



## EDITORIAL

# Genetic predisposition to sarcoidosis: another brick in the wall

Paolo Spagnolo\* and David A. Schwartz#

**T**he belief that a genetic susceptibility to the development of sarcoidosis exists is supported by several lines of evidence: 1) monozygotic twins are more often concordant for the disease than dizygotic twins; 2) sarcoidosis patients are more likely than healthy subjects to report a sibling or parent affected with the disease; 3) prevalence, incidence and severity of sarcoidosis vary widely amongst different races [1–4]. According to our current understanding of the disease pathophysiology, sarcoidosis is not due to defects in a single major gene or chemical pathway; instead, it is a complex disease that likely results from multiple genetic and environmental factors working together, each contributing a relatively small effect and few, if any, being absolutely required for the disease to occur. Genetics is also likely to contribute to the wide variety of clinical presentations and phenotypes observed in sarcoidosis. In this regard, some believe that sarcoidosis represents a family of diseases (sarcoidoses), including, among others: Löfgren syndrome, which is defined as the acute onset of fever, erythema nodosum, bilateral hilar lymphadenopathy and polyarthralgia; non-resolving/progressive lung disease; and granulomatous uveitis, each with potentially distinct genetic associations [5]. Berylliosis could also be considered as a subset of the broad grouping “sarcoidosis”.

Traditionally, genetic studies have used a “candidate gene case–control” approach, particularly in the context of rare diseases, like sarcoidosis, because of the difficulties in recruiting large numbers of pedigrees (linkage study) or even larger numbers of well phenotyped subjects (genome-wide association study; GWAS). In candidate gene case–control studies, the distribution of common genetic variations (single nucleotide polymorphisms; SNPs) in the gene(s) of interest is compared between unrelated, affected individuals and matched healthy controls. This hypothesis-based methodology requires understanding of the disease pathophysiology, judicious selection of candidate genes based on their biological plausibility, and knowledge of gene variations. Despite obvious limitations, this gene-hunting approach has enhanced our understanding of the genetic component of sarcoidosis by identifying a number of robust

associations, mostly with alleles located in the human leukocyte antigen area [6].

During the past few years, GWASs have revolutionised human genetics and led to the identification of thousands of loci that affect susceptibility to complex diseases [7, 8]. In just 5 years, the GWAS methodology has moved from extraordinary to commonplace. This hypothesis-free and unbiased approach is based on the data produced by the Human HapMap Project and the fact that genetic variance at one locus can predict with high probability genetic variance at adjacent loci, typically over distances of 30 000 base pairs of DNA [9]. Given its haplotypic structure, the human genome can be surveyed for common variants (those present in >5% of the population) associated with the risk of disease by simply genotyping approximately 500 000 accurately chosen markers, so-called tag SNPs [10]. Important insights from GWASs include identification of putative risk loci in or near genes not previously suspected of being involved in the pathogenesis of a given disease and associations with non-coding genomic regions. Conversely, GWASs identify loci and not sequence variants *per se*, and are unable to detect rare risk alleles. In addition, given the amount of data generated and the stringent threshold of statistical significance, GWASs require very large number of cases and controls. Table 1 summarises advantages and potential pitfalls of candidate gene case–control and GWASs.

The article by HOFMANN *et al.* [11] in this issue of the *European Respiratory Journal* adds another locus (and possibly a gene) to the (already) complex genetic architecture of sarcoidosis. These investigators performed a GWAS in a large cohort of German sarcoidosis patients and controls and identified a new sarcoidosis susceptibility locus at 12q13.3–q14.1. Fine-mapping (the genotyping of all known SNPs located in the genomic area surrounding the tag SNP) and sequencing (the reading one by one of the nucleotide bases in the DNA) of this genomic region pointed to rs1050045 in the 3′-untranslated region (3′ UTR) of *Osteosarcoma amplified 9 (OS9)* as the most likely candidate risk factor. This association has been validated in an independent German population and replicated by a meta-analysis of three independent cohorts of sarcoidosis patients from Germany, Czech Republic and Sweden. In addition, data analysis stratified by disease course revealed a stronger association of the lead SNP rs1050045 with acute sarcoidosis than with sarcoidosis as a whole. Given the stringent study methodology, consisting of a screening, validation and replication panel, this association is unlikely to be a spurious (false-positive) finding. Nevertheless,

\*Center for Rare Lung Diseases, Respiratory Disease Unit, Dept of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, Modena, Italy. #Dept of Medicine, University of Colorado Denver, Aurora, CO, USA.

CORRESPONDENCE: P. Spagnolo, University of Modena and Reggio Emilia, via del Pozzo 71, Modena, 41100, Italy. E-mail: paolo.spagnolo@unimore.it

**TABLE 1** Candidate gene case–control *versus* genome-wide association studies

	Candidate gene case–control studies	Genome-wide association studies
<b>Objectives</b>	Focus on a limited number of genes Look at functionally significant variants	Genome-wide evaluation of common variants
<b>Requirements</b>	Relatively small number of cases and controls Easily available genotyping technology	Large number of cases and controls Dense chip-based genotyping technology (SNP chip)
<b>Advantages</b>	Relatively inexpensive Less time-consuming May be the only way to study rare diseases	Permit "agnostic" genome-wide evaluation (prior information regarding gene function and gene variations are not required) Test a large number of variants (up to 1 million and more) Most robust associations have been with genes previously not suspected of association with a given disease May discover associations in regions not even known to harbor genes
<b>Limitations</b>	Biased (prior information regarding gene function and gene variations are required) Careful selection of candidate genes and variants to study Unable to discover new variations Can test only a limited number of variants Prone to spurious associations Require replication	Low statistical power (owing to the number of independent tests performed) Find loci not genes Require samples from a large number of cases and controls Expensive and difficult to organize Detect only alleles that are common (>5%) in a population Many associated variants are not causal Require replication

SNP: single-nucleotide polymorphism.

owing to their unbiased and hypothesis-free methodology, and by focusing almost exclusively on statistical evidence, GWAS tend to de-emphasise considerations on biological plausibility. The vast majority (>80%) of associated variants detected by GWAS reside in non-coding areas of the genomes (intergenic regions or introns) and have no established biological relevance. rs1050045 does not escape this generalisation being located in the 3' UTR. The implications of this finding are two-fold: on the one hand, the role of this variant and gene in sarcoidosis immuno-pathogenesis remain speculative, and on the other hand, intronic and intergenic regions may play a role in gene regulation [12]. The rs1050045 risk allele confers a relatively small risk elevation (OR 1.24 in the meta-analysis of the screening and validation stage and OR 1.14 in the meta-analysis of three independent populations of sarcoidosis patients from Germany, Czech Republic and Sweden), but small odds ratios do not discount the possibility that a given allele may be involved in a crucial disease pathway. Genetic contribution to sarcoidosis patho-biology is far more complicated than previously thought. As such, we should start thinking beyond the significance of genes/loci in isolation as the disease is more likely to result from a complex network of gene–gene and gene–environment interactions.

A number of variants have been convincingly associated with the risk of developing sarcoidosis [6, 13–20]. Yet, they confer relatively small increments in risk, thus confirming the multifactorial aetiology of the disease, and account for only a small proportion of familial clustering. A number of explanations for this missing heritability can be hypothesised, including: much larger numbers of common variants of smaller effect which have yet to be found; rarer variants (possibly with larger effects) or structural variants (insertion, deletion, duplication, translocation or inversion of segments of DNA), both of which are poorly

captured by current genotyping assays; low power to detect gene–gene interactions; and inadequate accounting for environmental factors [21]. The development of next generation sequencing technologies (whole genome, whole exome, targeted sequencing) is likely to rapidly increase the number of genetic variants associated with sarcoidosis [22]. Whole exome sequencing gives a full representation of all coding polymorphisms, and whole genome sequencing offers the advantage of capturing variations in coding and non-coding regions as well as identifying structural variants. Targeted sequencing captures both coding and non-coding variants in selected genes/loci, usually following-up a GWAS approach. Yet, these strategies cannot prove causation, which instead requires integration of genetic, gene expression and epigenetic data as well as functional (cellular and animal) studies. Nevertheless, while these research technologies are likely to be fruitful for the majority of complex diseases, they may not be sufficient in sarcoidosis for a number of reasons: the existence of a heterogeneous phenotype (ranging from indolent, self-limiting to progressive forms unresponsive to treatment) with potentially distinct genetic associations; the observation of clear ethnic-specific patterns of organ involvement, such as uveitis or cardiac involvement amongst Japanese patients, lupus pernio (a chronic rash consisting of papules and plaques usually found on the face) in Puerto Ricans, Löfgren syndrome (which is extraordinarily rare in Japan) in Scandinavians [23]; and the likelihood that sarcoidosis has more than one cause. In this latter case, a myriad of rare risk alleles could potentially be involved in disease pathogenesis by interacting with occupational or environmental factors (inorganic particulates and microbial antigens) [24]. This would also explain why a number of studies have obtained conflicting results. In addition, the results of GWAS are strongly influenced by the population studied.

Different associations have been reported in Caucasian and in African-American sarcoidosis patients, and further studies in non-Europeans are likely to reveal intriguing new findings [25].

GWASs have not explained as much of the genetic components of many diseases, including sarcoidosis, as was anticipated. As the power of the GWAS approach increases with access to larger datasets of more precisely defined phenotypes and as the methods to test for genetic associations expand to include copy number variants and rare alleles, more risk alleles and mechanisms worth exploring are likely to be identified. If this will be the case, we need to be ready to reconcile a far larger amount of genetic information with the (putative) immunopathogenesis of sarcoidosis. Disentangling the complex interaction between genetics and the environment in determining the variety of sarcoidosis phenotypes will then represent the next logical and challenging task. In this regard, it is imperative that meticulous databases of phenotypically well-defined patients continue to be constructed, as this will significantly reduce the number of subjects required to show meaningful genetic associations. In fact, relatively small studies based on accurate genotyping with exhaustively defined phenotype criteria are equally, if not more, able to detect the same effect as larger studies of a less stringent design. It is possible that genetics extends to determining not only overall susceptibility to sarcoidosis but also its distinct phenotypic routes, and that genes responsible for the development of the disease are different from those determining its wide spectrum of clinical manifestations (disease modifier genes). As such, it is essential that genetic data are always analysed according to clinical phenotype and not limited to a "generic" disease susceptibility. For the time being, the study by HOFMANN *et al.* [11] provides another brick in the wall of the genetics of sarcoidosis.

#### STATEMENT OF INTEREST

None declared.

#### REFERENCES

- Hunninghake GW, Costabel U, Ando M, *et al.* ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and other Granulomatous Disorders. *Sarcoidosis Vasc Diffuse Lung Dis* 1999; 16: 149–173.
- Spagnolo P, Luppi F, Roversi P, *et al.* Sarcoidosis: challenging diagnostic aspects of an old disease. *Am J Med* 2012; 125: 118–125.
- Rybicki BA, Iannuzzi MC, Frederick MM, *et al.* ACCESS Research Group. Familial aggregation of sarcoidosis – a case-control etiologic study of sarcoidosis (ACCESS). *Am J Respir Crit Care Med* 2001; 164: 2085–2091.
- Rybicki BA, Kirkey KL, Major M, *et al.* Familial risk ratio of sarcoidosis in African-American sibs and parents. *Am J Epidemiol* 2001; 153: 188–193.
- Spagnolo P, du Bois RM. Genetics of sarcoidosis. *Clin Dermatol* 2007; 25: 242–249.
- Grunewald J. Role of genetics in susceptibility and outcome of sarcoidosis. *Semin Respir Crit Care Med* 2010; 31: 380–389.
- Hindorff LA, Sethupathy P, Junkins HA, *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009; 106: 9362–9367.
- Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010; 363: 166–176.
- International HapMap Consortium. A haplomap type of the human genome. *Nature* 2005; 437: 1299–1320.
- Wang WY, Barratt BJ, Clayton DG, *et al.* Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; 6: 109–118.
- Hofmann S, Fischer A, Nothnagel M, *et al.* Genome-wide association analysis reveals 12q13.3–q14.1 as new risk locus for sarcoidosis. *Eur Respir J* 2013; 41: 888–900.
- Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; 360: 1759–1768.
- Sato H, Grutters JC, Pantelidis P, *et al.* HLA-DQB1\*0201: a marker for good prognosis in British and Dutch patients with sarcoidosis. *Am J Respir Cell Mol Biol* 2002; 27: 406–412.
- Rossmann MD, Thompson B, Frederick M, *et al.* ACCESS Group. HLA-DRB1\*1101: a significant risk factor for sarcoidosis in blacks and whites. *Am J Hum Genet* 2003; 73: 720–735.
- Valentonyte R, Hampe J, Huse K, *et al.* Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nat Genet* 2005; 37: 357–364.
- Rybicki BA, Walewski JL, Maliarik MJ, *et al.* The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. *Am J Hum Genet* 2005; 77: 491–499.
- Spagnolo P, Sato H, Grutters JC, *et al.* Analysis of BTNL2 genetic polymorphisms in British and Dutch patients with sarcoidosis. *Tissue Antigens* 2007; 70: 219–227.
- Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies ANXA11 as a new susceptibility gene for sarcoidosis. *Nat Genet* 2008; 40: 1103–1106.
- Franke A, Fischer A, Nothnagel M, *et al.* Genome-wide association analysis in sarcoidosis and Crohn's disease unravels a common susceptibility locus on 10p12.2. *Gastroenterology* 2008; 135: 1207–1215.
- Hofmann S, Fischer A, Till A, *et al.* A genome-wide association study reveals evidence of association with sarcoidosis at 6p12.1. *Eur Respir J* 2011; 38: 1127–1135.
- Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747–753.
- Yang IV, Schwartz DA. The next-generation of complex lung genetic studies. *Am J Respir Crit Care Med* 2012; 186: 1087–1094.
- Spagnolo P, Cullinan P, du Bois RM. Sarcoidosis. In: King TE, Schwarz MI, eds. *Interstitial Lung Disease*. 5th edn. Hamilton, B.C. Decker, 2011; pp. 433–498.
- Newman KL, Newman LS. Occupational causes of sarcoidosis. *Curr Opin Allergy Clin Immunol* 2012; 12: 145–150.
- Adrianto I, Lin CP, Hale JJ, *et al.* Genome-wide association study of African and European Americans implicates multiple shared and ethnic specific loci in sarcoidosis susceptibility. *PLoS One* 2012; 7: e43907.