Supplementary Figures/Tables:

Supplementary Table 1: Biomarkers included in study by group

Pathway	Biomarker	Abbreviation	Measurement in HCLD- controls (pg/ml)(IQR)	Measurement in HCLD+ cases (pg/ml) (IQR)	P value
Immune Activation	Beta-2-Microglobulin	β2M	2577312.1 (1026666.1)	2708684.2 (1890922.2)	0.166
	Granulocyte Colony Stimulating Factor	GCSF	104.0 (66.8)	81.3 (60.6)	0.016
	Interferon induced protein-10	IP-10	77.8 (57.5)	115.9 (125.6)	<0.001
	C- Reactive Protein	CRP	351638.7 (810154.9)	681709.5 (3846846.5)	0.001
	Interferon Gamma	IFN-γ	77.8 (38.4)	108.4 (77.5)	<0.001
Monocyte Activation	Soluble CD14	sCD14	1630166.3 (677911.7)	1959313.6 (863922.8)	<0.001
	Soluble CD163	sCD163	704196.9 (557007.0)	824710.1 (644962.8)	0.014
T-Cell Activation	Soluble CD40-Ligand	sCD40-L	2912.1 (1184.9)	4454.5 (2037.8)	<0.001
	Soluble CD27	sCD27	6318.9 (3315.5)	8367.9 (4744.5)	<0.001
	Soluble CD25	sCD25	423.1 (272.4)	648.7 (367.6)	<0.001
	Soluble CCL5	sCCL5	36568.1 (32328.7)	44381.6 (26809.5)	0.022
Endothelial Activation & Cellular Adhesion	E-Selectin	sE-Selectin	28610.0 (13313.1)	35903.1 (18481.5)	<0.001
	P-Selectin	sP-Selectin	28284.5 (12424.7)	30784.6 (13095.8)	0.011
	Vascular cell adhesion molecule 1	VCAM-1	830248.8 (553079.4)	1065117.6 (698897.2)	<0.001
	Intracellular cell adhesion molecule 1	ICAM-1	247118.8 (231948.2)	293854.2 (287375.8)	0.139
	Vascular endothelial growth factor	VEGF	69.3 (50.6)	49.3 (40.0)	0.027
Coagulation	D-Dimer and fibrin degradation products	D-Dimer	1306651.5 (1021554.2)	1903081.0 (2002126.2)	<0.001
Apoptosis	Fas	Fas	4517.6 (1440.8)	5094.7 (1870.0)	0.001
	Growth differentiation factor 15**	GDF-15	797.3 (877.2)	828.2 (1125.3)	0.864

Angiogenesis	Angiopoietin-1*	ANG-1	11669.8 (12856.8)	14868.1 (12936.1)	0.013
Extracellular matrix degradation	Matrix metalloproteinase-1	MMP-1	1232.7 (1139.6)	1577.7 (1811.6)	0.021
	Matrix metalloproteinase-3	MMP-3	3570.5 (3392.8)	3506.9 (3865.6)	0.929
	Matrix metalloproteinase-7	MMP-7	1173.8 (685.0)	1490.0 (939.7)	<0.001
	Matrix metalloproteinase-8	MMP-8	2868.5 (5383.5)	4013.2 (6760.7)	0.049
	Matrix metalloproteinase-10	MMP-10	736.7 (607.9)	1071.4 (844.9)	<0.001
	Matrix metalloproteinase-12*	MMP-12	21.6 (19.8)	28.5 (19.8)	0.002

* Indicates measurements falling below limit of detection. Assigned half the lowest measured value from the whole population (MMP-12 n=54, Angiopoietin 1 n=5). ** Indicates missing data (GDF-15 n=1). Differences between groups assessed by Kruskal Wallis non-parametric test

Supplementary Table 2 – Principal Component Analysis

	Whole Population (n=410)			HCLD Population (n=336)		
	Eigenvalue	% Variance	Cumulative Variance (%)	Eigenvalue	% Variance	Cumulative % Variance
Principal Component 1	5.155	20.622	20.622	5.075	20.3	20.3
Principal Component 2	2.61	10.442	31.064	2.758	11.03	31.331
Principal Component 3	2.158	8.631	39.695	2.02	8.081	39.412
Principal Component 4	1.85	7.402	47.096	1.672	6.686	46.098
Principal Component 5	1.325	5.302	52.398	1.355	5.422	51.52
Principal Component 6	1.29	5.16	57.558	1.282	5.127	56.647
Principal Component 7	1.037	4.15	61.707	1.05	4.199	60.846
Principal Component 8	0.991	3.962	65.67	1.019	4.076	64.922
Principal Component 9	0.894	3.577	69.246	0.953	3.813	68.734
Principal Component 10	5.155	3.291	72.537	0.858	3.431	72.165

Eigenvalues and percentage of variance explained for each factor derived from principal component analysis in whole population and HCLD population. Grey squares represent dimensions with eigenvalues < 1 (not included in downstream analysis).

\

Supplementary Table 3: Biomarkers and HCLD Status

Biomarker ($Log_{10,}$ Scaled)	Univariate Logistic Regression, OR (CI, P)	Adjusted Logistic Regression, OR (CI, P)
MMP-1	1.41 (1.09-1.84, p=0.009)	1.36 (1.04-1.79, p=0.028)
MMP-3	1.02 (0.79-1.31, p=0.902)	1.05 (0.78-1.44, p=0.754)
MMP-7	1.52 (1.18-1.98, p=0.001)	1.42 (1.07-1.90, p=0.015)
MMP-8	1.24 (0.96-1.61, p=0.095)	1.14 (0.88-1.48, p=0.340)
MMP-10	1.70 (1.29-2.25, p<0.001)	1.61 (1.20-2.19, p=0.002)
MMP-12*	1.25 (0.98-1.58, p=0.063)	1.25 (0.95-1.63, p=0.101)
Angiopoietin-1	1.49 (1.19-1.90, p=0.001)	1.53 (1.20-1.96, p=0.001)
Beta-2-Microglobulin	1.14 (0.89-1.46, p=0.302)	1.03 (0.75-1.42, p=0.837)
sCCL5	1.36 (1.06-1.75, p=0.015)	1.37 (1.05-1.80, p=0.023)
sCD14	1.96 (1.51-2.57, p<0.001)	2.23 (1.66-3.05, p<0.001)
sCD25	2.74 (2.01-3.81, p<0.001)	2.85 (2.00-4.19, p<0.001)
sCD27	2.26 (1.65-3.16, p<0.001)	2.05 (1.48-2.91, p<0.001)
sCD40-L	2.89 (2.12-4.06, p<0.001)	2.96 (2.12-4.25, p<0.001)
sCD163	1.33 (1.04-1.72, p=0.024)	1.32 (1.00-1.76, p=0.052)
CRP	1.55 (1.20-2.04, p=0.001)	1.48 (1.12-1.98, p=0.006)
IP-10/CXCL10	1.93 (1.42-2.69, p<0.001)	1.89 (1.36-2.72, p<0.001)
D-Dimer +	1.68 (1.27-2.25, p<0.001)	1.68 (1.25-2.29, p=0.001)
sE-Selectin	2.08 (1.57-2.81, p<0.001)	2.05 (1.52-2.82, p<0.001)
Fas	1.59 (1.23-2.08, p=0.001)	1.59 (1.21-2.12, p=0.001)
GCSF	0.75 (0.58-0.97, p=0.032)	0.68 (0.50-0.91, p=0.010)
GDF-15	1.04 (0.81-1.34, p=0.778)	0.99 (0.75-1.32, p=0.950)
ICAM-1	1.05 (0.81-1.33, p=0.701)	1.00 (0.75-1.30, p=0.981)
IFN-γ	1.75 (1.35-2.28, p<0.001)	2.63 (1.77-4.12, p<0.001)
sP-Selectin	1.29 (0.99-1.69, p=0.064)	1.22 (0.92-1.63, p=0.175)
VCAM-1	1.70 (1.29-2.30, p<0.001)	1.56 (1.18-2.10, p=0.003)
VEGF	0.79 (0.61-1.02, p=0.070)	0.73 (0.55-0.95, p=0.022)

Logistic regression results for all biomarkers included in the study. Abbreviations: OR=Odds Ratio, CI=Confidence Interval, P = P-Value). Adjusted analysis controlled for age, sex, study site, height for age z-scores, HIV viral suppression and having ever been treated for TB.

*MMP-12 dropped from downstream analysis due to high levels of missingness

Supplementary Table 4: Association of biomarkers with FEV1 z-score

Biomarker	Univariate Linear Regression β ± SE	P-Value	Adjusted Linear regression β ± SE	P-value
MMP-1	-0.063 ± 0.039	0.109	-0.034 ± 0.039	0.374
MMP-3	0.021 ± 0.039	0.595	0.013 ± 0.042	0.767
MMP-7	-0.059 ± 0.039	0.132	-0.021 ± 0.039	0.586
MMP-8	-0.122 ± 0.039	0.002	-0.097 ± 0.039	0.012
MMP-10	-0.161 ± 0.038	0	-0.132 ± 0.04	0.001
MMP-12	-0.022 ± 0.039	0.571	-0.026 ± 0.039	0.502
Angiopoietin-1	0.08 ± 0.039	0.041	0.077 ± 0.039	0.046
Beta-2-Microglobulin	-0.055 ± 0.039	0.161	-0.063 ± 0.044	0.151
sCCL5	-0.04 ± 0.039	0.311	-0.04 ± 0.039	0.304
sCD14	0.046 ± 0.039	0.238	0.05 ± 0.039	0.2
sCD25	-0.07 ± 0.039	0.073	-0.035 ± 0.041	0.39
sCD27	-0.071 ± 0.039	0.069	-0.04 ± 0.04	0.311
sCD40-L	-0.034 ± 0.039	0.389	-0.026 ± 0.039	0.508
sCD163	0.018 ± 0.039	0.64	0.023 ± 0.04	0.563
CRP	-0.149 ± 0.038	0	-0.128 ± 0.038	0.001
IP-10/CXCL10	-0.089 ± 0.039	0.023	-0.08 ± 0.04	0.048
D-Dimer +	-0.062 ± 0.039	0.113	-0.055 ± 0.039	0.165
sE-Selectin	-0.104 ± 0.039	0.008	-0.082 ± 0.039	0.037
Fas	0.083 ± 0.039	0.035	0.082 ± 0.039	0.033
GCSF	-0.106 ± 0.039	0.007	-0.11 ± 0.039	0.006
GDF-15	-0.028 ± 0.039	0.474	-0.009 ± 0.04	0.821
ICAM-1	-0.04 ± 0.039	0.312	-0.016 ± 0.039	0.679
IFN-γ	0.013 ± 0.039	0.735	-0.005 ± 0.042	0.905
sP-Selectin	-0.028 ± 0.039	0.473	-0.023 ± 0.039	0.55
VCAM-1	-0.105 ± 0.039	0.007	-0.09 ± 0.039	0.021
VEGF	-0.117 ± 0.039	0.003	-0.113 ± 0.039	0.004

Univariate and adjusted linear regression results for biomarkers and FEV1 z-score in the HCLD group. SE= Standard Error. Adjusted analysis controlled for age, sex, study site, height for age z-scores, HIV viral suppression and having ever been treated for TB.

Supplementary Table 5: Principal components associated with HCLD and FEV-1 z-score

	Adjusted Logistic Regression Results (OR (95% CI, p))	Adjusted Linear Regression Results (β (SE), p)
Principal Component 1	1.54 (1.33-1.80, p<0.001)	-0.044 ± 0.018, p=0.014
Principal Component 2	0.97 (0.81-1.16, p=0.728)	0.004 ± 0.024, p=0.873
Principal Component 3	1.44 (1.16-1.81, p=0.001)	-0.042 ± 0.029, p=0.138
Principal Component 4	0.62 (0.49-0.77, p<0.001)	0.096 ± 0.03, p=0.001
Principal Component 5	0.71 (0.54-0.92, p=0.012)	-0.092 ± 0.036, p=0.011
Principal Component 6	0.74 (0.58-0.94, p=0.014)	-0.031 ± 0.036, p=0.395
Principal Component 7	1.47 (1.11-1.95, p=0.007)	0.082 ± 0.038, p=0.031
Principal Component 8		0.004 ± 0.039, p=0.911

Logistic regression in whole population, linear regression in only HCLD+ participants. All analyses include age, sex, site, height for age z score, HIV viral control and having ever been treated for TB as confounders.

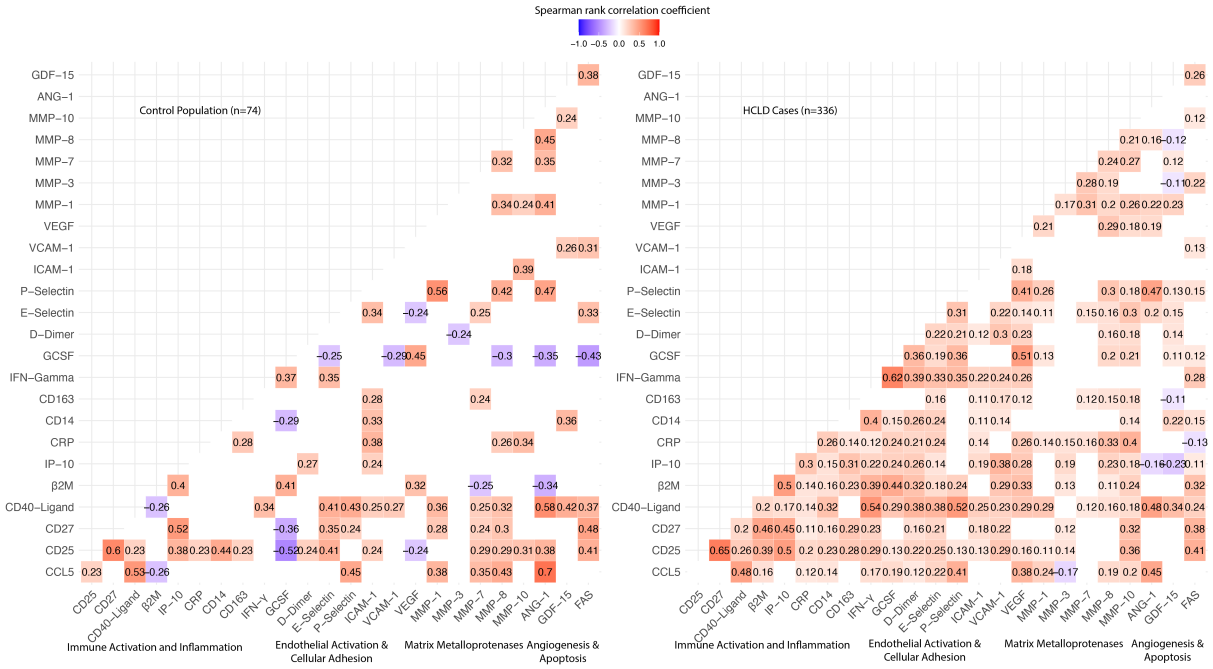
Supplementary Table 6: Sensitivity and specificity of biomarkers for HCLD.

	Threshold	Sensitivity	Specificity	AUC
MMP-1	3.297	0.77	0.387	0.586
MMP-3	3.613	0.473	0.595	0.497
MMP-7	3.196	0.784	0.452	0.635
MMP-8	3.25	0.405	0.753	0.573
MMP-10	2.919	0.622	0.658	0.643
MMP-12	1.45	0.77	0.509	0.617
Angiopoietin-1	3.808	0.284	0.908	0.593
Beta-2-Microglobulin	6.505	0.797	0.396	0.551
sCCL5	4.509	0.486	0.747	0.585
sCD14	6.324	0.851	0.435	0.685
sCD25	2.651	0.595	0.812	0.74
sCD27	3.837	0.608	0.705	0.688
sCD40-Ligand	3.526	0.716	0.812	0.768
sCD163	5.941	0.689	0.485	0.591
CRP	6.143	0.824	0.396	0.626
IP-10/CXCL10	1.918	0.568	0.729	0.665
D-Dimer +	6.165	0.635	0.643	0.648
sE-Selectin	4.493	0.622	0.679	0.687
Fas	3.715	0.757	0.47	0.627
GCSF	2.004	0.541	0.673	0.589
GDF-15	3.041	0.716	0.372	0.506
ICAM-1	5.593	0.797	0.339	0.555
IFN-γ	2.03	0.838	0.512	0.697
sP-Selectin	4.47	0.635	0.562	0.595
VCAM-1	5.867	0.432	0.815	0.641
VEGF	1.824	0.554	0.682	0.582
Principal Component 1	-0.485	0.797	0.643	0.76
Principal Component 2	-0.405	0.716	0.387	0.541
Principal Component 3	-1.035	0.473	0.821	0.637
Principal Component 4	0.358	0.649	0.622	0.642
Principal Component 5	-0.554	0.838	0.345	0.589
Principal Component 6	0.192	0.568	0.604	0.578
Principal Component 7	0.219	0.716	0.452	0.58

Sensitivity and specificity refer to values at the specified threshold. Biomarker levels are $\log_{10}$ transformed. AUC = Area Under the Curve

\

Supplementary Figure 1: Biomarker Correlations

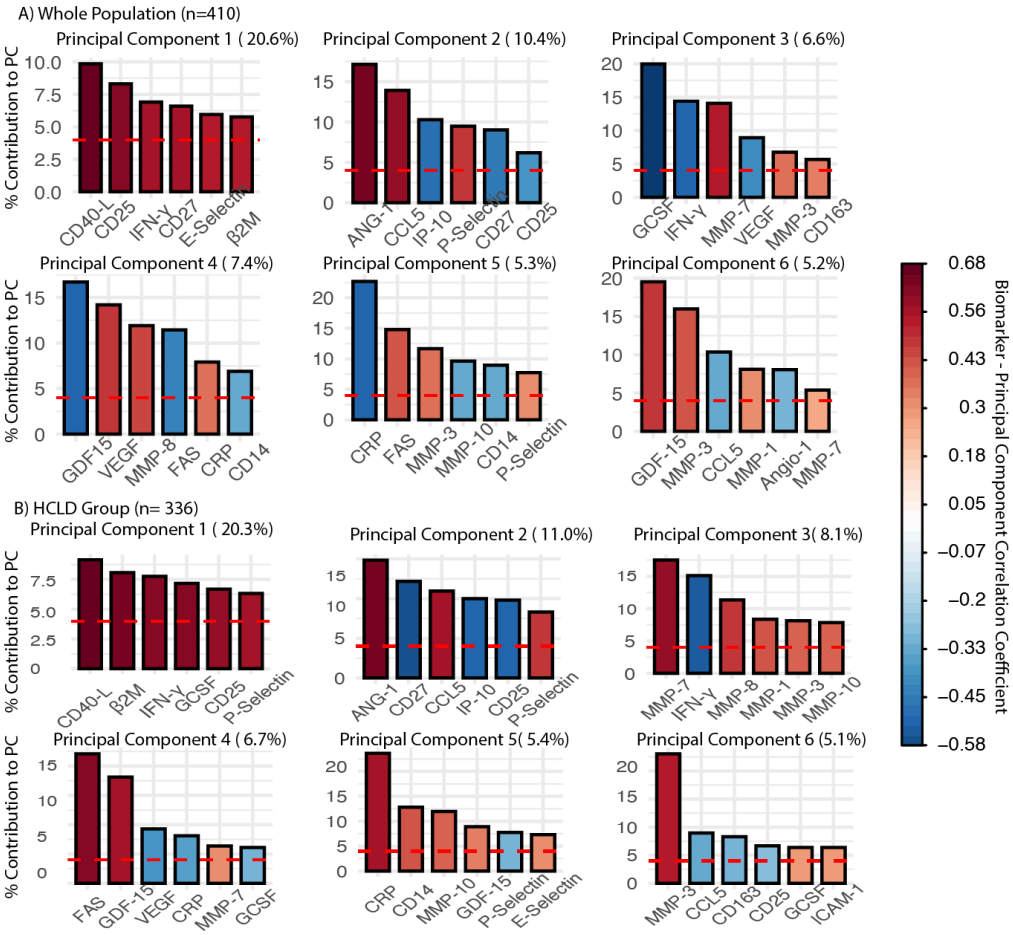


Spearman rank correlations between biomarkers in A) HCLD- controls B) HCLD+ cases.

Number within square indicates spearman rank correlation coefficient between biomarkers.

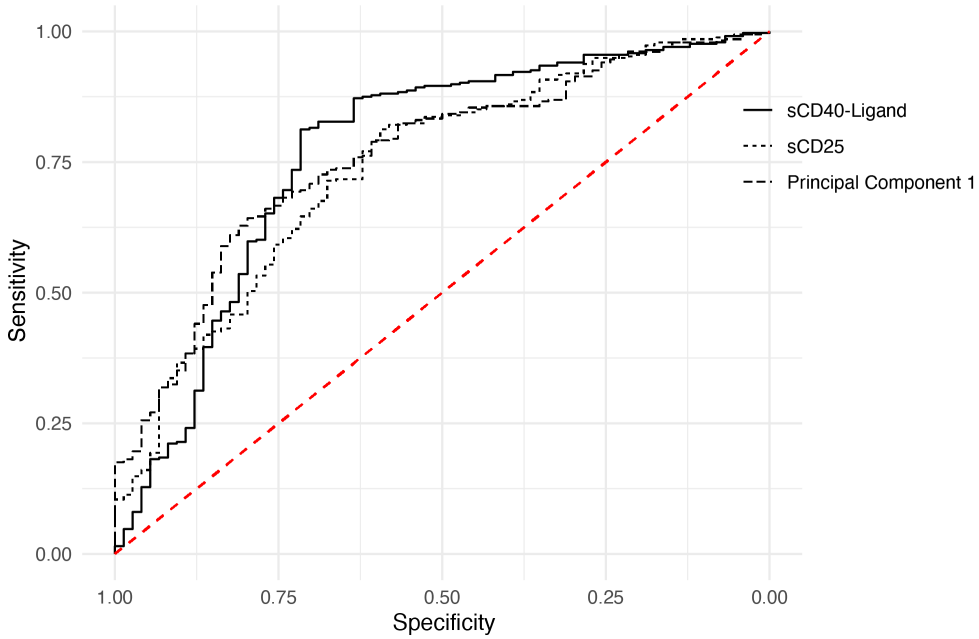
Blank squares represent no significant correlation between biomarker pairs ($p > .05$).

Supplementary Figure 2: Contribution of biomarkers principle components



Contribution of individual biomarkers to principal components derived from the A) Whole population and B) Participants with HCLD. Percentages in brackets show how much of variation of biomarker data is explained by each principal component. Top 6 biomarkers contributing to each component are shown. Dashed red line indicates contribution of biomarker if all equally contributing.

Supplementary Figure 3: Sensitivity and Specificity of markers for HCLD



ROC curve for top three performing variables. Only variables with AUC >0.7 included

Diagonal line represents no predictive ability. AUC = Area under the Curve