

## Zusammenfassung

**Einleitung:** Die Sarkoidose ist eine heterogene, granulomatöse Erkrankung, deren Diagnose sich als schwierig erweisen kann. Schwer zu diagnostizieren ist, da es keine genauen Biomarker für das Fortschreiten der Krankheit gibt. Daher haben wir ein Profil erstellt und DNA-Methylom, mRNAs und microRNAs profiliert und integriert, um molekulare Veränderungen zu identifizieren Sarkoidose und dem Fortschreiten der Krankheit assoziierte molekulare Veränderungen zu identifizieren, die die zugrunde liegenden Mechanismen der Krankheit und potenzielle genomische Biomarker aufklären zu können.

**Methoden:** Es wurden bronchoalveoläre Lavagezellen von 64 Sarkoidosepatienten und 16 gesunden Kontrollpersonen verwendet. Die DNA-Methylierung wurde mit Illumina HumanMethylationEPIC-Arrays, die mRNA durch RNA-Sequenzierung und miRNAs durch Sequenzierung kleiner RNAs. Lineare Modelle wurden angepasst, um die Auswirkungen der Diagnose und des Phänotyps zu testen, wobei Alter, Geschlecht und Rauchen berücksichtigt wurden. Wir haben ein überwachtes Multi-omics-Modell mit einer Untergruppe von Merkmalen aus jedem Datensatz verwendet.

**Ergebnisse:** Wir identifizierten 46.812 CpGs, 1.842 mRNAs und 5 miRNAs, die mit Sarkoidose assoziiert sind, und 1 mRNA, SEPP1 - ein Protein, das Zellen mit Selen versorgt und mit dem Fortschreiten der Krankheit assoziiert ist. Unser integriertes Modell unterstreicht die Bedeutung des PI3K/AKT1 Signalwegs bei Sarkoidose, der für die T-Zell- und mTOR-Funktion wichtig ist. Neuartige immunbezogene Gene und miRNAs, darunter LYST, RGS14, SLFN12L und hsa-miR-199b-5p, unterscheiden Sarkoidose von den Kontrollen. Unser integriertes Modell zeigte auch eine unterschiedliche Expression/Methylierung von IL20RB, ABCC11, SFSWAP, AGBL4, miR-146a-3p und miR-378b zwischen nicht-progredienter und progressiver Sarkoidose.

**Schlussfolgerungen:** Durch die Nutzung des DNA-Methyloms, des Transkriptom und der miRNA-Sequenzierung in Sarkoidose-BAL-Zellen konnten wir weit verbreitete molekulare Veränderungen nachweisen, die mit der Krankheit in Verbindung stehen und von denen viele an der Immunantwort beteiligt sind. Diese Moleküle können als diagnostische/prognostische Biomarker und/oder Zielmoleküle für Medikamente dienen, doch sind zur Bestätigung weitere Tests erforderlich.

## Einleitung

Die Sarkoidose ist eine heterogene Krankheit, die durch eine nicht kaseierende granulomatöse Entzündung gekennzeichnet ist, von der Schwarze und Frauen in besonderem Maße betroffen sind 1. Bei über 90 % der Betroffenen 2 mit pulmonaler Sarkoidose können Symptome asymptomatisch sein oder Remission oder Rückbildung zeigen; ein Fortschreiten kann jedoch zu Beeinträchtigungen und/oder Lungenfibrose führen, der Hauptursache für die Sterblichkeit 3. Der Verlauf der pulmonalen Sarkoidose ist unvorhersehbar, wobei mindestens 25 % der Patienten eine chronische oder fortschreitende Erkrankung entwickeln, die Behandlung erfordert 4. Sie wird durch eine Kombination von genetischen, umweltbedingten und immunologischen Faktoren des Wirtes ausgelöst, Die Ursache(n) der Sarkoidose sind derzeit unbekannt. Mehrere Indizien deuten auf einen antigenen Stimulus hin, darunter HLA-Allel-Assoziationen, umweltbedingte, saisonale und regionale Muster und eine Th1-vorherrschende Immunantwort, bei der CD4+ T-Zellen IFN- $\gamma$  und TNF- $\alpha$  sezernieren. Darüber hinaus werden abweichende und dysfunktionale Immunreaktionen mit Sarkoidose in Verbindung gebracht und durch genomweite Transkriptomstudien belegt 6. Zwar

haben frühere Studien viele Faktoren aufgeklärt, die zur Entstehung und zum Fortschreiten der Krankheit beitragen, doch gibt es noch viele Wissenslücken.

Epigenetische Mechanismen wie die DNA-Methylierung (DNAm) und mikroRNAs (miRNAs) vermitteln die Genexpression, werden durch Expositionen verändert und sind dynamisch und reversibel, was sie zu Kandidaten für die Genregulierung bei Sarkoidose sowie zu vielversprechenden Biomarkern und therapeutischen Zielen macht. Epigenetische Dysregulation wurde bei vielen Lungenerkrankungen festgestellt und ist wahrscheinlich für das Fortschreiten und die Manifestationen der Sarkoidose verantwortlich. Wir haben bereits Veränderungen der DNAm- und mRNA-Genexpression in bronchoalveolären Lavagezellen (BAL) von Patienten mit chronischer Berylliumkrankheit (einer durch Berylliumexposition verursachten granulomatösen Lungenerkrankung) und einer kleinen Stichprobe von Sarkoidosepatienten nachgewiesen 7, 8., obwohl andere miRNAs identifiziert haben, die mit der Krankheit in Verbindung stehen 9, 10. Darüber hinaus wurden miRNAs aus der BALF wie miR-27b, miR-192 und miR-221 mit dem Fortschreiten der pulmonalen Sarkoidose in Verbindung gebracht 11. Diese Studien waren durch kleine Stichprobengrößen, gezielte im Gegensatz zu genomweiten Ansätzen und eine mangelnde Integration verschiedener omics-Modalitäten begrenzt.

Wir stellen die Hypothese auf, dass mit einem größeren integrierten Ansatz, der mehrere genomische Datentypen nutzt, dass epigenomische Studien Risikofaktoren mit der Pathobiologie der Krankheit verknüpfen könnten, um den Krankheitsverlauf zu verstehen und Patienten auf der Grundlage des molekularen Profils zu klassifizieren; letztlich würde dies zielgerichtete Forschung zu Krankheitsmanifestationen und Behandlung ermöglichen. In einem ersten Schritt haben wir diese Studie durchgeführt, um die genomweite DNAm-, mRNA- und miRNA-Expression in Sarkoidose-BAL-Zellen zu erfassen, stratifiziert nach Krankheitsverlauf. Indem wir jeden Datensatz separat analysierten, konnten wir viele molekulare Merkmale identifizieren, die mit der Krankheit insgesamt assoziiert sind. Anschließend konstruierten wir ein spärliches Multi-omic Modell, das DNAm, mRNAs und miRNAs umfasst, konnten wir Merkmale identifizieren, die mit Sarkoidose und pulmonaler Progression assoziiert sind.

## **Methoden**

### **Studienpopulation**

Sarkoidose-BAL-Zellproben wurden aus dem Konsortium Genomic Research in Alpha-1 Antitrypsin-Mangel und Sarkoidose (GRADS)-Konsortium<sup>12</sup> (einschließlich National Jewish Health (NJH, n=17) und Nicht-NJH-GRADS-Fällen (n=39)) und Fällen aus dem Granuloma Biorepository am NJH (n=8). Kontrollpersonen ohne Lungenerkrankung in der Vorgeschichte kamen aus dem NJH-Lungenspenderkern (n=16). Alle Sarkoidosepatienten erfüllten die ATS/ERS-Kriterien<sup>13</sup> für die Biopsie zur Bestätigung der Diagnose einer Sarkoidose. Der nicht-progressive Phänotyp wurde definiert als keine neue Organbeteiligung, Lungenfunktionstests mit Bronchoskopie und Nukleinsäureaufbereitung. Die Bronchoskopie mit BAL wurde wie zuvor beschrieben durchgeführt.<sup>12,14</sup> Die Zellen wurden isoliert und bei -80°C in RLT-Puffer eingefroren. DNA und RNA wurden mit dem Qiagen AllPrep DNA/RNA extraction mini kit extrahiert. Die gereinigte genomische DNA wurde mit dem Zymo EZ-96 DNA-Methylierungs-Bisulfit-Konvertierungskit bisulfitiert, gefolgt von einer Ganzgenom-Amplifikation und enzymatischen Fragmentierung. Die DNA wurde denaturiert und mit Illumina Infinium HumanMethylationEPIC BeadChips hybridisiert, gefolgt von einer Einzelbasenverlängerung. Die hybridisierten BeadChips wurden gefärbt, gewaschen und mit dem iScan-System von Illumina gescannt. mRNA-Bibliotheken wurden aus 500 ng Gesamt-RNA mit TruSeq stranded mRNA library preparation kits (Illumina) hergestellt, und miRNA-Bibliotheken wurden mit Lexogen Small RNA-Seq library preparation kits erstellt.

Die RNA-Bibliotheken wurden mit einer durchschnittlichen Tiefe von 80M 150bp paired-end reads auf dem Illumina NovaSeq 6000 sequenziert. Die RNA-Seq-Reads von weiteren 28 Proben wurden aus dem GRADS-Konsortium<sup>6</sup>.

Weitere Einzelheiten sind in den ergänzenden Methoden beschrieben.

**Datenanalyse** Für jeden Datensatz wurden lineare Modelle an jedes Merkmal angepasst, um eine Auswirkung der Sarkoidose-Diagnose oder der progressiven bzw. nicht-progressiven Erkrankung zu testen, wobei Alter, Geschlecht und Rauchen berücksichtigt wurden.. Die RNA-seq-Daten wurden zusätzlich um ein Surrogat bereinigt, das einen Batch-Effekt des Sequenzierzentrums<sup>15</sup> erfasst (ergänzende Abbildung 1). Die Verteilungen der Nullteststatistiken wurden für jede Analyse mit Hilfe von Bacon abgeleitet, um die Inflation zu reduzieren <sup>16</sup>. Die P-Werte wurden an eine 5%ige Falschentdeckungsrate (FDR) angepasst, um Mehrfachtests mit Hilfe des Benjamini-Hochberg-Verfahrens zu berücksichtigen<sup>17</sup>. Die Anreicherung signifikanter Ergebnisse in der Gene Ontology (GO)-Ressource<sup>18</sup> und der Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>19</sup> wurde mit GOMeth<sup>20</sup> für DNAm-Daten und clusterProfiler<sup>21</sup> für mRNA-Daten durchgeführt. CpGs wurden mit Hilfe des annotatr R-Pakets<sup>22</sup> zu CpG-Inseln, -Regalen und -Ufern sowie Genelementen annotiert. Experimentell bestätigte miRNA-Zielgene wurden von MirTarBase<sup>23</sup> bezogen. Für jede miRNA wurden Zielgene ausgewählt, für die es mindestens zwei Beweisquellen gab (Reporter Assay, Western Blot, qPCR, Microarray, NGS, pSILAC, CLIP-Seq oder andere), darunter mindestens eine Quelle, die als starker Beweis gilt (Reporter Assay, Western Blot oder qPCR).

Datenintegrationsanalyse zur Entdeckung von Biomarkern mit latenten Komponenten (DIABLO) Methylierungs-M-Werte und normalisierte mRNA- und miRNA-Genexpressionswerte wurden als Input für ein DIABLO<sup>24</sup>-Modell verwendet, das im mixOmics R-Paket<sup>25</sup> implementiert wurde. Die Genexpressionswerte wurden mit DESeq2 <sup>26</sup> auf die Bibliotheksgröße normalisiert und mit einer Varianzstabilisierungstransformation (VST)<sup>27</sup> transformiert, um Heteroskedastizität zu berücksichtigen. Eine Teilmenge von Merkmalen aus jedem Datensatz für 2 Modellkomponenten wurden für die Modelleingabe mit LASSO-Regression und 5-fache Kreuzvalidierung 10 Mal wiederholt ausgewählt.

## Ergebnisse

**Demografische Daten der Studienpopulation** Wir rekrutierten 64 Sarkoidosefälle, darunter 26 progrediente und 38 nicht-progrediente, sowie 16 gesunde Kontrollpersonen. Die demografischen Daten zum Zeitpunkt der BAL sind in Tabelle 1 aufgeführt. Keine signifikante Unterschiede hinsichtlich Rasse, ethnischer Zugehörigkeit oder Alter festgestellt. Nicht-progressive Fälle waren häufiger weiblich als progressive Fälle. Progressive Fälle wiesen eine signifikant reduzierte FEV1 und FVC und mehr Erkrankungen im Stadium 2.

Die Genexpression, die für Immunwege steht, unterscheidet sich zwischen Fällen und Kontrollen und in geringerem Maße nach und in geringerem Maße bei den Phänotypen des Krankheitsverlaufs. Wir analysierten die Genexpressionsdaten, um die Unterschiede zwischen Fällen und Kontrollen sowie zwischen fortschreitender und nicht fortschreitender Krankheit. Beim Vergleich zwischen Sarkoidose und Kontrollen entdeckten wir 1.842 differentiell exprimierte Gene (DEGs), von denen 379 (20,6 %) bei Sarkoidose signifikant erhöht waren bei Sarkoidose signifikant erhöht waren (Abbildung 1A; ergänzende Tabelle 1). Die bedeutendsten hochregulierten Gene betrafen viele Gene, die mit Immunität, Zellmigration und Zelladhäsion zusammenhängen, darunter: EVL (Enah/Vasp-like) und SERPINA9 (Serp family A member 9). Die wichtigsten herunterregulierten DEGs sind SFTPA2 und SFTPC, Surfactant-Gene und Transkriptionsfaktoren wie CEBPD (CCAAT Enhancer Binding Protein Delta) und NKX2-1 (NK2 homeobox 1). Eine einzige DEG wurde in progressiven und nicht-



progressiven Fällen gefunden: SEPP1 (Abbildung 1B).

Als Nächstes testeten wir, ob Sarkoidose-DEGs in GO 18 und KEGG-Pathways überrepräsentiert waren. Pfade überrepräsentiert sind, indem wir clusterProfiler 21 verwendeten (Abbildung 1C-D; ergänzende Tabelle 2). Hochregulierte Gene waren für 257 GO- und 18 KEGG-Pfade angereichert. Zu den am stärksten angereicherten GO-Begriffen gehören der MHC-Klasse-II-Proteinkomplex, die Regulierung der mononuklearen Zellproliferation, die Zell-Zell-Adhäsion von Leukozyten und die Reaktion auf Interferon gamma. Zu den wichtigsten KEGG-Pfaden gehören Antigen processing and presentation, cell adhesion molecules, and multiple viral and autoimmune diseases including herpes simplex virus 1 infection, Epstein-Barr virus infection, viral myocarditis, Influenza A, tuberculosis, rheumatoid arthritis, and systemic lupus erythematosus.

Die herunterregulierten Gene wurden für 512 GO- und 0 KEGG-Pfade angereichert. Zu den am stärksten angereicherten GO-Begriffen gehören Komplementaktivierung, klassischer Signalweg, extrazelluläre Matrix, Regulation der humoralen Immunantwort, Immunglobulinkomplex, Zell-Zell-Verbindung, Wundheilung, Mesenchymentwicklung und Epithelzellproliferation.

DNA-Methylierung unterscheidet sich zwischen Fällen und Kontrollen und überschneidet sich mit der Genexpression Als Nächstes testeten wir auf differentielle Methylierung (DM) und identifizierten 46.812 DNAm-Stellen, die mit Sarkoidose assoziiert sind, von denen 38.504 zu 15.208 einzigartigen Genen gehören (Abbildung 2A; ergänzende Tabelle 3). Wir konnten keine DNAm-Stellen finden, die signifikant mit dem Fortschreiten der Krankheit assoziiert sind. Die Mehrzahl der CpGs war bei Sarkoidose hypermethyliert (35.154; 75,1 %). Signifikante Sonden waren angereichert für intronische Regionen (68,5 % gegenüber 60,8 %; exakter Test nach Fisher  $p < 2,2 \times 10^{-16}$ ) und FANTOM5-Enhancer (9,17 % gegenüber 4,14 %; exakter Test nach Fisher  $p < 2,2 \times 10^{-16}$ ) im Vergleich zu allen getesteten DNAm-Stellen. Hypomethylierte DM-Stellen waren signifikant für 687 GO-Terme angereichert, mit Top-Treffern wie Immunantwort, Leukozytenaktivierung und Abwehrreaktion sowie 88 KEGG-Pfaden, darunter virale Proteininteraktion mit Zytokin und Zytokinrezeptor, Chemokin-Signalweg, Th17-Zelldifferenzierung und T-Zell-Rezeptor-Signalweg (Abbildung 2B; ergänzende Tabelle 4). Hypermethylierte DM-Stellen waren signifikant für 146 GO-Terme angereichert, mit Top-Treffern wie Aktin-Zytoskelett-Organisation und 29 KEGG-Signalwegen, darunter Rap1-Signalisierung, Ras-Signalisierung und Phosphatidylinositol-Signalisierungssystem (Abbildung 2C). Zu den bedeutendsten hypomethylierten DM-Stellen gehören CpGs, die in einer vorhergesagten Promotorregion von RGS14, einem intronischen Enhancer von FGF18 und MGAT1 liegen. Top . hypermethylierte DM-Sites kartiert, um Gene wie LYST (lysosomal trafficking regulator), die Protein-Trafficking zu Lysosomen reguliert.

Als Nächstes wurden DEGs und differentieell methylierte Sonden (DMPs) bei Sarkoidose im Vergleich zu Kontrollen auf der Grundlage der Gen-ID überlagert. 961 DEGs (52,2%) wiesen DMs von 12.297 einzigartigen CpGs auf. 52,9 % der Assoziationen (41.674/78.761) zeigten kanonische umgekehrte Beziehungen: die Richtung der DNAm war der Genexpression entgegengesetzt (Abbildung 2D). Zu den Genen mit kanonischen Beziehungen und hochregulierter RNA gehört TGFB1. Zu den kanonischen Beziehungen mit herunterregulierter RNA gehören AQP1, FRMD4A, ARNT2 und BMP4. Zu den Genen mit nicht-kanonischen Beziehungen und herabregulierter RNA gehören ITGA9, eine Integrinkomponente des Rezeptors für VCAM1, LIMCH1, das den Zusammenbau von Stressfasern positiv reguliert und fokale Adhäsionen stabilisiert, und FGF18.

Die miRNA-Expression unterscheidet sich zwischen Fällen und Kontrollen und zielt auf DEGs ab. Als nächstes verglichen wir die miRNA-Expression bei Sarkoidose und Kontrollen und identifizierten 5 miRNAs (hsa-miR-143-3p, hsa-miR-199a-3p/hsa-miR-199b-3p, hsa-miR-199b-5p, hsa-miR-582-3p & hsa-miR-582-5p) bei Sarkoidose herunterreguliert (Abbildung 3A-B; ergänzende Tabelle 5).

Keine miRNAs wurden mit der Progression in Verbindung gebracht. Mit Hilfe von MirTarBase23 haben wir für jede signifikante miRNA experimentell validierte Zielgene ermittelt. Wir identifizierten 67 Zielgene, von denen 4 (AKT1, CD44, JAG1, PTGS2) von 2 DE miRNAs angegriffen werden (Abbildung 3C), und weitere 15 waren DEGs in unserer mRNA-Analyse, darunter TNF. Integriertes Modell zeigt Assoziationen mit Sarkoidose und mit der Progression, die in Einzelanalysen nicht gefunden wurden. Wir verwendeten DIABLO, um die drei Datensätze zu integrieren, einschließlich der Personen, die nach der Qualitätskontrolle in allen Datensätzen vorhanden waren (n = 65; 13 Kontrollen, 19 progressive, 33 nicht-progressive Sarkoidose). Mit dem Ziel, molekulare Merkmale zu bestimmen, die die drei Gruppen (Kontrollen, progressive und nicht-progressive Sarkoidose) voneinander abgrenzen, konstruierten wir ein Multi-omic Modell mit 2 latenten Variablen, die lineare Kombinationen von Input-Funktionen sind. Unter Verwendung einer 5-fachen Kreuzvalidierung, die 10-mal wiederholt wurde, stellten wir fest, dass die Auswahl von 5 DNAm-Stellen, 2 mRNAs und 13 miRNAs für die latente Variable 1 und 2 DNAm-Stellen, 2 mRNAs und 14 miRNAs für die latente Variable 2 die Vorhersagegenauigkeit des Modells maximierte und zu einer Clusterbildung auf der Grundlage der Diagnose führte (Abbildung 4A; ergänzende Tabelle 6). Die latente Variable 1 trennt Kontrollen von Sarkoidoseproben und die latente Variable 2 trennt progressive von nicht-progressiver Sarkoidose. Die in jeder latenten Variable enthaltenen Merkmale sind in Abbildung 4B und der ergänzenden Tabelle 7 dargestellt. Des Weiteren konstruierten wir Netzwerke auf der Grundlage von Korrelationen der Merkmalsgewichte für jede latente Variable (Abbildung 4C). Die beiden mRNAs, die zur latenten Variable 1 beitragen, waren SFTPB und SFTPD. Zu den DNAm-Stellen, die zur latenten Variable 1 beitragen, gehören DNAm-Stellen, die in den Fällen cg16962115 innerhalb eines Introns von LYST und cg05300241 innerhalb eines Introns von PIK3CD hypermethyliert sind. Die übrigen drei DNAm-Stellen waren hypomethyliert und umfassten cg21949194 innerhalb eines Enhancers von SOS1 (SOS Ras/Rac guanine nucleotide exchange factor 1), cg11370586 innerhalb des vorhergesagten Promotors von RGS14 und cg03526142 innerhalb eines Exons von SLFN12L. Alle DNAm-Stellen, die zur latenten Variable 1 beitragen, waren in unserer früheren Methylierungsanalyse DMPs. Neben der DE miR hsa-miR-199b-5p gehören auch hsa-miR-204-5p zu den miRNA-Merkmalen der latenten Variable 1. Zu den Merkmalen, die zur latenten Variable 2 beitragen, gehören die DNAm-Stellen cg05479174 innerhalb eines Exons von SFSWAP und cg06635176 innerhalb eines Introns von AGBL4. Die mRNA von IL20RB (Interleukin-20-Rezeptor-Untereinheit beta) war bei nicht-progressiver Sarkoidose hochreguliert, während die Expression von ABCC11 (ATP binding cassette subfamily C member 11) bei progressiver Sarkoidose höher war. miRNAs, die zur latenten Variable 1 beitragen, sind hsa-miR-146a-3p und hsa-miR-378b.

## Diskussion

In dieser Studie stellen wir die erste Anwendung der multimikrobiellen Integration bei Sarkoidose vor, indem wir kodierende und miRNA-Expressionsdaten mit DNA-Methylierungsdaten kombinieren, um ein multimikrobielles Netzwerk zu erstellen. Durch eine erste genomweite Profilierung von DNA-Methylierung, mRNA-Expression und miRNA-Expression bei Sarkoidose BAL identifizierten wir 1.846 DEGs, 46.812 DMPs und 5 DE miRNAs bei Sarkoidose im Vergleich zu gesunden Kontrollen sowie 1 DEG, die mit dem Fortschreiten der Sarkoidose assoziiert ist. Durch die Integration von Omic-Datensätzen definieren wir pathogene Moleküle/Gene bei Sarkoidose die mit konventionellen Modellierungsmethoden in Einzel-omics-Datensätzen nicht identifiziert werden konnten. Während unser Single-omics-Ansatz nur 1 DEG für die Progression nachwies, identifizierte das ultimische Modell mehrere progressionsassoziierte Methylierungs-, mRNA- und miRNA-Merkmale. Wir

identifizieren Moleküle und Signalwege, die mit der Krankheit assoziiert sind, und weisen auf neue molekulare Merkmale als potenzielle Treiber und Modifikatoren der Sarkoidose und zeigen damit das Potenzial von integrativen Ansätzen.

Unsere Einzelgenom-Analysen weisen auf eine Vielzahl von Genen in der Pathogenese der Sarkoidose hin, wobei viele dieser Gene an Prozessen beteiligt sind, die mit Entzündungen und Immunität (möglicherweise in Verbindung mit denen, die an viralen/atypischen bakteriellen Infektionen und Autoimmunität beteiligt sind) und der Entzündung der extrazellulären Matrix zusammenhängen. Während wir viele Assoziationen mit dem Fallstatus feststellten, entdeckten wir nur eine einzige omische Assoziation mit der Progression: SEPP1 mRNA-Spiegel.

Die Ergebnisse der GRADS-Studie zeigten, dass die SEPP1-spiegelverkehrt korreliert sind mit DLCO (%) und FVC (% pred) bei Sarkoidosepatienten korreliert 6. Diese Ergebnisse sprechen für einen progressiven Phänotyp, einschließlich der Lungenfunktionsveränderungen, die wir und andere in der Definition der fortschreitenden Lungenerkrankung verwendet haben.

In unserer Studie wurden auch 5 DE miRNAs nachgewiesen, die in Sarkoidosefällen herunterreguliert wurden, und zwar in der ersten Studie über genomweite miRNAs bei Sarkoidose. Während diese Assoziationen neu sind, wurden mehrere Zielgene dieser miRNAs bereits zuvor mit Sarkoidose in Verbindung gebracht, darunter die Gene AKT1, CD44, JAG1 und PTGS2. CD44 wurde zum Beispiel in Bereichen der Granulombildung und Fibrose gefunden<sup>28</sup> und wird unterschiedlich zwischen Löfgren-Syndrom und Nicht-Löfgren-Syndrom-Patienten<sup>29</sup> unterschiedlich exprimiert. Durch die Integration von Genen, die sowohl für DE und DM identifizierten wir biologisch relevante Gene bei Sarkoidose im Vergleich zu Kontrollen, darunter HIF1A und TGFB1. Eine frühere Studie zeigte, dass die HIF1A-Protein- und mRNA-Spiegel in Sarkoidose-Granulomen verringert waren<sup>30</sup>. Wir und andere haben Veränderungen bei TGFB1, das das für TGF- $\beta$  kodiert, bei Lungensarkoidose festgestellt<sup>31, 32</sup>. Außerdem wurden TGF- $\beta$ -Genotypen mit dem Schweregrad der Sarkoidose in Verbindung gebracht<sup>31</sup>.

Unser integriertes Modell zeigte mehrere neue Gene und miRNAs zwischen Sarkoidose und Kontrollen, darunter LYST, RGS14, SLFN12L und hsa-miR-199b-5p. Die ersten 3 Gene scheinen eine wichtige Rolle bei der Immunfunktion zu spielen. Insbesondere LYST, ein Regulator des Endosomen/Lysosomen LYST, ein Regulator des Endosomen-/Lysosomenverkehrs, kann TLR3- und TLR4-vermittelte Pfade<sup>33</sup> regulieren, Gene, die am angeborenen Immunsystem beteiligt sind und bei Sarkoidose eine Rolle spielen<sup>6, 34, 35</sup>. RGS14 wird exprimiert in Lymphozyten exprimiert und reguliert Chemokinrezeptoren zur Steuerung von Immunreaktionen auf exogene Wirkstoffe<sup>36</sup>. SLFN21L schließlich reguliert die Thymozytenentwicklung und wird bei der T-Zell-Aktivierung herunterreguliert, was auf eine Rolle als Regulator der Immunantwort hindeutet<sup>37</sup>.

Unsere integrierten Analysen zeigten miRNAs DE zwischen progressiver/nicht-progressiver Sarkoidose, einschließlich der zuvor beschriebenen. Insbesondere miR-146a-3p, das bei progressiver Sarkoidose hochreguliert ist, ist ein Indikator für Entzündung und oxidativen Stress, der auf TLR4 abzielen kann und zuvor in Sarkoidose-BALF sowie im Serum erhöht gefunden wurde<sup>38,39</sup>. miR-378b, das unter einer wurde in progressiven Fällen hochreguliert, wurde zuvor mit Sarkoidose in Verbindung gebracht<sup>40</sup>. Wir identifizierten auch neue Gene wie IL20RB, ABCC11, SFSWAP und AGBL4. Interessanterweise war die Expression von IL20RB in unserer Studie bei nicht-progressiver Sarkoidose höher als bei progressiver Sarkoidose, was darauf hindeutet, dass die Aktivierung dieses Entzündungswegs beim nicht-progressiven Phänotyp erhöht ist.



ABCC11 ist ein Gen, das die Makrophagendifferenzierung beeinflusst und TNF- $\alpha$  und IL17 durch TLR4-Signalisierung induziert 41 ; diese Ergebnisse sowie die oben genannten neuen Erkenntnisse unterstützen die Bedeutung von Genen der angeborenen Immunantwort bei der Sarkoidose. Sowohl unsere individuellen als auch unsere integrativen Analysen identifizieren Moleküle, die an der PI3K/AKT1-Signalübertragung beteiligt sind, einem Signalweg, der bereits in der Sarkoidose-Pathogenese erkannt wurde. So beobachten wir beispielsweise eine Hypermethylierung von PI3KCD (Phosphatidylinositol-4,5-bisphosphat-3-Kinase katalytische Untereinheit delta), die für eine Komponente von PI3K kodiert und mit AKT1 einen Komplex bildet, der die Differenzierung und Funktion von T-Zellen beeinflusst 42. Unsere miRNA-Ergebnisse zeigen, dass AKT1 ein Ziel von herunterregulierten miRNAs (miR-143-3p und miR-199a/miR-199b) bei Sarkoidose ist. Sowohl eine Herunterregulierung als auch eine Hochregulierung der PI3K/AKT-Signalwege wurde mit Sarkoidose in Verbindung gebracht 42, 43. Frühere Studien haben gezeigt, dass sowohl die verminderte proliferative Reaktion als auch die Erschöpfung der T-Zellen bei fortschreitender Sarkoidose zum Teil auf die Hemmung des PI3K/AKT1-Signalwegs zurückzuführen sind 42. Der PI3K/AKT1-Signalweg wurde auch mit der Aktivierung von mTOR in Verbindung gebracht 43, die mit der Granulombildung in Verbindung gebracht wurde 44. Interessanterweise wurde in der kürzlich durchgeführten GRADS-Studie eine PI3K-Aktivierung mit einem Endtyp der Sarkoidose in Verbindung gebracht, der durch eine hiläre Lymphadenopathie, eine pulmonale Retikulierung, eine geringere Multiorganbeteiligung und mehr Umweltassoziationen gekennzeichnet ist 6.

Obwohl unsere Stichprobengröße größer ist als die der meisten früheren Sarkoidose-omischen Studien, war die Aussagekraft unserer Analysen begrenzt, insbesondere bei den Phänotypenanalysen. In zukünftigen Studien wird es wichtig sein, dass die um sowohl die Stichprobengröße zu erhöhen als auch die Assoziationen mit quantitativen Messgrößen zu untersuchen, die besser auf die Krankheit abgestimmt sind, wie z. B. Lungenfunktionstestvariablen. Wir waren nicht in der Lage, die Zellproportionen bei Sarkoidosepatienten im Vergleich zu den Kontrollen zu korrigieren, da unsere Kontrollen keine Zelldifferenziale aufwiesen. Veränderte Zellanteile könnten die herunterregulierten Epithelgene in der Sarkoidose-BAL (z. B. verringerte Surfactant-Proteine bei Sarkoidose) und einige der DNAm-Stellen und DEGs erklären, die im Vergleich zwischen Sarkoidose und Kontrollen entdeckt wurden, obwohl Epithelzellen im Allgemeinen keinen großen Anteil der BAL ausmachen. Schließlich wurden in unseren multimikrobiellen Modellen demografische Informationen nicht berücksichtigt. Die Regression von demografischen Informationen wie Alter und Geschlecht könnte zu aussagekräftigeren Netzwerken führen. Trotz dieser Unzulänglichkeiten haben wir weit verbreitete molekulare Veränderungen festgestellt und multimikrobielle Netzwerke konstruiert, die mit Sarkoidose und deren Verlauf in Verbindung stehen.

Die in diesen Analysen entdeckten Moleküle geben Aufschluss über die Pathogenese der Krankheit und können nach Replikation/Validierung in weiteren Populationen auch therapeutisch oder als Biomarker genutzt werden.

.CC-BY-NC-ND 4.0 International license available under a (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. Das Copyright für diese Vorabdruckversion wurde am 27. Januar 2023 veröffentlicht.

<https://doi.org/10.1101/2023.01.26.525601>doi:bioRxiv preprint

[Translated with DeepL](#)

## Multi-Omic Signatures of Sarcoidosis and Progression in Bronchoalveolar Lavage Cells

Iain R. Konigsberg<sup>1\*†</sup>, Nancy W. Lin<sup>2,3\*†</sup>, Shu-Yi Liao<sup>2,3,4</sup>, Cuining Liu<sup>5</sup>, Kristyn MacPhail<sup>2</sup>, Margaret M. Mroz<sup>2</sup>, Elizabeth Davidson<sup>1</sup>, Clara I. Restrepo<sup>2</sup>, Sunita Sharma<sup>3</sup>, Li Li<sup>2,3</sup>, Lisa A. Maier<sup>2,3,4\*\*</sup>, & Ivana V. Yang<sup>1,3\*\*</sup>

<sup>1</sup> Department of Biomedical Informatics, School of Medicine, University of Colorado - Anschutz Medical Campus, Aurora, CO

<sup>2</sup> Division of Environmental and Occupational Health Sciences, Department of Medicine, National Jewish Health, Denver CO

<sup>3</sup> Division of Pulmonary and Critical Care Sciences, Department of Medicine, School of Medicine, University of Colorado - Anschutz Medical Campus, Aurora CO

<sup>4</sup> Department of Environmental and Occupational Health, Colorado School of Public Health, Aurora, CO

<sup>5</sup> Department of Biostatistics and Bioinformatics, Colorado School of Public Health, Aurora CO

SARKOIDOSE

\*, \*\*: These authors contributed equally

†: Correspondence: [iain.konigsberg@cuanschutz.edu](mailto:iain.konigsberg@cuanschutz.edu) and [linnancy@njhealth.org](mailto:linnancy@njhealth.org)

SELBSTHILFE

Supported by NIH awards R01HL140357, TL1TR0025331 and UL1TR001082; FSR grant 22-505-RFP.



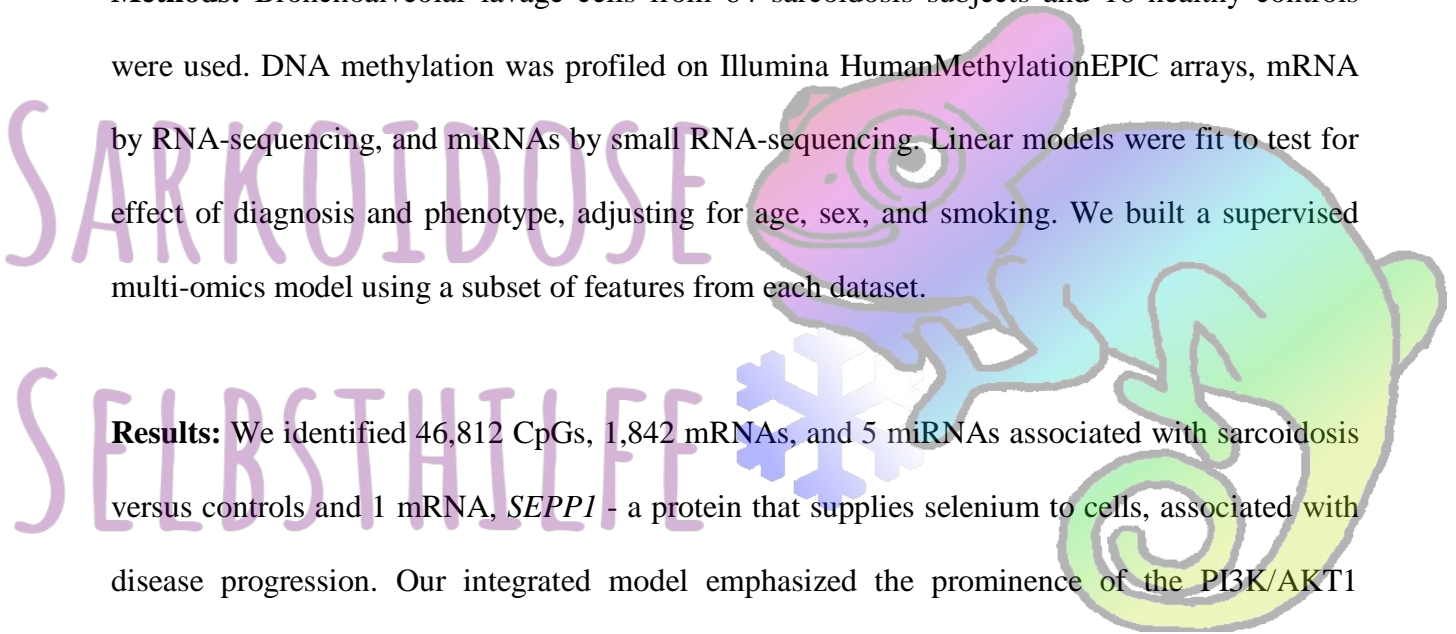


## Abstract

**Introduction:** Sarcoidosis is a heterogeneous, granulomatous disease that can prove difficult to diagnose, with no accurate biomarkers of disease progression. Therefore, we profiled and integrated the DNA methylome, mRNAs, and microRNAs to identify molecular changes associated with sarcoidosis and disease progression that might illuminate underlying mechanisms of disease and potential genomic biomarkers.

**Methods:** Bronchoalveolar lavage cells from 64 sarcoidosis subjects and 16 healthy controls were used. DNA methylation was profiled on Illumina HumanMethylationEPIC arrays, mRNA by RNA-sequencing, and miRNAs by small RNA-sequencing. Linear models were fit to test for effect of diagnosis and phenotype, adjusting for age, sex, and smoking. We built a supervised multi-omics model using a subset of features from each dataset.

**Results:** We identified 46,812 CpGs, 1,842 mRNAs, and 5 miRNAs associated with sarcoidosis versus controls and 1 mRNA, *SEPP1* - a protein that supplies selenium to cells, associated with disease progression. Our integrated model emphasized the prominence of the PI3K/AKT1 pathway in sarcoidosis, which is important in T cell and mTOR function. Novel immune related genes and miRNAs including *LYST*, *RGS14*, *SLFN12L*, and hsa-miR-199b-5p, distinguished sarcoidosis from controls. Our integrated model also demonstrated differential expression/methylation of *IL20RB*, *ABCC11*, *SFSWAP*, *AGBL4*, miR-146a-3p, and miR-378b between non-progressive and progressive sarcoidosis.



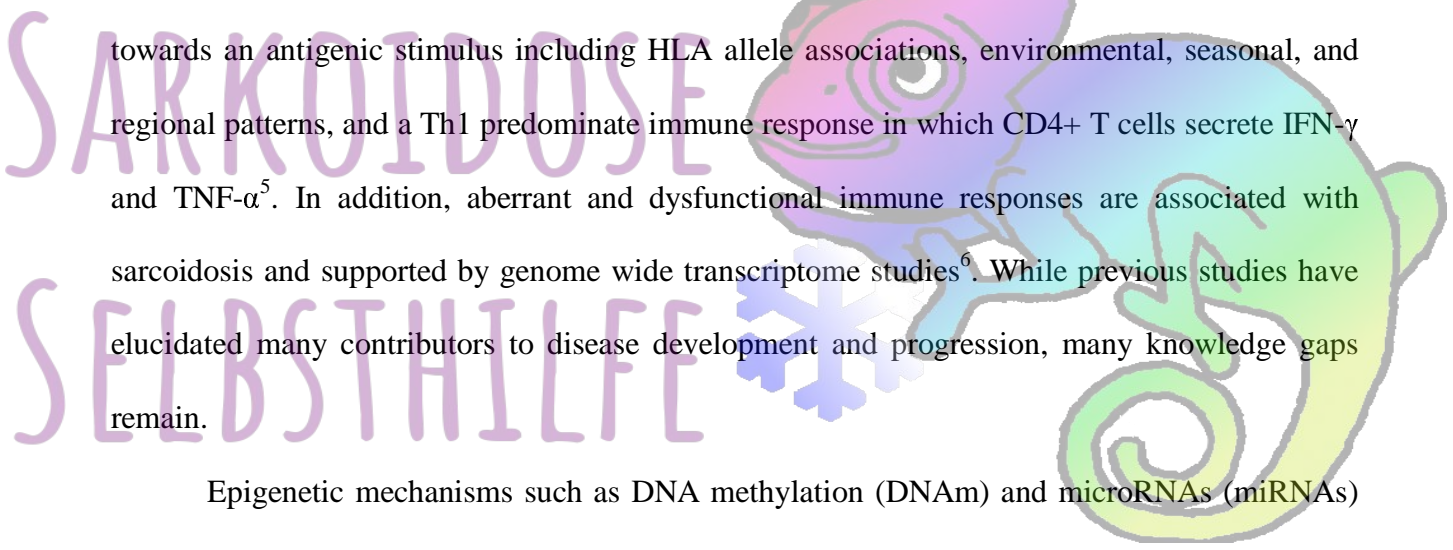
**Conclusions:** Leveraging the DNA methylome, transcriptome, and miRNA-sequencing in sarcoidosis BAL cells, we detected widespread molecular changes associated with disease, many which are involved in immune response. These molecules may serve as diagnostic/prognostic biomarkers and/or drug targets, although future testing will be required for confirmation.



## Introduction

Sarcoidosis is a heterogeneous disease characterized by non-caseating granulomatous inflammation that differentially impacts Black individuals and women<sup>1</sup>. The lungs are involved in over 90% of individuals<sup>2</sup>. Those with pulmonary sarcoidosis can be asymptomatic or demonstrate remission or resolution; however, progression can result in impairment and/or pulmonary fibrosis, the main cause of mortality<sup>3</sup>. The course of pulmonary sarcoidosis is unpredictable, with at least 25% of patients developing chronic or progressive disease requiring treatment<sup>4</sup>. Driven by a combination of genetic, environmental, and host immunologic factors, the underlying cause(s) of sarcoidosis are currently unknown. Multiple lines of evidence point towards an antigenic stimulus including HLA allele associations, environmental, seasonal, and regional patterns, and a Th1 predominate immune response in which CD4+ T cells secrete IFN- $\gamma$  and TNF- $\alpha$ <sup>5</sup>. In addition, aberrant and dysfunctional immune responses are associated with sarcoidosis and supported by genome wide transcriptome studies<sup>6</sup>. While previous studies have elucidated many contributors to disease development and progression, many knowledge gaps remain.

Epigenetic mechanisms such as DNA methylation (DNAm) and microRNAs (miRNAs) mediate gene expression, are modified by exposures, and are dynamic and reversible, making them candidates for gene regulation in sarcoidosis as well as promising biomarkers and therapeutic targets. Epigenetic dysregulation has been identified in many lung diseases and likely drives sarcoidosis progression and manifestations. We have previously demonstrated changes in DNAm and mRNA gene expression in bronchoalveolar lavage (BAL) cells from chronic beryllium disease (a granulomatous lung disease caused by beryllium exposure) patients and a small sample of sarcoidosis patients<sup>7,8</sup>. No other studies have evaluated epigenome-wide DNAm



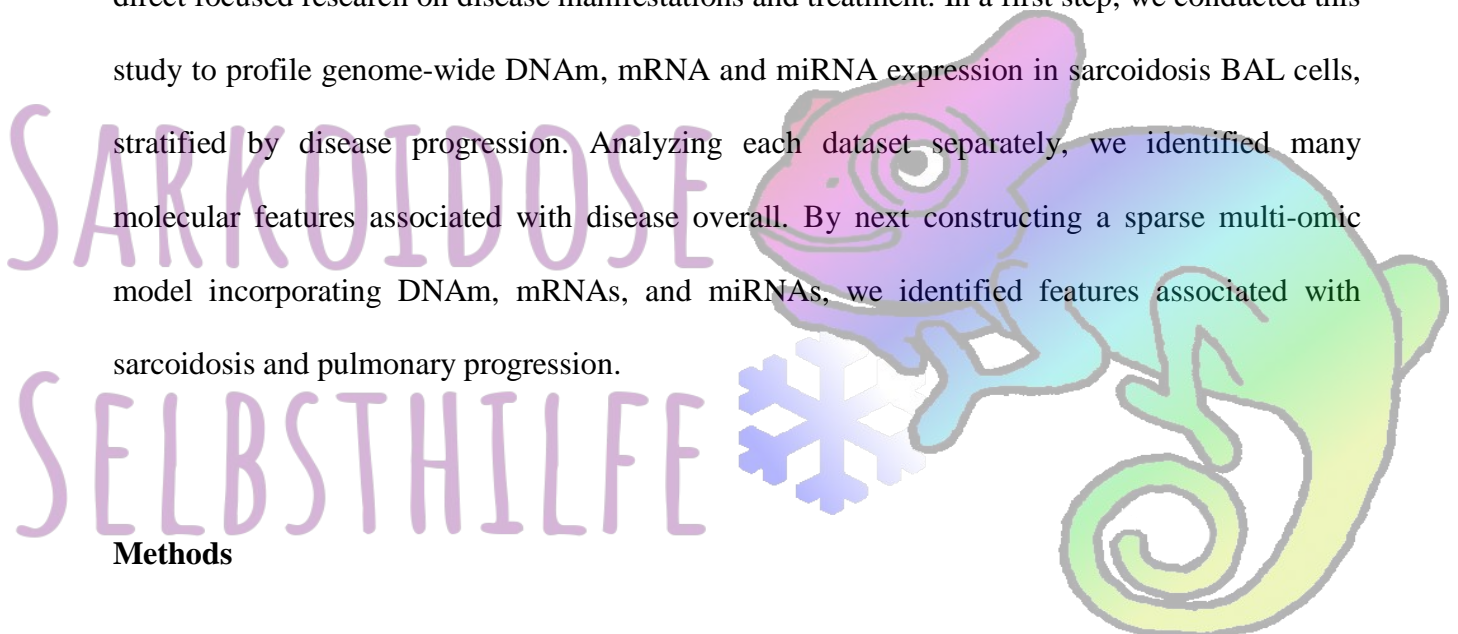
in sarcoidosis although others have identified miRNAs associated with disease<sup>9,10</sup>. Additionally, BALF miRNAs such as miR-27b, miR-192, and miR-221 have been associated with pulmonary sarcoidosis progression<sup>11</sup>. These studies were limited by small sample sizes, targeted as opposed to genome-wide approaches, and a lack of integration of different omics modalities.

With a larger integrated approach utilizing several genomic data types, we hypothesize that epigenomic studies could link risk factors to disease pathobiology to better understand disease course, and subclassify patients based on molecular profiling; ultimately this would direct focused research on disease manifestations and treatment. In a first step, we conducted this study to profile genome-wide DNAm, mRNA and miRNA expression in sarcoidosis BAL cells, stratified by disease progression. Analyzing each dataset separately, we identified many molecular features associated with disease overall. By next constructing a sparse multi-omic model incorporating DNAm, mRNAs, and miRNAs, we identified features associated with sarcoidosis and pulmonary progression.

## Methods

### *Study population*

Sarcoidosis BAL cell samples were obtained from the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) consortium<sup>12</sup> (including National Jewish Health (NJH, n=17) and non-NJH GRADS cases (n=39)) and cases from the Granuloma Biorepository at NJH (n=8). Controls with no history of lung disease were obtained from the NJH Donor Lung Core (n=16). All sarcoidosis subjects met the ATS/ERS criteria<sup>13</sup> for tissue biopsy confirmation of diagnosis of sarcoidosis. The non-progressive phenotype was defined as having either acute (i.e. consistent with Lofgren's syndrome) or non-acute disease presentation,





no new organ involvement, lung function testing with <10% decline in FVC or FEV<sub>1</sub>, <15% decline in DL<sub>CO</sub>, and stable chest imaging within 2 years after BAL. The progressive phenotype had a non-acute disease presentation; lung function testing with ≥10% decline in FVC or FEV<sub>1</sub>; or ≥15% decline in DL<sub>CO</sub>; worsening chest imaging; and/or if they required initiation of systemic immunosuppressive treatment any time up to 2 years after BAL. Non-NJH GRADS cases were phenotyped based on disease acuity, PFT, and treatment status only. See Supplemental Methods for more details.

### *Bronchoscopy and nucleic acid processing*

Bronchoscopy with BAL was performed as previously described.<sup>12,14</sup> Cells were isolated and frozen at -80C in RLT buffer. DNA and RNA were extracted using the Qiagen AllPrep DNA/RNA extraction mini kit. Purified genomic DNA was bisulfite-converted with the Zymo EZ-96 DNA Methylation bisulfite conversion kit, followed by whole-genome amplification and enzymatic fragmentation. DNA was denatured and hybridized to Illumina Infinium HumanMethylationEPIC BeadChips, followed by single base extension. Hybridized BeadChips were stained, washed, and scanned using Illumina's iScan System. mRNA libraries were prepared from 500 ng total RNA with TruSeq stranded mRNA library preparation kits (Illumina) and miRNA libraries were constructed using Lexogen Small RNA-Seq library preparation kits. RNA libraries were sequenced at an average depth of 80M 150bp paired-end reads on the Illumina NovaSeq 6000. RNA-Seq reads from an additional 28 samples were obtained from the GRADS consortium<sup>6</sup>. Additional details are outlined in the supplemental methods.

### *Data analysis*

For each dataset, linear models were fit to each feature testing for an effect of sarcoidosis diagnosis or progressive vs. non-progressive disease while adjusting for age, sex, and smoking

status. RNA-seq data was additionally adjusted for a surrogate variable capturing a batch effect of sequencing center<sup>15</sup> (**Supplementary Figure 1**). Null test statistic distributions were derived for each analysis using bacon to reduce inflation<sup>16</sup>. *P*-values were adjusted to a 5% false discovery rate (FDR) to account for multiple testing using the Benjamini-Hochberg procedure<sup>17</sup>. Enrichment of significant results in the Gene Ontology (GO) resource<sup>18</sup> and the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>19</sup> was performed using GOmeth<sup>20</sup> for DNAm data and clusterProfiler<sup>21</sup> for mRNA data. CpGs were annotated to CpG islands, shelves, and shores, and gene elements using the annotatr R package<sup>22</sup>. Experimentally confirmed miRNA target genes were obtained from MirTarBase<sup>23</sup>. For each miRNA, we retained target genes that had at least 2 sources of evidence (reporter assay, Western blot, qPCR, Microarray, NGS, pSILAC, CLIP-Seq, or other), including at least one source considered strong evidence (reporter assay, Western blot, or qPCR).

#### *Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO)*

Methylation M values and normalized mRNA and miRNA gene expression values were used as input for a DIABLO<sup>24</sup> model implemented in the mixOmics R package<sup>25</sup>. Gene expression values were normalized to library size using DESeq2<sup>26</sup> and transformed to account for heteroscedasticity with a variance stabilizing transformation (VST)<sup>27</sup>. A subset of features from each dataset for 2 model components were selected for model input using LASSO regression and 5-fold cross validation repeated 10 times.

## Results

### *Demographics of study population*

We recruited 64 sarcoidosis cases, including 26 progressive and 38 non-progressive, and 16 healthy controls. Demographic information at time of BAL is displayed in **Table 1**. No significant differences were observed in race, ethnicity, or age. Non-progressive cases were more likely female than progressive cases. Progressive cases presented with significantly reduced FEV<sub>1</sub> and FVC and more Stage 2 disease.

### *Gene expression representing immune pathways differs between cases and controls and to a lesser degree by disease progression phenotypes*

We analyzed gene expression data to assess differences between both cases and controls and progressive and non-progressive disease. In the sarcoidosis-control comparison, we detected 1,842 differentially expressed genes (DEGs), of which 379 (20.6%) were significantly increased in sarcoidosis (**Figure 1A; Supplementary Table 1**). Top significant upregulated genes involved many genes relating to immunity, cell migration, and cell adhesion including: *EVL* (Enah/Vasp-like) and *SERPINA9* (serpin family A member 9). Most significant downregulated DEGs include *SFTPA2* and *SFTPC*, surfactant genes and transcription factors such as *CEBPD* (CCAAT Enhancer Binding Protein Delta) and *NKX2-1* (NK2 homeobox 1). A single DEG was found in progressive vs non-progressive cases: *SEPP1* (**Figure 1B**).

We next tested whether sarcoidosis DEGs were overrepresented in GO<sup>18</sup> and KEGG pathways using clusterProfiler<sup>21</sup> (**Figure 1C-D; Supplementary Table 2**). Upregulated genes were enriched for 257 GO and 18 KEGG pathways. The most significantly enriched GO terms include MHC class II protein complex, regulation of mononuclear cell proliferation, leukocyte cell-cell adhesion, and response to interferon gamma. Top KEGG pathways include antigen

processing and presentation, cell adhesion molecules, and multiple viral and autoimmune diseases including herpes simplex virus 1 infection, Epstein-Barr virus infection, viral myocarditis, Influenza A, tuberculosis, rheumatoid arthritis, and systemic lupus erythematosus. Downregulated genes were enriched for 512 GO and 0 KEGG pathways. The most significantly enriched GO terms include complement activation, classical pathway, extracellular matrix, regulation of humoral immune response, immunoglobulin complex, cell-cell junction, wound healing, mesenchyme development, and epithelial cell proliferation.

#### *DNA methylation differs between cases and controls and overlaps with gene expression*

We next tested for differential methylation (DM) and identified 46,812 DNAm sites associated with sarcoidosis, of which 38,504 map to 15,208 unique genes (**Figure 2A; Supplementary Table 3**). We did not detect any DNAm sites significantly associated with disease progression. The majority of CpGs were hypermethylated in sarcoidosis (35,154; 75.1%). Significant probes were enriched for intronic regions (68.5% vs 60.8%; Fisher's exact test  $p < 2.2 \times 10^{-16}$ ) and FANTOM5 enhancers (9.17% vs 4.14%; Fisher's exact test  $p < 2.2 \times 10^{-16}$ ) relative to all tested DNAm sites. Hypomethylated DM sites were significantly enriched for 687 GO terms, with top hits including immune response, leukocyte activation, and defense response, and 88 KEGG pathways including viral protein interaction with cytokine and cytokine receptor, chemokine signaling pathway, Th17 cell differentiation, and T cell receptor signaling pathway (**Figure 2B; Supplementary Table 4**). Hypermethylated DM sites were significantly enriched for 146 GO terms, with top hits including actin cytoskeleton organization and 29 KEGG pathways including Rap1 signaling, Ras signaling, and phosphatidylinositol signaling system (**Figure 2C**). The most significant hypomethylated DM sites include CpGs that map to a predicted promoter region of *RGS14*, an intronic enhancer of *FGF18*, and *MGAT1*. Top



hypermethylated DM sites mapped to genes such as *LYST* (lysosomal trafficking regulator), which regulates protein trafficking to lysosomes.

We next overlapped DEGs and differentially methylated probes (DMPs) in sarcoidosis vs controls based on gene ID. 961 DEGs (52.2%) showed DM of 12,297 unique CpGs. 52.9% of associations (41,674/78,761) showed canonical inverse relationships: the directionality of DNAm was opposite that of gene expression (**Figure 2D**). Genes with canonical relationships and upregulated RNA include *TGFBI*. Canonical relationships with downregulated RNA include *AQP1*, *FRMD4A*, *ARNT2*, and *BMP4*. Genes with non-canonical relationships and downregulated RNA include *ITGA9*, an integrin component of receptor for *VCAM1*, *LIMCH1*, which positively regulates stress fiber assembly and stabilizes focal adhesions, and *FGF18*.

*miRNA expression differs between cases and controls and targets DEGs*

We next compared miRNA expression in sarcoidosis vs controls and identified 5 miRNAs (hsa-miR-143-3p, hsa-miR-199a-3p/hsa-miR-199b-3p, hsa-miR-199b-5p, hsa-miR-582-3p & hsa-miR-582-5p) downregulated in sarcoidosis (**Figure 3A-B; Supplementary Table 5**). No miRNAs were associated with progression. We derived experimentally validated target genes for each significant miRNA using MirTarBase<sup>23</sup>. We identified 67 target genes, of which 4 (*AKT1*, *CD44*, *JAG1*, *PTGS2*) are targeted by 2 DE miRNAs (**Figure 3C**), and a further 15 were DEGs in our mRNA analysis, including *TNF*.

*Integrated model reveals associations with sarcoidosis and with progression not found in individual analyses*

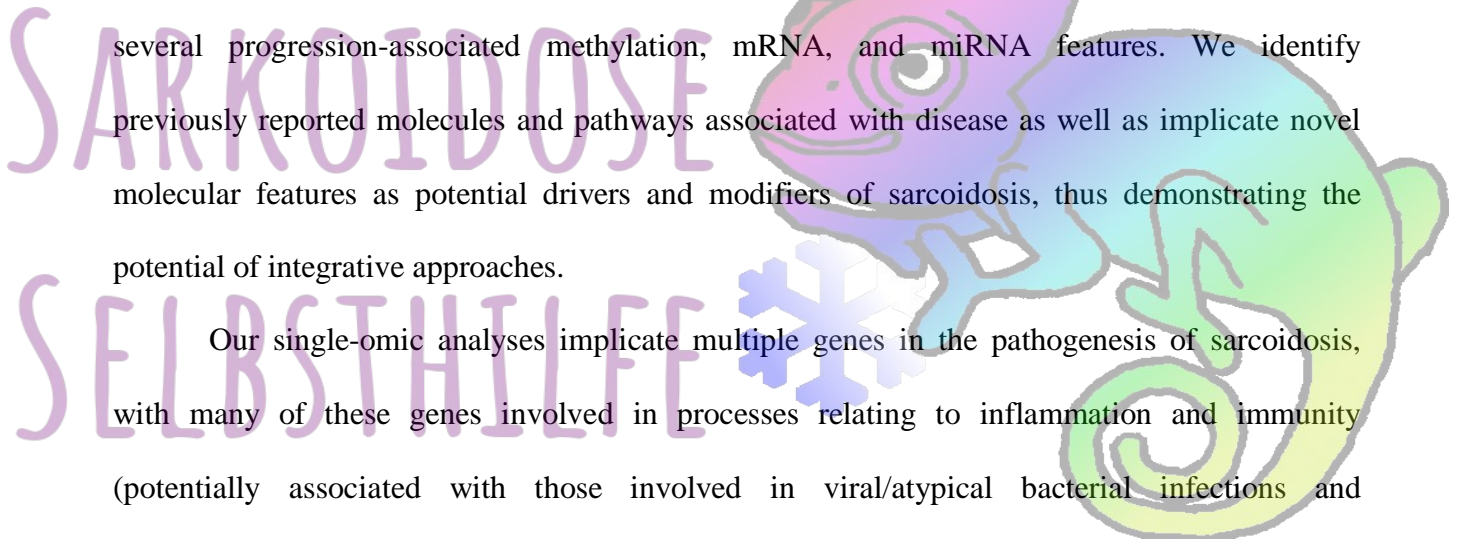
We used DIABLO to integrate the 3 datasets, including individuals present in all datasets after QC (n = 65; 13 controls, 19 progressive, 33 non-progressive sarcoidosis). With a goal to determine molecular features separating 3 groups (controls, progressive, and non-progressive

sarcoidosis), we constructed a multi-omic model with 2 latent variables, which are linear combinations of input features. Using 5-fold cross validation repeated 10 times, we determined that selecting 5 DNAm sites, 2 mRNAs, and 13 miRNAs for latent variable 1 and 2 DNAm sites, 2 mRNAs, and 14 miRNAs for latent variable 2 maximized the prediction accuracy of the model and resulted in clustering based on diagnosis (**Figure 4A**; **Supplementary Table 6**). Latent variable 1 separates controls from sarcoidosis samples and latent variable 2 separates progressive from non-progressive sarcoidosis. Features included in each latent variable are shown in **Figure 4B** and **Supplementary Table 7**. We further constructed networks based on correlations of feature weights for each latent variable (**Figure 4C**). The two mRNAs contributing to latent variable 1 were *SFTPB* and *SFTPD*. DNAm sites contributing to latent variable 1 include DNAm sites hypermethylated in cases cg16962115, within an intron of *LYST*, and cg05300241 within an intron of *PIK3CD*. The remaining 3 DNAm sites were hypomethylated and included cg21949194 within an enhancer of *SOS1* (SOS Ras/Rac guanine nucleotide exchange factor 1), cg11370586 within the predicted promoter of *RGS14*, and cg03526142 within an exon of *SLFN12L*. All DNAm sites contributing to latent variable 1 were DMPs in our previous methylation analysis. In addition to the DE miR hsa-miR-199b-5p, miRNA features on latent variable 1 include hsa-miR-204-5p. Features contributing to latent variable 2 include the DNAm sites cg05479174, within an exon of *SFSWAP*, and cg06635176 within an intron of *AGBL4*. *IL20RB* (interleukin 20 receptor subunit beta) mRNA was upregulated in non-progressive sarcoidosis, while *ABCC11* (ATP binding cassette subfamily C member 11) expression was higher in progressive sarcoidosis. miRNAs contributing to latent variable 1 include hsa-miR-146a-3p and hsa-miR-378b.

## Discussion

In this study, we present the first application of multiomic integration in sarcoidosis, leveraging coding and miRNA expression with DNA methylation data to construct a multiomic network. Through initial genome-wide profiling of DNA methylation, mRNA expression, and miRNA expression in sarcoidosis BAL, we identified 1,846 DEGs, 46,812 DMPs, and 5 DE miRNAs in sarcoidosis relative to healthy controls, as well as 1 DEG associated with sarcoidosis progression. By integrating omic datasets, we define pathogenic molecules/genes in sarcoidosis not identified using conventional modeling methods in single-omics datasets. While our single-omics approach only demonstrated 1 DEG for progression, the multiomic model identified several progression-associated methylation, mRNA, and miRNA features. We identify previously reported molecules and pathways associated with disease as well as implicate novel molecular features as potential drivers and modifiers of sarcoidosis, thus demonstrating the potential of integrative approaches.

Our single-omic analyses implicate multiple genes in the pathogenesis of sarcoidosis, with many of these genes involved in processes relating to inflammation and immunity (potentially associated with those involved in viral/atypical bacterial infections and autoimmunity) and extracellular matrix signaling. While we detected many associations with case status, we detected only a single omic association with progression: *SEPP1* mRNA levels. Results from the GRADS study demonstrated that *SEPP1* levels were inversely correlated with DLCO (%) and FVC (% pred) in sarcoidosis individuals<sup>6</sup>. These findings support a progressive phenotype, including lung function changes that we and others have used in the definition of progressive pulmonary disease.

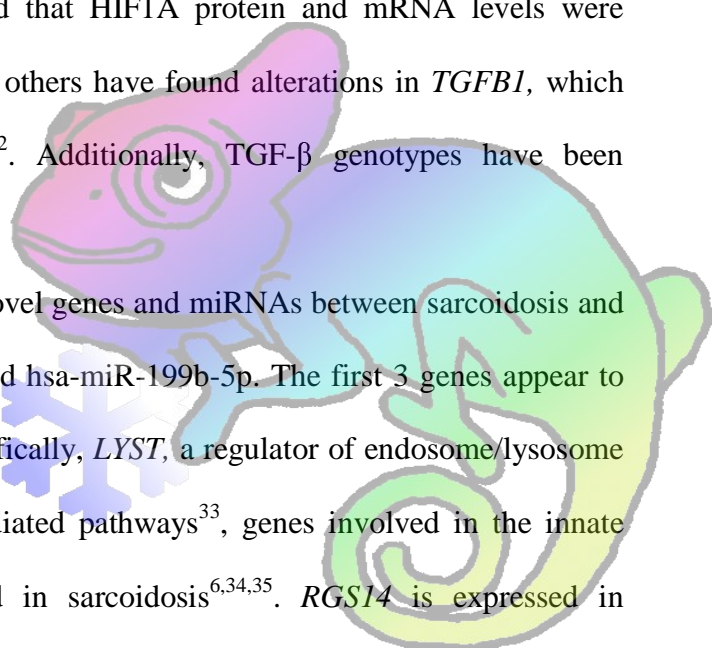


Our study also demonstrated 5 DE miRNAs downregulated in sarcoidosis cases in the first study of genome-wide miRNAs in sarcoidosis. While these associations are novel, multiple target genes of these miRNAs have been implicated in sarcoidosis previously, including the genes *AKT1*, *CD44*, *JAG1*, and *PTGS2*. For example, *CD44* has been found in areas of granuloma formation and fibrosis<sup>28</sup> and is differentially expressed between Lofgren syndrome versus non-Lofgren syndrome subjects<sup>29</sup>. Through integrating genes with evidence of both DE and DM, we identified biologically relevant genes in sarcoidosis versus controls, including *HIF1A* and *TGFBI*. A previous study showed that *HIF1A* protein and mRNA levels were decreased in sarcoidosis granulomas<sup>30</sup>. We and others have found alterations in *TGFBI*, which encodes TGF- $\beta$ , in pulmonary sarcoidosis<sup>31,32</sup>. Additionally, TGF- $\beta$  genotypes have been associated with sarcoidosis severity<sup>31</sup>.

Our integrated model revealed several novel genes and miRNAs between sarcoidosis and controls, including *LYST*, *RGS14*, *SLFN12L*, and hsa-miR-199b-5p. The first 3 genes appear to have important roles in immune function. Specifically, *LYST*, a regulator of endosome/lysosome trafficking, can regulate TLR3 and TLR4 mediated pathways<sup>33</sup>, genes involved in the innate immune system which have been implicated in sarcoidosis<sup>6,34,35</sup>. *RGS14* is expressed in lymphocytes and regulates chemokine receptors to control immune responses to exogenous agents<sup>36</sup>. Finally, *SLFN12L* regulates thymocyte development and is downregulated in T-cell activation, suggesting a role as an immune response regulator<sup>37</sup>.

Our integrated analyses demonstrated miRNAs DE between progressive/non-progressive sarcoidosis including those previously described. Specifically, miR-146a-3p upregulated in progressive sarcoidosis is an indicator of inflammation and oxidative stress that may target *TLR4* and was previously found elevated in sarcoidosis BALF, as well as serum<sup>38,39</sup>. miR-378b, which

SARKOIDOSE  
SELBSTHILFE





was upregulated in progressive cases, was previously found associated with sarcoidosis<sup>40</sup>. We also identified novel genes including *IL20RB*, *ABCC11*, *SFSWAP* and *AGBL4*. Interestingly, *IL20RB* expression was increased in non-progressive versus progressive sarcoidosis in our study, suggesting that the activation of this inflammatory pathway is increased in the non-progressive phenotype. *ABCC11* is a gene that influences macrophage differentiation and induces TNF- $\alpha$  and IL17 through TLR4 signaling<sup>41</sup>; these results as well as the novel findings above support the importance of innate immune response genes in sarcoidosis.

Both our individual and integrative analyses identify molecules involved in PI3K/AKT1 signaling, a pathway already recognized in sarcoidosis pathogenesis. For example, we observe hypermethylation of *PI3KCD* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta), which encodes a component of PI3K, and complexes with AKT1 to impact T cell differentiation and function<sup>42</sup>. Our miRNA results demonstrate *AKT1* as a target of downregulated miRNAs (miR-143-3p and miR-199a/miR-199b) in sarcoidosis. Both downregulation and upregulation of the PI3K/AKT signaling pathways have been associated with sarcoidosis<sup>42,43</sup>. Previous studies demonstrated both reduced proliferative response and exhaustion of T cells in progressive sarcoidosis is thought to be driven in part by inhibition of PI3K/AKT1 signaling<sup>42</sup>. The PI3K/AKT1 pathway has also been implicated in activation of mTOR<sup>43</sup>, which has been associated with granuloma formation<sup>44</sup>. Interestingly, in the recent GRADS study, PI3K activation was associated with an endotype of sarcoidosis characterized by hilar lymphadenopathy, pulmonary reticulation, less multiorgan involvement, and more environmental associations<sup>6</sup>.

While our sample size is larger than most previous sarcoidosis omic studies, power in our analyses was limited, especially in our phenotype analyses. In future studies, it will be important

to both increase sample sizes and investigate associations with better-powered quantitative measures related to disease such as pulmonary function testing variables. We were unable to correct for cell proportions in sarcoidosis subjects relative to controls as our controls lacked cell differentials. Altered cell proportions may explain the downregulated epithelial genes in sarcoidosis BAL (e.g. decreased surfactant proteins in sarcoidosis) and some of the DNAm sites and DEGs detected in sarcoidosis versus control comparisons, although in general epithelial cells are not a large proportion of BAL. Finally, our multiomic models did not take into account demographic information. Regressing out demographic information such as age and sex may lead to more predictive networks. Despite these shortcomings, we detected widespread molecular changes and constructed multiomic networks associated with sarcoidosis as well as progression. Molecules discovered in these analyses shed light on disease pathogenesis and may also be leveraged therapeutically or as biomarkers after replication/validation in additional populations.

SARKOIDOSE

SELBSTHILFE

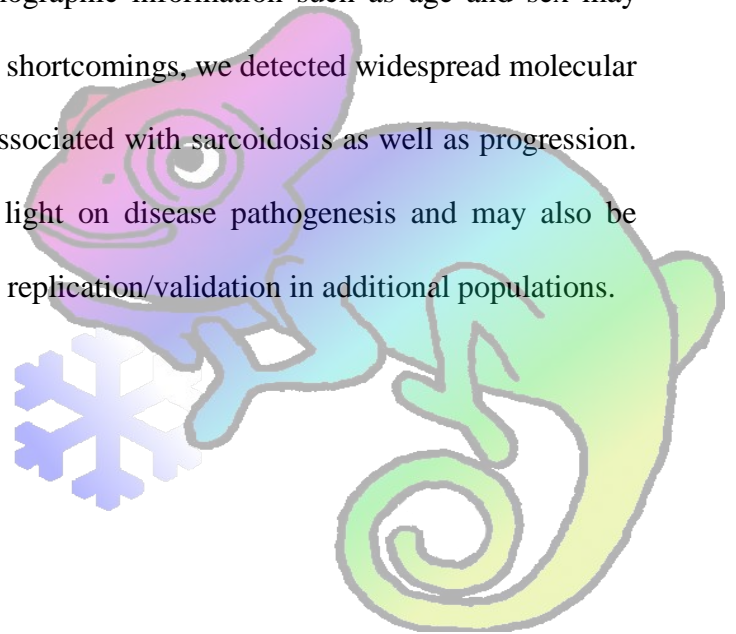


Table 1. Demographic information of control, progressive sarcoidosis, and non-progressive sarcoidosis cases at time of bronchoscopy with lavage.

	Control	Non-Progressive	Progressive	<i>p</i>
n	16	38	26	
Age (mean (SD))	55.62 (9.74)	51.68 (9.88)	51.19 (9.61)	0.316*
Sex = Male (%)	11 (68.8)	11 (28.9)	16 (61.5)	0.006**
Race (%)				0.242**
Asian	1 (6.2)	1 (2.6)	0 (0.0)	
Black	0 (0.0)	6 (15.8)	5 (19.2)	
White	15 (93.8)	31 (81.6)	21 (80.8)	
Ethnicity = Non-Hispanic (%)	15 (93.8)	36 (94.7)	25 (100.0)	0.441**
Smoking Status = Former (%)	0	12 (31.6)	6 (23.1)	0.0275**
FVC (mean (SD))		97.42 (11.25)	85.92 (14.21)	0.001 <sup>#</sup>
FEV <sub>1</sub> (mean (SD))		101.03 (12.19)	86.40 (15.04)	<0.001*
DLCO (mean (SD))		87.71 (21.81)	84.42 (12.64)	0.491*
Scadding Stage (%)				<0.001**
0		13 (34.2)	1 (3.8)	
1		20 (52.6)	1 (3.8)	
2		4 (10.5)	22 (84.6)	
3		1 (2.6)	2 (7.7)	

\*One-way ANOVA, \*\*Fisher's exact test, <sup>#</sup>Mann-Whitney U test. FVC: forced vital capacity. FEV<sub>1</sub>: forced expiratory volume over 1 second. DLCO: diffusing capacity of the lung for carbon monoxide.

Figure 1. Differentially expressed genes in sarcoidosis. A) Differentially expressed genes in sarcoidosis vs. controls. B) Differentially expressed genes in progressive vs. non-progressive sarcoidosis. C) Pathway enrichment of upregulated mRNAs. D) Pathway enrichment of downregulated mRNAs.

Figure 2. Differentially methylated sites in sarcoidosis. A) Differentially methylated sites in sarcoidosis vs. controls. B) Pathway enrichment of hypomethylated DNAm sites. C) Pathway enrichment of hypermethylated DNAm sites. D) Differentially expressed genes with DM DNAm sites.

Figure 3. Differentially expressed microRNAs in sarcoidosis. A) Differentially expressed miRNAs in sarcoidosis vs. controls. B) Distribution of significant miRNAs' expression in cases and controls. C) Sankey plot connecting DE miRNAs to target genes targeted by >1 DE miRNA. Connection width represents number of sources confirming relationship.

Figure 4. A sparse multi-omic model of sarcoidosis and progression. A) Projection of samples based on selected feature weights. B) Weights of selected features for latent variables. C) Network constructed from feature correlations on latent variables.

Supplementary Figure 1. Accounting for RNA-Sequencing batch effect. A) Principal component analysis of mRNA data, colored by sequencing location. B) Surrogate variable analysis of mRNA data, colored by sequencing location. C) Principal component analysis of mRNA data after regressing out SV1.

Supplementary Table 1. Differentially expressed genes associated with sarcoidosis.

Supplementary Table 2. Pathway over-representation analysis of DEGs.

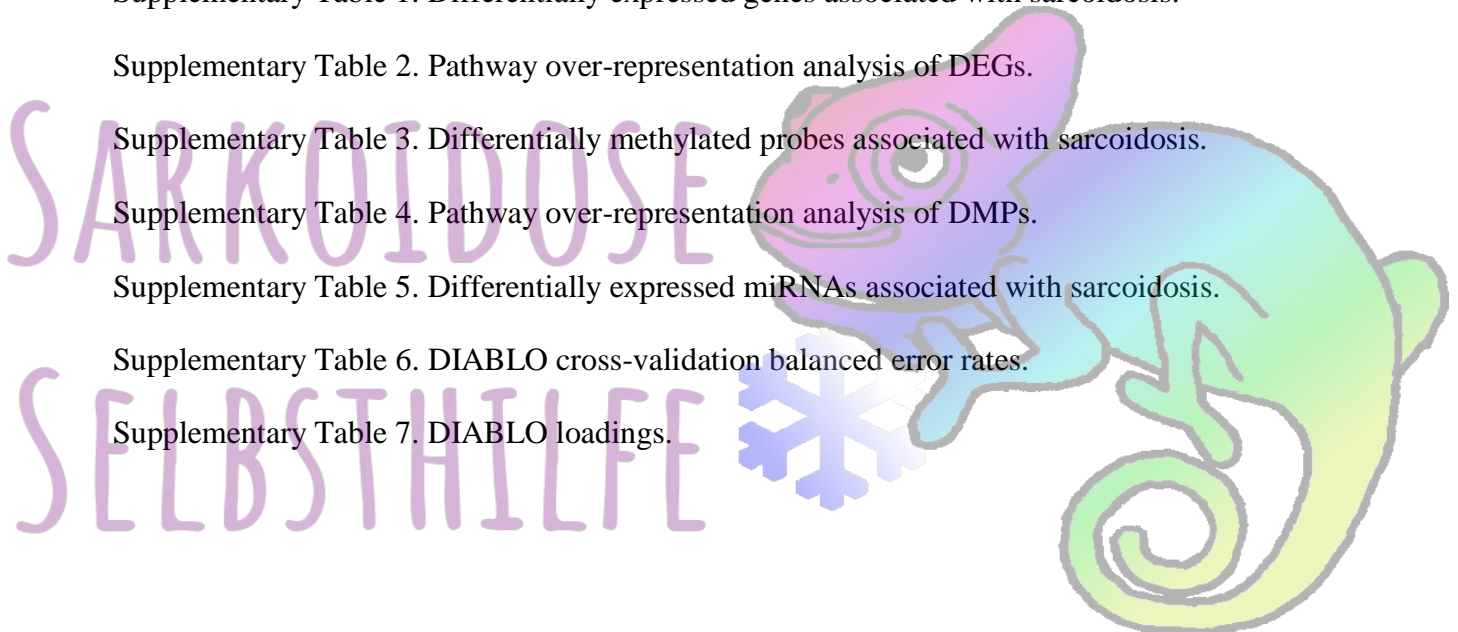
Supplementary Table 3. Differentially methylated probes associated with sarcoidosis.

Supplementary Table 4. Pathway over-representation analysis of DMPs.

Supplementary Table 5. Differentially expressed miRNAs associated with sarcoidosis.

Supplementary Table 6. DIABLO cross-validation balanced error rates.

Supplementary Table 7. DIABLO loadings.





## References

1. Baughman, R.P., Field, S., Costabel, U., Crystal, R.G., Culver, D.A., Drent, M., Judson, M.A., and Wolff, G. (2016). Sarcoidosis in America. Analysis Based on Health Care Use. *Annals of the American Thoracic Society* *13*, 1244-1252. 10.1513/AnnalsATS.201511-760OC.
2. Baughman, R.P., Teirstein, A.S., Judson, M.A., Rossman, M.D., Yeager, H., Bresnitz, E.A., DePalo, L., Hunninghake, G., Iannuzzi, M.C., Johns, C.J., et al. (2001). Clinical Characteristics of Patients in a Case Control Study of Sarcoidosis. *American Journal of Respiratory and Critical Care Medicine* *164*, 1885-1889. 10.1164/ajrccm.164.10.2104046.
3. Swigris, J.J., Olson, A.L., Huie, T.J., Fernandez-Perez, E.R., Solomon, J., Sprunger, D., and Brown, K.K. (2011). Sarcoidosis-related Mortality in the United States from 1988 to 2007. *American Journal of Respiratory and Critical Care Medicine* *183*, 1524-1530. 10.1164/rccm.201010-1679OC.
4. Gerke, A.K., Judson, M.A., Cozier, Y.C., Culver, D.A., and Koth, L.L. (2017). Disease Burden and Variability in Sarcoidosis. *Annals of the American Thoracic Society* *14*, S421-S428. 10.1513/AnnalsATS.201707-564OT.
5. Asano, M., Minagawa, T., Ohmichi, M., and Hiraga, Y. (1991). Detection of endogenous cytokines in sera or in lymph nodes obtained from patients with sarcoidosis. *Clin Exp Immunol* *84*, 92-96. 10.1111/j.1365-2249.1991.tb08129.x.
6. Vukmirovic, M., Yan, X., Gibson, K.F., Gulati, M., Schupp, J.C., DeLulii, G., Adams, T.S., Hu, B., Mihaljinec, A., Woolard, T.N., et al. (2021). Transcriptomics of bronchoalveolar lavage cells identifies new molecular endotypes of sarcoidosis. *Eur Respir J* *58*. 10.1183/13993003.02950-2020.
7. Yang, I.V., Konigsberg, I., MacPhail, K., Li, L., Davidson, E.J., Mroz, P.M., Hamzeh, N., Gillespie, M., Silveira, L.J., Fingerlin, T.E., and Maier, L.A. (2019). DNA Methylation Changes in Lung Immune Cells Are Associated with Granulomatous Lung Disease. *American Journal of Respiratory Cell and Molecular Biology* *60*, 96-105. 10.1165/rcmb.2018-0177OC.
8. Li, L., Konigsberg, I.R., Bhargava, M., Liu, S., MacPhail, K., Mayer, A., Davidson, E.J., Liao, S.Y., Lei, Z., Mroz, P.M., et al. (2022). Multiomic Signatures of Chronic Beryllium Disease Bronchoalveolar Lavage Cells Relate to T-Cell Function and Innate Immunity. *Am J Respir Cell Mol Biol* *67*, 632-640. 10.1165/rcmb.2022-0077OC.
9. Ascoli, C., Huang, Y., Schott, C., Turturice, B.A., Metwally, A., Perkins, D.L., Finn, P.W., and Group, A.R. (2018). A Circulating MicroRNA Signature Serves as a Diagnostic and Prognostic Indicator in Sarcoidosis. *Am J Respir Cell Mol Biol* *58*, 40-54. 10.1165/rcmb.2017-0207OC.
10. Zhou, T., Casanova, N., Pouladi, N., Wang, T., Lussier, Y., Knox, K.S., and Garcia, J.G.N. (2017). Identification of Jak-STAT signaling involvement in sarcoidosis severity via a novel microRNA-regulated peripheral blood mononuclear cell gene signature. *Sci Rep* *7*, 4237. 10.1038/s41598-017-04109-6.
11. Kiszalkiewicz, J., Piotrowski, W.J., Pastuszak-Lewandoska, D., Gorski, P., Antczak, A., Gorski, W., Domanska-Senderowska, D., Migdalska-Sek, M., Czarnecka, K.H., Nawrot,

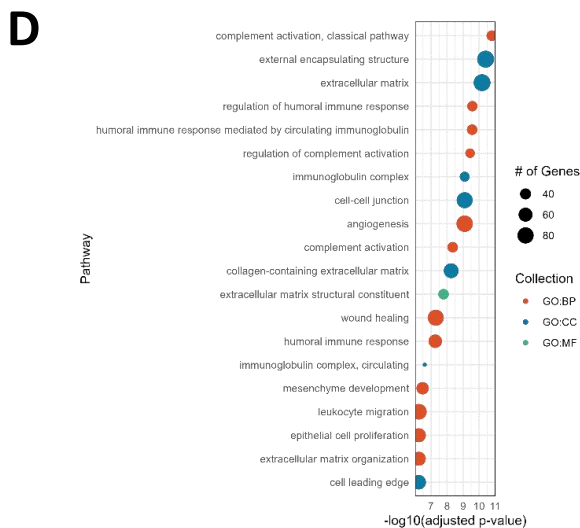
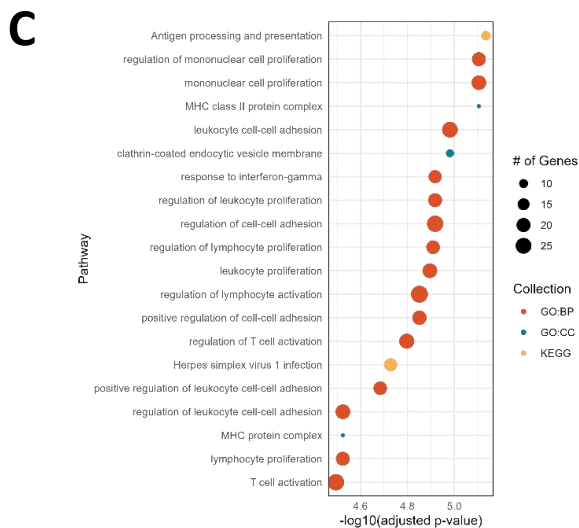
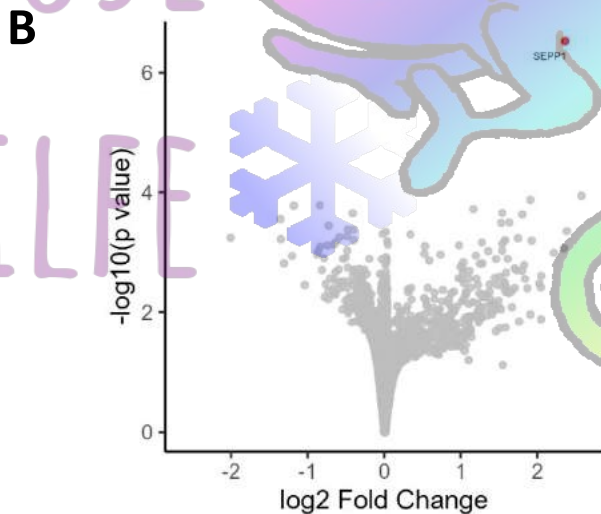
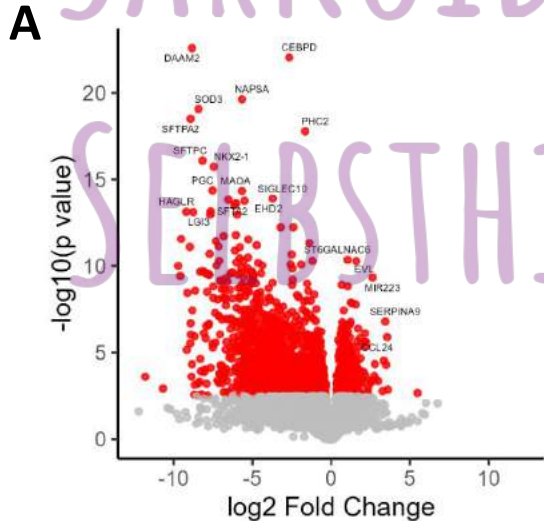
- E., and Brzezianska-Lasota, E. (2016). Altered miRNA expression in pulmonary sarcoidosis. *BMC Med Genet* 17, 2. 10.1186/s12881-016-0266-6.
12. Moller, D.R., Koth, L.L., Maier, L.A., Morris, A., Drake, W., Rossman, M., Leader, J.K., Collman, R.G., Hamzeh, N., Sweiss, N.J., et al. (2015). Rationale and Design of the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) Study. Sarcoidosis Protocol. *Ann Am Thorac Soc* 12, 1561-1571. 10.1513/AnnalsATS.201503-172OT.
  13. Crouser, E.D., Maier, L.A., Wilson, K.C., Bonham, C.A., Morgenthau, A.S., Patterson, K.C., Abston, E., Bernstein, R.C., Blankstein, R., Chen, E.S., et al. (2020). Diagnosis and Detection of Sarcoidosis. An Official American Thoracic Society Clinical Practice Guideline. *American Journal of Respiratory and Critical Care Medicine* 201, e26-e51. 10.1164/rccm.202002-0251ST.
  14. Maier, L.A., Sawyer, R.T., Tinkle, S.S., Kittle, L.A., Barker, E.A., Balkissoon, R., Rose, C., and Newman, L.S. (2001). IL-4 fails to regulate in vitro beryllium-induced cytokines in berylliosis. *Eur Respir J* 17, 403-415. 10.1183/09031936.01.17304030.
  15. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882-883. 10.1093/bioinformatics/bts034.
  16. van Iterson, M., van Zwet, E.W., and Heijmans, B.T. (2017). Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biology* 18. 10.1186/s13059-016-1131-9.
  17. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289-300.
  18. Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E., et al. (2021). The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Research* 49, D325-D334. 10.1093/nar/gkaa1113.
  19. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research* 49, D545-D551. 10.1093/nar/gkaa970.
  20. Maksimovic, J., Oshlack, A., and Phipson, B. (2021). Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biology* 22. 10.1186/s13059-021-02388-x.
  21. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2. 10.1016/j.xinn.2021.100141.
  22. Cavalcante, R.G., and Sartor, M.A. (2017). annotatr: genomic regions in context. *Bioinformatics* 33, 2381-2383. 10.1093/bioinformatics/btx183.
  23. Huang, H.Y., Lin, Y.C., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J., Zuo, H., et al. (2021). miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* 10.1093/nar/gkab1079.
  24. Singh, A., Shannon, C.P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S.J., and Le Cao, K.A. (2019). DIABLO: an integrative approach for identifying key molecular drivers

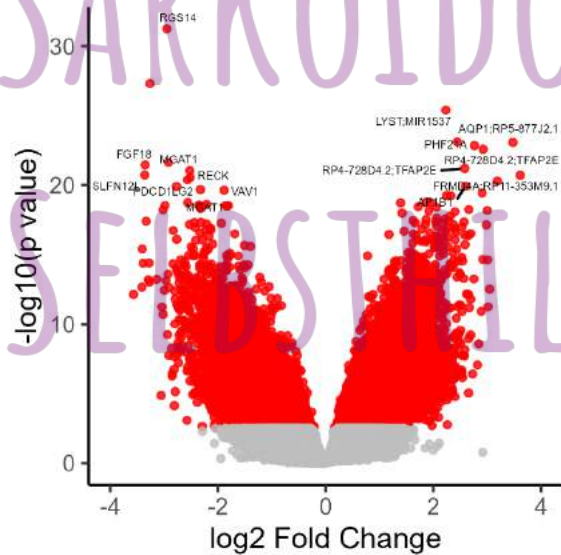
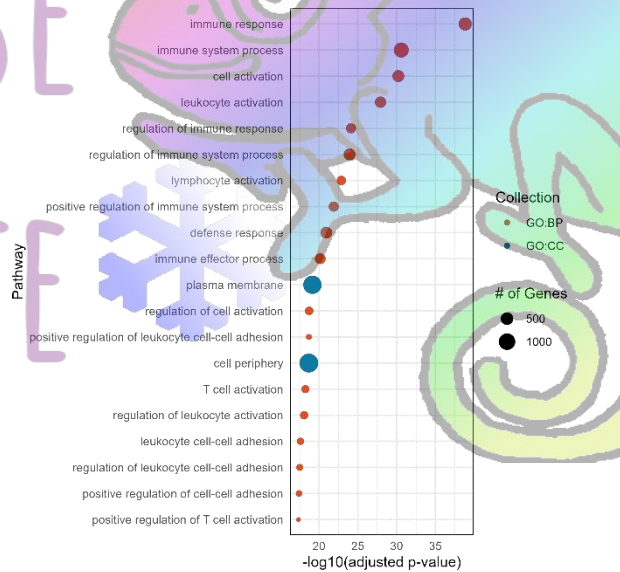
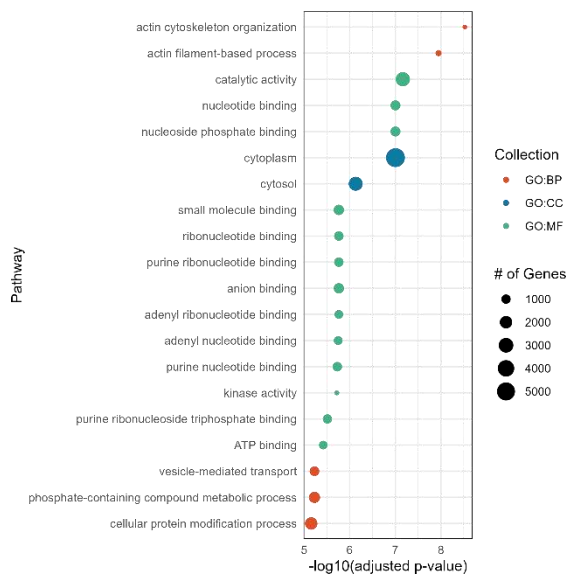
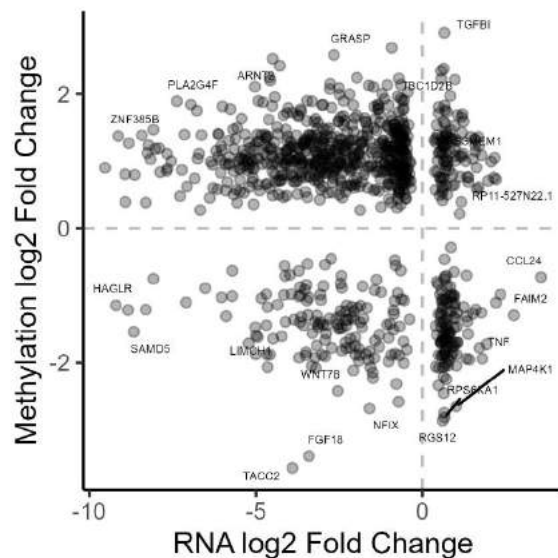
- from multi-omics assays. *Bioinformatics* 35, 3055-3062. 10.1093/bioinformatics/bty1054.
25. Schneidman, D., Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology* 13. 10.1371/journal.pcbi.1005752.
  26. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. 10.1186/s13059-014-0550-8.
  27. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106. 10.1186/gb-2010-11-10-r106.
  28. Aleksoniene, R., Besusparis, J., Gruslys, V., Jurgauskiene, L., Laurinaviciene, A., Laurinavicius, A., Malickaite, R., Norkuniene, J., Zablockis, R., Zurauskas, E., and Danila, E. (2021). CD31(+), CD38(+), CD44(+), and CD103(+) lymphocytes in peripheral blood, bronchoalveolar lavage fluid and lung biopsy tissue in sarcoid patients and controls. *J Thorac Dis* 13, 2300-2318. 10.21037/jtd-20-2396.
  29. Kaiser, Y., Lakshmikanth, T., Chen, Y., Mikes, J., Eklund, A., Brodin, P., Achour, A., and Grunewald, J. (2017). Mass Cytometry Identifies Distinct Lung CD4(+) T Cell Patterns in Lofgren's Syndrome and Non-Lofgren's Syndrome Sarcoidosis. *Front Immunol* 8, 1130. 10.3389/fimmu.2017.01130.
  30. Tzouvelekis, A., Ntoliou, P., Karameris, A., Koutsopoulos, A., Boglou, P., Koulelidis, A., Archontogeorgis, K., Zacharis, G., Drakopanagiotakis, F., Steiropoulos, P., et al. (2012). Expression of hypoxia-inducible factor (HIF)-1 $\alpha$ -vascular endothelial growth factor (VEGF)-inhibitory growth factor (ING)-4- axis in sarcoidosis patients. *BMC Res Notes* 5, 654. 10.1186/1756-0500-5-654.
  31. Jonth, A.C., Silveira, L., Fingerlin, T.E., Sato, H., Luby, J.C., Welsh, K.I., Rose, C.S., Newman, L.S., du Bois, R.M., Maier, L.A., and Group, A. (2007). TGF-beta 1 variants in chronic beryllium disease and sarcoidosis. *J Immunol* 179, 4255-4262. 10.4049/jimmunol.179.6.4255.
  32. Kobak, S., Akyildiz, M., Gokduman, A., Atabay, T., and Vural, H. (2021). Serum galectin-3 and TGF-beta levels in patients with sarcoidosis. *Reumatol Clin (Engl Ed)* 17, 562-565. 10.1016/j.reumae.2020.08.005.
  33. Westphal, A., Cheng, W., Yu, J., Grassl, G., Krautkramer, M., Holst, O., Foger, N., and Lee, K.H. (2017). Lysosomal trafficking regulator Lyst links membrane trafficking to toll-like receptor-mediated inflammatory responses. *J Exp Med* 214, 227-244. 10.1084/jem.20141461.
  34. Wiken, M., Grunewald, J., Eklund, A., and Wahlstrom, J. (2009). Higher monocyte expression of TLR2 and TLR4, and enhanced pro-inflammatory synergy of TLR2 with NOD2 stimulation in sarcoidosis. *J Clin Immunol* 29, 78-89. 10.1007/s10875-008-9225-0.
  35. Schurmann, M., Kwiatkowski, R., Albrecht, M., Fischer, A., Hampe, J., Muller-Quernheim, J., Schwinger, E., and Schreiber, S. (2008). Study of Toll-like receptor gene loci in sarcoidosis. *Clin Exp Immunol* 152, 423-431. 10.1111/j.1365-2249.2008.03621.x.
  36. Moratz, C., Harrison, K., and Kehrl, J.H. (2004). Regulation of chemokine-induced lymphocyte migration by RGS proteins. *Methods Enzymol* 389, 15-32. 10.1016/S0076-6879(04)89002-5.

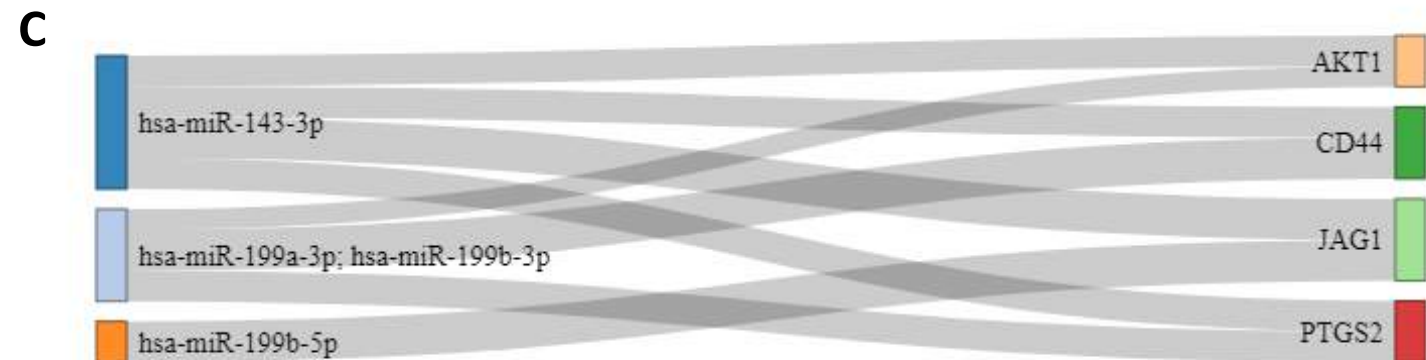
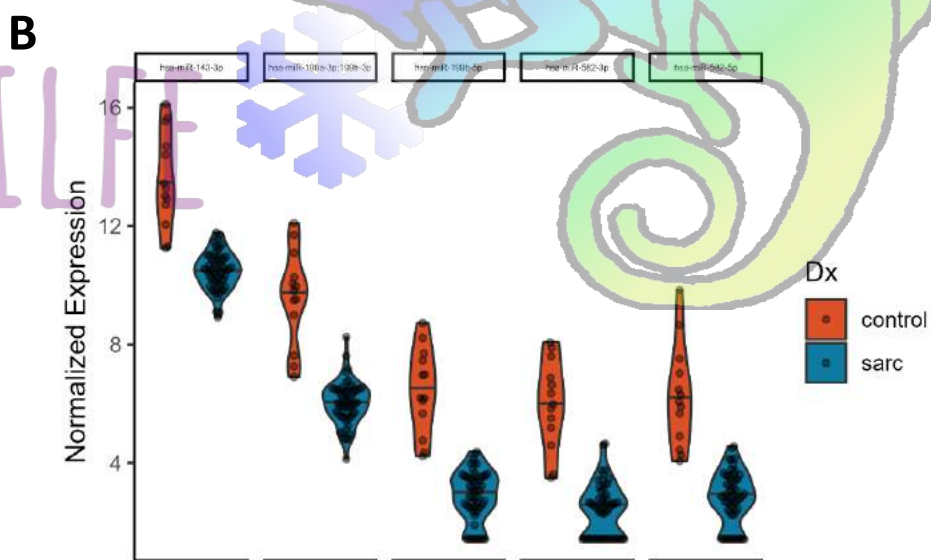
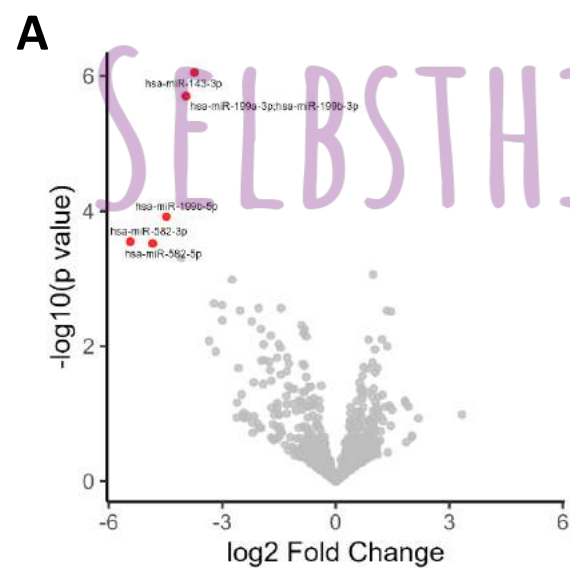


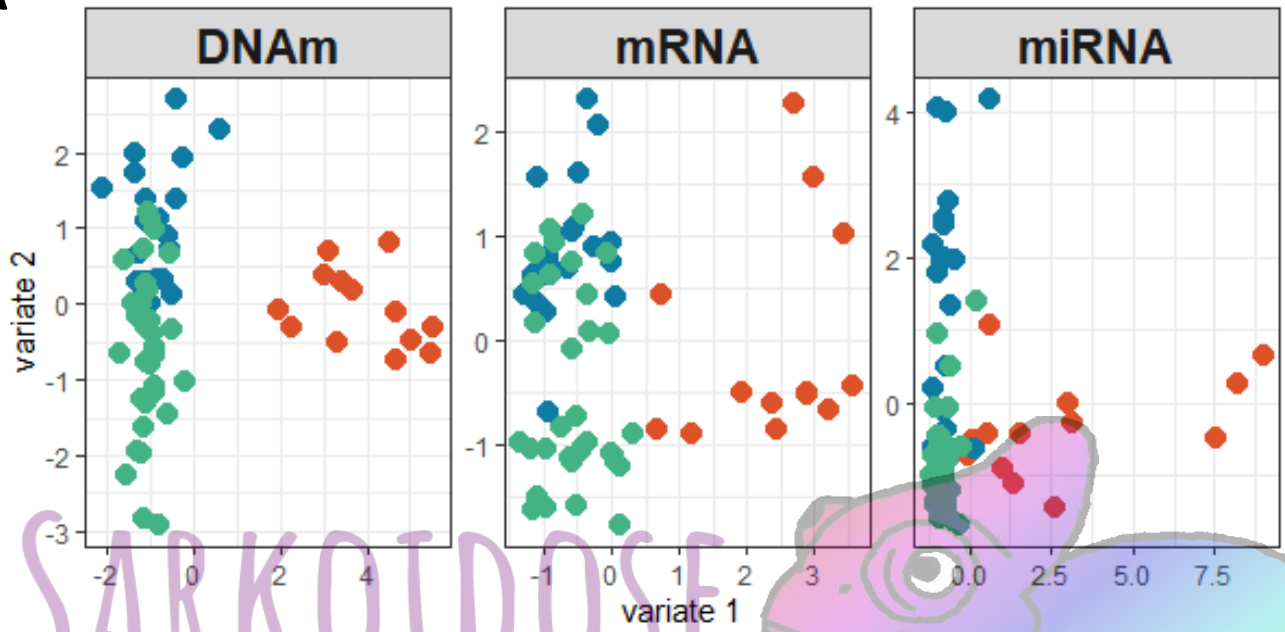
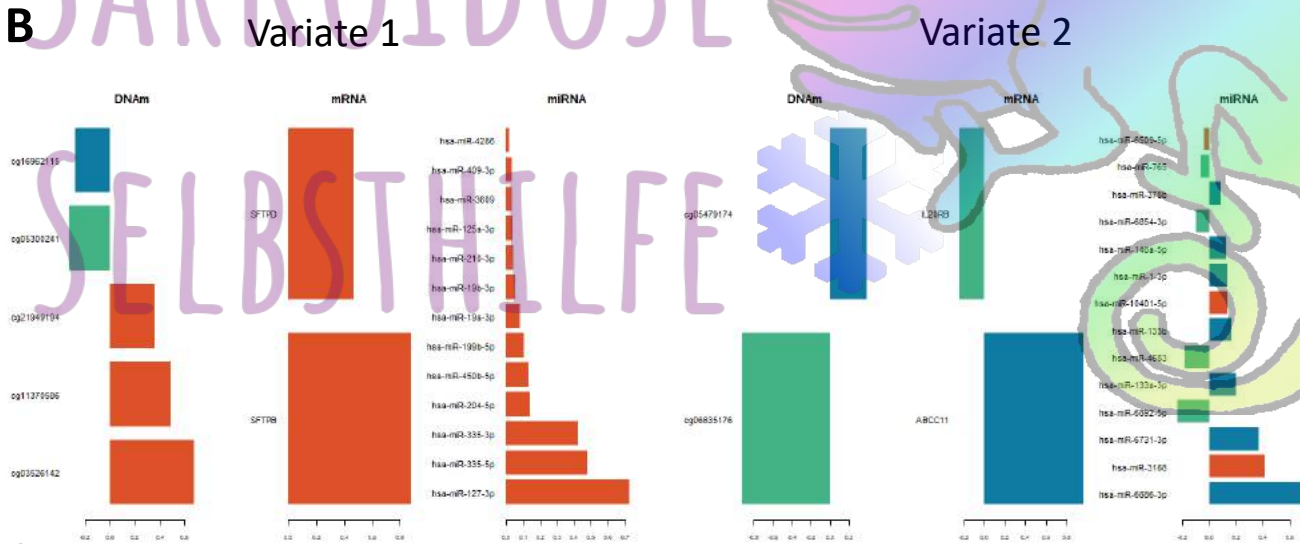
37. Puck, A., Aigner, R., Modak, M., Cejka, P., Blaas, D., and Stockl, J. (2015). Expression and regulation of Schlafen (SLFN) family members in primary human monocytes, monocyte-derived dendritic cells and T cells. *Results Immunol* 5, 23-32. 10.1016/j.rinim.2015.10.001.
38. Dyskova, T., Fillerova, R., Novosad, T., Kudelka, M., Zurkova, M., Gajdos, P., Kolek, V., and Kriegova, E. (2015). Correlation Network Analysis Reveals Relationships between MicroRNAs, Transcription Factor T-bet, and Deregulated Cytokine/Chemokine-Receptor Network in Pulmonary Sarcoidosis. *Mediators Inflamm* 2015, 121378. 10.1155/2015/121378.
39. Novosadova, E., Chabronova, A., Kolek, V., Petrek, M., and Navratilova, Z. (2016). The Serum Expression of Selected miRNAs in Pulmonary Sarcoidosis with/without Lofgren's Syndrome. *Mediators Inflamm* 2016, 1246129. 10.1155/2016/1246129.
40. Jazwa, A., Kasper, L., Bak, M., Sobczak, M., Szade, K., Jozkowicz, A., Sladek, K., and Dulak, J. (2015). Differential inflammatory microRNA and cytokine expression in pulmonary sarcoidosis. *Arch Immunol Ther Exp (Warsz)* 63, 139-146. 10.1007/s00005-014-0315-9.
41. Vogl, T., Tenbrock, K., Ludwig, S., Leukert, N., Ehrhardt, C., van Zoelen, M.A., Nacken, W., Foell, D., van der Poll, T., Sorg, C., and Roth, J. (2007). Mrp8 and Mrp14 are endogenous activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nat Med* 13, 1042-1049. 10.1038/nm1638.
42. Celada, L.J., Rotsinger, J.E., Young, A., Shaginurova, G., Shelton, D., Hawkins, C., and Drake, W.P. (2017). Programmed Death-1 Inhibition of Phosphatidylinositol 3-Kinase/AKT/Mechanistic Target of Rapamycin Signaling Impairs Sarcoidosis CD4(+) T Cell Proliferation. *Am J Respir Cell Mol Biol* 56, 74-82. 10.1165/rcmb.2016-0037OC.
43. Bhargava, M., Viken, K.J., Barkes, B., Griffin, T.J., Gillespie, M., Jagtap, P.D., Sajulga, R., Peterson, E.J., Dincer, H.E., Li, L., et al. (2020). Novel protein pathways in development and progression of pulmonary sarcoidosis. *Sci Rep* 10, 13282. 10.1038/s41598-020-69281-8.
44. Linke, M., Pham, H.T., Katholnig, K., Schnoller, T., Miller, A., Demel, F., Schutz, B., Rosner, M., Kovacic, B., Sukhbaatar, N., et al. (2017). Chronic signaling via the metabolic checkpoint kinase mTORC1 induces macrophage granuloma formation and marks sarcoidosis progression. *Nat Immunol* 18, 293-302. 10.1038/ni.3655.





**A****B****C****D**



**A****B****C**