HHS Public Access

Author manuscript

Ann Rheum Dis. Author manuscript; available in PMC 2023 August 28.

Published in final edited form as:

Ann Rheum Dis. 2020 September ; 79(9): 1234–1242. doi:10.1136/annrheumdis-2019-216599.

Machine learning algorithms reveal unique gene expression profiles in muscle biopsies from patients with different types of myositis

Iago Pinal-Fernandez, M.D., Ph.D.,
National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD; Johns Hopkins University School of Medicine, Baltimore, MD; Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

Maria Casal-Dominguez, M.D., Ph.D.,
National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD; Johns Hopkins University School of Medicine, Baltimore, MD

Assia Derfoul, Ph.D.,
National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

Katherine Pak, M.D.,
National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

Frederick W Miller, M.D., Ph.D.,
National Institute of Environmental Health Sciences, National Institutes of Health, Bethesda, MD

Jose C Milisenda, M.D.,
Clinic Hospital and the University of Barcelona, Barcelona, Spain

Josep M Grau-Junyent, M.D., Ph.D.,
Clinic Hospital and the University of Barcelona, Barcelona, Spain

Albert Selva-O'Callaghan, M.D., Ph.D.,
Vall d'Hebron Hospital and Autonomous University of Barcelona, Spain

Carme Carrion-Ribas, Ph.D.,
Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

Julie J. Paik, M.D., MHS,
Johns Hopkins University School of Medicine, Baltimore, MD

Jemima Albayda, M.D.,
Johns Hopkins University School of Medicine, Baltimore, MD

Address correspondence to: Andrew L. Mammen, M.D., Ph.D., or Iago Pinal-Fernandez, M.D., Ph.D. Muscle Disease Unit, Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, 50 South Drive, Room 1141, Building 50, MSC 8024, Bethesda, MD 20892. andrew.mammen@nih.gov or iago.pinalfernandez@nih.gov. Phone: 301-451-1199. Fax: 301-594-0305.

Competing interests: None

Lisa Christopher-Stine, M.D., M.P.H.,
Johns Hopkins University School of Medicine, Baltimore, MD

Thomas E. Lloyd, M.D., Ph.D.,
Johns Hopkins University School of Medicine, Baltimore, MD

Andrea M Corse, M.D.,
Johns Hopkins University School of Medicine, Baltimore, MD

Andrew L Mammen, M.D., Ph.D.
National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD; Johns Hopkins University School of Medicine, Baltimore, MD

Abstract

Objectives: Myositis is a heterogeneous family of diseases that includes dermatomyositis (DM), antisynthetase syndrome (AS), immune-mediated necrotizing myopathy (IMNM), inclusion body myositis (IBM), polymyositis, and overlap myositis. Additional subtypes of myositis can be defined by the presence of myositis-specific autoantibodies (MSAs). The purpose of this study was to define unique gene expression profiles in muscle biopsies from patients with DM, AS, IMNM, IBM, and the MSA-defined subtypes of myositis.

Methods: RNAseq was performed on muscle biopsies from 119 myositis patients with IBM or defined MSAs and 20 controls. Machine learning algorithms were trained on transcriptomic data and recursive feature elimination was used to determine which genes were most useful for classifying muscle biopsies into each type and MSA-defined subtype of myositis.

Results: The support vector machine learning algorithm classified the muscle biopsies with >90% accuracy. Recursive feature elimination identified genes most useful to the machine learning algorithm and that are only overexpressed in one type of myositis. For example, CAMK1G, EGR4, and CXCL8 are highly expressed in AS but not in DM or other types of myositis. Using the same computational approach, we also identified genes that are uniquely overexpressed in different MSA-defined subtypes. These included APOA4, which is only expressed in anti-HMGCR myopathy, and MADCAM1, which is only expressed in anti-Mi2-positive DM.

Conclusions: Unique gene expression profiles in muscle biopsies from patients with DM, AS, IMNM, IBM and different MSA-defined subtypes of myositis suggest that different pathological mechanisms underly muscle damage in each of these diseases.

Keywords

Myositis; Autoantibodies; Autoantigens; Skeletal Muscle; Interferons

INTRODUCTION

The idiopathic inflammatory myopathies (IIM) are a heterogeneous family of diseases that includes six major types: dermatomyositis (DM), antisynthetase syndrome (AS), immune-mediated necrotizing myopathy (IMNM), inclusion body myositis (IBM), polymyositis, and overlap myositis [1]. Furthermore, 50-80% of IIM patients have myositis-specific autoantibodies (MSAs) that define phenotypically distinct IIM subtypes[2 3].

Muscle biopsies from patients with each major type of myositis have distinctive pathological features. For example, perifascicular myofiber atrophy and/or necrosis is a characteristic feature of both DM and AS, IMNM biopsies have abundant scattered necrotic myofibers, and IBM muscle biopsies usually include myofibers with cytoplasmic vacuoles[4]. However, histologic features that can reliably distinguish between DM and AS have not been identified. Similarly, histologic features cannot reliably be used to distinguish between different MSA-defined subtypes of DM or IMNM. Thus, it remains unclear whether different pathological pathways lead to muscle damage in the different myositis types and MSA-defined subtypes.

The advent of gene chip microarray and next-generation sequencing technologies has facilitated the use of myositis muscle biopsy gene expression profiles to identify pathological pathways. For example, microarray analysis led to the discoveries that type I and type II IFN-inducible genes are upregulated in muscle biopsies from patients with DM[5] and IBM[6 7], respectively. However, disease-specific gene expression profiles have not been fully described in patients with IMNM, AS, or any of the autoantibody-defined subtypes of DM. Furthermore, little attention has been given to genes that are differentially expressed between patients with different types and subtypes of myositis.[8–11] In the current study, we trained machine learning algorithms to classify muscle biopsies using transcriptomic data from normal, IBM, and MSA-positive muscle biopsies; biopsies from the 20-50% of myositis patients who are MSA-negative were not included in this study. We then used recursive feature elimination to identify novel disease-specific gene expression patterns that may be pathologically relevant in DM, AS, IMNM, IBM, and MSA-defined subtypes of myositis.

MATERIALS AND METHODS

Patients, samples, and autoantibody testing

Muscle biopsies obtained from subjects enrolled in IRB-approved longitudinal cohorts from the NIH (IRB number 91-AR-0196), the Johns Hopkins Myositis Center (IRB number NA_00007454), the Clinic Hospital (Barcelona; IRB number HCB/2015/0479), and the Vall d'Hebron Hospital (Barcelona; IRB number PR (AG) 68/2008) were included in the study if the patients fulfilled IBM criteria according to Lloyd,[12] or had one of the following MSAs: anti-NXP2, -Mi2, -TIF1γ, -MDA5, -HMGCR, -SRP, or -Jo1. Autoantibody testing was performed as previously described for anti-HMGCR and by line blot for the others (EUROLINE Myositis Profile 4). Patients were classified as having the antisynthetase syndrome (AS) if they had autoantibodies against Jo-1 and fulfilled Connor's AS criteria, [13] in the DM group if they had autoantibodies recognizing Mi2, NXP2, TIF1γ or MDA5 and in the IMNM group if they tested positive for anti-SRP or anti-HMGCR autoantibodies. Normal muscle biopsies were obtained from the Johns Hopkins Neuromuscular Pathology Laboratory (n=10) and the Skeletal Muscle Biobank of the University of Kentucky (n=10).

Standard protocol approvals and patient consents.

This study was approved by the Institutional Review Boards at participating institutions and written informed consent was obtained from each participant.

Human muscle biopsy processing, human skeletal muscle cell culture, and mouse muscle injury

See Supplementary Methods.

RNA-sequencing

RNA-sequencing (RNA-seq) was performed as previously described.[14] Briefly, RNA was prepared using TRIzol. Libraries were prepared using the NeoPrep™ system according to the TruSeq Stranded mRNA Library Prep protocol (Illumina) and sequenced using the Illumina HiSeq 2500 or 3000. Reads were aligned using the STAR v.2.5 25, the abundance of each gene was quantified using StringTie v.1.3.3.26 and the differential gene expression was performed using DESeq2 v.1.20 (Supplementary Methods). The Benjamini-Hochberg correction was used to adjust for multiple comparisons and a corrected p-value (q-value) of 0.05 or less was considered statistically significant.

Pathway analysis

We used Ingenuity Pathway Analysis v.01-07 and genes with a q-value below 0.05 and an expression ratio greater than 2 in each group compared to the rest of the biopsies were included in the analysis. Immunologic pathways with a z-score over 2 were selected.

RNAseq-based classification

To find the ability of RNAseq data to classify different types of myositis we first tested several classification models. Next, we performed stratified cross-validation to estimate the accuracy of each model. All steps were performed using Python v.3.6.3. Numpy v.1.13.3 and Pandas v.0.20.3 were used for data wrangling and basic statistical calculations, respectively (Supplementary Methods).

Those genes with significantly differential expression levels in one group compared to the rest of the biopsies were included in each model. The sample was split into a training set containing 2/3 of the observations and a test set containing the remaining 1/3. The training set was used to build the classificatory models and the testing set to evaluate the accuracy of the model. The machine learning models were developed using the package Scikit-learn v.0.19.1. Models were built using 2/3 random resamples of the data and tested in the remaining 1/3. The accuracy of classifying each of the myositis subsets was determined based on the mean and 95% CI of one thousand resampling cycles (Supplementary Methods).

Recursive feature elimination was applied to the whole dataset to rank each gene according to how useful it was for the model to differentiate the different patient groups. The RFE technique was applied through its implementation in Scikit-learn v.0.19.1 (Supplementary Methods).

Statement of patient and public involvement

Neither patients nor the public were involved in the design, conduct, reporting, or dissemination of this research.

Data availability statement

Deidentified RNAseq data will be made available upon request to Dr. Andrew Mammen (andrew.mammen@nih.gov)

RESULTS

Machine learning models accurately classify muscle biopsies

Muscle biopsy specimens were available from 119 myositis patients including 39 with DM (11 anti-Mi2-, 12 anti-NXP2-, 11 anti-TIF1γ-, and 5 anti-MDA5-positive), 49 with IMNM (9 anti-SRP- and 40 anti-HMGCR-positive), 18 with anti-Jo1-positive AS, and 13 with IBM. Twenty normal muscle biopsy specimens were utilized as comparators. Expression levels of all genes were determined for each sample by RNAseq. Details regarding the patients and their muscle biopsy features are found in Supplementary Table 1. Expression levels of genes associated with immune cells, regenerating myofibers, and mature skeletal muscle are found in Supplementary Figure 1.

First, we identified those genes with statistically significant differential expression in controls and each major type of myositis compared to the rest of the groups. A total of 10,141 differentially expressed genes were identified and the top 10 for each group are listed in Table 1. For example, the interferon-inducible gene ISG15 is the top differentially expressed gene in both DM and normal muscle biopsies; it is expressed at 43-fold higher levels in DM compared to the rest of the groups and at 17-fold lower levels in normal biopsies compared to the rest of the groups.

To determine whether machine learning programs could use transcriptomic data to accurately classify patients into each major type of myositis or the control group, all differentially expressed genes were included in each of 10 machine learning models (Supplementary Methods). From among the models tested, the linear support vector machine (SVM) model performed the best with accuracies of 91% or greater to identify normal DM, AS, IMNM and IBM muscle biopsies. (Table 2).

Identifying genes with unique expression patterns in DM, AS, IMNM, and IBM

We expected that for each major type of myositis, those genes contributing most to the accuracy of the machine learning classification model would be involved in disease-specific pathological processes. To identify which among the thousands of differentially expressed genes used by the linear SVM model are most useful to classify a biopsy into each type of myositis, we used the recursive feature elimination technique.[15] This method systematically eliminates genes with the weakest role in the model, leaving those that are most important to classify muscle biopsies into the correct group. Table 3 lists the 10 genes whose expression levels have the greatest utility to identify samples as belonging to each type of myositis or control group. Figure 1 shows the expression levels of the 3 most important genes from each group.

We first sought to validate this approach by determining whether it would identify key genes already known to play roles in DM pathogenesis. As genes upregulated by type I IFN are

known to be expressed at high levels in DM muscle[5 16], we expected that expression levels of type I IFN-inducible genes should be important for the linear SVM model. Indeed, high expression levels of type 1 IFN-inducible genes MX1 and ISG15 were among the 3 most important features used to identify DM muscle biopsies (Table 3).

When applied to the AS group, recursive feature elimination identified CAMK1G (calcium/calmodulin-dependent protein kinase IG), EGR4 (early growth response protein 4), and CXCL8 (interleukin 8) as the 3 most important genes (Table 3). Each of these genes is expressed at markedly higher levels in AS than in the other groups (Figure 1).

High expression levels of MYH4 (myosin heavy chain 4) and JCHAIN (the joining chain of multimeric IgA and IgM) were among the 3 most important features used by the linear SVM model to identify samples as belonging to the IBM group (Table 3 and Figure 1). In addition, the low expression level of H19 (a noncoding RNA) in IBM compared to DM, AS, and IMNM (Figure 1) appeared to be important for IBM classification.

Expression levels of STAT1 (signal inducer and activator of transcription 1), MYH8 (myosin heavy chain 8), and PSMB9 (proteasome subunit beta 9) were the top features used to classify a muscle biopsy as IMNM (Table 3). Based on the patterns of expression (Figure 1), the model seems to rely both on the low expression of IFN-inducible genes STAT1 and PSMB9 (expressed at high levels in DM, AS, and IBM) as well as the high expression of MYH8 (expressed at low levels in normal muscle) to classify biopsies as IMNM.

The expression levels of ACTC1 (actin alpha cardiac muscle 1), LOC151121 (a non-coding gene), and SAA1 (serum amyloid A1) were the top features used to classify normal muscle biopsies (Table 3). Interestingly, normal muscle biopsies were characterized by low levels of ACTC1, which encodes a structural protein expressed during muscle regeneration[17] (Figure 1). Similarly, the SAA1 gene, which encodes the acute phase reactant serum amyloid A1, was expressed at low levels in normal muscles and high levels in all of the myositis groups. In contrast, LOC151121 was expressed at high levels in normal muscle but at low levels in all the myositis groups (Figure 1).

Identifying genes with unique expression patterns in the different subtypes of IMNM and DM

Using the same methodology, we next identified those genes that were most useful to classify biopsies according to the different autoantibody-defined subtypes within IMNM and DM. This revealed that APOA4 (apolipoprotein A4) was selectively expressed in IMNM patients with anti-HMGCR autoantibodies (Figure 2). Similarly, MADCAM1 (mucosal vascular addressin cell adhesion molecule 1) was exclusively detectable in DM patients with anti-Mi2 autoantibodies (Figure 2).

Pathway analysis

To gain further insight into the biological processes that distinguish each group compared to the others, we performed pathway analyses. For each analysis, we included the set of genes differentially expressed by at least two-fold in the type of myositis (or control) compared to the rest of the biopsies. Pathways annotated as related to the “cellular immune response”,

"cytokine signaling", and "humoral immune response" (i.e., immunologic pathways) were included in each analysis.

As expected, "interferon signaling" was the top over-represented immunologic pathway in DM (Figure 3). The AS and IBM biopsies shared the same top 3 over-represented pathways that were not included DM, IMNM, or control biopsies. These included the T cell pathways "ICOS-ICOSL signaling in T helper cells", "CD28 signaling in T helper cells", and the "Th1 pathway". No immunologic pathways were over-represented in IMNM biopsies. Rather, IMNM biopsies, like control biopsies, were notable for the under-representation of pathways that were important in DM, AS, and/or IBM.

Muscle regeneration genes are among the top differentially expressed genes in IMNM and are also overexpressed in other types of myositis

To classify biopsies as IMNM, linear SVM relied on the relative under expression of genes expressed at high levels in other forms of myositis (e.g., STAT1 and PSMB8)[16] rather than on genes that were uniquely overexpressed in IMNM. To further investigate pathological processes important for IMNM, we considered the known functions of the top 10 overexpressed genes in biopsies from these patients (Table 4). Interestingly, several of these are known to play a role in skeletal muscle differentiation and/or muscle repair. For example, ACTC1 encodes alpha-actin which is expressed in early adult skeletal muscle.[17] Similarly, TNC encodes an extracellular matrix protein that is expressed only in actively remodeling musculoskeletal tissue.[18]

To determine whether the other most overexpressed genes in IMNM play a role in muscle regeneration, we analyzed their expression levels in cultured human myoblasts as they differentiated to form myotubes. Each gene was expressed at low levels in myoblasts and at high levels in differentiating myotubes (Supplementary Fig 2). Similarly, these genes were expressed at low levels in healthy mouse muscle, but at high levels in regenerating mouse muscles following a muscle injury (Supplementary Fig 3). This pattern suggests that these genes are expressed as part of the muscle regeneration process induced by necrosis in IMNM muscle. Since regeneration is also a common feature of muscle biopsies from those with DM, AS, and IBM, we expected that muscle biopsies from each of these types of myositis should also have high levels of the genes overexpressed in IMNM. Indeed, even though they were not among the top 10 overexpressed genes in the other groups, each of these genes was highly expressed in the other types of myositis muscle but not control muscle (Supplementary Fig 4).

We next considered the known functions of the top 10 upregulated genes in DM, AS, and IBM compared to control muscle (Table 4). Consistent with prior studies, many of the top 10 differentially expressed genes in muscle biopsies from DM patients are inducible by interferon type I (e.g., ISG15[19 20], IFI6[21], and MX1[22]) (Table 4). Similarly, several of the most overexpressed genes in AS and IBM muscle biopsies are interferon type II inducible genes (e.g., PSMB8[23], GBP2, and GBP1[24 25]) (Table 4).

DISCUSSION

In this study, we showed that machine learning algorithms trained on transcriptomics data could accurately classify myositis muscle biopsies from DM, AS, IMNM, and IBM patients. This demonstrates that these IIM types have unique gene expression profiles. Indeed, by applying recursive feature elimination to the machine learning algorithms we identified novel gene markers (e.g., CAMK1G, EGR, and CXCL8) that are uniquely expressed in AS but not DM, even though these two diseases can be histologically indistinguishable. Moreover, we also identified genes (e.g., ACTC1 and SSA1) that are overexpressed in all types of myositis studied here but not in normal muscle. Finally, we confirmed previous observations related to the pathogenesis of myositis, including the role of interferon pathways in DM,[8 16] the prominence of muscle regeneration in IMNM,[26] and the presence of plasma cells in IBM (as evidenced by overexpression of JCHAIN, a plasma cell marker).[27 28]

We applied the same computational approach to identify genes that are uniquely upregulated in patients with different MSA-defined IIM subtypes. For example, although anti-SRP and anti-HMGCR myopathy muscle biopsies are histologically identical, we identified the APOA4 gene as being exclusively upregulated in the latter subtype of IMNM. Since statin exposure is a risk factor for developing anti-HMGCR myopathy but not other types of myositis[29], it is of interest that APOA4, which contributes to reverse cholesterol transport by facilitating the movement of cholesterol from the periphery to the liver for excretion[30], is only upregulated in anti-HMGCR myopathy muscle biopsies.

We also found that different MSA-defined DM subtypes had different gene expression profiles. For example, MADCAM1 was uniquely expressed in muscle biopsies from DM patients with anti-Mi2 autoantibodies. Of note, MADCAM1 is expressed on endothelial surfaces in the intestine where it mediates the migration of lymphocytes into the gut by binding to $\alpha_4\beta_7$ integrin found on the surface of CD4+ and CD8+ T-cells[31]. Since MADCAM1 recruits inflammatory cells to the gut in patients with colitis, we hypothesize that it could play a similar role in anti-Mi2-positive DM patients, who have more lymphocytic invasion of muscle fibers than DM patients with other autoantibodies[32]. This could have therapeutic implications since drugs that target the MADCAM1/$\alpha_4\beta_7$ pathway have already been developed.

This study was not designed to directly compare the performance of machine learning algorithms utilizing muscle biopsy transcriptomic data with the analysis of histologic features to diagnose different types of myositis. Still, the current study suggests that machine learning algorithms would fare favorably in such a comparison. For example, only 72% of biopsies from the included DM patients had perifascicular atrophy[32], the key feature required for histologic diagnosis of DM[33]. Nonetheless, the SVM algorithm diagnosed DM based on the muscle biopsy transcriptome with an accuracy of 92%. This raises the possibility that, with the availability of gene expression profile data collected from a large number of patients with different types of myopathy, machine learning algorithms could be diagnostically useful.

This study was limited in that we did not include muscle biopsies from all types of myositis. Indeed, we excluded biopsies from patients with polymyositis, overlap myositis, and MSA-negative forms of myositis. Furthermore, our analysis was restricted to gene expression data and did not include analyses of the corresponding proteins. Nonetheless, by applying machine learning algorithms to muscle biopsy transcriptomic data, we have demonstrated that DM, AS, IMNM, and IBM can be distinguished based on their unique gene expression patterns. Furthermore, by applying recursive feature elimination to these classification models, we not only confirmed known pathological pathways in IIM, such as the role of type I interferon in DM, we also identified novel genes that are uniquely upregulated in other types and MSA-defined subtypes of myositis. We expect this computational approach could be useful for analyzing transcriptomic data from other autoimmune conditions in which there are different types and subtypes of the disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

The authors thank Dr. Gustavo Gutierrez-Cruz, Dr. Stefania Dell'Orso and Faiza Naz from the NIAMS sequencing facility for all their technical collaboration in making the RNAseq libraries and sequencing them, and the University of Kentucky Center for Muscle Biology for providing normal human muscle samples for the study.

Funding:

This research was supported in part by the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases and the National Institute of Environmental Health Sciences of the National Institutes of Health. The Myositis Research Database and Dr. LC-S are supported by the Huayi and Siuling Zhang Discovery Fund. IPFś research was supported by a Fellowship from the Myositis Association. The authors also thank Dr. Peter Buck for support.

REFERENCES

1. Mariampillai K, Granger B, Amelin D, et al. Development of a New Classification System for Idiopathic Inflammatory Myopathies Based on Clinical Manifestations and Myositis-Specific Autoantibodies. JAMA Neurol 2018;75:1528–37. [PubMed: 30208379]
2. Love LA, Leff RL, Fraser DD, et al. A new approach to the classification of idiopathic inflammatory myopathy: myositis-specific autoantibodies define useful homogeneous patient groups. Medicine; analytical reviews of general medicine, neurology, psychiatry, dermatology, and pediatries 1991;70:360–74.
3. Betteridge Z, McHugh N. Myositis-specific autoantibodies: an important tool to support diagnosis of myositis. Journal of internal medicine 2016;280:8–23. [PubMed: 26602539]
4. Selva-O'Callaghan A, Pinal-Fernandez I, Trallero-Araguas E, et al. Classification and management of adult inflammatory myopathies. Lancet Neurol 2018;17:816–28. [PubMed: 30129477]
5. Greenberg SA, Pinkus JL, Pinkus GS, et al. Interferon-alpha/beta-mediated innate immune mechanisms in dermatomyositis. Annals of Neurology 2005;57:664–78. [PubMed: 15852401]
6. Ivanidze J, Hoffmann R, Lochmuller H, et al. Inclusion body myositis: laser microdissection reveals differential up-regulation of IFN-gamma signaling cascade in attacked versus nonattacked myofibers. Am J Pathol 2011;179:1347–59. [PubMed: 21855683]
7. Allenbach Y, Chaara W, Rosenzwajg M, et al. Th1 response and systemic treg deficiency in inclusion body myositis. PloS one 2014;9:e88788. [PubMed: 24594700]
8. Greenberg SA, Pinkus JL, Pinkus GS, et al. Interferon-alpha/beta-mediated innate immune mechanisms in dermatomyositis. Ann Neurol 2005;57:664–78. [PubMed: 15852401]

9. Hamann PD, Roux BT, Heward JA, et al. Transcriptional profiling identifies differential expression of long non-coding RNAs in Jo-1 associated and inclusion body myositis. Sci Rep 2017;7:8024. [PubMed: 28808260]
10. Raju R, Dalakas MC. Gene expression profile in the muscles of patients with inflammatory myopathies: effect of therapy with IVIg and biological validation of clinically relevant genes. Brain 2005;128:1887–96. [PubMed: 15857930]
11. Greenberg SA, Sanoudou D, Haslett JN, et al. Molecular profiles of inflammatory myopathies. Neurology 2002;59:1170–82. [PubMed: 12391344]
12. Lloyd TE, Mammen AL, Amato AA, et al. Evaluation and construction of diagnostic criteria for inclusion body myositis. Neurology 2014;83:426–33. [PubMed: 24975859]
13. Connors GR, Christopher-Stine L, Oddis CV, et al. Interstitial lung disease associated with the idiopathic inflammatory myopathies: what progress has been made in the past 35 years? Chest 2010;138:1464–74. [PubMed: 21138882]
14. Amici DR, Pinal-Fernandez I, Mazala DA, et al. Calcium dysregulation, functional calpainopathy, and endoplasmic reticulum stress in sporadic inclusion body myositis. Acta Neuropathol Commun 2017;5:24. [PubMed: 28330496]
15. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. Machine Learning 2002;46:389–422.
16. Pinal-Fernandez I, Casal-Dominguez M, Derfoul A, et al. Identification of distinctive interferon gene signatures in different types of myositis. Neurology 2019;93:e1193–e204. [PubMed: 31434690]
17. Gunning P, Ponte P, Kedes L, et al. Chromosomal location of the co-expressed human skeletal and cardiac actin genes. Proc Natl Acad Sci U S A 1984;81:1813–7. [PubMed: 6584914]
18. Fluck M, Mund SI, Schittny JC, et al. Mechano-regulated tenascin-C orchestrates muscle repair. Proc Natl Acad Sci U S A 2008;105:13662–7. [PubMed: 18757758]
19. Zhang D, Zhang DE. Interferon-stimulated gene 15 and the protein ISGylation system. J Interferon Cytokine Res 2011;31:119–30. [PubMed: 21190487]
20. D'Cunha J, Ramanujam S, Wagner RJ, et al. In vitro and in vivo secretion of human ISG15, an IFN-induced immunomodulatory cytokine. J Immunol 1996;157:4100–8. [PubMed: 8892645]
21. Kelly JM, Porter AC, Chernajovsky Y, et al. Characterization of a human gene inducible by alpha- and beta-interferons and its expression in mouse cells. EMBO J 1986;5:1601–6. [PubMed: 3017706]
22. Holzinger D, Jorns C, Stertz S, et al. Induction of MxA gene expression by influenza A virus requires type I or type III interferon signaling. J Virol 2007;81:7776–85. [PubMed: 17494065]
23. Hisamatsu H, Shimbara N, Saito Y, et al. Newly identified pair of proteasomal subunits regulated reciprocally by interferon gamma. J Exp Med 1996;183:1807–16. [PubMed: 8666937]
24. Hall JC, Casciola-Rosen L, Berger AE, et al. Precise probes of type II interferon activity define the origin of interferon signatures in target tissues in rheumatic diseases. Proc Natl Acad Sci U S A 2012;109:17609–14. [PubMed: 23045702]
25. Hall JC, Baer AN, Shah AA, et al. Molecular Subsetting of Interferon Pathways in Sjogren's Syndrome. Arthritis Rheumatol 2015;67:2437–46. [PubMed: 25988820]
26. Watanabe Y, Uruha A, Suzuki S, et al. Clinical features and prognosis in anti-SRP and anti-HMGCR necrotising myopathy. J Neurol Neurosurg Psychiatry 2016;87:1038–44. [PubMed: 27147697]
27. Greenberg SA, Bradshaw EM, Pinkus JL, et al. Plasma cells in muscle in inclusion body myositis and polymyositis. Neurology 2005;65:1782–87. [PubMed: 16344523]
28. Lim JH, Cho SJ, Park SK, et al. Stage-specific expression of two neighboring Crlz1 and IgJ genes during B cell development is regulated by their chromatin accessibility and histone acetylation. J Immunol 2006;177:5420–9. [PubMed: 17015728]
29. Mammen AL. Statin-Associated Autoimmune Myopathy. N Engl J Med 2016;374:664–9. [PubMed: 26886523]
30. Qu J, Ko CW, Tso P, et al. Apolipoprotein A-IV: A Multifunctional Protein Involved in Protection against Atherosclerosis and Diabetes. Cells 2019;8.

31. Berlin C, Berg EL, Briskin MJ, et al. Alpha 4 beta 7 integrin mediates lymphocyte binding to the mucosal vascular addressin MAdCAM-1. Cell 1993;74:185–95. [PubMed: 7687523]

32. Pinal-Fernandez I, Casciola-Rosen LA, Christopher-Stine L, et al. The Prevalence of Individual Histopathologic Features Varies according to Autoantibody Status in Muscle Biopsies from Patients with Dermatomyositis. J Rheumatol 2015;42:1448–54. [PubMed: 26443871]

33. Hoogendijk JE, Amato AA, Lecky BR, et al. 119th ENMC international workshop: trial design in adult idiopathic inflammatory myopathies, with the exception of inclusion body myositis, 10-12 October 2003, Naarden, The Netherlands. Neuromuscular disorders : NMD 2004;14:337–45. [PubMed: 15099594]

KEY MESSAGES

What is already known about this subject?

- Different types of myositis are likely to have unique pathological mechanisms.

What does this study add?

- Machine learning algorithms can be trained on transcriptomic data to classify muscle biopsies from patients with DM, AS, IMNM, and IBM.
- Recursive feature elimination can be used to determine which genes are most important for the machine learning algorithms to classify the muscle biopsies.
- Only antisynthetase syndrome muscle biopsies express high levels of CAMKG, EGR4, and CXCL8 (interleukin 8).
- APOA4, a gene involved in cholesterol metabolism, is uniquely over-expressed in anti-HMGCR myopathy, which can be triggered by statins.
- MADCAM1, which recruits lymphocytes to target tissues, is uniquely over-expressed in muscle biopsies from those with anti-Mi2-positive dermatomyositis.

How might this impact on clinical practice?

Gene expression profiling of muscle biopsies from individual myositis patients may identify specific pathologic pathways that could be used to tailor therapies.

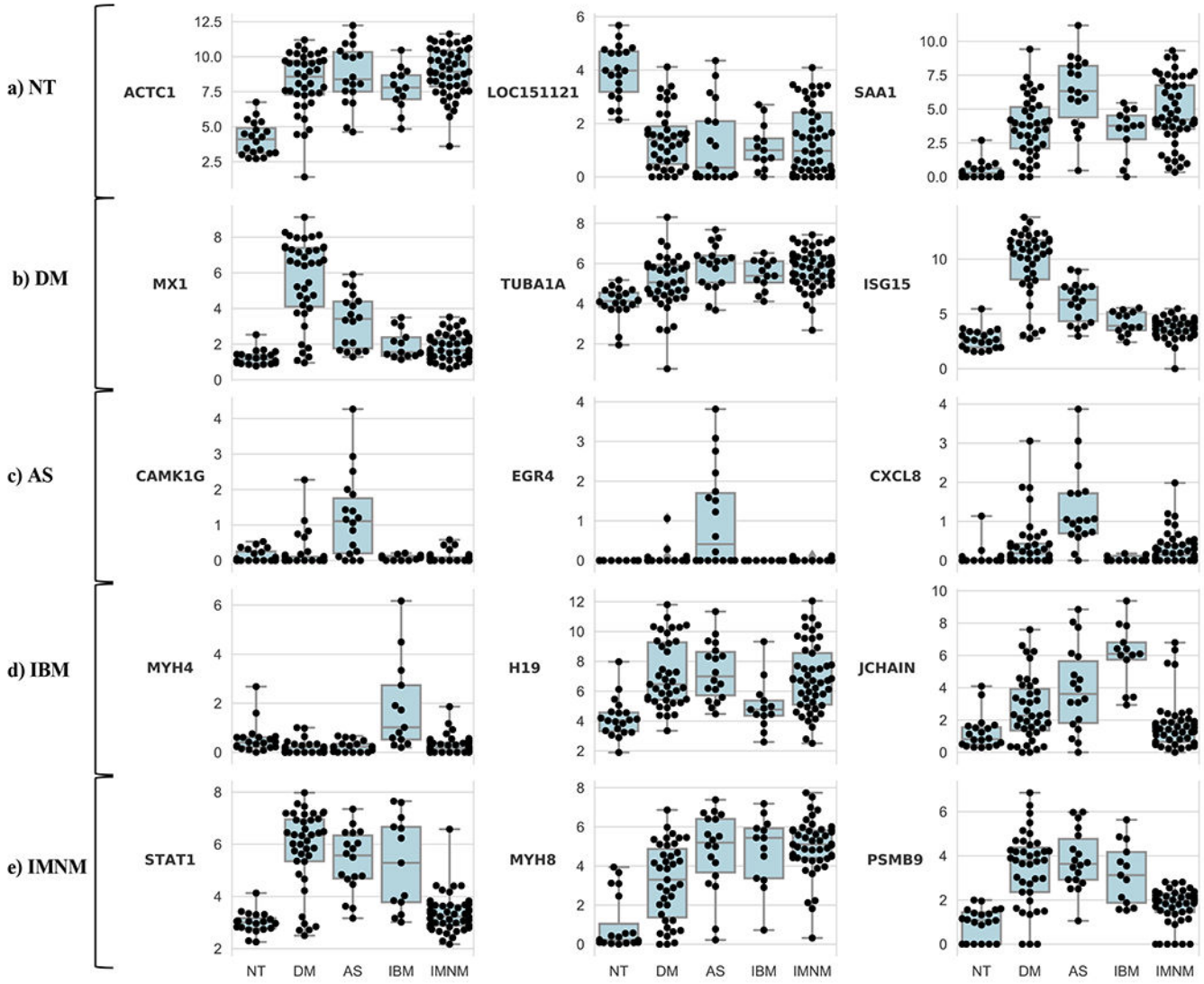

Figure 1. Expression levels of those genes most helpful to classify muscle biopsies into each type of myositis.

The expression levels of the top 3 genes used by the support vector machine model to classify muscle biopsies from normal tissue (NT), dermatomyositis (DM), immune-mediated necrotizing myositis (IMNM), antisynthetase syndrome (AS) or inclusion body myositis (IBM).

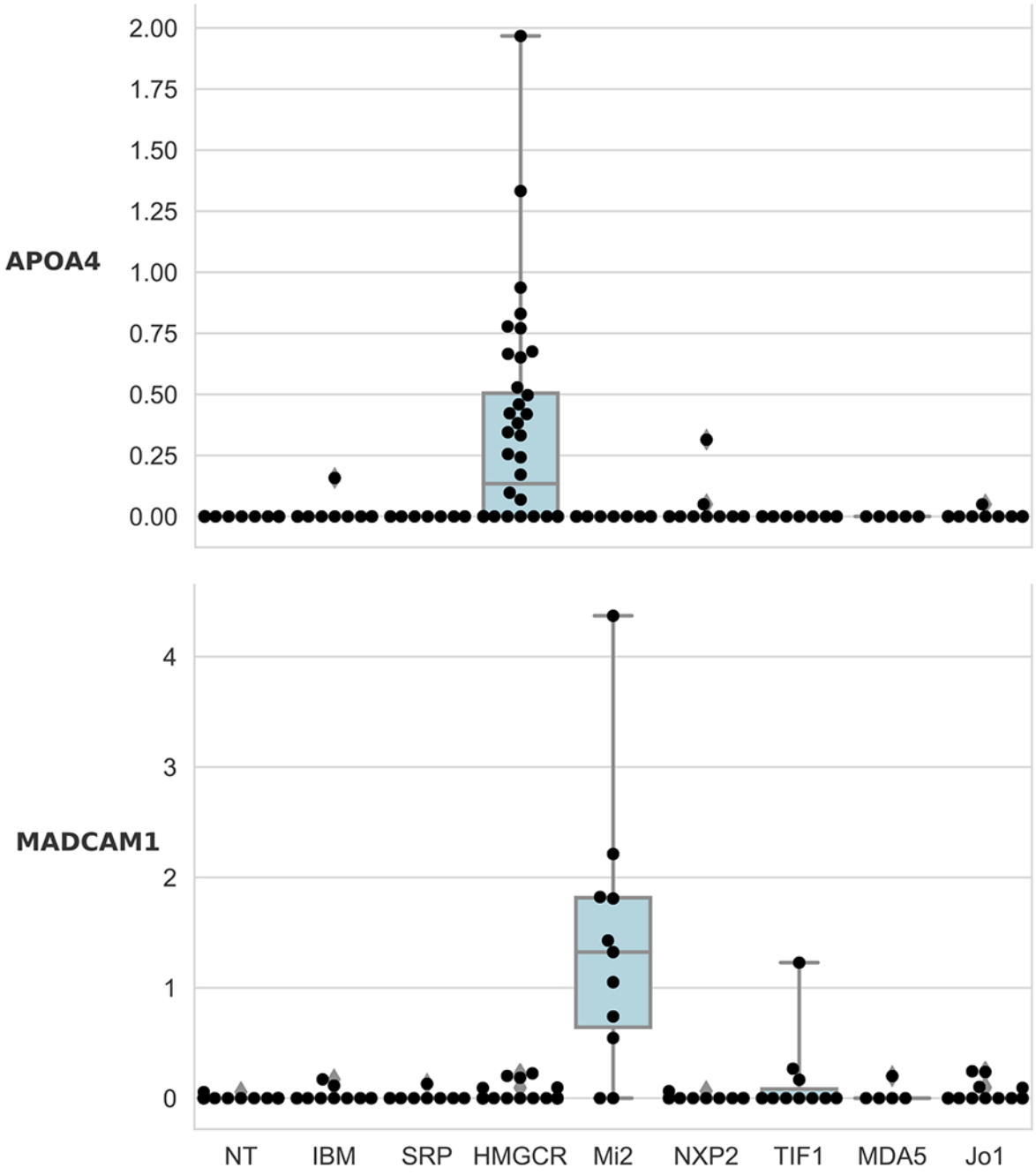

Figure 2. Genes selectively upregulated in different autoantibody-defined subtypes of myositis.

APOA4 and MADCAM1 are selectively overexpressed (log2[FPKM + 1]) in anti-HMGCR IMNM (q-value compared to SRP: 0.0009) and anti-Mi2 DM (q-value compared to other DM antibodies: 2.9E-9), respectively.

Normal tissue: NT; inclusion body myositis: IBM; anti-SRP IMNM: SRP; anti-HMGCR IMNM: HMGCR; anti-Mi2 DM: Mi2; anti-NXP2 DM: NXP2; anti-TIF1γ DM: TIF1; anti-MDA5 DM: MDA5; anti-Jo1 AS: Jo1.

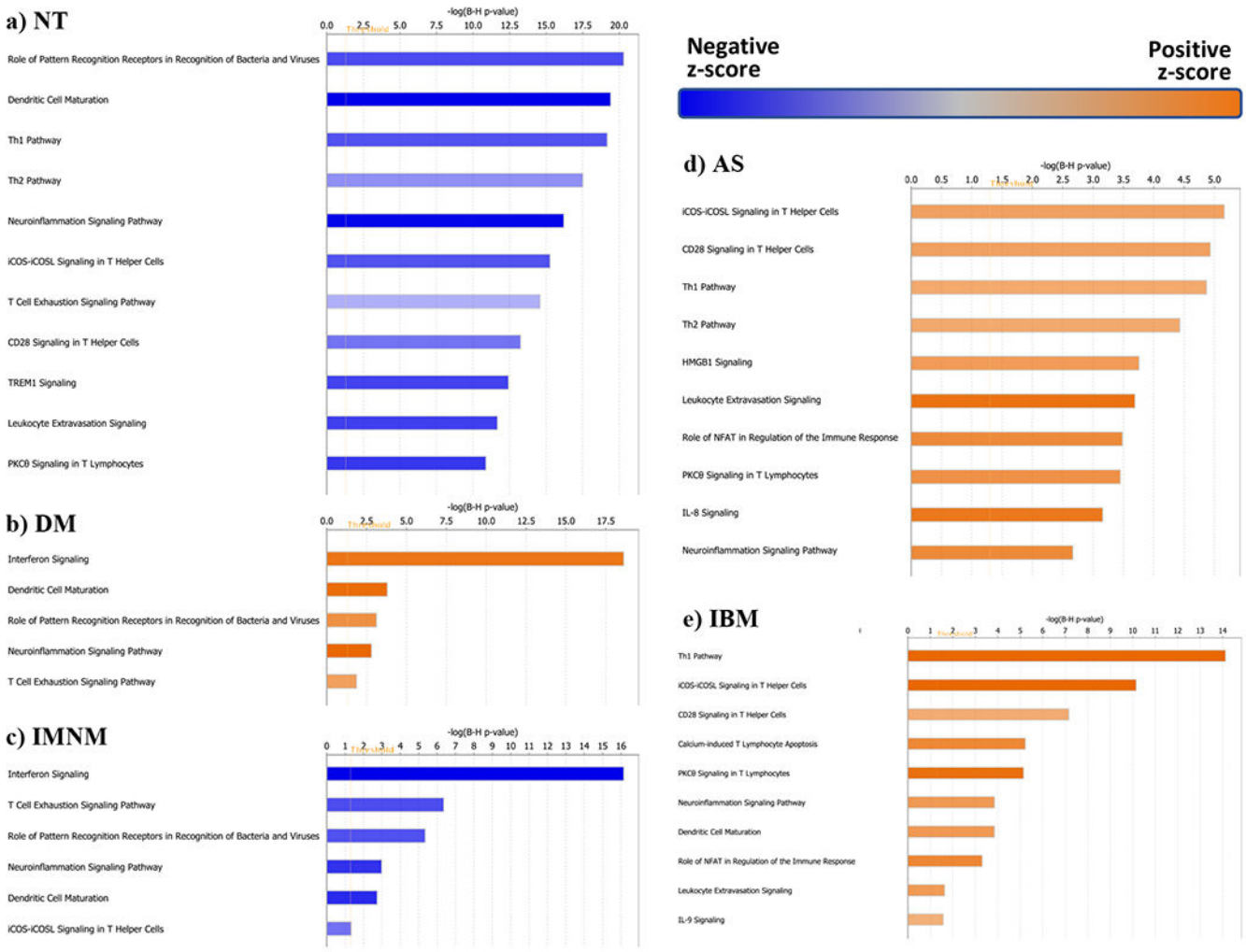

Figure 3. Pathway analysis in myositis and normal muscle biopsies.

The top 10 pathways of the different muscle biopsy groups are shown. NT, normal tissue; DM, dermatomyositis; IMNM, immune-mediated necrotizing myopathy; AS, antisynthetase syndrome; IBM, inclusion body myositis.

Table 1.

Genes differentially expressed in muscle biopsies from each major type of myositis and controls compared to the rest of the samples.

NT			DM (All)			Anti-Mi2			Anti-NXP2			Anti-MDA5			Anti-TIF1g			AS (Jo1)			IBM			IMNM (All)			Anti-HMGCR			Anti-SRP		
gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval
ISG15	−17	4.0E-40	**ISG15**	43	1.3E-139	**SCRT1**	14	1.4E-20	**ISG15**	12	4.2E-22	**ZFHX2**	11	1.3E-22	**MX1**	8	8.0E-16	**EGR4**	6	1.2E-09	**MYH4**	14	7.3E-19	**ISG15**	−12	3.0E-49	**ISG15**	−14	4.6E-51	**ISG15**	−6	4.5E-08
IFI6	−15	6.6E-39	**IFI6**	25	2.9E-107	**KCNJ4**	9	2.0E-16	**IFI6**	8	3.0E-16	**ISG15**	18	1.3E-22	**ISG15**	9	8.1E-16	**BRE-AS1**	4	3.2E-09	**ISG15**	−6	3.9E-12	**RSAD2**	−7	2.1E-38	**RSAD2**	−7	2.4E-38	**RSAD2**	−4	1.9E-07
PSMB8	−9	9.2E-32	**RSAD2**	12	1.8E-78	**COL11A2**	5	2.0E-16	**RSAD2**	6	1.0E-15	**DNAH1**	10	4.5E-16	**IFI6**	7	3.0E-13	**RNF165**	4	3.2E-09	**CRYBG3**	7	1.8E-09	**KLHDC7B**	−10	1.5E-31	**KLHDC7B**	−15	2.4E-32	**IFI6**	−5	1.2E-06
SECTM1	−10	2.3E-30	**MX1**	14	1.5E-75	**CHRM4**	11	1.0E-15	**KLHDC7B**	10	2.4E-14	**USP5**	5	2.8E-15	**HERC6**	5	1.8E-12	**CAMK1G**	6	8.2E-08	**AHNAK**	2	3.5E-09	**IFI6**	−7	6.5E-30	**MX1**	−6	4.5E-31	**IRF9**	−3	9.9E-05
ACTC1	−11	6.0E-30	**CMPK2**	9	1.2E-65	**MADCAM1**	8	1.2E-13	**IFIT2**	5	7.7E-13	**RRP7A**	6	3.0E-15	**SUSD2**	5	3.6E-11	**SAA1**	5	1.6E-07	**FCRL6**	7	3.5E-09	**CMPK2**	−4	9.0E-25	**IFI44L**	−6	1.1E-30	**STAT1**	−3	2.2E-04
IFI30	−12	5.6E-29	**MX2**	8	3.4E-55	**IFI6**	7	1.2E-13	**MX1**	5	8.6E-11	**AGPAT2**	9	1.1E-14	**DHX58**	4	7.8E-11	**ALPL**	3	4.2E-07	**GBP6**	7	6.4E-09	**MX1**	−5	1.2E-24	**CMPK2**	−5	6.8E-29	**IFI27**	−3	2.9E-04
SIGLEC1	−7	3.8E-28	**IFI27**	8	3.5E-55	**SPIB**	7	4.7E-13	**HERC5**	5	3.9E-10	**POU5F1P4**	9	1.3E-14	**MX2**	5	1.2E-10	**IL1RL1**	5	4.6E-07	**KIAA1147**	2	9.8E-09	**IFI27**	−5	5.8E-24	**OAS3**	−5	4.6E-28	**ZNFX1**	−2	2.9E-04
MX1	−9	3.0E-27	**OAS3**	8	2.4E-54	**ISG15**	7	1.9E-12	**NDUFB2-AS1**	4	3.9E-10	**HOXB-AS1**	11	2.4E-14	**IFI44**	5	1.4E-10	**SPP1**	5	5.9E-07	**PPM1L**	3	1.2E-08	**ZBP1**	−8	1.8E-23	**IFI44**	−5	1.4E-27	**DDX58**	−3	4.4E-04
OAS1	−9	3.3E-27	**HERC6**	7	7.6E-53	**MX1**	6	5.3E-12	**LOC101928053**	5	3.9E-10	**ACOT9**	5	3.0E-14	**RSAD2**	5	1.6E-10	**MIR6087**	4	1.5E-06	**LOC100128494**	5	3.5E-08	**DDX58**	−4	7.8E-23	**OAS2**	−4	5.6E-26	**NDUFS2**	3	4.4E-04
MX2	−7	5.9E-27	**OAS1**	9	8.2E-52	**COX6B2**	6	9.1E-12	**OAS3**	4	6.1E-10	**FRA10AC1**	5	4.8E-14	**HELZ2**	4	1.8E-10	**PBDC1**	2	1.9E-06	**KIAA0754**	5	3.5E-08	**IFI44L**	−5	2.2E-22	**IFI6**	−6	2.1E-25	**IFIT2**	−3	4.4E-04

NT: normal muscle tissue; DM: dermatomyositis; AS: antisynthetase syndrome; IBM: inclusion body myositis; IMNM: immune-mediated necrotizing myositis; FC: fold-change; qval: adjusted p-value. The name and location of the genes is indicated in Supplementary Table 2.

Table 2.

A comparison of machine learning models to classify muscle biopsies based on gene expression data.

Accuracy and 95% confidence interval in the 1000 test sets of the different machine learning models to classify muscle biopsies into normal muscle tissue (NT), dermatomyositis (DM), antisynthetase syndrome (AS), inclusion body myositis (IBM) or immune-mediated necrotizing myopathy (IMNM).

	NT	DM	AS	IBM	IMNM
Linear SVM	94.7 [87.2-100.0]	92.0 [85.1-97.9]	91.0 [85.1-95.7]	95.0 [91.5-100.0]	92.0 [85.1-97.9]
AdaBoost	91.5 [83.0-97.9]	89.6 [80.9-95.7]	89.1 [83.0-93.6]	91.9 [80.9-97.9]	85.8 [76.6-93.6]
Gaussian Process	94.2 [87.2-100.0]	82.9 [74.5-91.5]	87.2 [80.9-91.5]	91.0 [85.1-95.7]	79.6 [68.1-89.4]
Nearest Neighbors	91.5 [85.1-97.9]	87.8 [80.9-95.7]	87.2 [83.0-89.4]	90.6 [89.4-93.6]	77.4 [66.0-87.2]
Random Forest	89.7 [83.0-95.7]	85.6 [76.6-93.6]	85.7 [78.7-91.5]	90.4 [87.2-93.6]	78.3 [68.1-87.2]
Neural Network	89.1 [72.3-97.9]	83.5 [44.7-95.7]	87.4 [74.4-93.6]	91.1 [89.4-97.9]	71.6 [36.2-95.7]
Decision Tree	87.8 [76.6-95.7]	86.5 [76.6-93.6]	85.0 [74.5-91.5]	85.7 [76.6-93.6]	76.1 [57.4-89.4]
RBF SVM	85.1 [85.1-85.1]	82.6 [76.6-87.2]	87.2 [87.2-87.2]	89.4 [89.4-89.4]	64.0 [63.8-66.0]
Gaussian Naïve Bayes	85.1 [85.1-85.1]	80.2 [70.2-89.4]	86.4 [83.0-89.4]	89.3 [87.2-91.5]	66.1 [53.2-78.7]
QDA	86.5 [78.7-93.6]	63.5 [48.9-76.6]	75.5 [61.7-87.2]	80.4 [68.1-89.4]	63.1 [46.8-76.6]

SVM: support vector machines; RBF: radial basis function; AdaBoost: adaptative boosting; QDA: quadratic discriminant analysis. The models are sorted based on the average accuracy of all the groups.

Table 3.

The top 10 most useful genes to differentiate biopsy samples using the recursive feature elimination technique on the support vector machine model.

NT	DM	AS	IBM	IMNM
ACTC1	MX1	CAMK1G	MYH4	STAT1
LOC151121	TUBA1A	EGR4	H19	MYH8
SAA1	ISG15	CXCL8	JCHAIN	PSMB9
SOCS3	MCU	PROK2	CFAP126	KLF10
ANKRD1	HIST2H2AA3	NT5C3A	NT5C1A	MYBPH
NREP	IFI6	CXCL9	CCL13	ISG15
CCDC3	RARRES3	CAPN6	S100A9	MIR23A
PLEKHO1	CYB5R3	RAB13	COQ10A	COL3A1
SAA2	IGFN1	ANKRD28	DBNDD1	IGLL5
MYBPH	CDKN1A	C2ORF40	ZNF106	HIST1H2BD

NT: normal muscle tissue; DM: dermatomyositis; AS: antisynthetase syndrome; IBM: inclusion body myositis; IMNM: immune-mediated necrotizing myopathy; The name and location of the genes is indicated in Supplementary Table 2.

Table 4.

The top 10 up-regulated genes in each type of myositis compared to normal biopsies.

DM (All)			Anti-Mi2			Anti-NXP2			Anti-MDA5			Anti-TIF1g			AS (Jo1)			IBM			IMNM (All)			Anti-HMGCR			Anti-SRP		
gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval	gene	FC	qval
ISG15	101	1.06E-91	**IFI6**	62	2.6E-43	**ISG15**	110	1.5E-55	**ISG15**	163	1.2E-51	**ISG15**	84	1.5E-47	**PSMB8**	13	2.64E-25	**GBP2**	7	1.27E-18	**SERPINA3**	22	1.63E-28	**ACTC1**	18	3.6E-31	**SERPINA3**	24	1.8E-17
IFI6	67	2.73E-80	**ISG15**	67	5.9E-43	**IFI6**	70	7.8E-48	**IFI6**	72	2.6E-37	**IFI6**	62	1.7E-43	**ACTC1**	18	2.33E-23	**BIRC3**	7	4.43E-18	**ACTC1**	15	1.08E-27	**SERPINA3**	20	5.8E-25	**ACTC1**	12	9.8E-13
MX1	29	2.58E-56	**MX1**	32	1.9E-33	**RSAD2**	23	2.1E-33	**ZFHX2**	18	9.9E-25	**MX1**	41	9.0E-39	**GBP2**	7	1.11E-22	**PSMB8**	9	1.70E-16	**CHRNA1**	8	6.40E-21	**MYH3**	17	8.1E-21	**HP**	18	1.6E-12
RSAD2	18	1.25E-49	**OAS1**	25	9.7E-28	**MX1**	29	8.6E-33	**IFI35**	30	2.3E-24	**MX2**	22	6.9E-32	**SAA1**	37	1.11E-22	**GBP1**	11	3.50E-16	**IFITM10**	9	3.20E-20	**CHRNA1**	7	2.0E-19	**CHRNA1**	9	2.7E-12
MX2	17	2.48E-49	**MX2**	18	1.2E-27	**IFIT2**	22	5.6E-31	**ACP5**	35	2.2E-23	**OAS1**	29	3.8E-31	**SIK1**	10	3.05E-21	**CCL13**	20	4.96E-16	**TNC**	11	5.38E-20	**IFITM10**	9	2.6E-19	**TNC**	11	6.7E-12
OAS1	23	4.00E-48	**LY6E**	16	4.5E-26	**KLHDC7B**	65	1.1E-30	**SECTM1**	26	1.0E-22	**IFITM1**	12	2.7E-28	**NNMT**	14	9.65E-21	**ITGAL**	9	2.47E-15	**KRT80**	13	1.35E-19	**TNNT2**	16	1.7E-18	**DCLK1**	7	6.7E-12
IRF9	10	2.54E-43	**RSAD2**	17	8.1E-26	**OAS1**	26	4.7E-30	**ZBP1**	42	1.6E-22	**RSAD2**	18	2.0E-27	**MYH3**	23	9.83E-21	**GBP5**	17	3.95E-15	**TNNT2**	15	1.35E-19	**ANKRD1**	10	2.4E-18	**KRT80**	13	6.7E-12
IFITM1	10	8.62E-43	**CMPK2**	15	2.2E-25	**OAS3**	16	1.3E-26	**CLEC4GP1**	19	5.4E-21	**TYMP**	30	2.0E-27	**GADD45A**	8	9.83E-21	**HLA-DQA1**	13	2.41E-14	**MYH3**	13	2.66E-19	**CSPG4**	5	2.4E-18	**IFITM10**	10	1.0E-11
CMPK2	13	4.24E-42	**OAS3**	16	3.6E-25	**HERC5**	25	3.4E-26	**PSMB8**	19	3.3E-20	**IFI44**	18	4.1E-27	**GBP1**	12	9.83E-21	**CD8A**	16	3.45E-14	**ANKRD1**	10	2.04E-18	**TNC**	10.06	2.6E-18	**RUNX1**	6.98	3.0E-11
OAS3	14	2.27E-41	**KLHDC7B**	39	3.0E-23	**MX2**	15	4.8E-26	**IFI27**	18	2.2E-19	**DHX58**	13	6.2E-27	**IFI30**	16	1.02E-20	**HLA-DOA**	10	3.58E-14	**DCLK1**	6	2.45E-18	**KRT80**	11.97	9.9E-18	**TNNT2**	13.81	3.3E-11

DM: dermatomyositis; AS: antisynthetase syndrome; IBM: inclusion body myositis; IMNM: inclusion body myositis; FC: fold-change; qval: adjusted p-value. The name and location of the genes is indicated in Supplementary Table 2.