Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world-wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Brandon E. Shih Date

High-resolution metabolomics of diffuse large B-cell and follicular lymphomas: A pilot case-control study

By

Brandon E. Shih

Master of Science in Public Health

Department of Epidemiology

_____

Roberd M. Bostick, MD, MPH

Committee Chair

_____

Veronika Fedirko, PhD

Committee Member

High-resolution metabolomics of diffuse large B-cell and follicular lymphomas: A pilot case-control study

By

Brandon E. Shih

B.S., University of California, Los Angeles, 2014

Faculty Thesis Advisor

Roberd M. Bostick, MD, MPH

Veronika Fedirko, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Epidemiology

2021

Abstract

High-resolution metabolomics of diffuse large B-cell and follicular lymphomas: A pilot case-control study

By Brandon E. Shih

Background: Non-Hodgkin lymphoma (NHL) is a common cancer within the US, accounting for about 4% of all new cancer cases and 3% of all cancer deaths. The etiology of NHL is poorly understood with no useful biomarker to aid early diagnosis. To address this gap, we conducted a comprehensive metabolome-wide association study (MWAS) to identify metabolomic perturbations and pathways associated with NHL risk.

Methods: We used previously collected serum samples from a large population-based case-control study of incident NHL across six counties within the San Francisco Bay Area conducted between 2001 and 2005. For our pilot study, serum samples were available for 142 NHL cases frequency and matched to 142 controls by age and sex. Metabolome perturbations associated with the risk for two main NHL subtypes, diffuse large B-cell lymphoma (DLBCL, n = 75) and follicular lymphoma (FL, n = 67), were assessed with an untargeted HRM workflow using liquid chromatography-high resolution mass spectrometry with HILIC positive and C18 negative chromatography columns, followed by conditional logistic regression, pathway enrichment analysis, and chemical annotation.

Results: After HRM data processing and cleaning, 7,061 and 5,633 metabolic features were identified in serum samples with positive and negative ionization modes, respectively. Twenty-one metabolic pathways and ten metabolic pathways were found to be associated with DLBCL and FL risk, respectively. The only metabolic pathway associated with risk in both subtypes was tryptophan metabolism pathway. We confirmed thirteen metabolites within tryptophan pathways associated with either DLBCL or FL risk. Additionally, we confirmed nine metabolites within tyrosine metabolism associated with FL risk.

Discussion: The thirteen metabolites identified within tryptophan pathways suggest they might contribute to the suppression of anti-tumor immune responses, creating a favorable environment for tumor cells. The nine metabolites within tyrosine metabolism may also suggest a similar contribution to innate immune response and other important biological functions. Identification of metabolome perturbations and pathways associated with NHL risk from this study may contribute to the development of novel biomarkers for risk of NHL.

High-resolution metabolomics of diffuse large B-cell and follicular lymphomas: A pilot case-control study

By

Brandon E. Shih

B.S., University of California, Los Angeles, 2014

Faculty Thesis Advisor

Roberd M. Bostick, MD, MPH

Veronika Fedirko, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Department of Epidemiology

2021

Acknowledgements

Donghai Liang

Zeyuan Wang

Table of Contents

Chapter I: Background/Literature Review

Descriptive Epidemiology of NHL

Non-Hodgkin lymphoma (NHL) is a cancer arising from B-lymphocytes, T-lymphocytes, or natural killer cells of the immune system (1). The American Cancer Society estimates in 2021, there will be 81,560 new cases and 20,720 deaths from NHL in the US (2). However, these estimates vary by NHL subtype, age, sex, and ethnicity. The two most common types of NHL are diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) (3). In 2016, the US incidence rate of DLBCL was 5.6 per 100,000 persons per year while the incidence rate of FL was 2.7. Comparing the mortality rates, DLBCL had a higher mortality rate of 1.8 per 100,000 persons while FL had a morality rate of 0.5. Of the two subtypes, DLBCL is the more aggressive tumor with a relative 5-year survival of 64% compared to 5-year survival of 89% among FL patients (4,5). When comparing NHL estimates by age and sex, age-standardized rates in the UK showed men have consistently higher rates compared to women (6). As age increases, the divergence between sex-specific rates become more apparent as men have higher rates than women. Differences in age are also seen when examining specific NHL subtypes. For example, most patients with DLBCL have a median age at diagnosis over 70 years while the median age at diagnosis for FL is 65 years (6).

Molecular Basis of NHL

Each specific NHL subtype exhibit different gene expression profiles. Gene expression profiling of DLBCL reveal three molecular subgroups of DLBCL: germinal-center B-cell-like (GCB) DLBCL, activated B-cell-like (ABC) DLBCL, and primary mediastinal B-cell lymphoma (PMBCL) (7,8). While classified under DLBCL, each subgroup express different type of genes. GCB DLBCL derives from germinal-center B-cells and have been shown to express *BCL-6 or LMO-2* (7–9). However, these genes are shown to be downregulated in ABC DLBCL. ABC DLBCL derives from B cells in the process of differentiating into plasma cells and exhibit an upregulation in many of the genes normally expressed in

plasma cells, such as XBP-1 (7). The molecular pathways involved in the development of FL produces a different gene expression profile. About 90% of all FL cases express the anti-apoptotic BCL-2 protein but a small percentage of FL cases contain alterations leading to deregulated expression of the transcriptional repressor, BCL-6 (10) instead. Both of these proteins are linked to germinal center B cells (11). The subtle differences across NHL gene expression profiles make identification of subtypes important in determining appropriate treatment options and prognosis.

Treatment of NHL

The current standard treatment plan for DLBCL patients is a combination of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) chemotherapy (9,12). However, the addition of rituximab as maintenance after chemoimmunotherapy treatment was not beneficial (13). Treatments are available for NHL patients, but disparities among marginalized populations persist. The incidence of NHL is lower in minority groups but Black patients present with NHL at younger ages and with more advanced disease compared to White patients (14). Research also shows HIV-positive patients with NHL are presenting with more advanced disease, higher scores under prognostic models, and worse overall survival (OS) compared to HIV-negative patients (14–16). As immunotherapy has been used to treat different types of cancers by targeting PD-1 receptors, studies have also analyzed the association between PD-1/PD-L1 and NHL. One study found the potential prognostic impact PD-1/PD-L1 could have on NHL subtypes. For example, PD-L1+ DLBCL tumors had inferior OS rates than PD-L1- DLBC tumors (17). In comparison with FL subtypes, there have been contradicting evidence to determine the prognostic value of these receptors.

Analytic Epidemiology of NHL

Several epidemiologic studies have shown NHL is more common among immunocompromised patients (i.e. HIV patients) and can be caused by genetic mutations or infections. Because there are multiple NHL subtypes such as DLBCL and FL (18), the risk factors associated with each subtype varies

as well. For example, a pooled analysis of 4,667 DLBCL cases showed that factors most commonly associated with DLBCL included B-cell activating autoimmune disease (OR: 2.36, 95% CI: 1.80-3.09), hepatitis C virus (HCV) positivity (OR: 2.02, 95% CI: 1.47-2.76), first-degree family history of NHL (OR: 1.95, 95% CI: 1.54-2.47), and higher body mass index (BMI) of $\geq$ 30 kg/m^2 as a young adult (OR: 1.58, 95% CI: 1.12-2.23) (19). Several occupations were also identified as risk factors for NHL. Among women, being a field crop/vegetable farmer, a seamstress/embroider, and a hairdresser was associated with a higher DLBCL risk. Among men, working as a driver or material handling operator was associated with a higher risk for DLBCL (19,20). Both DLBCL and FL were shown to share several risk factors. A pooled analysis of 3,530 FL cases showed individuals with a first-degree family history of NHL (OR: 1.99, 95% CI: 1.55-2.54), higher BMI as a young adult (OR: 1.21, 95% CI: 1.09-1.35 per 5 kg/m^2 increase), and individuals who worked as a spray painter (OR: 2.66, 95% CI: 1.36-5.24) had a higher FL risk (19). Across sex, cigarette smoking was a stronger risk factor for FL among females than males (19). Aside from behavioral risk factors and patient demographics, observational studies have also found certain infections to be risk factors for NHL. Q fever, caused by the agent, *Coxiella burnetii*, has been known to be associated with B-cell NHL (21,22). Other known infections associated with NHL include HCV, Epstein-Barr virus (EBV), *Helicobacter pylori*, and human immuno-deficiency virus (HIV) (20,23).

There have also been studies analyzing the role of specific chemical exposures as risk factors for NHL but results have been inconsistent. One of these chemical exposures is trichloroethylene (TCE) and has been widely used as an industrial solvent to remove grease from metal. Because of their widespread use, occupation exposure to TCE and has been associated with higher risk of NHL (24). However, when analyzed by subtype, TCE was only found to be associated with FL and chronic lymphocytic leukemia (CLL) and not DLBCL (20,25). Similar associations of higher risk for NHL are found for another chlorinated solvent, carbon tetrachloride (26). Other occupational chemical exposures associated with a higher risk in NHL include chemicals found in herbicides and insecticides. A meta-analysis of six case-control studies, including the most recent Agricultural Health Study cohort, found glyphosate-based

herbicide exposure (i.e. Roundup and Ranger Pro) is associated with a higher risk of NHL (Summary RR: 1.41, 95% CI: 1.13-1.75) (20,27). Additionally, another meta-analysis of 44 cohort and case-control studies found a positive association between higher NHL risk and exposure to active ingredients and chemical groups in agricultural pesticides, including phenoxy herbicides (i.e. MCPA), carbamate insecticides, organophosphorus insecticides, and lindane, an organochlorine insecticide (28). Despite finding associations with all these categories of risk factors, the etiology of NHL has still yet to be determined.

Molecular Epidemiology of NHL

Because of the familial predisposition to lymphoma including NHL (29), multiple studies focused on germline mutations and used genome-wide association approach to identify potential genetic variants responsible for the development of NHL. Most genome-wide association studies (GWAS) have focused specifically on human leukocyte antigen (HLA) variants and found HLA-B (rs2523607) locus associated with DLBCL risk in European (OR: 1.32, 95% CI: 1.21-1.44) and Asian (OR: 3.05, 95% CI: 1.32-7.05) populations (30). When examining the association with FL risk, two GWAS studies in European populations found several HLA-DQB1 variants (rs10484561, rs7755224, and rs264712) associated with a higher risk of FL (30–32). Chronic inflammation also plays an important role when determining the etiology of NHL. One meta-analysis of 17 nested case-control studies reviewed inflammatory markers, such as interleukin 6 (IL-6), IL-10, and tumor necrosis factor α (TNF-α) and their association with NHL. The results of the meta-analysis showed higher blood concentrations of IL-6 (OR: 1.22, 95% CI: 0.97-1.54), IL-10 (OR: 1.24, 95% CI: 0.93-1.63), TNF-a (OR: 1.18, 95% CI: 1.04-1.34) and other inflammatory markers were associated with a higher risk of NHL (33).

Using metabolomics analysis can help identify potential biomarkers and understand biological and metabolic pathways (34) of NHL but not many studies have examined multiple -omics among different NHL subtypes, such as genomics for natural killer/T-cell lymphoma (35) or Burkitt lymphoma (36). Examples of high resolution metabolomic (HRM) approaches include identifying which metabolic

pathways were affected by traffic related air pollution (37) or metabolic pathways most affected in adolescents with elevated hepatic steatosis (38). These studies introduce the potential of using HRM to provide further insight into the exposure-disease etiology through molecular and genetic mechanisms within NHL. For example, a cross-sectional study of 175 individuals in Guangdong, China used metabolomic pathway identification to assess the association between occupational exposure to TCE with adverse health outcomes, including NHL and kidney and liver cancers (39).

To determine if individual metabolites or metabolic pathways can serve as early biomarkers of lymphoma, one study examining 66 cases and 96 controls found hypoxanthine and elaidic acid to be more abundant across all lymphoma subtypes but due to small sample sizes, but only obtained four partial least square-discriminant analysis (PLS-DA) models for DLBCL, chronic lymphocytic leukemia (CLL), multiple myeloma (MM), and Hodgkin lymphoma (HL) (40). Another study of 15 DLBCL cases and 34 controls found ascorbate and aldarate metabolism, glycine, serine and threonine metabolism, and aminoacyl-tRNA biosynthesis most impacted among NHL DLBCL cases (41). This study also identified 3-indolelactic acid, a product from tryptophan metabolism, which suggests higher levels of tryptophan among DLBCL patients. A third study from South Korea analyzed plasma samples from 5 cases and 6 healthy controls and found steroid hormone biosynthesis, ABC transporters, and arginine and proline metabolism but this study only examined 2 NHL and 3 MM cases (42). As evidenced by these studies, many of the current NHL metabolomic literature have insufficient population sizes and present different metabolic pathways associated with NHL risk. Using a HRM approach would be helpful to further understand the NHL disease-exposure pathogenesis and identify therapeutic targets for different NHL subtypes.

Chapter II: Manuscript

High-resolution metabolomics of diffuse large B-cell and follicular lymphomas: A pilot case-control study

Abstract

Background: Non-Hodgkin lymphoma (NHL) is a common cancer within the US, accounting for about 4% of all new cancer cases and 3% of all cancer deaths. The etiology of NHL is poorly understood with no useful biomarker to aid early diagnosis. To address this gap, we conducted a comprehensive metabolome-wide association study (MWAS) to identify metabolomic perturbations and pathways associated with NHL risk.

Methods: We used previously collected serum samples from a large population-based case-control study of incident NHL across six counties within the San Francisco Bay Area conducted between 2001 and 2005. For our pilot study, serum samples were available for 142 NHL cases frequency and matched to 142 controls by age and sex. Metabolome perturbations associated with the risk for two main NHL subtypes, diffuse large B-cell lymphoma (DLBCL, n = 75) and follicular lymphoma (FL, n = 67), were assessed with an untargeted HRM workflow using liquid chromatography-high resolution mass spectrometry with HILIC positive and C18 negative chromatography columns, followed by conditional logistic regression, pathway enrichment analysis, and chemical annotation.

Results: After HRM data processing and cleaning, 7,061 and 5,633 metabolic features were identified in serum samples with positive and negative ionization modes, respectively. Twenty-one metabolic pathways and ten metabolic pathways were found to be associated with DLBCL and FL risk, respectively. The only metabolic pathway associated with risk in both subtypes was tryptophan metabolism pathway. We confirmed thirteen metabolites within tryptophan pathways associated with either DLBCL or FL risk. Additionally, we confirmed nine metabolites within tyrosine metabolism associated with FL risk.

Discussion: The thirteen metabolites identified within tryptophan pathways suggest they might contribute to the suppression of anti-tumor immune responses, creating a favorable environment for tumor cells. The nine metabolites within tyrosine metabolism may also suggest a similar contribution to innate immune response and other important biological functions. Identification of metabolome perturbations and pathways associated with NHL risk from this study may contribute to the development of novel biomarkers for risk of NHL.

Introduction

In 2021, the American Cancer Society estimates there will be 81,560 new cases and 20,720 deaths from non-Hodgkin lymphoma (NHL) in the US (2). NHL is a lymphoid disorder arising from the cells of the immune system with varied patterns of clinical behavior and treatment response (1). The two most common types of NHL are diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) (3). The 2016 US incidence rate of DLBCL was 5.6 per 100,000 persons per year while the incidence rate of FL was 2.7. However, NHL rates have been steadily increasing, particularly in persons older than 55, over recent years without consensus as to the underlying cause (43). To explain the rising rates of NHL, several epidemiologic studies have shown NHL can be caused by genetic mutations or infections and are more common among immunocompromised patients. Other epidemiologic studies have hypothesized the role of industrialization and increased exposure to chemicals contributing to the rise of NHL rates, but the results have been inconsistent. For example, occupational exposure to trichloroethylene (TCE) is known to be associated with higher risk of NHL (24), however, evidence has been limited and more research is needed.

Using metabolomics analysis can identify potential biomarkers and help understand biological and metabolic pathways involved in development and progression of NHL (34). However, there are very few studies utilizing -omics approaches, with most of them having small sample sizes and including different subtypes of lymphoma (35,36). One pilot metabolomics study using gas chromatography-mass spectrometry (GC-MS) found hypoxanthine and elaidic acid to be more abundant across all lymphoma subtypes compared to controls. Fatty acids were found to be increased among multiple myeloma and Hodgkin lymphoma patients while 2-aminoadipic acid and 2-aminoheptanedioic acid, erythritol and threitol were increased among DLBCL and chronic lymphocytic leukemia (40). However, sample size in this study was very small with only thirteen DLBCL and eight FL cases. Another small untargeted metabolomics study (DLBCL = 15, controls = 34) using urine samples found ascorbate and aldarate metabolism, glycine, serine and threonine metabolism, and aminoacyl-tRNA biosynthesis most impacted among DLBCL cases (41). Both of these metabolomic studies had small sample sizes of mostly DLBCL

cases and found different metabolomic pathways associated with DLBCL risk, demonstrating the need for a larger metabolomics study and analysis of more than one NHL subtype. Using a high-resolution metabolomics (HRM) approach would be helpful to further understand the NHL disease-exposure pathogenesis and identify biomarkers of risk and therapeutic targets for different NHL subtypes. Our study was designed to conduct a comprehensive metabolomic pathway analysis and chemical identification to find metabolomic features and pathways associated with risk for two subtypes of NHL – DLBCL and FL. Using previously collected data and serum samples from a NHL San Francisco Bay Area Case-Control study (44), we aimed to compare metabolomic profiles of DLBCL and FL subtypes with healthy controls. Identifying any associated metabolites or metabolic pathways linked to an overall higher risk of DLBCL or FL could help understand which endogenous metabolic pathway(s) are perturbed by environmental chemicals or other exposures.

Methods

Parent NHL Case-Control Study Design

The design of the parent case-control study was previously published (44). In summary, cases were identified within approximately one month of diagnosis by the Northern California Cancer Center's rapid case ascertainment. Eligible patients were diagnosed with incident NHL from October 2001 to October 2005. All eligible patients were 20-84 years old and residents of one of the six Bay Area counties at time of diagnosis. There were 1,704 eligible cases interviewed during the study period. An additional 452 cases died prior to initial contact, 161 could not be located/moved, 155 were too ill, and 360 refused to participate. Pathology reports and diagnostic materials were obtained from the diagnosing pathology department and re-reviewed by an expert pathologist to confirm the diagnosis and provide a consistent classification of NHL subtypes using the REAL/WHO (45) lymphoma classification. Control participants were identified by random digit dial supplemented by random sampling of Centers of Medicare & Medicaid Services list for those older than 65 years old. Eligibility criteria for controls were identical to cases with the exception of an NHL diagnosis. Control participants were frequency matched to cases by

age within five years, sex, and county of residence. Among 1,939 eligible random-digit dialed controls, 1,313 completed in-person interviews, 85 were too sick to participate, and 541 refused to participate. Among the 1,475 eligible Medicare controls, 768 completed in-person interviews, 97 were too ill to participate, and 610 refused to participate.

Demographic and risk factor data including HIV status, body mass index (BMI), county location at diagnosis, self-reported race and ethnicity, and paternal and maternal ancestry were collected from all study participants during their in-person interviews. BMI was classified according to the National Institute of Health Guidelines. No proxy interviews were conducted. The median time from diagnosis to venipuncture was 6.5 months. Blood and/or buccal cell specimens were collected for eligible participants who agreed to biospecimen collection (87% of cases and 88% of controls). All study participants were approved by the review committee for human research at the University of California, San Francisco.

Pilot High-Resolution Metabolomics Study Design

We conducted a pilot high-resolution metabolomics (HRM) study on a randomly selected subset of NHL cases and matched controls that satisfied the following inclusion criteria: negative HIV status, available plasma sample not previously thawed, non-Hispanic White, and one of two main NHL subtypes (diffuse large B-cell lymphoma, DLBCL or follicular lymphoma, FL). Because the case-control study predominately included White study participants, we did not have sufficient power to include other race and ethnicities. The final metabolomics study included 75 DLBCL cases and 75 controls and 67 FL cases and 67 controls for a total of 142 cases and 142 controls. All controls were frequency matched to cases by age within five years, race/ethnicity, sex, and county of residence. All DLBCL and FL cases and control plasma samples were similarly collected, processed, and stored in -80°C until metabolomic analysis.

HRM Analysis and Data Processing

HRM untargeted profiling was conducted using a liquid chromatography with high-resolution mass spectrometry (LC-HRMS) techniques (Thermo Scientific™ Q-Exactive™ HF) at the HERCULES

High Resolution Metabolomics Core. To enhance the coverage of metabolic feature detection, each sample was analyzed in triplicates and in polar and non-polar analytical columns and analysis modes: hydrophilic interaction liquid chromatography (HILIC) with positivity electrospray ionization (ESI) and C18 hydrophobic reversed-phase chromatography with negative ESI. Samples were randomized into batches of twenty, consisting of ten cases and ten matched controls. For each sample, 65 mL of plasma was treated with acetonitrile containing a mixture of 14 stable isotope internal standards added to the aliquot at 2:1 to precipitate proteins. Samples were then analyzed with NIST 1950-calibrated reference pooled human plasma preceding and following each block according to standard operating procedures. Raw data was extracted and a quality control was performed using apLCMS (46) and xMSanalyzer (47) software packages. The processed data included a total of 15,321 metabolic features in HILIC+ and 11,188 metabolic features in C18- modes. Data were further filtered to remove metabolic features not detected in $\geq$50% in any of the groups (DLBCL, FL, or controls) and with median CV $\geq$ 30%, leaving 7,061 metabolic features remaining for HILIC+ and 5,633 metabolic features for C18-. The resulting analytical data contained individual metabolic features defined by mass-to-charge ratio (m/z), retention time (RT), and ion intensities. Technical replicates were averaged. Zero averaged values were replaced with the minimum value divided by 2. All data were log2-transformed. Batch correction was conducted using ComBat package in R (48). A flowchart of data cleaning and processing is shown in **Figure 1**.

Statistical Analysis

Conditional logistic regression model was used to estimate the strength of the association between each metabolic feature and DLBCL or FL risk, matching on age (continuous) and sex (men/women) and additionally adjusting for BMI (underweight, normal weight, overweight, obese class 1, and obese class 2). *P*-values were corrected using Benjamini-Hochberg (49) (FDR) method to account for multiple testing.

Metabolic Pathways and Metabolite Identification

The metabolic features identified to be statistically significantly (raw *p*-value < *0.05*) associated with risk for DLBCL and FL were used to conduct the pathway enrichment and metabolite annotation analyses. Metabolic pathway identification was performed using Mummichog (version 1.0.9) (50). Positive and negative ionization modes for both DLBCL and FL subtypes were analyzed separately using commonly used adducts (isotopic derivatives) to produce a collection of predicted metabolic pathways. For HILIC+, only the following adducts were considered: M[+], M + H[+], M + Na[+], M(C13) + H[1+] (in lieu of M + K[+]), M + 2H[2+], and M(C13) + 2H[2+]. For C18-, the following adducts were considered: M-H[−], M + Cl[−], M + ACN-H[−], M + HCOO[−], M(C13)-H[−], M-H2O-H[−], and M + Na-2H[−]). Once metabolic pathways were identified, they were examined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) to determine the associations between specific chemicals or pathways.

To supplement our pathway analysis, individual extracted ion-chromatograms (EIC) plots were generated and scored on a scale from 0 to 3 to identify plots with clean and prominent peaks. Those with clean and prominent peaks were compared to the Human Metabolome Database (HMDB) to identify confirmed and validated putative chemical identities. Additionally, metabolic features were assigned potential chemical identities using xMSAnnotator (51) package in R, and compared to an in-house list of confirmed and validated metabolites from the HERCULES High Resolution Metabolomics Core.

Results

Baseline Characteristics of Study Participants

The baseline characteristics of cases and controls are summarized in **Table 1**. The mean age of DLBCL cases and controls was 65 years and the mean age of FL cases and controls was 62 years. There were 39 (52%) DLBCL cases and 37 (55%) FL cases that were men. All 284 participants were White, non-Hispanic, with negative HIV status. Across both DLBCL and FL subtypes, most participants BMI profiles were of normal weight and overweight.

Metabolic Features Associated with NHL Subtypes Risk

A summary of data extraction and processes for HRM analysis is presented in **Figure 1**. After data quality filtering, 7,061 and 5,633 metabolic features in HILIC+ and C18- ionization modes, respectively, remained for further analysis. The strength of association between each metabolic feature and risk of NHL subtype was estimated using conditional logistic regression. With C18- ionization, we detected 574 statistically significant (raw p-value < 0.05) metabolic features for DLBCL and 446 statistically significant metabolic features for FL. With HILIC+ ionization, we detected 552 statistically significant metabolic features for DLBCL and 565 statistically significant metabolic features for FL. Manhattan plots were used to further illustrate the association of metabolic features with risk for DLBCL and FL (**Figures 2a-d**). Each set of significant features was used to determine metabolomic pathways using Mummichog and the resulting pathways associated with each subtype are listed in **Tables 2a-b**.

Metabolic Pathways Associated with NHL Subtypes Risk

Table 2a and **Table 2b** is divided by NHL subtype and includes metabolic pathways arranged alphabetically. For DLBCL, the most significant pathway was starch and sucrose for HILIC+ and pentose phosphate pathway for C18- ionization modes. The most significant pathway for FL was methionine and cysteine for HILIC+ and vitamin B6 (pyridoxine) metabolism for C18- ionization modes. The following pathways were identified with HILIC+ and C18- ionization modes associated with DLBCL risk: caffeine metabolism, hexose phosphorylation, N-Glycan biosynthesis, propanoate metabolism, sialic acid metabolism, and starch and sucrose metabolism. The only pathway overlapping between DLBCL and FL subtypes was tryptophan pathway. Given the role tryptophan pathway plays in inflammation, we focused on identifying putative chemicals from tryptophan pathway (**Tables 3-4**). In addition to tryptophan, we also identified putative chemicals from tyrosine pathway found to be associated with FL risk (**Table 5**).

From metabolic features found within tryptophan pathway that were statistically significantly associated with DLBCL risk, we were able to identify the following chemicals using the potential matches in the HMDB database: D-Galactose, Acetoacetyl-CoA, L-Formylkynurenine, Indole-3-

Acetaldehyde, Quinolinate, Anthranilate, and 5-Methoxyindoleacetate. Within the same pathway found to be associated with FL risk, we identified the following chemicals: L-Kynurenine, 3-Hydroxyanthranilate, 5-Phenyl-1,3-Oxazinane-2,4-Dione, 3-Hydroxykynurenamine, N-Methyltryptamine, and Indole-3-Acetamide. Among tyrosine pathway associated with FL risk, we identified 5-S-Cysteinyldopa, Caffeate, Choline Phosphate, L-Cysteine, alpha-N-Phenylacetyl-L-Glutamine, 2-Aminoacrylate, Tyramine O-Sulfate, 7,8-Dihydrobiopterin, and Salsolinol.

To supplement our pathway analysis, individual extracted ion-chromatograms (EIC) plots were generated and compared to known databases to identify confirmed and validated metabolites. Eight plots were found for metabolic features within the tryptophan pathway in DLBCL (**Figure 3**), six plots were found for metabolic features within the tryptophan pathway in FL (**Figure 4**), and two plots were found for metabolic features within the tyrosine pathway (**Figure 5**).

Discussion

In this analysis, we used HRM profiling to identify 571 significant metabolic features for DLBCL C18-, 446 significant metabolic features for FL C18-, 552 significant metabolic features for DLBCL HILIC+, and 565 significant metabolic features for FL HILIC. These metabolic features were used for pathway identification, and 21 metabolic pathways were associated with DLBCL risk and ten metabolic pathways were associated with FL risk. The only pathway found to be associated with both DLBCL and FL risk was tryptophan metabolism pathway.

Tryptophan is one of nine essential amino acids and acts as a precursor to kynurenine, which plays a crucial role in the regulation of immune response during infections, inflammation, and pregnancy (52). Tryptophan catabolism has also been an increasingly important factor in tumorigenesis. Produced by tumor or tumor-associated cells, indoleamine 2,3-dioxygenas 1 (IDO1) catabolizes tryptophan to suppress antitumor immune response, creating a more favorable environment for tumor cells to escape host immunity (53,54). One previous case-control study of 15 DLBCL cases and 34 controls found an upregulation of 3-indolelactic acid among DLBCL cases. Since 3-indolelactic acid is a byproduct of

tryptophan metabolism, this suggested an increase in tryptophan metabolism among DLBCL (41). While we did not find similar upregulation of 3-indolelactic acid, these results corroborate our findings of tryptophan metabolism perturbations among DLBCL cases. We also found metabolic features in tryptophan metabolism to be associated with FL risk. Additionally, we found kynurenine and its derivates to be associated with risk for DLBCL and FL. The increased levels of tryptophan metabolism and kynurenine from our samples suggest the possibility of using tryptophan metabolism as a potential biomarker and therapeutic target for DLBCL and FL cases. However, further research on the downstream effects of immunosuppressive tryptophan metabolism is required.

In addition to tryptophan metabolism, we also found metabolic features in tyrosine metabolism pathway to be associated with FL risk. While no studies have examined the use of tyrosine as a diagnostic marker for NHL, one previous study assessed the usefulness of tyrosine as a metabolic marker for colorectal cancer. The study found lower levels of tyrosine may be attributed to the metabolic disturbance resulting from colorectal cancer development and a panel with glutamine-leucine and tyrosine could be used as a biomarker for early detection of colorectal cancer (55). While our results showed an association between tryptophan and tyrosine metabolism with DLBCL and FL, other metabolomic studies found different associations such as adenine and guanine metabolism across various subtypes of lymphoma (40). Differences in our results suggest further research is required to elucidate the associations.

Animal studies have also explored the importance of taurine during chemotherapy to treat T-cell lymphoma. Taurine was one of the top metabolites associated with DLBCL risk in our study (OR: 0.12, 95% CI: 0.04-0.37) and is known as an amino acid with positive effects on digestion, endocrine, and immune systems. One previously published study found taurine played a role in enhancing adjuvant therapy by reducing toxicity effects from chemotherapy (56). Another study used nuclear magnetic resonance-based metabolic profiling to identify metabolites to differentiate between DLBCL and FL and found the most relevant metabolites were alanine in DLBCL cases and taurine in FL cases (57). While our results also found taurine in DLBCL, findings from our study and the NMR study suggest taurine may be an important metabolite associated with DLBCL and FL risk.

Strengths and Weaknesses

One of the major strengths from our study was being the largest study to utilize untargeted HRM to analyze metabolomic profiles of DLBCL and FL. By using an untargeted approach, we were able to examine all metabolic features and metabolic pathways associated with NHL risk without a priori hypotheses. We also performed our study within a homogenous study population by restricting participants to non-Hispanic White and HIV negative participants. By using a more well-defined population, we were able to reduce the impact of heterogeneity across participants that might have contributed to metabolite differences. However, this has also prevented us from generalizing our results to a broader NHL population. Additionally, another limitation in our analysis was in our study design. By performing a case-control study, plasma samples were collected after NHL diagnosis. Detected metabolomic pathways or metabolites in our study could have been present prior to NHL diagnosis or caused by NHL but we were unable to establish temporality which limits our ability to determine if they can be used as early predictors of NHL diagnosis. We were also unable to determine if metabolites changed across lymphoma stages or treatments because information on lymphoma disease stage or current treatment regimens was not available.

Despite these limitations, our study found hundreds of significant metabolic features across two NHL subtypes and identified 21 metabolic pathways associated with DLBCL risk and ten metabolic pathways associated with FL risk. All metabolic pathways were important because we do not know the specific drivers behind the rising NHL rates over recent years. Previous studies have shown the importance of tryptophan metabolism, tyrosine metabolism, and taurine, but more research on these pathways is warranted. If their potential applications as diagnostic or therapeutic biomarkers in other scenarios can be applied to NHL risk, we may have a better understanding of increasing NHL rates.

Funding

This research was funded by the Emory Health and Exposome Research Center: Understanding Lifetime Exposures (HERCULES), a P30 Core Center Grant from the National Institute of Environmental Health Sciences (P30 ES019776).

References

1. Bowzyk Al-Naeeb A, Ajithkumar T, Behan S, et al. Non-Hodgkin lymphoma. *BMJ* [electronic article]. 2018;362(August):1–7. (http://dx.doi.org/doi:10.1136/bmj.k3204)
2. Key Statistics for Non-Hodgkin Kymphoma. *Am. Cancer Soc.* (https://www.cancer.org/cancer/non-hodgkin-lymphoma/about/key-statistics.html). (Accessed September 7, 2020)
3. Cancer Stat Facts: Non-Hodgkin Lymphoma. *Natl. Cancer Inst. Surveillance, Epidemiol. End Results Progr.* (https://seer.cancer.gov/statfacts/html/nhl.html). (Accessed September 7, 2020)
4. Cancer Stat Facts: NHL – Diffuse Large B-Cell Lymphoma (DLBCL). *Natl. Cancer Inst. Surveillance, Epidemiol. End Results Progr.* (https://seer.cancer.gov/statfacts/html/dlbcl.html). (Accessed September 7, 2020)
5. Cancer Stat Facts: NHL - Follicular Lymphoma. *Natl. Cancer Inst. Surveillance, Epidemiol. End Results Progr.* (https://seer.cancer.gov/statfacts/html/follicular.html). (Accessed September 7, 2020)
6. Roman E, Smith AG. Epidemiology of lymphomas. *Histopathology*. 2011;58:4–14.
7. Frick M, Dorken B, Lenz G. The molecular biology of diffuse large B-cell lymphoma. *Ther. Adv. Hematol.* 2011;2(6):369–379.
8. Pasqualucci L. The genetic basis of diffuse large B cell lymphoma. *Curr. Opin. Hematol.* 2013;20(4):336–344.
9. Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma - treatment approaches in the molecular era. *Nat. Rev. Clin. Oncol.* 2014;11:12–23.
10. Bende R, Smit L, van Noesel C. Molecular pathways in follicular lymphoma. *Leukemia*. 2007;21:18–29.
11. Mahmoud HM, El-Sakhawy YN. Significance of BCL-2 and BCL-6 immunostaining in B-non Hodgkin's lymphoma. *Hematol. Rep.* 2011;3:e26.
12. Ansell SM. Non-Hodgkin Lymphoma: Diagnosis and Treatment. *Mayo Clin. Proc.* [electronic

article]. 2015;90(8):1152–1163. (http://dx.doi.org/10.1016/j.mayocp.2015.04.025)

13. Habermann TM, Weller EA, Morrison VA, et al. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J. Clin. Oncol.* 2006;24(19):3121–3127.

14. Becnel M, Flowers CR, Nastoupil LJ. Disparities in lymphoma on the basis of race, gender, HIV status, and sexual orientation. *Ann. Lymphoma*. 2017;1:1–8.

15. Chen Q, Ayer T, Nastoupil LJ, et al. Population-specific prognostic models are needed to stratify outcomes for African-Americans with diffuse large B-cell lymphoma. *Leuk. Lymphoma*. 2016;57(4):842–851.

16. Re A, Cattaneo C, Rossi G. HIV and Lymphoma: From Epidemiology to Clinical Management. *Mediterr. J. Hematol. Infect. Dis.* 2019;11:1–17.

17. Gravelle P, Burroni B, Péricart S, et al. Mechanisms of PD-1/PD-L1 expression and prognostic relevance in non-Hodgkin lymphoma: A summary of immunohistochemical studies. *Oncotarget*. 2017;8(27):44960–44975.

18. Lymphoma - Non-Hodgkin: Subtypes. *Am. Soc. Clin. Oncol.* 2019;(https://www.cancer.net/cancer-types/lymphoma-non-hodgkin/subtypes)

19. Chihara D, Nastoupil LJ, Williams JN, et al. New insights into the epidemiology of non-Hodgkin lymphoma and implications for therapy. *Expert Rev. Anticancer Ther.* 2015;15(5):531–544.

20. Moubadder L, McCullough LE, Flowers CR, et al. Linking Environmental Exposures to Molecular Pathogenesis in Non-Hodgkin Lymphoma Subtypes. *Cancer Epidemiol. Biomarkers Prev.* 2020;29(10):1844–1855.

21. Melenotte C, Mezouar S, Amara A Ben, et al. A transcriptional signature associated with non-Hodgkin lymphoma in the blood of patients with Q fever. *PLoS One*. 2019;14(6).

22. Melenotte C, Million M, Audoly G, et al. B-cell non-Hodgkin lymphoma linked to Coxiella burnetii. *Blood*. 2016;127:113–121.

23. Parkin D. Cancers attributable to infection in the UK in 2010. *Br. J. Cancer* [electronic article].

2011;105:S49–S56. (http://dx.doi.org/10.1038/bjc.2011.484)

24. Purdue MP, Bakke B, Stewart P, et al. A case-control study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma. *Environ. Health Perspect.* 2011;119:232–238.

25. Cocco P, Vermeulen R, Flore V, et al. Occupational exposure to trichloroethylene and risk of non-Hodgkin lymphoma and its major subtypes: a pooled InterLymph analysis. *Occup. Environ. Med.* 2013;70:795–802.

26. Callahan CL, Stewart PA, Friesen MC, et al. Case-control investigation of occupational exposure to chlorinated solvents and non-Hodgkin's lymphoma. *Occup. Environ. Med.* 2018;75:415–420.

27. Zhang L, Rana I, Shaffer RM, et al. Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: A meta-analysis and supporting evidence. *Mutat. Res. - Rev. Mutat. Res.* [electronic article]. 2019;781:186–206. (https://doi.org/10.1016/j.mrrev.2019.02.001)

28. Schinasi L, Leon ME. Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health*. 2014;11:4449–4527.

29. Cerhan JR, Slager SL. Familial predisposition and genetic risk factors for lymphoma. *Blood*. 2015;126(20):2265–2273.

30. Hernandez-Verdin I, Labreche K, Benazra M, et al. Tracking the genetic susceptibility background of B-cell non-Hodgkin' s lymphomas from genome-wide association studies. *Int. J. Mol. Sci.* 2021;22.

31. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat. Genet.* 2010;42(8):661–664.

32. Smedby KE, Foo JN, Skibola CF, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet.* 2011;7(4):1–11.

33. Makgoeng SB, Bolanos RS, Jeon CY, et al. Markers of immune activation and inflammation, and non-Hodgkin lymphoma: a meta-analysis of prospective studies. *JNCI Cancer Sprectrum*.

2019;0(0):1–13.

34. Pinu FR, Goldansaz SA, Jaine J. Translational Metabolomics: Current Challenges and Future Opportunities. *Metabolites*. 2019;9:1–25.

35. Xiong J, Zhao W-L. Advances in multiple omics of natural-killer/T cell lymphoma. *J. Hematol. Oncol.* 2018;11(134):1–7.

36. Panea RI, Love CL, Shingleton JR, et al. The whole-genome landscape of Burkitt lymphoma subtypes. *Blood*. 2019;134(19):1598–1607.

37. Liang D, Ladva CN, Golan R, et al. Perturbations of the arginine metabolome following exposures to traffic-related air pollution in a panel of commuters with and without asthma. *Environ. Int.* [electronic article]. 2019;127(February):503–513. (https://doi.org/10.1016/j.envint.2019.04.003)

38. Jin R, Banton S, Tran VT, et al. Amino acid metabolism is altered in adolescents with nonalcoholic fatty liver disease - An untargeted, high resolution metabolomics study. *J. Pediatr.* [electronic article]. 2016;172:14–19. (http://dx.doi.org/10.1016/j.jpeds.2016.01.026)

39. Walker DI, Uppal K, Zhang L, et al. High-resolution metabolomics of occupational exposure to trichloroethylene. *Int. J. Epidemiol.* 2016;45(5):1517–1527.

40. Barberini L, Noto A, Fattuoni C, et al. The metabolomic profile of lymphoma subtypes: A pilot study. *Molecules*. 2019;24:1–11.

41. Bueno Duarte GH, de Piloto Fernandes AMA, Silva AAR, et al. Gas chromatography-mass spectrometry untargeted profiling of non-Hodgkin's lymphoma urinary metabolite markers. *Anal. Bioanal. Chem.* 2020;412(27):7469–7480.

42. Medriano CAD, Na J, Lim K, et al. Liquid chromatography mass spectrometry-based metabolite pathway analyses of myeloma and non-Hodgkin's lymphoma patients. *Cell J.* 2017;19:44–54.

43. Müller AM, Ihorst G, Mertelsmann R, et al. Epidemiology of non-Hodgkin's lymphoma (NHL): trends, geographic distribution, and etiology. *Ann. Hematol.* 2005;84:1–12.

44. Skibola CF, Bracci PM, Halperin E, et al. Polymorphisms in the estrogen receptor 1 and vitamin C and matrix metalloproteinase gene families are associated with susceptibility to lymphoma. *PLoS*

One. 2008;3(7):e2816.

45. Jaffe ES, Harris NL, Diebold J, et al. World Health Organization classification of lymphomas: a work in progress. *Ann. Oncol.* 1998;9:25–30.

46. Yu T, Park Y, Johnson JM, et al. apLCMS - adaptive processing of high-resolution LC/MS data. *Bioinformatics*. 2009;25(15):1930–1936.

47. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*. 2013;14.

48. Johnson WE, Li C. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–127.

49. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* [electronic article]. 1995;57:289–300. (www.jstor.org/stable/2346101)

50. Li S, Park Y, Duraisingham S, et al. Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* 2013;9(7):e1003123.

51. Uppal K, Walker DI, Jones DP. xMSannotator: an R package for network-based annotation of high-resolution metabolomics data. *Anal. Chem.* 2017;89:1063–1067.

52. Le Floc'h N, Otten W, Merlot E. Tryptophan metabolism, from nutrition to potential therapeutic applications. *Amino Acids*. 2011;41:1195–1205.

53. Masaki A, Ishida T, Maeda Y, et al. Clinical significance of tryptophan catabolism in follicular lymphoma. *Hematol. Oncol.* 2020;38:742–753.

54. Masaki A, Ishida T, Maeda Y, et al. Clinical significance of tryptophan catabolism in Hodgkin lymphoma. *Cancer Sci.* 2018;109:74–83.

55. Li J, Li J, Wang H, et al. Tyrosine and glutamine-leucine are metabolic markers of early-stage colorectal cancers. *Gastroenterology* [electronic article]. 2019;157:257–259. (https://doi.org/10.1053/j.gastro.2019.03.020)

56. Dong J-F, Zheng X-Q, Rui H-B. Effect of taurine on immune function in mice with T-cell lymphoma during chemotherapy. *Asian Pac. J. Trop. Med.* 2017;10(11):1090–1094.

57. Barba I, Sanz C, Barbera A, et al. Metabolic fingerprinting of fresh lymphoma samples used to discriminate between follicular and diffuse large B-cell lymphomas. *Exp. Hematol.* [electronic article]. 2009;37:1259–1265. (http://dx.doi.org/10.1016/j.exphem.2009.08.006)

Tables

Table 1. Baseline characteristics of DLBCL and FL cases and controls, SF Bay Area Case-Control, 2001-2005

Characteristic	DLBCL		FL	
	Cases (n = 75)	Controls (n = 75)	Cases (n =67)	Controls (n =67)
Sex, n (%)				
Men	39 (52)	39 (52)	37 (55)	37 (55)
Women	36 (48)	36 (48)	30 (45)	30 (45)
Age, yrs, mean (SD)	65.04 (12.45)	64.83 (12.17)	61.99 (11.81)	61.91 (11.80)
BMI (kg/m^2), n (%)				
Underweight, < 18.5	2 (3)	2 (3)	3 (5)	0 (0.0)
Normal, 18.5 - 24.9	48 (64)	34 (45)	32 (48)	26 (39)
Overweight, 25 - 29.9	13 (17)	28 (37)	16 (24)	20 (30)
Obese, Class I, 30 – 34.9	7 (9)	7 (9)	13 (19)	16 (24)
Obese, Class II, 35+	5 (7)	4 (5)	3 (5)	4 (6)
County, n (%)				
1	15 (20)	10 (13)	17 (25)	11 (16)
2	11 (15)	14 (19)	13 (19)	13 (19)
3	5 (7)	6 (8)	7 (11)	6 (9)
4	21 (28)	23 (31)	13 (19)	17 (25)
5	5 (7)	9 (12)	7 (11)	7 (11)
6	18 (24)	13 (17)	10 (15)	13 (19)

Abbreviations: yrs, years; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; BMI, body mass index

Table 2. Mummichog metabolism pathways associated with DLBCL (a) and FL (b), SF Bay Area Case-Control, 2001-2005

a. DLBCL (HILIC+/C18-)

Metabolism Pathways	Overlap Size	Pathway Size	*p*-value	Pos/Neg
Caffeine metabolism	6	16	0.006	C18-
Caffeine metabolism	3	7	0.009	HILIC+
Chondroitin sulfate degradation	3	6	0.006	HILIC+
Galactose metabolism	5	23	0.026	HILIC+
Heparan sulfate degradation	3	6	0.006	HILIC+
Hexose phosphorylation	5	19	0.041	C18-
Hexose phosphorylation	4	19	0.048	HILIC+
N-Glycan biosynthesis	3	11	0.036	HILIC+
N-Glycan biosynthesis	4	13	0.037	C18-
Pentose phosphate pathway	11	40	0.004	C18-
Phosphatidylinositol phosphate metabolism	6	26	0.012	HILIC+
Propanoate metabolism	3	9	0.019	HILIC+
Propanoate metabolism	5	19	0.041	C18-
Sialic acid metabolism	7	26	0.018	C18-
Sialic acid metabolism	5	26	0.043	HILIC+
Starch and Sucrose Metabolism	4	10	0.004	HILIC+
Starch and Sucrose Metabolism	4	8	0.007	C18-
Tryptophan metabolism	11	67	0.018	HILIC+
Valine, leucine and isoleucine degradation	4	18	0.038	HILIC+

b. FL (HILIC+/C18-)

Metabolism Pathways	Overlap Size	Pathway Size	*p*-value	Pos/Neg
Methionine and cysteine metabolism	9	56	0.009	HILIC+
Porphyrin metabolism	3	19	0.067	C18-
Tryptophan metabolism	7	70	0.083	C18-
Tyrosine metabolism	11	84	0.020	HILIC+
Urea cycle/amino group metabolism	8	67	0.027	C18-
Vitamin B6 (pyridoxine) metabolism	4	12	0.003	C18-
Vitamin E metabolism	5	30	0.035	HILIC+
Vitamin D	1	1	0.040	HILIC+

Table 3. Putative chemical identification for metabolic features in tryptophan pathway associated with DLBCL risk, SF Bay Area Case-Control, 2001-2005

m/z	RT	Col	OR (95% CI)	Putative Identity	Adduct
139.0584	60	HILIC+	0.76 (0.61, 0.94)	Anthranilate	M(C13)+H[1+]
160.0757	36	HILIC+	0.46 (0.24, 0.85)	Indole-3-Acetaldehyde	M+H[1+]
167.0218	82	HILIC+	0.50 (0.32, 0.78)	Quinolinate	M[1+]
181.0718	31	HILIC+	0.67 (0.49, 0.91)	D-Galactose	M+H[1+]
195.0875	29	HILIC+	0.78 (0.66, 0.93)	CE2152 (Unidentified)	M+Na[1+]
203.0526	62	HILIC+	1.60 (1.11, 2.31)	D-Galactose	M+Na[1+]
206.0809*	26	HILIC+	0.61 (0.41, 0.92)	5-Methoxyindoleacetate	M+H[1+]
226.0121	61	HILIC+	1.60 (1.03, 2.48)	CE2949 (Unidentified)	M+Na[1+]
238.0921*	103	HILIC+	1.52 (1.02, 2.29)	L-Formylkynurenine	M(C13)+H[1+]
427.0760	95	HILIC+	1.56 (1.07, 2.29)	Acetoacetyl-CoA	M(C13)+2H[2+]

* Not confirmed by EIC plots

Table 4. Putative chemical identification for metabolic features in tryptophan pathway associated with FL risk, SF Bay Area Case-Control, 2001-2005

m/z	RT	Col	OR (95% CI)	Putative Identity	Adduct
188.0104*	283	C18-	0.42 (0.18, 0.95)	3-Hydroxyanthranilate	M+Cl[-]
188.0105	21	C18-	0.76 (0.58, 0.99)	3-Hydroxyanthranilate	M+Cl[-]
190.0510	23	C18-	2.33 (1.12, 4.87)	5-Phenyl-1,3-Oxazinane-2,4-Dione	M-H[-]
194.0464	20	C18-	0.83 (0.70, 0.99)	Indole-3-Acetamide	M+Na-2H[-]
194.0821	178	C18-	0.36 (0.13, 0.98)	N-Methyltryptamine	M+Na-2H[-]
207.0777	23	C18-	5.36 (2.10, 13.68)	L-Kynurenine	M-H[-]
225.0895	293	C18-	3.37 (1.29, 8.82)	3-Hydroxykynurenamine	M+HCOO[-]

* Not confirmed by EIC plots

Table 5. Putative chemical identification for metabolic features in tyrosine pathway associated with FL risk, SF Bay Area Case-Control, 2001-2005

m/z	RT	Col	OR (95% CI)	Putative Identity	Adduct
88.0393*	206	HILIC+	3.18 (1.55, 6.54)	2-Aminoacrylate	M+H[1+]
91.0275*	271	HILIC+	4.02 (1.11, 14.60)	Caffeate	M+2H[2+]
93.0448	104	HILIC+	0.21 (0.05, 0.82)	Choline Phosphate	M+2H[2+]
110.0293	37	HILIC+	0.45 (0.22, 0.92)	Tymsf, Tyramine O-Sulfate	M(C13)+2H[2+]
122.0271*	196	HILIC+	2.51 (1.34, 4.69)	L-Cysteine	M+H[1+]
181.1037*	54	HILIC+	1.22 (1.02, 1.46)	(-)-Salsolinol	M(C13)+H[1+]
240.0309*	25	HILIC+	0.62 (0.43, 0.91)	Tymsf, Tyramine O-Sulfate	M+Na[1+]
262.0918*	71	HILIC+	1.51 (1.00, 2.27)	7,8-Dihydrobiopterin	M+Na[1+]
265.1195*	37	HILIC+	0.69 (0.49, 0.98)	Alpha-N-Phenylacetyl-L-Glutamine	M+H[1+]
316.0754*	97	HILIC+	0.68 (0.51, 0.91)	5-S-Cysteinyldopa	M[1+]

* Not confirmed by EIC plots

Figures

Figure 1. HRM data extraction and processing flowchart, SF Bay Area Case-Control, 2001-2005

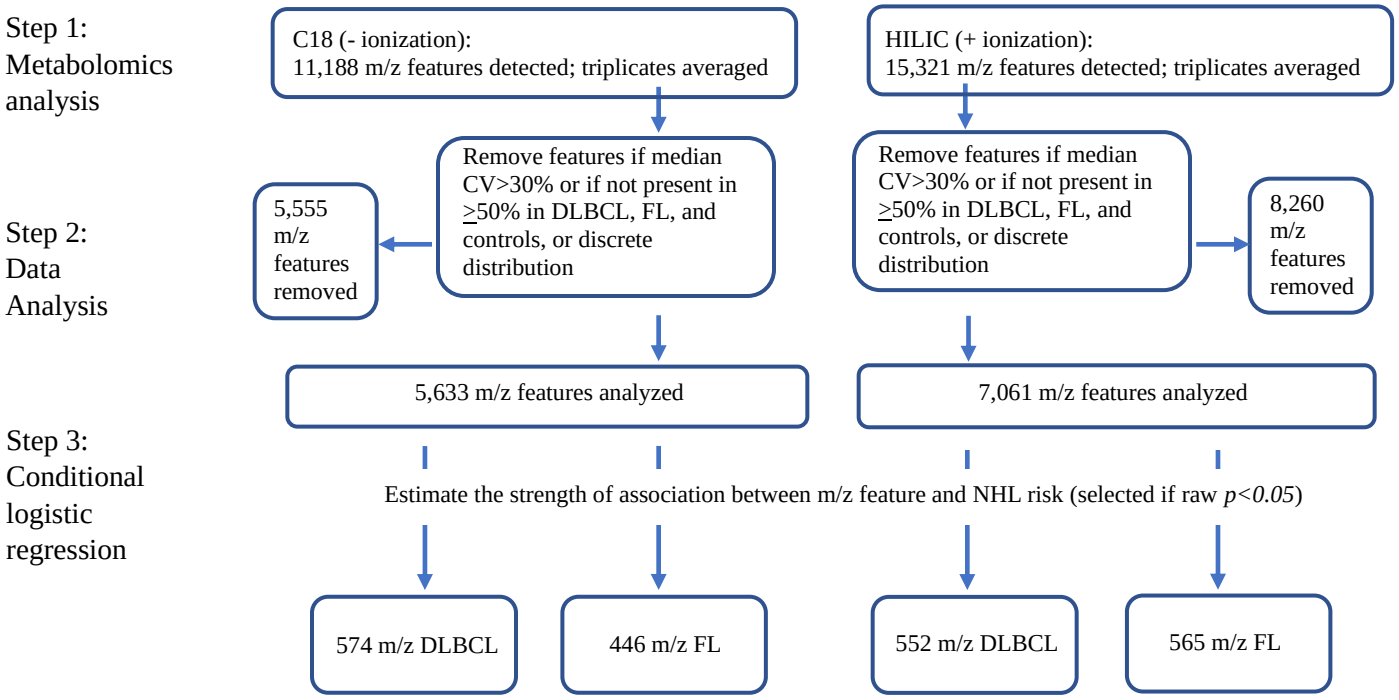
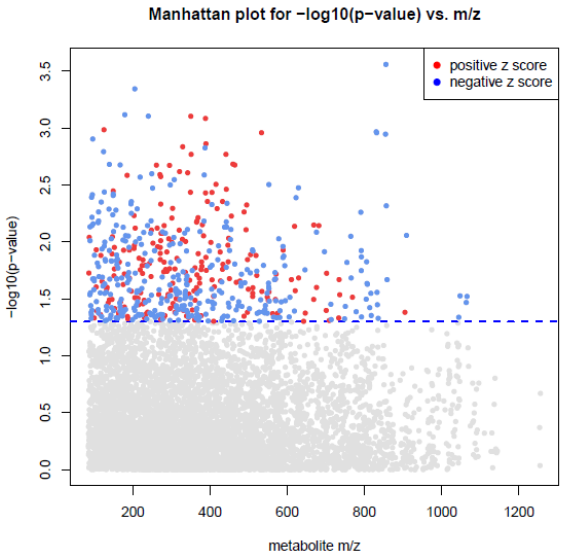
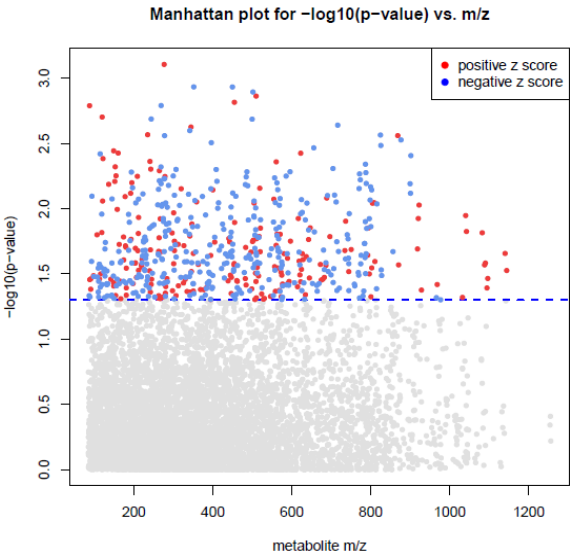


Figure 2. Manhattan plots of significant metabolic features associated with DLBCL (a. HILIC+, b. C18-) and FL (c. HILIC+, d. C18-) risk, SF Bay Area Case-Control, 2001-2005

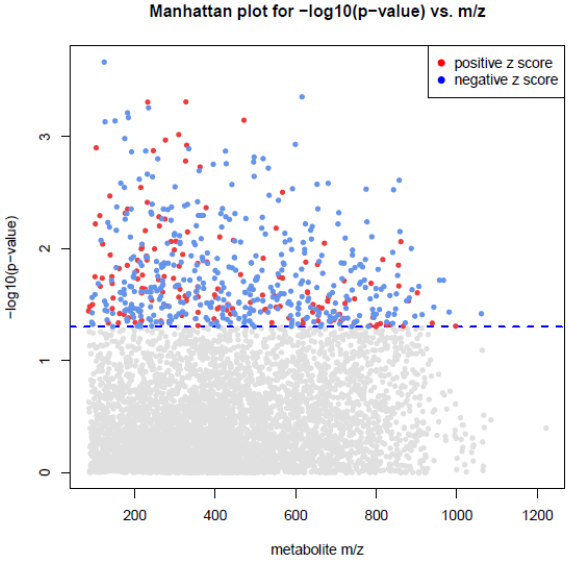
a. DLBCL (HILIC+), 552 significant metabolic features



c. FL (HILIC+), 565 significant metabolic features



b. DLBCL (C18-), 574 significant metabolic features



d. FL (C18-), 446 significant metabolic features

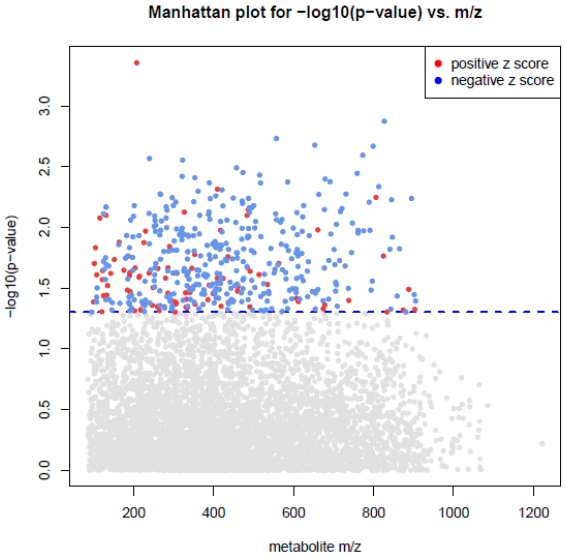
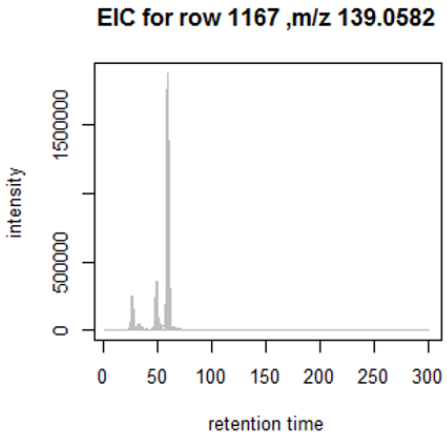
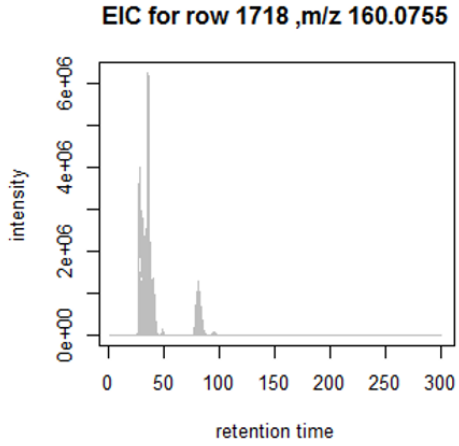


Figure 3. EIC plots for putative chemicals identified in tryptophan pathway and associated with DLBCL risk, SF Bay Area Case-Control, 2001-2005

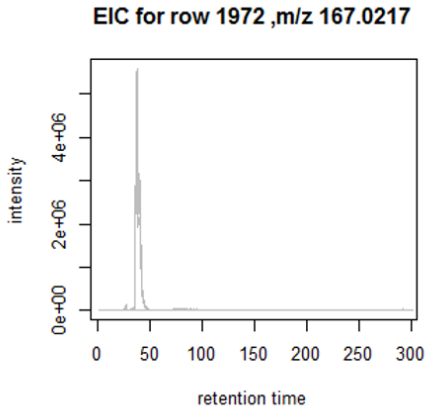
a. Anthranilate



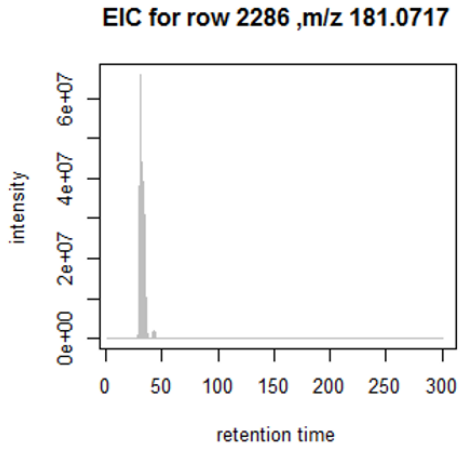
b. Indole-3-Acetalde



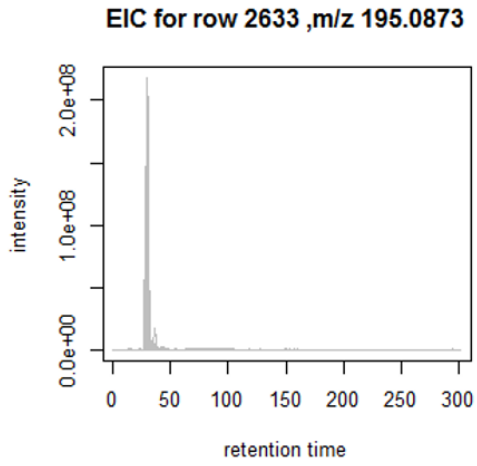
c. Quinolinate



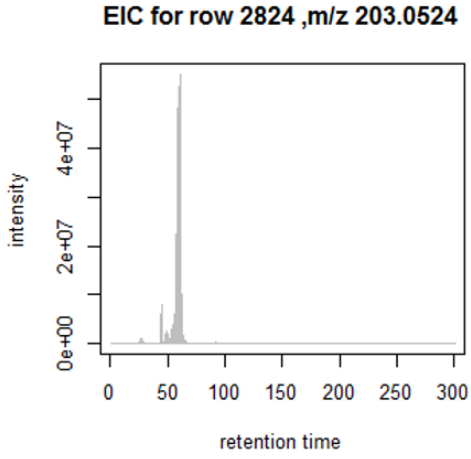
d. D-Galactose



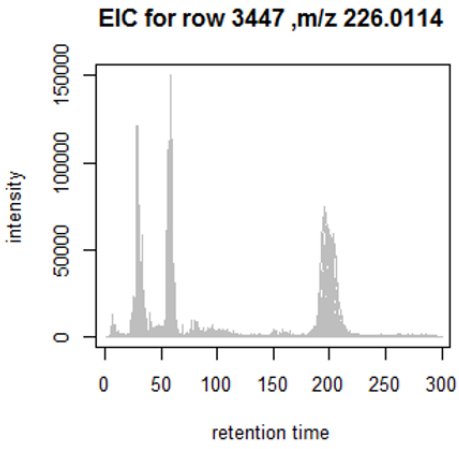
e. CE2152



f. D-Galactose



g. CE2949



h. Acetoacetyl-CoA

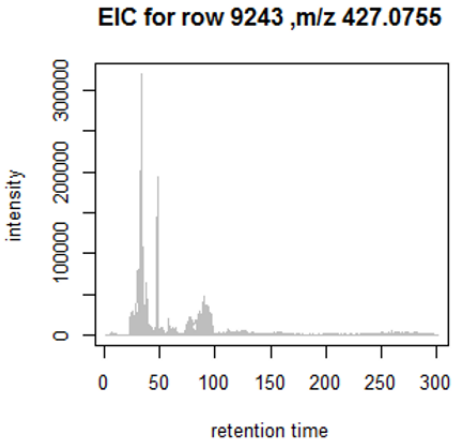
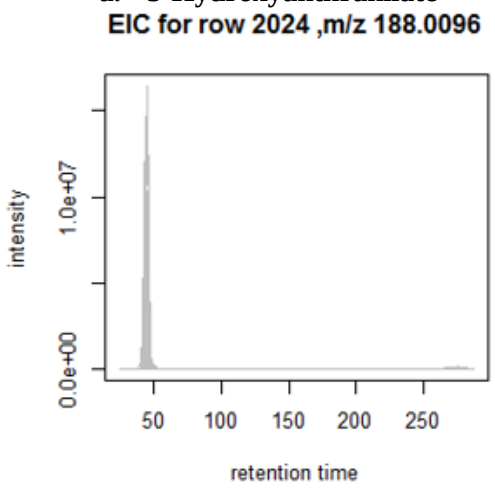
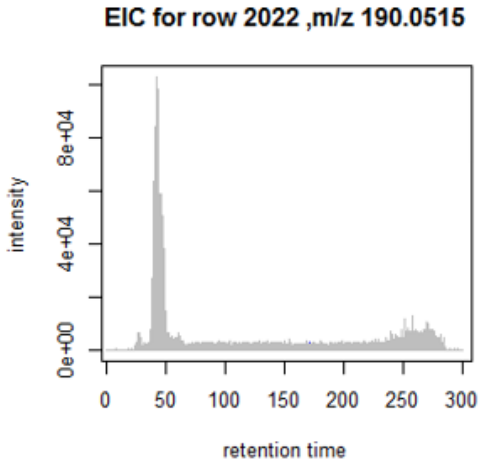


Figure 4. EIC plots for putative chemicals identified in tryptophan pathway and associated with FL risk, SF Bay Area Case-Control, 2001-2005

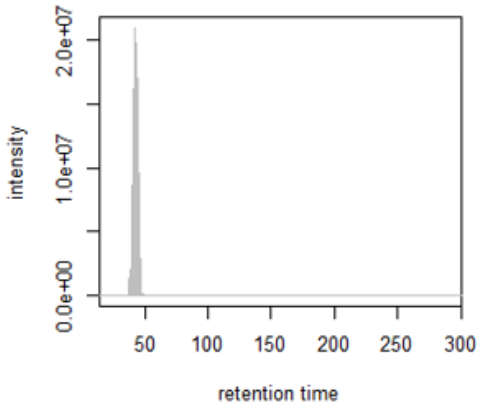
a. 3-Hydroxyanthranilate



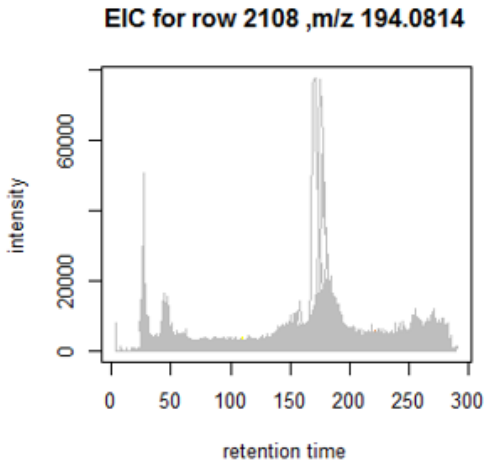
b. 5-Phenyl-1,3-Oxazinane-2,4-Dione



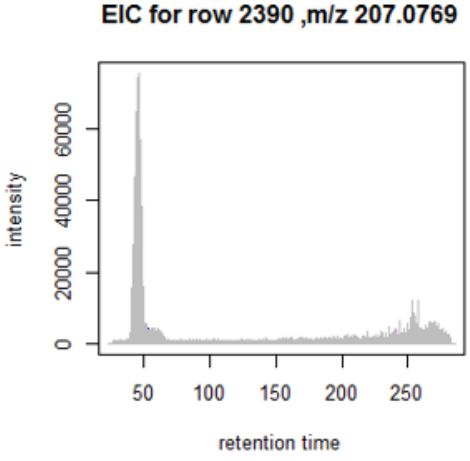
c. Indole-3-Acetamide



d. N-Methyltryptamine



e. L-Kynurenine



f. 3-Hydroxykynurenamine

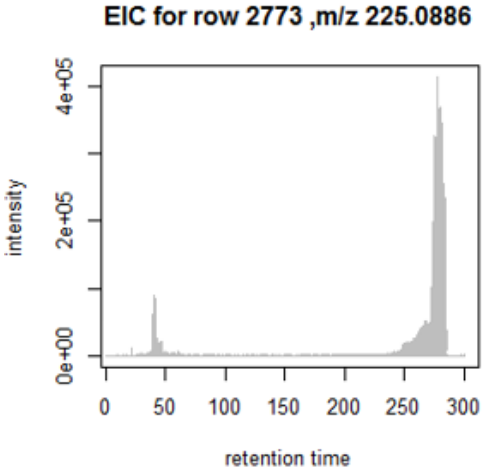
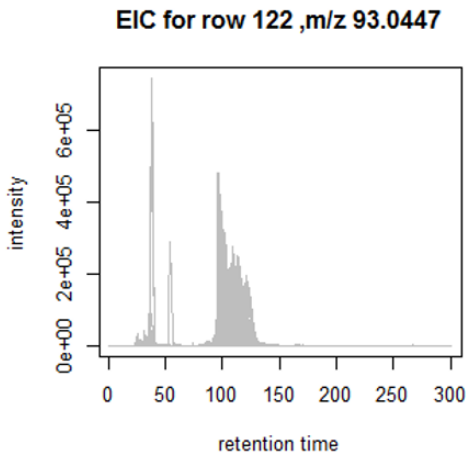
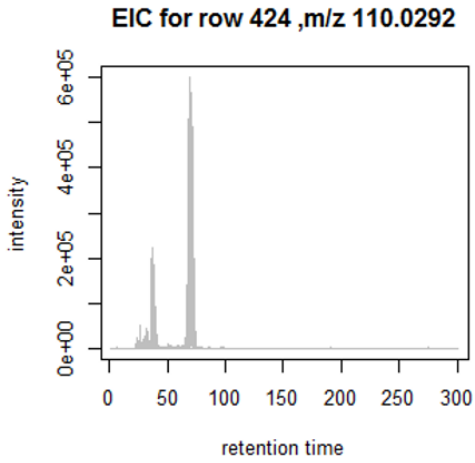


Figure 5. EIC plots for putative chemicals identified in tyrosine pathway and associated with FL risk, SF Bay Area Case-Control, 2001-2005

a. Choline Phosphate



b. Tymsf, Tyramine O-Sulfate



Chapter III: Future Directions

The findings from our pilot HRM study and evidence from previously published studies support a potential role of tryptophan and tyrosine metabolism in the development of NHL. Because tryptophan and tyrosine metabolism are also involved in the immune response during tumorigenesis, it is important for further research to confirm our results via targeted approaches for measuring metabolites found within these pathways, such as kynurenine. Because our case-control study was not able to collect information before NHL case diagnosis, post-diagnosis treatments, and NHL stage, future research should consider these important clinical covariates. Additionally, future work in a diverse population is warranted to generalize our results to broader NHL populations. Our study provides a starting point of which metabolic pathways to potentially target and requires further exploration. If subsequent studies support our findings, it may be possible to discover early diagnostic biomarkers for NHL risk.