Formosan Journal of Rheumatology 2024;38:40-52 10.6313/FJR.202406_38(1).0004

Original Article

Characterize the Taiwanese idiopathic inflammatory myositis cohort by hierarchical clustering

Pin-Yi Wang[1#], Ching-Yun Wang[2#], Wan-Yu Sung[1], Tsan-Teng Ou[1], Cheng-Chin Wu[1], Chia-Chun Tseng[1,3], Pei-Pei Chen[1], Hsin-Yu Ho[4], Jih-Jin Tsai[1,3,5,6], Jeng-Hsien Yen[1,7*]

[1] Division of Allergy, Immunology and Rheumatology, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[2] Department of Medical Education, Taichung Veterans General Hospital, Taichung, Taiwan
[3] School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[4] School of Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[5] Tropical Medicine Center, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[6] Division of Infectious Diseases, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[7] Graduate Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
[#] Equal contribution [*] Corresponding author

Objectives: This study aims to establish a clinico-seropathological classification system inclusive of 16 myositis specific antibodies (MSAs)/myositis associated antibodies (MAAs) to sub-classify patients from a southern Taiwanese cohort with idiopathic inflammatory myopathies (IIM), focusing on clinical myositis features and MSAs/MAAs. Additionally, the study extends to performing survival analysis on the sub-classified groups.

Methods: This study includes 108 Taiwanese adults with defined IIM in our single center and two affiliated hospitals from 2002 to 2022. Using a dataset that includes demographics, disease manifestations, laboratory examinations, and 16 MSAs/MAAs, patients were sub-classified through bioinformatics tools such as categorical principal component analysis and hierarchical cluster algorithms. The methodology aimed to condense complex clinical data into essential components, allowing for the identification of unique patient subgroups.

Results: A total of 108 patients, from 2002 to 2022, were clustered into the unique IIM subgroups with distinct pathophysiological profiles (Clusters 3 and 4), and confirmed conventional IIM phenotypes (Clusters 1 and 6) by integrating clinical and molecular features. Survival analysis highlighted a higher incidence of interstitial lung disease (ILD) and poorer overall survival among anti–melanoma differentiation-associated gene 5 (anti-MDA5) positive IIM patients, echoing the clinically amyopathic dermatomyositis (CADM) characteristics.

Conclusion: The integration of MSAs/MAAs profiles with clinical and molecular features has facilitated the identification of distinct IIM subgroups. Some of these subgroups present unique pathophysiological characteristics that warrant further study. Additionally, survival analysis based on specific MSAs/MAAs reactivities has confirmed the severe prognosis for anti-MDA5 positive IIM patients, consistent with the literature on anti-MDA5 positive CADM.

Corresponding author: Jeng-Hsien Yen
Division of Allergy, Immunology, and Rheumatology, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
Address: No. 100, Tzyou 1st Road Kaohsiung 807, Taiwan
Tel: +886-7-3121101 ext 6088; Fax: +886-7-3221505
E-mail: jehsye@kmu.edu.tw
Received: Mar 4, 2024
Revised: Apr 17, 2024
Accepted: Apr 30, 2024

Key words: Idiopathic inflammatory myositis;Myositis specific autoantibodies;Principal component analysis;Interstitial lung disease;Melanoma differentiation-associated protein 5

1. Introduction

The 2017 American College of Rheumatology (ACR) Board of Directors and the European League Against Rheumatism (EULAR) classification criteria for adult and juvenile idiopathic inflammatory myositis (IIM) is currently widely adopted due to its superior performance of predicting the disease among other conventional criteria. ACR/EULAR criteria involve several clinical characteristics (e.g. age of onset, muscle weakness, specific skin manifestations) and laboratory evidence (e.g. Anti-Jo1-autoantibody positivity, elevated serum levels of creatine kinase, elevated liver enzymes, and typical muscle biopsy findings).

Aligning with these criteria defines adult IIM, which has been recognized as a heterogeneous disease entity with diverse comorbidities and corresponding prognosis. These IIM clinical subgroups can be further classified by related comorbidities, biochemical examination, and even molecular evidence (e.g. Myositis Specific Antibodies, MSAs). For example, dermatomyositis (DM) with anti-MDA5 antibodies features interstitial lung disease (ILD) and other extramuscular manifestations, and has a relatively poor prognosis if rapidly-progressive ILD occurs. Importantly, the advance of molecular diagnosis[1-3], and bioinformatics prompted several attempts to fairly sub-classify the criteria-meeting IIMs based on the immense clinical information with more precise and less redundant approaches. For instance, principal component analysis (PCA) is a widely used dimensionality reduction methodology for heterogeneous clinical data, especially popular in Rheumatology[4]. In 2019, *Huiyi Zhu* et al.[5] collected 21 clinical features, including demographics, disease manifestations, and laboratory examinations, and sub-classified 720 dermatomyositis patients into six clusters. Those features were treated as categorical variables and sequentially processed through categorical principal component analysis (CATPCA) and hierarchical cluster algorithms. The aim of PCA is to condense a primary group of variables into a more concise collection of independent components that capture the majority of the data inherent in the initial variables. This methodology could substantially reduce redundant information while only weighing in crucial factors to subset a bulky dataset. Regarding rheumatological diseases based largely on clinical diagnosis, bioinformatics can kick in to help downsize the complexity generated by the miscellaneous clinical manifestations and offer a proper classification procedure. *Huiyi Zhu* et al. revealed subgroups of “classical DM with minimal organ impact", "DM with malignancy-related issues", "CADM with ILD", "DM emphasizing lung, muscle, and skin", "Overlapping syndromes", and "DM with leading cardiomyopathy."

On the other hand, *Junmei Zhang* et al.[6] adopted a similar analytic methodology to sub-classify 132 Chinese juvenile dermatomyositis (JDM) patients into four clusters in 2022. Importantly, myositis-specific antibodies (MSAs) were also integrated into the clustering algorithm of *Junmei Zhang's* work. Despite the breakthrough of these studies, they only incorporate a limited number of MSAs at the beginning of data extraction.

Enlightened by previous works, we wanted to establish a clinico-seropathological classification system, including 16 available MSAs/MAAs, in Asian IIM patients. This study was designed to explore the specifications of a southern Taiwanese IIM cohort and to sub-classify them based on clinical features and 16 MSAs/MAAs. Moreover, we also extended from pure subclassification to survival analysis of the clustering results, which were barely conducted in similar studies.

2. Materials and Methods

2.1 Population

A total of 108 patients diagnosed with IIM from the Department of Rheumatology inpatient wards of the Kaohsiung Medical University Hospital between 2002 and 2022 were enrolled in our study. The Bohan and Peter criteria were applied to diagnose IIM[7]. Ethics approval (KMUHIRB-E(I)-20230013) was obtained from the Institutional Review Board of Kaohsiung Medical University Hospital.

Determination of the disease activity of AS

The disease activity of AS was assessed using the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) [13] before (at baseline) and after 12 months of therapy with bDMARDs. The erythrocyte sedimentation rate (ESR) was determined using the Westergren method (Sed

Rate Screener 20/II; Greiner Bio-One, Kremsmünster, Austria), and the C-reactive protein (CRP) level was measured by an immunoturbidimetric method (Beckman Coulter, Inc., Brea, CA, USA). The serum level of IgA was determined by the nephelometer method (Beckman Coulter, Inc., Brea, CA, USA) based on the light scattered onto the antigen-antibody complexes.

2.2 Data extraction

The data recorded in the electronic medical record ranged from IIM diagnosis to subsequent follow-up. These included the following parameters: sex, age, IIM-related clinical manifestations, laboratory findings, MSAs/MAAs positivity, comorbidities, and immunosuppressive regimens. We performed the MSAs/MAAs profiles when diagnosing these patients or during the follow-up. All 108 patients underwent at least one testing for MSAs/MAAs profiles.

2.3 Data analysis

CATPCA: Categorical Principal Components Analysis (CATPCA) performs a dual role of quantifying categorical variables and reducing data dimensionality. It efficiently quantifies categorical variables within the chosen dimensions, allowing for optimal representation. This approach enables the modeling of nonlinear relationships between variables. We performed CATPCA on SPSS (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp).

Agglomerative hierarchical clustering: Based on CATPCA-transformed data, we performed agglomerative hierarchical clustering[8] using R software v4.2.3 (R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Cluster analysis, a widely used unsupervised learning method, identifies subgroups with similar characteristics. In this context, the "Ward D" method was explicitly selected. This method minimizes within-cluster variance while amalgamating clusters based on the smallest between-cluster distance. Final cluster selection was achieved according to the maximized Calinski and Harabasz Index (CH Index), which ensured a sizeable within-cluster variation and a slight within-cluster variation[9].

Survival analysis was conducted using the Kaplan-Meier method[10]. We implemented this analysis using the Python “lifelines[11].” We calculated and plotted survival probabilities over time using the Kaplan-Meier estimator, effectively addressing censored data. Furthermore, we employed the log-rank test to assess differences in survival curves between distinct subgroups, also facilitated by the lifelines package[12].

Finally, in the case of continuous variables that exhibited a normal distribution, contrasts between the clusters were evaluated through non-parametric examinations such as the Kruskal-Wallis test. Alternatively, for variables with categorical attributes, either the chi-square test or Fisher's exact test was applied. The threshold for statistical significance was set at $p<0.05$.

3. Results

3.1 Taiwanese IIM cohort characteristics

We selected 26 features frequently found in the IIM and MSAs/MAAs positivity as original variables for cluster analysis (Table 1). The selected clinical variables (e.g. muscle weakness and skin manifestations) aligned with the definitions of the 2017 ACR/EULAR classification criteria. Serum MSAs/MAAs in 108 patients, including anti-Mi-2α, anti-Mi-2 β, anti-TIF1-γ, anti-MDA5, anti-NXP2, anti-SAE1, anti-Ku, antiPM-Scl100, anti-PM-Scl75, anti-Jo-1, anti-SRP, anti-PL-7, anti-PL-12, anti-EJ, anti-OJ, anti-Ro-52, were examined by Euroline (Immunoblot, Euroimmun company). The distribution of demographics, some 2017 ACR/EULAR-encoded clinical features, and MAAs/MSAs positivity of the 108 Taiwanese IIM patients was displayed in (Table 1). Most of the patients are female, with an average age-of-onset of about 52 years old. 81.5% of patients had muscle enzyme (CPK, Creatine-phospho-kinase) elevation. Extra-muscular involvements were seen in part of the cohort. For example, 39.8% (n=43) patients had interstitial lung disease, while 31.5% (n=34) had esophageal involvement. Besides, our patients had a heterogeneous MSAs/MAAs expression profile, in which the top-third most prevalent MSAs/MAAs were Ro-52 (49.1%), TIF1-γ (20/4%), and Jo-1(18.5%). There is no missing data in any of the 26 enlisted variables.

3.2 Dimensional reduction of the heterogeneous clinical characteristics

As a result of CATPCA, the original 26 selected clinical variables were reduced to a compact 11 principal components (PCs) accounting for 73% data variance. Eventually, we preserved 13 clinical variables with vector loading greater than 0.6 from these PCs as our core variables for hierarchical analysis. These clinical variables are all categorical variables, including *Gottron*

papules, *Gottron signs*, *Heliotrope* signs/rash, esophageal involvement, symmetric proximal upper extremities muscle weakness, symmetric proximal lower extremities muscle weakness, and MSAs/MAAs positivity of *Mi2α*, *TIF1-γ*, *SAE1*, *Ku*, *PMScl100*, *PMScl75*, and *PL12*.

3.3 Six IIM clusters revealed by Agglomerative clustering algorithms

According to the maximized *CH* index during the "ward D" process, the 108 Taiwanese IIM patients were clustered into six subgroups, denoted from "Cluster 1" to "6". Based on the cluster results, we revealed four IIM clusters with evident MSAs features. Further analysis of the association between patients with these specific MSAs features and their clinical manifestations was displayed in Table 2. As for all recorded clinical features, there is no statistical significance among the six clusters. Cluster 1 (n=18) represents a group of IIM patients with ADM and/or ILD diagnosis and the highest anti-MDA5 antibody prevalence among the six IIM clusters. Other features are reportedly oldest average age-of-onset (**57.76**), relatively small DM prevalence, and reportedly highest occurrence of inverse *Gottron* papules (33%). Cluster 2 (n=22) specifies an IIM subgroup with the youngest average age-of-onset, elevated muscle enzymes (100%), and symmetrical lower extremities muscle weakness (100%) upon diagnosis. Other features include a high esophageal involvement rate (59%) and a high upper extremities muscle weakness rate (86%). Cluster 3(n=17) features co-expression of anti-Jo-1(100%) and anti-Ro-52(94%) MSAs/MAAs and is comorbid with ILD (71%). Other features include a high occurrence of elevated muscle enzyme (88%), lower extremities muscle weakness (65%), and myalgia (71%). Cluster 4(n= 8) features co-expression of anti-EJ and anti-Ro-52 MSAs/MAAs and is also comorbid with ILD (75%). Cluster 5 (n=31) is a less specified subgroup, composed of miscellaneous IIM features. Other features include a high occurrence of elevated muscle enzyme (77%) and facial rash (55%). Last, cluster 6 (n=12) represents DM patients (12/12, 100%) with the highest anti-*TIF1-γ* antibody prevalence and esophageal involvement rate (67%), and are comorbid with cancer (33%). Other features include expression of several conventional dermatomyositis clinical manifestations, such as upper/lower (100%/83%) extremities muscle weakness, myalgia (92%), *Gottron sign* (75%), *Heliotrope* signs/rash (92%), and facial rash (83%).

Table 1. Clinical Characteristics of 26 variables extracted from the Taiwanese IIM cohort

Characteristics	Number (% or STD)
Sex	Male: 41 (38%); Female:67 (62%)
Average age of onset[a]	52.39 (STD:16.322)
Elevated muscle enzymes	88 (81.5%)
Gottron papules	35(32.4%)
ILD	43(39.8%)
Esophageal involvement	34(31.5%)
Muscle weakness (Upper)	44(40.7%)
Muscle weakness (Lower)	46(42.6%)
Myalgia	51(47.2%)
Gottron signs	22(20.4%)
Heliotroph signs/rash	25(23.1%)
Inverse *Gottron* papules	7 (6.5%)
MSA/MAA positivity	
Mi-2α	2(1.9%)
Mi-2 β	7(6.5%)
TIF1-γ	22(20.4%)
MDA5	15(13.9%)
NXP2	6(5.6%)
SAE1	2(1.9%)
Ku	12(11.1%)
PM-Sc110	5(4.6%)
PM-sc175	11(10.2%)
Jo-1	20(18.5%)
SRP	9(8.3%)
PL-7	7(6.5%)
PL-12	3(2.8%)
EJ	12(11.1%)
OJ	3(2.8%)
Ro-52	53(49.1%)

Abbreviations: STD, standard deviation; ILD, interstitial lung disease; MAA, myositis-associated antibody; MSA, myositis-specific antibody

[a]Values are expressed as mean (standard deviation).

3.4 Clinical associations of 4 MSA-defining clusters derived from machine learning algorithm

According to the cluster results, 4 clusters, cluster 1(anti-MDA5+), cluster 3 (Anti-Jo-1 and Anti-Ro-52 co-expression), cluster 4(Anti-EJ and Anti-Ro-52 co-expression), and cluster 6(anti-TIF1-γ+) had relatively prominent MSAs features.

We employed logistic regression models to evaluate the association among 4 MSAs subgroups and clinical characteristics in individuals who tested positive for anti-MDA5, Anti-Jo-1 and Anti-Ro-52, Anti-EJ and Anti-Ro-52, and anti-TIF1-γ autoantibody. Table 3 and Table 4 summarize the clinical characteristics related to the

Table 2. Clinical characteristics of 108 patients with IIM based on the 6 clusters retrieved by CATPCA-based clustering algorithm

Cluster	1	2	3	4	5	6	P value
Number of pts	18	22	17	8	31	12	=1
Female	44%	68%	71%	25%	71%	58%	0.334
Average age of onset	57.76	44.05	55.53	51	51.61	56.92	=1
Elevated muscle enzymes	39%	100%	88%	100%	77%	100%	0.551
Gottron papules*	33%	23%	6%	0%	42%	75%	=1
IIM subtypes							
DM	11%	36%	35%	25%	16%	100%	=1
ADM	50%	0%	0%	13%	45%	0%	0.903
PM	0%	41%	41%	38%	16%	0%	=1
IMNM	0%	14%	0%	0%	0%	0%	=1
ASyS	11%	0%	12%	25%	3%	0%	=1
Overlap	6%	9%	6%	0%	13%	0%	=1
ILD	78%	9%	71%	75%	26%	0%	=1
Cancer	11%	18%	12%	13%	23%	33%	=1
Esophageal involvement*	6%	59%	18%	0%	29%	67%	=1
Myositis on biopsy	0%	41%	18%	0%	0%	0%	=1
Muscle weakness (Upper)*	6%	86%	41%	13%	13%	100%	=1
Muscle weakness (Lower)*	6%	100%	65%	0%	6%	83%	=1
Myalgia	11%	45%	71%	38%	42%	92%	0.789
Gottron signs*	17%	9%	6%	0%	23%	75%	=1
Heliotrope signs/rash*	6%	5%	6%	0%	26%	92%	=1
Shawl signs	11%	9%	0%	13%	13%	50%	=1
Vasculitis	0%	0%	0%	0%	0%	0%	=1
Facial rash	33%	18%	24%	25%	55%	83%	=1
Mechanics hands	17%	0%	29%	13%	19%	0%	0.802
Calcinosis	0%	0%	0%	0%	0%	0%	=1
Periungual hyperemia	22%	0%	0%	0%	6%	33%	0.832
Inverse *Gottron* papules	33%	0%	0%	0%	3%	0%	0.939
MSA positivity							
Mi-2α*	0%	0%	6%	0%	0%	8%	=1
Mi-2 β	11%	5%	0%	0%	3%	25%	=1
TIF1-γ*	11%	0%	0%	0%	35%	75%	=1
MDA5	67%	0%	0%	0%	6%	0%	=1
NXP2	0%	27%	0%	0%	0%	0%	=1
SAE1*	0%	0%	0%	0%	3%	8%	=1
Ku*	6%	27%	12%	13%	6%	0%	=1
PM-SCL100*	0%	0%	0%	0%	16%	0%	
PM-SCL75*	0%	18%	0%	0%	23%	0%	0.596
Jo-1	11%	0%	100%	0%	3%	0%	
SRP	0%	36%	0%	0%	3%	0%	
PL-7	0%	5%	6%	0%	13%	0%	
PL-12*	0%	0%	0%	0%	6%	8%	
EJ	6%	5%	0%	100%	6%	0%	0.156
OJ	11%	0%	0%	0%	3%	0%	
Ro-52	22%	27%	94%	100%	48%	25%	0.699

Abbreviations: Pts, patients; IIM, idiopathic inflammatory myositis; DM, dermatomyositis; ADM, Amyopathic dermatomyositis; PM, polymyositis; IMNM, immune-mediated necrotizing myopathy; ASyS, Antisynthetase syndrome; ILD, interstitial lung disease; MSA, myositis-specific antibody; *, selected variables involved in the hierarchical clustering process

Table 3. Univariate logistic regression analysis of the clinical associations of myositis-associated autoantibodies/ myositis-specific autoantibodies

Univariate analysis	More susceptible			Less susceptible		
		OR (95% CI)	*p* value		OR (95% CI)	*p* value
Anti-MDA5+	ILD	5.242 (1.545-17.787)	0.008**	Elevated muscle enzyme	0.060 (0.017-0.212)	0.000
	V sign	7.262 (2.170-24.307)	0.001**	Upper muscle weakness	0.083 (0.010-0.658)	0.018*
	Skin ulcer	10.909 (2.153-55.267)	0.004**	Lower muscle weakness	0.076 (0.010-0.603)	0.015*
	Periungual hyperemia	4.468 (1.124-17.749)	0.033*	Myalgia	0.234 (0.062-0.885)	0.032*
EJ+Ro-52	ILD	8.338 (1.704-40.804)	0.009**	Lower muscle weakness	0.116 (-)	0.043*
Jo1+Ro-52	ILD	4.645 (1.501-14.376)	0.008**			
	Myalgia	3.200 (1.041-9.835)	0.042*			
TIF1-γ	*Gottron* signs	3.888 (1.381-10.940)	0.010*			
	Heliotrope signs	8.907 (3.129-25.356)	0.000			
	Cancer	4.269 (1.500-12.153)	0.007**			
	Shawl signs	5.571 (1.794-17.301)	0.003**			
	Facial Rash	7.430 (2.483-22.235)	0.000			
	Generalized erythroderma	6.632 (1.035-42.480)	0.046*			

Abbreviations: ILD, interstitial lung disease; CI, confidence interval; OR, Odds ratio; * < 0.05; ** < 0.01

various MSAs/MAAs. Anti-MDA5 MSA existed in 15 patients (13.89%) and strongly correlated with ILD, V sign, skin ulcer, and periungual hyperemia. However, anti-MDA5 was negatively correlated with elevated muscle enzymes, upper and lower muscle weakness, and myalgia. Anti-Jo-1/ Anti-Ro-52 co-expression was present in 17 patients (15.74%) and had strong associations with ILD and Myalgia. Anti-EJ/Anti-Ro-52 co-expression was present in 11 patients (10.18%) and was associated with high ILD appearance. Patients with anti-EJ/Anti-Ro-52 co-expression had negative association of lower muscle weakness. Anti-TIF1-γ was present in 22 patients (20.37%) and had strong associations with skin manifestations (*Gottron signs*, *Heliotrope* signs/rash, Shawl signs, facial rash) and cancer.

3.5 Survival analysis of 4 MSA-defining clusters derived from machine learning algorithm

The median follow-up duration was 41.5 months (interquartile range [IQR]:

15.25 to 73.75 months). During analysis, 19 of the 108 patients (17.6%) had died. Comparisons of the 5-year overall survival among different MSAs subgroups using the log-rank test revealed significant differences (Figure 2). Patients with anti-MDA5 had a poor

5-year overall survival ($P < 0.005$; Fig. 2A); nonetheless, 5-year overall survival among those who tested positive for Anti-Jo-1/ Anti-Ro-52 co-expression, Anti-EJ/Anti-Ro-52 co-expression, and Anti-TIF1-γ showed no significant difference.

4. Discussion

Our study analyzed 26 clinical features and MSAs/ MAAs in 108 Taiwanese patients with IIM from 2002 to 2022. Using CATPCA and hierarchical clustering, the

Table 4. Multivariate logistic regression analysis of the clinical associations of myositis-associated autoantibodies/ myositis-specific autoantibodies

Multivariate analysis	More susceptible			Less susceptible		
		OR (95% CI)	*p* value		OR (95% CI)	*p* value
Anti-MDA5+	ILD	5.956 (1.659-21.387)	0.006**	Elevated muscle enzyme	0.043 (0.010- 0.187)	0.000
	V sign	8.039 (2.318-27.889)	0.001**	Upper muscle weakness	0.082 (0.010-0,657)	0.018*
	Skin ulcer	11.298 (2.173-58.752)	0.004**	Lower muscle weakness	0.076 (0.010-0.604)	0.015*
	Periungual hyperemia	4.376 (1.095-17.491)	0.037*	Myalgia	0.218 (0.056-0.854)	0.029*
EJ+Ro-52	ILD	9.459 (1.811-49.418)	0.008**	Lower muscle weakness	0.106 (0.013-0.883)	0.038*
Jo1+Ro-52	ILD	4.812 (1.488-15.556)	0.009**			
	Myalgia	3.955 (1.208-12.951)	0.023*			
TIF1-γ	Gottron signs	4.573 (1.482-14.112)	0.008**			
	Heliotrope signs	8.166 (2.801-23.806)	0.000			
	Cancer	3.808 (1.299-11.158)	0.015*			
	Shawl signs	7.357 (2.076-26.075)	0.002**			
	Facial Rash	12.110 (3.518-41.685)	0.000			

Multivariate logistic regression, adjusted with sex, and age of onset
Abbreviations: ILD, interstitial lung disease; CI, confidence interval; OR, Odds ratio; * < 0.05; ** < 0.01

patients were clustered into unique IIM subgroups with distinct pathophysiological profiles (Clusters 3 and 4), and confirmed conventional IIM phenotypes (Clusters 1 and 6) by integrating clinical and molecular features. Survival analysis showed that patients with anti-MDA5 antibodies had higher incidence of ILD and significantly poorer 5-year overall survival compared to other MSAs subgroups.

In Taiwan, the annual incidence of DM and PM was 7.1 and 4.4 per million[13]. The mean age at diagnosis of DM and PM was, respectively, 44 and 49.2. Moreover, the female-to-male incidence ratio of DM and PM in patients around 40 to 49 years old was 2.75 and 3.85. Among our 108 IIM patients with available MSAs/MAAs reports, the average age of onset was 52.39 ± 16.32, with a reportedly higher proportion of female IIM patients (female 62%, male 38%). Our patients' demographic features resembled the statistics revealed in the nationwide database study. Moreover, cancers were diagnosed in 21(19.4%) IIM

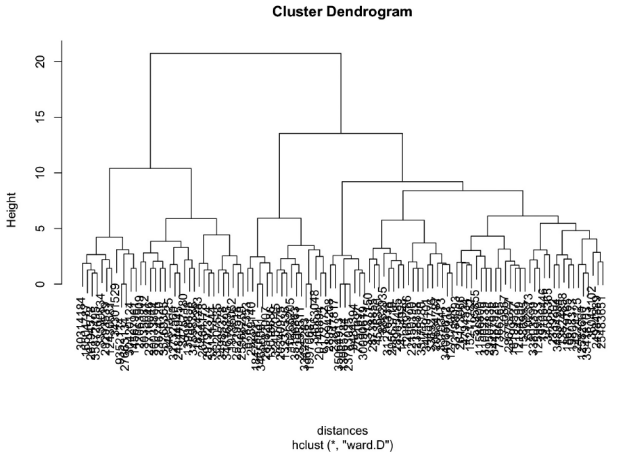
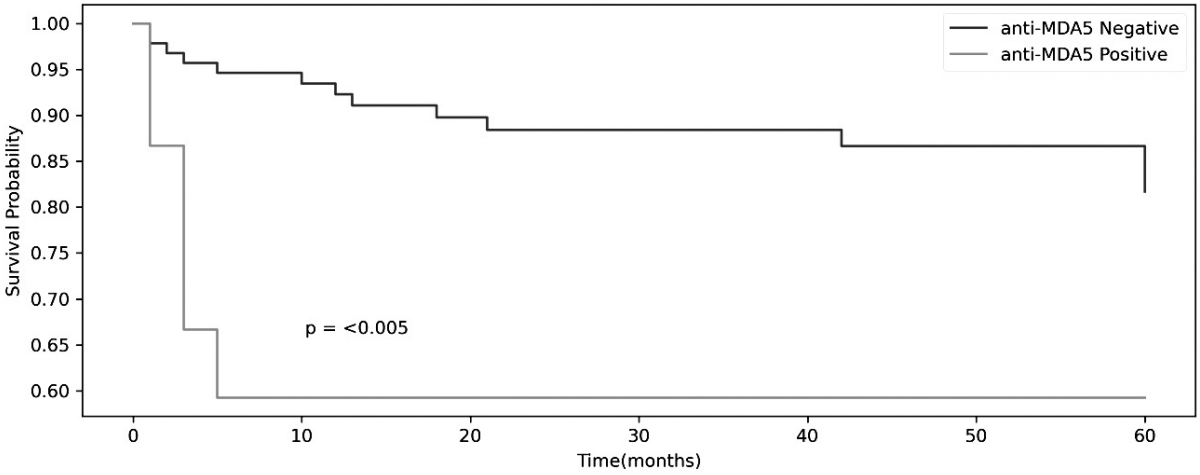


Figure 1. Agglomerative hierarchical clustering dendrogram of 108 IIM patients based on categorical principal components analysis. The figure displayed the combination process from 108 clusters to 1 cluster. The numbers below indicate the patient identification number

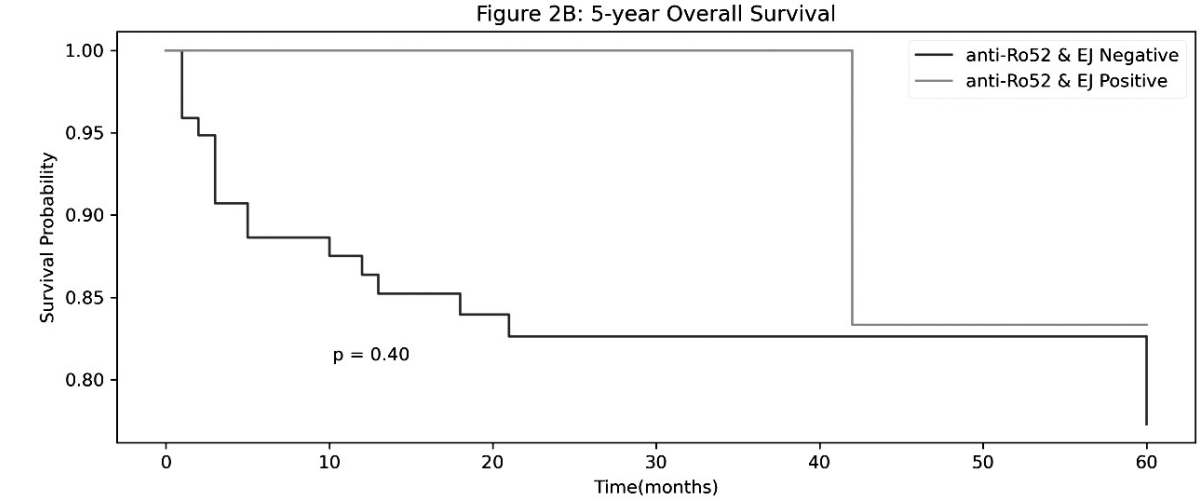
patients, of which the cancer occurrence was similar to the domestic trend of larger cohorts.

Among the six IIM clusters we demonstrated, some corresponded to well-known subgroups (Cluster 1 and Cluster 6), some resembled conventional IIM subgroups but had unique MSAs profiles (Cluster 3 and Cluster 4), while the rest were poorly identified subgroups with unspecific manifestations (Cluster 2 and Cluster 5).

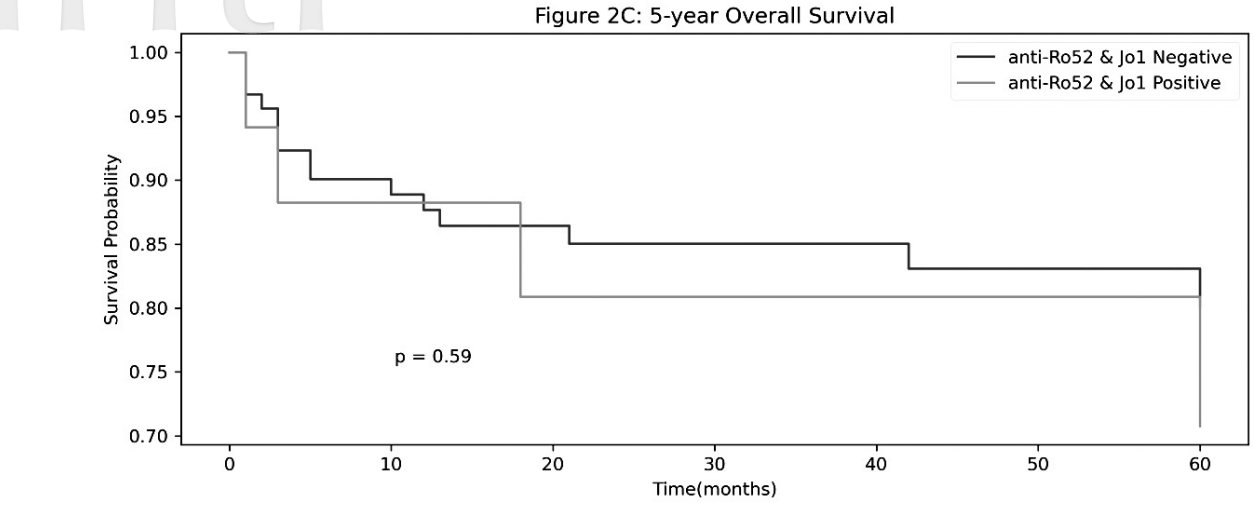
Cluster 1, ADM and/or ILD with anti-MDA5+, was an IIM subgroup with no myositis evidence on biopsy and a reportedly lowest occurrence rate (39%) of elevated muscle enzymes. Cluster 1 resembles a DM subgroup with anti-MDA5+ antibody-related phenotypes, ILD, and



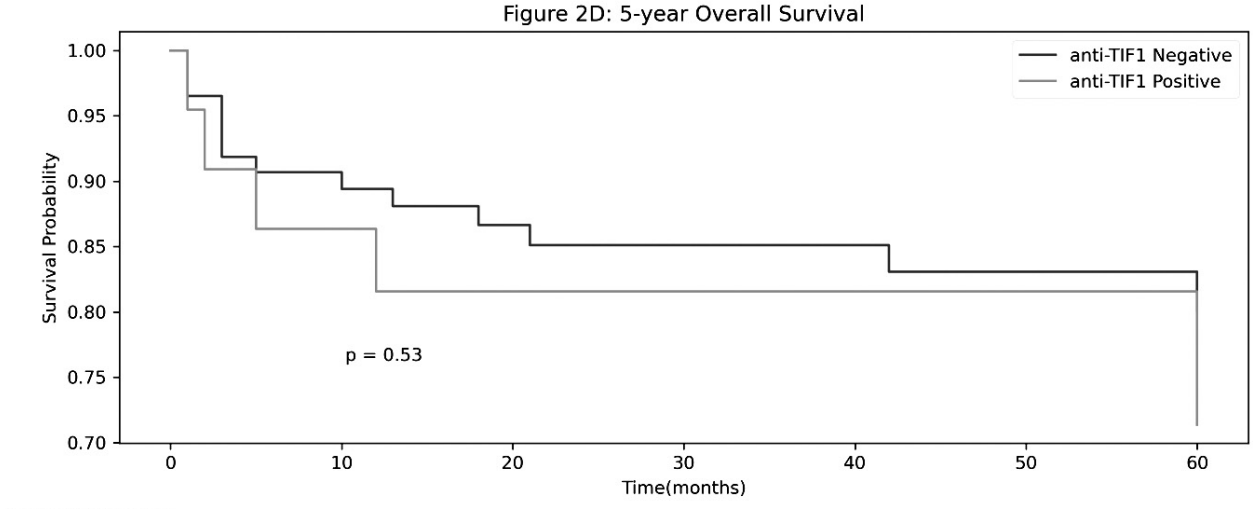
anti-MDA5 Negative							
At risk	93	79	65	58	53	43	0
Censored	0	8	19	25	30	39	80
Events	0	6	9	10	10	11	13
anti-MDA5 Positive							
At risk	15	8	7	6	4	1	0
Censored	0	1	2	3	5	8	9
Events	0	6	6	6	6	6	6



anti-Ro52 & EJ Negative							
At risk	97	77	64	56	51	39	0
Censored	0	8	18	25	30	42	79
Events	0	12	15	16	16	16	18
anti-Ro52 & EJ Positive							
At risk	11	10	8	8	6	5	0
Censored	0	1	3	3	5	5	10
Events	0	0	0	0	0	1	1



anti-Ro52 & Jo1 Negative							
At risk	91	73	61	54	47	36	0
Censored	0	8	18	24	31	41	76
Events	0	10	12	13	13	14	15
anti-Ro52 & Jo1 Positive							
At risk	17	14	11	10	10	8	0
Censored	0	1	3	4	4	6	13
Events	0	2	3	3	3	3	4



anti-TIF1 Negative							
At risk	86	69	56	48	43	35	0
Censored	0	8	19	26	31	38	72
Events	0	9	11	12	12	13	14
anti-TIF1 Positive							
At risk	22	18	16	16	14	9	0
Censored	0	1	2	2	4	9	17
Events	0	3	4	4	4	4	5

Figure 2. Kaplan-Meier survival curve and log-rank regression analysis on four MSA combinations (2A)Anti-MDA5 (2B) Anti-EJ and Anti-Ro52 (2C)Anti-Jo-1 and Anti-Ro52(2D)Anti-TIF1-γ

variable extra-muscular manifestations included[14].

Cluster 6, 12 DM patients, of which 8 (75%) were anti-TIF1-γ+ and 4 (33.3%) were diagnosed with cancers. Anti-TIF1-γ antibody is commonly seen in IIM patients and is a reliable risk factor for comorbid malignancy (HR 3.85, adjusted P value 0.003)[15].

According to *Connors* diagnostic criteria, clusters 3 and 4 resembled anti-synthetase syndrome (ASyS) by presenting with anti-aminoacyl tRNA synthetase (anti-ARS) autoantibodies and clinical symptoms of

ILD[16]. Cluster 3 mostly comprises DM (35%) and PM (41%) patients with co-expressed anti-Jo-1 and anti-Ro-52 MSAs/MAAs. In a previous study, anti-Ro-52 MAAs had a significantly higher occurrence rate (58%) in the anti-Jo-1+ IIM sera than in the anti-Jo-1 - IIM sera[17]. Though the underlying mechanism is still under debate, these two MSAs/MAAs might share regulatory pathways for IFN-α production and pro-myositis microenvironment[18]. Cluster 4 comprises anti-EJ and anti-Ro-52 dual positivity anti-MDA5 IIM patients with a reportedly high ILD occurrence (75%) and 100% muscle enzyme elevation. Despite the similar ILD comorbidity rate, Cluster 4 is distinct from Cluster 1(anti-MDA5+ CADM) in its reportedly higher muscle enzyme elevation rate (100% versus 39%) and muscular manifestations (Myalgia, 38% versus 11%). The evidence of co-expression of anti-EJ/Ro-52 MSAs/MAAs in IIM is still scarce. Anti-EJ is the second-most frequently occurring autoantibody in ASyS[19]. Anti-EJ-associated ILD has also been recognized as an ASyS phenotype[20] that fits the diagnosis criteria of variable connective tissue disease[21], mostly IIM(82.4%)[22]. As a result, Cluster 4 seemingly represents a distinct phenotype of ILD-IIM patients with muscular manifestations that do not account for anti-MDA5 MSA but for anti-EJ autoantibodies.

Anti-MDA5 antibodies, initially identified by Sato et al.[23], are linked to dermatomyositis, especially CADM. These antibodies react with the MDA5/IFIH1 protein, which is vital in the body's defense against viral RNA. While primarily found in CADM patients, these antibodies are also observed in traditional dermatomyositis cases. A key characteristic of patients with anti-MDA5 antibodies is their increased risk of developing rapidly progressive interstitial lung disease (RP-ILD)[24]. Diverse prevalence and clinical outcomes of these antibodies existed across different ethnic groups. For example, a comparison between Chinese and Japanese patients revealed a higher occurrence of anti-MDA5 antibodies in the Chinese cohort[25].

The IIM patients with anti-MDA5+ in our cohort were more prone to have ILD and less muscular involvement. Moreover, these anti-MDA5+ IIM patients had poorer overall survival than those who were anti-MDA5 negative, which is compatible with a recent Chinese cohort[1]. These features were in accordance with the well-characterized anti-MDA5+ CADM patients[24-26].

The development of RP-ILD in IIM patients contributes to the high mortality rates. Besides, specific MSAs/MAAs reactivities were associated with the prognosis of ILD in IIM patients. Specifically, IIM patients with solo Ro-52 positive or Ro-52/MDA5 dual reactivity have a higher risk of RP-ILD and poorer overall survival[27].

On the other hand, the higher probability of ILD (Odds ratio [OR] 8.338(1.704-40.804)) in anti-EJ and anti-Ro-52 double-positive patients had never been reported before. Besides, we reported less muscular presentation in these patients, which resembled anti-MDA5+ CADM. Despite being more likely to have ILD, our anti-EJ and anti-Ro-52 double-positive IIM patients did not have poorer prognosis than anti-MDA5+ patients. Furthermore, we also found a higher probability of ILD (4.645(1.501-14.376)) and myalgia (3.200(1.041-9.835)) in anti-Jo-1 and anti-Ro-52 double-positive patients. The high occurrence of muscular symptoms made this subgroup less similar to the well-characterized anti-MDA5+CADM patients. Of note, the survival analysis result of our anti-Jo-1 and anti-Ro-52 double-positive patients was also insignificant. These results suggested that not all IIM patients with ILD had a poor prognosis. Therefore, individualized therapy should be delivered according to the proper subclassification process for these patients.

Anti-TIF1-γ antibodies, identified as targeting the TIF1 family of nuclear transcription factors, primarily TIF1-γ, are significant in both juvenile and adult dermatomyositis[24, 25]. These antibodies, found in 20–30% of patients, are particularly specific to dermatomyositis and are rarely seen in other conditions like polymyositis. In juvenile cases, they are associated with chronic or polycyclic disease courses, while in adults, there is a strong correlation with malignancy, especially in those over 40 years of age. In a Japanese adult cohort, 65% of patients with these antibodies had malignancy, with a higher prevalence in older patients.

Moreover, anti-TIF1-γ antibodies are significantly linked to aggressive skin lesions in both adult and juvenile patients. Unique cutaneous manifestations include palmar hyperkeratotic papules and psoriasis-like lesions. However, they show a negative association with interstitial lung disease and features like Raynaud phenomenon and arthritis/arthralgia. Myositis symptoms tend to be mild, although there's an increased risk for dysphagia. This highlights the diagnostic importance of anti-TIF1-γ antibodies in dermatomyositis, especially for assessing adult cancer risk and understanding

disease progression in juveniles [28, 29]. In fact, our results supported the higher probability of cancer in the TIF1-γ+ IIM subgroup. However, no significant survival difference existed between anti-TIF1-γ+ and anti-TIF1-γ- patients.

Immune-mediated necrotizing myopathy (IMNM) mainly manifests with severe muscular symptoms[30]. Necrotic muscle fibers and C5b–9 (the membrane attack complex (MAC)) deposition on the sarcolemma are commonly seen in histopathology. IMNM patients with typical manifestations and autoantibodies (anti-HMGCR or anti-SRP autoantibodies) may not necessarily require muscle biopsy[31]; if the symptoms are atypical, antibodies are negative, or antibodies cannot be tested, muscle biopsy is one of the essential tools for diagnosing IMNM. In cluster 2, although the patients in this cluster presented heterogeneous data, a high proportion of muscular symptoms could be observed, including elevated muscle enzymes, proximal limb weakness, and dysphagia. In addition, the relatively high proportion of muscle biopsy-proven myositis could be attributed to the lack of anti-HMGCR antibodies in our MSAs/MAAs diagnostic kits.

Our cross-sectional study has a few limitations. First, given our purpose of incorporating the MSAs/MAAs positivity into a novel classification algorithm, selection bias existed due to excluding patients who did not receive an MSAs/MAAs examination. Second, our relatively small cohort (n=108) gave rise to a relatively poor PCA efficiency and difficulty in performing post-hoc analysis to test the inter-cluster significance of each clinical variable. Third, the unsupervised machine learning feature of CATPCA yielded unpredictable results, whose accuracy was hard to examine[32]. Fourth, we did not compare the performance of various hierarchical clustering methods, which probably led to entirely different phenotypic subgroups. Therefore, we need comparison group or validation cohort to test the reproducibility of the clustering results. And longitudinal data were need for long-term impact of different MSAs/MAAs profiles on patient outcomes.

In conclusion, aided by MSAs/MAAs profiles, we identified some unique IIM subgroups (Cluster 3 and 4) whose pathophysiology and prognosis deserve future exploration. On the other hand, we also successfully confirmed some conventional IIM phenotypes (Cluster 1 and 6) via the integration of clinical and molecular features. We also performed the survival analysis based on patients with specific MSAs/MAAs reactivities identified via our cluster dendrogram. Our cohort's higher ILD incidence and poorer overall survival for anti-MDA5 positive IIM patients echoed the well-characterized anti-MDA5 positive CAMD population.

Acknowledgments

The authors confirm their contribution to the paper as follows: study conception and design: C.Y.W, W.Y.S, J.H.Y, J.J.T; data collection: P.Y.W, T.T.O, C.C.W, C.C.T ; data curation: P.P.C, T.T.O, C.C.W, C.C.T; analysis and interpretation of results: C.Y.W, P.Y.W, W.Y.S, J.J.T, J.H.Y; draft manuscript preparation: C.Y.W, P.Y.W, W.Y.S, H.Y.H, J.J.T, J.H.Y. All authors reviewed the results and approved the final version of the manuscript.

References

1. Wen, L., et al., *Myositis-specific autoantibodies and their clinical associations in idiopathic inflammatory myopathies: results from a cohort from China*. Clin Rheumatol, 2022. **41**(11): p. 3419-3427.
2. To, F., et al., *Including myositis-specific autoantibodies improves performance of the idiopathic inflammatory myopathies classification criteria*. Rheumatology (Oxford), 2019. **58**(12): p. 2331-2333.
3. Bonroy, C., et al., *Positioning of myositis-specific and associated autoantibody (MSA/MAA) testing in disease criteria and routine diagnostic work-up*. J Transl Autoimmun, 2022. **5**: p. 100148.
4. Allenbach, Y., et al., *Different phenotypes in dermatomyositis associated with anti-MDA5 antibody: Study of 121 cases*. Neurology, 2020. **95**(1): p. e70-e78.
5. Zhu, H., et al., *Identification of 6 dermatomyositis subgroups using principal component analysis-based cluster analysis*. Int J Rheum Dis, 2019. **22**(8): p. 1383-1392.
6. Zhang, J., et al., *Identification of 4 subgroups in juvenile dermatomyositis by principal component analysis-based cluster analysis*. Clin Exp Rheumatol, 2022. **40**(2): p. 443-449.
7. Bohan, A. and J.B. Peter, *Polymyositis and dermatomyositis (first of two parts)*. N Engl J Med, 1975. **292**(7): p. 344-7.
8. Murtagh, F. and P. Legendre, *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* Journal of Classification, 2014. **31**(3): p. 274-295.
9. Caliński, T. and J. Harabasz, *A dendrite method for cluster analysis*. Communications in Statistics, 1974. **3**(1): p. 1-27.
10. Kaplan, E.L. and P. Meier, *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, 1958. **53**(282): p. 457-481.

11. Davidson-Pilon, C., *lifelines: survival analysis in Python*. The Journal of Open Source Software, 2019. **4**(40): p. 1317.
12. Bland, J.M. and D.G. Altman, The logrank test. BMJ, 2004. **328**(7447): p. 1073.
13. Kuo, C.F., et al., *Incidence, cancer risk and mortality of dermatomyositis and polymyositis in Taiwan: a nationwide population study*. Br J Dermatol, 2011. **165**(6): p. 1273-9.
14. Fiorentino, D., et al., *The mucocutaneous and systemic phenotype of dermatomyositis patients with antibodies to MDA5 (CADM-140): a retrospective study*. J Am Acad Dermatol, 2011. **65**(1): p. 25-34.
15. Wong, V.T., et al., *Myositis-specific autoantibodies and their clinical associations in idiopathic inflammatory myopathies*. Acta Neurol Scand, 2021. **143**(2): p. 131-139.
16. Witt, L.J., J.J. Curran, and M.E. Strek, *The Diagnosis and Treatment of Antisynthetase Syndrome*. Clin Pulm Med, 2016. **23**(5): p. 218-226.
17. Rutjes, S.A., et al., *Anti-Ro52 antibodies frequently co-occur with anti-Jo-1 antibodies in sera from patients with idiopathic inflammatory myopathy*. Clin Exp Immunol, 1997. **109**(1): p. 32-40.
18. Eloranta, M.L., et al., *A possible mechanism for endogenous activation of the type I interferon system in myositis patients with anti-Jo-1 or anti-Ro 52/anti-Ro 60 autoantibodies*. Arthritis Rheum, 2007. **56**(9): p. 3112-24.
19. Aggarwal, R., et al., *Patients with non-Jo-1 anti-tRNA-synthetase autoantibodies have worse survival than Jo-1 positive patients*. Ann Rheum Dis, 2014. **73**(1): p. 227-32.
20. Watanabe, K., et al., *Detection of antisynthetase syndrome in patients with idiopathic interstitial pneumonias*. Respir Med, 2011. **105**(8): p. 1238-47.
21. Marco, J.L. and B.F. Collins, *Clinical manifestations and treatment of antisynthetase syndrome*. Best Pract Res Clin Rheumatol, 2020. **34**(4): p. 101503.
22. Liu, Y., et al., *Clinical characteristics of patients with anti-EJ antisynthetase syndrome associated interstitial lung disease and literature review*. Respir Med, 2020. **165**: p. 105920.
23. Sato, S., et al., *Autoantibodies to a 140-kd polypeptide, CADM-140, in Japanese patients with clinically amyopathic dermatomyositis*. Arthritis Rheum, 2005. **52**(5): p. 1571-6.
24. Fujimoto, M., et al., *Recent advances in dermatomyositis-specific autoantibodies*. Curr Opin Rheumatol, 2016. **28**(6): p. 636-44.
25. Ghirardello, A., et al., *Myositis autoantibodies and clinical phenotypes*. Auto Immun Highlights, 2014. **5**(3): p. 69-75.
26. McPherson, M., et al., *Management of MDA-5 antibody positive clinically amyopathic dermatomyositis associated interstitial lung disease: A systematic review*. Semin Arthritis Rheum, 2022. **53**: p. 151959.
27. Gui, X., et al., *Anti-Ro52 antibodies are associated with the prognosis of adult idiopathic inflammatory myopathy-associated interstitial lung disease*. Rheumatology (Oxford), 2022. **61**(11): p. 4570-4578.
28. Izuka, S., et al., *Long-term risks of malignancy in myositis-specific antibody-positive idiopathic inflammatory myopathy*. Rheumatol Int, 2023. **43**(2): p. 335-343.
29. Fujimoto, M., et al., *Myositis-specific anti-155/140 autoantibodies target transcription intermediary factor 1 family proteins*. Arthritis Rheum, 2012. **64**(2): p. 513-22.
30. Allenbach, Y., et al., *Immune-mediated necrotizing myopathy: clinical features and pathogenesis*. Nat Rev Rheumatol, 2020. **16**(12): p. 689-701.
31. Watanabe, Y., et al., *Clinical features and prognosis in anti-SRP and anti-HMGCR necrotising myopathy*. J Neurol Neurosurg Psychiatry, 2016. **87**(10): p. 1038-44.
32. Prasad, S. and L.M. Bruce, *Limitations of Principal Components Analysis for Hyperspectral Target Recognition*. IEEE Geoscience and Remote Sensing Letters, 2008. **5**(4): p. 625-629.

臺灣特發性發炎性肌病變病患之階層式分群特徵

王品逸[1#] 王慶雲[2#] 宋婉瑜[1] 歐燦騰[1] 吳正欽[1] 曾家駿[1,3] 陳佩佩[1]
何昕諭[4] 蔡季君[1,3,5,6] 顏正賢[1,7*]

[1] 高雄醫學大學附設醫院 過敏免疫風濕內科
[2] 臺中榮民總醫院教學部
[3] 高雄醫學大學 醫學系 內科學
[4] 國立成功大學醫學院
[5] 高雄醫學大學附設醫院 熱帶疾病醫療暨防治中心
[6] 高雄醫學大學附設醫院 感染內科
[7] 高雄醫學大學 醫學院臨床醫學研究所
[#] 同等貢獻

目的：本研究旨在建立一個臨床血清病理分類系統，包括16種肌炎特異性抗體/肌炎相關抗體，以對台灣南部的特發性發炎性肌病變患者進行次分類，重點關注臨床肌炎特徵和肌炎特異性抗體/肌炎相關抗體。此外，該研究擴展到次分類組別進行生存分析。

方法：這個研究收集來自一個醫學中心及兩個附屬醫院從2002年到2022總共108名特發性發炎性肌病變的成人病患。使用包括人口統計學、疾病表現、實驗室檢查和16種肌炎特異性抗體/肌炎相關抗體的數據集，通過生物信息學工具（如類別主成分分析和階層式分群算法）對患者進行次分類。該方法旨在將複雜的臨床數據濃縮為必要組成部分，從而找出獨特的患者次族群。

結果：該研究通過整合從2002年到2022總共108名病患臨床和分子特徵，找出具有獨特病理生理學特徵的特發性發炎性肌病變次族群（第3和第4群）並確認了傳統的特發性發炎性肌病變表型（第1和第6群）。生存分析強調了抗MDA5陽性患者有更高間質性肺病(ILD)的發病率和更差的整體生存率，這與文獻上無肌變型皮肌炎(CADM)的特徵相呼應。

結論：肌炎特異性抗體/肌炎相關抗體配置文件與臨床和分子特徵的整合促進了不同特發性炎性肌病亞群的識別。其中一些亞群呈現獨特的病理生理學特徵，值得進一步研究。此外，基於特定肌炎特異性抗體/肌炎相關抗體反應性的生存分析確認了抗MDA5陽性炎性肌病患者的嚴重預後，與文獻中關於抗MDA5陽性CADM的報導一致。

關鍵詞：特發性發炎性肌病變；肌炎特異性抗體；主成分分析；間質性肺病；黑色素瘤分化相關蛋白質-5