Biomarker Signatures of Sickle Cell Disease Severity: Supplement Information

Study populations

The CSSCD was a 10-year, multi-center, longitudinal study designed to document the natural history of sickle cell disease [31,32]. The study began recruiting patients in 1979 from 23 centers resulting in over 3,600 patients in four age groups: newborns, children, adolescents, and adults. The study had numerous goals including examining the effects of the disease on growth and development, and the role of the disease in complicating other health events. To achieve these ends, patients underwent numerous diagnostic tests, many circulating biomarkers were measured at regular time intervals, and patients were followed to document various complications of the disease.

Participants in the Pulmonary Hypertension and the Hypoxic Response in Sickle Cell Disease (PUSH) trial and Sickle Cell Disease with Sildenafil Therapy (Walk-PHaSST) trial were used as more contemporary cohorts to replicate the findings in the CSSCD and generate robust data for future studies [33,34]. The multi-center PUSH study was designed to determine the prevalence, risk factors and clinical consequences of elevated systolic pulmonary artery pressure in children and adolescents with sickle cell disease in the US. Five hundred and ten children with between the ages of five and 20 years were enrolled and evaluated at steady state with an echocardiogram, six-minute walk test and nt-proBNP. Full follow-up evaluations at approximately two years were performed in 200 of the subjects, including 160 with HbSS [33]. The multicenter Walk-PHaSST study enrolled 720 adolescents and adults in the US and UK and evaluated them with echocardiogram, six-minute walk test and nt-proBNP. Follow-up for mortality was performed in 632 of the patients after about two and one-half years [35].

Biomarkers removed from the initial list:

1. Removed because dichotomous: Howell-Jody bodies; urine analysis: glucose, >5 RBC/HPF; >5 WBC/HPF; positive urine culture.
2. Removed because too homogeneous: liver function test: total protein; positive urine culture organism 1, 2; urine analysis: Casts/HPF, pH, protein, specific gravity; Differential: metamyeolcytes/myelocytes, atypical cells, basophils, nucleated RBC.
3. Removed because of high missing: SGPT, ISC's.
4. Removed because of correlation with other markers: Blood count: hematocrit, RBC, MCH; Differential: PMN.

Supplementary Table 1 – Distribution of biomarkers in clusters

Biomarker (median [IQR])	Overall (n = 2320)	Cluster 1 (n = 657)	Cluster 2 (n = 437)	Cluster 3 (n = 364)	Cluster 4 (n = 341)	Cluster 5 (n = 273)	Cluster 6 (n = 91)	Cluster 7 (n = 52)	Cluster 8 (n = 47)
Hemoglobin (g/dL)	8.9 [7.9, 10.4]	8.7 [7.9, 9.6]	10.9 [10.0, 11.8]	7.9 [7.2, 8.7]	8.3 [7.6, 9.0]	10.9 [10.1, 11.8]	8.6 [8.0, 9.6]	7.6 [6.5, 8.8]	7.6 [7.0, 8.3]
White blood cell count	10.7 [8.4, 13.6]	11.2 [9.1, 13.6]	8.2 [6.3, 10.3]	13.6 [10.9, 16.4]	11.7 [9.9, 14.1]	7.7 [6.1, 9.4]	12.6 [10.8, 14.8]	11.3 [9.2, 15.1]	13.7 [11.0, 17.3]
Mean corpuscular volume (fl)	86.0 [78.0, 92.0]	86.0 [80.0, 92.0]	77.0 [71.2, 84.0]	89.0 [85.0, 94.0]	92.0 [86.0, 98.0]	78.0 [73.0, 84.0]	87.0 [81.5, 93.0]	88.0 [83.8, 93.2]	94.0 [88.5, 101.0]
Differential eosinophils (%)	3.0 [1.0, 5.0]	3.0 [1.0, 5.0]	2.0 [1.0, 4.0]	3.0 [1.0, 6.0]	2.0 [1.0, 4.0]	2.0 [1.0, 4.0]	4.0 [2.0, 6.0]	2.5 [1.8, 5.0]	1.0 [1.0, 3.0]
Differential lymphocytes (%)	38.0 [29.0, 48.0]	39.0 [30.0, 49.0]	37.0 [28.0, 47.0]	32.0 [25.0, 43.0]	40.0 [32.0, 49.0]	40.0 [31.0, 51.0]	38.0 [25.0, 51.0]	35.0 [28.0, 46.8]	41.0 [32.0, 47.0]
Differential monocytes (%)	6.0 [4.0, 9.0]	7.0 [4.0, 10.0]	6.0 [4.0, 8.0]	6.0 [4.0, 9.0]	7.0 [4.0, 10.0]	5.0 [3.0, 7.0]	6.0 [5.0, 9.5]	6.0 [4.0, 8.0]	8.0 [4.2, 11.0]
Platelets (x 10^9/L)	390.0 [300.0, 499.0]	412.5 [331.0, 508.2]	305.0 [228.0, 379.0]	450.0 [364.5, 550.8]	415.0 [335.0, 510.0]	296.0 [224.0, 365.0]	696.5 [559.0, 822.8]	372.5 [292.5, 478.8]	420.0 [327.0, 555.0]
Reticulocytes (%)	8.5 [4.5, 14.2]	10.0 [6.4, 15.2]	3.3 [1.9, 5.0]	13.5 [9.4, 18.3]	11.3 [7.8, 15.4]	4.2 [2.8, 6.7]	10.2 [7.2, 16.6]	12.4 [7.5, 19.5]	18.6 [13.5, 26.1]
Blood urea nitrogen (mg/dL)	8.0 [6.0, 10.0]	8.0 [6.0, 10.0]	9.0 [8.0, 12.0]	7.0 [6.0, 8.0]	8.0 [6.0, 10.0]	9.0 [7.0, 11.0]	8.0 [5.2, 9.0]	12.0 [10.0, 17.0]	9.0 [7.0, 10.0]
Creatinine (mg/dL)	0.6 [0.5, 0.8]	0.5 [0.4, 0.7]	0.7 [0.5, 0.9]	0.5 [0.4, 0.7]	0.6 [0.5, 0.8]	0.7 [0.5, 0.9]	0.5 [0.4, 0.7]	0.8 [0.6, 1.1]	0.5 [0.4, 0.7]
Uric acid (mg/dL)	5.2 [4.2, 6.3]	4.9 [4.0, 6.0]	4.8 [4.0, 5.8]	5.4 [4.4, 6.2]	5.3 [4.4, 6.2]	5.1 [4.4, 6.2]	5.1 [4.3, 6.0]	8.6 [7.4, 11.0]	7.6 [6.2, 8.7]
Total bilirubin (mg/dL)	2.1 [1.3, 3.4]	2.3 [1.5, 3.6]	1.1 [0.8, 1.7]	3.0 [2.1, 4.4]	3.0 [2.1, 4.5]	1.1 [0.8, 1.7]	1.8 [1.2, 2.4]	2.8 [2.1, 4.0]	3.3 [2.5, 4.6]
AST (units/dL)	45.0 [31.0, 65.0]	49.0 [36.0, 70.0]	30.0 [22.0, 45.0]	50.0 [36.0, 68.0]	52.0 [40.0, 66.0]	36.5 [26.0, 62.5]	37.0 [28.0, 52.0]	51.5 [35.0, 80.0]	67.0 [57.5, 90.0]
Alkaline phosphatase (units)	136.0 [87.0, 192.0]	123.0 [86.0, 173.0]	117.0 [68.0, 179.0]	146.5 [98.2, 191.0]	150.0 [93.0, 204.8]	165.0 [91.0, 222.0]	143.5 [111.0, 189.0]	148.0 [106.0, 205.2]	121.0 [100.5, 175.5]
LDH (mg/dL)	354.0 [225.0, 539.0]	329.5 [201.5, 470.0]	255.0 [183.0, 335.0]	507.5 [362.8, 647.2]	555.0 [418.2, 668.8]	256.0 [147.5, 345.0]	343.5 [217.5, 439.5]	505.0 [351.0, 694.0]	550.0 [305.2, 934.2]
Albumin (g/dL)	4.4 [4.2, 4.6]	4.4 [4.2, 4.6]	4.4 [4.2, 4.6]	4.4 [4.2, 4.6]	4.5 [4.2, 4.7]	4.5 [4.3, 4.7]	4.4 [4.2, 4.6]	4.3 [4.1, 4.5]	4.3 [4.0, 4.6]
Fetal hemoglobin (%)	5.0 [2.2, 9.9]	5.8 [2.8, 10.6]	5.6 [2.0, 14.0]	4.6 [2.5, 8.6]	6.7 [3.7, 10.5]	2.0 [0.9, 4.8]	4.9 [2.4, 8.8]	3.4 [1.5, 8.1]	4.8 [1.9, 7.0]

Supplementary Table 2 –Demographic characteristics of sickle cell disease patients for all clusters

	n	Age (mean (sd))	Follow-up years[1] (mean (sd))	Sex = Male (%)	Fetal hemoglobin[2] (%, mean(sd))	Hemolytic Score[3]
Overall	2320	15.30 (12.15)	6.44 (2.32)	1217 (52.5)	8.47 (11.89)	0 (1)
Cluster 1	657	14.97 (11.30)	6.49 (2.27)	327 (49.8)	8.76 (10.83)	0.31 (1.13)
Cluster 2	437	15.46 (13.13)	6.50 (2.78)	229 (52.4)	11.56 (15.97)	-1.38 (1.09)
Cluster 3	364	14.91 (11.61)	6.17 (2.09)	188 (51.6)	7.23 (9.42)	0.86 (1.08)
Cluster 4	341	16.26 (11.72)	6.44 (2.10)	208 (61.0)	10.04 (13.81)	0.84 (1.11)
Cluster 5	273	15.04 (12.89)	6.45 (2.42)	144 (52.7)	3.99 (5.42)	-1.31 (0.99)
Cluster 6	91	12.69 (10.78)	6.63 (1.86)	41 (45.1)	7.42 (9.62)	0.47 (1.10)
Cluster 7	52	18.98 (14.57)	6.16 (1.92)	26 (50.0)	6.20 (7.44)	0.72 (1.23)
Cluster 8	47	15.40 (12.34)	6.56 (2.12)	28 (59.6)	4.94 (3.72)	1.28 (1.21)
Cluster 9	13	11.23 (7.62)	6.91 (1.31)	6 (46.2)	12.67 (20.89)	0.34 (0.89)
Cluster 10	12	11.00 (9.46)	6.43 (2.92)	6 (50.0)	7.98 (8.18)	-0.82 (1.30)
Cluster 11	9	12.22 (9.88)	6.96 (0.93)	2 (22.2)	2.87 (2.44)	-0.08 (1.28)
Cluster 12	9	22.11 (13.39)	7.94 (3.63)	5 (55.6)	11.70 (21.14)	0.74 (0.74)
Cluster 13	6	36.50 (26.37)	6.14 (1.90)	3 (50.0)	7.40 (6.59)	0.52 (NA)
Cluster 14	4	6.00 (2.94)	6.39 (1.36)	2 (50.0)	4.55 (4.37)	0.57 (2.06)
Cluster 15	2	37.50 (0.71)	5.07 (3.10)	1 (50.0)	1.75 (1.06)	0.73 (NA)
Cluster 16	2	16.00 (18.38)	6.70 (2.40)	0 (0.0)	3.15 (2.19)	//
Cluster 17	1	12.00 (NA)	7.15 (NA)	1 (100.0)	5.80 (NA)	//

[1]Time from baseline until death or date of last entry in any study dataset

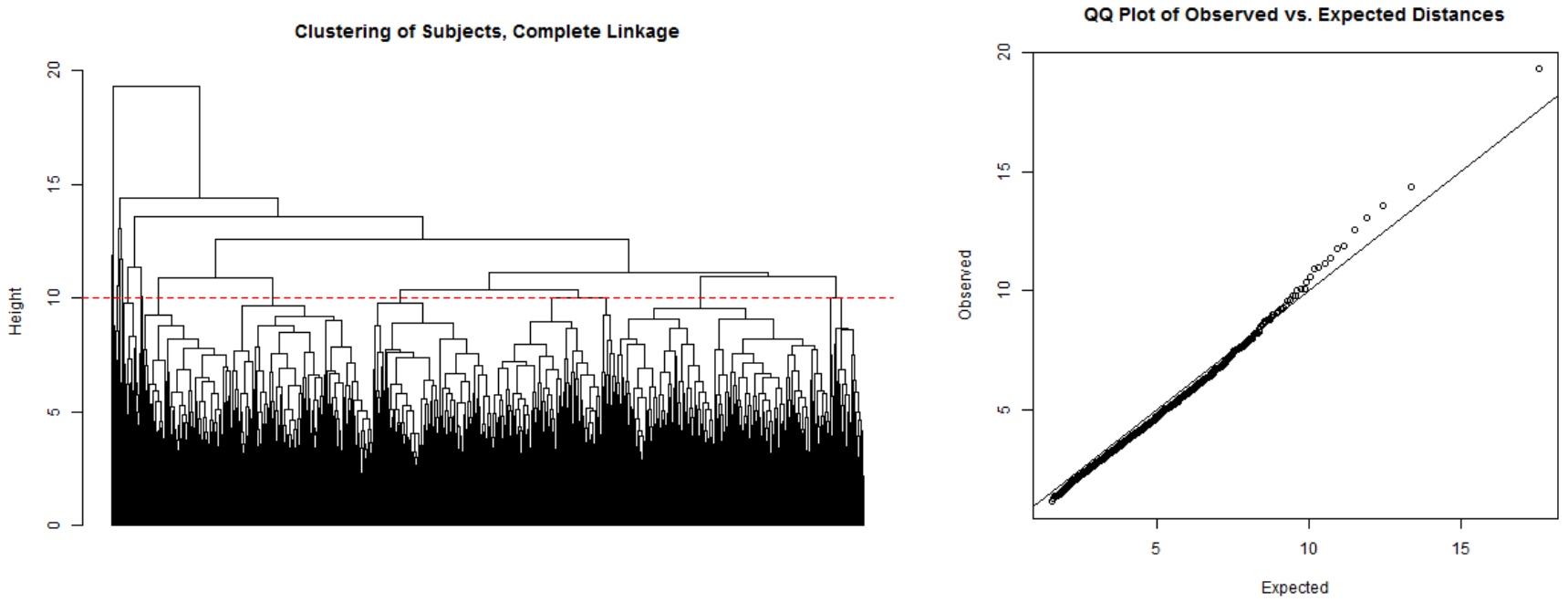
[2]Baseline fetal hemoglobin

[3] Hemolytic score at baseline is the first principal component from principal component analysis of AST, reticulocyte counts, LDH and hemoglobin [40]. A positive score denotes more hemolysis.

Supplementary Table 3 – Number of events for various complications by cluster

		Overall (n = 2320)	Cluster 1 (n = 657)	Cluster 2 (n = 437)	Cluster 3 (n = 364)	Cluster 4 (n = 341)	Cluster 5 (n = 273)	Cluster 6 (n = 91)	Cluster 7 (n = 52)	Cluster 8 (n = 47)
Death	Subjects	132	42	9	31	14	13	7	7	4
Stroke	Subjects	56	18	2	18	7	4	2	2	1
Seizure	Subjects	91	27	17	16	7	4	8	5	5
Leg Ulceration	Subjects	142	45	8	37	27	3	7	7	6
Acute Chest Syndrome	Subjects	700	221	85	128	107	58	44	16	18
Avascular Necrosis	Subjects	314	98	57	44	40	37	19	7	4
Acute Painful Episode	Subjects	942	294	139	152	153	108	41	15	19
At Least One Painful Episode	Subjects, records total	2257, 10207	637, 2966	418, 1788	356, 1601	335, 1533	264, 1197	91, 428	52, 225	47, 219
	Subjects, records with event	1632, 4324	500, 1331	261, 653	257, 698	241, 684	196, 484	64, 177	43, 110	34, 98
Severe Acute Episode	Subjects, records total	923, 3717	290, 1303	136, 516	151, 621	150, 670	102, 302	39, 137	15, 33	19, 68
	Subjects, records with event	639, 2282	214, 857	96, 298	96, 385	100, 373	64, 172	26, 89	13, 24	15, 46

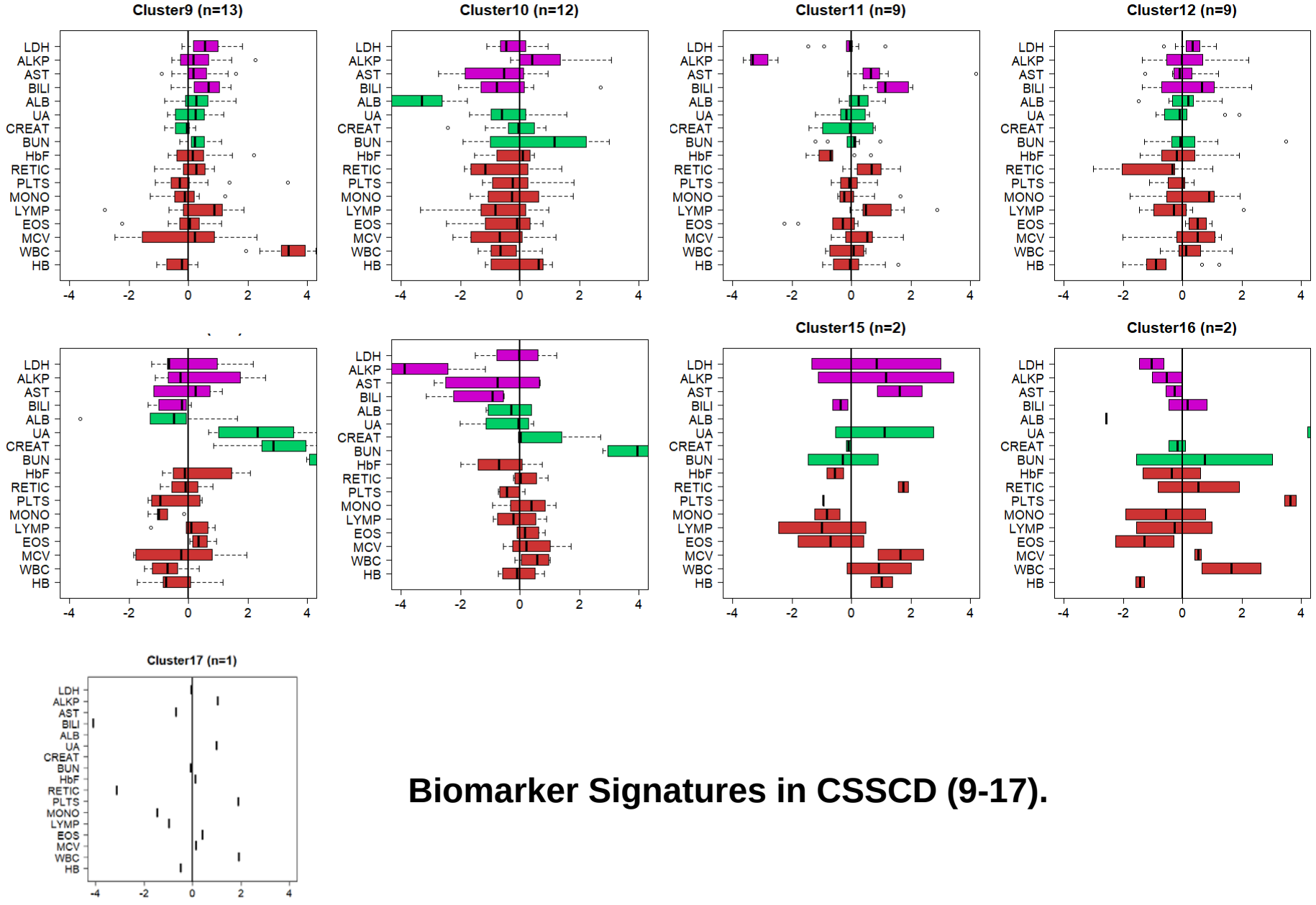
Supplement Figure 1



Supplement Figure 1. Results of hierarchical clustering analysis.
A hierarchical clustering algorithm with complete linkage and Euclidean distance was used to group the subjects by similarities in their biomarker profiles. Panel (a) shows a dendrogram displaying the steps of the clustering algorithm. The dotted line marks the height that indicates significant clusters at the 0.005 level which was determined using the permutation approach described next. Panel (b) shows a QQ plot of the observed dendrogram distances versus the expected distances under the null hypothesis of no clusters in the data. This QQ plot shows that there are clusters because at the bottom left part of the plot, the observed distances are shorter than the expected distances (suggesting that there are clusters of profiles more similar than random pairs) and then at the top right of the plot, the observed distances are longer than expected (suggesting that there are clusters of profiles more dissimilar than random pairs).

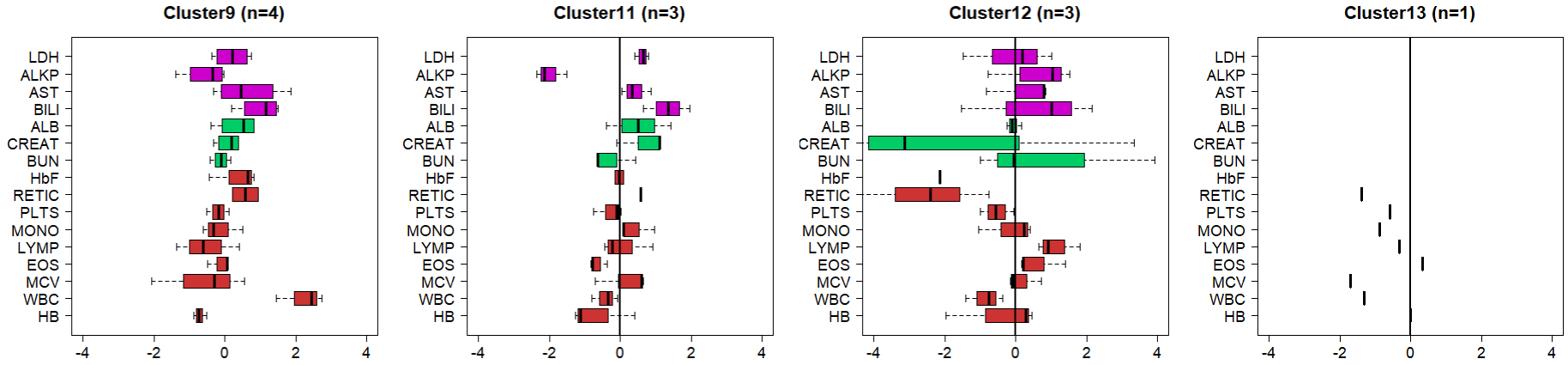
To generate the expected distances, we used reshuffling of each biomarker profile to generate a random data set with no clusters, and hierarchical clustering of the reshuffled data to generate a random dendrogram under the null hypothesis of no clusters in the data. The procedure was repeated multiple times and dendrograms were averaged to generate the reference distribution. The level of significance was chosen to have a good trade off between Type I error and cluster size.

Supplement Figure 2



Biomarker Signatures in CSSCD (9-17).

Supplement Figure 3
Biomarker Signatures in PUSH (9, 11, 12, 13)



Supplement Figure 4
Biomarker Signatures in WalkPHASST (9, 11, 12, 14, 15)