



- (51) **International Patent Classification:**
G01N 33/569 (2006.01) *C12Q 1/68* (2006.01)
- (21) **International Application Number:**
PCT/US2016/021035
- (22) **International Filing Date:**
4 March 2016 (04.03.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/128,436 4 March 2015 (04.03.2015) US
- (71) **Applicants:** **WAYNE STATE UNIVERSITY** [US/US]; 5057 Woodward Avenue, Detroit, Michigan 48202 (US). **HENRY FORD HEALTH SYSTEM** [US/US]; 1 Ford Place, Detroit, Michigan 48202 (US).
- (72) **Inventors:** **SAMAVATI, Lobelia**; 5057 Woodward Avenue, Detroit, Michigan 48202 (US). **TALWAR, Harvinder S.**; 5057 Woodward Avenue, Detroit, Michigan 48202 (US). **ROSATI, Rita**; 5057 Woodward Avenue, Detroit, Michigan 48202 (US). **FERNANDEZ-MADRID, Felix**; 5057 Woodward Avenue, Detroit, Michigan 48202 (US). **LI, Jia**; 1 Ford Place, Detroit, Michigan 48202 (US).
- (74) **Agents:** **WINGER, C. Rachal et al.**; Lee & Hayes PLLC, 601 W. Riverside Avenue, Suite 1400, Spokane, Washington 99201 (US).

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

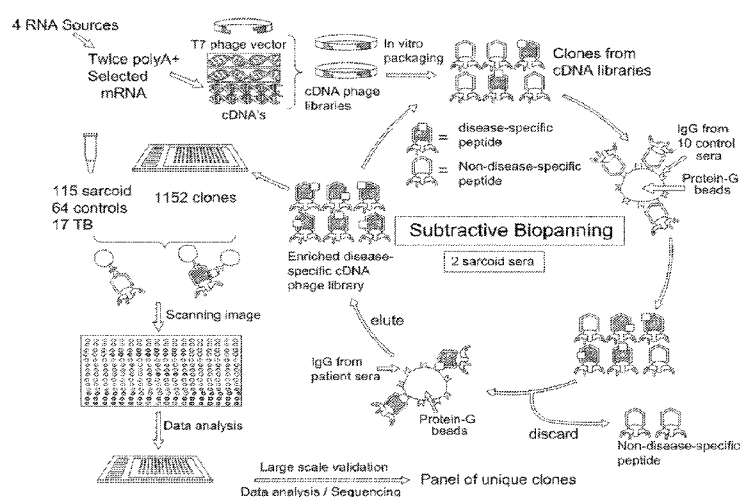
Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

[Continued on next page]

(54) **Title:** SYSTEMS AND METHODS TO DIAGNOSE SARCOIDOSIS AND IDENTIFY MARKERS OF THE CONDITION

FIG. 1



(57) **Abstract:** Systems and methods to diagnose sarcoidosis are described. In addition to diagnosing sarcoidosis, the systems and methods can distinguish sarcoidosis from tuberculosis. Further disclosed is a cDNA library and methods of its use for reliably identifying sarcoidosis markers.



Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

SYSTEMS AND METHODS TO DIAGNOSE SARCOIDOSIS AND IDENTIFY MARKERS OF THE CONDITION

CROSS REFERENCE TO RELATED APPLICATION

[0001] The present application claims the benefit of U.S. Provisional Patent Application Number 62/128,436, filed on March 4, 2015, which is incorporated herein by reference in its entirety as if fully set forth herein.

STATEMENT OF GOVERNMENT INTEREST

[0002] This invention was made in part using funds from United States federal grant number R21 HL104481-01A1 from the National Institutes of Health. Therefore, the United States government has certain rights in the invention.

FIELD OF THE DISCLOSURE

[0003] The current disclosure provides systems and methods to diagnose sarcoidosis. In addition to diagnosing sarcoidosis, the systems and methods can distinguish sarcoidosis from tuberculosis. Further disclosed is a cDNA library and methods of its use for reliably identifying sarcoidosis markers.

BACKGROUND OF THE DISCLOSURE

[0004] Sarcoidosis, also called sarcoid, is a disease involving abnormal collections of inflammatory cells (granulomas) that can form as nodules in multiple organs. The granulomas are most often located in the lungs or its associated lymph nodes. The disease seems to be caused by an immune reaction to an infection or some other trigger.

[0005] Diagnosis of sarcoidosis is challenging as the signs and symptoms of the condition are very broad, sometimes mimicking symptoms of other diseases. Further, symptoms can vary widely according to the organ system affected by the disorder. This variance can lead to a delay in diagnosis, or inappropriate treatment, therefore demonstrating a need for improved sarcoidosis diagnostic techniques.

[0006] The symptoms of sarcoidosis can also particularly resemble those caused by infection with tuberculosis. Thus, ability of a diagnostic to reliably distinguish between sarcoidosis and tuberculosis infection would allow faster treatment of each condition, resulting in better treatment outcomes.

SUMMARY OF THE DISCLOSURE

[0007] The present disclosure provides systems and methods to diagnose sarcoidosis in a subject. The systems and methods can distinguish a sarcoidosis subject from a healthy subject and/or a subject having tuberculosis. The systems and methods include diagnostic kits. The systems and methods also include a cDNA library to identify markers for sarcoidosis or tuberculosis diagnosis as well as methods of using the cDNA library to identify such markers, among others.

BRIEF DESCRIPTION OF THE FIGURES

[0008] FIG. 1 depicts a schematic diagram of identification of sarcoidosis antigens. The process of combining phage-display technology, protein microarray and bioinformatics tools to select a panel of novel clones for the diagnosis of sarcoidosis was used. A cDNA library was constructed from a pool of total RNA isolated from 20 bronchoalveolar cells (BAL) samples and 36 white blood cell (WBC) samples from sarcoid patients, and then combined with RNA extracts from cultured human monocytes and human embryonic fibroblasts. After digestion, the cDNA library was inserted into the T7 phage vector and packaged into T7 phages to generate a sarcoid cDNA-phage-display library. Several rounds of biopannings of the library were performed with pooled control sera for negative selection, and with sarcoid sera for positive enrichments. After four rounds of biopanning, enriched sarcoid specific peptide clones were cultured onto LB agar plates. A total of 1152 single colonies, including positive and negative clones were randomly picked and propagated into 96-well plates. Phage-clone lysates were then printed robotically onto coated glass slides to create a sarcoid-phage- protein microarray. Cy5 (red fluorescent dye)-labeled antihuman antibody was used to detect IgGs in human serum that were reactive to peptide clones, and a Cy3 (green fluorescent dye)-labeled antibody was used to detect the phage capsid protein in order to normalize for spotting. Thus, if a phage clone carried a peptide reactive to human IgG, the spot remains green suggesting an unreactive clone. A total of 115 sarcoid sera, 64 healthy control sera and 17 TB sera were tested on the 1152 phage peptide microarray. Bioinformatically analyzed data identified 259 antigens with the highest level of differentiation between sarcoidosis and healthy controls.

[0009] FIG. 2A shows a heatmap generated by applying meta-analysis using microarray analysis of 2 separate data sets derived from 115 sarcoidosis patients and 64 healthy controls. Data reflects 259 antigens expressed significantly differently between healthy controls and sarcoidosis subjects in immunoscreening using sera. The 259 antigens were further divided into three categories according to the AW-OC method. I: 78 antigens were consistently up- or down-regulated in sarcoidosis in both datasets; II: 115 antigens were up- or down-regulated in sarcoidosis in the second dataset only; III: 66 antigens were up- or down-regulated in sarcoidosis in the first dataset only. FIG. 2B shows receiver operating characteristics (ROC) curves demonstrating the performance of 32 classifiers to discriminate between healthy controls and sarcoidosis subjects.

[0010] FIGs. 3A-3J show receiver operating characteristics (ROC) curves for the top 10 sarcoidosis clones as follows: 3A (CCL21); 3B (Metap1); 3C (PC4); 3D (CLI_3190); 3E (TNFRSF21); 3F (CD14); 3G (DNAJC1); 3H (APBB1); 3I (FGFBP-2); and 3J (SH3YL1).

[0011] FIG. 4 shows a Venn diagram depicting differential phage clone significances among sarcoidosis, TB and healthy controls ($q < 0.01$). The Venn diagram shows the overlap between 259 sarcoidosis clones and 238 TB clones as compared to healthy controls, as well as 380 TB clones versus sarcoidosis. Forty seven clones could differentiate both sarcoidosis and TB from healthy controls. Five clones could not discriminate between TB and sarcoidosis.

[0012] FIG. 5A shows a heatmap generated from a microarray analysis using 3 data sets derived from 115 sarcoidosis patients, 64 control subjects and 17 TB patients. Fifty antigens showed significant differential expression among the three groups. FIG. 5B shows an enlarged version of clone identifiers to increase legibility of FIG. 5A.

[0013] FIG. 6 shows modified linkers distinguishing between the origins of each library. Each cDNA library was tagged with a modified linker: ECOR1/HindIII was used for BAL cDNA, ALA for WBC cDNA, LEU for MARC5 cDNA and THR for EL1 cDNA.

[0014] FIG. 7A shows a graphical representation of the output eluent phage titers as a function of biopanning (BP) showing exponential enrichment of the output eluent phage titers after the completion of each cycle of biopannings. FIG. 7B shows PCR amplification of the phage clones picked up from biopannings 3 & 4 (BP3 & BP4) showing retention of

diversity in the pool of immunoreactive phage.

[0015] FIG. 8 shows sequence analysis of the top 10 sarcoid phage clones using NCBI BLAST.

[0016] FIG. 9 shows sequence analysis of the top 10 TB phage clones using NCBI BLAST.

[0017] FIG. 10 shows an illustrative schematic for using computational tools as part of a process for diagnosing sarcoidosis, including an illustrative diagram of a computing device implementing the diagnostic framework.

[0018] FIG. 11 shows an illustrative process for diagnosing sarcoidosis.

DETAILED DESCRIPTION

[0019] Sarcoidosis is a multisystem granulomatous inflammatory disease. The disease is typically characterized by the formation of small, granular inflammatory lesions or granulomas (e.g., non-caseating granulomas) in a variety of organs, and/or the presence of immune responses (e.g., presence of CD4+ T lymphocytes and macrophages) in affected tissues or organs. Granulomatous inflammation may be attributed to the accumulation of monocytes, macrophages, and a pronounced Th1 response and activated T-lymphocytes, with elevated production of TNF α , IL-2, IL-12, IFN γ , IL-1, IL-6 or IL-15.

[0020] Exemplary subtypes of sarcoidosis include systemic sarcoidosis, Lofgren's syndrome, pulmonary sarcoidosis, cutaneous sarcoidosis, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis, musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues.

[0021] Systemic sarcoidosis is sarcoidosis with multiple organ involvement. Symptoms of systemic sarcoidosis include aches, arthritis, chills, dry mouth, enlarged lymph glands (e.g., armpit lump), fatigue, fever, loss of appetite, night sweats, nosebleed, pains, persistent cough, malaise, shortness of breath, weakness, and weight loss. Because systemic sarcoidosis involves multiple organs, symptoms described below for other more particular types of sarcoidosis can also be relevant to systemic sarcoidosis.

[0022] Lofgren's syndrome represents an acute presentation of systemic sarcoidosis, typically characterized by the triad of erythema nodosum, bilateral hilar denopathy and arthritis or arthralgias. It can also be accompanied by fever.

[0023] Pulmonary sarcoidosis refers to sarcoidosis that affects pulmonary tissues or organs (e.g., lungs). Symptoms of pulmonary sarcoidosis usually include normal, abnormal or deteriorating lung function; abnormal lung stiffness; bleeding from the lung tissue; cough; decreased lung volume; decreased vital capacity (full breath in, to full breath out); enlarged lymph nodes in the chest; granulomas in alveolar septa, bronchiolar, and/or bronchial walls; higher than normal expiratory flow ratios; an increased FEV₁/FVC ratio; limited amount of air drawn into the lungs; loss of lung volume; obstructive lung changes; pulmonary hypertension; pulmonary failure; scarring of lung tissue; and/or shortness of breath.

[0024] Cutaneous sarcoidosis is a complication of sarcoidosis with skin involvement. Cutaneous sarcoidosis includes annular sarcoidosis, erythrodermic sarcoidosis, hypopigmented sarcoidosis, ichthyosiform sarcoidosis, morpheaform sarcoidosis, mucosal sarcoidosis, papular sarcoid, scar sarcoid, subcutaneous sarcoidosis and ulcerative sarcoidosis. Symptoms of cutaneous sarcoidosis include erythema nodosum (e.g., raised, red, firm skin sores, cellulitis, furunculosis or other inflammatory panniculitis); hair loss; lupus pernio (e.g., scar or discoid lupus erythematosus); maculopapular eruptions; nodular lesions; papules (e.g., granulomatous rosacea, acne or benign appendageal tumors); skin lesions; skin plaques (e.g., psoriasis, lichen planus, nummular eczema, discoid lupus erythematosus, granuloma annulare, cutaneous T-cell lymphoma, Kaposi's sarcoma or secondary syphilis); skin rashes, and/or scars becoming more raised.

[0025] Neurosarcoidosis or neurosarcoid refers to sarcoidosis in which inflammation and abnormal deposits occur in the brain, spinal cord, and any other areas of the nervous system. Symptoms of neurosarcoidosis can include abnormal or loss of sense of smell; abnormal or loss of sense of taste; carpal tunnel syndrome; changes in menstrual periods; confusion; decreased hearing; delirium; dementia; disorientation; dizziness; double vision or other vision problems or changes; excessive thirst; excessive tiredness (e.g., fatigue); facial palsy, weakness or drooping; headache; high urine output; hypopituitarism; loss of bowel or bladder control; muscle weakness; paraplegia; psychiatric disturbances; radicular pain; retinopathy; seizures; sensory losses; speech impairment; and/or vertigo.

[0026] The systems and methods disclosed herein can be used to diagnose

sarcoidosis. In particular embodiments, the diagnosed sarcoidosis is systemic sarcoidosis, pulmonary sarcoidosis, cutaneous sarcoidosis, Lofgren's syndrome, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis, musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues. In more particular embodiments, the systems and methods disclosed herein can be used to diagnose pulmonary sarcoidosis, neurosarcoidosis, and/or ocular sarcoidosis.

[0027] Typically, a sarcoidosis patient will present with symptoms described above or clinical features set out in the Statement on Sarcoidosis published by the American Thoracic Society (Am. J. Respir. Crit. Care Med. 1999; 160(2):736-55). Sarcoidosis patients may often, however, be asymptomatic. Further the common symptoms of sarcoidosis are vague, and can sometimes be similar to symptoms of numerous other conditions including lymphoma and tuberculosis. Thus, diagnosis is difficult.

[0028] Currently, subjects with suspected sarcoidosis are typically assessed with a chest assessment for pulmonary involvement, as the vast majority of sarcoidosis subjects have pulmonary involvement. These assessments are generally based upon a bronchoscopy with biopsy; chest X-ray; CT scan; CT-guided biopsy; lung gallium (Ga) scan; mediastinoscopy; open lung biopsy; PET scan and/or a radiograph. Radiographs are typically assigned a stage of 0-4 according to the presence or absence of hilar adenopathy and parenchymal disease. Thus there are five stages: Stage 0: no visible intrathoracic findings; Stage 1: bilateral hilar lymphadenopathy (BHL), which may be accompanied by paratracheal adenopathy/lung fields are clear of infiltrates; Stage 2: bilateral hilar adenopathy (BHL) accompanied by parenchymal infiltration; Stage 3: parenchymal infiltration without bilateral hilar adenopathy (BHL); or Stage 4: advanced pulmonary fibrosis with evidence of honey-combing, hilar retraction, bullae, cysts, and emphysema.

[0029] The present disclosure provides significant advancements in the diagnosis of sarcoidosis because diagnosis can be achieved with, for example, a blood test and can distinguish sarcoidosis subjects from healthy subjects and/or subjects having tuberculosis.

[0030] The systems and methods disclosed herein were achieved by creating and

screening a complex cDNA library. Particularly, a heterologous cDNA library derived from bronchoalveolar cell (BAL) samples and total white blood cells (WBC) from sarcoidosis patients was developed. Both sarcoid-derived libraries were combined with cultured human monocytes and embryonic lung fibroblast cDNA libraries to build a complex sarcoidosis library (CSL). Differential biopanning for negative and positive selection was performed using sera from healthy controls to remove non-specific IgG, and sarcoidosis sera for selective enrichment. Four rounds of biopannings were performed and the selected phage libraries were used for microarray immunoscreening. Each cycle of biopanning included passing the entire phage library through protein G beads coated with IgG from pooled sera of healthy controls, then passing through beads coated with IgGs from individual serum of sarcoid subjects.

[0031] After biopanning, phage clones were randomly selected and amplified and their lysates were arrayed in quintuplicates onto slides (Grace Biolabs, OR) using the ProSys 5510TL robot (Cartesian Technologies, CA). It was tested whether this novel library representing relevant antigens would specifically recognize high IgG titer in sera of sarcoidosis subjects.

[0032] Using bioinformatics tools, a large number of markers with high sensitivity and specificity were identified that discriminate among the sera of patients with sarcoidosis, healthy controls and TB. Using the integrative-analysis method that combines results from two independent trials, clones that significantly differentiated sarcoidosis from controls were identified. Similarly, clones that differentially reacted with TB sera and not with sarcoidosis or control sera were identified. Furthermore, the top 10 discriminating antigens for TB and sarcoidosis were sequenced and homologies were identified in a public data base. These data indicates development of a unique library enabling the detection of highly significant antigens to discriminate between patients with sarcoidosis and tuberculosis.

[0033] An antigen is a substance that induces an immune response. Accordingly, the antigens detected from the library are markers useful for diagnosing sarcoidosis and TB.

[0034] The systems and methods diagnose sarcoidosis by assaying a sample obtained from a subject for the up- or down-regulation of one or more markers associated with sarcoidosis. The markers are selected from Small inducible cytokine A21 precursor

(CCL21); Methionine aminopeptidase 1 (Metap1); Activated RNA polymerase II transcription cofactor variant 4 (PC4); RNA methyltransferase (CLI_3190); Tumor necrosis factor receptor superfamily member 21 precursor (also known as death receptor 6 (DR6)) (TNFRSF21); Monocyte differentiation antigen CD14 (CD14); DnaJ (Hsp40) homolog subfamily C member 1 precursor (DNAJC1); Amyloid β A4 precursor protein-binding family B member 1-interacting protein (APBB1); Fibroblast growth factor binding protein 2 precursor (FGFBP-2); SH3 domain-containing YSC84 like protein 1 (SH3YL1); thioester reductase [*Pseudomonas fluorescens*] (PFWH6_0117); histidine kinase [*Pseudomonas fluorescens*] (PFL_3193); Homo sapiens chromatin modifying protein 4B (CHMP4B); hypothetical protein [*Porphyromonas somerae*] Peptidase family C39 mostly contains bacteriocin-processing endopeptidases from bacteria; truncated HIC1 protein [Homo sapiens] (H1C1); replication protein [*Mycobacterium*] (MVAC_06252); Homo sapiens ribosomal protein S2 (RPS2); triosephosphate isomerase [*Mycobacterium tuberculosis*] (tpiA); membrane protein [*Mycobacterium tuberculosis*] (Rv2563); serine/threonine protein kinase [*Mycobacterium tuberculosis*] (Rv0410C); PPE family protein [*Mycobacterium tuberculosis* RGTB423] (MRGA423_16320); rRNA methyltransferase [*Mycobacterium tuberculosis*] (Rv0881); peroxisome biogenesis factor 10 isoform 1 [Homo sapiens] (PEX10); sulfate ABC transporter permease [*Mycobacterium tuberculosis*] (CysU); and/or D-alpha-D-heptose-7-phosphate kinase [*Mycobacterium tuberculosis*] (hddA).

[0035] In particular embodiments, the systems and methods diagnose sarcoidosis by assaying a sample obtained from a subject for the up- or down-regulation of two or more; three or more; four or more; five or more; six or more; seven or more; eight or more; nine or more or ten or more markers associated with sarcoidosis disclosed herein. In further embodiments, the systems and methods diagnose sarcoidosis by assaying a sample obtained from a subject for the up- or down-regulation of two; three; four; five; six; seven; eight; nine or ten markers associated with sarcoidosis disclosed herein.

[0036] In one embodiment, the markers include (hereafter referred to by gene abbreviations for brevity) CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1. In another embodiment, the markers include CCL21; Metap1; PC4; CLI_3190; TNFRSF21; and APBB1. In another embodiment, the markers

include CCL21, PC4, CLI3190, DNAJC1, APBB1, FGFBP-2 and SH3YL1. In another embodiment, the markers include CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1. In another embodiment, the markers include CCL21; Metap1; CLI-3190; APBB1; and SH3YL1.

[0037] In other embodiments, the markers include CCL21 in combination with two, three, four, five, six, seven, eight or nine markers selected from: Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1.

[0038] In other embodiments, the markers include Metap1 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1.

[0039] In other embodiments, the markers include PC4 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1.

[0040] In other embodiments, the markers include CLI_3190 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1.

[0041] In other embodiments, the markers include TNFRSF21 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; CLI_3190; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1.

[0042] In other embodiments, the markers include CD14 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; DNAJC1; APBB1; FGFBP-2; and SH3YL1.

[0043] In other embodiments, the markers include DNAJC1 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; APBB1; FGFBP-2; and SH3YL1.

[0044] In other embodiments, the markers include APBB1 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; FGFBP-2; and SH3YL1.

[0045] In other embodiments, the markers include FGFBP-2 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; and SH3YL1.

[0046] In other embodiments, the markers include SH3YL1 in combination with two, three, four, five, six, seven, eight or nine markers selected from: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; and FGFBP-2.

[0047] In other embodiments, the markers exclude a marker selected from CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1. In other embodiments, the markers exclude CCL21. In other embodiments, the markers exclude Metap1. In other embodiments, the markers exclude PC4. In other embodiments, the markers exclude CLI_3190. In other embodiments, the markers exclude TNFRSF21. In other embodiments, the markers exclude CD14. In other embodiments, the markers exclude DNAJC1. In other embodiments, the markers exclude APBB1. In other embodiments, the markers exclude FGFBP-2. In other embodiments, the markers exclude SH3YL1. In other embodiments, the markers exclude one or more of Metap 1; TNFRSF21; and CD14. In other embodiments, the markers exclude Metap 1; TNFRSF21; and CD14.

[0048] Any of the embodiments described above can additionally include a marker selected from PFWH6_0117; PFL_3193; CHMP4B; hypothetical protein [Porphyromonas somerae] Peptidase family C39 mostly contains bacteriocin-processing endopeptidases from bacteria; H1C1; MVAC_06252; RPS2; tpiA; Rv2563; Rv0410C; MRGA423_16320; Rv0881; PEX10; CysU; and hddA. In particular embodiments, the additional marker includes PFWH6_0117. In particular embodiments, the additional marker includes PFL_3193. In particular embodiments, the additional marker includes CHMP4B. In particular embodiments, the additional marker includes hypothetical protein Peptidase family C39 mostly contains bacteriocin-processing endopeptidases from bacteria. In particular embodiments, the additional marker includes H1C1. In particular embodiments, the additional marker includes MVAC_06252. In particular embodiments, the additional marker includes RPS2. In particular embodiments, the additional marker includes tpiA. In particular embodiments, the additional marker includes Rv2563. In particular embodiments, the additional marker includes Rv0410C. In particular embodiments, the additional marker includes MRGA423_16320. In particular embodiments, the additional marker includes Rv0881. In particular embodiments, the additional marker includes PEX10. In particular embodiments, the additional marker includes CysU. In particular

embodiments, the additional marker includes hddA.

[0049] The systems and methods disclosed herein also allow distinguishing sarcoidosis from tuberculosis in a subject by assaying a sample obtained from a subject for the up- or down-regulation of one or more markers that distinguish sarcoidosis from tuberculosis. The markers include: Ferredoxin (*Mycobacterium tuberculosis*) (Fed A); WDFY3 protein (*Homo sapiens*) (WDFY3); Membrane protein (*Mycobacterium tuberculosis*) (MFS); Leucine rich PPR-motif containing protein (*Homo sapiens*) (LRPPRC); HLA-DR alpha (*Homo sapiens*) (HLA-DR); Transketolase (*Mycobacterium tuberculosis*) (TKT); Dihydroxy acid dehydratase (*Mycobacterium tuberculosis*) (Rv0189C); Chain A *Mycobacterium tuberculosis* (BfrA); Disabled homolog 2 isoform 2 (*Homo sapiens*) (DAB2); and Transcription elongation factor B polypeptide 2 isoform (*Homo sapiens*) (TCEB2).

[0050] In particular embodiments, the systems and methods distinguish sarcoidosis from tuberculosis in a subject by assaying a sample obtained from a subject for the up- or down-regulation of two or more; three or more; four or more; five or more; six or more; seven or more; eight or more; nine or more or ten or more markers that distinguish sarcoidosis from tuberculosis disclosed herein. In further embodiments, the systems and methods distinguish sarcoidosis from tuberculosis by assaying a sample obtained from a subject for the up- or down-regulation of two; three; four; five; six; seven; eight; nine or ten markers associated with sarcoidosis disclosed herein.

[0051] In one embodiment, the markers include (hereafter referred to by gene abbreviations for brevity) Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2. In another embodiment, the markers include HLA-DR; MFS; DAB2; BfrA; and WDFY3. In another embodiment, the markers include HLA-DR; MFS; DAB2; BfrA; or WDFY3. In another embodiment, the markers include HLA-DR; MFS; and DAB2. In another embodiment, the markers include HLA-DR; MFS; or DAB2.

[0052] In other embodiments, the markers include Fed A in combination with two, three, four, five, six, seven, eight or nine markers selected from: WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2.

[0053] In other embodiments, the markers include WDFY3 in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; MFS; LRPPRC;

HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2.

[0054] In other embodiments, the markers include MFS in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2.

[0055] In other embodiments, the markers include LRPPRC in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2.

[0056] In other embodiments, the markers include HLA-DR in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; LRPPRC; TKT; Rv0189C; BfrA; DAB2; and TCEB2.

[0057] In other embodiments, the markers include TKT in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; LRPPRC; HLA-DR; Rv0189C; BfrA; DAB2; and TCEB2.

[0058] In other embodiments, the markers include Rv0189C in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; BfrA; DAB2; and TCEB2.

[0059] In other embodiments, the markers include BfrA in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; DAB2; and TCEB2.

[0060] In other embodiments, the markers include DAB2 in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; and TCEB2.

[0061] In other embodiments, the markers include TCEB2 in combination with two, three, four, five, six, seven, eight or nine markers selected from: Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; and DAB2.

[0062] In other embodiments, the markers exclude a marker selected from Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2. In other embodiments, the markers exclude Fed A. In other embodiments, the markers exclude WDFY3. In other embodiments, the markers exclude MFS. In other embodiments, the markers exclude LRPPRC. In other embodiments, the markers exclude HLA-DR. In other embodiments, the markers exclude TKT. In other embodiments, the markers exclude

Rv0189C. In other embodiments, the markers exclude BfrA. In other embodiments, the markers exclude DAB2. In other embodiments, the markers exclude TCEB2.

[0063] "Up-regulation" or "up-regulated" means an increase in the presence of a protein and/or an increase in the expression of its gene. "Down-regulation" or "down-regulated" means a decrease in the presence of a protein and/or a decrease in the expression of its gene. "Its gene" in reference to a particular protein refers to a nucleic acid sequence (used interchangeably with polynucleotide or nucleotide sequence) that encodes the particular protein. This definition also includes various sequence polymorphisms, mutations, and/or sequence variants wherein such alterations do not substantially affect the identity or function of the particular protein. For example, in a sequence identity analysis, the test protein would share at least 80% sequence identity; at least 81% sequence identity; at least 82% sequence identity; at least 83% sequence identity; at least 84% sequence identity; at least 85% sequence identity; at least 86% sequence identity; at least 87% sequence identity; at least 88% sequence identity; at least 89% sequence identity; at least 90% sequence identity; at least 91% sequence identity; at least 92% sequence identity; at least 93% sequence identity; at least 94% sequence identity; at least 95% sequence identity; at least 96% sequence identity; at least 97% sequence identity; at least 98% sequence identity or at least 99% sequence identity with the particular protein.

[0064] "% sequence identity" refers to a relationship between two or more sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between protein (or nucleic acid) sequences as determined by the match between strings of such sequences. "Identity" (often referred to as "similarity") can be readily calculated by known methods, including those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, NY (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, NY (1994); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, NJ (1994); Sequence Analysis in Molecular Biology (Von Heijne, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Oxford University Press, NY (1992). Preferred methods to determine sequence identity are designed to give the best match between the sequences tested. Methods to

determine sequence identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR, Inc., Madison, Wisconsin). Multiple alignment of the sequences can also be performed using the Clustal method of alignment (Higgins and Sharp CABIOS, 5, 151-153 (1989) with default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Relevant programs also include the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wisconsin); BLASTP, BLASTN, BLASTX (Altschul, et al., J. Mol. Biol. 215:403-410 (1990); DNASTAR (DNASTAR, Inc., Madison, Wisconsin); and the FASTA program incorporating the Smith-Waterman algorithm (Pearson, Comput. Methods Genome Res., [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, N.Y. Within the context of this disclosure it will be understood that where sequence analysis software is used for analysis, the results of the analysis are based on the "default values" of the program referenced. "Default values" mean any set of values or parameters which originally load with the software when first initialized.

[0065] The function of a protein can be assayed by a relevant activity assay. Function is not substantially affected if there is no statistically significant difference in activity between the particular protein and the test protein. Exemplary activity assays include binding assays, or, if the protein is an enzyme, enzyme activity assays including, for example, protease assays, kinase assays, phosphatase assays, reductase assays, etc. Modulation of the kinetics of enzyme activities can be determined by measuring the rate constant K_M using known algorithms, such as the Hill plot, Michaelis-Menten equation, linear regression plots such as Lineweaver-Burk analysis, and Scatchard plot.

[0066] The term "gene" can include not only coding sequences but also regulatory regions such as promoters, enhancers, and termination regions. The term further can include all introns and other DNA sequences spliced from the mRNA transcript, along with variants resulting from alternative splice sites. Gene sequences encoding the particular protein can be DNA or RNA that directs the expression of the particular protein. These nucleic acid sequences may be a DNA strand sequence that is transcribed into RNA or an RNA sequence that is translated into the particular protein. The nucleic acid

sequences include both the full-length nucleic acid sequences as well as non-full-length sequences derived from the full-length protein. The sequences can also include degenerate codons of the native sequence. Portions of complete gene sequences are referenced throughout the disclosure as is understood by one of ordinary skill in the art.

[0067] Up- or down-regulation of the markers, as indicated elsewhere herein for particular markers can be assessed by comparing a value to a relevant reference level. For example, the quantity of one or more markers can be indicated as a value. The value can be one or more numerical values resulting from the assaying of a sample, and can be derived, e.g., by measuring level(s) of the marker(s) in the sample by an assay performed in a laboratory, or from a dataset obtained from a provider such as a laboratory, or from a dataset stored on a server. The markers disclosed herein can be a protein marker or a nucleic acid marker (gene encoding the protein marker).

[0068] In the broadest sense, the value may be qualitative or quantitative. As such, where detection is qualitative, the systems and methods provide a reading or evaluation, e.g., assessment, of whether or not the marker is present in the sample being assayed. In yet other embodiments, the systems and methods provide a quantitative detection of whether the marker is present in the sample being assayed, i.e., an evaluation or assessment of the actual amount or relative abundance of the marker in the sample being assayed. In such embodiments, the quantitative detection may be absolute or, if the method is a method of detecting two or more different markers in a sample, relative. As such, the term “quantifying” when used in the context of quantifying a marker in a sample can refer to absolute or to relative quantification. Absolute quantification can be accomplished by inclusion of known concentration(s) of one or more control markers and referencing, e.g., normalizing, the detected level of the marker with the known control markers (e.g., through generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of detected levels or amounts between two or more different markers to provide a relative quantification of each of the two or more markers, e.g., relative to each other. The actual measurement of values of the markers can be determined at the protein or nucleic acid level using any method known in the art. In some embodiments, a marker is detected by contacting a sample with reagents (e.g., antibodies or nucleic acid primers), generating complexes of reagent and

marker(s), and detecting the complexes.

[0069] The reagent can include a probe. A probe is a molecule that binds a target, either directly or indirectly. The target can be a marker, a fragment of the marker, or any molecule that is to be detected. In embodiments, the probe includes a nucleic acid or a protein. As an example, a protein probe can be an antibody. An antibody can be a whole antibody or a fragment of an antibody. A probe can be labeled with a detectable label. Examples of detectable labels include fluorescers, chemiluminescers, dyes, enzymes, enzyme substrates, enzyme cofactors, enzyme inhibitors, enzyme subunits, metal ions, and radioactive isotopes.

[0070] "Protein" detection includes detection of full-length proteins, mature proteins, pre-proteins, polypeptides, isoforms, mutations, post-translationally modified proteins and variants thereof, and can be detected in any suitable manner.

[0071] Those skilled in the art will be familiar with numerous specific immunoassay formats and variations thereof which can be useful for carrying out the methods disclosed herein. See, e.g., E. Maggio, *Enzyme-Immunoassay* (1980), CRC Press, Inc., Boca Raton, Fla; and U.S. Pat. Nos. 4,727,022; 4,659,678; 4,376,110; 4,275,149; 4,233,402; and 4,230,797.

[0072] Antibodies can be conjugated to a solid support suitable for a diagnostic assay (e.g., beads such as protein A or protein G agarose, microspheres, plates, slides or wells formed from materials such as latex or polystyrene) in accordance with known techniques, such as passive binding. Antibodies can be conjugated to detectable labels or groups such as radiolabels (e.g., ³⁵S, ¹²⁵I, ¹³¹I), enzyme labels (e.g., horseradish peroxidase, alkaline phosphatase), and fluorescent labels (e.g., fluorescein, Alexa, green fluorescent protein, rhodamine) in accordance with known techniques.

[0073] Examples of suitable immunoassays include immunoblotting, immunoprecipitation, immunofluorescence, chemiluminescence, electro-chemiluminescence (ECL), and/or enzyme-linked immunoassays (ELISA).

[0074] Antibodies may also be useful for detecting post-translational modifications of markers. Examples of post-translational modifications include tyrosine phosphorylation, threonine phosphorylation, serine phosphorylation, citrullination and glycosylation (e.g., O-GlcNAc). Such antibodies specifically detect the phosphorylated amino acids in marker

proteins of interest. These antibodies are well-known to those skilled in the art, and commercially available. Post-translational modifications can also be determined using metastable ions in reflector matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF). See U. Wirth et al., *Proteomics* 2002, 2(10):1445-1451.

[0075] Up- or down-regulation of genes also can be detected using, for example, cDNA arrays, cDNA fragment fingerprinting, cDNA sequencing, clone hybridization, differential display, differential screening, FRET detection, liquid microarrays, PCR, RT-PCR, quantitative real-time RT-PCR analysis with TaqMan assays, molecular beacons, microelectric arrays, oligonucleotide arrays, polynucleotide arrays, serial analysis of gene expression (SAGE), and/or subtractive hybridization.

[0076] As an example, Northern hybridization analysis using probes which specifically recognize one or more marker sequences can be used to determine gene expression. Alternatively, expression can be measured using RT-PCR; e.g., polynucleotide primers specific for the differentially expressed marker mRNA sequences reverse-transcribe the mRNA into DNA, which is then amplified in PCR and can be visualized and quantified. Marker RNA can also be quantified using, for example, other target amplification methods, such as transcription mediated amplification (TMA), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA), or signal amplification methods (e.g., bDNA), and the like. Ribonuclease protection assays can also be used, using probes that specifically recognize one or more marker mRNA sequences, to determine gene expression.

[0077] Further hybridization technologies that may be used are described in, for example, U.S. Pat. Nos. 5,143,854; 5,288,644; 5,324,633; 5,432,049; 5,470,710; 5,492,806; 5,503,980; 5,510,270; 5,525,464; 5,547,839; 5,580,732; 5,661,028; and 5,800,992 as well as WO 95/21265; WO 96/31622; WO 97/10365; WO 97/27317; EP 373 203; and EP 785 280.

[0078] Proteins and nucleic acids can be linked to chips, such as microarray chips. See, for example, U.S. Pat. Nos. 5,143,854; 6,087,112; 5,215,882; 5,707,807; 5,807,522; 5,958,342; 5,994,076; 6,004,755; 6,048,695; 6,060,240; 6,090,556; and 6,040,138. Microarray refers to a solid carrier or support that has a plurality of molecules bound to its surface at defined locations. The solid carrier or support can be made of any material.

As an example, the material can be hard, such as metal, glass, plastic, silicon, ceramics, and textured and porous materials; or soft materials, such as gels, rubbers, polymers, and other non-rigid materials. The material can also be nylon membranes, epoxy-glass and borofluorate-glass. The solid carrier or support can be flat, but need not be and can include any type of shape such as spherical shapes (e.g., beads or microspheres). The solid carrier or support can have a flat surface as in slides and micro-titer plates having one or more wells.

[0079] Binding to proteins or nucleic acids on microarrays can be detected by scanning the microarray with a variety of laser or CCD-based scanners, and extracting features with software packages, for example, Imogene (Biodiscovery, Hawthorne, CA), Feature Extraction Software (Agilent), Scanalyze (Eisen, M. 1999. SCANALYZE User Manual; Stanford Univ., Stanford, Calif. Ver 2.32.), or GenePix (Axon Instruments).

[0080] Embodiments disclosed herein can be used with high throughput screening (HTS). Typically, HTS refers to a format that performs at least about 100 assays, at least about 500 assays, at least about 1000 assays, at least about 5000 assays, at least about 10,000 assays, or more per day. When enumerating assays, either the number of samples or the number of protein or nucleic acid markers assayed can be considered.

[0081] Generally HTS methods involve a logical or physical array of either the subject samples, or the protein or nucleic acid markers, or both. Appropriate array formats include both liquid and solid phase arrays. For example, assays employing liquid phase arrays, e.g., for hybridization of nucleic acids, binding of antibodies or other receptors to ligand, etc., can be performed in multiwell or microtiter plates. Microtiter plates with 96, 384, or 1536 wells are widely available, and even higher numbers of wells, e.g., 3456 and 9600 can be used. In general, the choice of microtiter plates is determined by the methods and equipment, e.g., robotic handling and loading systems, used for sample preparation and analysis.

[0082] HTS assays and screening systems are commercially available from, for example, Zymark Corp. (Hopkinton, MA); Air Technical Industries (Mentor, OH); Beckman Instruments, Inc. (Fullerton, CA); Precision Systems, Inc. (Natick, MA), etc.. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s)

appropriate for the assay. These configurable systems provide HTS as well as a high degree of flexibility and customization. The manufacturers of such systems provide detailed protocols for the various methods of HTS.

[0083] As stated previously, obtained marker values can be compared to a reference level. Reference levels can be obtained from one or more relevant datasets. A "dataset" as used herein is a set of numerical values resulting from evaluation of a sample (or population of samples) under a desired condition. The values of the dataset can be obtained, for example, by experimentally obtaining measures from a sample and constructing a dataset from these measurements. As is understood by one of ordinary skill in the art, the reference level can be based on e.g., any mathematical or statistical formula useful and known in the art for arriving at a meaningful aggregate reference level from a collection of individual datapoints; e.g., mean, median, median of the mean, etc. Alternatively, a reference level or dataset to create a reference level can be obtained from a service provider such as a laboratory, or from a database or a server on which the dataset has been stored.

[0084] A reference level from a dataset can be derived from previous measures derived from a population. A "population" is any grouping of subjects or samples of like specified characteristics. The grouping could be according to, for example, clinical parameters, clinical assessments, therapeutic regimens, disease status, severity of condition, etc.

[0085] Subjects include humans, veterinary animals (dogs, cats, reptiles, birds, hamsters, etc.) livestock (horses, cattle, goats, pigs, chickens, etc.), research animals (monkeys, rats, mice, fish, etc.) and other animals, such as zoo animals (e.g., bears, giraffe, elephant, lemurs).

[0086] In particular embodiments, conclusions are drawn based on whether a sample value is statistically significantly different or not statistically significantly different from a reference level. A measure is not statistically significantly different if the difference is within a level that would be expected to occur based on chance alone. In contrast, a statistically significant difference or increase is one that is greater than what would be expected to occur by chance alone. Statistical significance or lack thereof can be determined by any of various methods well-known in the art. An example of a commonly

used measure of statistical significance is the p-value. The p-value represents the probability of obtaining a given result equivalent to a particular datapoint, where the datapoint is the result of random chance alone. A result is often considered significant (not random chance) at a p-value less than or equal to 0.05.

[0087] In one embodiment, values obtained about the markers and/or other dataset components can be subjected to an analytic process with chosen parameters. The parameters of the analytic process may be those disclosed herein or those derived using the guidelines described herein. The analytic process used to generate a result may be any type of process capable of providing a result useful for classifying a sample, for example, comparison of the obtained value with a reference level, a linear algorithm, a quadratic algorithm, a decision tree algorithm, or a voting algorithm. The analytic process may set a threshold for determining the probability that a sample belongs to a given class. The probability preferably is at least at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or higher.

[0088] In embodiments, the relevant reference level for a particular marker is obtained based on the particular marker in control subjects. Control subjects are those that are healthy and do not have sarcoidosis or tuberculosis. As an example, the relevant reference level can be the quantity of the particular marker in the control subjects.

[0089] In additional embodiments when more than one marker is assayed, values of the detected markers can be calculated into a score. Each value can be weighted evenly within an algorithm generating a score, or the values for particular markers can be weighted more heavily in reaching the score. For example, markers with higher sensitivity and/or specificity scores could be weighted more heavily than markers with lower sensitivity and/or specificity scores. For example, marker values for diagnosing sarcoidosis may be weighted as follows (from highest weight to lowest weight): CCL21; APBB1; Metap1; SH3YL; CLI_3190; PC4; DNAJC1; TNFRSF21; CD14; FGFBP-2. Markers may also be grouped into classes, and each class given a weighted score. For example, marker values for diagnosing sarcoidosis may be grouped into classes and weighted as follows (from highest weight to lowest weight): Class 1: CCL21 and APBB1; Class 2: Metap1 and SH3YL; Class 3: CLI_3190 and PC4; Class 4: DNAJC1 and TNFRSF21; and Class 5: CD14 and FGFBP-2. As another example, marker values for

diagnosing sarcoidosis may be grouped into classes and weighted as follows (from highest weight to lowest weight): Class 1: CCL21; APBB1; Metap1 and SH3YL; Class 2: CLI_3190; PC4; DNAJC1 and TNFRSF21; and Class 3: CD14 and FGFBP-2.

[0090] In particular embodiments, marker values for distinguishing sarcoidosis from tuberculosis may be weighted as follows (from highest weight to lowest weight): HLA-DR; MF5; BfrA; DAB2; WDFY3; FedA; TCEB2; Rv0189C; LRPPRC; TKT. Markers may also be grouped into classes, and each class given a weighted score. For example, marker values for diagnosing sarcoidosis may be grouped into classes and weighted as follows (from highest weight to lowest weight): Class 1: HLA-DR and MF5; Class 2: BfrA and DAB2; Class 3: WDFY3 and FedA; Class 4: TCEB2 and Rv0189C; and Class 5: LRPPRC and TKT. As another example, marker values for diagnosing sarcoidosis may be grouped into classes and weighted as follows (from highest weight to lowest weight): Class 1: HLA-DR; MF5; BfrA; and DAB2; Class 2: WDFY3; FedA; TCEB2; and Rv0189C; and Class 3: LRPPRC; and TKT.

[0091] Any marker or class of markers can be excluded from a particular value calculation. For example, in particular embodiments, Class 5 is excluded. In particular embodiments, Class 4 is excluded. In particular embodiments, Class 3 is excluded. In particular embodiments, Class 2 is excluded. In particular embodiments, Class 1 is excluded. In further embodiments, groups of classes can be excluded, for example, Classes 5 and 4; 5 and 3; 5 and 2; 4 and 3; 4 and 2; 3 and 2; etc.

[0092] Particular embodiments disclosed herein include obtaining a sample from a subject suspected of having sarcoidosis; assaying the sample for up- or down-regulation of one or more markers disclosed herein; determining one or more marker values based on the assaying; comparing the one or more marker values to a reference level; diagnosing sarcoidosis in the subject according to the up- or down regulation of a marker, as described elsewhere herein.

[0093] Particular embodiments also include distinguishing sarcoidosis from tuberculosis in a subject by obtaining a sample from a subject suspected of having sarcoidosis; assaying the sample for up- or down-regulation of one or more markers disclosed herein; determining one or more marker values based on the assaying; comparing the one or more marker values to a reference level; diagnosing sarcoidosis or

tuberculosis in the subject according to the up- or down regulation of a marker, as described elsewhere herein.

[0094] The sample can be any appropriate biological sample obtained from the subject, such as a blood sample, a serum sample, a saliva sample, a urine sample, bronchoalveolar lavage sample, etc. The sample also can be obtained from a biopsy of an affected tissue or organ, such as a lung biopsy, or lymph gland biopsy. The sample can include cells of affected tissue or organ.

[0095] A diagnosis according to the systems and methods disclosed herein can direct a treatment regimen. For example, a sarcoidosis diagnosis can direct treatment with a sarcoidosis treatment (e.g., lifestyle and behavioral interventions; corticosteroids; methotrexate or azathioprine; hydroxychloroquine or chloroquine; cyclophosphamide or chlorambucil; pentoxifylline and thalidomide; infliximab or adalimumab; colchicine; various nonsteroidal anti-inflammatory drugs (NSAIDs, e.g., ibuprofen or aspirin); organ transplantation). A tuberculosis diagnosis can direct treatment with a tuberculosis treatment (e.g., isoniazid (INH); rifampin (RIF); ethambutol (EMB); pyrazinamide (PZA)). A healthy diagnosis can direct further medical analysis if the subject's symptoms suggest further analysis is warranted. Administered treatments will be delivered in therapeutically effective amounts leading to an improvement or resolution of the treated condition, as assessed by a practicing physician, veterinarian or researcher.

[0096] The systems and methods disclosed herein include kits. Disclosed kits include materials and reagents necessary to assay a sample obtained from a subject for one or more markers disclosed herein. The materials and reagents can include those necessary to assay the markers disclosed herein according to any method described herein and/or known to one of ordinary skill in the art.

[0097] Particular embodiments include materials and reagents necessary to assay for up- or down-regulation of a marker protein in a sample. In particular embodiments, the kits include antibodies to marker proteins and/or can also include aptamers, epitopes or mimotopes. Other embodiments additionally or alternatively include oligonucleotides that specifically assay for one or more marker nucleic acids based on homology and/or complementarity with marker nucleic acids. The oligonucleotide sequences may correspond to fragments of the marker nucleic acids. For example, the oligonucleotides

can be more than 200, 175, 150, 100, 50, 25, 10, or fewer than 10 nucleotides in length. Collectively, any molecule (e.g., antibody, aptamer, epitope, mimotope, oligonucleotide) that forms a complex with a marker is referred to as a marker binding agent herein.

[0098] Embodiments of kits can contain in separate containers marker binding agents either bound to a matrix, or packaged separately with reagents for binding to a matrix. In particular embodiments, the matrix is, for example, a porous strip. In some embodiments, measurement or detection regions of the porous strip can include a plurality of sites containing marker binding agents. In some embodiments, the porous strip can also contain sites for negative and/or positive controls. Alternatively, control sites can be located on a separate strip from the porous strip. Optionally, the different detection sites can contain different amounts of marker binding agents, e.g., a higher amount in the first detection site and lesser amounts in subsequent sites. Upon the addition of test sample, the number of sites displaying a detectable signal provides a quantitative indication of the amount of marker present in the sample. The detection sites can be configured in any suitably detectable shape and can be, e.g., in the shape of a bar or dot spanning the width (or a portion thereof) of a porous strip.

[0099] In some embodiments the matrix can be a solid substrate, such as a "chip." See, e.g., U.S. Pat. No. 5,744,305. In some embodiments the matrix can be a solution array; e.g., xMAP (Luminex, Austin, Tex.), Cyvera (Illumina, San Diego, Calif.), RayBio Antibody Arrays (RayBiotech, Inc., Norcross, Ga.), CellCard (Vitra Bioscience, Mountain View, Calif.) and Quantum Dots' Mosaic (Invitrogen, Carlsbad, Calif.).

[0100] Additional embodiments can include control formulations (positive and/or negative), and/or one or more detectable labels, such as fluorescein, green fluorescent protein, rhodamine, cyanine dyes, Alexa dyes, luciferase, and radiolabels, among others. Instructions for carrying out the assay, including, optionally, instructions for generating a score, can be included in the kit; e.g., written, tape, VCR, or CD-ROM.

[0101] In particular embodiments, the kits include materials and reagents necessary to conduct an immunoassay (e.g., ELISA). In particular embodiments, the kits include materials and reagents necessary to conduct hybridization assays (e.g., PCR). In particular embodiments, materials and reagents expressly exclude equipment (e.g., plate readers). In particular embodiments, kits can exclude materials and reagents commonly

found in laboratory settings (pipettes; test tubes; distilled H₂O).

[0102] Numerous protein and gene sequence markers are disclosed herein. The disclosure is not limited to the particularly disclosed protein and gene sequences but instead also encompasses sequences including 80% sequence identity; 81% sequence identity; 82% sequence identity; 83% sequence identity; 84% sequence identity; 85% sequence identity; 86% sequence identity; 87% sequence identity; 88% sequence identity; 89% sequence identity; 90% sequence identity; 91% sequence identity; 92% sequence identity; 93% sequence identity; 94% sequence identity; 95% sequence identity; 96% sequence identity; 97% sequence identity; 98% sequence identity or 99% sequence identity.

[0103] When a protein sequence is provided, its gene sequences can be derived by one of ordinary skill in the art by, for example, consulting publicly available databases. In addition to the sequence identity parameters provided above, gene sequences that hybridize to derived sequences under high stringency conditions can also be included within the scope of the current disclosure. A gene or polynucleotide fragment "hybridizes" to another gene or polynucleotide fragment, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the polynucleotide fragment anneals to the other polynucleotide fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (incorporated by reference herein for its teachings regarding the same). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from distantly related organisms) to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of hybridization conditions to demonstrate that sequences hybridize uses a series of washes starting with 6XSSC, 0.5% SDS at room temperature for 15 min, then repeated with 2XSSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2XSSC, 0.5% SDS at 50°C for 30 min. Stringent conditions use higher

temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2XSSC, 0.5% SDS is increased to 60°C. Highly stringent conditions use two final washes in 0.1SSC, 0.1% SDS at 65°C. Those of ordinary skill in the art will recognize that these temperature and wash solution salt concentrations may need to be adjusted as necessary according to factors such as the length of the hybridizing sequences.

[0104] Also disclosed herein is a cDNA library including mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; and (ii) white blood cells obtained from sarcoidosis patients. In further embodiments, the cDNA library further includes mRNA isolated from (iii) human splenic monocytes; and/or (iv) embryonic lung fibroblasts. The cDNA library can be screened for markers associated with sarcoidosis or related disorders. The cDNA library can be a phage display library, a ribosome display library, or a nucleic acid display library. In particular embodiments, the cDNA library is a T7 phage display library. In particular embodiments, the cDNA library should be biopanned to negatively select and/or enrich for detection markers of interest. For example, biopanning with samples from control subjects can remove potential hits that are non-specific to the condition of interest, resulting in negative selection. Biopanning with samples from subjects of interest (e.g., subjects having a condition of interest) selects potential hits that are specific to the condition of interest, resulting in enrichment of the cDNA library for hits of potential interest. The systems and methods disclosed herein include biopanning a cDNA library including mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; (ii) white blood cells obtained from sarcoidosis patients; (iii) human splenic monocytes; and (iv) embryonic lung fibroblasts to negatively select for and/or enrich the library for hits of interest.

[0105] In embodiments, the cDNA library is differentially biopanned to identify markers for sarcoidosis. As described above, differential biopanning involves biopanning by negative selection using sera from control subjects to remove non-specific IgG, followed by biopanning by positive enrichment using sera from sarcoidosis patients.

[0106] Additional embodiments include adhering cDNA expression products from a negatively selected and enriched cDNA library to a microarray. Additional embodiments include exposing the microarray to samples from subjects of interest and control samples.

Additional embodiments include detecting cDNA expression products bound by molecules in samples from the subjects of interest. Additional embodiments include performing data analysis to identify molecules that bind cDNA expression products as markers of a condition of interest.

[0107] One embodiment includes detecting sarcoidosis or tuberculosis antigens by: (a) preparing a phage display library of sarcoidosis or tuberculosis antigens from cells of one or more subjects with sarcoidosis; (b) enriching the phage display library for sarcoidosis or tuberculosis antigens by biopanning; (c) selecting clones for amplification; (d) testing amplified clones for binding to antibodies in sera of sarcoidosis subjects; and (e) sequencing bound clones.

[0108] Another embodiment includes a library and method to identify sarcoidosis markers. One embodiment includes identifying proteins that bind to expression products of phage display clones derived from a library including mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; (ii) white blood cells obtained from sarcoidosis patients; (iii) human splenic monocytes; and/or (iv) embryonic lung fibroblasts. Another embodiment includes identifying proteins that bind to expression products of phage display clones derived from a library including mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; (ii) white blood cells obtained from sarcoidosis patients; (iii) human splenic monocytes; and (iv) embryonic lung fibroblasts. Following binding, identified proteins can be characterized and, in particular embodiments, synthesized.

[0109] These embodiments can be used to identify additional markers to diagnose systemic sarcoidosis, pulmonary sarcoidosis, cutaneous sarcoidosis, Lofgren's syndrome, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis, musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues.

[0110] In embodiments, diagnosis of sarcoidosis may be achieved in accordance with the previously disclosed methods through the use of a computing device to provide for a quicker, more reliable, and less labor intensive diagnosis.

[0111] FIG. 10 shows an illustrative schematic 1000 for diagnosing sarcoidosis in a subject 1002 on a computing device 1008, including an illustrative diagram 1028 of a

computing device 1008 implementing the diagnostic framework 1018. Sample biological material 1004 is collected from the subject 1002. That sample 1004 may be assayed for the presence of one or more markers. An indication of the up- or down-regulation of the markers is reflected by one or more marker values 1006 generated after assaying and analyzing the sample 1004. A computing device 1008 implementing the diagnostic framework 1018 will analyze and diagnose the subject 1002 as healthy, having sarcoidosis, or in some embodiments, having tuberculosis. The diagnosis is published to a user via a graphical user interface 1026.

[0112] In embodiments, to enhance security, subject privacy, and compliance with government regulations, subject data like the subject's marker values 1006 may be deleted after it is used to generate a computer assisted diagnosis. Thus, the sample information will no longer exist as standalone information on the one or more computing devices 1028 implementing the diagnostic framework 1018. Thus, the only subject data available to the computing device 1008 will be integrated into the diagnosis provided by the one or more computing devices.

[0113] FIG. 10 includes an illustrative diagram 1028 of the computing device 1008. The computing device 1008 may contain one or more processing unit(s) 1012 and memory 1014, both of which may be distributed across one or more physical or logical locations. The processing unit(s) 1012 may include any combination of central processing units (CPUs), graphical processing units (GPUs), single core processors, multi-core processors, application-specific integrated circuits (ASICs), programmable circuits such as Field Programmable Gate Arrays (FPGA), and the like. One or more of the processing unit(s) 1012 may be implemented in software and/or firmware in addition to hardware implementations. Software or firmware implementations of the processing unit(s) 1012 may include computer- or machine-executable instructions written in any suitable programming language to perform the various functions described. Software implementations of the processing unit(s) 1012 may be stored in whole or part in the memory 1014.

[0114] Additionally, the functionality of the computing devices 1008 can be performed, at least in part, by one or more hardware logic components. For example, and without limitation, illustrative types of hardware logic components that can be used include Field-

programmable Gate Arrays (FPGAs), Application-specific Integrated Circuits (ASICs), Application-specific Standard Products (ASSPs), System-on-a-chip systems (SOCs), Complex Programmable Logic Devices (CPLDs), etc.

[0115] Computing device 1008 may be connected to a network through one or more network connectors 1016 for receiving and sending information. The network may be implemented as any type of communications network such as a local area network, a wide area network, a mesh network, and ad hoc network, a peer-to-peer network, the Internet, a cable network, a telephone network, and the like. In embodiments, the computing device 1008 have a direct connection to one or more other devices (e.g. devices that output subject 1002 information, like marker values 1006, in electrical or electronic form) without the presence of an intervening network. The direct connection may be implemented as a wired connection or a wireless connection. A wired connection may include one or more wires or cables physically connecting the computing device 1008 to another device. For example, the wired connection may be created by a headphone cable, a telephone cable, a SCSI cable, a USB cable, an Ethernet cable, or the like. The wireless connection may be created by radio frequency (e.g., any version of Bluetooth, ANT, Wi-Fi IEEE 802.11, etc.), infrared light, or the like.

[0116] The computing device 1008 may be a supercomputer, a network server, a desktop computer, a notebook computer, a collection of server computers such as a server farm, a cloud computing system that uses processing power, memory, and other hardware resources distributed across multiple geographic locations, or the like. The computing device 1008 may include one or more input/output components(s) such as a keyboard, a pointing device, a touchscreen, a microphone, a camera, a display, a speaker, a printer, and the like.

[0117] Memory 1014 of the computing device 1008 may include removable storage, non-removable storage, local storage, and/or remote storage to provide storage of computer-readable instructions, data structures, program modules, and other data. The memory 1014 may be implemented as computer-readable media. Computer-readable media includes non-volatile computer-readable storage media, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules, or other data.

Computer-readable storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other non-transmission medium that can be used to store information for access by a computing device.

[0118] The computing device 1008 includes multiple modules that may be implemented as instructions stored in the memory 1014 for execution by processing unit(s) 1012 and/or implemented, in whole or in part, by one or more hardware logic components or firmware. The diagnostic framework 1018 is contained within the computing device 1008 and may be implemented as instructions stored in the memory 1014 for execution by the processing unit(s) 1012, by hardware logic components, or both.

[0119] A scoring module 1012 obtains from an external source an indication of the expression of the tested markers in a sample 1004 as one or more marker value(s) 1006. The marker values 1006 can be obtained from a microarray or any machine connected to the computing device 1008 either directly or through the network connectors 1016. The marker values 1006 may also be previously saved or stored on a separate computing device or computer-readable media prior to being transferred to the scoring module 1020. The marker values 1008 may also be inputted directly by a user, including a physician or laboratory technician, through any appropriate I/O method. Exemplary I/O methods include any methods making use of the previously mentioned input/output components such as a keyboard, camera, microphone, touchscreen, or scanner.

[0120] The scoring module 1020 also obtains a reference level corresponding to the one or more marker values 1006. As with the marker values 1006, the reference levels can be calculated, as previously explained, and stored in a reference level database 1024, on the computing device 1008. Those having skill in the art will appreciate, however, that the one or more reference levels 1024 may, in other embodiments, be obtained either directly or through the network connectors 1016 from one or more separate computing devices, machines, or computer readable media. The reference levels may also be directly inputted by the user.

[0121] The scoring module 1020 may partially process, normalize, rewrite,

anonymize, or otherwise modify the marker values 1006 or reference levels 1024. The scoring module 1020 will generate a score based at least in part on the one or more marker values 1006. In some embodiments this score is equivalent to the one or more marker values. In other embodiments, the score will be generated based at least in part on the marker values 1006 and a weight associated with each corresponding marker. For example, markers with higher sensitivity, specificity, or both could be weighted more heavily than markers with lower sensitivity or specificity. Alternative scores may be generated based on any other previously discussed analytic process.

[0122] The scoring module 1020 provides the generated score to a diagnostic module 1022. The diagnostic module compares the score to the reference level and diagnoses the subject 1002 based on a result of the comparison as having sarcoidosis, not having sarcoidosis, or in some embodiments, having tuberculosis. The diagnosis is published to the user via a graphical user interface 1026.

[0123] Illustrative Process: For ease of understanding, the processes discussed in this disclosure are delineated as separate operations represented as independent blocks. However, these separately delineated operations should not be construed as necessarily order dependent in their performance. The order in which the process is described is not intended to be construed as a limitation, and any number of the described process blocks may be combined in any order to implement the process, or an alternate process. Moreover, it is also possible that one or more of the provided operations is modified or omitted.

[0124] FIG. 11 shows an illustrative process 1100 for diagnosing sarcoidosis.

[0125] At 1102, one or more reference levels are received, as well as an indication of the expression of relevant markers in a sample. The indication of the one or more marker values may be received from a clinician who assayed the sample for the value, or they may be received from a database where the values from a previously performed assay have been stored.

[0126] At 1104, a score is generated at least partly based on the marker value. The score may be the same as the marker value, or it may be additionally based on a weight corresponding to each tested marker, or based in part on any other previously disclosed analytic process. Note that there may be a score for each marker, or there may be a

single score based on an aggregation of data related to multiple marker values.

[0127] At 1106, the score is compared to one or more reference levels.

[0128] At 1108, a subject is diagnosed based on a result of the comparison 1106 as being healthy, having sarcoidosis, or in some embodiments, having tuberculosis.

[0129] In embodiments, the subjects diagnosed with sarcoidosis or tuberculosis using the methods disclosed herein can be effectively treated with the appropriate therapy. As an example, treating subjects with sarcoidosis includes delivering therapeutically effective amounts of an appropriate drug to alleviate one or more symptoms of sarcoidosis or tuberculosis.

[0130] Particular embodiments include:

[0131] Embodiment 1. A method of diagnosing sarcoidosis in a subject including assaying a sample derived from a subject for the presence of one or more markers selected from Small inducible cytokine A21 precursor (CCL21); Methionine aminopeptidase 1 (Metap1); Activated RNA polymerase II transcription cofactor variant 4 (PC4); RNA methyltransferase (CLI_3190); Tumor necrosis factor receptor superfamily member 21 precursor (TNFRSF21); Monocyte differentiation antigen CD14 (CD14); DnaJ (Hsp40) homolog subfamily C member 1 precursor (DNAJC1); Amyloid β A4 precursor protein-binding family B member 1-interacting protein (APBB1); Fibroblast growth factor binding protein 2 precursor (FGFBP-2); or SH3 domain-containing YSC84 like protein 1 (SH3YL1); and diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers, as compared to a reference level for each marker.

[0132] 2. A method of embodiment 1 including assaying the sample for the presence of CCL21; Metap1; CLI_3190; APBB1; and SH3YL1; and diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers.

[0133] 3. A method of embodiment 1 including assaying the sample for the presence of CCL21; Metap1; PC4; CLI_3190; TNFRSF21; and APBB1; and diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers.

[0134] 4. A method of embodiment 1 including assaying the sample for the presence

of CCL21; PC4; CLI_3190; DNAJC1; APBB1; FGFBP-2; and SH3YL1; and diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers.

[0135] 5. A method of embodiment 1 including assaying the sample for the presence of CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; and SH3YL1; and diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers .

[0136] 6. A method of distinguishing sarcoidosis from tuberculosis in a subject including assaying a sample obtained from the subject for the presence of one or markers selected from Ferredoxin (Fed A); WDFY3 protein (WDFY3); Membrane protein (MFS); Leucine rich PPR-motif containing protein (LRPPRC); HLA-DR alpha (HLA-DR); Transketolase (TKT); Dihydroxy acid dehydratase (Rv0189C); Chain A Mycobacterium tuberculosis (BfrA); Disabled homolog 2 isoform 2 (DAB2); or Transcription elongation factor B polypeptide 2 isoform (Homo sapiens) (TCEB2); and diagnosing the subject as healthy, having sarcoidosis or having tuberculosis based on the up- or down-regulation of the one or more markers as compared to a reference level for each marker.

[0137] 7. A method of embodiment 6 including assaying the sample for the presence of Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; and TCEB2; and diagnosing the subject as healthy, having sarcoidosis or having tuberculosis based on the up- or down-regulation of the one or more markers.

[0138] 8. A method of embodiment 6 including assaying the sample for the presence of HLA-DR; MFS; DAB2; BfrA; and WDFY3; and diagnosing the subject as healthy, having sarcoidosis or having tuberculosis based on the up- or down-regulation of the one or more markers.

[0139] 9. A method of embodiment 6 including assaying the sample for the presence of HLA-DR; MFS; and DAB2; and diagnosing the subject as healthy, having sarcoidosis or having tuberculosis based on the up- or down-regulation of the one or more markers.

[0140] 10. A kit for diagnosing sarcoidosis in a subject wherein the kit includes a protein that binds CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1; and a detectable label.

[0141] 11. A kit according to embodiment 10 including one or more proteins that bind

CCL21; Metap1; CLI_3190; APBB1; or SH3YL1; and a detectable label.

[0142] 12. A kit according to embodiment 10 including one or more proteins that bind CCL21; Metap1; PC4; CLI_3190; TNFRSF21; or APBB1; and a detectable label.

[0143] 13. A kit according to embodiment 10 including one or more proteins that bind CCL21, PC4; CLI_3190; DNAJC1; APBB1; FGFBP-2; or SH3YL1, and a detectable label.

[0144] 14. A kit according to embodiment 10 including one or more proteins that bind CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1; and a detectable label.

[0145] 15. A kit for distinguishing sarcoidosis from tuberculosis in a subject wherein the kit includes a protein that binds Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2; and a detectable label.

[0146] 16. A kit according to embodiment 15 including one or more proteins that bind Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2; and a detectable label.

[0147] 17. A kit according to embodiment 15 including one or more proteins that bind HLA-DR; MFS; DAB2; BfrA; or WDFY3; and a detectable label.

[0148] 18. A kit according to embodiment 15 including one or more proteins that bind HLA-DR; MFS; or DAB2; and a detectable label.

[0149] 19. A kit according to any one of embodiments 10-18 wherein the proteins include antibodies, epitopes or mimotopes.

[0150] 20. A kit for diagnosing sarcoidosis in a subject wherein the kit includes a nucleic acid that binds a gene encoding CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1; and a detectable label.

[0151] 21. A kit according to embodiment 20 including one or more nucleic acids that bind a gene encoding CCL21; Metap1; CLI_3190; APBB1; or SH3YL1; and a detectable label.

[0152] 22. A kit according to embodiment 20 including one or more nucleic acids that bind a gene encoding CCL21; Metap1; PC4; CLI_3190; TNFRSF21; or APBB1; and a detectable label.

[0153] 23. A kit according to embodiment 20 including one or more nucleic acids that bind a gene encoding CCL21, PC4; CLI_3190; DNAJC1; APBB1; FGFBP-2; or SH3YL1;

and a detectable label.

[0154] 24. A kit according to embodiment 20 including one or more nucleic acids that bind a gene encoding CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1; and a detectable label.

[0155] 25. A kit for distinguishing sarcoidosis from tuberculosis in a subject wherein the kit includes one or more nucleic acids that bind a gene encoding Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2; and a detectable label.

[0156] 26. A kit according to embodiment 25 including one or more nucleic acids that bind a gene encoding Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2; and a detectable label.

[0157] 27. A kit according to embodiment 25 including one or more nucleic acids that bind a gene encoding HLA-DR; MFS; DAB2; BfrA; or WDFY3; and a detectable label.

[0158] 28. A kit according to embodiment 25 including one or more nucleic acids that bind a gene encoding HLA-DR; MFS; or DAB2; and a detectable label.

[0159] 29. A kit according to any one of embodiments 10-28 wherein the detectable label is a radioactive isotope, enzyme, dye, fluorescent dye, magnetic bead, or biotin.

[0160] 30. A kit according any one of claims 10-29 wherein the kit further includes reagents to perform an enzyme-linked immunosorbent assay (ELISA), a radioimmunoassay (RIA), a Western blot, an immunoprecipitation, an immunohistochemical staining, flow cytometry, fluorescence-activated cell sorting (FACS), an enzyme substrate color method, and/or an antigen-antibody agglutination.

[0161] 31. A method of diagnosing sarcoidosis in a subject including

obtaining a sample from a subject;

assaying the sample for one or more markers selected from CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1

obtaining a value based on the assay;

comparing the value to a reference level; and

diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers as demonstrated by the value and the reference level.

[0162] 32. A method according to embodiment 31 including assaying the sample for one or more markers selected from CCL21; Metap1; CLI_3190; APBB1; or SH3YL1.

[0163] 33. A method according to any one of embodiments 31 or 32 including assaying the sample for one or more markers selected from CCL21; Metap1; PC4; CLI_3190; TNFRSF21; or APBB1.

[0164] 34. A method according to any one of embodiments 31-33 including assaying the sample for one or more markers selected from CCL21; PC4; CLI_3190; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

[0165] 35. A method according to any one of embodiments 31-34 including assaying the sample for one or more markers selected from CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1

[0166] 36. A method of distinguishing sarcoidosis from tuberculosis in a subject including:

obtaining a sample derived from the subject;

assaying the sample for one or more markers selected from Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2;

obtaining a value based on the assay;

comparing the value to a reference level; and

diagnosing the subject as healthy, having sarcoidosis or having tuberculosis based on the up- or down-regulation of the one or more markers as demonstrated by the value and the reference level.

[0167] 37. A method according to embodiment 36 including assaying the sample for one or more markers selected from Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

[0168] 38. A method according to embodiment 36 or 37 including assaying the sample for one or more markers selected from HLA-DR; MFS; DAB2; BfrA; or WDFY3.

[0169] 39. A method according to any one of embodiments 36-38 including assaying the sample for one or more markers selected from HLA-DR alpha (HLA-DR); Membrane protein (MFS); or Disabled homolog 2 isoform 2 (DAB2).

[0170] 40. A method according to any one of embodiments 31-39, wherein assaying the sample for one or more markers include contacting the sample with a probe including a detectable label, wherein the probe binds the marker.

[0171] 41. A method of any one of embodiments 31-40, wherein obtaining a value

based on the assay includes analyzing the binding of the probe to the marker in the sample.

[0172] 42. A method of any one of embodiments 31-41, wherein analyzing the binding of the probe to the marker in the sample includes quantitating the amount of the marker in the sample.

[0173] 43. A method of any one of embodiments 31-42, wherein the sample is a tissue sample, a cell sample, a whole blood sample, a serum sample, a plasma sample, a saliva sample, a sputum sample, or a urine sample.

[0174] 44. A method of any one of embodiments 31-43 wherein the value is a score.

[0175] 45. A method of any one of embodiments 31-44 wherein the score is a weighted score.

[0176] 46. A cDNA library including mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; and (ii) white blood cells obtained from sarcoidosis patients.

[0177] 47. A cDNA library of embodiment 46 further including mRNA isolated from (iii) human splenic monocytes; and/or (iv) embryonic lung fibroblasts.

[0178] 48. A cDNA library of any one of embodiments 46 or 47 wherein the cDNA library is a phage display library.

[0179] 49. A cDNA library of any one of embodiments 46-48 wherein the phage display library is a T7 phage display library.

[0180] 50. A cDNA library of any one of embodiments 46-49 wherein cDNA from each cell type is linked to an identifying sequence or tag.

[0181] 51. A cDNA library of any one of embodiments 46-50 wherein the identifying sequence or tag is a modified linker selected from ECOR1/HindIII; ALA; LEU; and THR.

[0182] 52. A cDNA library of any one of embodiments 46-51 following biopanning.

[0183] 53. A cDNA library of any one of embodiments 46-52 wherein the biopanning includes negative selection and/or enrichment.

[0184] 54. A method of identifying markers to diagnose sarcoidosis including adhering cDNA expression products from a cDNA library of any one of embodiments 46-52 to a microarray; exposing the microarray to samples from sarcoidosis subjects and control subjects; detecting cDNA expression products bound by molecules in the samples from sarcoidosis subjects but not by samples from control subjects; performing data analysis

to identify bound molecules that reliably diagnose sarcoidosis.

[0185] 55. A method of detecting sarcoidosis or tuberculosis antigens by preparing: (a) a phage display library of sarcoidosis or tuberculosis antigens from cells of one or more subjects with sarcoidosis; (b) enriching the phage display library for sarcoidosis or tuberculosis antigens by biopanning; (c) selecting clones for amplification; (d) testing amplified clones for binding to antibodies in sera of sarcoidosis subjects; and (e) sequencing bound clones.

[0186] 56. A method of embodiment 55 wherein the cells are bronchoalveolar cells (BAL) and white blood cells.

[0187] 57. A method of any one of embodiments 55 or 56 wherein the sarcoidosis subject has systemic sarcoidosis, pulmonary sarcoidosis, cutaneous sarcoidosis, Lofgren's syndrome, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis, musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues.

[0188] 58. A method of any one of embodiments 55-58 wherein the detected sarcoidosis antigens are specific to systemic sarcoidosis, pulmonary sarcoidosis, cutaneous sarcoidosis, Lofgren's syndrome, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis, musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues.

[0189] 59. A kit to practice a method of any one of embodiments 31-45 or 54-58.

[0190] 60. A method of identifying markers for sarcoidosis including preparing a cDNA library including mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; and (ii) white blood cells obtained from sarcoidosis patients, (iii) human splenic monocytes; and (iv) embryonic lung fibroblasts; biopanning the cDNA library to isolate clones expressing antigens for sarcoidosis from the cDNA library; and identifying the antigens as markers for sarcoidosis.

[0191] 61. The method of embodiment 60, wherein biopanning includes differential biopanning.

[0192] 62. The method of any one of embodiments 60 or 61, wherein differential biopanning includes using sera from healthy control subjects to remove non-specific IgG.

[0193] 63. The method of any one of embodiments 60-62, wherein differential

biopanning further includes using sarcoidosis sera for positive enrichment.

[0194] 64. The method of any one of embodiments 60-63, wherein identifying the antigens includes immobilizing the clones on a microarray; contacting the antigens in the clones with sera of sarcoidosis patients; and analyzing binding of the antigens to the sera.

[0195] 65. The method of any one of embodiments 60-64, wherein analyzing binding of the antigens to the sera includes quantifying the binding of the antigens to the sera.

[0196] 66. The method of any one of embodiments 60-65, wherein the analyzing binding of the antigens to the sera includes comparing the binding of the antigens to the sera of sarcoidosis patients with the binding of the antigens to the sera of healthy subjects.

[0197] 67. The method of any one of embodiments 60-66, further includes identifying markers for tuberculosis, the method further including obtaining the clones expressing the antigens identified as markers for sarcoidosis, contacting the clones with sera from tuberculosis patients to identify clones expressing antigens for tuberculosis, and identifying the antigens as markers for tuberculosis.

[0198] 68. A microarray including a protein that binds CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

[0199] 69. A microarray including a protein that binds Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

[0200] 70. A microarray including a nucleic acid that binds to a gene encoding CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

[0201] 71. A microarray including a nucleic acid that binds a gene encoding Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

[0202] 72. A microarray including one or more of the following proteins: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

[0203] 73. A microarray including one or more of the following proteins Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

[0204] 74. The microarray of any one of embodiments 68-73, wherein the protein or the nucleic acid on the microarray includes a label that can be detected.

[0205] 75. The microarray of any one of embodiments 68, 69, or 72-74, wherein the microarray includes two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, or nine or more of the proteins on the microarray.

[0206] 76. The microarray of any one of embodiments 70, 71, or 74, wherein the microarray includes two or more, three or more, four or more, five or more, six or more, seven or more, eight or more, or nine or more of the nucleic acids on the microarray.

[0207] 77. A kit comprising the microarray of any one of embodiments 68-76.

[0208] 78. A method of treating a subject having symptoms of sarcoidosis including:
diagnosing whether the subject has sarcoidosis, the diagnosing including,
obtaining a sample from the subject,
contacting the microarray of any one of embodiments 68, 70, or 72 with the sample from the subject,
detecting up- and/or down-regulation of a protein or nucleic acid on the microarray as compared to a reference level, thereby diagnosing the subject has sarcoidosis; and
treating the subject with a drug that alleviates the symptoms of sarcoidosis.

[0209] 79. The method of embodiment 78, wherein the drug is a nonsteroidal anti-inflammatory.

[0210] 80. The method of embodiment 79, wherein the drug is corticosteroids, methotrexate, azathioprine, hydroxychloroquine, chloroquine, cyclophosphamide, chlorambucil, pentoxifylline, thalidomide, infliximab, adalimumab, or colchicine.

[0211] 81. A method of treating a subject having symptoms of tuberculosis and sarcoidosis including:
diagnosing that the subject has tuberculosis, the diagnosing including,
obtaining a biological sample from the subject,
contacting the protein microarray of embodiment 69, 71, or 73 or with the biological sample from the subject,
detecting up- and/or down-regulation of a protein or nucleic acid on the microarray, thereby diagnosing the subject has tuberculosis; and
treating the subject with a drug that alleviate the symptoms of tuberculosis.

[0212] 82. The method of embodiment 81, wherein the drug is isoniazid (INH), rifampin (RIF), ethambutol (EMB), or pyrazinamide (PZA). The Examples below are included to demonstrate particular embodiments of the disclosure. Those of ordinary skill in the art should recognize in light of the present disclosure that many changes can be

made to the specific embodiments disclosed herein and still obtain a like or similar result without departing from the spirit and scope of the disclosure.

[0213] 83. A computer-implemented method of diagnosing subjects having sarcoidosis, the computer-implemented method including:

- receiving at a computer system a value representing an expression of one or more of the following markers in a subject sample: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1;

- generating a score based at least in part on the one or more values and a weight associated with each of the one or more corresponding markers;

- comparing the score to a reference level; and

- diagnosing the subject as having sarcoidosis or not having sarcoidosis based on a result of the comparison.

[0214] 84. A computing device for diagnosing sarcoidosis including:

- a processing unit;

- a memory;

- a user interface;

- a scoring module configured to:

- receive a value representing an expression of each of one or more of the following markers: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1; and

- generate a score based at least in part on the one or more values and a weight associated with each of the corresponding markers; and

- a diagnostic module configured to:

- compare the score to a reference level;

- diagnose the subject as having sarcoidosis or not having sarcoidosis based on a result of the comparison; and

- publish the diagnosis to the user interface.

[0215] 85. A computer-implemented method of distinguishing sarcoidosis from tuberculosis in a subject, the computer-implemented method including:

- receiving at a computer system a value representing an expression of each of one or more of the following markers in a subject sample: Fed A; WDFY3; MFS; LRPPRC;

HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2;

generating a score based at least in part on the value and a weight associated with each of the one or more markers;

comparing the score to a reference level;

diagnosing the subject as healthy, having sarcoidosis, or having tuberculosis based on a result of the comparison; and

publishing a result.

[0216] 86. A computing device for diagnosing sarcoidosis including:

a processing unit;

a memory;

a user interface;

a scoring module configured to:

receive a reference level, and a value representing an expression each of one or more of the following markers in a subject sample: Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2; and

generate a score based at least in part on the value and a weight associated with each of the one or more markers; and

a diagnostic module configured to:

compare the score to a reference level;

diagnose the subject as having sarcoidosis or not having sarcoidosis based on a result of the comparison; and

publish the diagnosis to the user interface.

[0217] 87. The computer-implemented method of embodiments 83 or 85, the method includes receiving a reference level.

[0218] 88. The computer device of embodiments 84 or 86, wherein the scoring module receives a reference level.

[0219] Examples. Significance. Aberrant immune responses are a major cause of a vast array of human diseases. Sarcoidosis is an inflammatory disease of unknown etiology sharing similarities with non-infectious and infectious granulomatous diseases, including Mycobacteria tuberculosis. Tuberculosis (TB) remains a major global health problem. There is a tremendous need to develop accurate tests to diagnose sarcoidosis

and TB. A highly sensitive and specific T7 phage antigen library derived from bronchoalveolar lavage cells and leukocytes of sarcoidosis subjects was developed. This complex cDNA library was biopanned and a microarray was constructed to immunoscreen sera from healthy, sarcoidosis and TB subjects. A panel of specific antigens to classify sarcoidosis from healthy controls and subjects with TB was identified.

[0220] Introduction. Sarcoidosis is an inflammatory granulomatous disease of unknown etiology affecting multiple organs, such as lungs, skin, CNS, and eyes. Common features shared by patients with sarcoidosis are the presence of non-caseating granuloma, a lack of cutaneous reaction to tuberculin skin testing (PPD) and increased local and circulating inflammatory cytokines. In addition, there is evidence of abnormal immune function that presents as cutaneous anergy accompanied by hypergammaglobulinemia. Sarcoidosis shares striking clinical and pathological similarities with infectious granulomatous diseases, especially *Mycobacteria tuberculosis* (MTB). Iannuzzi et al., *N. Engl. J. Med.* 2007; 357(21): 2153-65; Prince et al., *J. Allergy Clin. Immunol.* 2003; 111(2 Suppl): S613-23. Although there is mounting evidence of the presence of nonviable bacterial components (including MTB and *Propionibacterium acnes*) in sarcoidosis tissue (Gupta et al., *Eur. Respir. J.* 2007; 30(3): 508-16; Chen et al., *Am. J. Respir. Crit. Care Med.*; 181(4): 360-73; Negi et al., *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2012; 25(9): 1284-97) all attempts to isolate viable MTB or other microbial pathogens from sarcoidosis tissue have failed. Hunninghake et al. *Sarcoidosis Vasc Diffuse Lung Dis* 1999; 16(2): 149-73; Chen et al. *J. Immunol.* 2008; 181(12): 8784-96.

[0221] Intradermal injection of the Kveim-Siltzbach suspension (a granulomatous splenic tissue suspension) induces granuloma formation weeks later in sarcoidosis patients suggesting the presence of antigen(s) in granuloma tissue and host immunoreactivity to these antigens. Proteomics, genomics, transcriptomics, and high throughput technology clearly suggest that early immune reaction to diverse antigens is highly prevalent in a large number of rheumatic, neoplastic, and inflammatory diseases such as sarcoidosis. Several studies using state-of-the-art technologies have attempted to identify sarcoidosis antigens or to identify the underlying genetic and environmental factors (Hajizadeh et al., *J. Clin. Immunol.* 2007; 27(4): 445-54; Chen et al., *Proc. Am.*

Thorac. Soc. 2007; 4(1): 101-7; Zhang et al., Respiratory research 2013; 14: 18) yet unifying environmental or genetic factors as initiators of this disease have not been found. Hunninghake et al., Sarcoidosis Vasc Diffuse Lung Dis 1999; 16(2): 149-73; Dubaniewicz, Autoimmunity reviews 2010; 9(6): 419-24; Eishi et al., Journal of Clinical Microbiology 2002; 40(1): 198-204; Oswald-Richter & Drake, Seminars in respiratory and critical care medicine 2010; 31(4): 375-9. These studies reported a number of markers or variations in gene expression signatures, which, however, failed to discriminate between sarcoidosis and other inflammatory or granulomatous diseases. Koth et al., Am. J. Resp. Crit. Care 2011; 184(10): 1153-63; Maertzdorf et al. Proc. Natl. Acad. Sci. U S A 2012; 109(20): 7853-8. This is partly due to the fact that several inflammatory diseases may respond to various antigens with activation of a similar transcriptome and/or inflammatory gene expression profiles.

[0222] Because non-caseating granulomas, cutaneous anergy and hypergammaglobulinemia suggest an immune dysfunction in this disease, it was hypothesized that sarcoidosis is triggered by a group of unknown antigens represented in the host immune cells. To identify the elusive antigen(s), a heterologous cDNA library derived from bronchoalveolar cell (BAL) samples and total white blood cells (WBC) from sarcoidosis patients was developed. Both sarcoid-derived libraries were then combined with cultured human monocytes and embryonic lung fibroblast cDNA libraries to build a complex sarcoidosis library (CSL). Furthermore, antibody recognition and random plaque selection was used during biopanning of the cDNA libraries to minimize the confounding effects of autoantibodies unrelated to sarcoidosis. It was tested whether this novel library representing relevant antigens could specifically recognize high IgG titer in sera of sarcoidosis subjects. This approach has been successfully applied in biomarker discovery for the diagnosis of lung, head and neck and breast cancer. Fernandez-Madrid et al., Cancer research 2004; 64(15): 5089-96; Fernandez-Madrid et al., Clinical cancer research : an official journal of the American Association for Cancer Research 1999; 5(6): 1393-400; Lin et al., Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007; 16(11): 2396-405. A feature that distinguishes the described methods from previous studies is that the exquisite power of antibody recognition present

in the sera of sarcoidosis patients was used to interrogate the potential antigens presented in the macrophages and monocytes.

[0223] The present study describes a novel approach to identify sarcoidosis antigens and to detect serum antibodies on high-throughput arrays. Sera from 3 cohorts (sarcoidosis, controls, and TB) were used for immunoscreening. Using bioinformatics tools, a large number of biomarkers with high sensitivity and specificity that can discriminate among the sera of patients with sarcoidosis, healthy controls and MTB was identified. Using the integrative-analysis method that combines results from two independent trials, clones that significantly differentiate sarcoidosis from controls were identified. Similarly, clones that differentially react with TB sera and not with sarcoidosis or control sera were identified. Furthermore, the top 10 discriminating antigens for TB and sarcoidosis were sequenced and homologies were identified in a public data base. These data indicate that a unique library enabling the detection of highly significant antigens to discriminate between patients with sarcoidosis and tuberculosis was developed.

[0224] **Materials and Methods. Chemicals.** All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless specified otherwise. LeukoLOCK filters and RNAlater were purchased from Life Technologies (Grand Island, NY). The RNeasy Midi kit was obtained from Qiagen, (Valencia, CA). The T7 mouse monoclonal antibody was purchased from Novagen (San Diego, CA). Alexa Fluor 647 goat anti-human IgG and AlexFluor goat anti-mouse IgG antibodies were purchased from Life Technologies (Grand Island, NY).

[0225] **Patient selection.** This study was approved by the Institutional Review Board at Wayne State University and the Detroit Medical Center. Patients were recruited at the center for Sarcoidosis and Interstitial Lung Diseases (SILD), which is a referral center for patients with sarcoidosis and other ILDs. Three sources of patient derived materials have been used in this study: A) a BAL cDNA library was derived from BAL cells obtained during diagnostic bronchoscopy from newly diagnosed patients with sarcoidosis (n=20); B) a leukocyte cDNA library were developed from sarcoidosis patients who were followed in outpatient setting with various stages of sarcoidosis (n=36); and C) sera collected from 3 groups: 1) healthy controls, who were volunteers recruited from the community; 2) subjects with biopsy confirmed sarcoidosis who were followed in an outpatient setting;

and 3) sera from subjects with culture positive TB collected at the Detroit Department of Health and Wellness Promotion. Subjects were included who had a diagnosis of sarcoidosis as proven by tissue biopsy per guidelines (Costabel & Hunninghake, Eur Respir J 1999; 14(4): 735-7) and have a negative PPD. TB subjects were included who had a positive TB culture and were HIV negative. Subjects were excluded, who were positive for HIV or were receiving high dose immune suppressive medication that was defined as prednisone more than 15 mg alone or in combination with immune modulatory medications. Subjects who had positive PPD or quantiferon test were excluded from the sarcoidosis group. All study subjects signed a written informed consent.

[0226] Bronchoalveolar lavage: BAL cells were obtained, after informed consent, during diagnostic bronchoscopy from subjects with active sarcoidosis as previously described. Rastogi et al., American journal of respiratory and critical care medicine 2011; 183(4): 500-10. BAL cells were suspended in 500µl of RNAlater and stored at -80°C.

[0227] Collection of total leukocytes from sarcoid subjects. Leukocytes from 36 sarcoid subjects were isolated using whole blood with LeukoLOCK filters as previously described. Glatt et al., Current pharmacogenomics and personalized medicine 2009; 7(3): 164-88.

[0228] Human macrophage (EL-1) and human lung embryonic fibroblast (MRC-5) cell cultures. Both cell lines were obtained from ATCC and cultured as per ATCC recommendations. From each cell line 1-2 mg RNA was isolated to construct the cDNA library.

[0229] Serum collection. Using standardized phlebotomy procedures blood samples were collected and allowed to clot and then centrifuged at 2500 rpm for 10 min. Supernatants were stored at -80°C.

[0230] Construction of T7 phage display cDNA libraries. Total RNA was isolated using the RNeasy Midi kit (Qiagen, Valencia, CA). Integrity of the RNA samples was assessed using the Agilent 2100 bioanalyzer. Total RNA, in the amount of 1-2 mg, was subjected to two cycles of polyA purification to minimize ribosomal RNA contamination as suggested by the manufacturer (Qiagen, Valencia, CA). The construction of phage cDNA libraries was performed using Novagen's Orient Express cDNA Synthesis (Random Primer System) and Cloning system as per manufacturer's suggestions (EMD Biosciences-

Novagen). Each library was cloned using modified linkers that allow identification of the phage clones. Chatterjee et al., Cancer research 2006; 66(2): 1181-90. The number of clones in each of the 4 libraries was titrated by plaque assay as per manufacturer's instructions (EMD Biosciences-Novagen). Finally, the same number of phages from each BAL, WBC, EL-1 and MRC5 library was pooled to generate a complex sarcoid library (CSL).

[0231] Biopanning of T7 phage displayed cDNA library with human sera. Differential biopanning for negative and positive selection was performed using sera from healthy controls to remove the non-specific IgG, and sarcoidosis sera for selective enrichment according to manufacturer's suggestions (T7Select System, TB178; EMD Biosciences-Novagen). Protein G Plus-agarose beads (Santa Cruz Biotechnology) were used for serum IgG immobilization. Four rounds of biopannings were performed and the selected phage libraries were used for microarray immunoscreening. Each cycle of biopanning included passing the entire phage library through protein G beads coated with IgG from pooled sera of healthy controls, then passing through beads coated with IgGs from individual serum of sarcoid subjects. Microarray construction and immunoscreening. Informative phage clones were randomly picked and amplified after several rounds of biopannings and their lysates were arrayed in quintuplicates onto nitrocellulose FAST slides (Grace Biolabs, OR) using the ProSys 5510TL robot (Cartesian Technologies, CA). The nitrocellulose slides were then blocked with a solution of 1% BSA in PBS for 1 hour at room temperature followed by another hour of incubation with serum at a dilution of 1:300 in 1X PBS or plasma at a dilution of 1:100 as primary antibodies, together with mouse anti-T7 capsid antibody (0.15µg/mL) and BL21 E. coli cell lysates (5µg/mL). BL21 E. coli cell lysates were added to remove antibodies specific to E.coli from the serum. The microarrays were then washed three times at room temperature with a solution of PBS/0.1% Tween20 for 4 minutes. Secondary antibodies included goat anti-human IgG Alexa Fluor 647 (red fluorescent dye) 1µg/mL and goat anti-mouse IgG Alexa Fluor 532 (green fluorescent dye) 0.05µg/mL. After 1 hour incubation in the dark, the microarrays were washed 3 times with a solution of PBS/0.1% Tween20 for 4 minutes at room temperature, and 2 times in PBS for 4 minutes at room temperature and then air dried.

[0232] Sequencing of phage cDNA clones. Individual phage clones were PCR

amplified using T7 phage forward primer 5' GTTCTATCCGCAACGTTATGG 3' (SEQ ID NO. 75) and reverse primer 5' GGAGGAAAGTCGTTTTTTGGGG 3' (SEQ ID NO. 76) and sequenced by Genwiz (South Plainfield, NJ), using T7 phage sequence primer TGCTAAGGACAACGTTATCGG (SEQ ID NO. 77).

[0233] Data acquisition and pre-processing. Following the immunoreaction, the microarrays were scanned in an Axon Laboratories 4100 scanner (Palo Alto, CA) using 532 and 647nm lasers to produce a red (Alexa Fluor 647) and green (Alexa Fluor 532) composite image. Using the ImaGene 6.0 (Biodiscovery) image analysis software, the binding of each sarcoid specific peptide with IgGs in each serum was then analyzed and expressed as a ratio of red-to-green fluorescent intensities. The microarray data were further read into the R environment v2.3.0 (Team RDC. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna (Austria). 2004) and processed by a sequence of pre-processing, including background correction, omission of poor quality spots and log2 transformations. Within array loess normalization was performed for each spot and summarized by median of triplicates and followed by between array quantile normalization.

[0234] Statistical analysis. A microarray analysis was performed using sera from sarcoid and healthy controls in two independent sets of experiments. Technical and biological sources of variation were expected in the design of the experiment. As opposed to pooling all datasets, one powerful and robust method is to integrate results from individual datasets. Obtaining a higher confidence list of markers than by using individual datasets was expected. To detect differentially expressed antigens between sarcoidosis samples and healthy controls, an integrative analysis of two datasets was performed. Limma's empirical Bayes moderated t-test identified fold-changes in expression of antigens that differed significantly between sarcoidosis and controls for each dataset separately. Then an integrative-analysis method -- an adaptively-weighted method with one-sided correction (AW-OC) (Li & Tseng, The Annals of Applied Statistics 2011; 5(2A): 994-1019) was performed to combine the statistics from both datasets. The integrative method was designed to test whether an antigen is consistently up- or down regulated in sarcoidosis subjects in both datasets. False Discovery Rate (FDR) was estimated using the Benjamini-Hochberg method. Benjamini & Hochberg. J. R. Stat. Soc. Ser. B 1995; 57:

289–300.

[0235] To identify a panel of markers that classify sarcoidosis samples and controls, a strategy of univariate marker selection followed by multivariate modeling was used. The top antigens differentially expressed in the two groups were selected using the above described AW-OC approach. The top genes that were consistently up- or down- regulated in both datasets were used. The top markers were then required by the supervised classification models to achieve the most sensitivity and specificity in differentiating sarcoid and controls. The multivariate classification models chosen for this study were K-nearest neighbors (KNN) and support vector machine (SVM). The cross-validation technique was used to prevent the overfitting of data analysis due to a large number of antigens used to discriminate between sarcoid and control subjects. The study was performed in two nested 10-fold cross-validation loops, an inner loop to select the optimal number of antigens and an outer loop to measure the optimized model performance with estimation of the area under the receiver operating characteristic (AUROC) sensitivity and specificity. The receiver operating characteristic curves were estimated through 10-fold cross-validation. A moderated t-test was carried out to identify the significant clones between healthy controls, sarcoidosis and tuberculosis.

[0236] Results. Generation of cDNA libraries representative of sarcoidosis antigens. Both PBMCs and alveolar macrophages (AMs) play an important role in initiation of sarcoidosis granuloma. It has been shown that extracts from sarcoidosis BAL cells and peripheral blood monocytes (PBMCs) are able to initiate a Kveim-like reaction. Siltzbach & Ehrlich, *The American Journal of Medicine* 1954; 16(6): 790-803; Holter et al., *The American Review of Respiratory Disease* 1992; 145(4 Pt 1): 864-71. Therefore, total BAL cells and WBCs from patients with biopsy proven sarcoidosis were used to develop a cDNA antigen library. BAL cells and WBC were used as sources of antigens in order to increase the diversity of sarcoidosis antigens. To increase the chance of identifying sarcoidosis antigen(s), RNA was isolated from BAL samples obtained from 20 patients with active sarcoidosis to generate the BAL cDNA library. The patients' characteristics are shown in Table 1 (left panel). The LeukoLock system was used to isolate RNA from total leukocytes (WBC) obtained from a different cohort of 36 sarcoidosis subjects to build the WBC cDNA library. The patients' characteristics are shown in Table 1 (right panel).

Table 1. Subject Demographics, Chest X-Ray Stages, and organ involvements

BAL derived RNA		Leukocyte derived RNA	
Age (Mean±SEM)	30±8	Age (Mean±SEM)	36±11.2
BMI (Mean±SEM)	27.7 ± 8.7	BMI (Mean±SEM)	31±5.4
Gender, N (%)		Gender, N (%)	
Male	7 (33)	Male	12 (33)
Female	13 (67)	Female	24 (67)
Race, N (%)		Race, N (%)	
African American	17(87)	African American	32 (88)
White	3 (13)	White	4 (12)
CXR stage, N (%)		CXR Stage, N (%)	
1	2 (6)	1	1 (3)
2	14(67)	2	13(41)
3	4 (27)	3	12(37)
4	0	4	6 (19)
Lung	18	Lung	33
Extrapulmonary	16	Extrapulmonary	31
Neuro-ophthalmologic	6	Neuro-ophthalmologic	11
Skin	6	Skin	13
Liver	2	Liver	4
Heart	1	Heart	2
Prednisone	1	Prednisone	3
IMD	0	IMD	14
Smoking		Smoking	
None	12	None	26

Age, BMI and disease duration values are presented as means and variability in SD or range where indicated. N=Number of patients and percent shown in parentheses. IMD=Immunomodulatory drugs

[0237] Two other sources of cDNA, one from cultured human splenic monocytes (EL-1) and another from lung embryonic fibroblasts (MRC5) were used to generate two additional libraries. These sources were added to increase the chance of discovering potential sarcoidosis antigens. Each cDNA underwent two cycles of PolyA selection to minimize ribosomal contamination. These four libraries were developed as described in the Materials and Methods section. Each library was cloned using modified linkers; ECOR1/HindIII was used for BAL cDNA, ALA for WBC cDNA, LEU for MARC5 cDNA and THR for EL1 cDNA (FIG. 6). The use of these linkers enabled identification of the original library for each antigen.

[0238] Differential biopanning of sarcoidosis phage cDNA display libraries. The four phage cDNA display libraries (BAL, WBC, EL-1 and MARC5) were combined to generate a complex sarcoidosis library (CSL). To isolate a large panel of antigens, differential biopanning of the T7 phage cDNA display library was performed on the combined complex sarcoid library. A negative biopanning selection was done using 10 pooled sera from healthy controls to remove non-specific IgG, while 2 sarcoidosis sera were used for positive selective enrichment. One serum was obtained from a woman (P51) with systemic sarcoidosis who had uveitis and another serum was collected from a male subject (P197) who had active systemic sarcoidosis with renal involvement. Both patients had pulmonary involvements. Each clone was derived either from P51 or from P197. The titer of the complex library was assessed (FIG. 7A) and individual phage clones were amplified by PCR (FIG. 7B).

[0239] High-throughput protein microarray immunoreaction to select sarcoidosis specific antigens. A total of 1152 potential antigen antigens were randomly selected from the two highly enriched pools of T7 phage cDNA libraries (FIG. 1). These antigen antigens were robotically spotted on nitrocellulose Fast slides and were hybridized with sera of sarcoidosis patients or healthy controls. The binding of each of the arrayed potential sarcoidosis-specific peptides with antibodies in sera was quantified with Alexa Fluor 647 (red-fluorescent dye)-labeled goat anti-human antibody. The amount of phage particles at each spot throughout the microarray was detected using a mouse monoclonal antibody to the T7 capsid protein and quantified using Alexa Fluor 532 (green-fluorescent dye)-labeled goat anti-mouse antibody (FIG. 1). To correct for any small variation in the amount of antibody binding in each spot that may be due to different amounts of phage spotted on the microarray, the ratio of intensity of Alexa Fluor 647 over Alexa Fluor 532 was calculated for each spot. Following immunoreaction, the microarray data were processed by a sequence of transformations and then analyzed. The intra-assay reproducibility was assessed by comparing the results among five replicates printed within the same chip for each clone.

[0240] Selection of a panel of antigens and estimation of neural network classifier performance in sarcoidosis. A novel aspect of the described work was the integration of data from two independent trials of printing allowing the development of two data sets

obtained from two independent cohorts of sarcoidosis patients and healthy controls utilized for hybridization. To generate the first dataset, sera from 54 sarcoidosis subjects and 45 healthy controls were immune-screened against 1152 sarcoidosis specific peptides. In a second dataset, sera from 19 healthy controls and 61 sarcoidosis subjects were similarly immune-screened with 1152 potential sarcoidosis specific antigens. Sera used in both data sets for hybridization had not been previously used for biopanning or selection of clones. Table 2 shows the clinical characteristics of sarcoidosis and healthy control subjects.

Table 2. Patient characteristics		Control Subjects
Age	29.7±13.4 y	33±7.4
BMI	29±10.4	28±3.6
Gender, N		
Female	87 (75)	48(75)
Male	28 (25)	16(25)
Race, N		
African American	107 (89)	44(69)
White	8(11)	20(31)
CXR stage, N		
0	3(2)	NA
1	18 (15)	NA
2	49 (43)	NA
3	45(39)	NA
Organ Involvements,		
Neuro-ophthalmologic	33(28)	NA
Lung	109(94)	NA
Skin	50(45)	NA
Multiorgan	70(52)	NA

- Some Patients had multiple organ involvements
- NA= Not Applicable

[0241] Within array loess normalization was performed for each spot and summarized by median of triplicates and followed by between array quantile normalization. After preprocessing, 1101 antigens common in both datasets were used for further analysis. Univariate and multivariate analyses were performed. Limma's empirical Bayes moderated t-test was used to identify fold-changes in expression of antigens that differed significantly between sarcoidosis and controls for each dataset separately. Then both datasets were combined using an integrative-analysis method -- an adaptively-weighted

method with one-sided correction (AW-OC). Li & Tseng, The Annals of Applied Statistics 2011; 5(2A): 994-1019. Out of the 1101 potential antigen, 259 showed a strong differentiation between sarcoidosis and healthy control subjects with adjusted p value (q value) <0.05 and FDR (false discovery rate) <0.05. FIG. 2A shows the heatmap of the 259 significant antigens that were differentially expressed in both datasets. Seventy eight markers out of 259 were consistently up- or down-regulated in sarcoidosis subjects. FIG. 2B shows the AUROC for this classifier. KNN method performed slightly better than SVM. Using the highly significant 32 antigens selected by AW.OC and KNN methods to classify sarcoidosis and healthy controls (AW.OC+KNN), the area under the curve (AUROC) was 0.78, with a sensitivity of 89% and a specificity of 83% estimated after 10-fold cross-validation (FIG. 2B).

[0242] Characterization of 10 most significant sarcoid antigens. Based on the results of AW-OC integrative-analysis, the top 10 high performance antigens that predict sarcoidosis were identified. To further characterize the performance of each clone, the AU-ROC, and sensitivity and specificity given the optimal cutoff of the clones was calculated. FIG. 3 depicts the ROC curves for individual sarcoid antigens and their adjusted p value (q value). As shown, each antigen has a different specificity and sensitivity as well as ROC to predict the presence of sarcoidosis. ROC for these antigens ranged from the highest of 0.84 to the lowest of 0.7. Nine of 10 antigens were clearly up-regulated, whereas one was down-regulated. To further characterize the identified antigens, these 10 highest ranked antigens were sequenced. After obtaining the sequences of clones, the Expasy program was used to translate the cDNA sequences to protein sequences. Protein blast using Blastn and tblastn algorithms of the BLAST program were applied to identify the highest homology to identified proteins or peptides and these results were compared with corresponding nucleotide sequences using nucleotide blast. The predicted amino acid in frame with phage T7 gene 10 capsid proteins was also determined. Five Antigens (PC4, SAMDHI, DNAJC1, TPT1 and SH3YL1) among the top 10 fit the definition of an epitope containing known gene products in the reading frame known genes. The other five contained peptides coded by the inserted gene fragments leading to out of frame peptides, which fits the definition of mimotopes. Among the 10 high performance clones, nine were up-regulated and only

one was down-regulated in sarcoidosis versus healthy controls. FIG. 8 shows the full length of proteins and genes of 10 sarcoidosis clones. Without being bound by theory, as sarcoidosis sera reacted to these out of frame peptides, it is likely that these clones represent sarcoidosis antigens produced as a result of altered reading frames or alternative splicing. Interestingly, when a similar technique was applied to discovery of cancer antigens, numerous out of frame peptides were discovered. Lin et al., Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007; 16(11): 2396-405. Table 3 shows the 10 most significant sarcoidosis antigens, gene names and q-values.

Clone	Up-Regulated in Sarcoidosis Vs Healthy	Gene Name	q Value AUC	Sensitivity // Specificity %, 95% CI
P51_BP3_287 (MRC5)	Small inducible cytokine A21 precursor	CCL21	1.9×10^{-20} 0.84	78 // 82
P51_BP3_281 (BAL)	Methionine aminopeptidase 1	Metap1	1.0×10^{-20} 0.78	70 // 82
P51_BP4_388 (EL-1)	Activated RNA polymerase II transcription cofactor variant 4	PC4	0.00045 0.75	70 // 74
P51_BP4_596 (WBC)	RNA methyltransferase	CLI_3190	0.00045 0.72	72 // 74
P51_BP4_566 (WBC)	Tumor necrosis factor receptor superfamily member 21 precursor. Also known as death receptor 6 (DR6)	TNFR SF21	0.0009 0.74	70 // 71

P51_BP3_283 (WBC)	Monocyte differentiation antigen CD14	CD14	0.0009 0.74	68 // 65
P51_BP3_47 (EL-1)	DnaJ (Hsp40) homolog subfamily C member 1 precursor	DNAJC 1	0.002 0.72	60 // 82
P197_BP4_885 (BAL)	Amyloid β A4 precursor protein-binding family B member 1-interacting protein	APBB1	0.007 0.79	75 // 82
P51_BP4_577 (BAL)	Fibroblast growth factor binding protein 2 precursor	FGFBP- 2	0.009 0.70	64 // 68
Clone	Down-Regulated In Sarcoidosis vs Healthy Controls	Gene Name	q Value & AUC	Sensitivity & Specificity %, 95% CI
P197_BP4_755 (BAL)	SH3 domain-containing YSC84 like protein 1	SH3YL1	1.0×10^{-20} 0.77	65 // 82

[0243] Complex sarcoidosis library detects novel antigens in the sera of tuberculosis patients. In view of the clinical and pathological similarities between MTB and sarcoidosis, a most useful clinical antigen(s) should discriminate between these two conditions. To this end, using the antigens identified by biopanning the CSL library a microarray was constructed, then this construct was interrogated with sera from 17 culture positive MTB subjects. Using a moderate t-test and a q value <0.05 in this system, 238 clones differentially expressed between TB and healthy controls and 380 clones differentially expressed between TB and sarcoidosis were identified. FIG. 4 shows a Venn diagram depicting the overlap between 259 sarcoidosis markers, 238 TB vs. control and 380 TB vs. sarcoidosis markers. Clearly, 47 clones differentiate both sarcoidosis and TB from healthy controls, while 5 of them cannot differentiate sarcoidosis from TB significantly. From these clones, 164 were found to be TB specific, and different from both healthy controls and sarcoidosis clones. FIG. 5 show the heatmap of 50

significant clones differentially expressed in all three groups. Similarly to the sarcoidosis antigens, the specificity and sensitivity of TB clones was analyzed to predict the presence of TB (Table 4). Finally, 10 TB antigens were sequenced and sequence homologies were searched using the same algorithm as previously described. Table 4 shows the 10 TB-specific antigens as compared to healthy controls as well as sarcoidosis.

Clone	Up-Regulated in TB vs Sarcoidosis Subjects	Gene Name	q Value	AUC	Sensitivity // Specificity %, 95% CI
P51_BP3_174 (MRC5)	Ferredoxin (Mycobacterium tuberculosis)	Fed A	4.9×10^{-15}	0.87	88 // 83
P51_BP4_610 (BAL)	WDFY3 protein (Homosapiens)	WDFY 3	4.1×10^{-12}	0.92	88 // 84
P51_BP3_ 266 (EL-1)	Membrane protein (Mycobacterium tuberculosis)	MFS	6.7×10^{-10}	0.9	82 // 93
P51_BP3-166 (BAL)	Leucine rich PPR- motif containing protein (Homosapiens)	LRPP RC	1.3×10^{-9}	0.81	71 // 90
P51_BP4_704 (BAL)	HLA-DR alpha (Homosapiens)	HLA- DR	1.1×10^{-8}	0.89	94 // 83
P197_BP4_ 763 (BAL)	Transketolase (Mycobacterium tuberculosis)	TKT	2.7×10^{-6}	0.86	82 // 76
P51-BP4_ 563 (BAL)	Dihydroxy acid dehydratase (Mycobacterium tuberculosis)	Rv018 9C	1.04×10^{-6}	0.85	76 // 86
Clone	Down-Regulated in TB vs Sarcoidosis				

	Subjects				
P51_BP3_113 (BAL)	Chain A Mycobacterium tuberculosis	BfrA	1.2×10^{-10}	0.9	88 // 85
P51_BP3_200 (BAL)	Disabled homolog 2 isoform 2 (Homosapiens)	DAB2	1.5×10^{-9}	0.92	82 // 91
P51_BP4_622 (BAL)	Transcription elongation factor B polypeptide 2 isoform (Homosapiens)	TCEB2	6.9×10^{-7}	0.89	82 // 89

[0244] After sequence analysis and homology search, one identical sequence between TB and sarcoidosis clone was identified. Although the identified clone's name was different: P51_BP3_287 versus P51_BP3_174, and they performed differently in sarcoidosis versus TB as indicated in q value (compare Table 3 and Table 4). However, using NCBI blast databases (mycobacterium toxoid and the universal blast) on the same sequence, two different proteins could be identified. FIG. 9 shows the full length of protein and genes of 10 TB antigens. Surprisingly, TB clones show much higher sensitivity and specificity; similarly the AUROC was larger for the majority of TB antigens (Table 4).

[0245] Discussion. The described work was inspired by the classic observation that the intradermal injection of a suspension of granulomatous splenic tissue (Kveim-Siltzbach test) induces granuloma formation weeks later in patients with sarcoidosis, suggesting the presence of antigen(s) in granuloma tissue and host immunoreactivity to those antigen(s). Kveim-like effects have also been observed using non-viable BAL cell extracts or PBMCs derived from sarcoidosis subjects. Several studies have attempted to identify specific antigens that can discriminate sarcoidosis from normal subjects or from patients with other granulomatous diseases such as TB (Hajizadeh et al., J. Clin. Immunol. 2007; 27(4): 445-54; Chen & Moller, Proc. Am. Thorac. Soc. 2007; 4(1): 101-7) but, most of these studies used limited proteomics or genomics to search for tissue antigens. Hajizadeh et al., J. Clin. Immunol. 2007; 27(4): 445-54; Richter et al., Am. J. Resp. Crit.

Care 1999; 159(6): 1981-4; Song et al., The Journal of Experimental Medicine 2005; 201(5): 755-67. Here, using novel high throughput technology, the current gap was overcome by constructing phage-protein microarrays in which peptides derived from a unique sarcoidosis cDNA library were expressed as a sarcoidosis phage fusion protein. The phage-protein microarrays were screened to identify phage-peptide clones that bind antibodies in serum samples from patients with sarcoidosis but not in those from controls. Importantly, the same microarray constructs were immune-screened using sera of culture positive TB patients.

[0246] The average length of identified peptides for sarcoidosis antigens was between 9-130 amino acids (AA), while the average peptide length for TB antigens was 9-209 AA. Among 10 sarcoidosis specific phage peptides, 5 expression sequence tags with in frame epitopes were identified. Five other reactive antigens were relatively short out of frame peptides meeting the criteria to be considered as mimotopes (mimetic sequence of a true epitope). Similarly, among 10 sequenced TB specific phage peptides, 5 in frame epitopes with full length in frame proteins with homology to known human sequences were identified. Five other sequences were relatively short peptides with homology to various known MTB proteins (Table 4).

[0247] Interestingly, TB antigens had much higher specificity and sensitivity as compared to antigens selective to sarcoidosis as indicated by higher AUCs (Table 4). Although the significance of mimotopes is not clear, it has been shown that some out of frame peptides are immunogenic and can activate MHC class I molecules. Due to smaller peptide sequences of mimotopes, they may have homology with diverse proteins. Prior studies using similar techniques in various cancers had similarly identified out of frame peptides. Lin et al., Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007; 16(11): 2396-405; Wang et al., Autoantibody signatures in prostate cancer. N. Engl. J. Med. 2005; 353(12): 1224-35; Chatterjee et al., Cancer Research 2006; 66(2): 1181-90. Detection of mimotopes in the described methods may be due to out of frame peptide synthesis secondary to altered ribosomal function, or may correspond to open reading frames, or generation of displayed peptides due to competition for binding during phage selection during phage insertion.

[0248] Although the primary goal was to identify the immune signature in sarcoidosis, a panel of antigens differentially expressed in sarcoidosis and tuberculosis as compared to healthy subjects was also identified. Tables 3 and 4 summarize the 10 most significant clones identified in sarcoidosis and tuberculosis respectively.

[0249] In recent years several groups have attempted to identify specific signatures to distinguish between tuberculosis and sarcoidosis using transcriptomics or gene expression profilings. Koth et al., *Am. J. Resp. Crit. Care* 2011; 184(10): 1153-63; Maertzdorf et al., *Proc. Natl. Acad. Sci. U S A* 2012; 109(20): 7853-8; Berry et al., *Nature* 2010; 466(7309): 973-7. Yet most of these methods led to the discovery of a series of markers or expression signatures that failed to discriminate between these two diseases. Koth et al., *American journal of respiratory and critical care medicine* 2011; 184(10): 1153-63; Stone et al., *PLoS One* 2013; 8(1): e54487. This is partly due to the fact that several inflammatory or infectious diseases such as CD, lupus, sarcoidosis and tuberculosis may respond to various antigens with activation of similar transcriptomes and/or inflammatory gene expression profiles. For instance, Maertzdorf et al. found more similarity in the activated pathways than differences between sarcoidosis and MTB. *Proc. Natl. Acad. Sci. U S A* 2012; 109(20): 7853-8. Their results in sarcoidosis were similar to those results by Berry indicating the importance of the interferon pathway (IFN) signature in MTB. Maertzdorf et al., *Proc. Natl. Acad. Sci. U S A* 2012; 109(20): 7853-8; Berry et al., *Nature* 2010; 466(7309): 973-7. In addition, considerable pathway overlap was identified between lupus, sarcoidosis and TB. Maertzdorf et al., *Proc. Natl. Acad. Sci. U S A* 2012; 109(20): 7853-8. However, despite similar genetic or transcriptomic signatures, these diseases are clinically entirely different and require different therapy. Tuberculosis, a global infectious disease caused by the intracellular bacterium *Mycobacterium tuberculosis* remains a worldwide health problem (<http://www.who.int>). One barrier for eradication of tuberculosis besides the lack of effective vaccination is the lack of reliable antigen to evaluate the activity of the disease and its response to treatment. Nahid et al., *Am. J. Resp. Crit. Care* 2011. 184(8): 972-9. Standard methods to diagnose TB and to monitor response to treatment rely on sputum microscopy and culture. The current CDC/NIH roadmap emphasizes the need for development of new TB antigens as alternative methods. Nahid et al., *Am. J. Resp. Crit. Care* e 2011. 184(8): 972-9. In view

of this background, perhaps surprisingly, the described microarray platform could discriminate tuberculosis from sarcoidosis and healthy controls. In addition to antigens for sarcoidosis, more than 300 clones specifically for tuberculosis were detected. Interestingly, a considerable number of these clones were TB specific and related to bacterial growth of mycobacterium tuberculosis, and its metabolism (Table 4). Recently a tremendous effort has been put toward elucidating the antibody response to MTB antigens, which has implications for the development of new antigens to diagnose and monitor successful treatment, as well as to develop effective vaccination. Kunnath-Velayudhan et al., Proc. Natl. Acad. Sci. U S A 2010; 107(33): 14703-8. Yet, a consistent immune response to MTB has not been found. Most other studies searching for antigens in TB have identified unspecific markers primarily involving host response such as C-reactive protein or serum amyloid A and others, but not MTB specific antigens. Agranoff et al., Lancet 2006; 368(9540): 1012-21; De Groote et al., PLoS One 2013; 8(4): e61002. MTB has the ability to survive within host macrophages, largely escaping immune surveillance and maintaining its ability for replication and person to person transmission. Meena & Rajni, The FEBS journal 2010; 277(11): 2416-27.

[0250] The primary goal of the described project was to discover antigens related to sarcoidosis. Yet, in addition specific antigens for TB were detected. These results are surprising, as the question remains, how can the sarcoidosis library detect TB specific antigens? Lungs are environmentally highly exposed to numerous bacteria, and the described library is predominantly derived from BAL cells that contain all types of immune cells, including macrophages that might have integrated messages from MTB. Without being bound by theory, this could be the reason why the CSL was able to detect TB specific antigens. Still, the major question is why BAL cells of patients with sarcoidosis can harbor MTB messages, yet respond to PPD skin testing with anergy, as all donors with sarcoidosis were PPD negative.

[0251] Similar to gene-expression profiling and the pattern-recognition approaches utilizing serum proteomics, the described methods may have the limitations of background signals, and sample-selection bias. To minimize these problems, an integrative-analysis method, an adaptively-weighted statistical method on two sets of data acquired in two independent experiments was applied. The discriminatory power of

antibody signatures was validated by analyzing data from two completely different cohorts of patients.

[0252] In summary, a novel T7 phage display library derived from macrophages from BAL, monocytes from blood leukocytes of patients with sarcoidosis that may display a significant segment of the universe of potential sarcoidosis and MTB antigens that can be specially recognized by high IgG antibodies in sarcoidosis and MTB sera was developed. The described results support the hypothesis that sarcoidosis sera can recognize antigens presented in sarcoidosis materials. Current study of the antibody response can advance how proteomics can be used to harness immunity to identify and treat diseases, because it investigates antibody–antigen interactions and also evaluates the effects on antibody responses of pathogen and host characteristics.

[0253] Standard reference works setting forth the general principles of immunology include Abbas et al., *Cellular and Molecular Immunology* (6th Ed.), W.B. Saunders Co., Philadelphia, 2007; Janeway et al., *Immunobiology. The Immune System in Health and Disease*, 6th ed., Garland Publishing Co., New York, 2005; Delves et al. (eds.) *Roitt's Essential Immunology* (11th ed.) Wiley-Blackwell, 2006; Roitt et al., *Immunology* (7 th ed.) C.V. Mosby Co., St. Louis, Mo. (2006); Klein et al., *Immunology* (2nd ed), Blackwell Scientific Publications, Inc., Cambridge, Mass., (1997).

[0254] Additionally, methods particularly useful for polyclonal and monoclonal antibody production, isolation, characterization, and use are described in the following standard references: Harlow et al., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1988); Harlow et al., *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998; *Monoclonal Antibodies and Hybridomas: A New Dimension in Biological Analyses*, Plenum Press, New York, N.Y. (1980); Zola et al., in *Monoclonal Hybridoma Antibodies: Techniques and Applications*, CRC Press, 1982).

[0255] As will be understood by one of ordinary skill in the art, each embodiment disclosed herein can comprise, consist essentially of or consist of its particular stated element, step, ingredient or component. Thus, the terms “include” or “including” should be interpreted to recite: “comprise, consist of, or consist essentially of.” As used herein, the transition term “comprise” or “comprises” means includes, but is not limited to, and

allows for the inclusion of unspecified elements, steps, ingredients, or components, even in major amounts. The transitional phrase “consisting of” excludes any element, step, ingredient or component not specified. The transition phrase “consisting essentially of” limits the scope of the embodiment to the specified elements, steps, ingredients or components and to those that do not materially affect the embodiment. As used herein, a material effect would cause a statistically-significant reduction in the ability to diagnose a sarcoidosis subject from a healthy subject or a sarcoidosis subject from a tuberculosis subject.

[0256] Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term “about.” Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. When further clarity is required, the term “about” has the meaning reasonably ascribed to it by a person skilled in the art when used in conjunction with a stated numerical value or range, i.e. denoting somewhat more or somewhat less than the stated value or range, to within a range of $\pm 20\%$ of the stated value; $\pm 19\%$ of the stated value; $\pm 18\%$ of the stated value; $\pm 17\%$ of the stated value; $\pm 16\%$ of the stated value; $\pm 15\%$ of the stated value; $\pm 14\%$ of the stated value; $\pm 13\%$ of the stated value; $\pm 12\%$ of the stated value; $\pm 11\%$ of the stated value; $\pm 10\%$ of the stated value; $\pm 9\%$ of the stated value; $\pm 8\%$ of the stated value; $\pm 7\%$ of the stated value; $\pm 6\%$ of the stated value; $\pm 5\%$ of the stated value; $\pm 4\%$ of the stated value; $\pm 3\%$ of the stated value; $\pm 2\%$ of the stated value; or $\pm 1\%$ of the stated value.

[0257] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[0258] The terms “a,” “an,” “the” and similar referents used in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., “such as”) provided herein is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the invention.

[0259] Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member may be referred to and claimed individually or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group may be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[0260] Certain embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Of course, variations on these described embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventor expects skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

[0261] Furthermore, numerous references have been made to patents, printed publications, journal articles and other written text throughout this specification (referenced materials herein). Each of the referenced materials are individually incorporated herein by reference in their entirety for their referenced teaching.

[0262] In closing, it is to be understood that the embodiments of the invention disclosed herein are illustrative of the principles of the present invention. Other modifications that may be employed are within the scope of the invention. Thus, by way of example, but not of limitation, alternative configurations of the present invention may be utilized in accordance with the teachings herein. Accordingly, the present invention is not limited to that precisely as shown and described.

[0263] The particulars shown herein are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of various embodiments of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for the fundamental understanding of the invention, the description taken with the drawings and/or examples making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

[0264] Definitions and explanations used in the present disclosure are meant and intended to be controlling in any future construction unless clearly and unambiguously modified in the following examples or when application of the meaning renders any construction meaningless or essentially meaningless. In cases where the construction of the term would render it meaningless or essentially meaningless, the definition should be taken from Webster's Dictionary, 3rd Edition or a dictionary known to those of ordinary skill in the art, such as the Oxford Dictionary of Biochemistry and Molecular Biology (Ed. Anthony Smith, Oxford University Press, Oxford, 2004).

CLAIMS

What is claimed is:

1. A kit for diagnosing sarcoidosis in a subject wherein the kit comprises a protein that binds Small inducible cytokine A21 precursor (CCL21); Methionine aminopeptidase 1 (Metap1); Activated RNA polymerase II transcription cofactor variant 4 (PC4); RNA methyltransferase (CLI_3190); Tumor necrosis factor receptor superfamily member 21 precursor (TNFRSF21); Monocyte differentiation antigen CD14 (CD14); DnaJ (Hsp40) homolog subfamily C member 1 precursor (DNAJC1); Amyloid β A4 precursor protein-binding family B member 1-interacting protein (APBB1); Fibroblast growth factor binding protein 2 precursor (FGFBP-2); or SH3 domain-containing YSC84 like protein 1 (SH3YL1) and a detectable label.
2. A kit for diagnosing sarcoidosis in a subject wherein the kit comprises a nucleic acid that binds a gene encoding CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1 and a detectable label.
3. A kit for distinguishing sarcoidosis from tuberculosis in a subject wherein the kit comprises a protein that binds Ferredoxin (Fed A); WDFY3 protein (WDFY3); Membrane protein (MFS); Leucine rich PPR-motif containing protein (LRPPRC); HLA-DR alpha (HLA-DR); Transketolase (TKT); Dihydroxy acid dehydratase (Rv0189C); Chain A Mycobacterium tuberculosis (BfrA); Disabled homolog 2 isoform 2 (DAB2); or Transcription elongation factor B polypeptide 2 isoform (Homo sapiens) (TCEB2) and a detectable label.
4. A kit for distinguishing sarcoidosis from tuberculosis in a subject wherein the kit comprises a nucleic acid that binds a gene encoding Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2 and a detectable label.
5. The kit according to any one of claims 1 or 3 wherein the proteins comprise antibodies, epitopes or mimotopes.
6. The kit according to any one of claims 1-4 wherein the detectable label is a radioactive isotope, enzyme, dye, fluorescent dye, magnetic bead, or biotin.
7. The kit according any one of claims 1-4 wherein the kit comprises reagents to perform an enzyme-linked immunosorbent assay (ELISA), a radioimmunoassay (RIA), a Western blot, an immunoprecipitation, an immunohistochemical staining, flow cytometry,

fluorescence-activated cell sorting (FACS), an enzyme substrate color method, and/or an antigen-antibody agglutination.

8. A method of diagnosing sarcoidosis in a subject comprising

obtaining a sample from the subject;

assaying the sample for one or more markers selected from CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1 and a detectable label

obtaining a value based on the assay;

comparing the value to a reference level; and

diagnosing the subject as healthy or having sarcoidosis based on the up- or down-regulation of the one or more markers as demonstrated by the value and the reference level.

9. A method of distinguishing sarcoidosis from tuberculosis in a subject comprising

obtaining a sample from the subject;

assaying the sample for one or markers selected from Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2;

obtaining a value based on the assay;

comparing the value to a reference level; and

diagnosing the subject as healthy, having sarcoidosis or having tuberculosis based on the up- or down-regulation of the one or more markers as demonstrated by the value and the reference level.

10. The method according to any one of claims 8 or 9, wherein the sample is a tissue sample, a cell sample, a whole blood sample, a serum sample, a plasma sample, a saliva sample, a sputum sample, or a urine sample.

11. The method according to any one of claims 8 or 9, wherein assaying the sample for one or more markers comprises contacting the sample with a probe comprising a detectable label and that binds the one or more markers.

12. The method according to any one of claims 8 or 9, wherein obtaining a value based on the assay comprises quantitating the amount of the marker in the sample.

13. The method according to any one of claims 8 or 9 wherein the value is a score.

14. The method according to claim 13 wherein the score is a weighted score.

15. A cDNA library comprising mRNA isolated from (i) bronchoalveolar cells (BAL) of

sarcoidosis patients; and (ii) white blood cells obtained from sarcoidosis patients.

16. The cDNA library according to claim 15 further comprising mRNA isolated from (iii) human splenic monocytes; and/or (iv) embryonic lung fibroblasts.

17. The cDNA library according to claim 15 or 16 wherein the cDNA library is a phage display library.

18. The cDNA library according to claim 17 wherein the phage display library is a T7 phage display library.

19. The cDNA library according to claim 15 or 16 wherein cDNA from each cell type is linked to an identifying sequence or tag.

20. The cDNA library according to claim 19 wherein the identifying sequence or tag is a modified linker selected from ECOR1/HindIII; ALA; LEU; and THR.

21. The cDNA library according to claims 15 or 16 obtained following biopanning.

22. The cDNA library according to claim 21 wherein the biopanning comprises negative selection and/or enrichment.

23. A method of identifying markers to diagnose sarcoidosis comprising adhering cDNA expression products from a cDNA library according to claim 22 to a microarray; exposing the microarray to samples from sarcoidosis subjects and control subjects; detecting cDNA expression products bound by molecules in the samples from sarcoidosis subjects but not by samples from control subjects; and performing data analysis to identify bound molecules that diagnose sarcoidosis.

24. A method comprising detecting sarcoidosis or tuberculosis antigens by preparing: (a) a phage display library of sarcoidosis or tuberculosis antigens from cells of one or more subjects with sarcoidosis; (b) enriching the phage display library for sarcoidosis or tuberculosis antigens by biopanning; (c) selecting clones for amplification; (d) testing amplified clones for binding to antibodies in sera of sarcoidosis subjects; and (e) sequencing bound clones.

25. A method according to claim 24 wherein the cells are bronchoalveolar cells (BAL) and white blood cells.

26. A method according to claim 24 wherein the sarcoidosis subject has systemic sarcoidosis, pulmonary sarcoidosis, cutaneous sarcoidosis, Lofgren's syndrome, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis,

musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues.

27. A method according to claim 26 wherein the detected sarcoidosis antigens are specific to systemic sarcoidosis, pulmonary sarcoidosis, cutaneous sarcoidosis, Lofgren's syndrome, neurosarcoidosis, cardiac sarcoidosis, ocular sarcoidosis, hepatic sarcoidosis, musculoskeletal sarcoidosis, renal sarcoidosis, or sarcoidosis with the involvement of other organs or tissues.

28. A kit to practice a method according to any one of claims 8-14 or 23-27.

29. A method of identifying markers for sarcoidosis comprising preparing a cDNA library comprising mRNA isolated from (i) bronchoalveolar cells (BAL) of sarcoidosis patients; and (ii) white blood cells obtained from sarcoidosis patients, (iii) human splenic monocytes; and (iv) embryonic lung fibroblasts; biopanning the cDNA library to isolate clones expressing antigens for sarcoidosis from the cDNA library; and identifying the antigens as markers for sarcoidosis.

30. The method according to claim 29, wherein biopanning comprises differential biopanning.

31. The method according to claim 30, wherein differential biopanning comprises using sera from control subjects to remove non-specific IgG.

32. The method according to claim 31, wherein differential biopanning further comprises using sarcoidosis sera for positive enrichment.

33. The method according to claim 29, wherein identifying the antigens comprises immobilizing the clones on a microarray; contacting the antigens in the clones with sera of sarcoidosis patients; and analyzing binding of the antigens to the sera.

34. The method according to claim 33, wherein analyzing binding of the antigens to the sera comprises quantifying the binding of the antigens to the sera.

35. The method according to claim 33, wherein the analyzing binding of the antigens to the sera comprises comparing the binding of the antigens to the sera of sarcoidosis patients with the binding of the antigens to the sera of control subjects.

36. The method according to claim 29, further comprising identifying markers for tuberculosis, the method further comprising obtaining the clones expressing the antigens identified as markers for sarcoidosis, contacting the clones with sera from tuberculosis

patients to identify clones expressing antigens for tuberculosis, and identifying the antigens as markers for tuberculosis.

37. A microarray comprising a protein that binds CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

38. A microarray comprising a protein that binds Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

39. A microarray comprising a nucleic acid that binds to a gene encoding CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

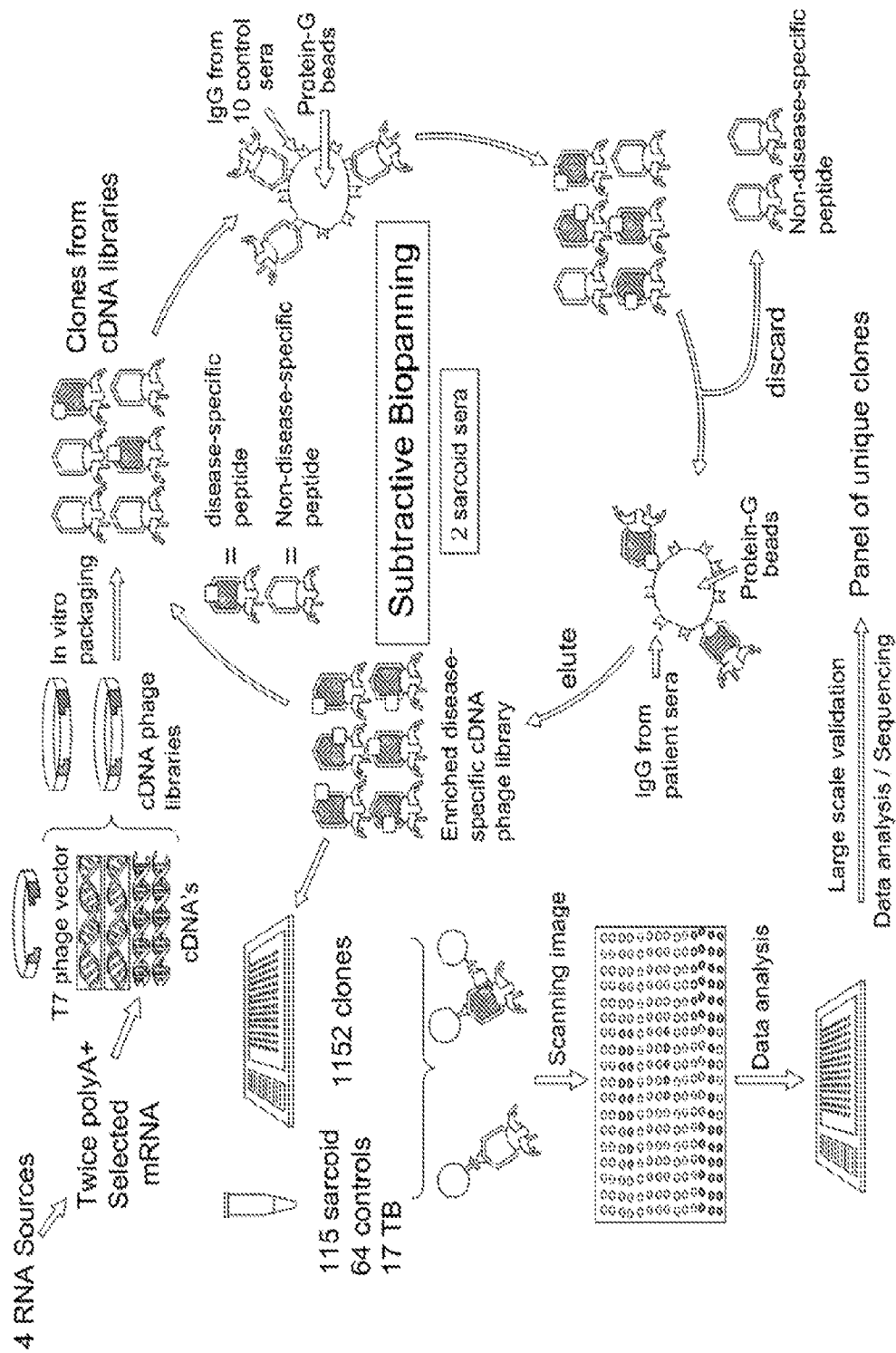
40. A microarray comprising a nucleic acid that binds a gene encoding Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

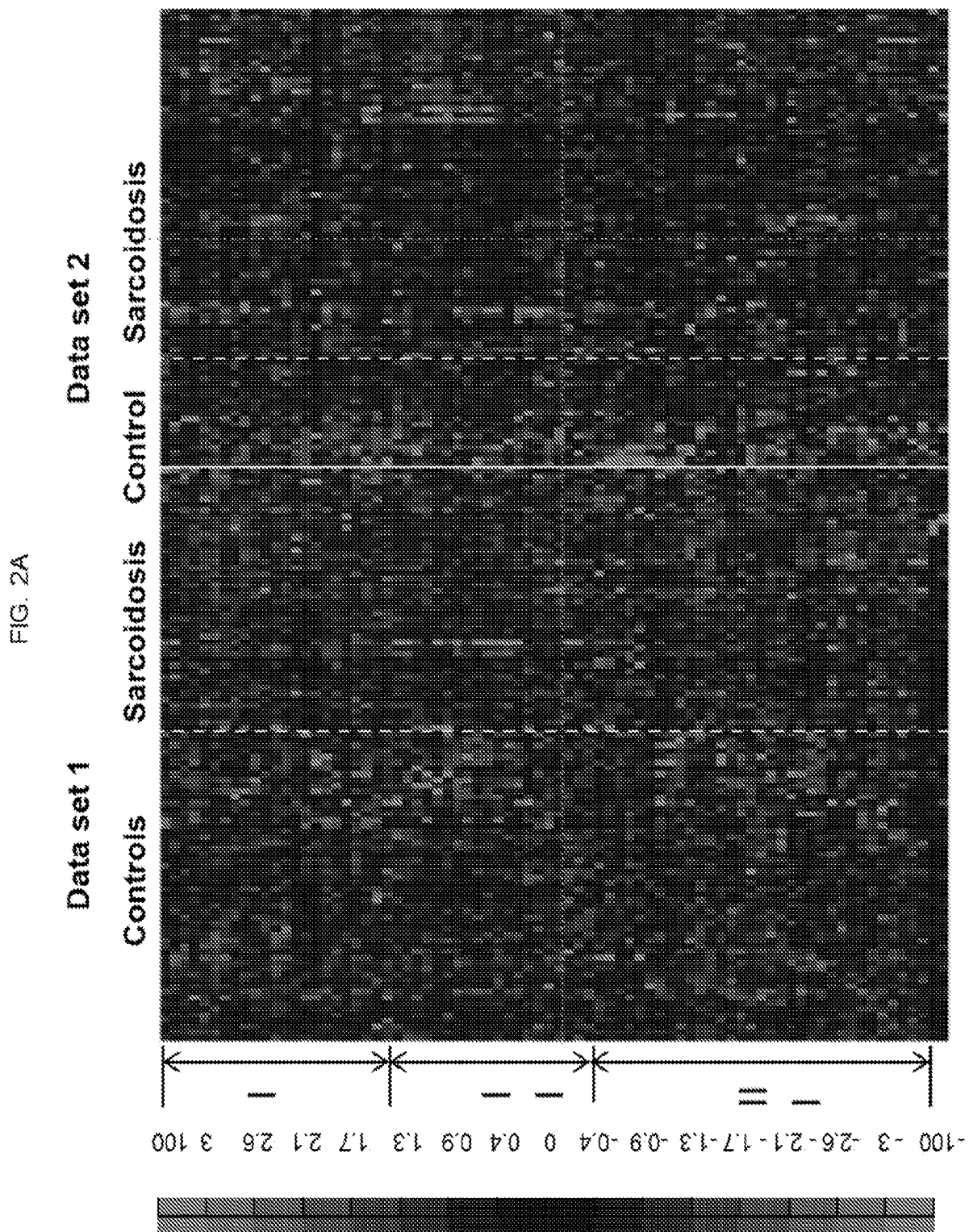
41. A microarray comprising one or more of the following proteins: CCL21; Metap1; PC4; CLI_3190; TNFRSF21; CD14; DNAJC1; APBB1; FGFBP-2; or SH3YL1.

40. A microarray comprising a nucleic acid that binds a gene encoding Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2

42. A microarray comprising one or more of the following proteins Fed A; WDFY3; MFS; LRPPRC; HLA-DR; TKT; Rv0189C; BfrA; DAB2; or TCEB2.

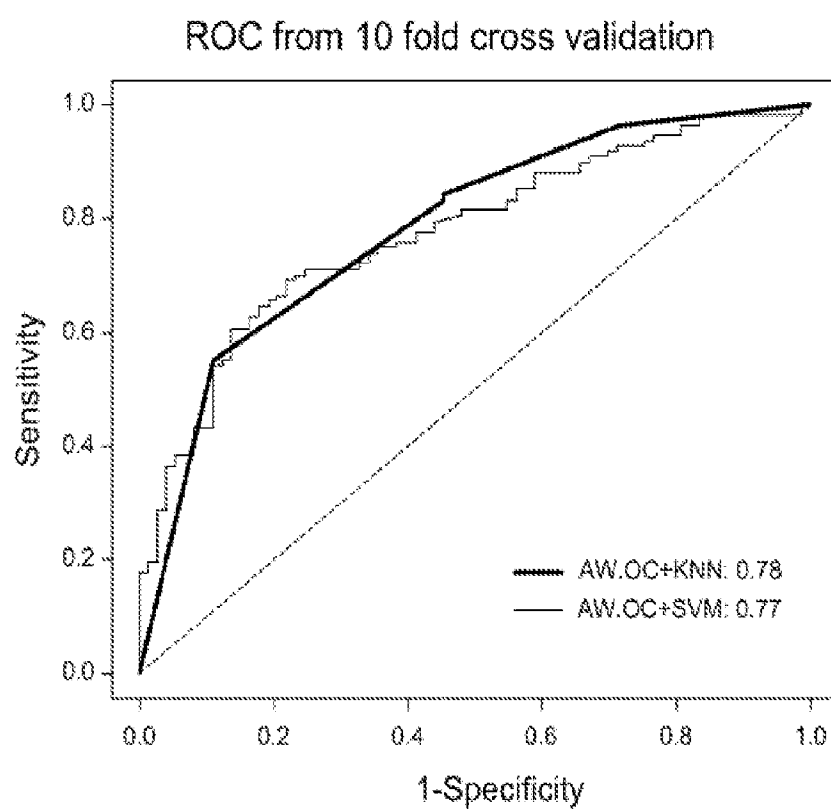
FIG. 1





3/24

FIG. 2B



4/24

FIG. 3A

P51BP3-287 (CCL21)

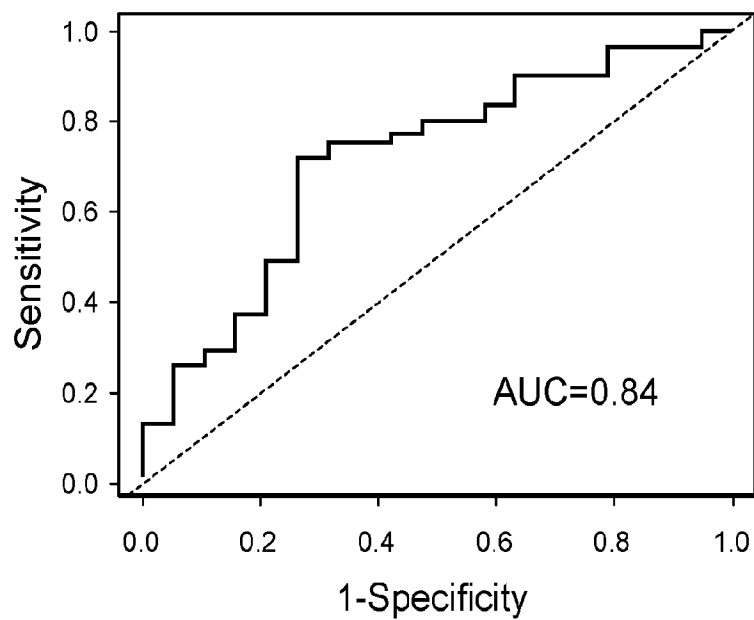
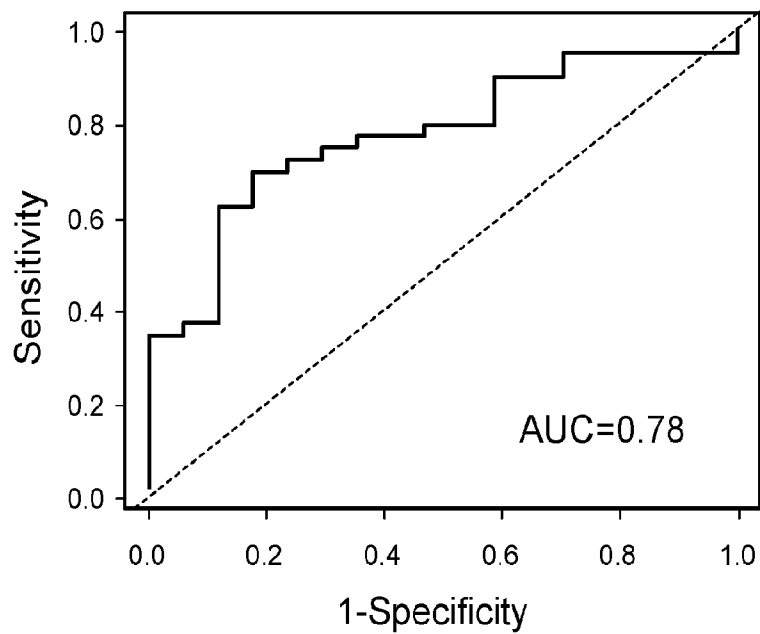


FIG. 3B

P51BP3-281 (Metap1)



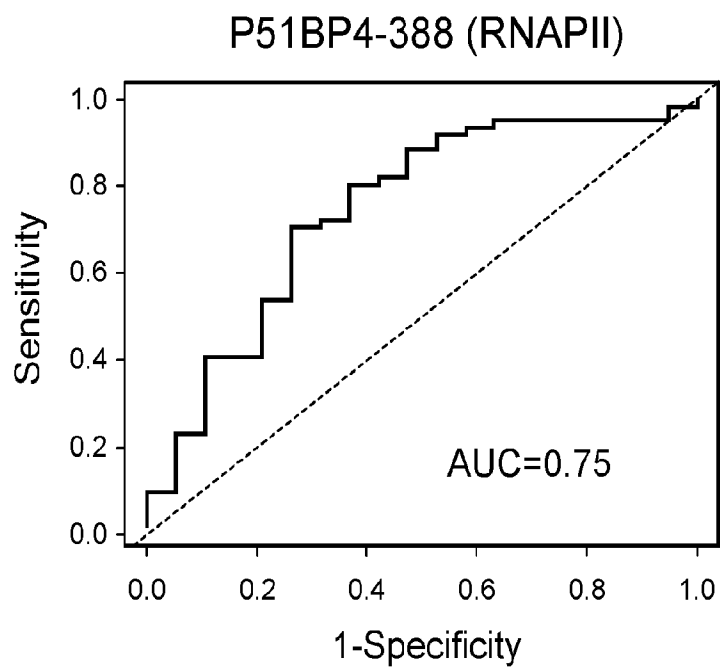
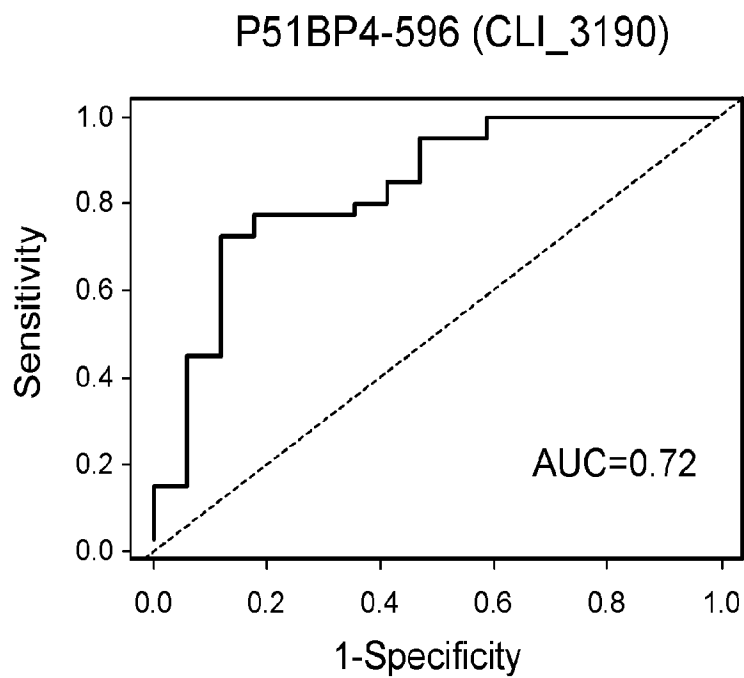
5/24
FIG. 3C

FIG. 3D



6/24

FIG. 3E

P51BP4-566 (TNFRSF21)

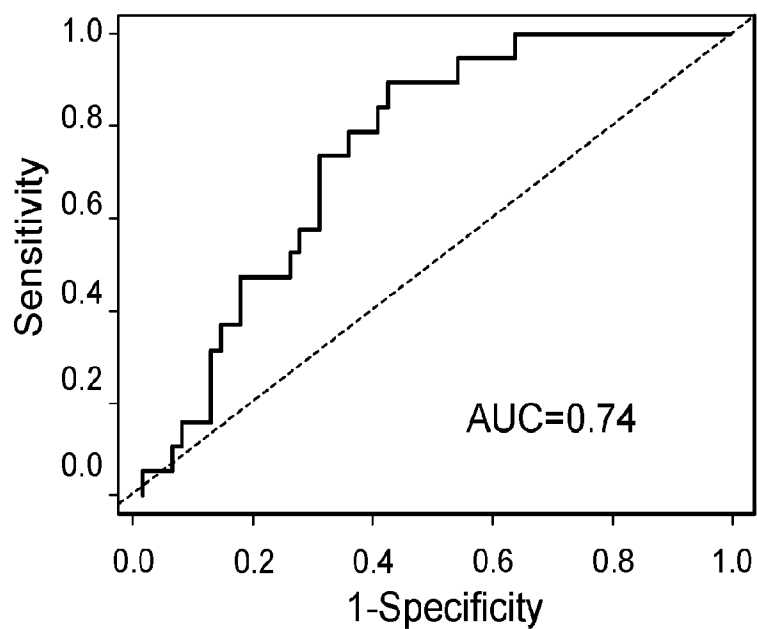
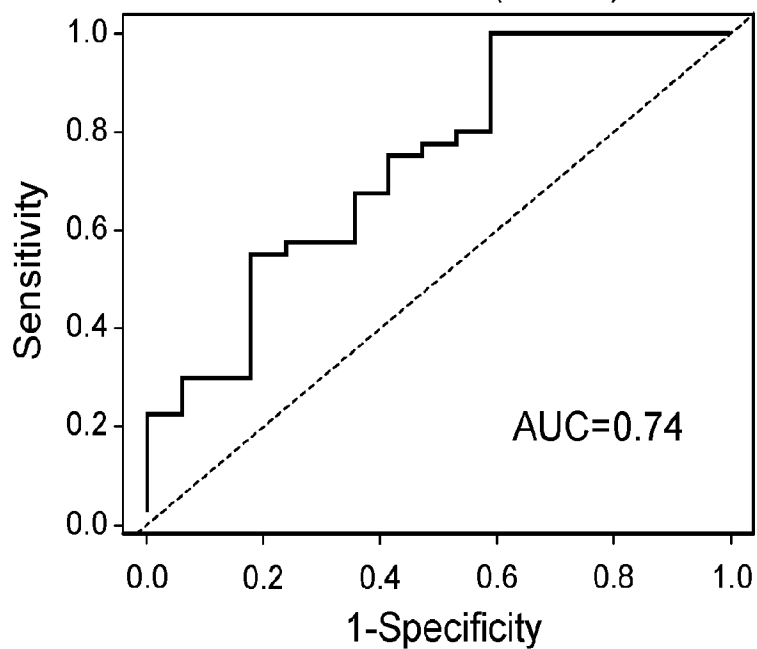


FIG. 3F

P51BP3-283 (CD14)



7/24

FIG. 3G

P51BP3-47 (DNAJC1)

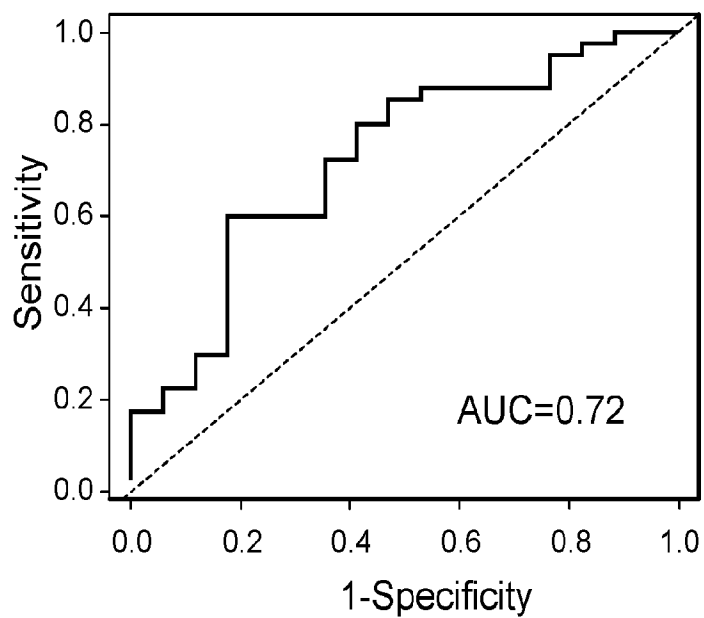
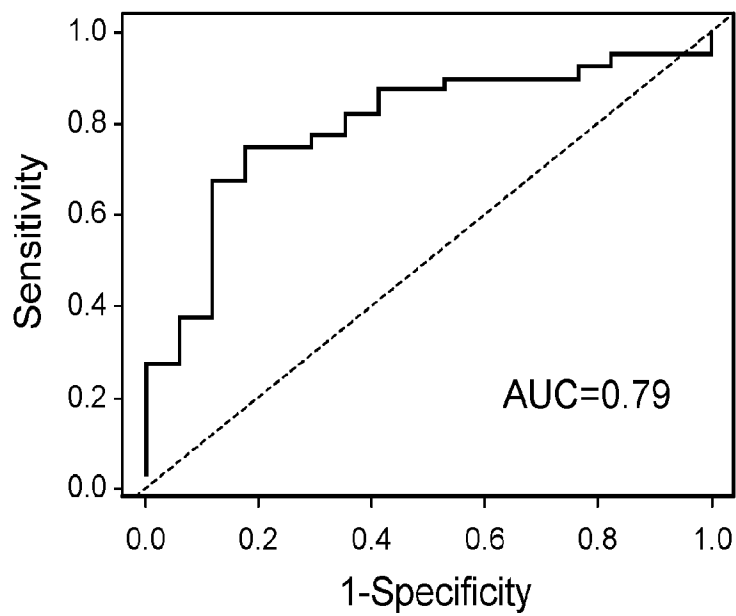


FIG. 3H

P197BP4-885 (APBB1)



8/24

FIG. 3I

P51BP4-577 (FGFBP-2)

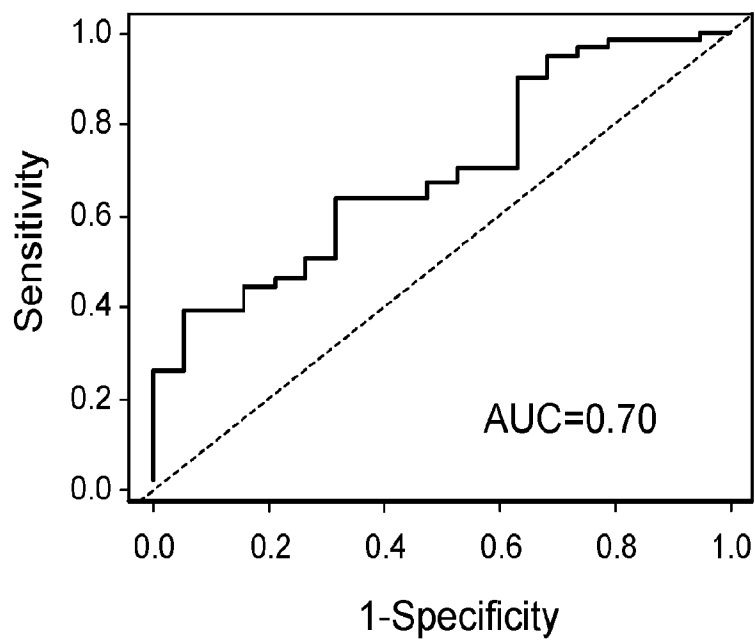
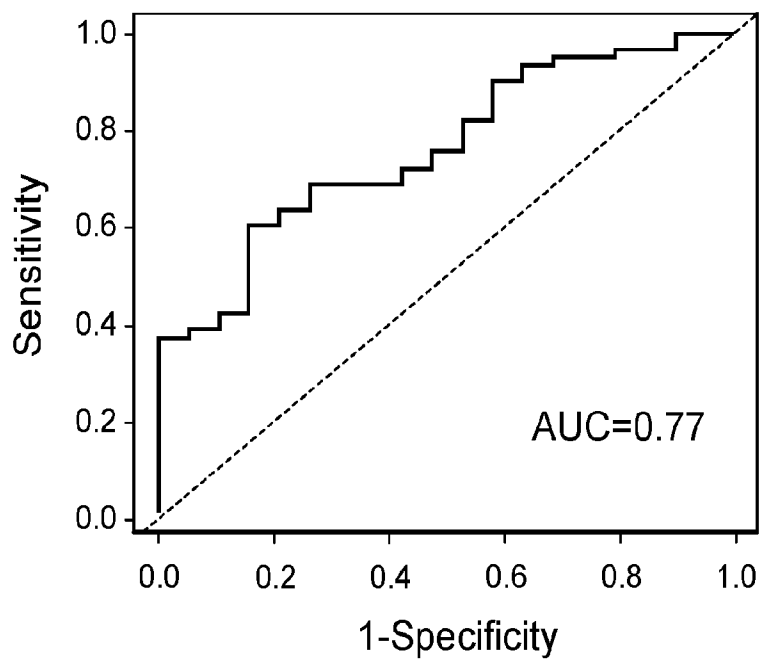


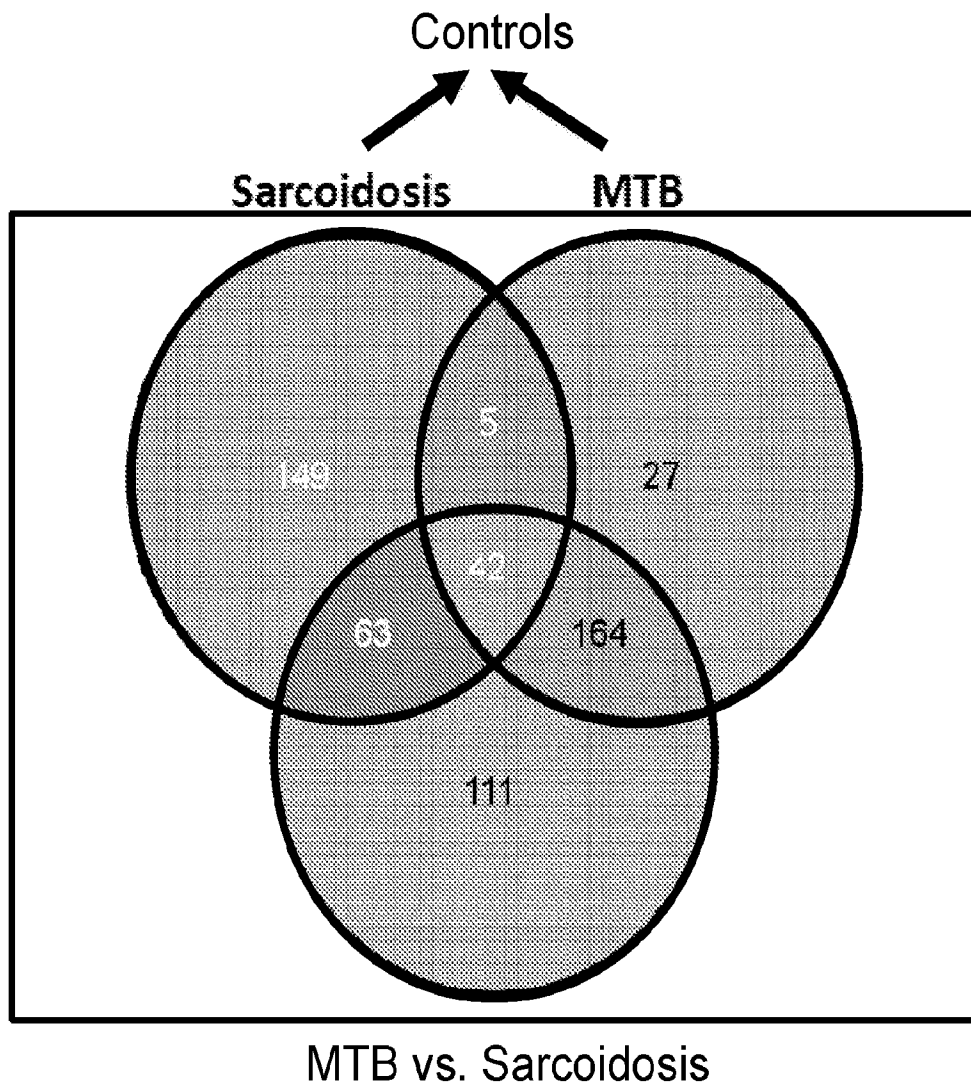
FIG. 3J

P197BP4-755 (SH3YL1)



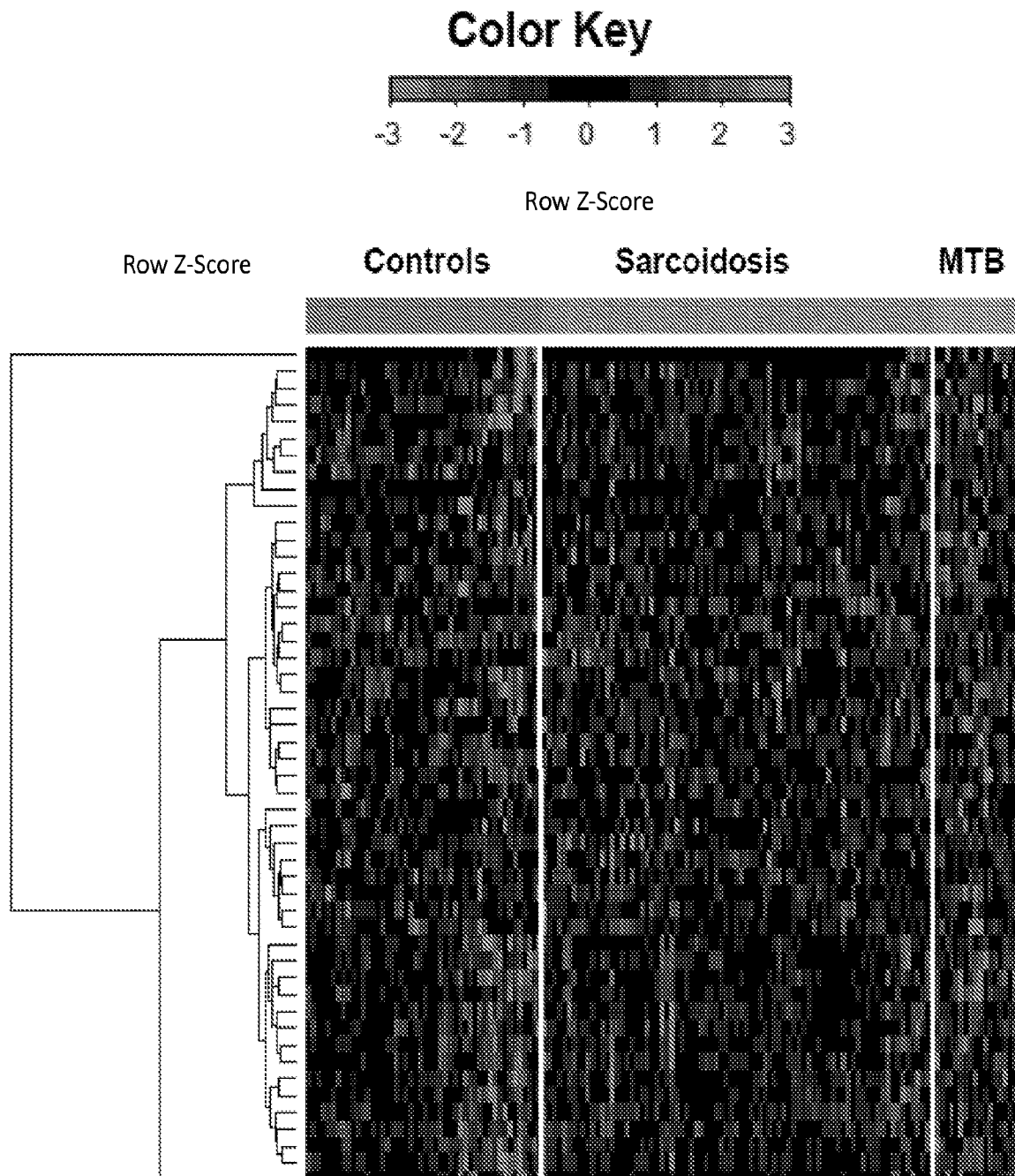
9/24

FIG. 4



10/24

FIG. 5A



11/24

FIG. 5B

P52-BP3-278
P197-BP4-888
P197-BP4-881
P197-BP4-888
P51-BP4-442
P51-BP4-453
P51-BP4-485
P197-BP4-751
P197-BP4-393
P51-BP4-438
P51-BP3-331
P197-BP4-903
P197-BP4-763
P197-BP4-392
P197-BP4-762
P197-BP4-370
P197-BP4-1060
P197-BP4-927
P197-BP4-363
P197-BP4-364
P51-BP4-517
P51-BP4-423
P51-BP4-513
P51-BP4-378
P51-BP4-716
P51-BP4-661
P51-BP4-582
P51-BP4-705
P197-BP4-1038
P197-BP4-342
P197-BP4-1026
P197-BP4-359
P197-BP4-345
P197-BP4-766
P51-BP3-34
P51-BP3-266
P51-BP3-22
P51-BP3-200
P51-BP3-288
P51-BP3-11
P51-BP3-178
P51-BP3-35
P51-BP3-140
P51-BP3-128
P51-BP4-887
P51-BP3-222
P51-BP3-281
P51-BP3-232
P51-BP3-184

FIG. 6

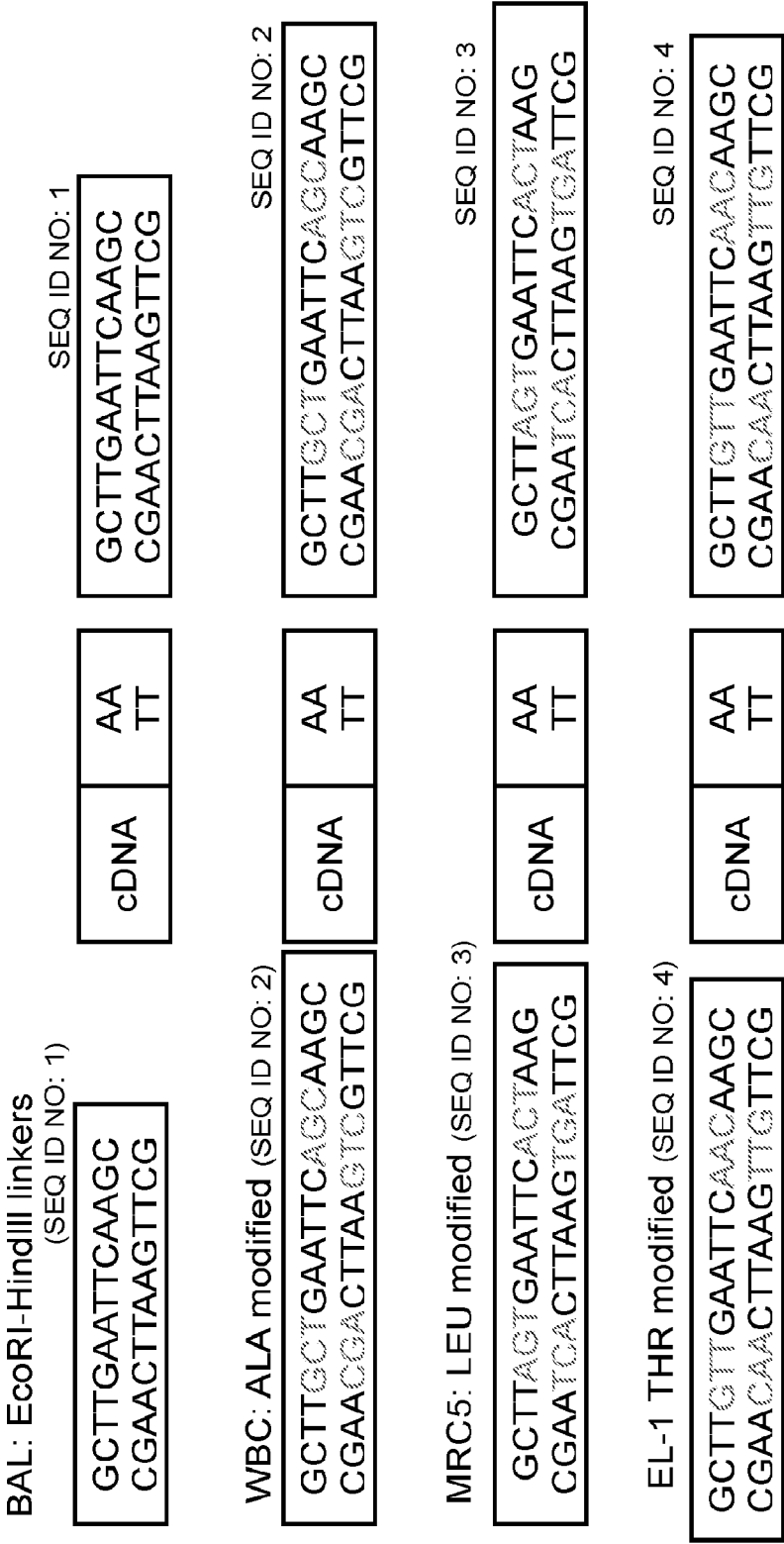


FIG. 7B

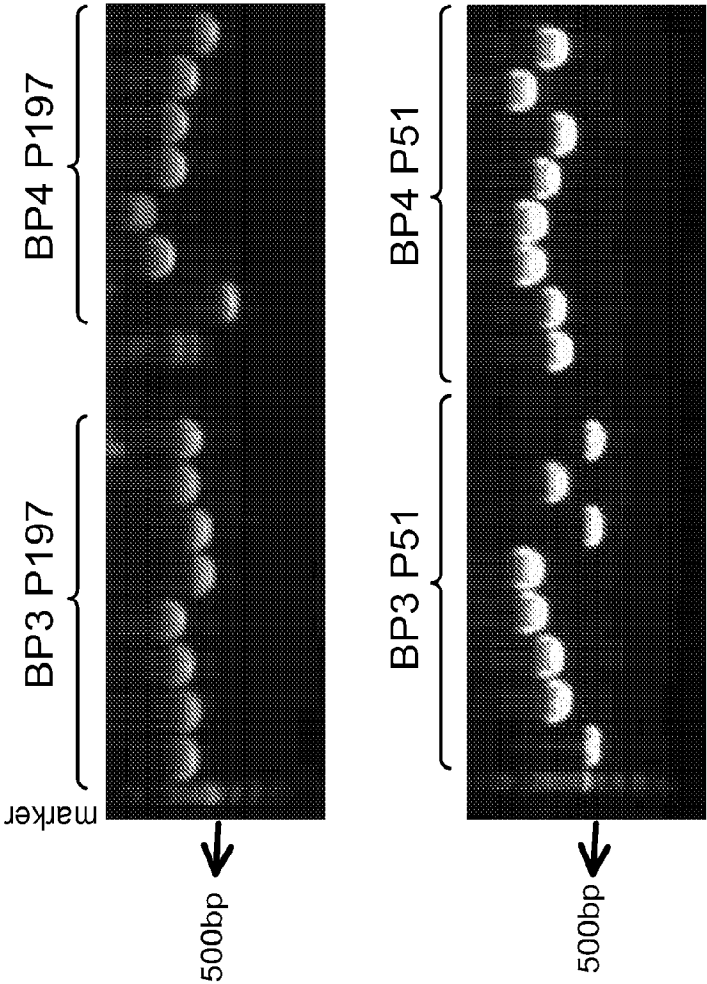
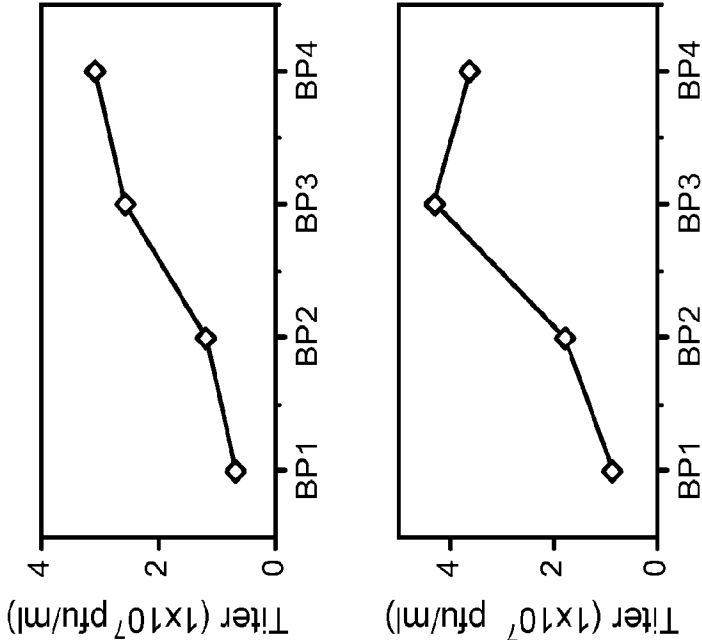


FIG. 7A



14/24

FIG. 8

Rank	Clone and peptide size	Peptide sequences of mimotopes in-frame with T7 108 gene	Description of the sequences that Mimotopes mimic	Region of similarity of peptide
1	PSIBP3_287 (9aa)	IQHQHLGQI (SEQ ID NO: 5)	Small inducible cytokine A21 precursor [Oncorhynchus mykiss] Sequence ID: gbAA031336.1 (SEQ ID NO: 6)	Id=77 (100%) Gaps=0/7 (0%) Length=110 Query 2 IQHQHLGQ 8 (SEQ ID NO: 5, Amino Acids 2-8) QHQHLLGQ (SEQ ID NO: 8) Sbjct 46 QHQHLLGQ 52 (SEQ ID NO: 6, Amino Acids 46-52) Id=78 (76%) Gaps=0/8 (0%) Length=333 Query 1 IQHQHLGQ 9 (SEQ ID NO: 5, Amino Acids 1-9) IQH HL QH (SEQ ID NO: 9) Sbjct 225 IQHQHLEQI 333 (SEQ ID NO: 7, Amino Acids 325-333)
2	PSIBP3_281 (16aa)	AGISRELVDK LAAALE (SEQ ID NO: 10)	Chain A, Human Meap1 Sequence ID: p28161.1 (SEQ ID NO: 11)	Id=111/11 (100%) Gaps=0/11 (0%) Length=326 Query 8 ELVDKLAALALE 16 (SEQ ID NO: 10, Amino Acids 6-16) ELVDKLAALALE (SEQ ID NO: 12) Sbjct 310 ELVDKLAALALE 320 (SEQ ID NO: 11, Amino Acids 310-320)
3	PSIBP4_388 (42aa)	SEYWMDFEG EMKPGKKGIS LNPEQWSQLK EQISDIDDAV RKL (SEQ ID NO: 13)	activated RNA polymerase II transcription coactivator 4 variant [Homo sapiens] Sequence ID: g222429221.1 (SEQ ID NO: 14)	Id=41/41 (100%) Gaps=0/41 (0%) Length=134 Query 2 EYWMDFEGEMKPGKKGISLNPEQWSQLKEQISDIDDAVRKL 42 (SEQ ID NO: 13, Amino Acids 2-42) EYWMDFEGEMKPGKKGISLNPEQWSQLKEQISDIDDAVRKL (SEQ ID NO: 15) Sbjct 94 EYWMDFEGEMKPGKKGISLNPEQWSQLKEQISDIDDAVRKL 134 (SEQ ID NO: 14, Amino Acids, 94-134)

15/24

FIG. 8 cont.

4	PS1BP4_586 (19aa)	SENTKISRVK LAGGGVCL (SEQ ID NO: 16)	RNA, methyltransferase [Clostridium botulinum] Sequence ID: <u>W019P_0110612.1</u> (SEQ ID NO: 17) wingless-type 3,0,4,11V integration site family, member 78 [Homo sapiens] Sequence ID: <u>AB024711.1</u> (SEQ ID NO: 18)	Id=9:10 (90%) Gaps=0:10 (0%) Length=161 Query 1 SENTKISRVK 10 (SEQ ID NO: 16, Amino Acids 1-10) SENTKIS VK (SEQ ID NO: 20) Subject 71 SENTKISVK 80 (SEQ ID NO: 17, Amino Acids 71-80)
				Id=9:10 (90%) Gaps=0:10 (0%) Length=177 Query 10 KLAGGGVCL 19 (SEQ ID NO: 16, Amino Acids 10-19) KLAGGG CL (SEQ ID NO: 21) Subject 168 KLAGGGVCL 177 (SEQ ID NO: 18, Amino Acids 168-177)
				Id=17:19 (88%) Gaps=0:19 (0%) Length=67020 Query 1 SENTKISRVKLAGGGVCL 19 (SEQ ID NO: 16, Amino Acids 1-19) --NTKISRVKLAGGGVCL (SEQ ID NO: 22) Subject 6818 TRNTKISRVKLAGGGVCL 66239 (SEQ ID NO: 19, Amino Acids 6618-66239)
5	PS1BP4_566 (21aa)	GISLILVLVVI VVLAKPLDPL (SEQ ID NO: 23)	tumor necrosis factor receptor superfamily member 21 precursor [Homo sapiens] Sequence ID: <u>U019P_022267.1</u> (SEQ ID NO: 24)	Id=10:10 (100%) Gaps=0:10 Length=345 Query 4 LILVLVVI 13 (SEQ ID NO: 23, Amino Acids 4-13) LILVLVVI (SEQ ID NO: 25) Subject 48 LILVLVVI 37 (SEQ ID NO: 24, Amino Acids 48-57)
6	PS1BP4_283 (18aa)	SGAPKLRRPK PDLPKNS (SEQ ID NO: 26)	PREDICTED: monocyte differentiation antigen CD14 [Ochobotia pinnatepa] Sequence ID: <u>W019P_064566263.1</u> (SEQ ID NO: 27) Homo sapiens hypoxia inducible gene-14 Sequence ID: <u>AB038117.1</u> (SEQ ID NO: 28)	Id=10:13 (77%) Gaps=1:13 (7%) Length=420 Query 1 KL-PPKPDLPK 16 (SEQ ID NO: 26, Amino Acids 5-16) KL R P PDLPK (SEQ ID NO: 29) Subject 154 KLVRAPQPDLPK 366 (SEQ ID NO: 27, Amino Acids 354-366)
				Id=16:18 (82%) Gaps=0:18 (0%) Length=146 Query 1 SGAPKLRRKPDLPKNS 18 (SEQ ID NO: 26, Amino Acids 1-18) G PKLRRKPDLPKNS (SEQ ID NO: 30) Subject 134 QGPKLRRKPDLPKNS 187 (SEQ ID NO: 28, Amino Acids 134-187)

FIG. 8 cont.

7	FSBP3_47 (17aa)	SAK YKETRLK EKEDALTRIE LEFLQKQKRV KKPKPEFPVY YPLEITYIQS YDHGTSIEIE EQMDDWLEN KNEIQKQAP EWTEEDLSOL TRSMVKFPQG TFQRWENIAH ELGRSVYDVI TKAKLAALAE (SEQ ID NO: 31)	<i>DnaJ</i> (Hsp40) homolog, subfamily C, member 1 (Homo sapiens) Sequence ID: 911563.2 (SEQ ID NO: 32)	Id=123123 (100%) Gaps=0 (0%) Length=634 2 AKYKETRLKENEDALTRIELELOKOKVKKPKPEFPVYITLETITYQSYDHGTSIEIE query 63 AKYKETRLKENEDALTRIELELOKOKVKKPKPEFPVYITLETITYQSYDHGTSIEIE AKYKETRLKENEDALTRIELELOKOKVKKPKPEFPVYITLETITYQSYDHGTSIEIE objct 320 62 EQMDDWLENKNEIQKQAPFWTEEDLSOLTRSMVKFPQCTPGRWKIAHELGRSVYDVI query 124 EQMDDWLENKNEIQKQAPFWTEEDLSOLTRSMVKFPQCTPGRWKIAHELGRSVYDVI EQMDDWLENKNEIQKQAPFWTEEDLSOLTRSMVKFPQCTPGRWKIAHELGRSVYDVI objct 340
8	P19BTP4_885 (25aa)	AKFPFPFPERR SGCTALAPSF RLVLS (SEQ ID NO: 34)	PREDICTED: amyloid hem A4 precursor protein-binding family B member 1-interacting protein [<i>Malopusinus undulatus</i>] Sequence ID: 166372_905152.1.1 (SEQ ID NO: 35) Human mRNA for translationally controlled tumor protein Sequence ID: emb316664.1 (SEQ ID NO: 36)	Id=818 (100%) Gaps=0 (0%) Length=653 Query 4 PFPFERS 11 (SEQ ID NO: 34, Amino Acids 4-11) Objct 531 PFPFERS 118 (SEQ ID NO: 35, Amino Acids 511-518)
9	FSBP4_577 (17aa)	SGK YIIQLR RLNHPA (SEQ ID NO: 39)	Growth growth factor-binding protein 2 precursor (Homo sapiens) Sequence ID: 114155.1 (SEQ ID NO: 40)	Id=919 (100%) Gaps=0 (0%) Length=223 Query 6 LQELRLHH 14 (SEQ ID NO: 39, Amino Acids 6-14) LQELRLHH (SEQ ID NO: 41) Objct 96 LQELRLHH 104 (SEQ ID NO: 40, Amino Acids 96-104)

FIG. 8 cont.

10	P197BP4_755 (103aa)	NEANZESFIL VLRVCYNFLF LWSEEGSCLIF ERKEFNKEFY CQDIRAYDIL FQDTPRPAQA ELYELDSFTIE KYENEGQRIN AGFREQRRLP IKTIYOKSDL Q3 (SEQ ID NO: 42)	SH3 domain-containing YSC94- like protein 1 isoform 3 [Homo sapiens] Sequence ID: U01252.1 (SEQ ID NO: 43)	Id-66:79 (64%) Caper-3:79 (5%) Length-227 Query 24 SLEGSCLIERKETNKKFYCQDIRAYDILFQDTPRPAQAE LYELDSFTIEKYENEGQRIN 82 Sbjct 67 SLEGSCLIERKETNKKFYCQDIRAYDILFQDTPRPAQAE LYELDSFTIEKYENEGQRIN 128 Query 82 A- GPEEQRLPTKTIYOK 99 (SEQ ID NO: 42, Amino Acids 24-99) A REQ+ K + K (SEQ ID NO: 44) Sbjct 127 ARKAAREQRSSAKELPTK 143 (SEQ ID NO: 43, Amino Acids 67-145)
----	------------------------	---	---	--

FIG. 9

Rank	Clone and Peptide size	Peptide sequences of mimotopes in-frame with T7 10B gene	Description of the sequences that mimotopes mimic	Region of similarity of peptide
1	PSIRP3_174 (9aa)	IQHQHLQOI (SEQ ID NO: 45)	(SEQ ID NO: 46) Mimotopes [Mycobacterium tuberculosis] Sequence ID: <u>mlh17_00001</u> 313.1	E3=6.6 (100%) Gaps=0.6 (0%) Length=103 Query 2 QHQHLQ 7 (SEQ ID NO: 45, Amino Acids 2-7) QHQHLQ (SEQ ID NO: 47) Sbjct 36 QHQHLQ 41 (SEQ ID NO: 46, Amino Acids 56-61)
2	PSIRP4_610 (161aa)	LYDPHPNPPIEV RNYSLKPGY RWREQLVFRS KLTMTAFDR KDNARPAEVT ALGSKDHSRI LVQDSRGKVES WSVSDQPGRS AADHWYKDEQ QDSGSGC6VRP SLTERRHCRN CGLFCQKCSR FQSEIKRLKISS PVRVCNCCYN LQHERGSGPRN C (SEQ ID NO: 48)	W2F73 protein (Homo sapiens) Sequence ID: <u>255AAH1214</u> 21 (SEQ ID NO: 49)	E3=158/163 Gaps=2/163 (1%) Length=437 Query 1 LYDPHPNPPIEVNYSLKPGYRWREQLVFRSRLIMETHLAFDRKDNARPAEVTALGSKD 60 L PHPNPPIEVNYSLKPGYRWREQLVFRSRLIM HIAFDRKDNARPAEVTALGSKD Sbjct 275 LSHPTNPPIEVNYSLKPGYRWREQLVFRSRLD4--HTAFDRKDNARPAEVTALGSKD 334 Query 41 HSRLLVDSRGKVESVSDQPGRSAADHWYKDEGDSGSGSVFSLTERRHCRNCCQ 124 HSRLLVDSRGKVESVSDQPGRSAADHWYKDEGDSGSGSVFSLTERRHCRNCCQ Sbjct 335 HSRLLVDSRGKVESVSDQPGRSAADHWYKDEGDSGSGSVFSLTERRHCRNCCQ 394 Query 121 LFCQKCSRFPQSEIKRLKISSPVRVCNCCYNLQHERGSGKPRMC 163 LFCQKCSRFPQSEIKRLKISSPVRVCNCCYNLQHERGSGKPRMC GPRMC (SEQ ID NO: 50) Sbjct 393 LFCQKCSRFPQSEIKRLKISSPVRVCNCCYNLQHERGSGKPRMC 437 LFCQKCSRFPQSEIKRLKISSPVRVCNCCYNLQHERGSGKPRMC (SEQ ID NO: 49, Amino Acids 275-437)

19/24

FIG. 9 cont.

3	P51BP3_266 (27aa)	VDKSVLISLGR KKYGAVGSL QSTGGH (SEQ ID NO: 51)	membrane protein [Mycobacterium tuberculosis] Sequence ID: P51BP_02364 37113 (SEQ ID NO: 52)	Id=6-6 (100%) Gaps=0-6 (0%) Length=962 Query 3 VLSLG 10 (SEQ ID NO: 51, Amino Acids 5-10) VLISLG (SEQ ID NO: 53) Sbjct 802 VLSLG 807 (SEQ ID NO: 52, Amino Acids 802-807)
4	P51BP3_166 (177aa)	SASTTEPDEFQ DILACRLNQK EQAYDIFLNAX EQNVFNATFY SNLIKLLMSED YFTQAMEVKA FAETHIRGFTL NDAAAGRLIT QVRDYLKEA VTLLKTVLDQ QYPSRLAVTR VIQALAMKCD VENIEVVQ KMLMGLDSDG LSKMFVFNIA LAQIKNNNDA AKLAAALE (SEQ ID NO: 54)	lecithase c1c4 PF8 motif containing protein [Homo sapiens] Sequence ID: gi123456789 31 (SEQ ID NO: 55)	Id=170-170 (100%) Gaps=0-170 (0%) Length=1273 Query 1 SASTTEPDEFQDILACRLNQKGAVIDFLNAKEQNVFNATFYNLKLMSDVFYQA 60 SASTTEPDEFQDILACRLNQKGAVIDFLNAKEQNVFNATFYNLKLMSDVFYQA 963 Sbjct 906 SASTTEPDEFQDILACRLNQKGAVIDFLNAKEQNVFNATFYNLKLMSDVFYQA 963 Query 61 MEVKAFATHKGFTLNDAAANSRLHTQVERDYLYKEAVTTLKTVLDQQQTPSRLAVTRVI 129 MEVKAFATHKGFTLNDAAANSRLHTQVERDYLYKEAVTTLKTVLDQQQTPSRLAVTRVI 170 Sbjct 966 MEVKAFATHKGFTLNDAAANSRLHTQVERDYLYKEAVTTLKTVLDQQQTPSRLAVTRVI 1025 (SEQ ID NO: 54, Amino Acids 1-170) Query 121 QALAMKGVENIEVVQKMLNGLEDSDGLSKMVFNNIALAQKNNNDAA 170 QALAMKGVENIEVVQKMLNGLEDSDGLSKMVFNNIALAQKNNNDAA 560 Sbjct 1026 QALAMKGVENIEVVQKMLNGLEDSDGLSKMVFNNIALAQKNNNDAA 1073 (SEQ ID NO: 55, Amino Acids 906-1075)

20/24

FIG. 9 cont.

5	PS1BP4_704 (49aa)	D A P S P L P E T T E M V V C A L G L I V G L V G I I G T I F I K G V R K S N A A E R R G P L (SEQ ID NO: 57)	HLA-DR alpha [Homo sapiens] Sequence ID: g55A022827 UAF88132.1 (SEQ ID NO: 58)	Id=4849 (92%) Cqpr=0.49 (0%) Length=59 Query 1 D A P S L P E T T E N V V C A L G L I V G L V G I I G T I F I K O V E K S N A A E R R G P L 49 Amino Acids 1-49) D A P S L P E T T E N V V C A L G L I V G L V G I I G T I F I K O V E K S N A A E R R G P L (SEQ ID NO: 59) Sbjct 2 D A P S L P E T T E N V V C A L G L I V G L V G I I G T I F I K O V E K S N A A E R R G P L 50 (SEQ ID NO: 58, Amino Acids 2-50)
6	PS17BP4_76 3 (17aa)	D L S S E V A T H Q P I I A C L P (SEQ ID NO: 60)	transketolase [Mycobacterium tuberculosis] Sequence ID: r088P_0167 21253.1 (SEQ ID NO: 61)	Id=53 (100%) Cqpr=0.3 (0%) Length=51 Query 3 T H Q P I 12 (SEQ ID NO: 60, Amino Acids 8-12) T H Q P I (SEQ ID NO: 62) Sbjct 433 T H Q P I 437 (SEQ ID NO: 61, Amino Acids 453-457)
7	PS1BP4_563 (14aa)	L Q A C F P Q I L R G S L (SEQ ID NO: 63)	aldehyde-acid dehydrogenase [Mycobacterium tuberculosis] Sequence ID: r088P_02264 14312.1 (SEQ ID NO: 64)	Id=67 (88%) Cqpr=0.7 (0%) Length=575 Query 8 I L K G S L A 14 (SEQ ID NO: 63, Amino Acids 8-14) I L K G S L A (SEQ ID NO: 65) Sbjct 400 I L K G S L A 406 (SEQ ID NO: 64, Amino Acids 400-406)

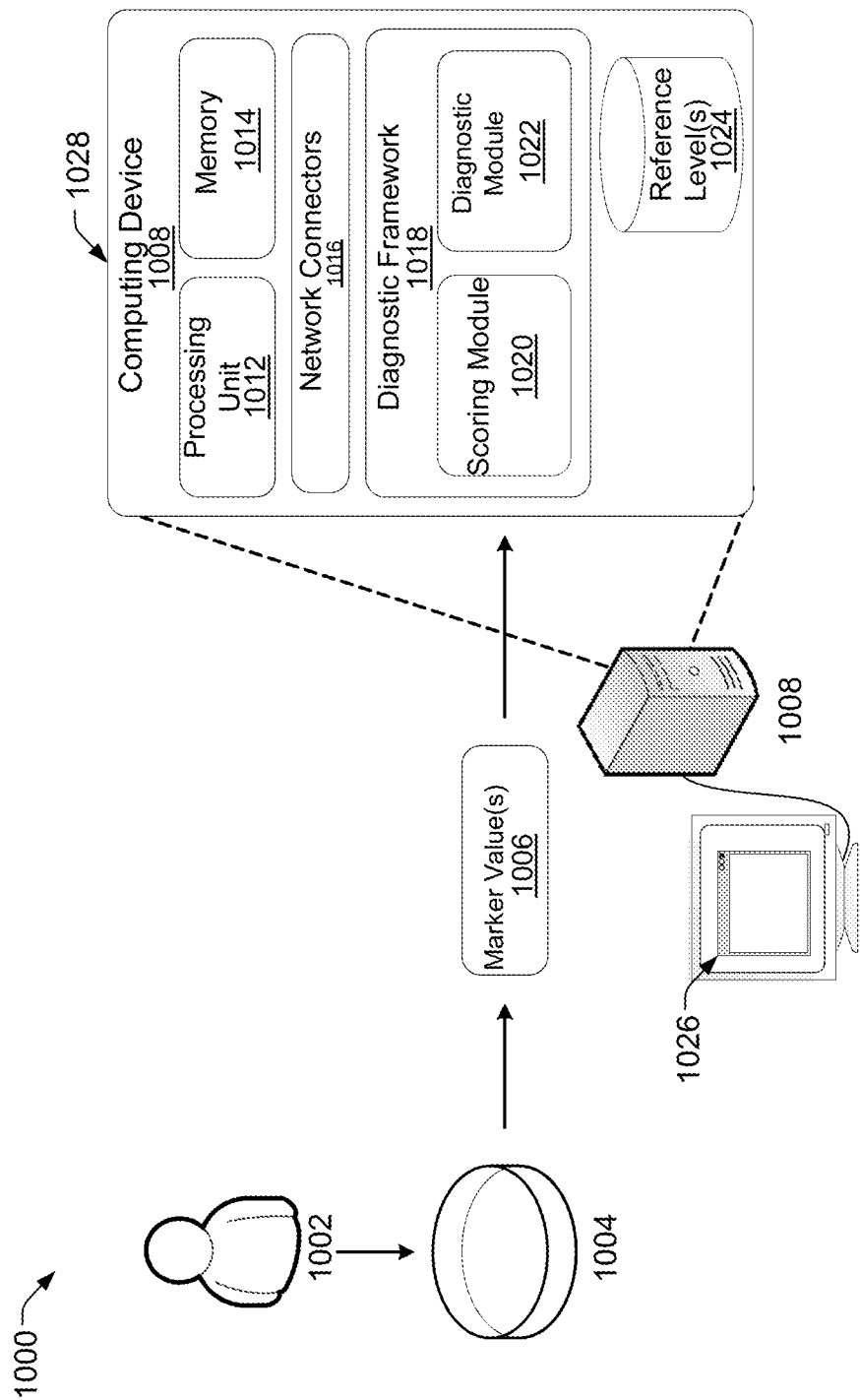
FIG. 9 cont.

\$	PSIBP3_113 (16aa)	AGISRELVDNL AAALE (SEQ ID NO: 66)	Class A, Mycobacterium Tuberculosis Bacterioferritin, Hfr Sequence ID: <u>pfba0222a</u> (SEQ ID NO: 67)	Id=77 (100%) Gaps=0/7 (0%) Length=174 Query 10 KLAALLE 16 (SEQ ID NO: 66, Amino Acids 10-16) KLAALLE (SEQ ID NO: 68) Subject 162 KLAALLE 16% (SEQ ID NO: 67, Amino Acids 162-168)
9	PSIBP3_200 (209aa)	SSSTPLSNQPL NQDNDYFQQQ FDQISNR TQKQ EAQAGP WPFSSSQTPA VRTQNGVSR EQNGFSVKSSP NPFVGSPPKGL SIQNGVKQDLE SSVQSSPHDSI AIIFPQSTKFG RGR TAKSSA NDL L ASDIFAP FVSEFSQAAP TQPTA IQNP LDELKTSAPAP VGFVGLGGV TVTL PQAGRW XTASLVFNQSP SWXLAALLE (SEQ ID NO: 69)	disabled homolog 2 isoform 2 (from sapient) Sequence ID: <u>MF07_001211</u> <u>800.11</u> (SEQ ID NO: 70)	Id=201/208 (97%) Gaps=0/208 (0%) Length=749 Query 1 SSSTPLSNQPLNGDIDYFQQQFDQISNR TCKQEAQAGPWFSSSQTPA VRTQNGVSR 60 SSSTPLSNQPLNGDIDYFQQQFDQISNR TCKQEAQAGPWFSSSQTPA VRTQNGVSR 60 SSSTPLSNQPLNGDIDYFQQQFDQISNR TCKQEAQAGPWFSSSQTPA VRTQNGVSR 364 Subject 363 Query 61 QNGFSVKSSPWFVGSPPKGLSIQNGVKQDLESSVQSSPHDSIAIIFPQSTKFGGRLE 120 QNGFSVKSSPWFVGSPPKGLSIQNGVKQDLESSVQSSPHDSIAIIFPQSTKFGGRLE 120 Subject 365 QNGFSVKSSPWFVGSPPKGLSIQNGVKQDLESSVQSSPHDSIAIIFPQSTKFGGRLE 424 Query 121 AKSSANDLLASDIFAPFVSEFSQAAPSTQPTALQNPFLDLFKTSAPAPVGLVGLGGVT 180 AKSSANDLLASDIFAPFVSEFSQAAPSTQPTALQNPFLDLFKTSAPAPVGLVGLGGVT 180 Subject 423 AKSSANDLLASDIFAPFVSEFSQAAPSTQPTALQNPFLDLFKTSAPAPVGLVGLGGVT 484 Query 181 VTLPQAGPWXTASLVFNQSPSWXLAAL 208 (SEQ ID NO: 69, Amino Acids 1-208) VTLPQAGPWXTASLVFNQSPS A + (SEQ ID NO: 71) Subject 483 VTLPQAGPWXTASLVFNQSPSWXLAAL 512 (SEQ ID NO: 70, Amino Acids 305-512)

FIG. 9 cont.

10	P51BP4_622 (122aa)	AAAMDVFLM IRRHRTFTFD AKESSTVFEIK RIVEQILKEPP DEQRLVKDDQ LDDCKTLCE CGFTSQIARQ APATVGLAEK ADDTPEALCIE PFSSPFELPDV MEPQDSGSSA NEQAVQ (SEQ ID NO: 72)	transcription elongation factor B polypeptide 2 isoform a [Homo sapiens] Sequence ID: <u>U08392.1</u> (SEQ ID NO: 73)	Id=118/118 (100%) Query=Q118 (2%) Length=118 Query: 5 MDVFLMRHRTFTFDAKESSTVFEIKRIVEQILKEPPDEQRLVKDDQLDDCKTLGEC 84 Subject 1: MDVFLMRHRTFTFDAKESSTVFEIKRIVEQILKEPPDEQRLVKDDQLDDCKTLGEC 60 Query: 61 GFTSQIARQAPATVGLAFRADDITFEALCIETSSPFELPDVMEKPDGSSANEQAVQ 122 Subject 61: GFTSQIARQAPATVGLAFRADDITFEALCIETSSPFELPDVMEKPDGSSANEQAVQ 118 (SEQ ID NO: 72, Amino Acids 5-122) (SEQ ID NO: 74) (SEQ ID NO: 73, Amino Acids 1-118)
----	----------------------------------	---	--	--

FIG. 10



24/24

FIG. 11

