Phenotypes of organ involvement in sarcoidosis

Jonas Christian Schupp[1], Sandra Freitag-Wolf[2], Elena Bargagli[3], Violeta Mihailović-Vučinić[4], Paola Rottoli[3], Aleksandar Grubanovic[1], Annegret Müller[1,5], Arne Jochens[2], Lukas Tittmann[6], Jasmin Schnerch[1,7], Carmela Olivieri[3], Annegret Fischer[8], Dragana Jovanovic[4], Snežana Filipovic[4], Jelica Videnovic-Ivanovic[4], Paul Bresser[9], René Jonkers[9], Kate O'Reilly[10], Ling-Pei Ho[11], Karoline I. Gaede[12,13], Peter Zabel[12,13], Anna Dubaniewicz[14], Ben Marshall[10], Robert Kieszko[15], Janusz Milanowski[15], Andreas Günther[16], Anette Weihrich[17], Martin Petrek[18,19], Vitezslav Kolek[18,19], Michael P. Keane[20,21], Sarah O'Beirne[20,21], Seamas Donnelly[20,21], Sigridur Olina Haraldsdottir[22], Kristin B. Jorundsdottir[22], Ulrich Costabel[23], Francesco Bonella[23], Benoît Wallaert[24], Christian Grah[25], Tatjana Peroš-Golubičić[26], Mauritio Luisetti[27,†], Zamir Kadija[27], Stefan Pabst[28], Christian Grohé[29], János Strausz[30], Martina Vašáková[31,32], Martina Sterclova[31,32], Ann Millar[33], Jiří Homolka[34], Alena Slováková[34], Yvonne Kendrick[11], Anjali Crawshaw[11], Wim Wuyts[35], Lisa Spencer[36], Michael Pfeifer[37], Dominique Valeyre[38], Venerino Poletti[39], Hubertus Wirtz[40], Antje Prasse[1,41], Stefan Schreiber[8,42], Michael Krawczak[2] and Joachim Müller-Quernheim[1]

@ERSpublications
Five new clinical phenotypes of sarcoidosis have been identified by analysing organ manifestations of 1932 patients http://ow.ly/UYLC30jpUkq

Cite this article as: Schupp JC, Freitag-Wolf S, Bargagli E, *et al*. Phenotypes of organ involvement in sarcoidosis. *Eur Respir J* 2018; 51: 1700991 [https://doi.org/10.1183/13993003.00991-2017].

ABSTRACT Sarcoidosis is a highly variable, systemic granulomatous disease of hitherto unknown aetiology. The GenPhenReSa (Genotype–Phenotype Relationship in Sarcoidosis) project represents a European multicentre study to investigate the influence of genotype on disease phenotypes in sarcoidosis.

The baseline phenotype module of GenPhenReSa comprised 2163 Caucasian patients with sarcoidosis who were phenotyped at 31 study centres according to a standardised protocol.

From this module, we found that patients with acute onset were mainly female, young and of Scadding type I or II. Female patients showed a significantly higher frequency of eye and skin involvement, and complained more of fatigue. Based on multidimensional correspondence analysis and subsequent cluster analysis, patients could be clearly stratified into five distinct, yet undescribed, subgroups according to predominant organ involvement: 1) abdominal organ involvement, 2) ocular–cardiac–cutaneous–central nervous system disease involvement, 3) musculoskeletal–cutaneous involvement, 4) pulmonary and intrathoracic lymph node involvement, and 5) extrapulmonary involvement.

These five new clinical phenotypes will be useful to recruit homogenous cohorts in future biomedical studies.

Published online Jan 25, 2018; republished Jan 30, 2018 with amendments to author forenames.

This article has supplementary material available from erj.ersjournals.com

Received: May 15 2017 | Accepted after revision: Oct 25 2017

Affiliations: [1]Dept of Pneumology, University Medical Centre, Faculty of Medicine, Freiburg, Germany. [2]Institute of Medical Informatics and Statistics, Christian-Albrechts-Universität, Kiel, Germany. [3]Respiratory Diseases and Lung Transplant Unit, AOUS – Medical and Surgical Sciences and Neuroscience Dept, Siena University, Siena, Italy. [4]University Hospital of Lung Diseases, Clinical Centre of Serbia, Belgrade, Serbia. [5]Dept of Cardiology, Pneumology, Angiology and Intensive Care Medicine, University Hospital, Aachen, Germany. [6]Popgen Biobank and Institute of Epidemiology, Christian-Albrechts-Universität, Kiel, Germany. [7]Dept of Haematology and Oncology, University Medical Centre, Freiburg, Germany. [8]Institute of Clinical Molecular Biology, Christian-Albrechts-Universität, Kiel, Germany. [9]Pulmonology Dept, Academic Medical Center Amsterdam, Amsterdam, The Netherlands. [10]University Hospital, Southampton, UK. [11]Oxford Sarcoidosis Service, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [12]Medical Hospital, Research Center Borstel, Borstel, Germany. [13]Leibniz Center for Medicine and Biosciences, Airway Research Center North, Member of the German Center for Lung Research, Borstel, Germany. [14]Dept of Pulmonology, Medical University of Gdansk, Gdansk, Poland. [15]Dept of Pneumonology, Oncology and Allergology, Medical University of Lublin, Lublin, Poland. [16]University Hospital, Giessen, Germany. [17]Private Practice, Neuruppin, Germany. [18]Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic. [19]University Hospital Olomouc, Olomouc, Czech Republic. [20]University College Dublin, Dublin, Ireland. [21]St Vincent's University Hospital, Dublin, Ireland. [22]University Hospital, Reykjavik, Iceland. [23]Ruhrlandklinik, Westdeutsches Lungenzentrum am Universitätsklinikum Essen gGmbH, Universitätsklinik Essen, Essen, Germany. [24]University Hospital, Lille, France. [25]Hospital Berlin-Havelhöhe, Berlin, Germany. [26]University Hospital Jordanovac, Zagreb, Croatia. [27]Pulmonology Unit, Foundation IRCCS Policlinico San Matteo, Pavia, Italy. [28]University Hospital, Bonn, Germany. [29]Evangelische Lungenklinik, Berlin, Germany. [30]National Koranyi Institute, Budapest, Hungary. [31]Thomayer Hospital, Prague, Czech Republic. [32]First Faculty of Medicine, Charles University, Prague, Czech Republic. [33]University Hospital, Bristol, UK. [34]Prague General Hospital, Charles University, Prague, Czech Republic. [35]University Hospital, Leuven, Belgium. [36]Aintree University Hospital, Liverpool, UK. [37]University Hospital, Regensburg, Germany. [38]Service de Pneumologie, Groupe Hospitalier Avicenne-Jean Verdier-René Muret, Bobigny, France. [39]Pulmonary Unit, Dept of Thoracic Diseases, Azienda USL Romagna, G.B. Morgagni–L. Pierantoni Hospital, Forlì, Italy. [40]Dept of Pneumology, University Hospital, Leipzig, Germany. [41]Dept of Pneumology, Hannover Medical School, Hannover, Germany. [42]Dept of Internal Medicine, Christian-Albrechts-Universität, Kiel, Germany.

Correspondence: Joachim Müller-Quernheim, Dept of Pneumology, University Medical Centre, Albert Ludwigs University, Killianstrasse 5, 79106 Freiburg, Germany. E-mail: joachim.mueller-quernheim@uniklinik-freiburg.de

Introduction

Sarcoidosis is a highly variable, systemic granulomatous disease of hitherto unknown aetiology. Even though the disease affects people worldwide, sarcoidosis cohorts are not homogeneous, and differ in terms of age, sex, ethnicity, type of onset and organ involvement [1–5]. The clinical course of sarcoidosis ranges from spontaneous resolution to disabling chronic disease, with lung transplantation as the last resort. More serious complications are pulmonary fibrosis, renal failure, cardiac involvement, neurosarcoidosis, defacing lupus pernio and loss of sight due to posterior uveitis. This phenotypic variability may reflect the difference in genetic background, environmental exposure profile and socioeconomic status of sarcoidosis patients [6–10]. Onset of the disease is either acute or subacute. Acute onset is characterised by fever, fatigue, weight loss and/or with the triad of erythema nodosum, bihilar lymphadenopathy and arthritis called Löfgren syndrome [11, 12], whereas subacute onset is typically nonspecific, with cough, dyspnoea and/or chest pain being the most common symptoms [13, 14].

Previous epidemiological surveys, such as the American ACCESS [2], the German–Swiss WATL [3] and the retrospective single-centre MUSC [15] studies, describe this clinical heterogeneity of sarcoidosis, and highlight an urgent need for further detailed clinical phenotypes to better understand the pathomechanisms of the disease and to plan future clinical therapeutic studies [16, 17]. The GenPhenReSa (Genotype–Phenotype Relationship in Sarcoidosis) project, a European multicentre study designed to investigate the influence of genotypes on disease phenotypes in sarcoidosis, addresses this issue by the detailed characterisation of over 2000 European sarcoidosis patients. Here, we present the results of the phenotype module of GenPhenReSa.

Methods

Subjects

31 study centres providing tertiary pulmonary care either as an academic or national centre (supplementary table S1) consecutively screened their prevalent sarcoidosis patients for the following

This study is registered at the German Clinical Trials Register with identifier number DRKS00000045.

Support statement: This study was supported by German Federal Ministry for Education and Research grant 01EY1103 and German Research Foundation grant MU 692/8-1 and Clusterlab XI EXC 306/2. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

inclusion criteria: 1) diagnosis of sarcoidosis according to the consensus statement of the American Thoracic Society, European Respiratory Society (ERS) and World Association of Sarcoidosis and Other Granulomatous Disorders [1], 2) a documented course of disease over at least 2 years prior to recruitment start, thereby providing all parameters of the phenotyping protocol, and 3) Caucasian parentage with a local ancestry over at least two generations to minimise confounding the planned genetic studies by population substructure. Individuals with granulomatous disease that could not be unequivocally diagnosed as sarcoidosis were excluded. Patients who met these criteria were asked for informed consent to participate in this study. In total, 2163 patients with sarcoidosis were included in the study over a period of 3 years.

As the definition of organ involvement in sarcoidosis varies widely, we chose to use the criteria of the ACCESS study [18]. Patients were phenotyped at the 31 study sites according to a standardised protocol that was both validated and field tested [13]. A custom-built software programme was used for data management. The phenotyping protocol included more than 200 clinical parameters covering all organ manifestations. To harmonise pulmonary function data, the ERS-modified European Community for Steel and Coal reference values were applied to the entire cohort [19, 20]. Chest radiographs were grouped according to Scadding [21] at the corresponding study centre. The observational study did not influence the clinical processes or any therapeutic decisions made by the attending physicians, who were solely responsible for the choice and initiation of treatment.

The GenPhenReSa study was aimed at generating the requested data *via* regular processes of patient care. Staff at all recruiting centres were trained in phenotyping their own patients either during a visit to Freiburg (Germany) or during a visit by the Freiburg team to their centres. Data management was undertaken by the Popgen biobank in Kiel (Germany), which is audited by the independent centre for data protection of the state of Schleswig-Holstein [22]. All phenotype data were checked for plausibility by the Freiburg team before logging the data in the Popgen database. 203 patients had to be excluded because of incomplete or contradictory data. A positive vote by the Ethics Committee of the University of Freiburg was obtained prior to the initiation of the study, which was also registered in the German Clinical Trials Register (identifier DRKS00000045).

Statistical analysis

Patient characteristics were summarised in the form of frequencies (categorical data) or means and standard deviations (continuous data). In case of missing values, frequencies are given for available data only. Intergroup differences were assessed for statistical significance using a Chi-squared test or Wilcoxon rank sum test, as appropriate. In an explorative manner, an unsupervised data mining method was used to identify new possible phenotype groups of patients according to organ involvement as follows. Multiple correspondence analysis (MCA) was applied to extract principal components, which were then subjected to an agglomerative hierarchical clustering (hierarchical clustering on principal components (HCPC)) analysis. MCA was chosen as it allows an unbiased analysis of patterns of relationships between several variables in a categorical data set. Ward's minimum variance method with Euclidean distance was used and the total inertia (the sum of between- and within-group variance) was calculated at each aggregating step [23]. The inertia gain is the increase in within-group variance when moving from one partition to the next. The optimal number of clusters Q is given when the increase of between-inertia between Q–1 and Q clusters is greater than that between Q and Q+1 clusters. The stability of the final cluster solution was assessed by bootstrapping (k=10 000) using Jaccard coefficients (JCs) [24]. Statistical analyses were performed using R [25] (packages FactoMineR [26] and fpc [27]). A p-value <0.05 was considered statistically significant without adjustment for multiple testing.

Results

Patient characteristics

2163 Caucasian patients with sarcoidosis were included in the study. Their mean±SD age was 47.0 ±12.1 years and 1290 (59.6%) were female. Noncaseating granulomas could be histologically confirmed in 1987 patients (94.6%) (table 1). A total of 327 sarcoidosis patients (15.1%) had at least one organ involvement other than lung or lymph nodes histologically confirmed. Lung involvement (n=1664 (92.6%)) was by far the most common type of organ involvement, followed by mediastinal and/or hilar lymphadenopathy (n=1355 (77.0%)), and skin (n=342 (16.1%)), eye (n=163 (7.8%)) and joint (n=158 (7.5%)) involvement. The most common presenting symptoms were fatigue (n=1112 (61.8%)), cough (n=873 (47.9%)) and dyspnoea (n=818 (44.4%)). The most frequent radiological type was Scadding I (n=764 (36.1%)); 347 of the sarcoidosis patients (16.4%) did not have sarcoidosis-associated chest radiography findings (Scadding type 0) and 83 (3.9%) had signs of lung fibrosis (Scadding type IV). Pulmonary function was only marginally impaired, on average, with forced expiratory volume in 1 s (FEV_1) being the most reduced measure (92.8% predicted) (table 1).

TABLE 1 Characteristics of the sarcoidosis patients

	All patients	Acute onset	Subacute onset	p-value[#]
Subjects n	2163	829	1286	
Age years	47.0±11.9	44.6±11.5	48.0±12.3	$<10^{-6}$
Male/female	40.4/59.6	35.5/64.5	43.3/56.7	<0.001
Height cm	169.0±10.0	168.5±9.9	169.3±10.0	<0.05
Weight kg	78.4±16.2	78.3±15.5	78.4±16.5	0.63[NS]
Smoking history				$<10^{-6}$
Never-smoker	71.2	76.6	67.6	
Ex-smoker	18.8	13.2	22.4	
Current smoker	10.0	10.3	10.0	
Pack-years	6.6±10.2	5.5±9.2	7.0±10.6	0.17[NS]
Chest radiograph type (Scadding)				$<10^{-6}$
0	16.4	21.4	13.3	
I	36.1	50.9	27.0	
II	33.0	21.4	40.6	
III	10.6	5.3	13.7	
IV	3.9	1.0	15.3	
TLC % pred	97.8±16.6	99.5±15.4	96.7±17.2	<0.001
FEV_1 % pred	92.8±20.4	96.5±17.9	90.7±21.6	$<10^{-6}$
Biopsy with granulomas	94.6	91.4	96.6	$<10^{-6}$
Pulmonary involvement (radiography)	75.5	67.7	80.5	$<10^{-6}$
Pulmonary involvement (histology)	73.5	79.4	70.0	<0.001
Bronchial involvement	70.6	78.4	65.8	$<10^{-5}$
Intrathoracic lymph nodes	77.0	75.7	78.0	0.27[NS]
Extrathoracic lymph nodes	11.3	6.2	15.9	$<10^{-6}$
Arthritis	9.6	15.5	5.6	$<10^{-6}$
Skin involvement	16.1	16.4	16.0	0.88[NS]
Eye involvement	7.8	6.9	8.3	0.27[NS]
CNS involvement	3.4	2.7	4.0	0.15[NS]
Lacrimal gland involvement	3.9	3.1	4.4	0.19[NS]
Cardiac involvement	3.2	1.5	4.4	<0.001
Hepatic involvement	4.9	2.5	6.6	$<10^{-4}$
Splenic involvement	3.9	1.7	5.4	$<10^{-4}$
Renal involvement	3.3	3.4	3.2	0.90[NS]
Musculoskeletal involvement	7.5	10.6	5.7	$<10^{-4}$
Gastrointestinal involvement	0.6	0.3	0.9	0.14[NS]
Genital involvement	0.2	0.1	0.3	0.93[NS]
Need for therapy	61.3	60.3	61.9	0.50[NS]

Data are presented as mean±SD or %, unless otherwise stated. TLC: total lung capacity; FEV_1: forced expiratory volume in 1 s; CNS: central nervous system; NS: nonsignificant. [#]: significance levels comparing acute *versus* subacute onset.

Sex and age differences

The characteristics of male and female sarcoidosis patients were found to differ considerably in several aspects. Thus, the age at inclusion in the study peaked much earlier in male patients (between the third and fourth decades) than in females (fifth decade) (figure 1 and supplementary table S2). Male patients had slightly worse lung function. Female patients predominantly presented with Scadding type I (37.6%); male patients predominantly presented with Scadding type II (39.7%). Female patients suffered significantly more frequently from eye (9.2% *versus* 5.6%; $p<0.01$), salivary gland (4.7% *versus* 2.6%; $p<0.05$) and skin (18.9% *versus* 11.9%; $p<10^{-4}$) involvement. Symptoms also showed a sex difference, in that females suffered significantly more often from fatigue (66.4% *versus* 55.0%; $p<10^{-5}$), arthralgia (54.7% *versus* 49.1%; $p<10^{-6}$) and chest pain (23.4% *versus* 18.6%; $p<0.05$), but less from fever or subfebrile temperatures (18.5% *versus* 23.0%; $p<0.05$).

Another important factor influencing the clinical phenotype of patients with sarcoidosis is age (figure 1). A female/male ratio of 1:1 was observed in patients aged ⩽40 years, whereas a female/male ratio of 2:1 was observed in patients aged >40 years. Younger patients presented predominantly with Scadding type I (n=318 (44.7%)), whereas higher frequencies of Scadding type III or IV were noted in older patients. Patients aged ⩽40 years were significantly more prone to eye, intrathoracic lymph node and bronchial involvement, but less prone to heart involvement (supplementary table S3). Patients aged ⩽40 years

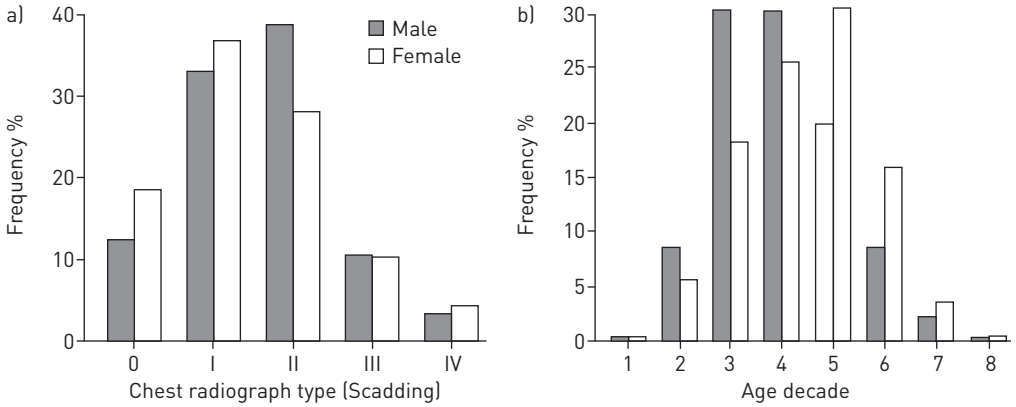

FIGURE 1 Frequency distributions of a) chest radiograph type according to Scadding [21] and b) age at baseline. The percentages are given separately for both sexes.

suffered significantly more from fever or subfebrile temperatures, but less from dyspnoea compared with patients aged >40 years.

Acute versus *subacute onset*

Patients with acute onset of disease compared with patients with subacute onset were significantly younger (44.6 *versus* 48.0 years; $p<10^{-6}$), female (64.5% *versus* 56.7%; p<0.001) (table 1) and had a slightly but significantly better lung function (*e.g.* FEV_1 % pred 96.5% *versus* 90.4%; $p<10^{-6}$). Patients with acute sarcoidosis more often had a radiologically normal lung (67.7% *versus* 80.5%; $p<10^{-6}$) (figure 2), but a higher percentage of histologically proven lung involvement (79.4% *versus* 70.0%; p<0.001). They more often had bronchial and musculoskeletal involvement, but less frequent cardiac, hepatic or splenic involvement (for frequencies and levels of significance, see table 1). Patients with acute sarcoidosis presented with a different spectrum of symptoms than patients with subacute sarcoidosis: they reported significantly more fatigue, fever, night sweats and arthralgia, but less cough and dyspnoea (supplementary table S4). No significant intergroup differences in terms of comorbidities were observed, including Inflammatory bowel disease: 12 (0.6%) of the sarcoidosis patients suffered from Crohn's disease and 10 (0.5%) had ulcerative colitis. For details, see supplementary table S4.

Treatment

1303 sarcoidosis patients (61.3%) were in need of treatment. Treatment regimens varied and were in the hands of the recruiting centres. Of the drug-treated patients, 92.7% received prednisolone or equivalents, 22.8% methotrexate, 11.2% azathioprine, 2.5% chloroquine, 1.3% infliximab, 0.8% cyclophosphamide and 0.3% cyclosporine; none received thalidomide. Reasons to start medication as stated by the treating physicians included clinical symptoms of cough and/or dyspnoea in 56.3%, clinical symptoms of fatigue, fever or weight loss in 63.1%, extrapulmonary organ involvement in 28.4%, and loss of pulmonary function in 23.5%.

Organ manifestation phenotypes

Organ involvement in sarcoidosis is characterised by a pronounced heterogeneity and seems to manifest randomly. In order to indentify underlying patterns of organ involvement, we performed MCA followed by HCPC analysis (figure 3). Patients with >50% missing data on organ involvement were excluded (supplementary table S5); the remaining 1932 patients were analysed in the MCA and HCPC analysis. The MCA led to the following main dimensions (supplementary table S6). The first and second dimensions, which explain 28% and 22% of the total variance, separated patients with arthritis and/or musculoskeletal involvement from patients with abdominal affection. The third dimension (19% explained variance) then segregated patients with of eye, skin, heart, central nervous system (CNS) or salivary gland involvement, with the fourth residual cluster of pulmonary–lymphonodal sarcoidosis patients. Subsequently, the HCPC identified five phenotypes (table 2 and see figure 4 for dendrogram), which are presented in detail, with frequencies always given for affected *versus* nonaffected patients:

1) Abdominal (renal/splenic/hepatic) involvement (n=133, JC 0.83). In patients with liver involvement, 36.7% also had spleen involvement (*versus* 2.1% without; $p<10^{-6}$), 10% also had kidney involvement (*versus* 2.6% without; p<0.001) and more often intrathoracic lymph node involvement (87.5% *versus* 76.8%; p<0.05). These patients typically experienced weight loss (21.4% *versus* 7.8%; p<0.001) and night

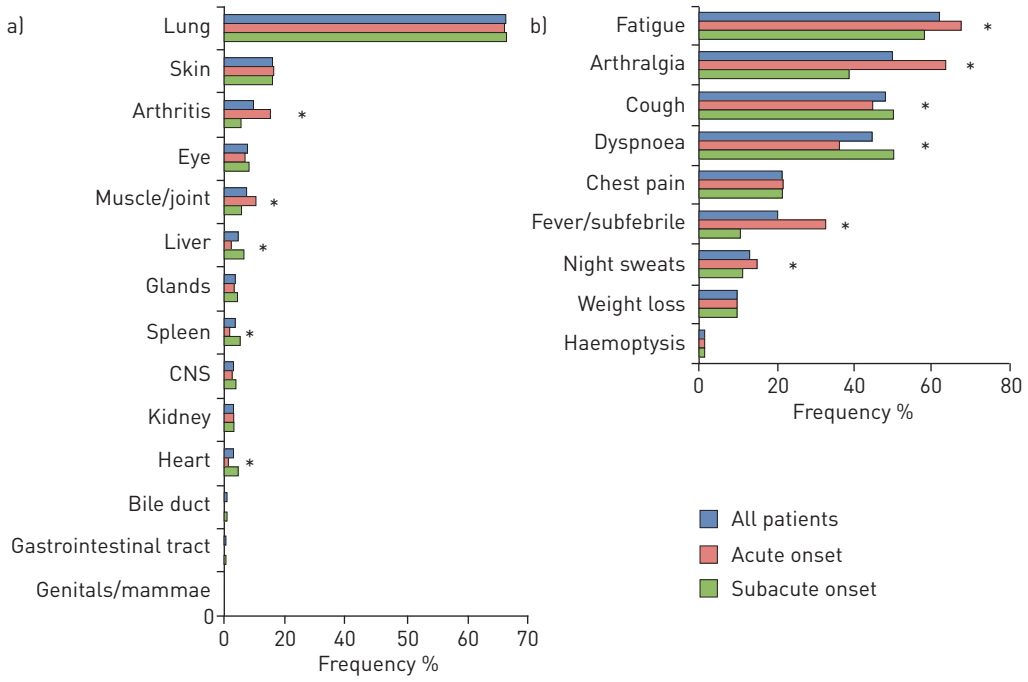

FIGURE 2 Frequency distributions of a) organ involvement and b) symptoms. CNS: central nervous system. *: p<0.05, comparing patients with acute *versus* subacute onset.

sweats (26.5% *versus* 11.4%; p<0.001). The most common organs involved in patients with splenic involvement were liver (45.2% *versus* 3.0%; $p<10^{-6}$) and kidney (11.0% *versus* 2.7%; p<0.001). The most prevalent symptom in patients with splenic and renal involvement was weight loss (19.3% *versus* 8.2%; p<0.01 and 22.5% *versus* 8.4%; p<0.01, respectively). A similar spectrum was observed for renal involvement. Patients with splenic, hepatic or renal involvement also presented significantly more often with impaired lung function, *e.g.* with a reduction of about –5% FEV1 % pred.

2) Ocular–cardiac–cutaneous–CNS ("OCCC"; eye/heart/skin/salivary glands/CNS) involvement (n=240, JC 0.82). Patients with ocular involvement more frequently showed skin (37.5% *versus* 14.2%; $p<10^{-6}$), CNS (13.8% *versus* 2.5%; $p<10^{-6}$) and heart (8.9% *versus* 2.7%; $p<10^{-4}$) sarcoidosis. Patients with neurosarcoidosis suffered more often from ocular (32.4% *versus* 7.1%; $p<10^{-6}$) and cardiac (11.6% *versus* 2.8%; p<0.001) involvement. An increased frequency of ocular (18.8% *versus* 6%; $p<10^{-6}$) and cardiac involvement (6.1% *versus* 2.5%; p<0.01) was observed in patients with skin sarcoidosis. Patients with eye or skin sarcoidosis also suffered more often from involvement of the salivary glands (both p<0.05; for eye: 11.3% *versus* 3.1%).

In cardiac, CNS and skin sarcoidosis, fatigue was more prevalent than in patients without these organ manifestations (all p<0.05). Interestingly, the rate of arthralgia was also increased in patients with eye, cardiac or skin involvement (65.4% *versus* 49.7%, 71.7% *versus* 50.5% and 64.6% *versus* 48.6%, respectively; all p<0.01). A significantly reduced frequency of fever or subfebrile temperature (12.2% *versus* 21.3%; p<0.05) was observed in patients with eye involvement.

3) Musculoskeletal–cutaneous involvement (n=189, JC 0.86). Patients with arthritis or musculoskeletal involvement presented significantly more often with an acute onset (both $p<10^{-6}$), and suffered more from fever or subfebrile temperature, night sweats, weight loss and arthralgia (all at least p<0.05) than patients without (for frequencies and levels of significance, see supplementary table S7). Patients with arthritis or musculoskeletal involvement less often showed involvement of the lungs or bronchi, but more often of the skin, intrathoracic lymph nodes and kidneys (all at least p<0.05). Patients with musculoskeletal sarcoidosis also suffered significantly more often from eye involvement ($p<10^{-3}$), whereas there was only a trend towards higher eye involvement in patients with arthritis. There was no significant difference between patients with or without arthritis or musculoskeletal involvement in terms of age, sex and lung function data.

4+5) Pulmonary–lymphonodal (n=1257, JC 0.99) and extrapulmonary sarcoidosis (n=113, JC 0.92). The remaining patients clustered in a group of no extrathoracic sarcoid involvement and a group of extrapulmonary sarcoidosis. Patients with lung involvement had worse lung function, and suffered more often from bronchial and intrathoracic lymph node involvement, but less often from skin or musculoskeletal involvement (all at least p<0.05; for frequencies and levels of significance, see

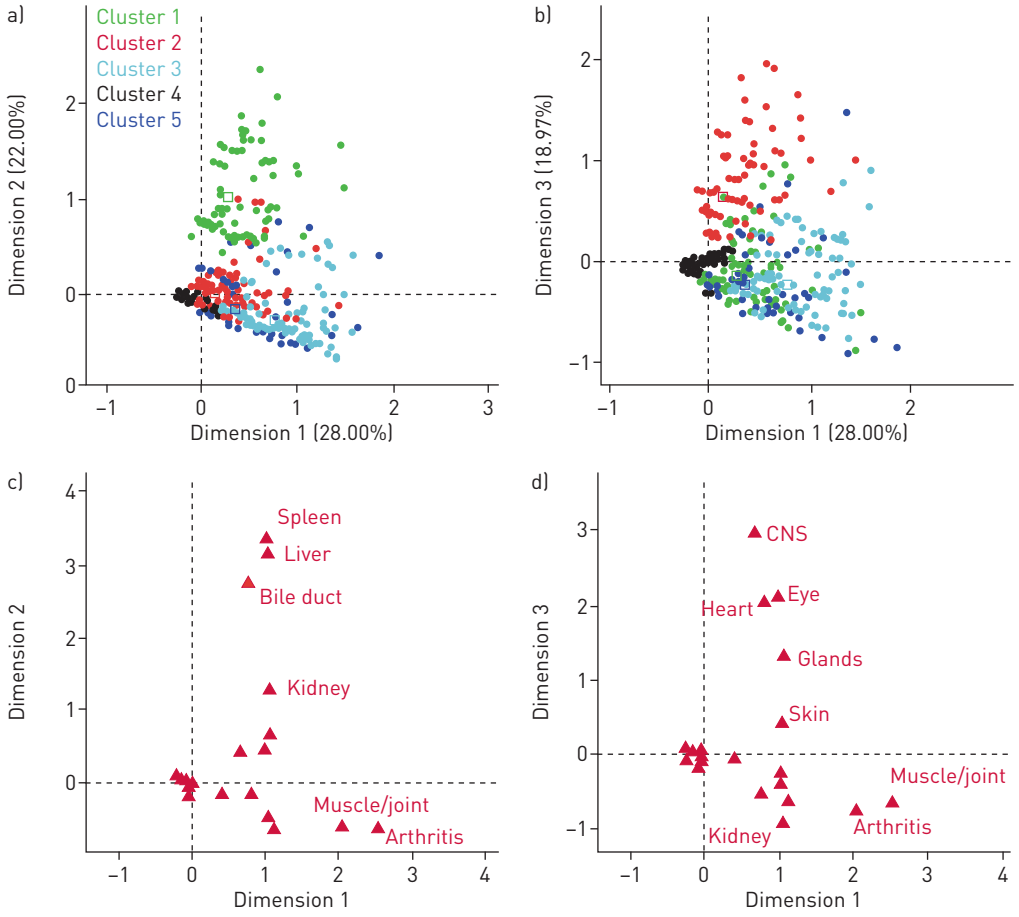

FIGURE 3 Multiple correspondence analysis (MCA) and clusters of organ involvement that allow analysis of patterns of relationships between categorical variables (different organ involvement). CNS: central nervous system. a, b) Scatter plots of patients based on different dimensions derived from MCA coloured by cluster membership with explained variance given in brackets: a) dimension 1 *versus* dimension 2 and b) dimension 1 *versus* dimension 3. Cluster 1: abdominal; cluster 2: ocular–cardio–cutaneous–CNS; cluster 3: musculoskeletal–cutaneous; cluster 4: pulmonary–lymphonodal; cluster 5: extrapulmonary. Data are presented as dots for individual patients; open squares represent cluster centres. c, d) Scatter plots of variables contributing to different dimensions: c) dimension 1 *versus* dimension 2 and d) dimension 1 *versus* dimension 3. Labelling of variables close to the zero-point is omitted due to space constraints.

supplementary table S8). They presented significantly more often with dyspnoea, cough, chest pain and fatigue (all at least $p<0.01$). Interestingly, the extrapulmonary phenotype had the highest frequencies for kidney involvement.

See table 2 for more details of the five new phenotypes.

To evaluate cluster stability, we performed bootstrapping using the JC values. Clusters with JC >0.75 are generally considered as valid, stable clusters, which all five clusters fulfil [27]. The pulmonary–lymphonodal, musculoskeletal–cutaneous and extrapulmonary clusters even had JC >0.85, indicative of high cluster stability.

Discussion

The GenPhenReSa study is a large-scale European multicentre sarcoidosis study of over 2000 patients that has documented patient characteristics in great detail. Phenotype analysis identified five novel distinct phenotype groups.

The sex distribution in our study showed a female predominance with a 3:2 female/male ratio, which is in line with the ACCESS study and a recent French study [2, 28]. However, in patients aged ⩽40 years or with subacute onset, the ratio was almost 1:1. Age at study inclusion peaked in females during the fifth decade and in males during the fourth decade. This later peak and the predominance of females were similarly observed in the ACCESS and MUSC cohorts, as well as in a recent Swedish study [2, 15, 29]. As

TABLE 2 Characteristics (organ involvement and symptoms) of the patients[#] clustered into the five phenotypes

	Abdominal	OCCC	Musculoskeletal–cutaneous	Pulmonary–lymphonodal	Extrapulmonary
Subjects n	133	240	189	1257	113
Acute onset	15.3	29.2	62.5	36.2	38.7
Need for therapy	63.2	75.5	57.1	68.8	33.3
CNS	4.0	26.7	2.2	0.0	0.0
Arthritis	11.1	1.9	89.6	0.0	12.7
Eye	3.2	55.1	12.2	0.0	6.5
Glands	8.9	23.8	7.3	0.0	2.7
Heart	5.4	20.8	2.3	0.0	6.6
Intrathoracic lymph nodes	12.6	26.4	12.8	11.1	10.1
Kidney	11.4	1.7	3.8	1.8	19.6
Liver	67.7	2.1	2.8	0.0	3.7
Lung	99.1	99.5	94.2	100.0	11.6
Muscle/joint	7.6	3.4	71.3	0.0	12.7
Skin	14.3	29.0	42.1	11.6	30.4
Spleen	57.6	0.0	2.2	0.0	2.7
Arthralgia	45.9	60.0	89.5	48.9	40.5
Cough	43.2	52.2	48.9	56.2	37.2
Dyspnoea	46.8	56.5	47.2	48.8	35.4
Fatigue	73.5	73.5	66.2	66.8	41.6
Fever/subfebrile	26.5	11.3	35.9	23.2	24.0
Night sweats	25.5	13.9	25.2	12.6	8.5
Weight loss	24.2	12.6	17.1	8.6	9.6
FEV_1 % pred	89.9±21.6	92.5±19.8	93.8±18.6	90.9±20.1	100.5±23.7
TLC % pred	94.1±16.9	96.5±17.3	99.2±16.0	96.7±16.5	102.2±17.7

Data are presented as % or mean±SD, unless otherwise stated. OCCC: ocular–cardiac–cutaneous–central nervous system (CNS); FEV_1: forced expiratory volume in 1 s; TLC: total lung capacity. [#]: n=1932.

shown in the current study, sarcoidosis is not only a disease of young adults; it is also frequently diagnosed in middle-aged and elderly patients.

Pulmonary manifestation is by far the most common organ involvement in sarcoidosis, as shown in our study and a number of previous studies [2, 3, 30–32]. Compared with the ACCESS study [2], the distribution of organ involvement differs slightly. In the ACCESS study, more eye (11.8% *versus* 8.1%) and hepatic (11.5% *versus* 4.5%), yet less renal (0.7% *versus* 3.1%) and musculoskeletal (0.9% *versus* 7.3%) involvement was recorded than in the current study. These differences may be attributable to ethnic differences (*e.g.* African-Americans suffer more often from ocular involvement), as proven in the ACCESS study [2]. The distribution of the chest radiography types (Scadding) was similar to the ACCESS and WATL study populations [2, 3]. The higher percentage of Scadding type 0 in our study is explained by the fact that many recruiting centres are tertiary care institutions and in many countries patients with extrathoracic sarcoidosis are referred to pulmonologists for exclusion of pulmonary involvement. Notably, the rate of inflammatory bowel disease (0.5% ulcerative colitis and 0.6% Crohn's disease) was ~5–10 times higher than the corresponding prevalence rates in Europe [33]. The simultaneous occurrence of sarcoidosis and inflammatory bowel disease might be due to overlapping genetic risk profiles [34]. In patients with subacute sarcoidosis, we found a slightly higher prevalence of cancer in the patients' history, which is in accordance with a previous study by Askling *et al.* [35] and corroborates the notion that chronic inflammatory diseases, such as sarcoidosis and Crohn's disease, cause an increase in the incidence of neoplasia [36].

Many previous empirical studies distinguished pulmonary from extrapulmonary sarcoidosis [7, 37–39]. Some single associations between two involved organs have been published, *e.g.* liver and spleen or CNS and eye [40, 41]. Utilising MCA, we were able to comprehensively analyse associations in organ involvement and to identify five distinct clinical phenotypes of organ involvement: an abdominal, an "OCCC", a musculoskeletal–cutaneous, a pulmonary–lymphonodal and an extrapulmonary phenotype. The classical Löfgren syndrome [12] is probably concealed behind the musculoskeletal–cutaneous phenotype, but we cannot distinguish between patients with arthritis, muscle, bone or joint sarcoidosis in our database. Skin involvement was part of the "OCCC" and musculoskeletal–cutaneous phenotypes. We

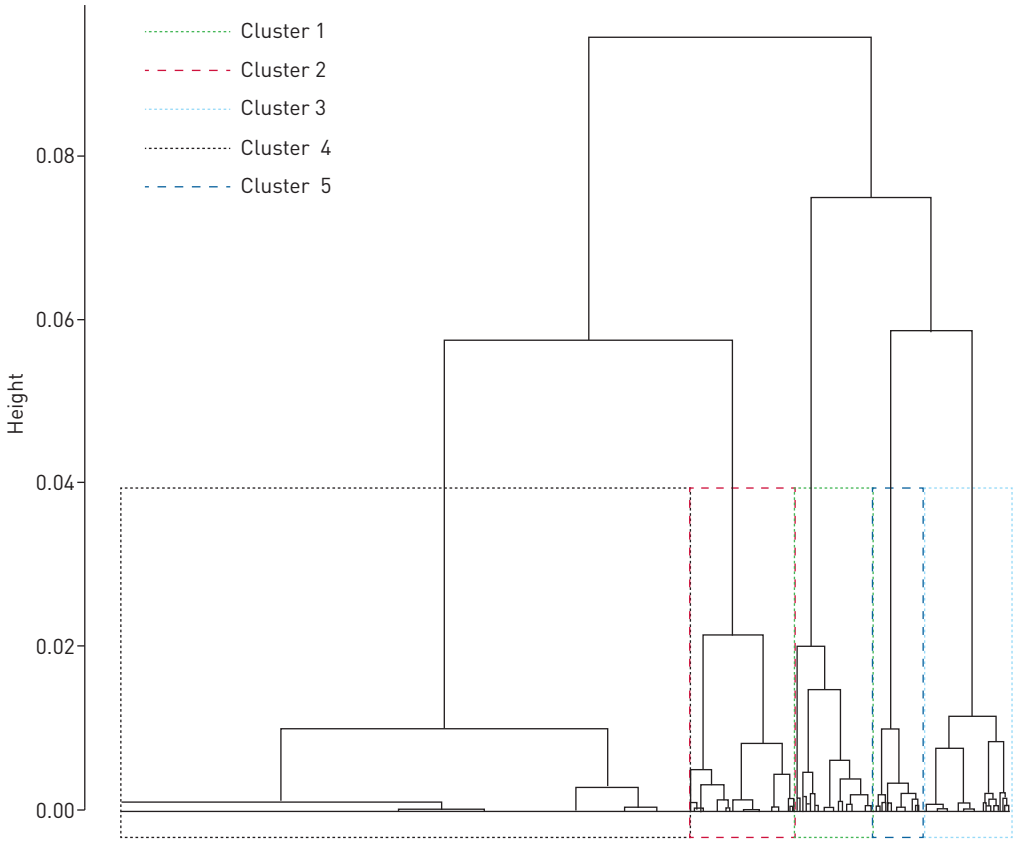

FIGURE 4 Dendrogram of the hierarchical clustering on principal components analysis leading to five clusters of organ involvement. Cluster 1: abdominal; cluster 2: ocular–cardio–cutaneous–central nervous system; cluster 3: musculoskeletal–cutaneous; cluster 4: pulmonary–lymphonodal; cluster 5: extrapulmonary. The height corresponds to the inertia gain.

have to assume that erythema nodosum predominantly arises in the musculoskeletal–cutaneous phenotype (classical Löfgren-like phenotype) and the other skin occurrences arise in the "OCCC" phenotype, but we have no data on skin involvement subtypes.

Cardiac involvement is part of the "OCCC" phenotype. A hypothetical conclusion might be that the clinically apparent cardiac sarcoidosis is mainly sarcoidosis of the electrical conduction system of the heart, as the conduction system is the structure, besides the myocardium itself, that is most often affected [42].

It is tempting to speculate that these five phenotypes are the sequelae of different aetiological agents causing sarcoidosis, which are either inhaled, ingested or sufficiently lipophilic to pass through the skin and/or the blood–brain barrier [7]. Furthermore, different genetic risk profiles might predispose to one of these five phenotypes. However, simple confounders such as access to diagnostic technologies might also lead to these clusters. Cluster analysis cannot deliver answers as to why these five phenotypes arise and thus further research is required.

This study also has several limitations. First, recruitment bias needs to be considered [43]. Investigators were mainly pulmonologists, so an overrepresentation of pulmonary sarcoidosis is likely. In addition, the centre recruiting the most patients was Belgrade (Serbia), resulting in an overrepresentation of Serbian patients. However, this subcohort (for basic characteristics, see supplementary table S9) did not shift the results of the MCA and did not influence the new phenotypes (supplementary figure S1). Second, all study centres were tertiary referral centres and hence an overrepresentation of patients with complicated, multiorgan sarcoidosis is possible. In this context, it has to be noted that an epidemiological study was not intended. Third, the assignment of the sarcoidosis patients to all variables was dependent on the use of a somewhat rigid phenotyping software tool, thus the full phenotypic spectrum of the disease "sarcoidosis" is potentially not covered in total, although over 200 traits have been documented. In addition, we phenotyped the patients according to the then up-to-date ACCESS criteria of organ involvement and methods of diagnosing them [18]. These criteria have since been updated [44] and new technologies will lead to better detection of organ involvement; therefore, frequencies and clusters of organ involvement will

change over time. Fourth, it is well known that the extent and frequency of organ involvement in sarcoidosis varies in different ethnicities [15]; our newly described phenotypes are therefore only applicable in Caucasian cohorts and might be different in, for example, African or Asian populations. Fifth, our new clinical phenotypes do not constitute clearly defined groups of patients, but more likely represent the extremes of a continuum, albeit of scientific usefulness. Sixth, disease duration might influence disease phenotype; therefore, we evaluated the influence of disease duration on MCA and found it to be negligible (supplementary figure S2).

In conclusion, the phenotype of sarcoidosis in patients of Caucasian descent is highly variable, and depends on age, sex and type of onset. We defined an abdominal, an "OCCC", a musculoskeletal–cutaneous, a pulmonary–lymphonodal and an extrapulmonary phenotype of organ affection. These new phenotypes can now be used in clinical and biomedical studies to obtain homogenous and clearly defined subcohorts of sarcoidosis patients. These results will be the basis of our efforts to link specific genotypes with phenotypes.

Acknowledgements

It is with great regret that we have to thank Philip Quanjer (Emeritus Professor of Physiology, Leiden University, The Netherlands) posthumously for his expertise on lung function data and his critical revision of the manuscript.

References

1 Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 1999; 160: 736–755.
2 Baughman RP, Teirstein AS, Judson MA, *et al.* Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med* 2001; 164: 1885–1889.
3 Loddenkemper R, Kloppenborg A, Schoenfeld N, *et al.* Clinical findings in 715 patients with newly detected pulmonary sarcoidosis – results of a cooperative study in former West Germany and Switzerland. WATL Study Group. Wissenschaftliche Arbeitsgemeinschaft für die Therapie von Lungenkrankheitan. *Sarcoidosis Vasc Diffuse Lung Dis* 1998; 15: 178–182.
4 Hosoda Y, Yamaguchi M, Hiraga Y. Global epidemiology of sarcoidosis. What story do prevalence and incidence tell us? *Clin Chest Med* 1997; 18: 681–694.
5 Morimoto T, Azuma A, Abe S, *et al.* Epidemiology of sarcoidosis in Japan. *Eur Respir J* 2008; 31: 372–379.
6 Müller-Quernheim J, Schürmann M, Hofmann S, *et al.* Genetics of sarcoidosis. *Clin Chest Med* 2008; 29: 391–414.
7 Kreider ME, Christie JD, Thompson B, *et al.* Relationship of environmental exposures to the clinical phenotype of sarcoidosis. *Chest* 2005; 128: 207–215.
8 Kajdasz DK, Judson MA, Mohr LC, *et al.* Geographic variation in sarcoidosis in South Carolina: its relation to socioeconomic status and health care indicators. *Am J Epidemiol* 1999; 150: 271–278.
9 Yigla M, Badarna-Abu-Ria N, Goralnik L, *et al.* Sarcoidosis in residents of northern Israel of Arabic and Jewish origin: a comparative study. *Respirology* 2006; 11: 586–591.
10 Anantham D, Ong SJ, Chuah KL, *et al.* Sarcoidosis in Singapore: epidemiology, clinical presentation and ethnic differences. *Respirology* 2007; 12: 355–360.
11 Consensus conference: activity of sarcoidosis. Third WASOG meeting, Los Angeles, USA, September 8–11, 1993. *Eur Respir J* 1994; 7: 624–627.
12 Löfgren S. Primary pulmonary sarcoidosis. I. *Acta Med Scand* 2009; 145: 424–431.
13 Prasse A, Katic C, Germann M, *et al.* Phenotyping sarcoidosis from a pulmonary perspective. *Am J Respir Crit Care Med* 2008; 177: 330–336.
14 Valeyre D, Prasse A, Nunes H, *et al.* Sarcoidosis. *Lancet* 2014; 383: 1155–1167.
15 Judson MA, Boan AD, Lackland DT. The clinical course of sarcoidosis: presentation, diagnosis, and treatment in a large white and black cohort in the United States. *Sarcoidosis Vasc Diffuse Lung Dis* 2012; 29: 119–127.
16 Pereira CAC, Dornfeld MC, Baughman R, *et al.* Clinical phenotypes in sarcoidosis. *Curr Opin Pulm Med* 2014; 20: 496–502.
17 Cozier YC. Assessing the worldwide epidemiology of sarcoidosis: challenges and future directions. *Eur Respir J* 2016; 48: 1545–1548.
18 Judson MA, Baughman RP, Teirstein AS, *et al.* Defining organ involvement in sarcoidosis: the ACCESS proposed instrument. ACCESS Research Group. A Case Control Etiologic Study of Sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 1999; 16: 75–86.
19 Stocks J, Quanjer PH. Reference values for residual volume, functional residual capacity and total lung capacity. ATS Workshop on Lung Volume Measurements. Official Statement of The European Respiratory Society. *Eur Respir J* 1995; 8: 492–506.
20 Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society *Eur Respir J Suppl* 1993; 16: 5–40.
21 Scadding JG. Prognosis of intrathoracic sarcoidosis in England. A review of 136 cases after five years' observation. *Br Med J* 1961; 2: 1165–1172.
22 Krawczak M, Nikolaus S, von Eberstein H, *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006; 9: 55–61.
23 Husson F, Josse J, Pagès J. Principal component methods – hierarchical clustering – partitional clustering: why would we need to choose for visualizing data? Technical Report. 2010. http://math.agrocampus-ouest.fr/infoglueDeliverLive Date last accessed: November 10, 2017.
24 Hennig C. Cluster-wise assessment of cluster stability. *Comput Stat Data Anal* 2007; 52: 258–271.

25 R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing, 2013.
26 Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw* 2008; 25: 1–18.
27 Hennig C. fpc: Flexible Procedures for Clustering. 2015. https://cran.r-project.org/web/packages/fpc/fpc.pdf Date last accessed: November 10, 2017.
28 Jamilloux Y, Maucort-Boulch D, Kerever S, *et al.* Sarcoidosis-related mortality in France: a multiple-cause-of-death analysis. *Eur Respir J* 2016; 48: 1700–1709.
29 Arkema EV, Grunewald J, Kullberg S, *et al.* Sarcoidosis incidence and prevalence: a nationwide register-based assessment in Sweden. *Eur Respir J* 2016; 48: 1690–1699.
30 Neville E, Walker AN, James DG. Prognostic factors predicting the outcome of sarcoidosis: an analysis of 818 patients. *Q J Med* 1983; 52: 525–533.
31 Henke CE, Henke G, Elveback LR, *et al.* The epidemiology of sarcoidosis in Rochester, Minnesota: a population-based study of incidence and survival. *Am J Epidemiol* 1986; 123: 840–845.
32 Pietinalho A, Ohmichi M, Hiraga Y, *et al.* The mode of presentation of sarcoidosis in Finland and Hokkaido, Japan. A comparative analysis of 571 Finnish and 686 Japanese patients. *Sarcoidosis Vasc Diffuse Lung Dis* 1996; 13: 159–166.
33 Trallori G, Palli D, Saieva C, *et al.* A population-based study of inflammatory bowel disease in Florence over 15 years (1978–92). *Scand J Gastroenterol* 1996; 31: 892–899.
34 Fischer A, Nothnagel M, Franke A, *et al.* Association of inflammatory bowel disease risk loci with sarcoidosis, and its acute and chronic subphenotypes. *Eur Respir J* 2011; 37: 610–616.
35 Askling J, Grunewald J, Eklund A, *et al.* Increased risk for cancer following sarcoidosis. *Am J Respir Crit Care Med* 1999; 160: 1668–1672.
36 Jess T, Horváth-Puhó E, Fallingborg J, *et al.* Cancer risk in inflammatory bowel disease according to patient phenotype and treatment: a Danish population-based cohort study. *Am J Gastroenterol* 2013; 108: 1869–1876.
37 Rybicki BA, Sinha R, Iyengar S, *et al.* Genetic linkage analysis of sarcoidosis phenotypes: the sarcoidosis genetic analysis (SAGA) study. *Genes Immun* 2007; 8: 379–386.
38 Judson MA, Hirst K, Iyengar SK, *et al.* Comparison of sarcoidosis phenotypes among affected African-American siblings. *Chest* 2006; 130: 855–862.
39 Reynolds HY. Sarcoidosis: impact of other illnesses on the presentation and management of multi-organ disease. *Lung* 2002; 180: 281–299.
40 Judson MA. Hepatic, splenic, and gastrointestinal involvement with sarcoidosis. *Semin Respir Crit Care Med* 2002; 23: 529–542.
41 Baughman RP, Weiss KL, Golnik KC. Neuro-ophthalmic sarcoidosis. *Eye Brain* 2012; 4: 13–25.
42 Roberts WC, McAllister HA, Ferrans VJ. Sarcoidosis of the heart. A clinicopathologic study of 35 necropsy patients (group 1) and review of 78 previously described necropsy patients (group 11). *Am J Med* 1977; 63: 86–108.
43 Judson MA, Thompson BW, Rabin DL, *et al.* The diagnostic pathway to sarcoidosis. *Chest* 2003; 123: 406–412.
44 Judson MA, Costabel U, Drent M, *et al.* The WASOG sarcoidosis organ assessment instrument: an update of a previous clinical tool. *Sarcoidosis Vasc Diffus Lung Dis* 2014; 31: 19–27.