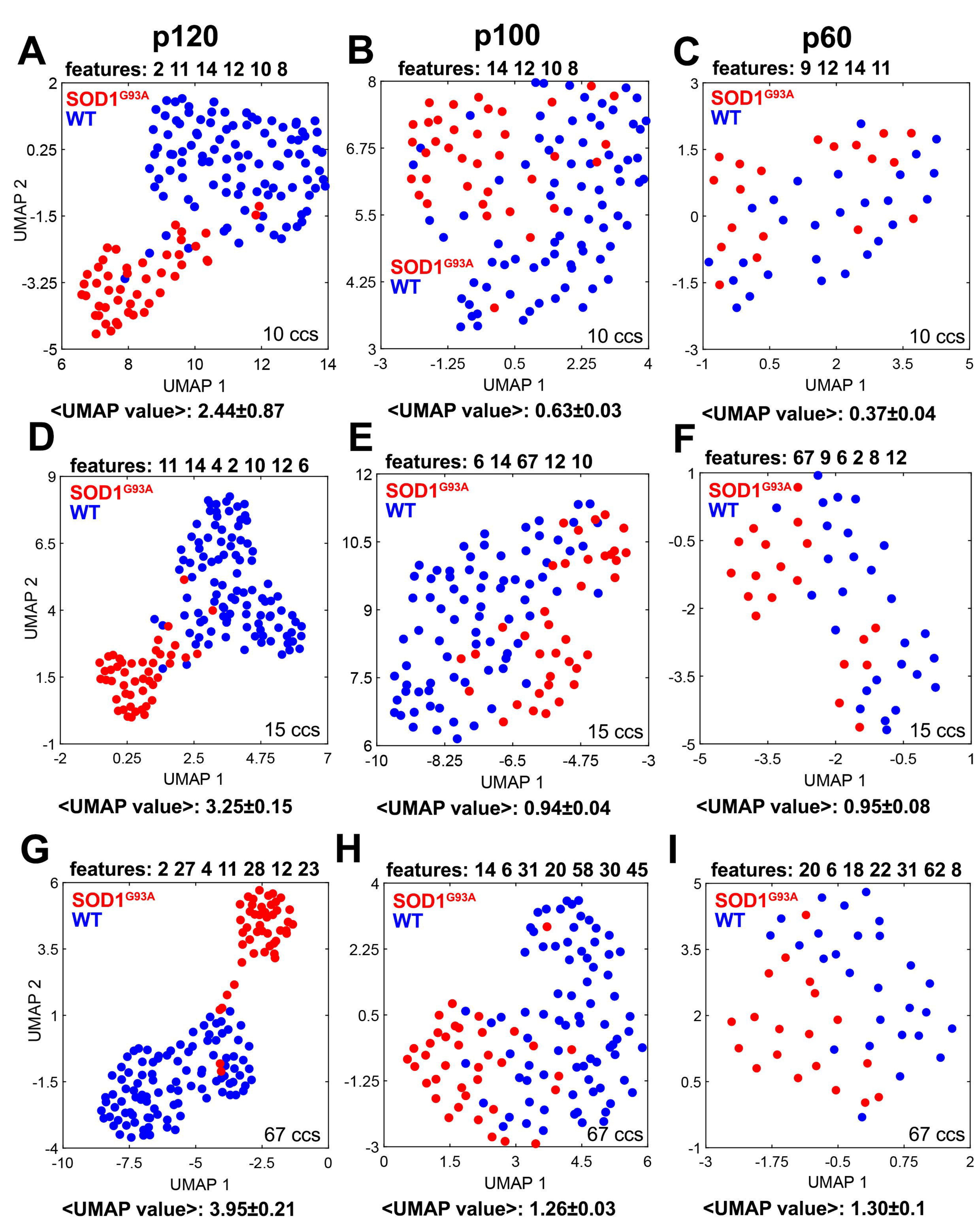
Weydt, P., Hong, S. Y., Kliot, M., & Möller, T. (2003). Assessing disease onset and progression in the SOD1 mouse model of ALS. *NeuroReport*, *14*(7), 1051–1054. https://doi.org/10.1097/01.wnr.0000073685.00308.89

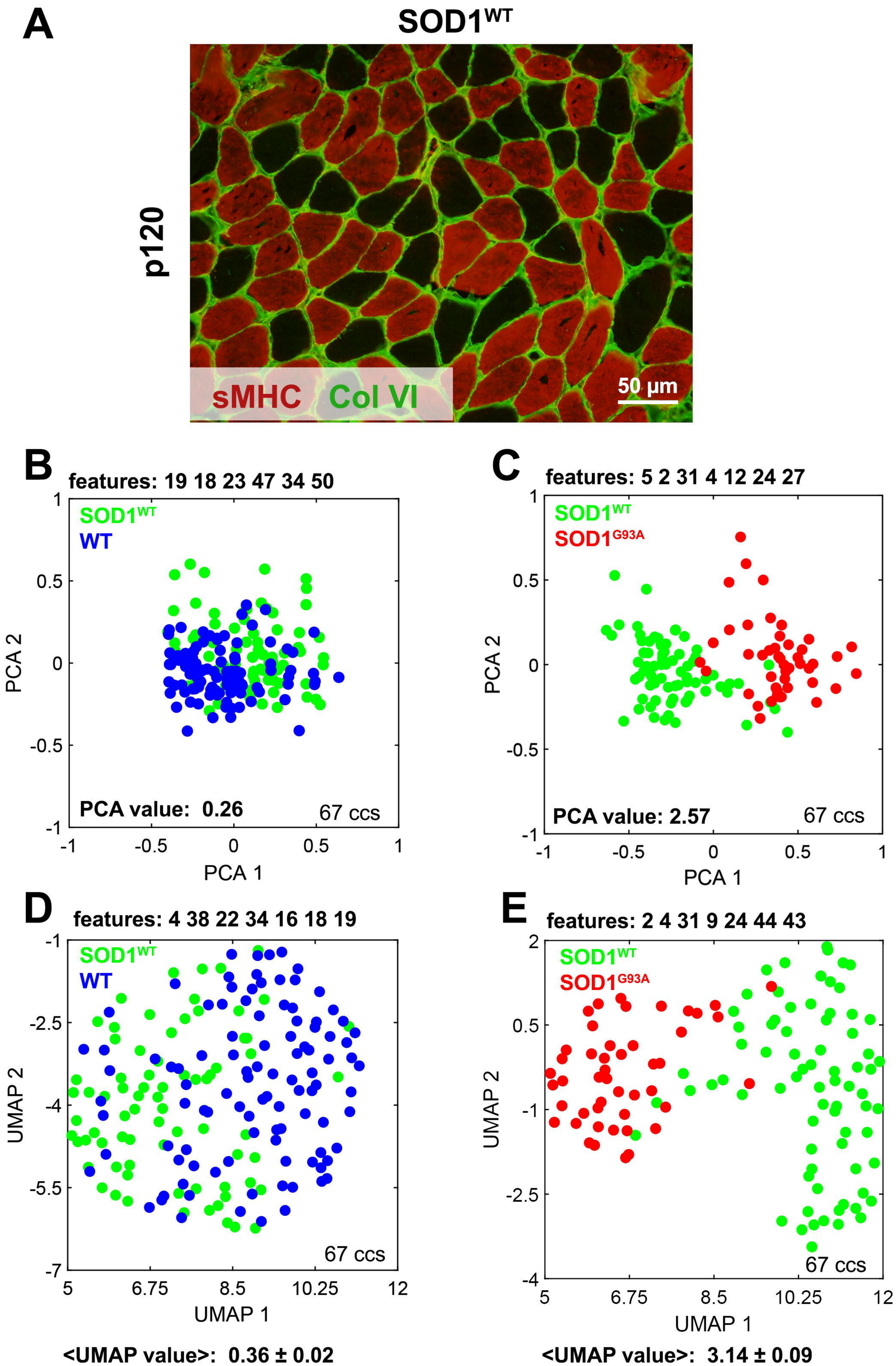
Wong, P. C., Cai, H., Borchelt, D. R., & Price, D. L. (2002). Genetically engineered mouse models of neurodegenerative diseases. *Nature Neuroscience*, *5*(7), 633–639. https://doi.org/10.1038/nn0702-633


WT
SOD1G93A
p60
p100
p120
A
B
C
D
E
F
sMHC Col VI
50 µm
sMHC Col VI


A
NDICIA segmentation module
B
C
D
E
F
NDICIA features extraction module
67 features/image


A
p120
features: 2 11 12
SOD1G93A
WT
PCA value: 2.6
10 ccs
B
p100
features: 8 14 10
PCA value: 0.74
10 ccs
C
p60
features: 8 10 14 12 9 13
PCA value: 0.33
10 ccs
D
features: 67 11 12 14 2 4
PCA value: 2.94
15 ccs
E
features: 67 10 14 9 12 6
PCA value: 0.95
15 ccs
F
features: 67 12 9 14 6
PCA value: 0.88
15 ccs
G
features: 67 11 12 24 27 2 4
PCA value: 3.49
67 ccs
H
features: 24 67 14 26 12 6 30
PCA value: 1.12
67 ccs
I
features: 20 10 12 9 18 62 45
PCA value: 1.18
67 ccs
PCA 1
PCA 2


A
p120
features: 2 11 14 12 10 8
SOD1G93A
WT
10 ccs
UMAP 2
UMAP 1
<UMAP value>: 2.44±0.87
B
p100
features: 14 12 10 8
SOD1G93A
WT
10 ccs
UMAP 1
<UMAP value>: 0.63±0.03
C
p60
features: 9 12 14 11
SOD1G93A
WT
10 ccs
UMAP 1
<UMAP value>: 0.37±0.04
D
features: 11 14 4 2 10 12 6
SOD1G93A
WT
15 ccs
UMAP 2
UMAP 1
<UMAP value>: 3.25±0.15
E
features: 6 14 67 12 10
SOD1G93A
WT
15 ccs
UMAP 1
<UMAP value>: 0.94±0.04
F
features: 67 9 6 2 8 12
SOD1G93A
WT
15 ccs
UMAP 1
<UMAP value>: 0.95±0.08
G
features: 2 27 4 11 28 12 23
SOD1G93A
WT
67 ccs
UMAP 2
UMAP 1
<UMAP value>: 3.95±0.21
H
features: 14 6 31 20 58 30 45
SOD1G93A
WT
67 ccs
UMAP 1
<UMAP value>: 1.26±0.03
I
features: 20 6 18 22 31 62 8
SOD1G93A
WT
67 ccs
UMAP 1
<UMAP value>: 1.30±0.1


A
SOD1WT
p120
sMHC
Col VI
50 µm
B
features: 19 18 23 47 34 50
SOD1WT
WT
PCA value: 0.26
67 ccs
PCA 1
PCA 2
C
features: 5 2 31 4 12 24 27
SOD1WT
SOD1G93A
PCA value: 2.57
67 ccs
PCA 1
PCA 2
D
features: 4 38 22 34 16 18 19
SOD1WT
WT
67 ccs
UMAP 1
UMAP 2
<UMAP value>: 0.36 ± 0.02
E
features: 2 4 31 9 24 44 43
SOD1WT
SOD1G93A
67 ccs
UMAP 1
UMAP 2
<UMAP value>: 3.14 ± 0.09