Supporting Information for:

Identifying Differentiation Stage of Individual Primary Hematopoietic Cells from Mouse Bone Marrow Using TOF-Secondary Ion Mass Spectrometry

Jessica F. Frisz[†], Ji Sun Choi[‡], Robert L. Wilson[†], Brendan A. C, Harley[‡,§], Mary L. Kraft[‡†*]

[†]Department of Chemistry, [‡]Department of Chemical and Biomolecular Engineering, and [§]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

Supplemental Methods

Hematopoietic cell isolation and preparation. Euthanization and tissue collection were performed using IACUC approved methods. After isolation, the bones were gently crushed with a mortar and pestle, washed with a solution of phosphate buffered saline (PBS) + 2% fetal bovine serum (FBS) (PBS/FBS), and filtered with a 40 μm sterile filter to isolate whole bone marrow. All subsequent steps were then performed in a PBS/FBS solution on ice. Red blood cells were then lysed with ACK lysis buffer (Invitrogen, Carlsbad, CA) and re-suspended in PBS/FBS with Fc receptor-blocking antibody to reduce non-specific antibody binding. Aliquots of the mononuclear bone marrow cell population were then stained with distinct antibody cocktails to facilitate identification of the HSPC, CLP, and B cell populations using standardized sorting protocols.[1-2] HSPCs (Lin^- $Sca\text{-}1^+c\text{-}Kit^+$) were identified with a cocktail of antibodies: PE-conjugated Sca-1 (1:100), APC-conjugated c-Kit (1:200), a FITC-conjugated Lineage (Lin) cocktail (CD5, B220, Mac-1, CD8a, Gr-1, Ter-119; 1:200), and a propidium iodide (PI) nuclear stain. CLPs ($Lin^-IL\text{-}7R\alpha^+Sca\text{-}1^{med}c\text{-}Kit^{med}$) were identified with a cocktail of antibodies: PE-conjugated Sca-1 (1:200), APC-conjugated c-Kit (1:200), PE-Cy 7-conjugated IL-7Rα (0.2 mg/mL), a fluorescein isothiocyanate- (FITC)-conjugated Lin cocktail (1:200), and a 4',6-diamidino-2-phenylindole (DAPI) nuclear stain. Mature B cells ($B220^+IgM^+$) were isolated with a cocktail of antibodies: FITC-conjugated B220 (1:200), eFluor 450-conjugated IgM (0.2 mg/mL), and a PI nuclear stain. All antibodies were supplied by eBioscience (San Diego, CA). Distinct HSPC, CLP, and B-cell populations were then isolated using a BD FACS Aria II flow cytometer operated by the BD FACS Diva software.

FACS plots for HSPCs, CLPs, and B-cells showed populations consistent with previous results in the literature.[1-5] Cells were collected on polylysine-coated substrates on ice and were then chemically fixed in 4% glutaraldehyde diluted with PBS within 4 hr after collection, followed by osmication in 0.4% osmium tetroxide diluted with triple-distilled water and filtered using a 0.22 μm syringe filter directly prior to use. Cells were rinsed for 20 min in triple distilled water and allowed to air dry. To facilitate locating the cells in the TOF-SIMS, optical maps of each sample substrate were created with an optical microscope (Leica DM6000 B, Q-Imaging EXi Blue Fluorescence Microscope) that was operated in reflectance mode.

Identification of outliers in the spectra acquired from the two groups of mice. Spectra that exhibit unusual variation that might be caused by sample contamination or high levels of inorganic ions that alter the relative intensity of the other mass peaks in the spectra can compromise the predicative ability of a multivariate model.[6-7] To detect the spectra that exhibited unusual variance compared to the rest of the cells that were harvested from the same set of mice, PCA was performed on the mass peaks between 1 and 300 *m*/z in the spectra. The upper mass of 300 *m*/z was selected because peaks with *m*/*z* > 300 have poor detection reproducibility using our instrumentation.[8] Each mass peak was normalized to the total intensity of the peaks in the spectrum and autoscaled to the spectra in the data set. Two separate PCA models with two principal components (PCs) were constructed: one for the cell spectra from the old mice, and another for the cell spectra from the young mice. The plot of the Hotelling T^2 statistic versus the Q residual contributions (**Fig. S1**) was used to identify the samples that had a Q statistic greater than the 95% confidence limit, which indicates the sample exhibited unusual variance that was not captured by the model. For the data set consisting of the cells harvested from the five old mice, the spectra from 6 HSPCs, 5 CLPs, and 1 B cell (arrows, **Fig. S1A**) were identified as outliers. Spectra from 3 HSPCs, 2 CLPs, and 1 B cell (arrows, **Fig. S1B**) that were harvested from the group of young mice were identified as outliers and excluded from further analysis. These outlier spectra were excluded from further analysis.

PLS-DA Model Construction. PLS-DA was performed on the peaks in the *m*/*z* 50 to 300 range. This range was selected because ions with *m*/*z* < 50 are not chemically specific[9] and peaks with *m*/*z* > 300 have poor detection reproducibility using our instrumentation.[8] Two different peak sets were employed to analyze the HC spectra: 1) a cell-related peak set consisting of the mass peaks that are known to be produced by phosphocholine, fatty acids, and amino acids[10] but not common surface contaminants (*m*/*z* 73, 133, 147, 207, 221, and 281)[11] (**Table S1**); and 2) a complete peak set that consisted of all mass peaks from *m*/*z* 50 to 300 except those related to the aforementioned contaminants. Previous studies indicated that the cell-related peak set is applicable to chemically fixed biological samples.[10,12] Each peak was normalized to the total intensity of the remaining peaks in the spectrum and autoscaled to the data set. The PLS-DA model was constructed using the minimum number of latent variables (LVs) required to capture at least 80% of the variance in the test spectra. Identification plots were generated to determine which cells exceed the classification threshold for the indicated class, where the classification threshold was estimated using Bayes' Theorem.[13] Variable importance in projection (VIP) score plots that show the magnitude that each mass peak contributes to identifying the indicated cell population were also generated. See the Supplemental Information and **Figure S2** for details of the construction of specific PLS-DA models.

PLS-DA Model for Identification of the Differentiation Stage of Hematopoietic Cells Isolated from the Same Mice. After removing the outlier spectra, the remaining normalized spectra in the old mice data set were divided into two sets that were nearly equal in size: an old mice calibration set consisting of 15, 13, and 15 randomly selected B cells, CLPs, and HSPCs, and an old mice test set consisting of spectra from the remaining 15, 12, and 13 B cells, CLPs, and HSPCs. Using the cell-related peak set, a preliminary PLS-DA model was constructed and refined by excluding the mass peaks with Q contributions above 30% (*m*/*z* 90, 98, 102, 110, 111, 120, 122, 188, 196, and 200) (**Fig. S2A**). The peaks with these high Q residual contributions had intensities that varied significantly between cells but were not indicative of differentiation status. Using the remaining spectra, a PLS-DA model was constructed that consisted of five LVs that captured 61.2% of the variance in the calibration data, and

81.4% of the variance in the test spectra. The PLS-DA model was then used to predict the differentiation status of the cells in the test set. When this process was repeated using the complete peak set described above, a PLS-DA model with 4 LVs that captured 43.1% and 83.4% of the variance in the calibration and test spectra, respectively, was constructed and used to predict the differentiation status of the cells in the test set.

Preparation of Samples of Population-Specific Antibody Cocktails. A separate sample was prepared for each aforementioned antibody cocktail used to isolate the HC populations. A 20 μL droplet of the antibody cocktail was spotted onto a 5 x 5 mm silicon substrate and allowed to air dry. The samples were chemically fixed as described above, but osmuim tetroxide treatment was omitted. The omission of osmium tetroxide is not expected to affect which peaks contribute significantly to identifying each antibody cocktail because because osmuim tetroxide does not react with proteins,[14-15] and only affects the intensity of the peaks that contain phosphate moieties does not produce new peaks in positive ion spectra over the *m*/*z* 1 – 300 range.

PLS-DA Model of Spectra from HC Population-Specific Antibody Cocktails. PLS-DA of the spectra acquired from the population-specific antibody cocktail samples was performed using the same peaks and preprocessing conditions as those used to construct the model shown in **Figure 2**. The calibration set consisted of 15 spectra of the B cell-specific antibody cocktail, 15 spectra of the CLP-specific antibody cocktail, and 14 spectra of the HSPC-specific antibody cocktail. The resulting PLS-DA model consisted of two LVs that captured 81.3% of the variance in the test set.

Construction of Principal Component Analysis (PCA) Models of each HC Population. Each spectrum was filtered so that it consisted of only the cell-related peaks within the *m*/*z* 50 to 300 range, and each peak was normalized to the total intensity of the remaining peaks in the spectrum. A data set consisting of the spectra of cells that were harvested from the old and young mice was created for each HC population. Each peak in the data set was autoscaled, and a separate PCA model was constructed for the B cells, CLPs, and HSPCs. The PCA model contained the minimum number of PCs that were required to separate the cells harvested from the young and old mice. Scores plots that show the

projection of each sample onto the PCs and loading plots that show the contribution of each peak to each PC were generated.

PLS-DA Model of Spectra from Hematopoietic Cells Isolated from Different Mice. Data preprocessing was identical as that described for PLS-DA model construction using data from the same mice. The calibration set of data that was used to for this experiment consisted of all the (non-outlier) spectra that had been acquired from the five old mice. These data were preprocessed using the same procedure as that described for the construction of the other PLS-DA model. This new calibration set consisted of the spectra from 30 B cells, 25 CLPs, and 29 HSPCs that were isolated from the five old mice. The preliminary PLSDA model that was constructed using the cell-related peaks in the calibration set of spectra from the old mice was refined by excluding the mass peaks with Q contributions above 60% (*m/z* 71, 90, 138, 188, 190, 194, 196, 200, 212, 231, 240, and 241) (**Fig. S2B**). The refined PLS-DA model consisted of 6 LVs that captured 65.9% and 82.3% of the variance in the calibration and test spectra, respectively. This model was then used to predict the differentiation status of 20 test B cells, 20 test CLPs, and 14 test HSPCs that were harvested from the three young mice. This process of PLS-DA model construction and testing was repeated using the same spectra for the calibration and test set, but the complete peak set (described above) was employed. The PLS-DA model constructed using the complete peak set consisted of 4 LVs that captured 39.8% and 81.0% of the spectral variance in the calibration and test data sets, respectively.

Supplemental Figures and Tables

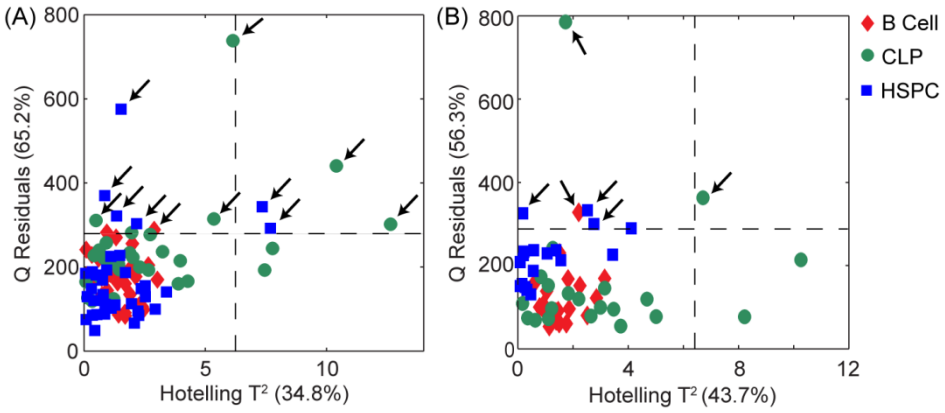

Figure S1. Identification of outlier spectra from cells harvested from the old (A) and young (B) mice. Using the spectra acquired for each group of cells, a PC model with two principal components was constructed using the mass peaks in the 1 – 300 *m*/*z* range. Outlier samples exhibited spectral variation that was not captured by the PCA model. The outlier spectra (indicated with black arrows) have Q residual values that are greater than the 95% confidence interval (horizontal dashed line). These samples were removed from the data set and excluded from further analysis.

Table S1. List of the mass peaks in the cell-specific peak set and the biomolecules that are related to these peaks.[10,9] Peaks that are produced by common surface contaminants were omitted.[11]

Mass (*m*/*z*)	Related Biomolecule
51	phenylalanine
53	lipid ($C_4H_5^+$)
54	lipid fragment ($^{13}C^{12}C_3H_5^+$)
55	lipid fragment ($C_4H_7^+$), arginine, lysine, cysteine, valine, leucine, histidine
56	lipid fragment ($C_3H_6N^+$), lysine, methionine, glutamine, asparagine, threonine, isoleucine
57	lipid fragment ($C_4H_9^+$), valine, serine, alanine, arginine, cysteine, aspartic acid, leucine, threonine, glycine
58	lipid fragment ($C_3H_8N^+$), isoleucine
59	lipid fragment ($^{13}C^{12}C_2H_8N^+$), arginine, valine
60	lipid fragment ($C_3H_{10}N^+$), serine
61	methionine
67	lipid fragment ($C_5H_7^+$)
68	lipid fragment ($C_4H_6N^+$), proline
69	lipid fragment ($C_5H_9^+$), isoleucine, histidine, lysine
70	lipid fragment ($C_4H_8N^+$), proline, leucine, glutamic acid, asparagine, arginine
71	lipid fragment ($C_5H_{11}^+$)
72	lipid fragment ($C_4H_{10}N^+$), valine
74	lipid fragment ($C_4H_{12}N^+$) threonine
76	cysteine, glycine
77	phenylalanine
81	lipid fragment ($C_6H_9^+$), alanine
82	lipid fragment ($C_5H_8N^+$), histidine
83	lipid fragment ($C_6H_{11}^+$)
84	lipid fragment ($C_5H_{10}N^+$), lysine, glutamine, glutamic acid
85	lipid fragment ($C_6H_{13}^+$)
86	lipid fragment ($C_5H_{12}N^+$), isoleucine, leucine
87	asparagine
88	lipid fragment ($C_5H_{14}N^+$), aspartic acid
90	alanine
91	lipid fragment ($C_7H_7^+$), serine, phenylalanine, methionine
93	lipid fragment ($C_7H_9^+$)
95	lipid fragment ($C_7H_{11}^+$), histidine
97	lipid fragment ($C_7H_{13}^+$)
98	lipid fragment ($C_5H_8NO^+$), glycine
100	lipid fragment ($C_5H_{10}NO^+$)
101	glutamine
102	lipid fragment ($C_5H_{12}NO^+$), glutamic acid

103	phenylalanine
104	lipid fragment ($C_5H_{14}NO^+$), lysine
106	serine
107	tyrosine
110	histidine
111	lipid fragment ($C_8H_{15}^+$)
112	alanine
116	serine, tyrosine, threonine, proline
117	tryptophan, proline
118	valine
120	phenylalanine, methionine, threonine, glycine
122	cysteine
123	tyrosine
128	serine
130	glutamine, tryptophan, glutamic acid
132	isoleucine, leucine
133	asparagine
134	alanine, aspartic acid
136	tyrosine
138	proline
143	tryptophan
146	lipid fragment ($C_5H_9NPO_2^+$)
148	lipid fragment ($C_5H_{11}NPO_2^+$), glutamic acid
150	lipid fragment ($C_5H_{13}NPO_2^+$), methionine
155	asparagine, aspartic acid
156	histidine
159	tryptophan
165	tyrosine
166	lipid fragment ($C_5H_{13}NPO_3^+$), methionine, phenylalanine
168	lipid fragment ($C_5H_{15}NPO_3^+$)
175	arginine, glutamic acid
177	asparagine
178	glycine
179	alanine
182	lipid fragment ($C_5H_{13}NPO_4^+$), tyrosine
184	lipid fragment ($C_5H_{15}NPO_4^+$)
188	tryptophan
190	lipid fragment ($C_7H_{13}NPO_3^+$)
194	lipid fragment ($C_7H_{17}NPO_3^+$)
196	lipid fragment ($C_6H_{15}NPO_4^+$)
198	lipid fragment ($C_6H_{17}NPO_4^+$)
200	glycine
205	tryptophan
206	lipid fragment ($C_5H_{14}NPO_4Na^+$)
210	lipid fragment ($C_7H_{17}NPO_4^+$)
212	lipid fragment ($C_7H_{19}NPO_4^+$)
219	aspartic acid, cysteine
224	lipid fragment ($C_7H_{15}NPO_5^+$)
226	lipid fragment ($C_7H_{17}NPO_5^+$)
231	proline
235	valine

238	lipid fragment ($C_8H_{17}NPO_5^+$)
239	threonine
240	lipid fragment ($C_8H_{19}NPO_5^+$)
241	cysteine
246	lipid fragment ($C_8H_{18}NPO_4Na^+$ or $C_8H_{18}NPO_5Li^+$)
252	lipid fragment ($C_8H_{15}NPO_6^+$)
254	lipid fragment ($C_8H_{17}NPO_6^+$)
256	lipid fragment ($C_8H_{19}NPO_6^+$)
263	isoleucine, leucine
279	aspartic acid, glutamine
282	lipid fragment ($C_9H_{17}NPO_7^+$)

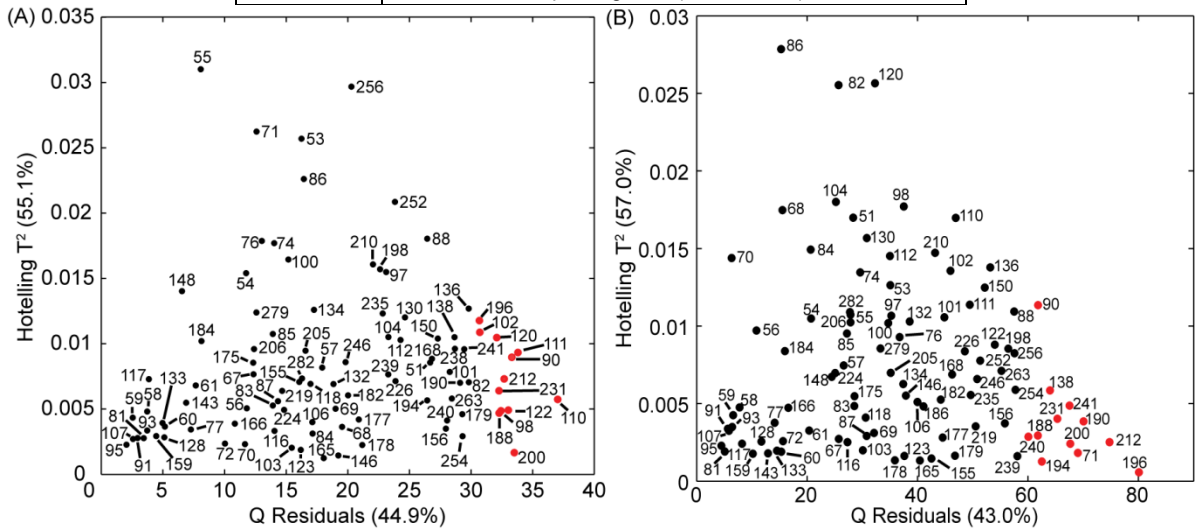

Figure S2. Plots show the contributions of the peaks in the cell-related peak set to the overall Q residual and Hotelling's T^2 statistic in the preliminary PLS-DA models. Peaks with high Q residual contributions and low Hotelling's T^2 contributions vary in a manner that is not related to HC differentiation status. (A) The preliminary PLS-DA model that was used to test the differentiation status of cells that were harvested from the same aged mice as the calibration set was refined by excluding the mass peaks with Q contributions above 30% (red). (B) The preliminary PLS-DA model that was used to test the differentiation status of cells that were harvested from young mice was refined by excluding the mass peaks with Q contributions above 60% (red).

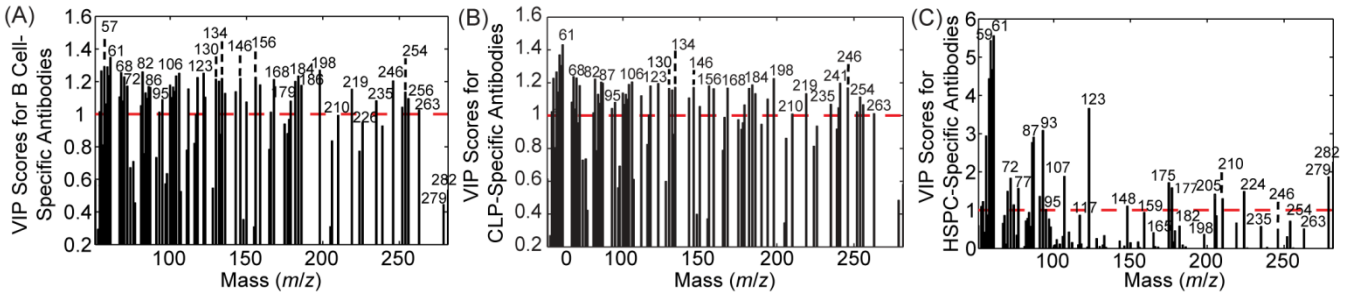

Figure S3. The VIP score plots show the importance of each mass peak towards the identification of the (A) B cell-specific antibodies, (B) CLP-specific antibodies, and (C) HSPC-specific antibodies that were used to isolate each cell population using flow cytometry.

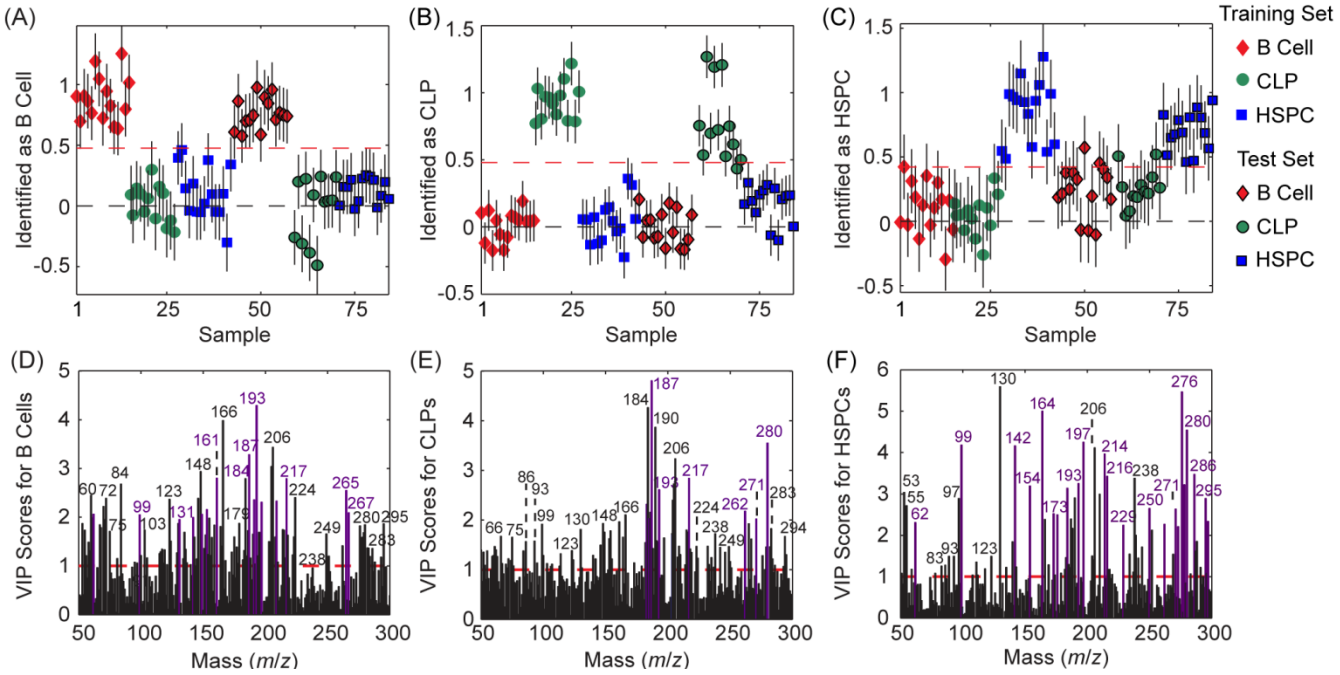

Figure S4. The (A) B cell, (B) CLP, and (C) HSPC identification plots for the PLS-DA model constructed using the complete peak set show the cells that exceed the threshold (red dashed line) for identification as the indicated cell type. The VIP plots for the (D) B cells, (E) CLPs, and (F) HSPCs show that many peaks from unknown sources were important towards identifying the indicated population (purple peaks).

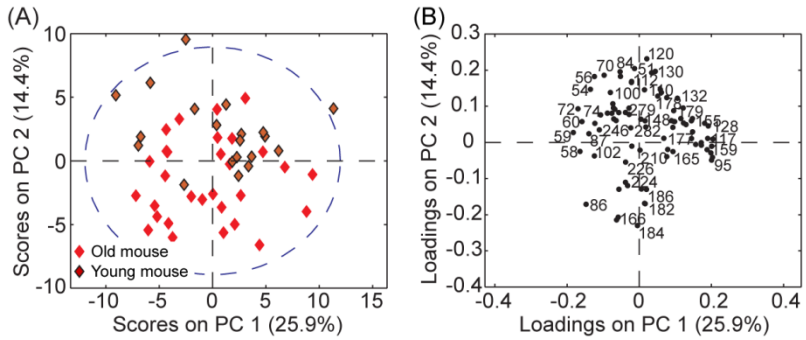

Figure S5. (A) PC score plot shows that PC1 captured the largest percentage of spectral variation in the B cell spectra, but did not separate the B cell spectra according to the age of the mice that the cells were harvested from. The elliptical region that is outlined with the dashed blue line represents the 95% confidence limit of the PC model. (B) The loadings plot shows the contribution of each peak to the variation captured by each PC.

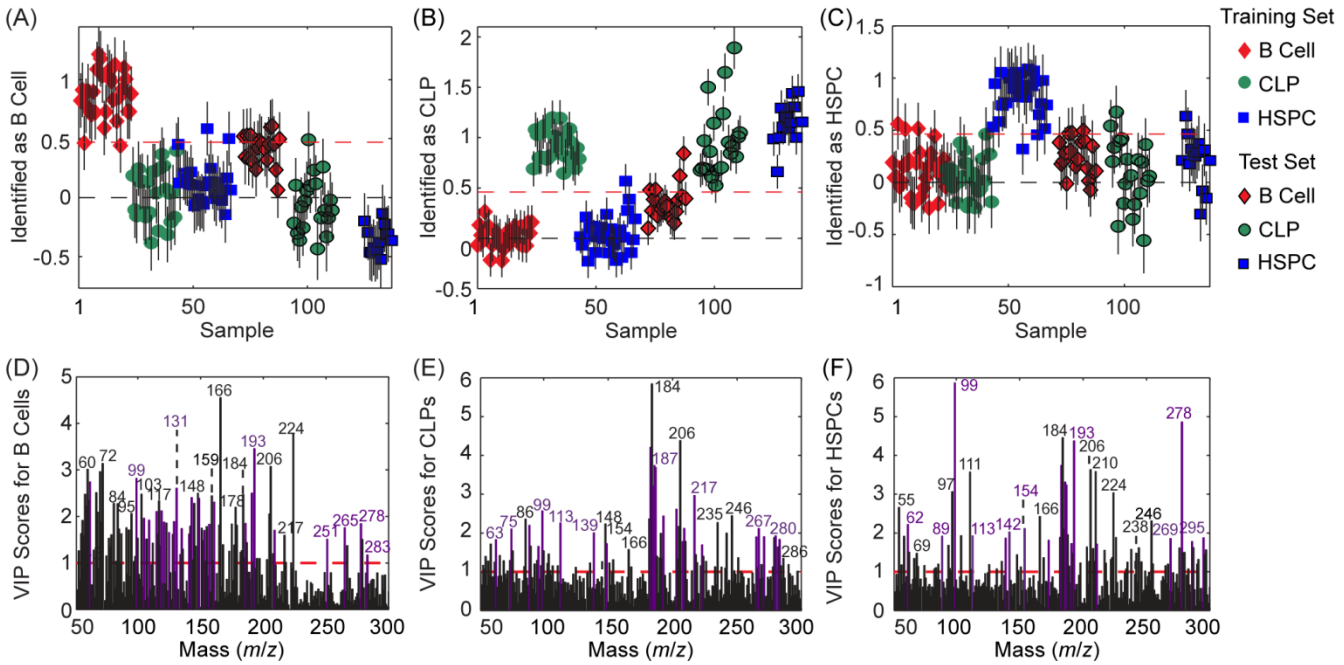

Figure S6. The (G) B cell, (H) CLP, and (I) HSPC identification plots for the PLS-DA model that was constructed using the complete peak set show that the inclusion of the peaks from unknown sources was detrimental to identifying each cell population. The VIP plots for the (J) B cells, (K) CLPs, and (L) HSPCs shows that many of the peaks that were important towards identifying the cells in each population were not included in the cell-related peak set (purple peaks). The parent molecules that produced these fragments are not known.

REFERENCES

(1) Jia, Y.; Loison, F.; Hattori, H.; Li, Y.; Erneux, C.; Park, S. Y.; Gao, C.; Chai, L.; Silberstein, L. E.; Schurmans, S.; Luo, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4739-44.
(2) Rumfelt, L. L.; Zhou, Y.; Rowley, B. M.; Shinton, S. A.; Hardy, R. R. *J. Exp. Med.* **2006**, *203*, 675-87.
(3) Challen, G. A.; Boles, N.; Lin, K. K.; Goodell, M. A. *Cytometry A* **2009**, *75*, 14-24.
(4) Yang, L.; Bryder, D.; Adolfsson, J.; Nygren, J.; Mansson, R.; Sigvardsson, M.; Jacobsen, S. E. *Blood* **2005**, *105*, 2717-23.
(5) Hardy, R. R.; Hayakawa, K. *Annu. Rev. Immunol.* **2001**, *19*, 595-621.
(6) Baker, M. J.; Brown, M. D.; Gazi, E.; Clarke, N. W.; Vickerman, J. C.; Lockyer, N. P. *Analyst* **2008**, *113*, 175-179.
(7) Graham, D. J.; Wagner, M. S.; Castner, D. G. *Appl. Surf. Sci.* **2006**, *252*, 6860-6868.
(8) Vaezian, B.; Anderton, C. R.; Kraft, M. L. *Anal. Chem.* **2010**, *82*, 10006-10014.
(9) Kulp, K. S.; Berman, E. S. F.; Kinze, M. G.; Shattuck, D. L.; Nelson, E. J.; Wu, L.; Montgomery, J. L.; Felton, J. S.; Wu, K. J. *Anal. Chem.* **2006**, *78*, 3651-3658.
(10) Anderton, C. R.; Vaezian, B.; Lou, K.; Frisz, J. F.; Kraft, M. L. *Surf. Interface Anal.* **2012**, *44*, 322-333.
(11) Oran, U.; Unveren, E.; Wirth, T.; Unger, W. E. S. *Appl. Sur. Sci.* **2004**, *227*, 318-324.
(12) Xia, N.; Castner, D. G. *J. Biomed. Mat. Res.* **2003**, *67A*, 179-190.
(13) Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, R. S., *PLS_Toolbox 4.0 for use with MATLAB*. Eigenvector Research, Inc.: Wenatchee, 2006.

(14) Norman, L.; Oetama, R.; Dembo, M.; Byfield, F.; Hammer, D.; Levitan, I.; Aranda-Espinoza, H. *Cell. Mol. Bioeng.* **2010**, *3*, 151-162.
(15) Hayes, T. L.; Lindgren, F. T.; Gofman, J. W. *J. Cell Biol.* **1963**, *19*, 251-255.