

Epigenetics and sarcoidosis

Iain R. Konigsberg^{1,2}, Lisa A. Maier^{2,3,4} and Ivana V. Yang^{1,2,5}

¹Human Medical Genetics and Genomics Program, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ²Dept of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ³Dept of Medicine, National Jewish Health, Denver, CO, USA. ⁴Dept of Environmental and Occupational Health, Colorado School of Public Health, Aurora, CO, USA. ⁵Dept of Epidemiology, Colorado School of Public Health, Aurora, CO, USA.

Corresponding author: Iain R. Konigsberg (iain.konigsberg@cuanschutz.edu)



Shareable abstract (@ERSpublications)

In this review, we summarise recent findings relating to epigenetic modifications in lung disease with a focus on sarcoidosis, and propose guidelines and rationale for future epigenomic studies of sarcoidosis https://bit.ly/3d7ls1B

Cite this article as: Konigsberg IR, Maier LA, Yang IV. Epigenetics and sarcoidosis. *Eur Respir Rev* 2021; 30: 210076 [DOI: 10.1183/16000617.0076-2021].

Copyright ©The authors 2021

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 16 March 2021 Accepted: 6 April 2021

Abstract

Epigenetic modifications are emerging as important regulatory mechanisms of gene expression in lung disease, given that they are influenced by environmental exposures and genetic variants, and that they regulate immune and fibrotic processes. In this review, we introduce these concepts with a focus on the study of DNA methylation and histone modifications and discuss how they have been applied to lung disease, and how they can be applied to sarcoidosis. This information has implications for other exposure and immunologically mediated lung diseases, such as chronic beryllium disease, hypersensitivity pneumonitis, and asbestosis.

Introduction to epigenetics

Epigenetics refers to potentially heritable changes in gene expression that are not the result of changes to the underlying DNA sequence. Environmental exposures and stimuli influence epigenetic processes that in turn mediate gene expression, often by modifying chromatin and its 3-dimensional structure. As such, the epigenome (the sum of epigenetic modifications across the individual's genome) is influenced both by the environment and underlying genetics. Unlike mutations to DNA sequence, epigenetic marks can be dynamic and reversible, rendering them excellent targets for therapeutics, as well as biomarkers to track disease course and distinguish subgroups in heterogeneous disease. Interestingly they may also be inherited and pass on traits and exposures from parents or even grandparents to offspring. These observations have been difficult to explain owing to a global loss of the majority of known epigenetic marks present in parental chromosomes during primordial germ cell development and after fertilisation. However, potential mechanisms of transgenerational epigenetic inheritance are being unravelled [1]. One of the key recent observations is that transcription factors (TFs) may act as carriers of epigenetic information during germ cell and pre-implantation development [2]. Epigenetic marks have been associated with many biological processes such as ageing and dosage compensation between sexes, in which the non-coding RNA Xist is transcribed from the inactivated X chromosome and is required for stable silencing. Xist expression is shown to be regulated by both DNA methylation and histone modifications [3]. Three of the most studied epigenetic processes are DNA methylation, modification of histone tails, and non-coding RNAs such as microRNAs (miRNAs) and long non-coding RNAs. Given the recent review of miRNAs in sarcoidosis [4], we will mainly focus on DNA methylation and histone modifications in this review (figure 1). Definitions of the commonly used terms in epigenetic studies are given in table 1.

DNA methylation

DNA methylation refers to the addition of a methyl group to cytosine bases in DNA, which can alter accessibility of genetic loci by transcriptional machinery as well as recruit downstream proteins for chromatin remodelling. Methylation is often concentrated in gene bodies, repeat sequences and intergenic regions, with methylation in the promoters of housekeeping genes being rare to allow for gene



FIGURE 1 Epigenetic regulation of DNA accessibility. Histone acetyltransferases (HATs) and methyltransferases (HMTs) add acetyl (Ac) and methyl (Me) group specific residues on histone tails to modulate chromatin compaction. DNA methyltransferases (DNMTs) methylate CpG sites. Promoter methylation blocks binding of RNA polymerase (POL) and transcription factors to negatively regulate gene expression. HDAC: histone deacetylases.

transcription to occur [5]. Increased DNA methylation in the context of regulatory elements often indicates decreased accessibility to that locus by proteins and other molecules and leads to repression of transcription. However, recent studies have identified TFs that bind methylated DNA [6]. DNA methyltransferases (DNMTs) target CpG dinucleotides, which are loci characterised by a cytosine followed by a guanine in the DNA sequence, and catalyse the transfer of a methyl group to the 5th carbon of cytosine, forming 5-methylcytosine. Cytosine methylation rarely occurs in a non-CpG context, but is mainly restricted to stem cells and brain tissue [7]. Sites of epigenetic variability are often marked with CpG islands, conserved GC-rich regions of the genome with greater-than-average concentrations of CpGs, and more recently identified nearby elements, CpG island shores [8]. CpG islands are often found in the regulatory regions of genes such as cis promoters and more distant regulatory elements such as enhancers [9]. Demethylation of CpGs can occur either actively, through successive oxidation reactions catalysed by the ten-eleven translocase family of enzymes and eventual replacement with cytosine through base excision repair [10], or passively, in which newly synthesised DNA is not methylated by DNMTs. DNA methylation can be assayed genome wide through sodium bisulfite conversion. Sodium bisulfite causes deamination of cytosine to uracil (which is ultimately amplified as thymine), but occurs orders of magnitude more slowly in 5-methylcytosine [11]. Next-generation sequencing and array-based genotyping can then be performed, with thymines mapping to expected cytosine positions representing unmethylated

TABLE 1 Commonly used terms in epigenetic studies	
Term	Definition
CpG site	 DNA sequence where a cytosine nucleotide is followed by a guanine nucleotide CpG is shorthand for 5'—C—phosphate—G—3', that is, cytosine and guanine separated by only one phosphate group in the 5' to 3' direction The cytosine in CpG sites can be dynamically methylated and demethylated
CpG island (CGI)	A CpG-dense region of DNA that often overlaps regulatory sequences Methylation of CpG islands can dynamically regulate gene expression, with decreased methylation often associated with increased expression
Chromatin	DNA wrapped around histone core proteins for the purpose of tight packaging of DNA into a small volume to fit into the nucleus of a cell and protect the DNA structure and sequence
Differentially methylated position	A single CpG site that is statistically significantly differentially methylated among study groups
Differentially methylated region	A group of CpG sites across a region of DNA that is statistically significantly differentially methylated among study groups
DNA methylation	DNA methylation is a process by which methyl groups are added to the DNA molecule Methylation can change transcriptional activity of a DNA segment without changing the sequence
Enhancer	A region of DNA that binds to transcription factors, which then interact with promoter regions to upregulate gene expression
Epigenome	Collection of epigenetic marks throughout the genome; genome-wide distribution of transcriptional regulators believed to mediate the memory of past cellular events
Epigenome-wide association study (EWAS)	A study of the epigenome that looks for association between epigenetic variation and a high-level trait, such as disease status across individuals, or for association of a difference in organisation of a genomic regulator
Genome-wide association study (GWAS)	A study that looks for association between genetic variation and a high-level trait such as disease across individuals, typically scanning millions of genetic variants genome-wide for association signals
Histone modification	A histone modification is a covalent post-translational modification (PTM) to histone proteins The PTMs made to histones can impact gene expression by altering chromatin structure or recruiting histone modifiers
Promoter	A region of DNA upstream of gene transcription start sites bound by RNA polymerase and transcription factors to facilitate gene expression

sites. The most commonly used platform for studies of DNA methylation in human samples is the Infinium MethylationEPIC BeadChip that interrogates 850k CpGs [12] and Illumina has recently released the Infinium Mouse Methylation BeadChip with coverage for 285k CpGs in the mouse genome.

Histone modifications

Histone proteins form octamers around which DNA is wrapped to form nucleosomes [13]. Compaction of DNA into chromatin is dependent upon this process. Chromatin density can be dynamically altered through chemical modifications to the protruding tails of certain histones by histone-modifying enzymes. Histone tails can be modified in many ways, including methylation, acetylation, ubiquitylation, and SUMOylation (addition of small ubiquitin-like modifiers), as well as the removal of chemical groups that result from these processes [14]. Acetylation of histone tails is often associated with gene activation, as acetylation of lysine residues neutralises a positive charge that increases affinity between histones and DNA, resulting in decompaction of chromatin at that locus. A common acetylation mark, that of lysine 27 on the histone H3 (H3K27ac) is a mark of active enhancers and active promoters. Histone acetyltransferases acetylate histone tails, histone deacetylases (HDACs) remove acetyl groups from histone tails, and bromodomain proteins are chromatin readers that recognise and bind acetylated histones and play a key role in transmission of epigenetic memory across cell divisions and transcription regulation. Methylation is associated both with activation and repression of genes depending on the level of methylation at different residues in histone tails. For example, H3K4me1 is present at poised and active enhancers while H3K4me3 marks poised or active promoters. Similarly, histone methyltransferases add the methyl groups to histone tails while histone demethylases remove them. The Encyclopedia of DNA

Elements (ENCODE) and Roadmap Epigenomics projects have created maps of histone modifications in cell lines and primary cells/tissues, providing important public resources for the research community [15]. Histone modifications can be queried genome wide through chromatin immunoprecipitation with DNA-sequencing (ChIP-seq) [16]. ChIP-seq involves cross-linking of proteins of interest to genomic DNA, fragmentation, pulling down proteins of interest with antibodies, and purification and amplification of bound DNA, which is then sequenced to provide a genome-wide view of protein binding.

Recent epigenomic advances

There are many new technologies that enable epigenomic research. While ChIP-seq provides detailed information on histone modifications associated with different chromatin states, it is costly and time-consuming to perform multiple ChIP-seq experiments. Because of this, the field has moved to broad assessment of chromatin accessibility [17], with the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) more commonly being used. ATAC-Seq utilises a transposase which preferentially inserts in open chromatin to ligate sequencing adapters to these regions. Sequencing then produces "pile-ups" of reads aligning to open chromatin regions of the genome [18]. ATAC-seq has been optimised for many sample types, including frozen cells/tissues not amenable to more traditional ChIP-seq protocols [19].

The advent of single cell sequencing, in which nucleic acid is measured in thousands of suspended single cells from tissue, has been applied to epigenomics, with single-cell methylation, histone modification, and non-coding RNA sequencing methodologies. Single cell techniques allow for the quantification of thousands of cellular features in thousands of cells within the same experiment. Protocols have also been developed and are being implemented for large scale experiments to assess multiple -omes [20]. Another important technology in epigenomics is high-resolution contact mapping, which assays chromatin contacts genome wide [21]. In this assay, chromatin is crosslinked using formaldehyde, fragmented, and proximity ligated. The resulting fragments are amplified and sequenced, providing a map of chromatin contacts.

Epigenetics and the immune system

Epigenetic regulation is essential in the development and function of the immune system, including cellular differentiation. This is seen in the differentiation of naïve CD4⁺ helper T-cells (Th) into distinct Th subsets [22, 23]. Cell fate is determined by the presence of signalling molecules that can induce and repress expression of certain transcription factors, as their active transcription drives differentiation. For example, the cytokine interleukin (IL)-12 drives differentiation of Th1 cells, which are characterised by the active transcriptional state of the TF T-bet. Th2 differentiation, on the other hand, is induced by IL-4, with transcription of GATA-3 marking mature Th2 cells. These processes are regulated by DNA methylation and histone modifications [22]. Specifically, methylation of a DNaseI-hypersensitive region in the 3' end of IL-4 has been associated with Th1 differentiation, while demethylation of a proximal promoter and regulatory elements in the IL-4 gene results in Th2 differentiation [24, 25]. Th1/Th2 cytokine production is affected by methylation of the promoter of interferon (IFN)- γ , impacting TF binding [26, 27]. Similarly, the TF FOXP3, regulated by DNA methylation, is critical for differentiation of regulatory T-cells [23].

Exposures and epigenetics

Environmental exposures often result in epigenetic changes that can drive, modify or indicate disease. As such, epigenetic changes may mediate the effects of exposures on human health. Epigenetics has already been shown to play a significant role in the development, maintenance and responses of the respiratory system while studies linking exposure to air pollution, epigenetics and disease are emerging [28]. Tobacco smoke has long been known to be linked to decreased lung function and risk of COPD and lung cancer [29]. Exposure of airway and bronchial epithelial cells to cigarette smoke results in histone modifications associated with transcriptional repression and generalised genome-wide demethylation [30]. Exposure to air pollution has been linked to increased morbidity and mortality [31]. Inhalation of particulate matter from traffic-related air pollution and other byproducts produced in certain occupations have been associated with declining lung health and respiratory disease [32].

Introduction to sarcoidosis

Sarcoidosis is a granulomatous disease affecting an estimated ~60 out of 100000 Americans per year, with higher prevalence observed in African–Americans and women [33]. Outcomes are variable and associated with race, sex and lower income [34]. Women tend to present at older ages and with less severe disease as measured by the Scadding scale [35]. While sarcoidosis is a systemic disease, affecting almost any organ and multiple organs in the body, the most prominent and severe manifestations are related to lung, heart and neurologic involvement as they are associated with increased morbidity and mortality. Pulmonary sarcoidosis is present in 90% of affected individuals [36, 37] and can be asymptomatic; if symptoms manifest, coughing, wheezing, and difficulty breathing are common, as with many interstitial lung

diseases. Pulmonary fibrosis, which tends to occur later in disease, is the main cause of mortality in affected individuals. The course of disease is difficult to predict, with at least 25% of patients developing chronic or progressive disease requiring treatment [38] while acute disease often resolves within months without the need for treatment. Patients with chronic disease are older and more often male [35]. Treatment is generally initiated in symptomatic patients experiencing impaired organ involvement, quality of life or risk of complication and generally consists of corticosteroids such as prednisolone, or immunosuppressants such as methotrexate and azathioprine in steroid-intolerant patients and as steroid-sparing agents in long-term therapy for patients with chronic disease [39].

At the molecular level, sarcoidosis appears to have many similarities to other granulomatous diseases, including chronic beryllium diseases (CBD) [35]. The pathologic hallmark of sarcoidosis is the noncaseating granuloma, characterised by a cluster of immune cells with a core of macrophages, histiocytes and multinucleated giant cells surrounded by a layer of lymphocytes. A Th1 predominate immune response is noted in sarcoidosis, with increased IFN- γ and tumour necrosis factor (TNF)- α in sera, lymph nodes and bronchoalveolar lavage (BAL) [40]. Fibroblasts and collagen can be found surrounding and extending from granulomas and likely drive fibrosis of affected organs. Granulomas are thought to form in response to foreign substances, including certain antigens, organisms and foreign bodies; antigen presentation drives T-cell differentiation and promotes an inflammatory response, including macrophage recruitment, which in turn results in release of chemokines to attract cells to the site of inflammation. Ultimately, granulomatous inflammation is thought to be an immunological mechanism to isolate and sequester antigens/pathogens/ foreign substances in the body. Some granulomatous diseases, such as tuberculosis, are triggered by bacterial or fungal infection [41], while others, such as CBD and hypersensitivity pneumonitis, are driven by occupational and or environmental exposures [42].

Despite an ultimately unknown aetiology, studies have identified many genetic, environmental and host immunologic risk factors associated with sarcoidosis development, course and outcomes [43].

Genetic risk factors

Familial clustering of cases has long indicated a likely genetic component to disease. Family studies suggest that between one-third and two-thirds of the risk of disease is due to genetic variation [44, 45]. Early genetic studies focused on the human leukocyte antigen (HLA) region, which showed both protective and risk-increasing alleles associated with sarcoidosis and related phenotypes in a variety of populations [46–53]. The ACCESS study showed that the HLA-DRB1*1101 allele was associated with disease in both European and African–Americans [49]. Interestingly, this allele has since been shown to significantly interact with insecticide exposure at work to confer increased sarcoidosis risk [54]. A number of studies have now identified interactions between genetic variants and exposures in sarcoidosis [55, 56].

Multiple association studies have identified further genetic associations with disease outside of the HLA region. Variants have been identified in genes such as *ANXA11* [57–62], *BTNL2* [63, 64], *XAF1* [65] and *NOTCH4* [66]. The GenPhenReSa consortium genotyped over 19000 individuals on the Illumina Immunochip, and found novel associations mapping to the genes *ATXN2/SH2B3*, *IL12B*, *MANBA/NFKB1*, *FAM117B* and *IL23R* [67]. A recent genome-wide association study (GWAS) in a Japanese cohort identified signals in *CCL24*, *STYXL1-SRRM3* and *IL23R* [68]. Many of the aforementioned genes have roles in immune response. However, most of the identified variants in these studies have small effect sizes, explaining a small proportion of disease risk, and are population specific. Furthermore, many identified genetic variants are intronic or intergenic and require further functional studies to elucidate their role in disease.

Genetic architecture has also been shown to affect disease presentation. Löfgren syndrome, a distinct phenotype of sarcoidosis characterised by joint pain and erythema nodosum, has a favourable prognosis and often resolves within 2 years. A recent study comparing sarcoidosis patients with and without Löfgren syndrome identified multiple genetic loci which confer risk for one phenotype but not the other [69].

Environmental risk factors

Although many studies have focused on identifying causal antigens that may drive sarcoidosis, without a known and replicable antigenic stimulus, cell and animal models have been difficult to develop and often do not recapitulate key features of disease [70, 71]. Regardless, many associations between disease and environmental factors have been observed. Studies in multiple populations have observed variable incidence of disease by season with incidence peaking in winter months in the USA [72, 73], as well as geographical clustering, such as in North Eastern USA, suggesting a likely environmental cause or stressor [74]. The ACCESS study observed associations between sarcoidosis and agricultural employment,

insecticide exposure, and work environments with mould/mildew exposures [75]. Multiple studies have also linked disease to wood combustion, with firefighters displaying high rates of disease [76, 77] and wood stove and fireplace use [78], as well as silica exposure [79–81]. While these exposures may directly drive disease, they may also predispose individuals to develop disease by compromising the host immune response.

Studies have implicated a possible microbial antigen in disease [82–84], with multiple groups finding mycobacterial DNA and proteins in sarcoidosis tissue samples [82, 84]. Despite this, there is no evidence of active mycobacterial infection in sarcoidosis patients. A recent 16S sequencing study interrogated pulmonary microbiota in sarcoidosis BAL compared to other interstitial lung diseases [85] and found no significant differences in composition or diversity between groups.

Immunological risk factors

Aside from genetic variation in immune genes such as the HLA region, host immunity is altered in disease. Recent studies have shown that Th17.1 cells are the predominant $CD4^+$ T-cell subset in disease, originating from Th17 cells and expressing IFN- γ , and are associated with inflammation and autoimmunity [86]. Sarcoidosis BAL is heavily enriched for Th17.1 cells while Th17 and Th1 percentages did not differ from controls [87]. BROOS *et al.* [88] showed that in sarcoidosis BAL, 48% of Th cells are Th17.1s, 17% are Th1s, and that Th17.1 cells are also enriched in the mediastinal lymph nodes.

T-cells in disease have also been shown to have impaired function, consistent with cell exhaustion. CD4⁺ T-cells in peripheral blood show increased expression of PDCD1, a gene which encodes PD1. This upregulation was confirmed in BAL and blood at the protein level [89]. Regulatory T-cells have also shown a diminished ability to counteract granulomatous inflammation in sarcoidosis compared to other granulomatous diseases [90]. Furthermore, biased T-cell receptor variable region gene repertoires have been observed in disease [91, 92].

Evidence for epigenetic regulation in sarcoidosis

The advancement of epigenomics, along with associated technologies and methodologies, allows for novel methods to study complex diseases for which traditional methods have proven ineffective. This is especially true in the respiratory system, which has many endogenous and exogenous inputs and is the site of many exposures. Several features of the disease indicate that epigenetic mechanisms may drive sarcoidosis (figure 2). 1) Gene expression studies demonstrate significant changes in transcriptional profiles in sarcoidosis transbronchial biopsies [93] and peripheral blood cells [94, 95]. 2) Differentiation of naïve CD4⁺ T-cells into Th1, Th2, Th17 and T-regulatory lineages, critical in sarcoidosis, is regulated by TFs via DNA methylation and histone modifications [23]. 3) Other immune-mediated diseases implicate epigenetic regulation of immune function and immunopathogenesis. Specifically, studies of DNA methylation in lupus [96], rheumatoid arthritis [97, 98], and in CD4⁺ T-cells in juvenile idiopathic arthritis [99], revealed new genes and regulatory elements associated with disease and disease severity. 4) Genes that are important in epigenetic regulation, including HDAC and other histone and chromatin modifying genes are associated with sarcoidosis [93]. Demographic differences in sarcoidosis presentation also implicate possible underlying epigenetic mechanisms. The relevance of epigenetics in sarcoidosis is also supported by findings in other lung diseases. Differential methylation has been found in tuberculosis, with those infected patients with active disease (5–10% of patients) demonstrating a different transcriptomic profile than infected individuals without signs of disease [100]. While asthma and atopy are Th2 driven, they are also epigenetically regulated [101]. Another important feature common to epigenetics and sarcoidosis is the strong environmental influence. Finally, sarcoidosis has a variable disease course, with some individuals developing more progressive lung disease and even fibrosis while others do not. Another fibrotic lung disease, idiopathic pulmonary fibrosis, is associated with extensive DNA methylation and gene expression differences in lung tissue [102], suggesting that this is likely going to be the case for progressive forms of sarcoidosis.

Despite these strong lines of evidence for the potential of epigenetic regulation in sarcoidosis, investigations to date have been limited to studies of miRNAs [103–109], reviewed in [4], with one epigenome-wide association study (EWAS) of DNA methylation [110] and none interrogating histone modifications. However, we believe that epigenetic regulation of gene expression in sarcoidosis is a promising area of research that needs to be developed, and we use the Applications of Epigenomics to Sarcoidosis section to highlight a few potential directions.

In a 2019 study, our group investigated coupled differential methylation and expression in sarcoidosis as well as CBD [110]. In a comparison to BAL of nine progressive *versus* 15 remitting sarcoidosis patients,



FIGURE 2 Features common to sarcoidosis aetiology that can be mediated through epigenetic modifications, providing the rationale for the study of epigenetics in sarcoidosis. Th: T-helper cell.

we found 516 CpGs near 420 unique genes significantly associated with disease progression (false discovery rate-adjusted p<0.05 and >5% methylation change). Interestingly, many differentially methylated sites were within or near differentially expressed genes, many of which are involved in immunity (*HLA*, chemokines, *MSMB*, *KIR3DL2*). We also observed considerable directional overlap between sites in sarcoidosis and CBD [110]. KOTH *et al.* [94] observed a similar pattern in blood, with sarcoidosis transcriptomic profiles largely overlapping those of tuberculosis patients.

Applications of epigenomics to sarcoidosis

An increased awareness and understanding of epigenetics as a driver and modifier of lung disease has advanced the field. By characterising lung disease signatures of epigenetic modification, we can better study and elucidate the relationship between genetics, environment and disease, including disease pathogenesis (figure 3). We have long known that genetics and environment interact to drive many diseases, and the development of epigenetics is allowing us to explore this interaction in novel and exciting ways. The National Heart, Lung, and Blood Institute conducted a workshop to define next steps in sarcoidosis research [111] and epigenomics was defined as one of the priority areas in the application of -omics and systems biology to sarcoidosis research [112].

Endotyping

Complex diseases and their clinical phenotypes are influenced by multiple genetic and environmental components, as well as their interactions. In many cases, modifying contributory factors can impact the resultant phenotype. As such, variable genotypes and phenotypes are often observed in complex disease. This can make study difficult, as a disease may result from multiple inputs and pathways. For example, asthma has long been studied and approached as a single disease, even though phenotypes of reversible airway obstruction can be achieved through multiple pathways. Transcriptional profiling of asthma has revealed two major groups of patients: those with Th2 driven inflammation (Th2 high) and those without (Th2 low) [113]. Corticosteroid treatment has shown to be effective in Th2-high individuals but not Th2-low asthma [114], highlighting the importance of subclassifying diseases based on driving mechanisms.

Subtypes of sarcoidosis have already been defined based on clinical characteristics and computed tomography features [115]. A recent study involving RNA-sequencing of 209 sarcoidosis BAL samples



FIGURE 3 Conceptual overview of the proposed importance of epigenetics in sarcoidosis. Epigenetic marks influence expression of genes that regulate differentiation of T-cell subtypes, leading to pulmonary sarcoidosis, as well as other immune-mediated lung diseases, like chronic beryllium diseases and hypersensitivity pneumonitis. Epigenetic marks may be important in disease biology and severity, and represent novel biomarkers of disease as well as therapeutic targets. Th: T-helper cell; Treg: regulatory T-cell.

found expression of many genes associated with disease, as well as specific clinical features [116]. Weighted gene co-expression network analysis followed by K-means clustering of genes within modules of interest showed four potential endotypes of disease relating to: 1) hilar lymphadenopathy and acute lymphocytic inflammation; 2) extraocular organ involvement and phosphoinositide 3-kinase (PI3K) activation; 3) chronic disease; and 4) multiorgan involvement with increased immune response. Similar to transcriptional endotypes of disease, the field will be able to define epigenetic endotypes of sarcoidosis disease; this work has begun in asthma [117].

Biomarkers

Omics studies have identified multiple protein and gene expression signatures associated with sarcoidosis. Unsurprisingly, many of the molecules upregulated in disease relate to processes such as T-cell skewing, granulomatous processes, and macrophage and lymphocyte activity. Studies of protein quantification of exhaled breath condensate have shown multiple proteins associated with disease, such as TNF- α , TGF- β 1, neopterin and angiotensin converting enzyme (ACE) [118, 119]. Protein studies of blood serum have shown that SAA, ACE, lysozyme, chitotriosidase, CXCL9, CXCL10, CXCL11, CCL18, matrix metalloprotein (MMP)7, periostin and soluble IL-2 receptor (sIL-2R) levels are associated with disease [120–126], and that ACE and sIL-2R levels correlate with lung function increases in methotrexate treatment [127]. Chitotriosidase, ACE, sIL-2R, neopterin and TNF- α levels have also been shown to be increased and predictive of disease in BAL fluid, with sIL-2R levels being predictive of severity [128–131].

However, many current biomarkers are limited by lack of specificity as they are associated with broad immune activities upregulated in multiple diseases. To move past this issue, recent studies have derived signatures of disease consisting of multiplex signals from many molecules. In a proteomic screen, BHARGAVA *et al.* [132] observed 272 proteins significantly differentially abundant between sarcoidosis cases and controls. These proteins were enriched for both known (phagosome maturation and redox balance) and novel (integrin-linked kinase signalling) pathways in disease. CROUSER *et al.* [133] showed gross transcriptomic differences in lung tissue and identified novel dysregulated genes in disease, among them *MMP12* and *ADAMDEC1.* KOTH *et al.* [94] performed a similar analysis in blood and built a classification model predicting disease with 92% specificity and sensitivity. Another group used peripheral blood gene expression to build a 20-gene signature distinguishing sarcoidosis with and without complications that may affect treatment [134]. These protein and gene expression studies pave the way for incorporating epigenetics into biomarker studies of sarcoidosis, as has been done in lung cancer [135].

Therapeutic targets

Current sarcoidosis therapeutics suppress inflammation, in a relatively nonspecific manner, but have no effect on fibrosis. Epigenomic studies can implicate new molecules to target with therapeutics, which can be evaluated both *in vivo* and *in vitro*. The dynamic nature of epigenetic modifications provide

opportunities to experimentally modulate biological pathways, which has proven invaluable in disease study and treatment. Specifically, DNA methylation changes have been shown to drive tumour formation and malignant progression [136], and thus have established basic mechanisms for disease pathogenesis and targets for intervention in cancer. DNMT inhibitors have been approved for the treatment of myelo-dysplastic syndrome and are in clinical trials for treatment of solid tumours [137]. Interestingly, a recent study showed that DNMT and HDAC inhibitors induce cryptic transcription of thousands of treatment-induced non-annotated transcription start sites [138]. While currently available DNMT inhibitors lack specificity for gene(s) of interest, locus-specific therapies are currently being developed using genome editing tools based on the CRISPR/dCas9 system [139, 140]. Additionally, current US Food and Drug Administration approved and in development histone mark modifying drugs are effective in targeting specific gene loci and pathways [141] and treating lung cancer [142].

Epigenome-wide association studies

EWAS, based largely on the methodology of GWAS, compare the epigenomes between cases and controls (although quantitative traits can be interrogated as well) to identify statistically significant associations between certain marks and disease. However, interpretation of the results of these studies has been confounded by multiple issues and has prompted a set of guidelines for the second generation of EWAS [143]. Some of the challenges for studies of DNA methylation include accounting for the effect of genetic variants on methylation levels, distinguishing inherent changes in DNA methylation within a cell *versus* changes observed due to differences in cell proportions, and interpretation of DNA methylation data in the absence of other epigenetic data. Going forward, epigenetic studies in sarcoidosis should be designed addressing these principles. It will be important to integrate DNA methylation with genetic variants, chromatin accessibility, 3-dimensional genome information, and TF binding data to provide insight into the function of the epigenome.

Methylation quantitative trait loci studies

As methylation is one mechanism through which genetic variation can impact disease, genomic and methylomic data can be integrated to identify possible functional mechanisms driving genetic associations with disease. Genetic loci associated with methylation of specific sites, known as DNA methylation quantitative trait loci (meQTLs), are heavily enriched for *cis* interactions, with genetic variants often affecting methylation of proximal CpGs. meQTLs have been shown to be enriched for intronic regions and miRNA binding sites, with the associated CpG sites enriched for promoter and enhancer regions [144]. meQTLs also show co-localisation with expression quantitative loci, and PIERCE *et al.* [145] show that in a subset of these co-localised signals, expression or methylation is mediating the effect on the other. The same analytical principles can be used to identify associations of methylation and expression quantitative methylations in Puerto Rican children and adolescents with asthma. CpGs were found to be on average 378 kb from related genes and enriched in enhancers, indicating the ability of expression quantitative methylation studies to identify methylation-expression relationships relevant to disease.

Important considerations for future epigenomic studies of sarcoidosis

There are a myriad of practical concerns relating to undertaking epigenomic research. We attempt to address common considerations in designing these experiments, as well as highlight how these considerations may apply to sarcoidosis studies.

Tissue/cells of interest

Epigenetic marks are often cell- or tissue-specific, making it important to carefully consider the tissue of study. The use of samples such as blood and exhaled breath condensates is appropriate for biomarker studies, as an easily accessible tissue obtained noninvasively allows for screening of many individuals. However, to better understand molecular pathogenesis of disease, specific affected tissues and cell types should be interrogated. In studies of bulk tissue, epigenomic signals are derived from many cells. This risks confounding, especially in diseases with known alterations in cell proportions such as sarcoidosis. Multiple approaches have been implemented to account for cellular heterogeneity in epigenomic data. One approach is to perform statistical deconvolution of epigenomic profiles by relying on cell type specific features. Deconvolution approaches have been widely used in peripheral blood profiling studies [147], and more recently in complex tissues [148]. Reference-free methods overcome the issue of well-typed markers of disease, which are often difficult to establish in hard-to-access tissues such as the lung. The second approach is to adjust for cell proportion differences. The third approach is to isolate cell types of interest based on flow cytometry; the concern with this approach is that cell sorting may affect cellular epigenomic patterns. The recently developed single cell technologies provide the best solution to

identification of relevant cell populations albeit they have technical limitations and are currently cost prohibitive [149].

Sample size

Most previous -omic studies in sarcoidosis have been limited by small numbers. Generally, studies of >1000 subjects are adequately powered to detect most methylation differences in peripheral blood [150]. Required sample sizes will be smaller for target tissues such as BAL. While we have demonstrated that EWAS methodologies have sufficient power to detect gross differences in DNA methylation in BAL between progressive and remitting cases of disease (n=24), in order to detect stable and replicable signatures that can be used for purposes such as biomarkers and endotyping, larger numbers of individuals are necessary. Sarcoidosis research consortia such as ACCESS and GRADS have pooled samples from multiple centres, providing greater numbers to be leveraged for large-scale -omic research. Additionally, signals derived from a population with varied geography, genetics and exposures are more likely to provide generalisable and validated results across additional populations that may differ in these factors.

Timing of sampling

Epigenetic marks are dynamic. Epigenomic assays at a single time-point only provide a snapshot of molecular events at that moment in time. It is thus important to quantify epigenetic marks at temporally relevant times in disease development. For a disease such as sarcoidosis, where investigators are interested in predicting disease course, longitudinal studies should be leveraged. By quantifying epigenomic marks at multiple time-points in disease course or treatment, we can hopefully add to the understanding of disease pathogenesis, and identify epigenomic signatures related to disease stage and predictive of progression.

Conclusions

Sarcoidosis is a complex disease driven by a combination of genetic, environmental and host immunologic factors. While previous studies have elucidated many of these risk factors, there are still many gaps in our knowledge of sarcoidosis development and progression. Epigenomic studies will provide a means of linking risk factors to disease pathobiology to better understand disease course, aid in antigen studies, and subclassify patients based on molecular profiling to better direct research and treatment.

Provenance: Submitted article, peer reviewed.

Conflict of interest: I.R. Konigsberg has nothing to disclose. L.A. Maier reports grants from National Institutes of Health, during the conduct of the study; and grants from National Institutes of Health, University of Cincinnati under a Mallinckrodt Foundation, MNK14344100 and ATYR1923-C-002, outside the submitted work. L.A. Maier is a member of the FSR Scientific Advisory Board. She does not receive any compensation for this activity. I.V. Yang reports grants from NIH, during the conduct of the study; and personal fees from Eleven P15, outside the submitted work. I.V. Yang has a patent "Circulating Biomarkers of Preclinical Pulmonary Fibrosis" pending.

References

- 1 Boskovic A, Rando OJ. Transgenerational epigenetic inheritance. Ann Rev Genet 2018; 52: 21–41.
- 2 Kremsky I, Corces VG. Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. *Genome Biol* 2020; 21: 118.
- 3 Kelsey AD, Yang C, Leung D, *et al.* Impact of flanking chromosomal sequences on localisation and silencing by the human non-coding RNA XIST. *Genome Biol* 2015; 16: 208.
- 4 Pattnaik B, Sryma PB, Mittal S, *et al.* MicroRNAs in pulmonary sarcoidosis: a systematic review. *Respir Investig* 2020; 58: 232–238.
- 5 Chen ZX, Riggs AD. DNA methylation and demethylation in mammals. J Biol Chem 2011; 286: 18347–18353.
- 6 Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* 2016; 17: 551–565.
- 7 Pinney SE. Mammalian non-CpG methylation: stem cells and beyond. Biology (Basel) 2014; 3: 739–751.
- 8 Irizarry RA, Ladd-Acosta C, Wen B, *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009; 41: 178–186.
- 9 Hnisz D, Abraham BJ, Lee TI, *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 2013; 155: 934–947.
- **10** Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* 2012; 13: 7–13.
- 11 Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 1992; 89: 1827–1831.

- 12 Pidsley R, Zotenko E, Peters TJ, *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 2016; 17: 208.
- 13 Huang RC, Bonner J. Histone, a suppressor of chromosomal RNA synthesis. *Proc Natl Acad Sci USA* 1962; 48: 1216–1222.
- 14 Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011; 21: 381–395.
- 15 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; 518: 317–330.
- 16 O'Geen H, Echipare L, Farnham PJ. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol Biol* 2011; 791: 265–286.
- 17 Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet 2014; 15: 709–721.
- 18 Buenrostro JD, Wu B, Chang HY, *et al.* ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 2015; 109: 21.29.1–21.29.9.
- 19 Corces MR, Trevino AE, Hamilton EG, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 2017; 14: 959–962.
- 20 Chappell L, Russell AJC, Voet T. Single-cell (multi)omics technologies. Ann Rev Genom Hum Genet 2018; 19: 15–41.
- 21 Di Pierro M, Cheng RR, Lieberman Aiden E, *et al. De novo* prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci USA* 2017; 114: 12126–12131.
- 22 Russ BE, Prier JE, Rao S, *et al.* T cell immunity as a tool for studying epigenetic regulation of cellular differentiation. *Front Genet* 2013; 4: 218.
- 23 Huehn J, Polansky JK, Hamann A. Epigenetic control of FOXP3 expression: the key to a stable regulatory T-cell lineage? *Nat Rev Immunol* 2009; 9: 83–89.
- 24 Tykocinski LO, Hajkova P, Chang HD, *et al.* A critical control element for interleukin-4 memory expression in T helper lymphocytes. *J Biol Chem* 2005; 280: 28177–28185.
- 25 Lee DU, Agarwal S, Rao A. Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene. *Immunity* 2002; 16: 649–660.
- 26 Young HA, Ghosh P, Ye J, *et al.* Differentiation of the T helper phenotypes by analysis of the methylation state of the IFN-gamma gene. *J Immunol* 1994; 153: 3603–3610.
- 27 Jones B, Chen J. Inhibition of IFN-gamma transcription by site-specific methylation during T helper cell development. *EMBO J* 2006; 25: 2443–2452.
- 28 Yang IV, Lozupone CA, Schwartz DA. The environment, epigenome, and asthma. J Allergy Clin Immunol 2017; 140: 14–23.
- 29 National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General. Atlanta, Centers for Disease Control and Prevention (US), 2014.
- **30** Liu F, Killian JK, Yang M, *et al.* Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene* 2010; 29: 3650–3664.
- **31** Samek L. Overall human mortality and morbidity due to exposure to air pollution. *Int J Occup Med Environ Health* 2016; 29: 417–426.
- 32 Brunekreef B, Beelen R, Hoek G, *et al.* Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: the NLCS-AIR study. *Res Rep Health Eff Inst* 2009: 5–71.
- 33 Baughman RP, Field S, Costabel U, *et al.* Sarcoidosis in America. Analysis based on health care use. *Ann Am Thorac Soc* 2016; 13: 1244–1252.
- 34 Harper LJ, Gerke AK, Wang XF, et al. Income and other contributors to poor outcomes in U.S. patients with sarcoidosis. Am J Respir Crit Care Med 2020; 201: 955–964.
- 35 Schupp JC, Freitag-Wolf S, Bargagli E, *et al.* Phenotypes of organ involvement in sarcoidosis. *Eur Respir J* 2018; 51: 1700991.
- 36 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.
- 37 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.
- 38 Gerke AK, Judson MA, Cozier YC, *et al.* Disease burden and variability in sarcoidosis. *Ann Am Thorac Soc* 2017; 14: Suppl. 6, S421–S428.
- **39** James WE, Baughman R. Treatment of sarcoidosis: grading the evidence. *Expert Rev Clin Pharmacol* 2018; 11: 677–687.
- 40 Asano M, Minagawa T, Ohmichi M, *et al.* Detection of endogenous cytokines in sera or in lymph nodes obtained from patients with sarcoidosis. *Clin Exp Immunol* 1991; 84: 92–96.
- **41** Gideon HP, Phuah J, Myers AJ, *et al.* Variability in tuberculosis granuloma T cell responses exists, but a balance of pro- and anti-inflammatory cytokines is associated with sterilisation. *PLoS Pathogens* 2015; 11: e1004603.

- **42** Samuel G, Maier LA. Immunology of chronic beryllium disease. *Curr Opin Allergy Clin Immunol* 2008; 8: 126–134.
- 43 Moller DR, Rybicki BA, Hamzeh NY, *et al.* Genetic, immunologic, and environmental basis of sarcoidosis. *Ann Am Thorac Soc* 2017; 14: Suppl. 6, S429–S436.
- 44 Sverrild A, Backer V, Kyvik KO, et al. Heredity in sarcoidosis: a registry-based twin study. Thorax 2008; 63: 894–896.
- 45 Rossides M, Grunewald J, Eklund A, *et al.* Familial aggregation and heritability of sarcoidosis: a Swedish nested case-control study. *Eur Respir J* 2018; 52: 1800385.
- 46 Abe S, Yamaguchi E, Makimura S, *et al.* Association of HLA-DR with sarcoidosis. Correlation with clinical course. *Chest* 1987; 92: 488–490.
- 47 Hedfors E, Lindstrom F. HLA-B8/DR3 in sarcoidosis. Correlation to acute onset disease with arthritis. *Tissue Antigens* 1983; 22: 200–203.
- 48 Krause A, Goebel KM. Class II MHC antigen (HLA-DR3) predisposes to sarcoid arthritis. *J Clin Lab Immunol* 1987; 24: 25–27.
- 49 Rossman MD, Thompson B, Frederick M, et al. HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites. Am J Hum Genet 2003; 73: 720–735.
- 50 Bogunia-Kubik K, Tomeczko J, Suchnicki K, *et al.* HLA-DRB1*03, DRB1*11 or DRB1*12 and their respective DRB3 specificities in clinical variants of sarcoidosis. *Tissue Antigens* 2001; 57: 87–90.
- 51 Foley PJ, McGrath DS, Puscinska E, *et al.* Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. *Am J Respir Cell Mol Biol* 2001; 25: 272–277.
- 52 Swider C, Schnittger L, Bogunia-Kubik K, *et al.* TNF-alpha and HLA-DR genotyping as potential prognostic markers in pulmonary sarcoidosis. *Eur Cytokine Netw* 1999; 10: 143–146.
- 53 Levin AM, Adrianto I, Datta I, *et al.* Association of HLA-DRB1 with sarcoidosis susceptibility and progression in African Americans. *Am J Respir Cell Mol Biol* 2015; 53: 206–216.
- 54 Rossman MD, Thompson B, Frederick M, *et al.* HLA and environmental interactions in sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2008; 25: 125–132.
- 55 Rivera NV, Patasova K, Kullberg S, *et al.* A gene-environment interaction between smoking and gene polymorphisms provides a high risk of two subgroups of sarcoidosis. *Sci Rep* 2019; 9: 18633.
- 56 Li J, Yang J, Levin AM, *et al.* Efficient generalised least squares method for mixed population and family-based samples in genome-wide association studies. *Genet Epidemiol* 2014; 38: 430–438.
- 57 Hofmann S, Franke A, Fischer A, *et al.* Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nat Genet* 2008; 40: 1103–1106.
- 58 Feng X, Zang S, Yang Y, *et al.* Annexin A11 (ANXA11) gene polymorphisms are associated with sarcoidosis in a Han Chinese population: a case-control study. *BMJ Open* 2014; 4: e004466.
- 59 Levin AM, Iannuzzi MC, Montgomery CG, *et al.* Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun* 2013; 14: 13–18.
- 60 Li Y, Pabst S, Kubisch C, *et al.* First independent replication study confirms the strong genetic association of ANXA11 with sarcoidosis. *Thorax* 2010; 65: 939–940.
- **61** Mrazek F, Stahelova A, Kriegova E, *et al.* Functional variant ANXA11 R230C: true marker of protection and candidate disease modifier in sarcoidosis. *Genes Immun* 2011; 12: 490–494.
- 62 Sikorova K, Kishore A, Rapti A, *et al.* Association of TGF-beta3 and ANXA11 with pulmonary sarcoidosis in Greek population. *Expert Rev Respir Med* 2020; 14: 1065–1069.
- 63 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.
- 64 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.
- 65 Levin AM, Iannuzzi MC, Montgomery CG, *et al.* Admixture fine-mapping in African Americans implicates XAF1 as a possible sarcoidosis risk gene. *PLoS One* 2014; 9: e92646.
- 66 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.
- 67 Fischer A, Ellinghaus D, Nutsua M, *et al.* Identification of immune-relevant factors conferring sarcoidosis genetic risk. *Am J Respir Crit Care Med* 2015; 192: 727–736.
- 68 Meguro A, Ishihara M, Petrek M, *et al.* Genetic control of CCL24, POR, and IL23R contributes to the pathogenesis of sarcoidosis. *Commun Biol* 2020; 3: 465.
- 69 Rivera NV, Ronninger M, Shchetynsky K, *et al.* High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *Am J Respir Crit Care Med* 2016; 193: 1008–1022.
- 70 Locke LW, Schlesinger LS, Crouser ED. Current sarcoidosis models and the importance of focusing on the granuloma. *Front Immunol* 2020; 11: 1719.
- 71 Jeny F, Grutters JC. Experimental models of sarcoidosis: where are we now? *Curr Opin Pulm Med* 2020; 26: 554–561.

- 72 Demirkok SS, Basaranoglu M, Coker E, *et al.* Seasonality of the onset of symptoms, tuberculin test anergy and Kveim positive reaction in a large cohort of patients with sarcoidosis. *Respirology* 2007; 12: 591–593.
- 73 Ungprasert P, Crowson CS, Matteson EL. Seasonal variation in incidence of sarcoidosis: a population-based study, 1976–2013. *Thorax* 2016; 71: 1164–1166.
- 74 Dumas O, Abramovitz L, Wiley AS, *et al.* Epidemiology of sarcoidosis in a prospective cohort study of U.S. women. *Ann Am Thorac Soc* 2016; 13: 67–71.
- 75 Newman LS, Rose CS, Bresnitz EA, *et al.* A case control etiologic study of sarcoidosis: environmental and occupational risk factors. *Am J Respir Crit Care Med* 2004; 170: 1324–1330.
- 76 Kern DG, Neill MA, Wrenn DS, *et al.* Investigation of a unique time-space cluster of sarcoidosis in firefighters. *Am Rev Respir Dis* 1993; 148: 974–980.
- 77 Prezant DJ, Dhala A, Goldstein A, *et al.* The incidence, prevalence, and severity of sarcoidosis in New York City firefighters. *Chest* 1999; 116: 1183–1193.
- 78 Kajdasz DK, Lackland DT, Mohr LC, *et al.* A current assessment of rurally linked exposures as potential risk factors for sarcoidosis. *Ann Epidemiol* 2001; 11: 111–117.
- 79 Armbruster C, Dekan G, Hovorka A. Granulomatous pneumonitis and mediastinal lymphadenopathy due to photocopier toner dust. *Lancet* 1996; 348: 690.
- 80 Drent M, Bomans PH, Van Suylen RJ, et al. Association of man-made mineral fibre exposure and sarcoidlike granulomas. Respir Med 2000; 94: 815–820.
- 81 Graff P, Larsson J, Bryngelsson IL, *et al.* Sarcoidosis and silica dust exposure among men in Sweden: a case-control study. *BMJ Open* 2020; 10: e038926.
- 82 Drake WP, Pei Z, Pride DT, *et al.* Molecular analysis of sarcoidosis tissues for mycobacterium species DNA. *Emerg Infect Dis* 2002; 8: 1334–1341.
- 83 Oswald-Richter KA, Beachboard DC, Seeley EH, *et al.* Dual analysis for mycobacteria and propionibacteria in sarcoidosis BAL. *J Clin Immunol* 2012; 32: 1129–1140.
- 84 Song Z, Marzilli L, Greenlee BM, *et al.* Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis. *J Exp Med* 2005; 201: 755–767.
- 85 Becker A, Vella G, Galata V, *et al.* The composition of the pulmonary microbiota in sarcoidosis an observational study. *Respir Res* 2019; 20: 46.
- 86 Lexberg MH, Taubner A, Albrecht I, et al. IFN-gamma and IL-12 synergize to convert in vivo generated Th17 into Th1/Th17 cells. Eur J Immunol 2010; 40: 3017–3027.
- 87 Ramstein J, Broos CE, Simpson LJ, *et al.* IFN-gamma-producing T-helper 17.1 cells are increased in sarcoidosis and are more prevalent than T-helper type 1 cells. *Am J Respir Crit Care Med* 2016; 193: 1281–1291.
- 88 Broos CE, Koth LL, van Nimwegen M, *et al.* Increased T-helper 17.1 cells in sarcoidosis mediastinal lymph nodes. *Eur Respir J* 2018; 51: 1701124.
- 89 Braun NA, Celada LJ, Herazo-Maya JD, *et al.* Blockade of the programmed death-1 pathway restores sarcoidosis CD4(+) T-cell proliferative capacity. *Am J Respir Crit Care Med* 2014; 190: 560–571.
- **90** Taflin C, Miyara M, Nochy D, *et al.* FoxP3+ regulatory T cells suppress early stages of granuloma formation but have little impact on sarcoidosis lesions. *Am J Pathol* 2009; 174: 497–508.
- **91** Grunewald J, Janson CH, Eklund A, *et al.* Restricted V alpha 2.3 gene usage by CD4+ T lymphocytes in bronchoalveolar lavage fluid from sarcoidosis patients correlates with HLA-DR3. *Eur J Immunol* 1992; 22: 129–135.
- 92 Wiken M, Ostadkarampour M, Eklund A, *et al.* Antigen-specific multifunctional T-cells in sarcoidosis patients with Lofgren's syndrome. *Eur Respir J* 2012; 40: 110–121.
- 93 Lockstone HE, Sanderson S, Kulakova N, et al. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. Am J Respir Crit Care Med 2010; 181: 1367–1375.
- 94 Koth LL, Solberg OD, Peng JC, *et al.* Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med* 2011; 184: 1153–1163.
- 95 Su R, Li MM, Bhakta NR, *et al.* Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes. *Eur Respir J* 2014; 44: 985–993.
- **96** Jeffries MA, Dozmorov M, Tang Y, *et al.* Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus. *Epigenetics* 2011; 6: 593–601.
- **97** Liu Y, Aryee MJ, Padyukov L, *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 2013; 31: 142–147.
- 98 Whitaker JW, Shoemaker R, Boyle DL, *et al.* An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med* 2013; 5: 40.
- 99 Ellis JA, Munro JE, Chavez RA, *et al.* Genome-scale case-control analysis of CD4+ T-cell DNA methylation in juvenile idiopathic arthritis reveals potential targets involved in disease. *Clin Epigenet* 2012; 4: 20.
- 100 Mayer AS, Hamzeh N, Maier LA. Sarcoidosis and chronic beryllium disease: similarities and differences. Sem Respir Crit Care Med 2014; 35: 316–329.

- 101 Davidson EJ, Yang IV. Role of epigenetics in the development of childhood asthma. *Curr Opin Allergy Clin Immunol* 2018; 18: 132–138.
- **102** Yang IV, Pedersen BS, Rabinovich E, *et al.* Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 1263–1272.
- **103** Crouser ED, Julian MW, Crawford M, *et al.* Differential expression of microRNA and predicted targets in pulmonary sarcoidosis. *Biochem Biophys Res Commun* 2012; 417: 886–891.
- 104 Jazwa A, Kasper L, Bak M, *et al.* Differential inflammatory microRNA and cytokine expression in pulmonary sarcoidosis. *Arch Immunol Ther Exp (Warsz)* 2015; 63: 139–146.
- 105 Dyskova T, Fillerova R, Novosad T, *et al.* Correlation network analysis reveals relationships between microRNAs, transcription factor T-bet, and deregulated cytokine/chemokine-receptor network in pulmonary sarcoidosis. *Mediators Inflamm* 2015; 2015: 121378.
- **106** Kiszalkiewicz J, Piotrowski WJ, Pastuszak-Lewandoska D, *et al.* Altered miRNA expression in pulmonary sarcoidosis. *BMC Med Genet* 2016; 17: 2.
- 107 Novosadova E, Chabronova A, Kolek V, et al. The serum expression of selected miRNAs in pulmonary sarcoidosis with/without Lofgren's syndrome. *Mediators Inflamm* 2016; 2016: 1246129.
- 108 Kishore A, Navratilova Z, Kolek V, *et al.* Expression analysis of extracellular microRNA in bronchoalveolar lavage fluid from patients with pulmonary sarcoidosis. *Respirology* 2018; 23: 1166–1172.
- 109 Kachamakova-Trojanowska N, Jazwa-Kusior A, Szade K, *et al.* Molecular profiling of regulatory T cells in pulmonary sarcoidosis. *J Autoimmun* 2018; 94: 56–69.
- 110 Yang IV, Konigsberg I, MacPhail K, *et al.* DNA methylation changes in lung immune cells are associated with granulomatous lung disease. *Am J Respir Cell Mol Biol* 2019; 60: 96–105.
- 111 Maier LA, Crouser ED, Martin WJ 2nd, *et al.* Executive summary of the NHLBI workshop report: leveraging current scientific advancements to understand sarcoidosis variability and improve outcomes. *Ann Am Thorac Soc* 2017; 14: Suppl. 6, S415–S420.
- 112 Crouser ED, Fingerlin TE, Yang IV, *et al.* Application of "omics" and systems biology to sarcoidosis research. *Ann Am Thorac Soc* 2017; 14: Suppl. 6, S445–S451.
- **113** Fahy JV. Type 2 inflammation in asthma present in most, absent in many. *Nat Rev Immunol* 2015; 15: 57–65.
- 114 Poole A, Urbanek C, Eng C, *et al.* Dissecting childhood asthma with nasal transcriptomics distinguishes subphenotypes of disease. *J Allergy Clin Immunol* 2014; 133: 670–678.
- **115** Papiris SA, Georgakopoulos A, Papaioannou AI, *et al.* Emerging phenotypes of sarcoidosis based on 18F-FDG PET/CT: a hierarchical cluster analysis. *Expert Rev Respir Med* 2020; 14: 229–238.
- 116 Vukmirovic M, Yan X, Gibson KF, et al. Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. medRxiv 2020; preprint [https://doi.org/10.1101/2020.07.24.20161448].
- 117 Nicodemus-Johnson J, Myers RA, Sakabe NJ, *et al.* DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight* 2016; 1: e90151.
- 118 Ahmadzai H, Cameron B, Chui J, *et al.* Measurement of neopterin, TGF-beta1 and ACE in the exhaled breath condensate of patients with sarcoidosis. *J Breath Res* 2013; 7: 046003.
- 119 Rozy A, Czerniawska J, Stepniewska A, et al. Inflammatory markers in the exhaled breath condensate of patients with pulmonary sarcoidosis. J Physiol Pharmacol 2006; 57: Suppl. 4, 335–340.
- 120 Bons JA, Drent M, Bouwman FG, *et al.* Potential biomarkers for diagnosis of sarcoidosis using proteomics in serum. *Respir Med* 2007; 101: 1687–1695.
- 121 Lieberman J. Elevation of serum angiotensin-converting-enzyme (ACE) level in sarcoidosis. *Am J Med* 1975; 59: 365–372.
- 122 Selroos OB. Biochemical markers in sarcoidosis. Crit Rev Clin Lab Sci 1986; 24: 185–216.
- 123 Grosso S, Margollicci MA, Bargagli E, *et al.* Serum levels of chitotriosidase as a marker of disease activity and clinical stage in sarcoidosis. *Scand J Clin Lab Invest* 2004; 64: 57–62.
- 124 Su R, Nguyen ML, Agarwal MR, *et al.* Interferon-inducible chemokines reflect severity and progression in sarcoidosis. *Respir Res* 2013; 14: 121.
- 125 Isshiki T, Matsuyama H, Yamaguchi T, *et al.* Plasma matrix metalloproteinase 7, CC-chemokine ligand 18, and periostin as markers for pulmonary sarcoidosis. *Respir Investig* 2020; 58: 479–487.
- 126 Arger NK, Ho M, Woodruff PG, *et al.* Serum CXCL11 correlates with pulmonary outcomes and disease burden in sarcoidosis. *Respir Med* 2019; 152: 89–96.
- 127 Vorselaars AD, van Moorsel CH, Zanen P, *et al.* ACE and sIL-2R correlate with lung function improvement in sarcoidosis during methotrexate therapy. *Respir Med* 2015; 109: 279–285.
- **128** Bargagli E, Margollicci M, Perrone A, *et al.* Chitotriosidase analysis in bronchoalveolar lavage of patients with sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2007; 24: 59–64.
- 129 Rothkrantz-Kos S, van Dieijen-Visser MP, Mulder PG, *et al.* Potential usefulness of inflammatory markers to monitor respiratory functional impairment in sarcoidosis. *Clin Chem* 2003; 49: 1510–1517.
- 130 Prasse A, Katic C, Germann M, *et al.* Phenotyping sarcoidosis from a pulmonary perspective. *Am J Respir Crit Care Med* 2008; 177: 330–336.

- **131** Fehrenbach H, Zissel G, Goldmann T, *et al.* Alveolar macrophages are the main source for tumour necrosis factor-alpha in patients with sarcoidosis. *Eur Respir J* 2003; 21: 421–428.
- **132** Bhargava M, Viken KJ, Barkes B, *et al.* Novel protein pathways in development and progression of pulmonary sarcoidosis. *Sci Rep* 2020; 10: 13282.
- **133** Crouser ED, Culver DA, Knox KS, *et al.* Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med* 2009; 179: 929–938.
- 134 Zhou T, Zhang W, Sweiss NJ, *et al.* Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. *PLoS One* 2012; 7: e44818.
- **135** Jurmeister P, Bockmayr M, Seegerer P, *et al.* Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. *Sci Transl Med* 2019; 11: eaaw8513.
- 136 Yu DH, Waterland RA, Zhang P, *et al.* Targeted p16Ink4a epimutation causes tumorigenesis and reduces survival in mice. *J Clin Investig* 2014; 124: 3708–3712.
- 137 Saba HI. Decitabine in the treatment of myelodysplastic syndromes. Ther Clin Risk Manag 2007; 3: 807–817.
- **138** Brocks D, Schmidt CR, Daskalakis M, *et al.* DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* 2017; 49: 1052–1060.
- 139 Xu X, Tao Y, Gao X, *et al.* A CRISPR-based approach for targeted DNA demethylation. *Cell Discov* 2016; 2: 16009.
- 140 Vojta A, Dobrinic P, Tadic V, *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* 2016; 44: 5615–5628.
- 141 Arrowsmith CH, Bountra C, Fish PV, *et al.* Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Dis* 2012; 11: 384–400.
- 142 Huffman K, Martinez ED. Pre-clinical studies of epigenetic therapies targeting histone modifiers in lung cancer. *Front Oncol* 2013; 3: 235.
- 143 Lappalainen T, Greally JM. Associating cellular epigenetic models with human phenotypes. *Nat Rev Genet* 2017; 18: 441–451.
- 144 Smith AK, Kilaru V, Kocak M, *et al.* Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genom* 2014; 15: 145.
- 145 Pierce BL, Tong L, Argos M, *et al.* Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat Commun* 2018; 9: 804.
- 146 Kim S, Forno E, Zhang R, *et al.* Expression quantitative trait methylation analysis reveals methylomic associations with gene expression in childhood asthma. *Chest* 2020; 158: 1841–1856.
- 147 Houseman EA, Accomando WP, Koestler DC, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformat* 2012; 13: 86.
- **148** Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 2017; 18: 220.
- 149 Prakadan SM, Shalek AK, Weitz DA. Scaling by shrinking: empowering single-cell "omics" with microfluidic devices. Nat Rev Genet 2017; 18: 345–361.
- **150** Mansell G, Gorrie-Stone TJ, Bao Y, *et al.* Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genom* 2019; 20: 366.