Science Translational Medicine
AAAS

Supplementary Materials for

Electronic health record signatures identify undiagnosed patients with common variable immunodeficiency disease

Ruth Johnson *et al.*

Corresponding author: Manish J. Butte, mbutte@mednet.ucla.edu

Sci. Transl. Med. **16**, eade4510 (2024)
DOI: 10.1126/scitranslmed.ade4510

The PDF file includes:

Figs. S1 to S10
Tables S1 to S3

Other Supplementary Material for this manuscript includes the following:

MDAR Reproducibility Checklist

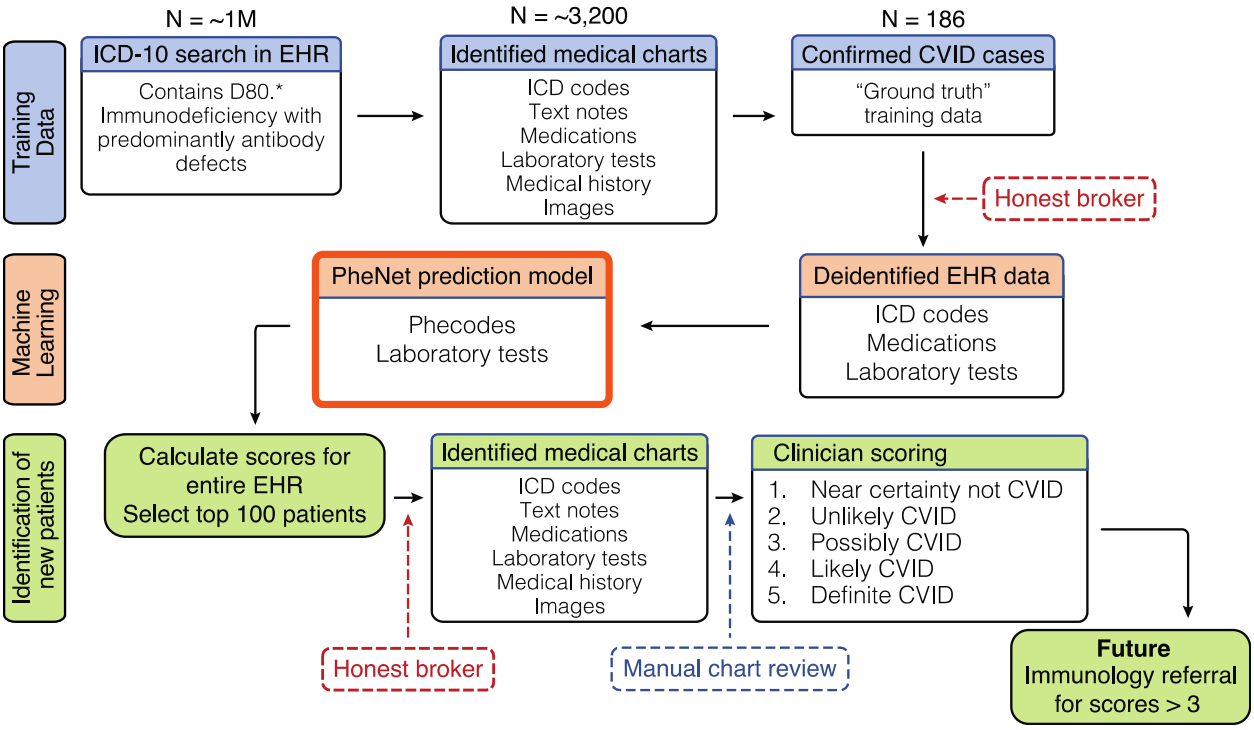


Fig. S1. Overview of curation process for the "ground truth" CVID cohort and process of identifying new putative patients with CVID. We provide a flowchart describing the EHR review process for constructing a well-curated list of clinically diagnosed patients with CVID. We then demonstrate how this cohort is used for training a prediction model which is then used to identify undiagnosed CVID patients in a discovery cohort. A manual chart review is performed on the patients with the highest risk score with the future goal of highly probable patients with CVID being referred to an immunologist.

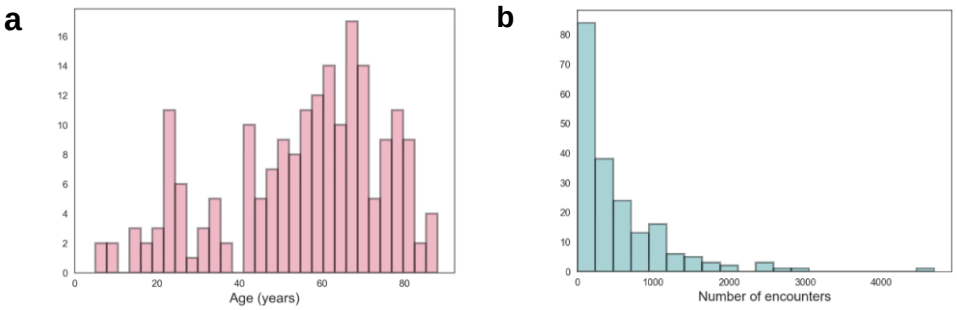


Fig. S2: Overview of UCLA "ground truth" CVID cohort. a) Distribution of ages in the CVID case cohort. We show the age of patients from their most recent encounter (up to a cutoff date in 2019). **b**) shows the distribution of the number of encounters recorded in the EHR within the case cohort. Both these histograms count the number of individuals in the case cohort on the vertical axis.

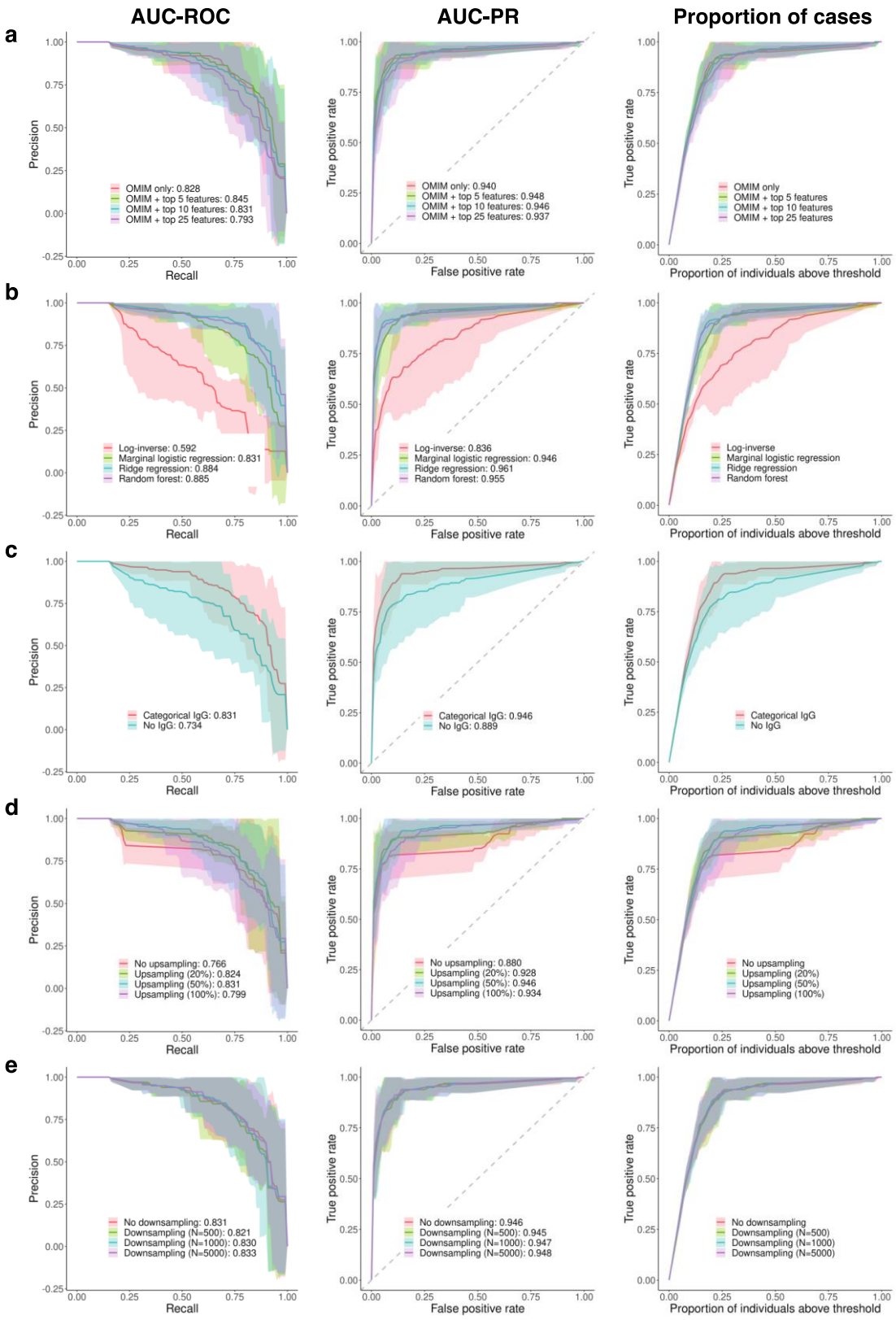


Fig. S3: Exploration of model parameters for PheNet. We show the Area Under the Curve for the Receiver Operating Curve (AUC-ROC) and Area Under the Curve for the Precision Recall curve (AUC-PR) for the PheNet model using matched case (n=197) and control (n=1,106) cohorts with 5-fold cross-validation. We also show the proportion of CVID cases captured within the varying percentiles of PheNet scores. We varied the **a)** number of additional phecode features in addition to OMIM-selected features, **b)** prediction model, **c)** inclusion of immunoglobulin G (IgG) tests, **d)** upsampling, and **e)** downsampling.

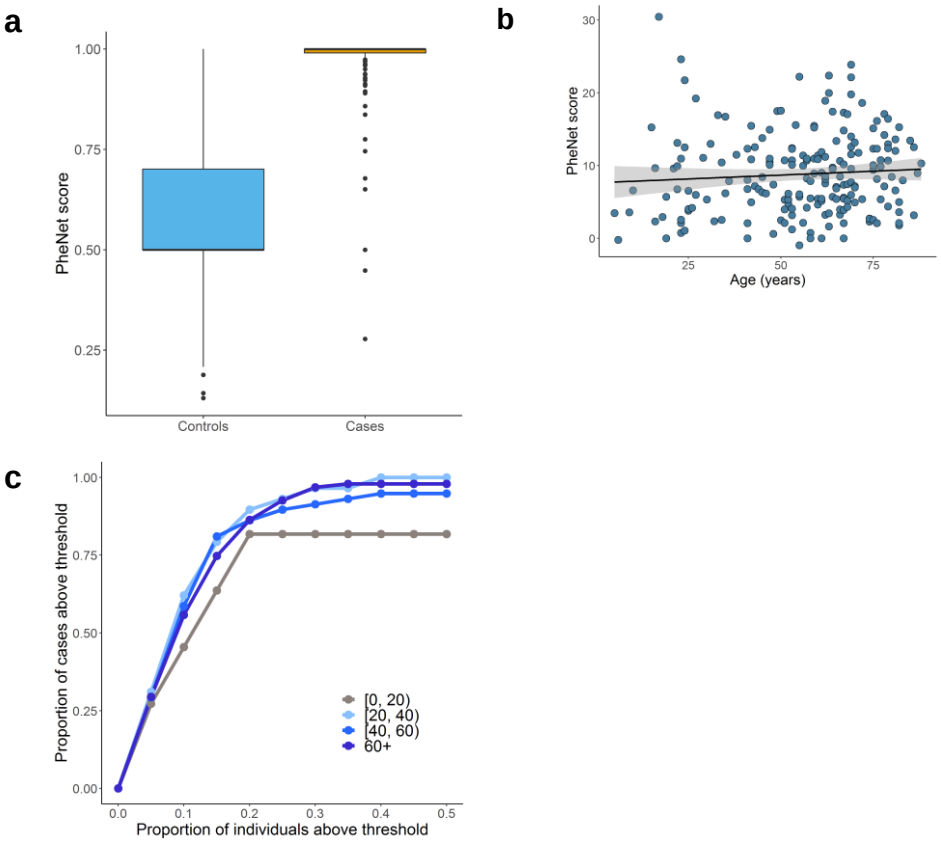


Fig. S4: PheNet scores for cases and controls. a) Normalized distribution of PheNet scores within the case-cohort and the control-cohort trained using 5-fold cross-validation. Using a Cochran-Armitage test, we found that the scores in the case-cohort are significantly higher than those in the control-cohort ($p < 2.2 \times 10^{-16}$). **b**) PheNet scores did not vary by age across our case cohort. **c**) PheNet performance is better for older subjects. ROC curves showing PheNet scores calculated on ages 0-20; 20-40; 40-60; and 60+.

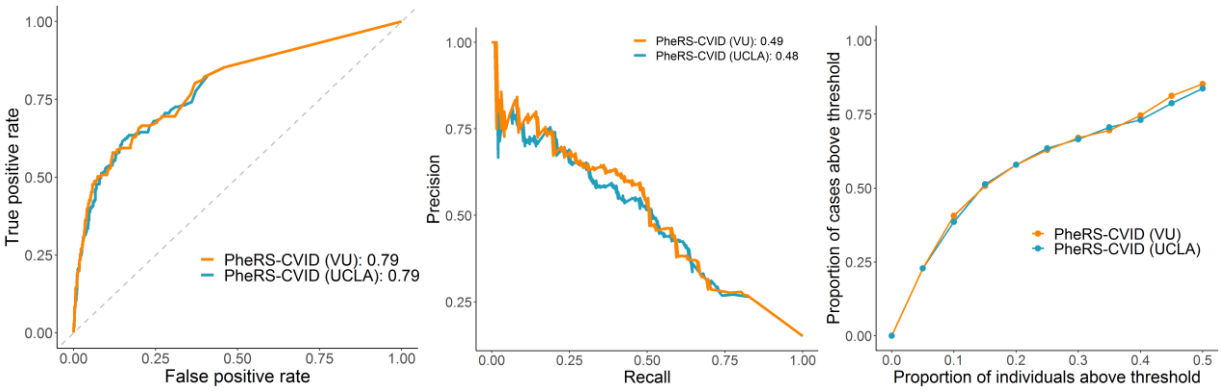


Fig. S5: Comparing PheRS performance using models trained at UCLA and Vanderbilt. We show AUC-ROC, AUC-PR, and calibration curves for the PheRS models trained at UCLA and Vanderbilt (VU). Models were trained and tested using our matched case and control cohorts. Because the model is unsupervised, no test-train split was needed.

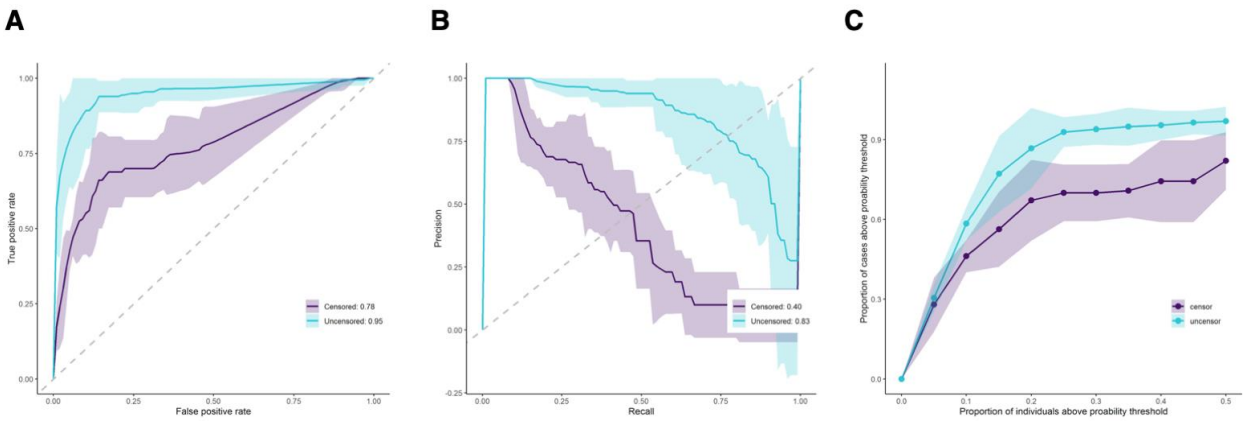


Fig. S6: PheNet captures CVID patients with censored data. Risk scores for CVID were calculated for each patient using PheNet with 5-fold cross-validation across case (n=197) and control (n=1,106) cohorts using both "censored" data that only included information before diagnosis and "uncensored" data that used the entire medical record regardless of diagnosis date. **A)** AUC-ROC and **B)** AUC-PR curves for each model. **C)** report the percentage of CVID cases captured at varying percentile cutoffs.

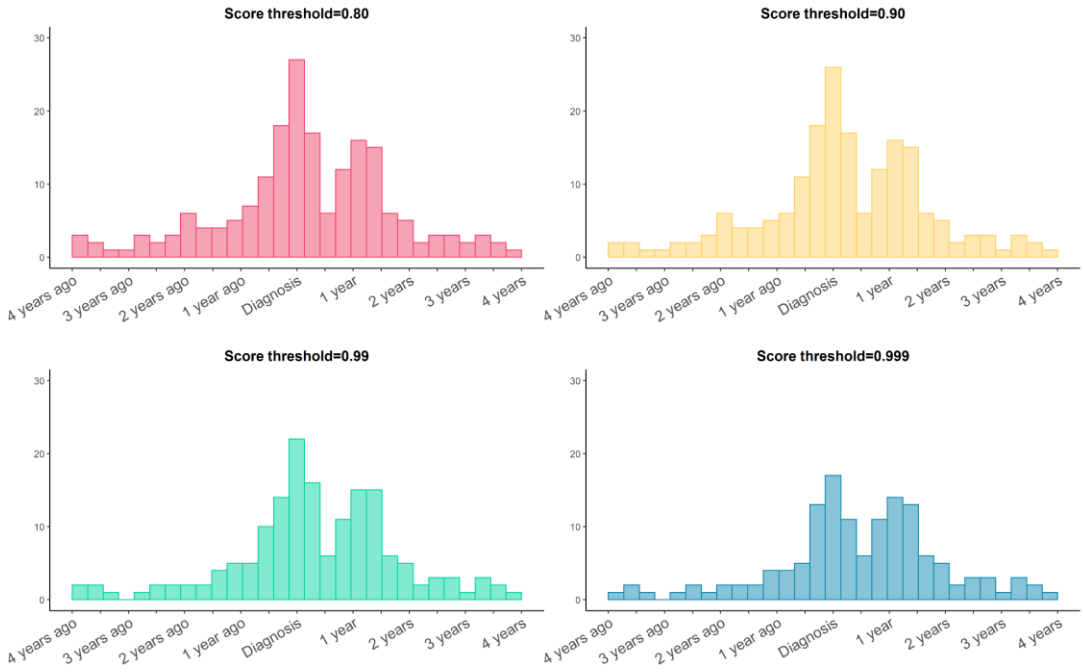


Fig. S7: Different score thresholds capture various numbers of patients with CVID before diagnosis. Training on the full UCLA Health population (n=~880K), we estimated PheNet scores for patients with CVID using only information prior to their ICD-based diagnosis, and 5-fold cross-validation. We show the distribution of times when patients pass the scoring threshold at 0.80, 0.90, 0.99, and 0.999. We only show those patients with CVID with at least one year of recorded UCLA EHR data prior to their diagnosis (n=58).

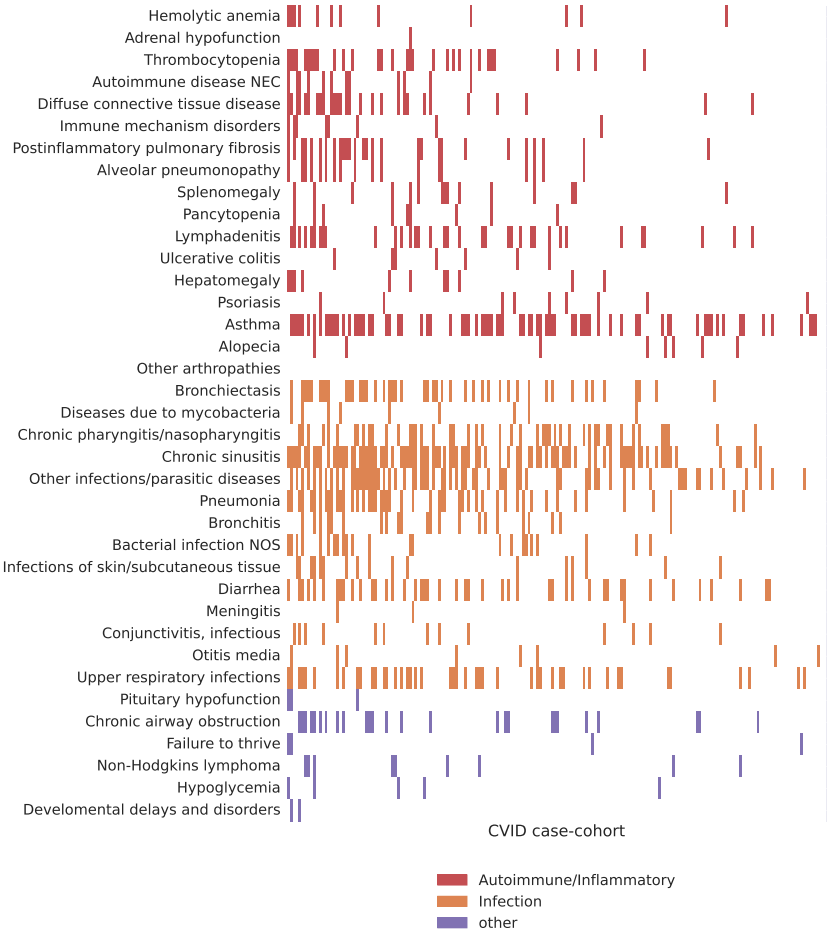


Fig. S8: EHR signatures of individuals in the CVID case cohort. Each row shows a clinical feature from the PheNet model and each column is a patient's EHR profile. Individuals in the "ground truth" CVID cohort of UCLA (n=186) are shown where the lowest to highest PheNet scores are displayed right to left. Boxes are colored according to phenotype category (autoimmune / inflammation, infection, other).

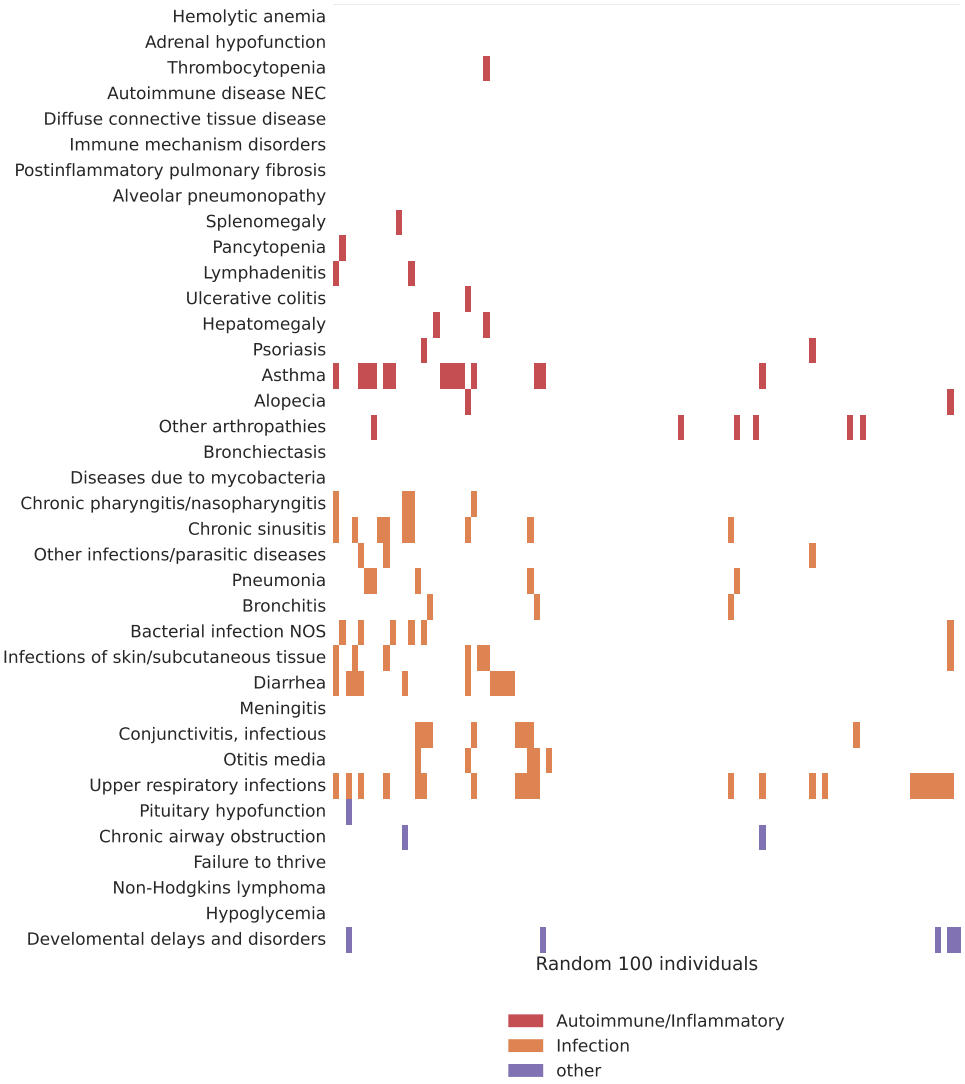
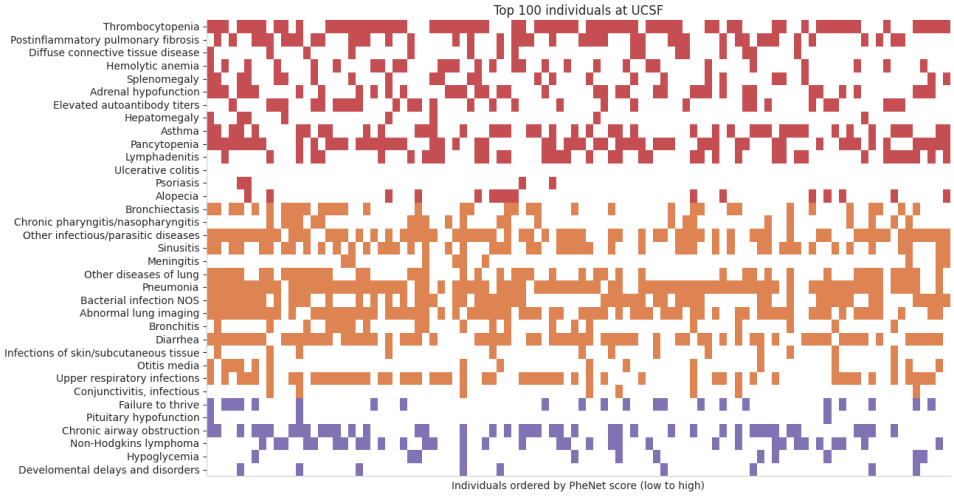
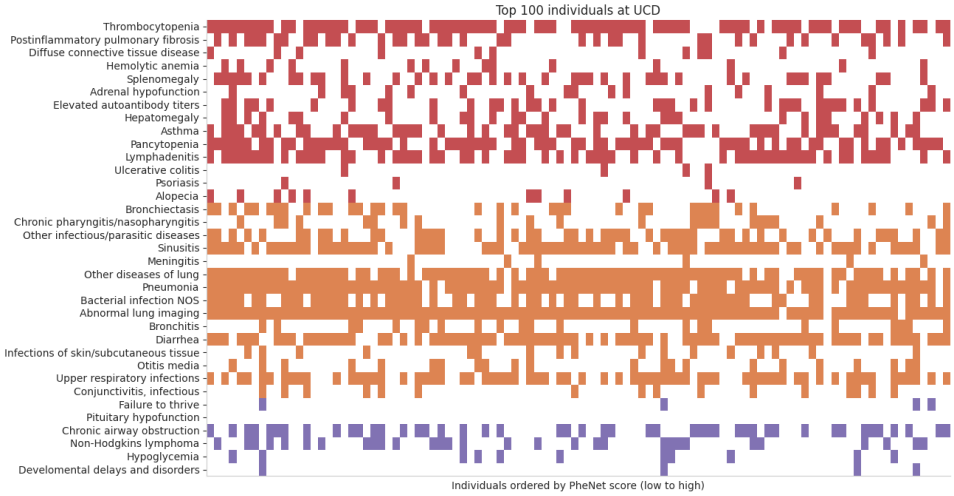


Fig. S9: EHR signatures of individuals in a randomly selected sample. Each row shows a clinical feature from the PheNet model and each column is a patient's EHR profile. The 100 individuals randomly sampled from the EHR are shown where the highest risk score is on the left. Boxes are colored according to phenotype category (autoimmune / inflammation, infection, other).

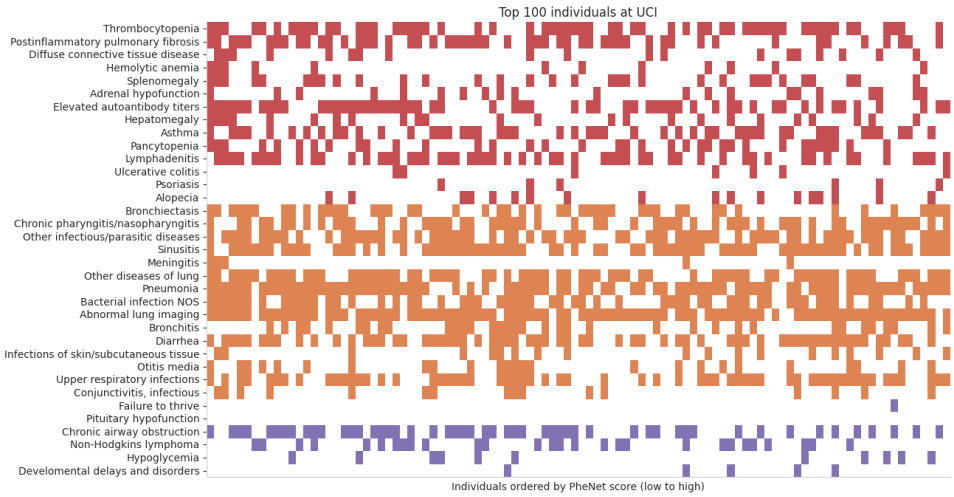
a) UC San Francisco



b) UC Davis



c) UC Irvine



d) UC San Diego

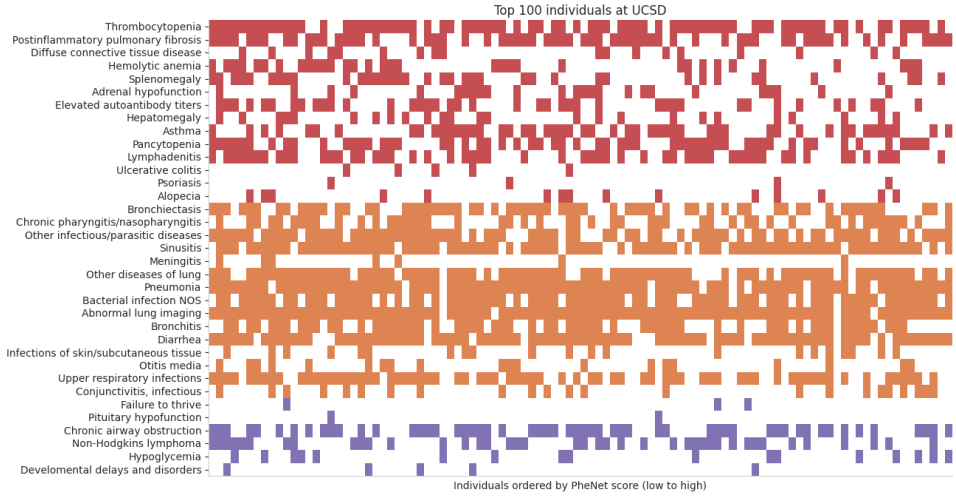


Fig. S10. EHR signatures of the top-100-ranked individuals drawn from four independent hospitals in the University of California system. Each row shows a clinical feature from the PheNet model, trained on the UCLA database, and then later applied to independent hospitals of the University of California system. Each column is a subject's EHR profile, where the lowest to highest risk scores are displayed left to right. Boxes are colored according to phenotype category (autoimmune, infection, other). **A)** University of California San Francisco; **B)** University of California Davis; **C)** University of California Irvine; **D)** University of California San Diego.

Table S1. The UCLA case cohort

Age	IgG prior to therapy	Clinical
8	391	Panhypogammaglobulinemia and poor strep pneumo response, 6% switched memory, 463 B cells/ uL
11	450	Kabuki Syndrome KMT2D, ALPS-like phenotype with Evans syndrome; no recurrent infections. 0/23 pneumococcal serotypes
14	228	Mild eczema, recurrent ear infections; Older sister with hypogam and recurrent sinopulmonary infections; Poor HiB and tetanus responses
16	391	Pneumonias, hypogam as toddler, sinus infections, poor pneumococcal responses 10/23
18	48	Panhypogammaglobulinemia; few sinopulmonary infections, hospitalized for gastritis, failure to thrive; low switched memory 0.2% of PBL
18	390	LRBA deficiency. enteropathy, hypogammaglobulinemia, granulomatous lung disease and Evans syndrome; autoimmune hepatitis;
18	586	TACI R202H, recurrent sinus infections and pneumonia; very low switched memory B cells (0.4% of PBL); brother and father affected
19	407	Early childhood with many infections and eczema; diagnosed age 7; Recurrent otitis media, sinusitis and bronchitis. Multiple courses of antibiotics yearly. Tonsillectomy and adenoidectomy and myringotomy tubes x 2. low switched memory B cells; 7/23 pneumococcal serotypes
23	514	Recurrent pneumonias as child, erythrocytosis, 10/23 pneumococcal serotypes
24	221	Recurrent ITP, Evans syndrome, mastoiditis, low memory B cells, absent tetanus T cell response
24	320	Warts, autoimmune encephalitis and stroke; utterly absent tetanus, HiB and pneumococcal titers; mother affected
24	503	Sinus infections; GLILD; Very low switched memory 0.2%; high CD21 low B cells 21%; low HiB, 0/23 pneumococcal serotypes, low tetanus antibodies
24	451	22q11DS plus TACI C104R heterozygous; severe autoimmunity; chronic urticaria; Evans syndrome; B cell deficiency (20-50 / µL); 77% CD21 low B cells
25	270	Hypogam and poor vaccine responses, recurrent sinopulmonary infections
26	213	ITP (+steroids), severe hypogam, no response to tetanus and pneumococcal (0/14 pneumococcal serotypes), alopecia
27	290	Pancytopenia, Evans syndrome, 7/23 pneumococcal serotypes, low switched memory (2%)
27	320	Frequent sinus infections; tonsillectomy; low tetanus titer; 2/23 pneumococcal serotypes after pneumovax, 12/23 after Prevnar
27	184	Bronchiectasis, pneumonias, sinus infections, panhypogammaglobulinemia
57	413	Hypogam, arthritis, 10/23 pneumococcal serotypes, sepsis
28	341	Chronic cough and sinusitis in college, poor pneumococcal (0/23 serotypes), low switched memory
28	467	Recurrent sinus infections, lymphoproliferation, Father with CVID; “poor pneumococcal response”
28	590	Severe AIHA and Evans syndrome since infancy, splenectomy, +lymphoproliferation, 0/23 pneumococcal serotypes
29	214	Frequent sinus infections; cough; croup; panhypogammaglobulinemia; 0/23 pneumococcal serotypes
31	49	CTLA4 deficiency; neutropenia, AIHA, CNS vasculitis; sepsis
31	164	Sinus infections, diarrhea, otitis, no thrush, short stature, 0/23 pneumococcal serotypes
33	100	Pneumonias; bronchiectasis; panhypogammaglobulinemia, low B cells (38 / µL), intermittent CD4 cytopenia,
33	380	STAT1 gof, infections, cellulitis, facial candida, thrush, non-TB mycobacterial infection, onychomycosis, multiple pneumonias; deceased of myocarditis
33	560	Pneumonias starting at 15, GLILD; lymphoproliferative phenotype, panhypogammaglobulinemia; mild T and B penia
33	565	Chronic diarrhea, fungal skin infections; 2/23 pneumococcal serotypes, low switch memory 1.3%

33	<35	ITP, recurrent pneumonias; warts, high T cells, normal number of B cells, total panhypogammaglobulinemia; no switched memory B cells, 0/23 pneumococcal serotypes
35	475	Chronic diarrhea, one pneumonia; enteritis with prominent lymphoid aggregates in colon; panhypogammaglobulinemia; 0/23 pneumococcal serotypes
35	526	Vasculitis, warts; retroperitoneal fibrosis; osteomyelitis; mild B cell-penia, recent T cell-penia, low NK cells
37	510	Otitis and sinus infections since childhood, rhinitis, viral meningitis; normal switched memory B cells; “low titers” after vaccine
38	434	XLP, lymphoma age 12; recurrent infections
39	321	Diagnosed age 12 after meningitis, 0% switched memory B cells
39	300	Age 8 multiple sinopulmonary infections, poor wound healing, diagnosed as teenager
43	384	Recurrent ear infections, bronchiectasis, asthma, history of recurrent pulmonary MAC infection requiring lobectomy), recurrent sinusitis; low 50 B cells, 3/23 pneumococcal serotypes
44	276	Lymphadenopathy, no steroids, +sinus, eczema, very low switched memory mem B cells, panhypogammaglobulinemia; 0/23 pneumococcal serotypes
44	488	Recurrent sinus infections and surgeries; Utterly absent panhypogammaglobulinemia; no B cell memory, 13% CD21 low B cells
44	496	AIHA starting age 17, ITP, chronic diarrhea, lymphadenopathy
44	22	Pneumonia started at age 21, chronic diarrhea; sigmoid cancer in 40s; absent strep titers; utterly absent panhypogammaglobulinemia
44	255	NFKB1 deficiency, ITP, recurrent otitis, sinus, chronic diarrhea; nodular regenerative hyperplasia of liver; tetanus low
46	438	ITP; few infections; treated with steroids prior to diagnosis; hypogam persisted at least one year after cessation of steroids, 4/23 pneumococcal serotypes
46	441	Age 15, recurrent pneumonias and sinus infections
46	523	SLE-scleroderma, received rituxan, wound healing, panhypogammaglobulinemia, ILD
46	626	Sinus infections, pneumonias, initially diagnosed as IgM deficiency, “poor pneumococcal response”, lymphopenia
47	410	Starting age 12 multiple otitis, few pneumonias
48	253	Pneumonias as child; GLILD; granulomas on vocal cords; 5/23 pneumococcal serotypes
48	540	IPEX-like enteritis, Evans syndrome, bronchiectasis
48	510	Frequent sinopulmonary infections; Low CD4 T cells, low switched memory 0.3% of PBL, very high CD21 low B cells (48%)
48	590	Seizures, inflammatory pseudotumor, recurrent pneumonia requiring lobectomy
49	250	Recurrent sinus infections, diagnosed in teens
50	351	Allergic rhinitis, bronchitis, otitis as child, no steroids
50	369	8-10 years of frequent sinus infections, bronchitis and UTIs; 1/23 pneumococcal serotypes
51	537	Chronic sinus infections, brother with hypogam; 3/12 pneumococcal serotypes
51	539	TACI; multiple sinus infections; chronic diarrhea, multiple pneumonias, legionella, coccidioidomycosis; low HiB titer, 1/23 pneumococcal
51	33	Age 22 recurrent pneumonia and sinus infections; utterly panhypogammaglobulinemia
52	226	Recurrent infections, GLILD; +ITP, lymphadenopathy, Arthritis, splenectomy for ITP; panhypogammaglobulinemia; high T cells; "No response to vaccines"
53	99	Age 26 recurrent sinopulmonary infections, pancytopenia, GLILD dx on biopsies, very low T cells
53	460	Age 20-30s multiple bronchitis, pneumonias, hospitalizations; inadequate vaccine responses
54	369	Recurrent sinusitis, pneumonias. 2/23 pneumococcal serotypes
54	465	Age 30s sinus infections, UTI, vomiting/diarrhea, 11/23 pneumococcal serotypes
54	494	Sinus infections, pneumonias with hospitalizations; 2/14 pneumococcal serotypes
54	494	Frequent sinus infections, asthma, ITP; no steroids; 6/23 pneumococcal serotypes
54	180	Age 26 recurrent sinus, otitis as child

55	291	Recurrent ITP, +treatment with steroids prior to diagnosis, lymphadenitis consistent sarcoidosis; 1/23 pneumococcal serotypes, low tetanus antibody response
55	33	Eczema, sinus infections for 10 years, utterly panhypogammaglobulinemia
55	428	Recurrent shingles; 0/14 pneumococcal serotypes, good tetanus response
55	472	Recurrent sinusitis and otitis; +ITP, hemolytic anemia, very low T (150) and B cells; +treatment with steroids prior to diagnosis
55	598	Recurrent sinus infections, 1/23 pneumococcal serotypes
55	156	recurrent sinus infections with bronchitis and pneumonias, Rhinitis, Hypothyroidism, Skin infections, Urticaria
56	73	Innumerable sinus and bronchial infections, chronic liver disease, ITP
56	599	CTLA4 deficiency; Pneumonias; GLILD; recurrent MRSA infections of the skin, recurrent otitis media, ITP requiring splenectomy, allergic rhinitis; 0.3% switched memory B cells; 4/12 pneumococcal serotypes; no detectable tetanus vaccine response
56	518	Recurrent lung infections; Bechet's; Raynaud's
57	375	Multiple hospitalizations for dyspnea and coughing, Low B cells 53, 4/23 pneumococcal serotypes post vaccine
58	516	Dyspnea, aspergillus lung infection, ILD; treated with steroids but weaned down to 5 mg before pneumococcal vaccine challenge 5/23 serotypes; deceased from pneumonia
58	615	Chronic sinus infections, 7/15 pneumococcal serotypes post vaccine
59	287	Lifelong history of recurrent infections, mostly sinopulmonary in nature. 2/23 pneumococcal after vaccine
59	300	sinusitis and pneumonia about 10 times per year requiring systemic antibiotics; frequent UTI's, no autoimmunity, 64 B cell / µL, no steroids, normal switched memory B cells; 18% CD21 low B cells
59	484	Sinus infections since childhood; son affected; low B cells (64 / µL) with low switched memory (0.5% of PBL)
59	542	TREX1 deficiency AD, lung granulomas biopsy confirmed, 11/23 pneumococcal serotypes post vaccine
60	90	bronchitis, sinusitis, pneumonia during childhood; dental infection; severe panhypogammaglobulinemia
60	200	Recurrent bacterial respiratory infections including sinusitis, bronchitis and pneumonia, Psoriatic arthritis; severe panhypogammaglobulinemia; 0 B cells
60	500	Chronic sinusitis, chronic otitis media arthritis, asthma, Chronic EBV, lymphopenia
60	457	Recurrent infections, soft tissue infections, dental abscess, 5/23 pneumococcal serotypes post vaccine, tetanus protective
60	560	Otitis in childhood with tubes; Recurrent sinus infections (pansinusitis), chronic rhinitis, two sinus surgeries; recurrent bronchitis; autoimmunity (SLE, Hashimoto's), no steroids, history of MGUS, fibromyalgia, H pylori infection; 5-6/14 pneumococcal serotypes
60	570	Recurrent sinus infections and pneumonias, bronchiectasis, esophageal candidiasis
61	20	Recurrent sinusitis, pneumonia, meningitis, chronic diarrhea requiring hospitalization since teens; utterly absent panhypogammglobulinemia
61	596	Recurrent otitis as child, recurrent sinus infections in 40s; interstitial cystis, GERD, MS (not on steroids); 6/23 pneumococcal serotypes
62	381	4-6 sinus infections per year in early 20s, Viral URIs, multiple episodes of bronchitis, IBS, Chronic rhinitis; 8/23 pneumococcal serotypes
62	514	Recurrent tonsilitis as a child, Chronic sinusitis in 20s; Very low T cells in 60s (230 / µL); very low switched memory (0.4% of PBL)
62	560	Recurrent respiratory infections in 30s, bronchiectasis, diarrhea, asthma, frequent gastrointestinal infections; 4/23 serotypes after vaccination; 0% switched memory B cells
62	588	Frequent sinopulmonary infections in 20s, "poor vaccine titers" after vaccination
63	275	Developed sinus and fungal infections in 20s, frequent bronchitis, lymphadenopathy, chronic diarrhea, osteomyelitis, IBS, 0/23 pneumococcal serotypes 0/23 protective post vaccination

63	493	Bronchiectasis, recurrent respiratory infections, Chronic sinusitis, ITP, MALT Lymphoma, CVID diagnosed 1995, Asthma, Arthritis
63	389	Recurrent upper respiratory tract infections, intestinal bacterial overgrowth, Urticaria, mildly low T cells and B cells
64	134	Multiple pneumonias in 40s, Bronchiectasis, Recurrent URIs, Chronic cough
64	503	RTEL1-deficiency; Lung infections and pulmonary fibrosis; recurrent ITP since childhood
64	580	Frequent infections in childhood including pneumonias and viral infections; recurrent sinus infections s/p sinus surgery, oligoarthritis, Asthma, Psoriasis; 9/23 pneumococcal serotypes; mildly low B cells (82 / µL)
65	254	Sinusitis, Asthma, RA, ankylosing spondylitis, recurrent C difficile and chronic diarrhea; 1/14 pneumococcal serotypes after vaccine, T cell count mildly low (514). Total B cell numbers slightly low (91) normal switched memory and no increase in immature B cells
65	330	Pneumonias and sinusitis during early 20s, Chronic diarrhea, Granulomatous disease, 50% pneumococcal serotypes after vaccination, infection-associated amputations, COPD
65	449	Chronic sinusitis, Recurrent otitis media with Aspergillus niger, Chronic pain syndrome, chronic fevers dental infections, Deceased of sepsis
65	530	Bronchitis and sinus infections, chronic C diff, chronic cough, IBS, interstitial cystitis, UTIs
65	304	Bronchiectasis, arthritis, recurrent sinusitis, Chronic diarrhea, “poor vaccine antibodies and despite booster”
66	293	As a child, many gastroenteritis infections and respiratory infections; ulcerative colitis in 40s; panhypogammaglobulinemia; Pre- and post-titers of diphtheria were undetectable and HiB whereas tetanus did show boosting (>4 fold)
66	332	Lymphadenopathy without clonality, Frequent infections in childhood, low CD4 and CD8 T cells prior to cancer
66	506	Dermatitis, recurrent gastrointestinal problems since age 20 y, Multiple pneumonia hospitalizations, Asthma, Arthritis, recurrent sinus infections, Frequent strep pharyngitis in teens, ILD
66	613	Starting at age 10 y, recurrent sinus infections per year, multiple bronchitis, arthritis. Low B (23 / µL) and T cells (252 / µL), ITP, dental infections; very low switched memory (0.02% of PBL), very high CD21 low B cells (56%)
67	516	Bronchiectasis, pneumonias, 10/23 pneumococcal titers, tetanus protective
67	527	No sinopulmonary infections as child but in 40s developed recurrent sinusitis (proteus, pseudomonas, staph) and otitis; autoimmune thyroiditis, AIHA, Sjogren’s; very low B cells (13 / µL); 1/23 pneumococcal titers; deceased of ovarian cancer
67	621	Chronic diarrhea, no respiratory infections, very low classed switched B cells (0.4% of PBL), low T cells, 14/23 pneumococcal serotypes; losses of Ig ruled out
68	306	Pneumonia, SLE/MCTD (not on steroids, no rituximab), absent switched memory B cells; 0/23 pneumococcal serotypes after vaccination, tetanus protective
68	500	TACI S144X and C193R; Recurrent sinus infections and bronchitis, bronchiectasis, chronic diarrhea, osteomyelitis x 2
68	513	Recurrent infections, mycobacterium pneumonia 2-4 times, multiple osteomyelitis, cellulitis, recurrent staph infections, 4/23 pneumococcal serotypes, HiB non-protective
68	526	Recurrent sinusitis and bronchitis; cellulitis, diverticulitis, GERD
68	523	Recurrent sinus infections, bronchitis, sinus surgery
68	586	Recurrent sinus infections in the 50s including four sinus surgeries, GI symptoms including chronic diarrhea, allergic rhinitis, 9/23 pneumococcal serotypes
68	190	Pneumonias in 40s and 50s, few sinus infections, psoriasis, asthma, allergic rhinitis, osteoporosis, GERD, rheumatoid arthritis, +steroids prior to diagnosis but IgG 190 was off steroids for months
69	320	Recurrent sinus infections with sinus surgery, Osteoarthritis, Addison’s disease, Stroke, SLE
69	400	Recurrent lung infections, Bronchiectasis, Chronic cough, Pancreatitis, chronic sinusitis, chronic UTI, SLE, Asthma, Antiphospholipid syndrome, Colitis
69	423	Chronic respiratory infections, bronchiectasis, arthritis, lupus, fibromyalgia, Recurrent sinus infections with sinus surgeries, “Low pneumococcal titers”

69	452	Recurrent bronchitis and sinusitis, COVID 2x, Shingles 3x, Vasculitis, IBS, Eczema, osteoarthritis
69	581	Asthma, Recurrent sinus infections, eye infections, chronic cough, Gastroenteritis
69	607	sinus infections, recurrent respiratory infections x 30 years, recurrent thrush, Cellulitis, chronic UTI, 4/23 pneumococcal serotypes
69	619	Chronic Sinusitis x 30 y, Asthma, Fibromyalgia, Hashimotos, Hypertension
69	364	Chronic sinus infections with multiple surgeries, Multiple episodes of sepsis due to multifocal pneumonia, Fibromyalgia
70	319	Sinus infections and bronchitis 5 times a year, ILD, Seronegative SLE, Asthma, Osteoporosis
70	500	Recurrent UTI, recurrent respiratory infection, chronic sinusitis, Bronchiectasis, Sjogren's, Hashimoto's thyroiditis
70	400	Chronic URI and sinus infections, Bronchiectasis, Asthma, 9/23 pneumococcal serotypes
71	353	Recurrent otitis in childhood, Multiple episodes of bronchitis and pneumonia, recurrent skin and nail infections, CAD, Psoriasis, chronic nephrolithiasis, multiple GI infections, 5/23 pneumococcal serotypes
71	415	Recurrent sinus infections since 40s, Rhinitis, Sjogren's, 2/23 pneumococcal serotypes
71	451	Recurrent sinus infections, Asthma, GERD, Arthritis, Osteoporosis, Dermatitis, Chronic cough, 4/23 pneumococcal serotypes
72	450	Bronchiectasis, COPD, asthma, chronic mycobacterial infections, Basal cell carcinoma, Osteoarthritis, 3/14 pneumococcal serotypes
72	523	Allergic Rhinitis, Multiple otitis and hearing Loss since age 3, chronic sinusitis, Hypothyroidism, Asthma, Hypothyroidism, 1/23 pneumococcal serotypes
72	563	Recurrent sinus infections approximately every 6 weeks, Rhinitis, Asthma,
72	599	Recurrent Infections, Osteoarthritis, Lymphopenia, Recurrent sinus infections despite therapy, Low B cells
73	322	Starting age 12, recurrent sinusitis, recurrent otitis, Asthma, Rhinitis
73	523	Chronic bronchitis in mid 30s, pneumonias with multiple hospitalizations throughout life
73	527	Chronic lung infections, Bronchiectasis, Pseudomonas lung disease, Rhinovirus, COPD, Recurrent MAC, Eczema, Asthma, "low pneumococcal serotypes" one month after vaccine
73	411	Recurrent pneumonia/ bronchiectasis exacerbations due to MSSA and pseudomonas, pulmonary MAC, pulmonary nocardiosis, Chronic sinusitis, antibiotic use 2-3 times a year, Arthritis, Hashimoto thyroiditis, URIs, Deceased from COVID-19
73	300	No recurrent infections, infrequent URIs, myasthenia gravis, MCTD, Sjogren's, pneumococcal 0/23 serotypes
74	363	Age 18 chronic sinopulmonary infections, asthma, recurrent sinus infections, Breast cancer, Hypothyroidism. ~12 months of steroids for cervical spine injury prior to diagnosis, but stopped for 6 months subsequent IgG 363, repeat 375. Pneumococcal 3/23 serotypes.
74	520	Chronic diarrhea, Gastritis, Recurrent sinusitis, chronic pharyngitis, chronic UTI Non-Hodgkin's lymphoma 2015. Elevated IgM consistently. CVID diagnosis was 8 y before lymphoma diagnosis
76	104	Chronic sinus infections in age 60s, seronegative RA, no Ig losses with panhypogammaglobulinemia, 0% switched memory B cells, 0/23 pneumococcal serotypes
76	370	Respiratory infections; sinusitis; anemia, Asthma, COPD / Emphysema; 11/23 pneumococcal serotypes
76	520	Sinus infections in childhood; recurrent scalp infections; Osteoarthritis, Chronic UTI, Neuropathy, IBS, Vasculitis
76	408	Recurrent respiratory infections, recurrent diarrhea; rubella infections several times, recurrent mouth sores; fatigue; fibromyalgia, osteoporosis
77	413	Recurrent respiratory infections in childhood; Asthma, Allergic rhinitis, IBS, Psoriatic Arthritis, Sepsis; No response to pneumovax
77	633	Chronic GI infections; Seronegative autoimmune arthritis, anemia; Low B cell memory, Low Class switched memory B cells, Low Plasmablasts

78	356	Frequent pharyngitis requiring tonsillectomy; recurrent otitis; Pneumonias twice a year as well as bronchitis; Bronchiectasis; low T cells; absent plasmablasts; elevated CD21 low B cells (12%)
78	<35	2-3 sinopulmonary infections per year, arthritis, osteomyelitis, thrombocytopenia; 1/14 pneumococcal serotypes
79	229	Recurrent infections as a child, Pneumonia 2x, chronic pruritus, COPD
80	86	Recurrent pharyngitis requiring tonsillectomy, age 20-30s recurrent sinus infections; bronchiectasis; chronic MAC and aspergillus, 3/23 pneumococcal titers
80	300	Recurrent skin and respiratory infections; Bronchiectasis; MAI; Pseudomonas; dermatitis, IDDM, Adrenal insufficiency. 8/23 pneumococcal titers
80	423	Bronchiectasis; Macrocytosis, Numerous skin infections; Hashimotos, Asthma, Polyarthritis, undifferentiated connective tissue disease, low memory B cells,
80	556	Recurrent sinopulmonary infections with antibiotic 3-4 times a year, hospitalized with pneumonia several times; COPD, Arthritis, Asthma, Emphysema; pneumococcal 7/23 serotypes, only 1% switched memory B cells
81	512	Chronic ear infections; Chronic Cough; Bronchiectasis; Squamous cell carcinoma of tongue, recent CKD
81	560	History of recurrent sinusitis in childhood; T2DM with peripheral Neuropathy, Giant cell arteritis, Parkinson's, recent development of CKD, GI Infections
81	460	Recurrent sinus infections since childhood; Leukopenia, Thrombocytopenia; Ankylosing Spondylitis, Diabetes, Breast cancer
81	548	Chronic Rhinitis, Bronchiectasis, MAI/MAC infections; Pseudomonas respiratory infections; Asthma, GERD
82	330	Chronic Rhinitis, Diverticulosis, Chronic Cough, GI Infections; PMH Carcinoid Tumor
82	403	Chronic Bronchitis, Sinusitis, COPD, Asthma, Chronic URI, Chronic GI infections, Diarrhea
82	535	Childhood recurrent respiratory infections; Bronchiectasis; Crohn's disease x 20 years
82	640	Recurrent Pneumonia; Neutropenia, Arthritis, Recurrent UTI, Sjogren's 2017, Chronic Cough, Adenopathy, Low B cells
83	554	Chronic Bronchitis, chronic rhinitis; Neutropenia, B Cell Lymphoma, Alopecia, Diabetes Mellitus, Low B Cells
83	590	Multiple Pneumonias with hospitalizations
83	630	Scleroderma, Osteoporosis, CKD, ILD, Chronic Upper respiratory Infections; recurrent GI infections
83	509	Recurrent Infections during early life (Respiratory, Skin, and sinus), Chronic Rhinitis; Leukopenia, CKD in 70s (mild proteinuria), no steroids
84	240	Chronic Diarrhea, GI infections; chronic sinusitis, Multiple pneumonias; COPD
84	300	Recurrent Respiratory infections, many bacterial skin infections, Hypothyroidism, Breast cancer 2004, pneumococcal 3/23 serotypes
84	355	Chronic Sinus infections, Osteoarthritis, Fibromyalgia, Atrophic gastritis
84	491	Recurrent Pneumonia; COPD, Bronchiectasis; Osteoarthritis, CREST, History of breast cancer, FVT, Hypothyroid
86	446	Recurrent lung infections, Chronic Rhinitis, Bronchiectasis, Leukopenia, Transaminitis, history of Basal cell carcinoma
88	482	Chronic Rhinitis, recurrent pneumonia; Asthma; Colonic Polyps
89	200	Recurrent URIs, Sinusitis, Mycosis Fungoides, no Steroids, 6/23 Pneumococcal serotypes
90	340	Chronic Sinusitis, in elderly years CKD, no autoimmunity
91	472	Chronic sinusitis; Rhinitis; 8/23 pneumococcal serotypes
92	426	Respiratory infections since 60s; COPD
59	121	Thymoma, 2/23 pneumococcal serotypes after vaccination
37	130	Pneumonia w empyema, +/- SLE; progressive diminishment of B cells, 3/23 pneumococcal serotypes

Table S2.

	Top 100	Random 100
Age (years)		
Mean	57.4	43.4
Median	61	42
Sex (%)		
Male	72	54
Female	28	46
Number of ICD codes		
Mean unique	242.0	30.8
Median	210	18
Medical record length		
Mean years	15.5	7.07

Demographics of top 100 patients identified by PheNet and 100 randomly sampled patients. We show a summary of the top 100 individuals with the highest PheNet score out of the discovery cohort (n = ~880K) and a control group of 100 randomly sampled patients from the patient population. We provide summary statistics on patients' age, self-reported sex, number of unique ICD codes, and the number of years recorded in the EHR.

Table S3

Feature	OMIM-derived	Phenotype	Frequency (case cohort)	Frequency (all patients)	Log-OR	OR	Category
279.1		Immunity deficiency	41.62%	0.48%	3.758	42.84	
283	x	Hemolytic anemias	4.57%	0.16%	3.333	28.01	Autoimmunity
255.2	x	Adrenal hypofunction	0.51%	0.01%	3.258	26.00	Autoimmunity
496.3	x	Bronchiectasis	23.35%	0.56%	2.995	20.00	Infection
287.31		Thombocytopenia	9.64%	0.24%	2.953	19.16	Autoimmunity
31		Diseases due to mycobacteria	4.57%	0.13%	2.886	17.92	Infection
279.2		Autoimmune disease NEC	6.09%	0.15%	2.746	15.57	Autoimmunity
709.7		Diffuse connective tissue disease	14.21%	0.51%	2.572	13.09	Autoimmunity
279.8		Immune mechanism disorders	4.06%	0.07%	2.483	11.98	Autoimmunity
502		Postinflammatory pulmonary fibrosis	15.23%	0.79%	2.336	10.34	Autoimmunity
IGG		Igg < 600	NA	NA	2.266	9.64	
253.5	x	Failure to Thrive	1.52%	0.10%	1.997	7.37	Other
504		Alveolar pneumonopathy	9.64%	0.52%	1.976	7.21	Autoimmunity
472		Chronic pharyngitis and nasopharyngitis	24.37%	1.96%	1.890	6.62	Infection
475	x	Chronic sinusitis	47.72%	4.45%	1.709	5.52	Infection
136		Other infectious and parasitic diseases	34.52%	3.13%	1.701	5.48	Infection
579.2	x	Splenomegaly	8.12%	1.02%	1.363	3.91	Autoimmunity
284.1	x	Pancytopenia	4.57%	0.81%	1.172	3.23	Autoimmunity
496	x	Chronic airway obstruction	13.71%	2.16%	1.160	3.19	Other
202.2	x	Non-Hodgkins lymphoma	4.57%	0.74%	1.059	2.88	Other
289.4	x	Lymphadenitis	17.77%	3.24%	1.048	2.85	Autoimmunity
480	x	Pneumonia	25.38%	4.67%	0.966	2.63	Infection
555.2	x	Ulcerative colitis	3.55%	0.56%	0.935	2.55	Autoimmunity
497	x	Bronchitis	12.18%	2.60%	0.840	2.32	Infection
41	x	Bacterial infection NOS	12.18%	2.86%	0.827	2.29	Infection
573.3	x	Hepatomegaly	6.09%	1.43%	0.789	2.20	Autoimmunity
696.4	x	Psoriasis	5.08%	1.33%	0.779	2.18	Autoimmunity
495	x	Asthma	42.13%	9.56%	0.762	2.14	Autoimmunity
251.1	x	Hypoglycemia	2.54%	0.72%	0.686	1.98	Other
686	x	Infections of skin and subcutaneous tissue	8.63%	2.52%	0.630	1.88	Infection
561.1	x	Diarrhea	23.86%	6.08%	0.590	1.80	Infection
320	x	Meningitis	1.52%	0.32%	0.574	1.77	Infection
369.5	x	Conjunctivitis, infectious	6.60%	3.08%	0.133	1.14	Infection
704.1	x	Alopecia	4.57%	2.11%	0.124	1.13	Autoimmunity

381.1	x	Otitis media	4.06%	1.92%	0.120	1.13	Infection
555	x	Inflammatory bowel disease	0.00%	0.00%	0.000	1.00	Autoimmunity
465	x	Acute upper respiratory infections	23.86%	12.86%	-0.078	0.92	Infection
315	x	Developmental delays and disorders	1.02%	1.14%	-0.746	0.47	Other
716	x	Other arthropathies	0.00%	0.02%	-8.989	0.00	Autoimmunity

Phenotype features and lab values scored by PheNet, Frequencies, and Categorization. Category indicates a broad grouping applied to each phenotype by clinicians. Autoimmunity here implies both autoimmune and inflammatory phenotypes. Other implies phenotypes that were not classified as either of the other two categories.