Undergraduate Honors Theses

2021-03-16

Lymphocyte Proteomics for Monitoring Long Term Immune System Dynamics

Michaela McCown

Honors Thesis

LYMPHOCYTE PROTEOMICS FOR MONITORING
LONG TERM IMMUNE SYSTEM DYNAMICS

by Michaela McCown

Advisor: Samuel H. Payne

Faculty Reader: Ryan T. Kelly

Honors Coordinator: Craig Coleman

ABSTRACT

LYMPHOCYTE PROTEOMICS FOR MONITORING
LONG TERM IMMUNE SYSTEM DYNAMICS

Michaela A. McCown

Department of Plant and Wildlife Sciences

Bachelor of Science

Time dependent change in lymphocytes provides a necessary background for monitoring the development of Chronic Lymphocytic Leukemia from asymptomatic to malignant. Several bulk proteomics studies characterize differences between tumor cells and healthy B cells, yet the transition to symptomatic disease is not well understood. This could be shown by a longitudinal study beginning with asymptomatic patients if the rare tumor cells in early stages can be isolated from small volume blood draws and effectively characterized with few cells, such as was done here using healthy B cells to provide a baseline for such studies to discriminate routine fluctuation from pathogenic changes. Lymphocyte proteomic data from two subjects and two collection days show that B lymphocytes change significantly with time. Most proteins were identified broadly in cell types, collection dates, and subjects. This study shows that the overall correlation between B lymphocytes from the same subject a month apart is nearly as low as the correlation between B and T lymphocytes, though time dependent changes to differentially expressed proteins are not as extreme as the differences between B and T

cells. These results show that lymphocyte subtypes can be collected and monitored by global proteomics to identify change over time. This could allow characterization of circulating tumor cells from patients with Chronic Lymphocytic Leukemia to identify variability in each stage and differential expression as the disease progresses.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Chronic Lymphocytic Leukemia (CLL) demonstrates cancer as a microevolutionary process where change over time is vital to treatment. Mutations in B cells lead first to a small circulating population of clonal B cells, then a growing population of circulating tumor cells, and after a long presymptomatic phase the cancer becomes malignant.[1] The disease is often discovered when a complete blood count assay reports abnormally high lymphocyte counts as part of a routine checkup. Following diagnosis, the disease is observed during a long asymptomatic period in which treatment has either shown no benefit or has yet to be demonstrated sufficiently for broad use, though several therapies are approaching the possibility of early treatment.[2] Once treatment is necessary, there are three common options—ibrutinib, acalabrutinib, or venetoclax and obinutuzumab—that manage the disease in spite of their side effects.[3] Resistance is often acquired to these treatments and, though more guided control may ameliorate the effects, additional biomarkers and novel treatments are needed.[4]

Many studies show characteristics of CLL tumor cells in contrast with healthy lymphocytes using global proteomics.[5–9] Proteomics can identify complex pathways, as opposed to isolated proteins, that are differentially expressed in diseased samples. This is particularly advantageous in the complex interactions in cancer, provided high accuracy and careful experimental design.[10] Recent proteomic research has made considerable progress understanding CLL. Effects of most established prognostic marker, IGHV mutation status, are explained using advancing global proteomics techniques.[6] Proteomic distinctions between B cells from elderly subjects and B cells

from young donors indicate that CLL cells exaggerate characteristics of aging.[8] Further analysis shows that a number of the misregulated pathways are targeted by existing therapeutics, which could be similarly beneficial in treating CLL.[7] Each study reveals more about the pathogenicity of CLL and leads towards more effective treatment options.[5]

While the differences between circulating CLL tumor and healthy B lymphocyte protein expression profiles are well characterized,[6,8] the changes that accompany the transition to malignance are not. Regulatory changes in disease give a greater level of understanding to the mechanism of disease and, by revealing those mechanisms, suggest avenues for potential treatment options. However, many studies are unable to show how much the expression changes throughout the disease or even in healthy tissues. For any disease, understanding changes across the natural history provides important insight for developing improved treatment options and identifying prognostic markers. CLL's asymptomatic stage provides an opportunity to collect samples prior to and as the tumor cells become malignant. Before those changes can be monitored, an analysis must be established that can characterize rare cell types like circulating tumor cells in early CLL with a minimally invasive sample collection, such as small volume blood draws, that will allow frequent sampling. Likewise, analysis of CLL's malignant changes requires an understanding of normal variability in B lymphocytes.

This study shows proteome changes in healthy individuals under normal conditions by comparing lymphocytes collected one month apart. This demonstrates an effective method of small sample analysis in select populations of blood cells. Samples

from two subjects were collected one month apart, sorted into B and T cells (Figure 1).

Samples were sorted into B and T cells provide the level of distinction between the two

lymphocyte subpopulations as context for the variability between time points. These

cells were analyzed by mass spectrometry and
compared to reveal proteome level differences.
Over one month, the B cell proteome changes
less than the amount that proteins differ
between B and T cells. This provides a
background to what level of change is aberrant
and what variation represents disease, and
more fundamentally shows the flexibility in
phenotype of immune cells in healthy
individuals.
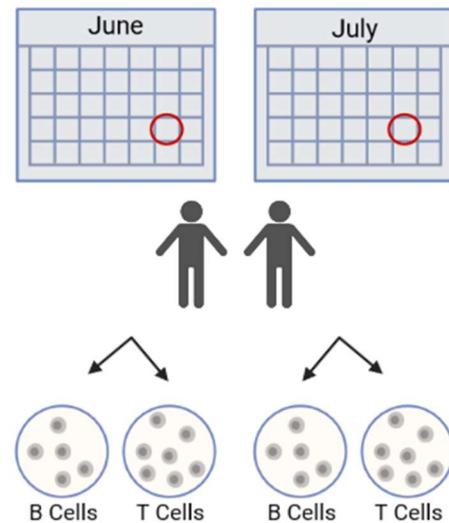
Figure 1. Experimental Design. As
shown, blood was drawn from each
of two subjects twice, one month
apart in June and July. Each sample
was processed identically and
sorted into 145 cell samples for
proteomic analysis by mass
spectrometry to show the variation
across a short period of time.

METHODS

Lymphocyte Collection

At each sampling time, whole blood samples were collected from the Student

Health Center (IRB Authorization X19045: Understanding the Progression of Chronic

Lymphocytic Leukemia). Samples were from two healthy adult male volunteers.

Samples were kept on ice for transport and prior to processing. These samples were

collected in June and July of 2020, thirty days apart. Each time, one 4ml EDTA tube of

whole blood was taken for lymphocyte isolation. Additional blood was sent to LabCorp for two clinical tests, a complete blood count and comprehensive metabolic panel.

White blood cells were isolated by lysing the red blood cells and pelleting the intact cells. Cells were processed in triplicate from 200 µl whole blood. Along with sample replicates, four single color controls were prepared for FACS sorting: cells only, APC-positive, FITC-positive, and PI-positive. 100 µl whole blood was used for each control. Propidium iodide (PI) marks dead or dying cells by staining exposed DNA, so dead cells were necessary for the PI-positive control and month-old refrigerated blood was used. Each sample was mixed with 4ml of red blood cell lysis buffer in 5 ml falcon tubes and incubated for 5 minutes at room temperature. Debris was separated from intact white blood cells by centrifugation at 250 g for 5 minutes and the supernatant was decanted.

Lymphocytes were marked using fluorescent immunostaining. T lymphocytes and B lymphocytes were marked using CD3 anti-human FITC and CD19 anti-human APC (BioLegend), respectively. Cells were resuspended in 100 µl of phosphate buffer saline (PBS) with 5 µl of each immunostain. 10 µl PI was added to samples and the appropriate control tube to allow for the exclusion of dead cells. Samples were incubated for 20 minutes in the dark at room temperature. Excess dye was removed by washing with 2 ml PBS, then removed by centrifuging at 250 g for 5 minutes and decanting the supernatant. The cells were resuspended in 300 µl, bringing the cell suspension to an appropriate concentration for FACS sorting. This procedure is available on protocols.io.[11] Cells were then FACS sorted to samples of 145 cells in 384 well plates that were prewashed with water, mobile phase A, 0.01% n-dodecyl β-D-

4

maltoside (DDM) in stages and prefilled with 4 µl to minimize cell damage from

sorting.

Protein Quantification by Mass Spectrometry

      Samples were analyzed using standard global proteomics for low input samples.

Protein was extracted prepared using automated digestion and nanoLC-MS/MS as

described by Liang et al.[12] Cells were sonicated for 5 minutes and centrifuged at 1500

rpm for 1 minute to collect. The OT-2 liquid handler (Opentrons, Brooklyn, NY)

carried out further sample processing as reported previously.[12] In brief, the liquid

handler added the following: 1 µl dithiothreitol, then 2 µl every 15 minutes during a 1

hour incubation at 70 °C, 1 µl iodoacetamide prior to a 30 minute incubation at 25°C, 1

µl lys-C prior to a 3 hour incubation at 37°C, 1 µl trypsin, 2 µl water every 3 hours

during a 12 hour incubation at 37°C, 1 µl water and 1 µl 5% formic acid prior to a 1

hour incubation at 25°C. At that time, samples were stored at -20°C till thawing for

nanoLC-MS/MS.

      Samples were run from each cell type for each subject in five replicates on an

Orbitrap Exploris 480 (Thermo Fisher) mass spectrometer using a Nanospray Flex ion

source at 2.0 kV. Runs used a maximum injection time of 250 ms and an exclusion

duration of 90s. For additional settings, see the methodological details explained in

reference 12. The raw files will be made available on a public server when submitted

for publication. Raw mass spectrometry files were processed using the FragPipe

interface with MSFragger[13] and IonQuant[14] using default settings. Match Between Runs

was not used. False discovery rates are 0.01. The database search was conducted with a UniProt FASTA file (Swiss-Prot, reviewed, and downloaded July 2020).
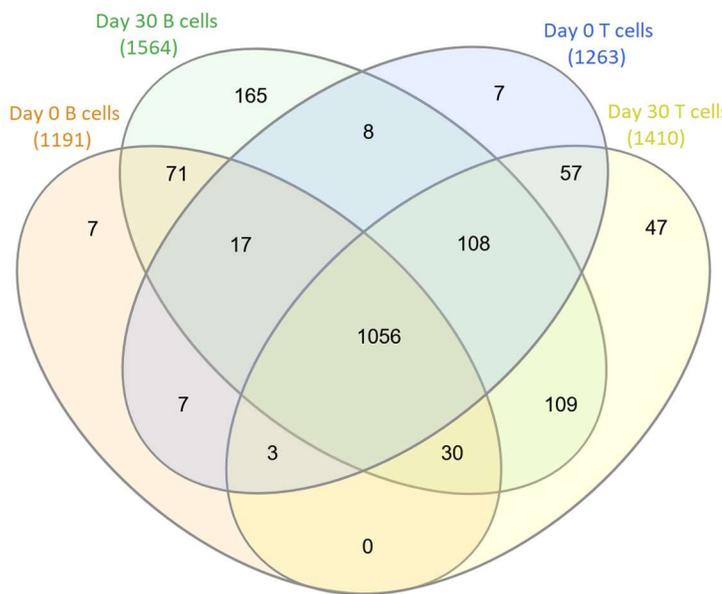
Relationship Analysis

Specific analyses were conducted in Python using Jupyter Notebooks. My code can be found on GitHub at github.com/PayneLab/Lymphocytes2 following a standard dissemination pattern.[15] Two additional tools were used for figure creation; Figure 1 was created with BioRender.com and Figure 2 using InteractiVenn.net with output from a Jupyter Notebook. The quantification tables were read in as Pandas dataframes, then normalized by log 2 transformation. Because batch effects will shift the signal intensity between runs, each run was adjusted by its median value to create a normalized dataset of relative intensities centered at zero. Identifications by sample are counted by the number of protein groups with a non-null value. Gene set enrichment analysis[16,17] was done using the cell type classifier ProteomicsDB[18] to verify these proteins as lymphocyte-representative expressions. Proteins were filtered by cell type for those that are present in at least half the samples. Spearman correlation coefficients were calculated to show reproducibility within a cell type. To better display the change between B lymphocytes collected at each time, the fold change was calculated for each protein and graphed on a log 2 scale. These figures were created using Seaborn and pyplot. Significance of relationships was determined by t-test for independent samples with Bonferroni correction.

RESULTS

Global proteomics of lymphocyte samples collected one month apart demonstrate time dependent change. Comparison of B and T lymphocytes provide context for the magnitude of that change and will allow long term analysis of variation during disease development. Proteomic data was analyzed from two subjects at each time point in five replicates for a total of forty datasets.

Identifications

FragPipe identifies a total of 2426 unique proteins from global proteomics analysis in these datasets. 1080-1607 proteins are identified in each sample. Confirming the effectiveness of this coverage, gene set enrichment analysis by tissue type shows that these proteins are highly similar to lymphoblastoid cells (Adjusted p value = 2.103470e-08). In analyzing reproducibility, proteins used are consistently identified in at least half the replicates. Of the proteins consistent in any group, most are shared across groups for a common set of 1056 proteins (Figure 2). The next largest regions are in samples from the

Figure 2. Core Identifications. The number of proteins identified in half the replicates from each group of samples.

second time point (165 were only in July B cells and 109 in both cell types in July), which has more identifications overall. By contrast with the central overlap, relatively few proteins are uniquely identified in any category. The proteins not identified may represent insufficient sample or lower instrument sensitivity as well as genuine absence. This indicates a large core set of proteins can be accurately compared for differential expression.

Reproducibility

Both cell types have high technical reproducibility as shown by an average Spearman correlation coefficient of 0.91 between sample replicates. This provides a baseline for variation between replicates. When comparing all samples of a cell type, combining both time points and subjects, reproducibility drops to 0.87 for B cells and 0.89 for T cells. Correlation is lowest between samples of different cell types at an average correlation of 0.85. This data distinguishes the subpopulations of B and T cells, while reflecting the similarity between the two types of lymphocytes.

Time Dependent Change

The Spearman correlation coefficients show the extent to which comprehensive patterns in protein expression persist across samples. B lymphocytes from the first sampling correlate with corresponding samples from the second date at a coefficient of 0.88. This is more variable than technical replicates (p=1.95e-4), and near the variability between B and T lymphocytes from the same sampling time (p=.06)

showing that protein expression changes in one month is to a similar extent as defining two cell types (Figure 3). More data is necessary to show whether the time comparison correlation is significantly higher than the cell type comparison. The range of correlation is broad, reaching lower than cross cell types and nearly as high as technical replicates.
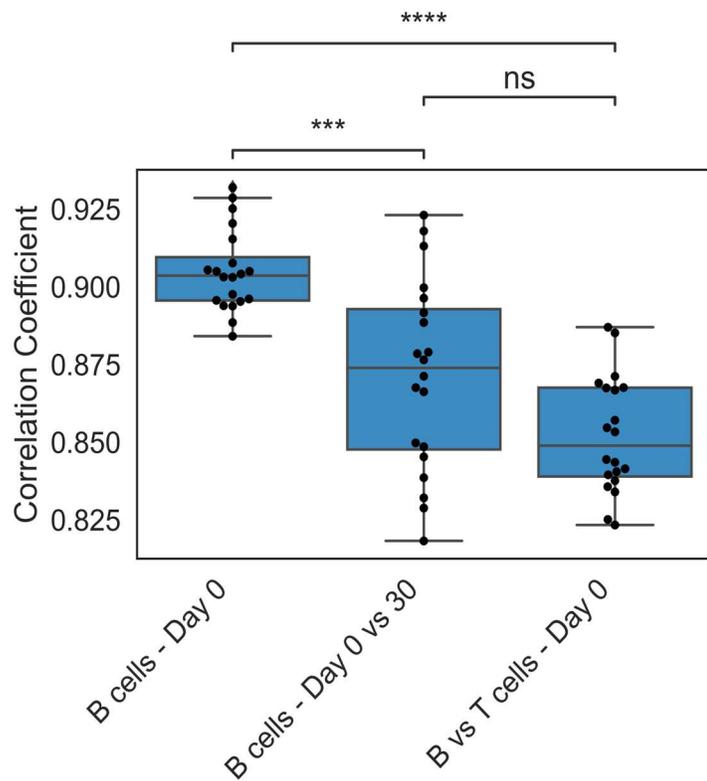
Figure 3. Magnitude of Time Dependent Change. Correlation coefficients are plotted for three key comparisons: technical replicates of B cells from day 0, B cells from day 0 compared to day 30, and B cells compared to T cells from the same time (day 0). Each subject's data was compared separately.

Specific Expression

Most proteins regardless of sample type show only moderate change, with a small differentiating set of proteins. The variable proteins across time are fewer and have milder changes than between cell types. As well as by overall correlation patterns, time dependent change is shown in terms of differentially expressed proteins. At $p < 0.05$, 67 proteins from B lymphocytes in Subject 1 change significantly in abundance between sampling times (Figure 4a). For comparison, 105 proteins are significantly

different in abundance using Subject 1's first sampling between B and T lymphocytes (Figure 4b). Both comparisons show that the majority of proteins are similarly expressed while a subset of proteins is not. The most extreme time dependent changes are lower in both magnitude and significance than the extreme changes between cell types (Figure 4).
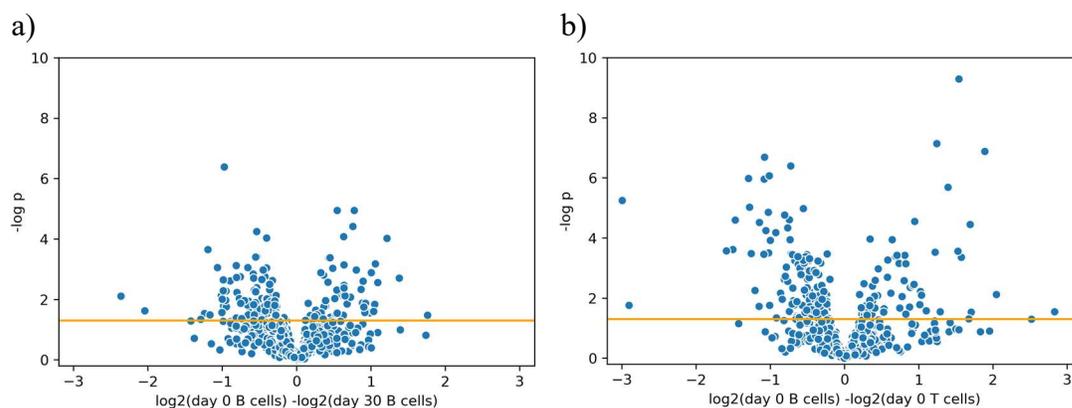
Figure 4. Differential Expression Profile. Variably expressed proteins from day 1 to 30 in Subject 1 B cells (a) and between B and T cells from Subject 1 at day 1 (b) are identified using fold change against P value.

DISCUSSION

I present a comparison of repeated sampling of global lymphocyte proteomics in the context of cell type variability. This characterization of short-term variation will aid in discriminating routine fluctuation from pathogenic changes and indicates that small sample analysis yields sufficient coverage depth for time dependent changes. A population across time points has variability near that of the different cell types. B lymphocytes show more slightly lower correlation and higher expression shifts than do T lymphocytes. Additional research is necessary to show to what extent the adaptation

10

influences expression changes, yet as B cells give rise to a variety of clonal populations it is unsurprising to see more variability.

At one month apart, these data show routine variation and seasonal effects are unlikely. Seasonal effects do change health metrics on blood and disease frequency.[19,20] Another example of seasonal change is shown in proteins, metabolites, and clinical tests on blood samples.[21] Blood pressure also varies across seasons, correlated with outdoor temperature.[22] This study shows variation from two points in the summer with similar weather. Additional data will show to what extent the patterns shown in higher level blood measurements are mirrored in lymphocyte proteomics. Long-term sampling and characterization of key immune cells will allow us to better understand immune system dynamics.

The ability to characterize such small samples is necessary for longitudinal CLL research. Small sample proteomics are rapidly advancing, yet lymphocytes are still an unusual application.[23] For common cell types or intact tissues, better results may be obtained from bulk analysis such as Mayer's 20 μg of protein extract identifying 6,945 proteins[8] or Johnston's 100 μg of cell lysate for 8694 proteins,[7] each of which would require millions of cells.[24] However, the reduction in number of proteins identified is justified by the diverse applications that will allow only restricted sample size due to collection of rare cell types. Limited blood draws do not yield millions of rare cells for microgram samples as required for bulk proteomics, but could give a few hundred. These few cells would be effectively characterized by small scale assays like this lymphocyte characterization.

In conclusion, time dependent change in lymphocytes is nearly as high as the variation between lymphocyte subpopulations. This baseline indicates changes that occur in healthy subjects' B lymphocytes over time. The methods demonstrated here effectively characterize the proteome from 145-cell samples sorted by surface antigens and will allow longitudinal analysis of even rare cell types. The baseline variability and the monitoring methods will allow analysis of lymphocyte population changes during the course of disease.

REFERENCES

(1)   D'Arena, G.; Musto, P. Monoclonal B-Cell Lymphocytosis. *Transl. Med. UniSa* **2014**, *8*, 75–79.

(2)   Muchtar, E.; Kay, N. E.; Parikh, S. A. Early Intervention in Asymptomatic Chronic Lymphocytic Leukemia. *Clin. Adv. Hematol. Oncol. HO* **2021**, *19* (2), 92–103.

(3)   Shadman, M.; Goodrich, A. Improving Outcomes for Patients With Chronic Lymphocytic Leukemia. *J. Adv. Pract. Oncol.* **2020**, *11* (3), 312–315. https://doi.org/10.6004/jadpro.2020.11.3.19.

(4)   Skånland, S. S.; Mato, A. R. Overcoming Resistance to Targeted Therapies in Chronic Lymphocytic Leukemia. *Blood Adv.* **2021**, *5* (1), 334–343. https://doi.org/10.1182/bloodadvances.2020003423.

(5)   Thurgood, L. A.; Chataway, T. K.; Lower, K. M.; Kuss, B. J. From Genome to Proteome: Looking beyond DNA and RNA in Chronic Lymphocytic Leukemia. *J. Proteomics* **2017**, *155*, 73–84. https://doi.org/10.1016/j.jprot.2017.01.001.

(6)   Thurgood, L. A.; Dwyer, E. S.; Lower, K. M.; Chataway, T. K.; Kuss, B. J. Altered Expression of Metabolic Pathways in CLL Detected by Unlabelled Quantitative Mass Spectrometry Analysis. *Br. J. Haematol.* **2019**, *185* (1), 65–78. https://doi.org/10.1111/bjh.15751.

(7)   Johnston, H. E.; Carter, M. J.; Larrayoz, M.; Clarke, J.; Garbis, S. D.; Oscier, D.; Strefford, J. C.; Steele, A. J.; Walewska, R.; Cragg, M. S. Proteomics Profiling of CLL Versus Healthy B-Cells Identifies Putative Therapeutic Targets and a Subtype-Independent Signature of Spliceosome Dysregulation. *Mol. Cell.*

Proteomics MCP **2018**, *17* (4), 776–791.

https://doi.org/10.1074/mcp.RA117.000539.

(8)   Mayer, R. L.; Schwarzmeier, J. D.; Gerner, M. C.; Bileck, A.; Mader, J. C.;

Meier-Menches, S. M.; Gerner, S. M.; Schmetterer, K. G.; Pukrop, T.; Reichle,

A.; Slany, A.; Gerner, C. Proteomics and Metabolomics Identify Molecular

Mechanisms of Aging Potentially Predisposing for Chronic Lymphocytic

Leukemia. *Mol. Cell. Proteomics MCP* **2018**, *17* (2), 290–303.

https://doi.org/10.1074/mcp.RA117.000425.

(9)   Alsagaby, S. A.; Alhumaydhi, F. A. Proteomics Insights into the Pathology and

Prognosis of Chronic Lymphocytic Leukemia. *Saudi Med. J.* **2019**, *40* (4), 317–

327. https://doi.org/10.15537/smj.2019.4.23598.

(10)   Bhawal, R.; Oberg, A. L.; Zhang, S.; Kohli, M. Challenges and Opportunities in

Clinical Applications of Blood-Based Proteomics in Cancer. *Cancers* **2020**, *12*

(9), 2428. https://doi.org/10.3390/cancers12092428.

(11)   McCown, M. A.; Tueller, J. A.; Weber, S.; Payne, S. H. Lymphocyte Collection

Protocol. **2021**. https://doi.org/10.17504/protocols.io.bnsmmec6.

(12)   Liang, Y.; Acor, H.; McCown, M. A.; Nwosu, A. J.; Boekweg, H.; Axtell, N. B.;

Truong, T.; Cong, Y.; Payne, S. H.; Kelly, R. T. Fully Automated Sample

Processing and Analysis Workflow for Low-Input Proteome Profiling. *Anal.*

Chem. **2021**, *93* (3), 1658–1666. https://doi.org/10.1021/acs.analchem.0c04240.

(13)   Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.;

Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide

Identification in Shotgun Proteomics. *Nat. Methods* **2017**, *14* (5), 513–520.

14

https://doi.org/10.1038/nmeth.4256.

(14)   Yu, F.; Haynes, S. E.; Teo, G. C.; Avtonomov, D. M.; Polasky, D. A.;

Nesvizhskii, A. I. Fast Quantitative Analysis of TimsTOF PASEF Data with

MSFragger and IonQuant. *Mol. Cell. Proteomics* **2020**, *19* (9), 1575–1585.

https://doi.org/10.1074/mcp.TIR120.002048.

(15)   Boekweg, H.; McCown, M. A.; Payne, S. H. Simple and Efficient Data Analysis

Dissemination for Individual Laboratories. *J. Proteome Res.* **2020**, *19* (10),

4191–4195. https://doi.org/10.1021/acs.jproteome.0c00454.

(16)   Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.;

Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.;

Mesirov, J. P. Gene Set Enrichment Analysis: A Knowledge-Based Approach for

Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. U. S. A.*

2005, *102* (43), 15545–15550. https://doi.org/10.1073/pnas.0506580102.

(17)   Kuleshov, M. V.; Jones, M. R.; Rouillard, A. D.; Fernandez, N. F.; Duan, Q.;

Wang, Z.; Koplev, S.; Jenkins, S. L.; Jagodnik, K. M.; Lachmann, A.;

McDermott, M. G.; Monteiro, C. D.; Gundersen, G. W.; Ma'ayan, A. Enrichr: A

Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update.

Nucleic Acids Res. **2016**, *44* (W1), W90-97. https://doi.org/10.1093/nar/gkw377.

(18)   Schmidt, T.; Samaras, P.; Frejno, M.; Gessulat, S.; Barnert, M.; Kienegger, H.;

Krcmar, H.; Schlegl, J.; Ehrlich, H.-C.; Aiche, S.; Kuster, B.; Wilhelm, M.

ProteomicsDB. *Nucleic Acids Res.* **2018**, *46* (D1), D1271–D1281.

https://doi.org/10.1093/nar/gkx1029.

(19)   Fares, A. Factors Influencing the Seasonal Patterns of Infectious Diseases. *Int. J.*

Prev. Med. **2013**, *4* (2), 128–132.

(20) Liu, B.; Taioli, E. Seasonal Variations of Complete Blood Count and

Inflammatory Biomarkers in the US Population - Analysis of NHANES Data.

PLoS ONE **2015**, *10* (11). https://doi.org/10.1371/journal.pone.0142382.

(21) Sailani, M. R.; Metwally, A. A.; Zhou, W.; Rose, S. M. S.-F.; Ahadi, S.;

Contrepois, K.; Mishra, T.; Zhang, M. J.; Kidziński, Ł.; Chu, T. J.; Snyder, M. P.

Deep Longitudinal Multiomics Profiling Reveals Two Biological Seasonal

Patterns in California. *Nat. Commun.* **2020**, *11* (1), 4933.

https://doi.org/10.1038/s41467-020-18758-1.

(22) Alpérovitch, A.; Lacombe, J.-M.; Hanon, O.; Dartigues, J.-F.; Ritchie, K.;

Ducimetière, P.; Tzourio, C. Relationship between Blood Pressure and Outdoor

Temperature in a Large Sample of Elderly Individuals: The Three-City Study.

Arch. Intern. Med. **2009**, *169* (1), 75–80.

https://doi.org/10.1001/archinternmed.2008.512.

(23) Kelly, R. T. Single-Cell Proteomics: Progress and Prospects. *Mol. Cell.*

Proteomics **2020**. https://doi.org/10.1074/mcp.R120.002234.

(24) Kuse, R.; Schuster, S.; Schübbe, H.; Dix, S.; Hausmann, K. Blood Lymphocyte

Volumes and Diameters in Patients with Chronic Lymphocytic Leukemia and

Normal Controls. *Blut* **1985**, *50* (4), 243–248.

https://doi.org/10.1007/BF00320301.