CHARACTERIZATION OF ENVIRONMENTAL AND GENETIC FACTORS IN MULTIPLE-CASE LYMPHOID CANCER FAMILIES

by

SAMANTHA JEAN JONES

B.Sc., Brock University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies the acceptance, a dissertation entitled:

Characterization of environmental and genetic factors in multiple-case lymphoid cancer families

submitted by Samantha Jones in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Medical Genetics

Examining Committee:

Angela R. Brooks-Wilson, Professor, Medical Genetics, UBC
Supervisor

Jan Friedman, Professor, Medical Genetics, UBC
Supervisory Committee Member

David Huntsman, Professor, Pathology and Laboratory Medicine, UBC
University Examiner

Wan Lam, Professor, Pathology and Laboratory Medicine, UBC
University Examiner

John R. McLaughlin, Professor, Epidemiology, University of Toronto
External Examiner

Additional Supervisory Committee Members:

Inanc Birol, Professor, Medical Genetics, UBC
Supervisory Committee Member

Carolyn Brown, Professor, Medical Genetics, UBC
Supervisory Committee Member

Abstract

Lymphoid cancers, the fifth most common cancer type in Canada, comprise a clinically and biologically heterogeneous group of neoplasms. Established risk factors include older age, male sex, compromised immune function, and family history of lymphoproliferative disorders. The hygiene hypothesis, according to which a relative lack of infectious exposure during early childhood may cause aberrant immune development and subsequent adult-onset immune-related diseases, including allergies, autoimmune conditions and some lymphoid cancers, provides a framework for understanding these risk factors. Susceptibility genes involved in immune function and DNA repair have been identified; however, there remains a large gap in our knowledge of genetic and environmental factors for familial lymphoid cancers.

We examined familial aggregation, age of onset and environmental factors in more than 200 multiple-case lymphoid cancer families. Familial lymphoid cancer ages of onset were substantially earlier than comparable population data and showed an anticipation effect after controlling for ascertainment biases. Relative to the general population, families were enriched for some combinations of lymphoid cancers (e.g., HL/HL, CLL/CLL). Familial cases were more likely to have allergies and a tonsillectomy, which may indicate defective immune regulation. The risk of lymphoma tended to decrease with later birth order and larger sibship sizes. Measures of family structure and crowding relate to the hygiene hypothesis as they affect age and extent of exposures to infections, with low birth order and smaller sibships correlating with higher risk. These associations underscore the complex etiology of familial lymphoma and suggest that lymphoid cancers in multiple-case families may be different from sporadic cases.

Genome-wide scans have identified few risk alleles with small effect sizes in multiple-case families. We performed a genome-wide Identity-by-Descent analysis using 1.8 million markers on a well-characterized multi-generational family with 4 lymphoid cancer cases. Three interesting candidate variants were found (*MYC, EPHA1, MMS19*), but no compelling high-penetrance disease variant segregated with lymphoid cancer. Identifying genetic factors in rare lymphoid cancer families will aid in uncovering key networks involved in cancer susceptibility.

Lymphoma has an important familial component. Establishing genetic and environmental associations facilitates a better understanding of lymphomagenic mechanisms and effective approaches to cancer prevention and clinical management.

Lay Summary

Lymphoid cancers are a diverse group of neoplasms. They are the fifth most common cancer type in Canada; however, multiple-case families are rare. To understand the role that environmental factors and genetic factors may play on the risk of lymphoid cancer in a familial setting, we characterized over 200 multiple-case lymphoid cancer families. Familial cases had an earlier age of onset and were enriched for some subtypes of lymphoid cancer. The risk of lymphoid cancer decreased among later birth order and larger sibship size. Familial cases were more likely to show indicators of defective immune regulation than their unaffected siblings. These observations suggest familial lymphoid cases may be different than non-familial lymphoid cancer cases. Multiple-case families may have an underlying susceptibility factor that affects lymphoid cancer risk. Understanding the mechanisms by which environmental and lifestyle factors affect lymphoid cancer among families may improve cancer prevention.

Preface

I, Samantha Jones, conducted all investigations presented in this thesis, with the exceptions listed below, under the supervision of Dr. Angela Brooks-Wilson at the British Columbia Cancer Research Centre in Vancouver, Canada. All studies in this thesis were conceived, designed, performed, analyzed and interpreted by Samantha Jones, while Dr. Angela Brooks-Wilson contributed to the design and interpretation.

All studies in this thesis were approved by the University of British Columbia / BC Cancer Research Ethics Board certificate number H05-60107. Samantha Jones completed the Tri-Council Policy Statement 2 (TCPS2) for ethical conduct of research involving humans and/or human biological materials (certificate issued on July 8th, 2014).

- As study coordinators, Amy Williams, Diane Salema, Ruth Thomas, Dr. Johanna Schuetz, and Amy English enrolled families and completed telephone interviews with family members.
- Dr. Graham W. Slack (oncology pathologist) verified the lymphoid cancer subtype using histopathology slides.

Chapter 1: (Introduction) I wrote the content in Chapter 1.
Chapter 2: (Methods) I wrote the content in Chapter 2.

Chapter 3: (Cooccurrence) A version of Chapter 3 is published as: **Jones SJ**, Voong J, Thomas R, English A, Schuetz J, Slack GW, Graham J, Connors JM, Brooks-Wilson A. Nonrandom occurrence of lymphoid cancer types in 140 families. *Leukemia & Lymphoma* **58**(9): 2134-2143 (2017). DOI: 10.1080/10428194.2017.1281412.

- I generated the research questions, constructed the dataset and performed data processing. I performed data analyses and figure/table generation with Jackson Voong.
- I contributed to study design and data interpretation with Jackson Voong, Dr. Jinko Graham, Dr. Joseph M. Connors and Dr. Angela Brooks-Wilson.
- I wrote the first draft of the manuscript with input from Dr. Angela Brooks-Wilson. All authors contributed to editing the article.

Chapter 4: (Age of onset) A version of Chapter 4 is published as: **Jones SJ**, Voong J, Thomas R, English A, Schuetz J, Slack GW, Graham J, Connors JM, Brooks-Wilson A. Nonrandom occurrence of lymphoid cancer types in 140 families. *Leukemia & Lymphoma* **58**(9): 2134-2143 (2017). DOI: 10.1080/10428194.2017.1281412. A second manuscript that addresses ascertainment bias is in preparation by Samantha Jones.

Part A:

- I generated the research questions, constructed the dataset and performed data processing. I performed data analyses and figure/table generation with Jackson Voong.
- I contributed to study design and data interpretation with Jackson Voong, Dr. Jinko Graham, Dr. Joseph M. Connors and Dr. Angela Brooks-Wilson.
- I wrote the first draft of the manuscript. All authors contributed to editing the article.

Part B:

- I generated the research questions, constructed the dataset, and independently performed all data processing, analyses, and generated tables/figures. I interpreted the data with Dr. Angela Brooks-Wilson. I wrote the first draft of the manuscript.

Chapter 5: (Hygiene hypothesis) Studies described in Chapter 5 are published as: **Jones SJ**, Stroshein S, Liu D, Spinelli J, Connors JM, Brooks-Wilson A. Birth order, sibship size, childhood environmental and immune-related disorders and risk of lymphoma in lymphoid cancer in families. *Cancer Epidemiology, Biomarkers and Prevention* (2019). DOI: 10.1158/1055-9965.EPI-19-1204.

- I provided the initial motivation for the project and generated the research questions and study design. Dr. Joseph M. Connors and Dr. Angela Brooks-Wilson also contributed to study design.
- I performed a majority of the statistical analyses (chi-square goodness-of-fit test, standard logistic regression with generalizes estimating equation, stepwise model selection, and subtype analyses). I completed the standard logistic regression analysis with Sumara Stroshein. Amy Williams completed the permutation analyses. I analyzed and interpreted

the data with Sumara Stroshein, Dr. John Spinelli and Dr. Angela Brooks-Wilson. Dongmeng Liu, Dr. John Spinelli, and Dr. Susan Slager provided statistical guidance.

- I generated figures/tables and wrote the first draft of the manuscript. All authors contributed to editing the article.

Chapter 6: (Family 133) This analysis will be used to prepare a manuscript for publication in a peer-reviewed journal.

- I contributed to study design with Dr. Angela Brooks-Wilson. Genomic DNA was extracted from whole blood by Stephen Leach. I extracted tumour DNA and RNA from formalin-fixed paraffin-embedded tissues, and performed validation measures for all samples.
- SNP genotyping was performed at McGill University/Genome Quebec Innovation Centre (Montreal, Quebec, Canada). I performed quality control (QC) measures and called high-quality SNPs for identity-by-descent detection.
- Whole exome sequencing was performed at the Genome Sciences Centre by the Sequencing core laboratories. Variant calling was performed by the Bioinformatics core.
- I independently performed all analyses (kinship coefficients, phasing and Identity-by-descent segment detection using Beagle, Refined IBD, FastIBD, KING and Golden Helix). I extracted, filtered and prioritized variants.
- I generated all figures and tables.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	amino acid
ALC	Absolute lymphocyte count
ALL	Acute lymphoblastic leukemia
Alt	Alternate allele
AoO	Age of onset
ASD	Autism spectrum disorder
BAM	Binary alignment file
BC	British Columbia
BCC	British Columbia Cancer
BL	Burkitt lymphoma
BMI	Body mass index
bp	Base pair
CADD	Combined annotation-dependent depletion
CHL	Classic Hodgkin lymphoma
CI	Confidence interval
CiNA	Cancer in North America
CLL	Chronic lymphocytic leukemia
cM	Centimorgan
CNS	Central nervous system
dbNSFP	Database for nonsynonymous SNPs' functional predictions
DLBCL	Diffuse large B-cell lymphoma
DLCL	Diffuse large-cell lymphoma
DNA	Deoxyribonucleic acid
DP	Depth of coverage
ds	Double strand
DZ	Dizygotic twins
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
Eph	Erythropoietin-producing human hepatocellular
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embedded
FL	Follicular lymphoma
FWER	Familywise error rate

gDNA	Genomic DNA
GEE	Generalized estimating equation
gnomAD	Genome Aggregation Database
GWAS	Genome wide association study
H.	*Helicobacter*
HCL	Hairy cell leukemia
HCV	Hepatitis C virus
HHV-8	Human herpesvirus 8
HIV	Human immunodeficiency virus
HL	Hodgkin lymphoma
HLA	Human leukocyte antigen
HRS	Hodgkin/Reed-Sternberg
IBD	Identity by descent
IBD1	Identity by descent on one chromosome
IBD2	Identity by descent on two chromosomes
ICD	International statistical classification of disease
IgE	Immunoglobulin E
IgM	Immunoglobulin M
LCFC	Lymphoid Cancer Families Consortium
LCFS	Lymphoid Cancer Family Study
LD	Linkage disequilibrium
LP	Lymphocyte predominant
LPD	Lymphoproliferative disorder
LPL	Lymphoplasmacytic lymphoma
LRR	Leucine-rich repeat
M.	*Mycoplasma*
MAF	Minor allele frequency
MALT	Mucosa associated lymphoid tumour
MBL	Monoclonal B-cell lymphocytosis
MC	Mixed cellularity
MCL	Mantle cell lymphoma
MF	Mycosis fungoides
MGUS	Monoclonal gammopathy of undetermined significance
MM	Multiple myeloma
MZ	Monozygotic twins

MZL	Marginal zone lymphoma
NAACCR	North American Association of Central Cancer Registries
NER	Nucleotide excision repair
NGS	Next generation sequencing
NHL	Non-Hodgkin lymphoma
NK	Natural killer
NLPHL	Nodular lymphocyte predominant Hodgkin lymphoma
NOS	Not otherwise specified
NS	Nodular sclerosis
OR	Odds ratio
QC	Quality control
QIC	Quasilikelihood under the Independence model Criterion
RA	Rheumatoid arthritis
Ref	Reference allele
RNA	Ribonucleic acid
SD	Standard deviation
SEER	Surveillance, Epidemiology, and End Results
SES	Socioeconomic status
SLE	Systemic lupus erythematosus
SLL	Slow lymphocytic lymphoma
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SS	Sjögren syndrome
SVS	SNP and Variation Suite
THRLBCL	T-cell/histiocyte rich diffuse large B-cell lymphoma
TOPMed	Trans-Omics for Precision Medicine
UBC	University of British Columbia
USA	United States of America
UTR	Untranslated region
WBC	White blood cell
WES	Whole exome sequencing
WHO	World health organization
WM	Waldenström macroglobulinemia

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my research supervisor, Dr. Angela Brooks-Wilson, who supported me throughout my PhD, giving me the freedom and trust to explore my scientific interests. Your remarkable expertise in genetics and willingness (and patience) to train me is truly appreciated.

I would also like to thank my committee members Drs. Inanc Birol, Carolyn Brown and Jan Friedman for their guidance and encouraging remarks. Your expertise and critique were appreciated and invaluable to my accomplishments.

I would like to thank past and present members of the ABW lab: Jessica Nelson, Stephen Leach, Sneha Ralli and Rawnak Hoque; study coordinators Ruth Turnbull, Jessica Nelson, Diane Salema and Amy Williams. Your knowledge, help and willingness to share your experiences and information made for a much smoother experience.

My project has benefited from the help of some of the most amazing undergraduate trainees – Sumara Stroshein, Lindsay Woof, and Jackson Voong were incredible Co-op students, and I enjoyed every aspect of supervising and mentoring you. I am also grateful for the help of a Directed Studies student, Sasha Uvarov and many volunteers: Alita Lin, Danielle Curtis and Scott Robinson. Thank you for your detective work and being data-gurus.

In addition, I would like to acknowledge the BC Cancer Agency, BC Foundation for Non-Animal Research, TRFI, CCSRI, CIHR, ICHG, iGSN, Medical Genetics, and UBC for funding part of my work, tuition and conference travels.

Finally, I would like to express my gratitude to my family and friends who were extraordinarily encouraging throughout the years. Your unwavering support, sincere dose of reality, and home cooked meals have kept me alive and (relatively) sane.

Dedication

I dedicate this work to my inspirational teachers and mentors who have taught me the value of education, hard work, critical thinking and a good night's sleep.

I'd also like to dedicate this to my brothers. Without your constant competitiveness, I wouldn't be where I am today.

Chapter 1: Introduction

1.1 Haematological malignancies

Hematopoiesis is the process by which blood cells form and mature. All blood cells arise in the bone marrow from a common pluripotent hematopoietic stem cell, and undergo a series of developmental steps to differentiate into all lineages of mature blood cells. Immature blood cells may mature in the bone marrow, or other parts of the body (e.g., thymus), depending on their function. Blood cell production is normally an organized and controlled process based on the body's need. Mature blood cells may circulate throughout the body via the blood and lymphatic vessels, or reside in lymphatic tissues concentrated in lymph nodes, thymus, spleen and in most major organs (1,2).

The disruption of normal cell fate may form an immature blood cell which can develop into a hematological malignancy. Hematological malignancies are cancers that affect blood, bone marrow, and lymph nodes (1,2). The cell's development is arrested and does not mature further, but is instead replicated, resulting in the proliferation of abnormal blood cells (1–3). Different stages of the hematopoietic process may give rise to a distinct type of cancer. Consequently, hematological malignancies include a large number of genetically and clinically diverse diseases. There are three main types of hematological malignancies: leukemia, lymphoma and myeloma (1,2).

Historically, hematological malignancies were classified by their locations in the body, cellular morphology and the natural disease progression. In lymphoma, the cancerous cells tend to aggregate and form masses or tumours in lymphatic tissues, whereas in leukemia, the cells circulate in the blood and bone marrow (1–4). Myeloma is also a tumour of the bone marrow, which arises from plasma cells and produces a distinctive protein (1,5,6).

1.1.1 Leukemia

Leukemia is a cancer found in blood and bone marrow, and is caused by the rapid production of abnormal white blood cells (WBC) (2,3). The high number of abnormal WBCs are unable to fight infection and impair the ability of the bone marrow to produce red blood cells and platelets. Leukemias are subdivided into 4 groups based on 2 criteria: the type of precursor cell they develop from (lymphoid or myeloid) and how quickly the disease progresses (acute or

chronic) (2,3). Each of the 4 subtypes (acute lymphocytic leukemia, acute myeloid leukemia, chronic lymphocytic leukemia, and chromic myeloid leukemia) are further distinguished by morphological differences, cytogenetic abnormalities, immunophenotype, and clinical features (1,3,7).

This thesis examines multiple-case families with lymphomas and myelomas, but not leukemias. For this reason, leukemias will not be discussed in depth.

1.1.2 Lymphoma

Lymphoma is a type of blood cancer that affects the lymphatic system (1,2,4,7). Abnormal lymphocyte production may impair immune function. Lymphomas are subdivided into 2 categories: Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL) (1,4). HL is characterized by the presence of Hodgkin/Reed-Sternberg (HRS) cells, while these cells are absent in NHL (4,8). HRS cells are derived from B-cells and are unusually large in appearance and may be multinucleated (4,8). By contrast, NHLs are derived from B-lymphocytes or T-lymphocytes; however, most NHLs are of B-cell origin. Lymphomas are classified based on cell size, cell differentiation and cleavage, low-grade or high-grade clinical behaviours, immunophenotype and molecular genotype with morphology and clinical features (1,4,7) .

1.1.2.1 Hodgkin lymphoma

HL accounts for 15% of all lymphoid cancers and is further subdivided into Classical HL (CHL) and nodular-lymphocyte predominant HL (NLPHL) (7,8). Globally, there are an estimated 65,950 new HL cases every year, which vary by age, sex, ethnicity, geographic location and socioeconomic status (7,9,10).

1.1.2.1.1 Classic Hodgkin lymphoma (CHL)

CHL accounts for approximately 90% of all HL cases (**Table 1**) and is characterized by the presence of HRS cells (7–9,11). Based on histological features, CHL is subdivided into four subtypes that vary with respect to age of onset and sex (7,8,11,12).

Nodular sclerosis (NS)

NS CHL is the most common subtype of HL, accounting for 60-70% of all HL cases (7,8). NS HL originates in the lymph nodes of the deep tissue in the centre of the chest (mediastinum)

or neck. In NS, the involved lymph nodes contain HRS cells mixed with normal WBCs. NS is more common in women than in men, usually affects teens and young adults and adults (15 to 34 years of age) (7,8). Most cases are curable (7,8).

Mixed cellularity (MC)

MC HL accounts for 20-30% of all HL cases (7,8). The disease is more common in men than in women, and primarily affects children (under the age of 14) and older adults (55 to 74 years of age) (8). The involved lymph nodes are usually in the upper half of the body, which contain many HRS cells with a mixed cellular background that varies greatly. This subtype is usually diagnosed at a more advanced stage of disease (7,8).

Lymphocyte rich

Lymphocyte rich accounts for less than 5% of HL cases (7,8). The disease may be diffuse or nodular in form and is characterized by the presence of numerous normal-appearing lymphocytes and classic HRS cells. It usually develops in the lymph nodes in the neck, armpits, and above the collarbone (7). Lymphocyte rich HL is typically diagnosed at an early stage in middle-aged adults and has a low relapse rate (8). There is a 2:1 male predominance for the disease.

Lymphocyte depleted

Lymphocyte depleted is rarely diagnosed and accounts for about 1% of all HL cases (7,8). The involved lymph nodes contain abundant HRS cells and few normal lymphocytes. This subtype is usually aggressive and is diagnosed at an advanced stage (8). Lymphocyte depleted HL is more common in men than in women, and typically affects young adults (30 to 37 years of age) (8).

1.1.2.1.2 Nodular lymphocyte predominant HL (NLPHL)

NLPHL accounts for nearly 5% of all HL cases (7,8). NLPHL is characterized by scattered lymphocyte predominant (LP) tumour cells (not HRS cells), also known as "popcorn cells" (7–9,11). NLPHL is more common in men and adults (30-49 years of age) (7,8). NLPHL is indolent (slow-growing) and has a high cure rate (8). There is a 3:1 male predominance for the disease.

1.1.2.2 Non-Hodgkin lymphoma

The NHLs are a heterogeneous group of more than 60 malignancies with varying clinical and biological features (7,13). NHL is divided into B-cell and T-cell neoplasms based on histological characteristics and are further classified by clinical features (13).

1.1.2.2.1 B-cell NHL

B-cell lymphomas represent approximately 85-90% of NHL cases (7).

Burkitt's lymphoma (BL):

BL is a high-grade mature B-cell lymphoma and represents ~1% of NHL cases (7,14). It is one of the fastest growing lymphomas and tends to occur in organs and tissues other than the lymph nodes (extranodal sites), often spreading to the brain or spinal cord. BL develops in children or young adults and occurs most often in young boys (14). There are three subtypes of BL: 1) Endemic BL, which occurs in Africa and is almost always linked with Epstein-Barr virus; 2) Immunodeficiency-associated BL, which occurs in individuals with weakened immune systems (e.g., human immunodeficiency virus); and 3) Sporadic BL, the most common type of BL in high-income countries (7,14). BL is easily cured.

Chronic lymphocytic leukemia (CLL)/Small lymphocytic lymphoma (SLL):

CLL and SLL are essentially the same disease, the only difference being where the cancer primarily occurs. For CLL, most of the malignant clones are located in the bloodstream and the bone marrow (lymph nodes and spleen may also be involved) (7,15). In contrast, for SLL, most of the malignant clones appear in the lymph nodes (7,15). CLL/SLL can be detected during routine blood tests and most individuals are asymptomatic. Asymptomatic cases are offered active surveillance (watchful waiting), while aggressive cases may be treated with combinations of chemotherapy, radiation therapy or a stem cell transplant. CLL occurs more frequently in men and typically affects individuals over the age of 70 (7,15).

Diffuse large B-cell lymphoma (DLBCL):

DLBCL is the most common lymphoma subtype which represents ~30% of all NHL cases (7,16). It is an aggressive subtype and originates in the lymph nodes (16). The average age of diagnosis for DLBCL is the mid-60s, and it is more common in men than in women. At the time of diagnosis, approximately 50-60% of DLBCL cases have spread beyond the lymph nodes above

and below the diaphragm or to different lymphatic organs such as the spleen, liver or bone marrow (16). Chemotherapy is used as the main treatment; however, a stem cell transplant may be used for recurrent cases (16).

T-cell/histiocyte rich large B-cell lymphoma (THRLBCL) is a rare morphological variant of DLBCL (1-3% of all B-cell lymphomas) with abundant nonneoplastic T-cells and histiocytes (17). THRLBCL is more common in men and middle-aged adults (40 years of age) (17). It is often challenging to diagnose due to its similarity with other lymphoid cancers such as NLPHL (17).

Follicular lymphoma (FL):

FL is the second most common subtype of lymphoma, comprising 20-30% of all NHL cases (7,18). FL is indolent (slow-growing) and affected individuals may be asymptomatic (18). FL usually occurs in individuals over 50 years of age and is slightly more common in women than in men (7,18). FL responds well to treatment; however, most cases relapse. Although rare, FL can transform into DLBCL (18).

Hairy cell leukemia (HCL):

HCL is a low-grade B-cell lymphoma and is often left untreated until it reaches a progressive stage (19). HCL is more common in men than in women, and typically affects individuals over the age of 50 years. About 1% of NHLs or 2% of all leukemias are HCL (7,19). HCL gets its unusual name because of the hair-like projections on the surface of the malignant cell. HCL is called a leukemia because the lymphocytes are found in the blood, but they may also collect in the bone marrow and spleen (7,19).

Lymphoplasmacytic lymphoma (LPL)/ Waldenström macroglobulinemia (WM):

LPL is a chronic neoplasm that occurs in B-cells that normally mature into antibody producing plasma cells (20,21). The malignant clone over-produces antibodies which accumulate in the blood and impair circulation (21). WM is characterized as a subset of LPL that has detectable immunoglobulin M (IgM) paraproteins, and accounts for approximately 95% of LPL cases. LPL/WM typically affects individuals over the age of 60 and accounts for 1-2% of all hematological tumours (20,21).

Mantle cell lymphoma (MCL):

MCL is a rare B-cell lymphoma which comprises approximately 5% of all NHLs (7,22). MCL is mostly indolent, but has aggressive variations. It typically affects men over the age of 60 (22). The disease is called “mantle cell lymphoma” because the tumour cells originate from the “mantle zone,” which is the outer edge of the lymph node. MCL is frequently treated using chemotherapy (22).

Marginal zone lymphoma (MZL):

MZL is a group of indolent B-cell NHLs which account for 8% of all NHL cases (7,23). MZL is more common in women than in men and typically affects individuals over the age of 60 (7). There are 3 types of MZL: 1) mucosa-associated lymphoid tissue (MALT) lymphoma, 2) Nodal MZL, and 3) Splenic MZL.

1. **Mucosa-associated lymphoid tissue (MALT) lymphoma:**

MALT (or extranodal) lymphoma is the most common MZL, accounting for 70% of all MZL cases (7,23). MALT lymphoma is a B-cell NHL which develops in the mucosa membranes of lymphatic tissue. MALT lymphomas affect individuals in their 60s, and men and women equally (23). It is usually indolent, but in rare cases may become aggressive. Most MALT lymphomas originate in the stomach, and nearly 60% of these cases have a history of *Helicobacter pylori* (23). Treatment varies by stage of disease (early vs. late) and location in the body (stomach vs. lungs) (7,23).

2. **Nodal MZL:**

Nodal MZL is usually slow growing, but has the ability to transform into an aggressive lymphoma (although unlikely). It occurs within the lymph nodes and accounts for 10% of MZL cases (7,24). It typically affects individuals over the age of 60 years, and is more common in women than in men.

3. **Splenic MZL:**

Splenic MZL is a low-grade B-cell lymphoma which develops in the spleen, and less commonly in the bone marrow or blood. It represents ~20% of annual MZL cases (7,25). It is most common among individuals over the age of 50 and affects men and women equally.

1.1.2.2.2 T- and NK-cell NHL

T-cell NHLs represent less than 10% of NHL cases, while NK-cell lymphomas represent less than 1% of NHL cases (7). There are several types of T-cell lymphoma, but they are relatively rare in the population and among families in this study.

1.1.3 Multiple myeloma (MM)

Myeloma is a cancer of plasma cells (2,5–7). Plasma cells are WBCs which produce antibodies which fight infection and disease in your body. Myeloma is associated with the abnormal and uncontrolled growth of plasma cells which are primarily located in the bone marrow (6). Abnormal plasma cells overproduce monoclonal immunoglobulins which interfere with the production of normal healthy blood cells in the bone marrow, causing the immune system to weaken and be susceptible to infections (2,6). Because myeloma occurs at many sites in the bone marrow it is often referred to as multiple myeloma (2,5).

Monoclonal gammopathy of undetermined significance (MGUS) is a common premalignant plasma cell disorder. Individuals diagnosed with MGUS have a 1% annual risk of progression to MM (26). Individuals with MGUS are also at risk of developing WM (26–28).

MM and MGUS occur more frequently in men, older individuals (over 60 years of age) and among persons with an African American ethnicity (6,7,26,29).

1.1.4 Hematopoietic fate

Lymphoma is a cancer that starts in white blood cells (lymphocytes). A lymphoid cancer may arise from any lymphatic tissue, including the lymph nodes, spleen, bone marrow, thymus, adenoids, tonsils or digestive tract (13). The type of lymphoma is dependent on the type of affected lymphocyte (B-cell or T-cell), the maturity of the cell, and several other factors (**Figure 1**) (7,13,30).

The progression of lymphoma has yet to be completely understood. From a hematopoietic stem cell origin to a fully differentiated mature cell, the accumulation of genetic mutations (including translocations) can lead to the transformation and subsequent clonal expansion of a lymphocyte, resulting in a malignant lymphoma (30). Hereditary factors may increase the likelihood of disease development for some families and individuals.

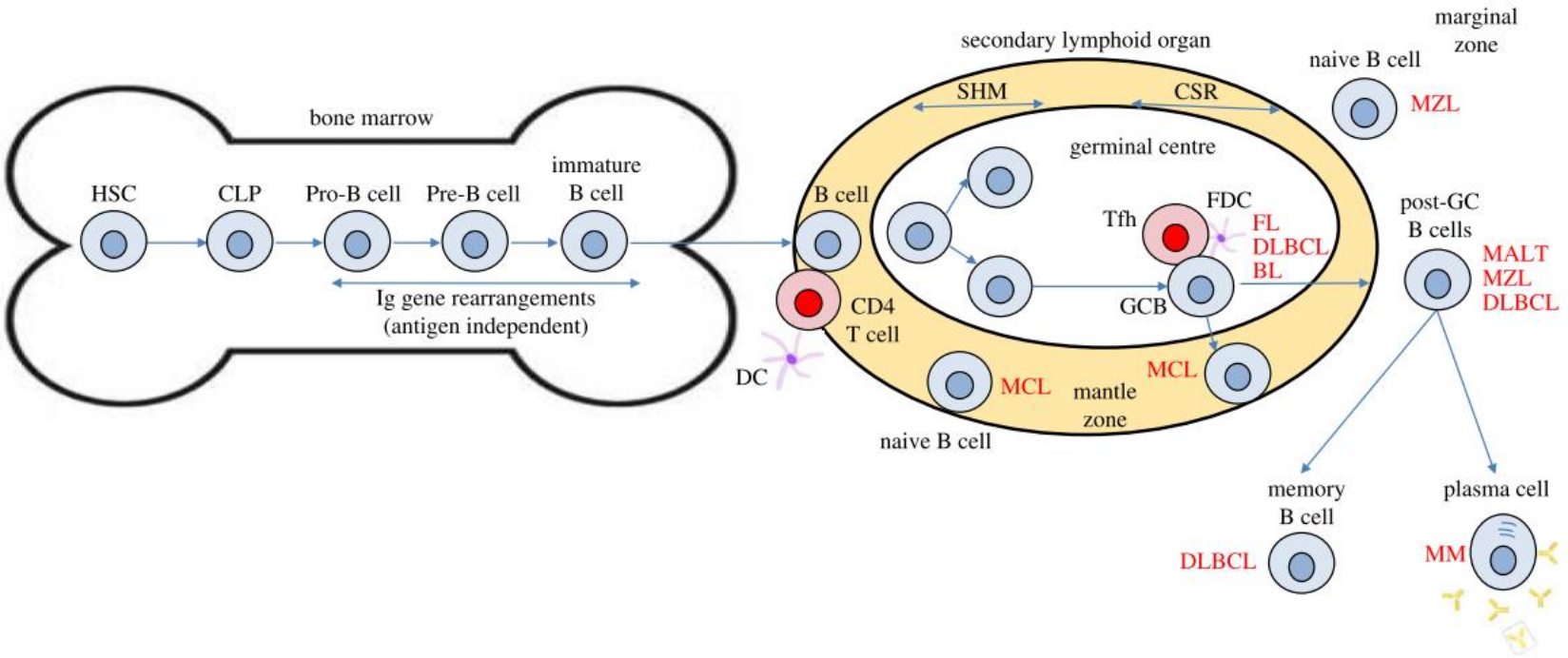


Figure 1: **B-cell development and the origins of B-cell lymphoma.**
B-cell development commences in the bone marrow where immunoglobulin (Ig) genes are rearranged to generate a B-cell receptor. B-cells enter the periphery and congregate in lymphoid tissue where antigen-dependent B-cell development occurs. Upon encountering an antigen, native B-cells become activated and generate germinal centers in which somatic hypermutation and class switch recombination take place. The status of the Ig genetic regions and cell surface markers/proteins of lymphoid cancer cells are indicative of the stage of B-cell development.
Notes: Acronyms in red denote the presumed cell of origin of the indicated lymphoma subtype.
Abbreviations: Ig, immunoglobulin; DC, dendritic cell; CLP, common lymphoid progenitor; HSC, haemopoietic stem cell; GC, germinal center B cell; FDC, follicular dendritic cell; Tfh, follicular helper T cell.
Subtype abbreviations: MALT, mucosal-associated lymphoid tissue lymphoma; MCL, marginal zone lymphoma; MM, multiple myeloma; MZL, marginal zone lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; BL, Burkitts lymphoma.
Figure reproduced from: Malcolm, *et al* (2016) (30) (with permission).

1.2 Hierarchical classification and subtypes

The World Health Organization (WHO) classification of hematopoietic and lymphoid tumours is currently the most extensively used system for the diagnosis of malignant lymphomas (31). The WHO divides lymphomas based on the lineage from which they are derived (B-, T- or NK-cell) and then stratifies the subtypes within each linage based on a combination of morphology, immunophenotype, genetic features, and clinical features (31). Currently, lymphoid cancers are subdivided into 5 subgroups:

1) Mature B-cell neoplasms,

2) Mature T and NK neoplasms,

3) Hodgkin lymphoma,

4) Posttransplant lymphoproliferative disorders, and

5) Histiocytic and dendritic cell neoplasms (32).

For the purposes of this thesis, lymphoid cancers were categorized according to 2008 and 2016 WHO criteria and include: 1) Mature B-cell neoplasms, 2) Mature T and NK neoplasms, and 3) Hodgkin lymphoma.

1.2.1 Mature B-cell neoplasms

Mature B-cell neoplasms, which comprise over 90% of mature lymphoid (B- and T-cell) neoplasms and 83% of all lymphoid cancers (B- and T-cell and HL), consist of MM and NHL subtypes (**Table 1**). The most common mature B-cell subtypes are DLBCL, FL and CLL, which account for 35%, 25% and 10% of lymphomas in western countries, respectively (32,33). Mature B-cell lymphomas may be aggressive (e.g., DLBCL) or indolent (e.g., FL and CLL) (32). The distribution of NHL subtypes varies by geographic region (33).

Multiple Myeloma

According to the WHO, myeloma is classified as a mature B-cell neoplasm (**Table 1**); however, for the purposes of this thesis (and current peer-reviewed publications), MM is considered a separate entity.

Table 1: Mature lymphoid neoplasms according to the 2016 WHO classification of lymphoid neoplasms.

Entity	% observed in the population		
Mature B-cell neoplasms	**83**		
Burkitt lymphoma (BL)		~1	
Chronic lymphocytic leukemia (CLL)		22	
Diffuse large B-cell lymphoma (DLBCL)		23	
Follicular lymphoma (FL)		12	
Hairy cell leukemia (HCL)		~1	
Lymphoplasmacytic lymphoma (LPL)/ Waldenström macroglobulinemia (WM)		2	
Mantle cell lymphoma (MCL)		2.5	
Multiple myeloma (MM)		20	
Marginal zone lymphoma (MZL)		7	
Mucosa-associated lymphoid tissue (MALT)			70
Nodal MZL			10
Splenic MZL			20
Other/ not otherwise specified (NOS)		9.5	
Mature T- & NK-cell	**9**		
Mycosis fungoides (MF)		26	
Primary cutaneous lymphoma (PCL)		9	
Peripheral T-cell lymphoma (PTCL)		52	
Peripheral T-cell lymphoma, NOS			33
Anaplastic large cell lymphoma (ALCL)			22
Cutaneous T-cell lymphoma, NOS			20
Angioimmunoblastic T-cell lymphoma (AITL)			15
Other			10
Other / NOS		13	
Hodgkin lymphoma	**8**		
Classic Hodgkin lymphoma (CHL)		93	
Nodular sclerosis (NS)			65
Lymphocyte rich (LR)			<5
Mixed cellularity (MC)			25
Lymphocyte-depleted (LD)			<1
CHL, NOS			
Nodular lymphocyte predominant (NLPHL)		7	
HL, NOS			
Total	100%		

Notes: "Other" denotes rarer histological subtypes. Population estimated based on SEER (USA) data.
Subtype abbreviations: NOS, not otherwise specified.

1.2.2 Mature T and NK neoplasms

Mature T- and NK-cell neoplasms are less common and account for less than 10% of mature lymphoid neoplasms and 8% of all lymphoid cancers (**Table 1**). Most T- and NK-cell lymphomas are aggressive (32).

1.2.3 Hodgkin lymphoma

HL accounts for 15% of all lymphoid cancers and is further subdivided into CHL and NLPHL. CHL accounts for approximately 90-95% of all HL cases (**Table 1**) and is characterized by the presence of HRS cells (9,11). Based on histological features, CHL is subdivided into four subtypes that vary with respect to age of onset and sex (11,12). NLPHL is uncommon (representing ~5-10% of all HL cases) and is characterized by LP cells (9,11).

1.2.4 Classification of diseases (ICD)

The International Statistical Classification of Disease (ICD) is a global system for healthcare providers to classify and code all diagnoses, symptoms, and procedures. ICD-10 is the 10^{th} revision that was endorsed by the World Health Assembly in 1990 and implemented in Canada between 2001-2018; it represents notable progress in cancer prognostic and diagnostic parameters, which are largely driven by technological advances (31,32). The newest version, ICD-11, was released in 2018 and is expected to replace ICD-10 in Canada by 2022.

The introduction of automated blood counters and the use of flow cytometry in routine clinical practice had a major effect on the diagnostic criteria for CLL. Improved ability to distinguish clonal from reactive processes led to a lowering of the absolute lymphocyte count (ALC) required for CLL diagnosis (15×10^9 ALC/L to 5×10^9 ALC/L). Increased use of automated blood counters in combination with the lowering of ALC diagnostic threshold led to a temporary dramatic increase in the number of individuals incidentally diagnosed with CLL after discovering an elevated ALC from routine blood tests (34). Since individuals with circulating lymphocyte clones were likely to progress to CLL, a B-cell count was used in replacement of ALC (which includes neutrophils, eosinophils, monocytes, basophils, and T-lymphocytes as well as B-lymphocytes). The presence of fewer than 5×10^9 B-cells/L (but greater than the normal level) was defined monoclonal B-lymphocytosis (MBL), an asymptomatic CLL precursor state (35,36). Some research groups observed a decrease in CLL incidence as a result of the change in guidelines in 2008 (37–39), while others have not (40).

1.3 Familial lymphoid cancers

Reports of familial clustering of lymphomas and leukemias have been extensively documented. An emerging theme of lymphoma etiology is that there are both shared susceptibility factors among all lymphomas, and other factors that differ between histological subtypes (41).

Heritability Estimates:

Heritability estimates and genetic correlation attributable to the additive effects of common single nucleotide polymorphisms (SNPs) vary by type of lymphoid cancer (42). Variance explained by SNPs accounts for 21-48% of estimated HL heritability (43–45). Twenty-three independent risk loci explain an estimated 17% of SNP heritability for MM (46–48). Heritability estimates vary by NHL subtype; specifically, common SNPs explain approximately 8% of MZL (49), 21% of FL (49), 9-16% of DLBCL (42,49,50), and 16-34% of CLL (42,49,51–54). There are no heritability estimates available for rare or uncommon lymphoid subtypes (e.g., MCL or NLPHL).

1.4 Descriptive epidemiology

There were an estimated 14,660 new cases of lymphoma (NHL, HL, CLL and MM) and 4,890 deaths from lymphoma in Canada in 2017, the last year for which complete data is available (55–60). Collectively, lymphoma is the fifth most common cancer among Canadians (61). NHL, HL and MM account for 3.4%, 0.1% and 1.9% of all cancer deaths among Canadians (55–58,61). The lifetime probability to developing NHL, HL and MM are 2.4%, 0.2% and 0.9%, respectively (61). The projected 5-year net survival of NHL, HL, CLL and MM is 68%, 86%, 83% and 44%, respectively (61).

1.4.1 Trends in incidence

Lymphoma (NHL, HL and CLL) age-adjusted incidence rates increased annually by 3-4% in high-income countries for 2 decades before stabilizing or declining in the mid-1990s (33,62–72). Since the mid-1990s, age-adjusted incidence rates have declined by 1-5% annually in high-income countries, including Canada, USA, Japan and parts of Europe (33,62–65,70–73). Improved case ascertainment and advancements in diagnosis or change in diagnostic practices may explain a small portion of earlier generalized rises (33,62,69–72,74). The human immunodeficiency virus (HIV) epidemic (62,70–72,75), hepatitis C virus (HCV) infection

(62,70,76), and blood transfusions and transplantations (70) explained a limited proportion of elevated NHL trends. The prevalence of underlying etiological factors may explain the remaining portion of these trends (63,72).

Age-adjusted incidence rates for MM increased annually since 1985 (and continue to rise) in high-income countries, including Canada, USA, Great Britain and other European countries (61,66,67,77–79). Increasing incidence rates may be partially attributed to improved case ascertainment (79–81), improvements in diagnosis (77,79,80,82), and change of environmental factors (80).

1.4.2 Sex

Lymphoid cancers are slightly more common in men than in women (**Figure 2**) (9,29,33,83,84). Most NHL subtypes are characterized by a slight predominance of men, with the exception of MCL (70% male) (33,85), and FL, which is slightly more common in women than in men (18,33). With the exception of NS (7,8), HL cases are more frequently observed in men than in women. Lymphocyte rich HL and NLPHL subtypes are characterized by a 2:1 and 3:1 male predominance, respectively (7,8).

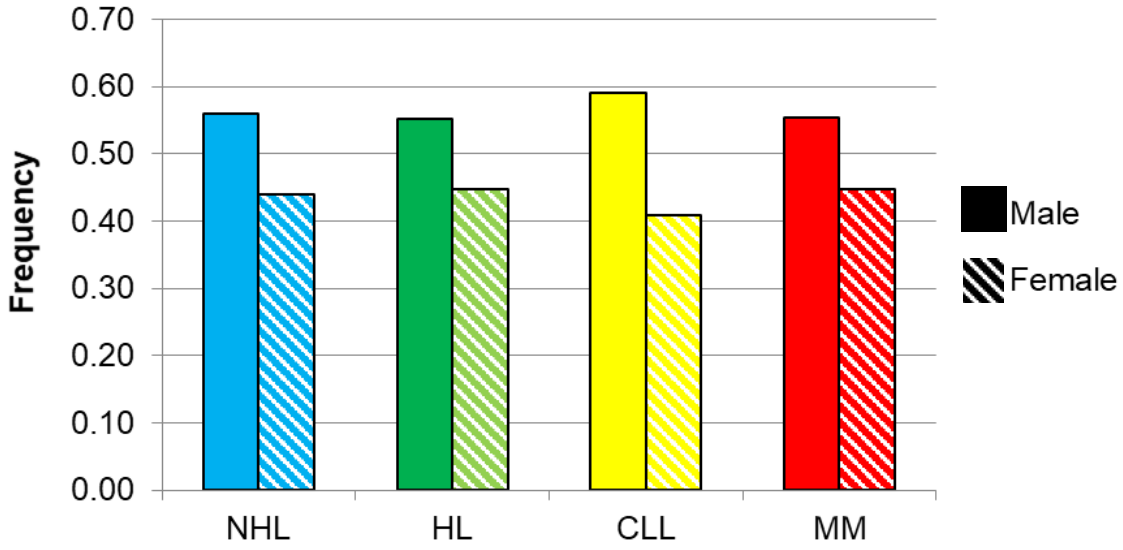


Figure 2: Frequency of male and female population cases of NHL, HL, CLL and MM.

Notes: Created using SEER (USA) population data (86); comparable population data is unavailable for Canada.
Abbreviations: SEER, Surveillance, Epidemiology, and End Results.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

1.4.3 Variation by age

NHL, CLL and MM occur very rarely in children (less than 1% of lymphoma cases; **Figure 3**) (87). Incidence and mortality rates for NHL, CLL and MM increase steadily with older age, peaking between the sixth and seventh decade of life (29,33,87,88). Unlike other lymphoid neoplasms, HL is more frequently a disease of young people between the ages of 15-35 (9,87). HL typically has bimodal age of onset with the first peak at 25-35 years and a second peak after 50-55 years of age (89). HL accounts for 15% of all cancers in young adults globally (9,83).

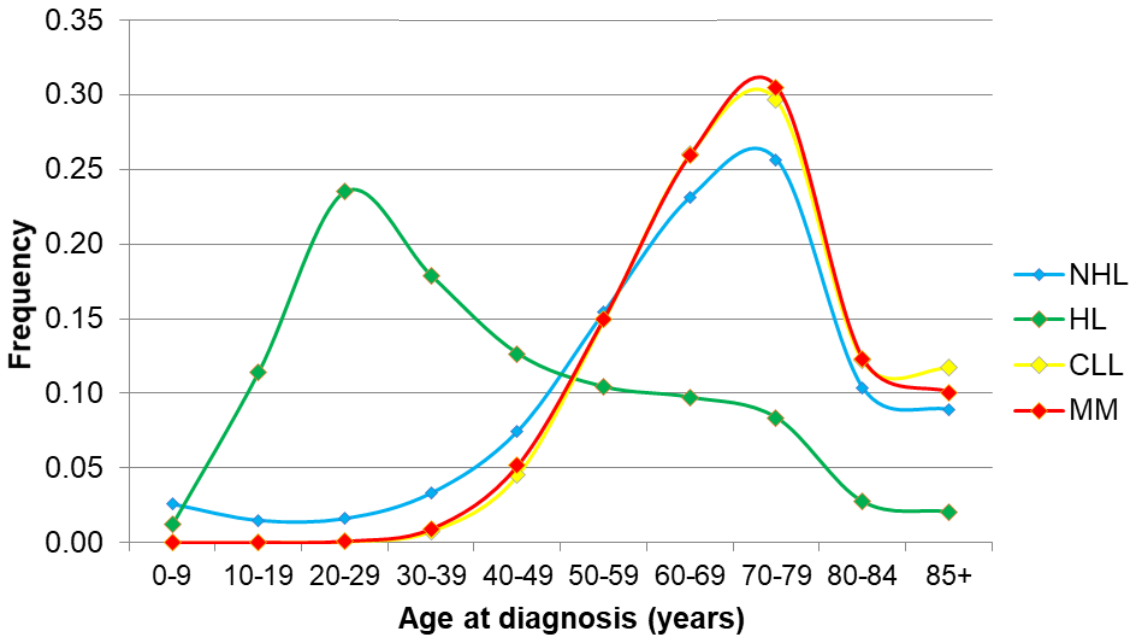


Figure 3: Frequency of age of diagnosis for SEER population NHL, HL, CLL and MM cases.

Notes: Created using SEER (USA) population data (86); comparable population data is unavailable for Canada.
Abbreviations: SEER, Surveillance, Epidemiology, and End Results.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

1.4.4 Race/ethnicity

Racial differences have been observed among many cancer types, including breast, colon, prostate and lymphoid (90). NHL (33,62), HL (83) and CLL (88,91) incidence is highest in Caucasian populations, followed by African American and Asian populations (**Figure 4**). The highest global NHL incidence is in North America and is correlated with high socioeconomic status (SES) (33,92). Elevated rates of CLL have also been reported among specific ethnic groups (e.g.,

Jewish and Ashkenazi Jewish) (91,93) and geographic regions (e.g., Latvia and Russia) (94). MM incidence is highest among African American populations, followed by Caucasian and Asian populations (29,95). These trends persist for generations in migrants to other countries, which suggests a genetic basis for the ethnic variation in incidence (91,96–99).

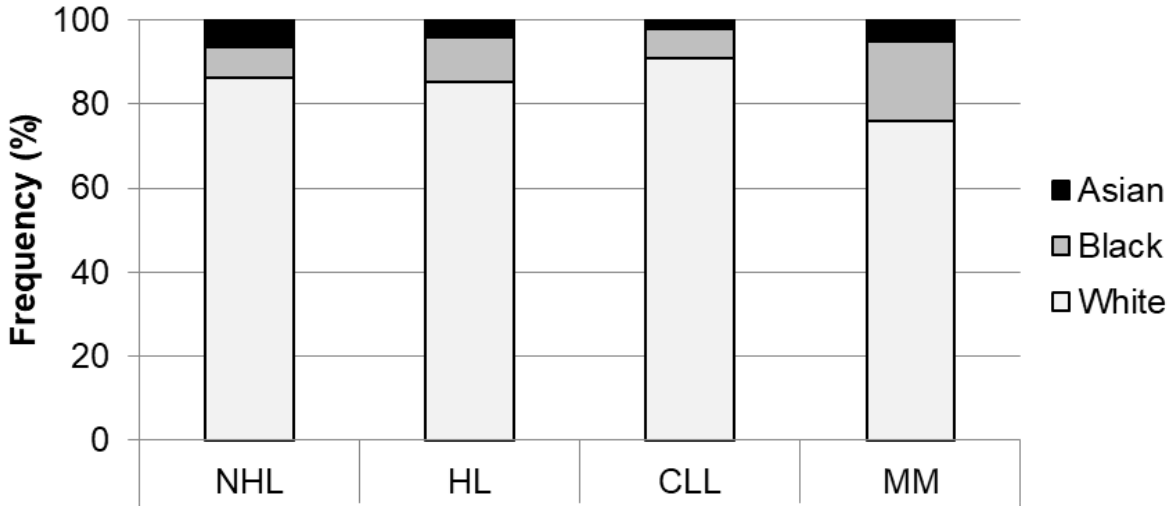


Figure 4: Race/ethnicity population-based frequencies of NHL, HL, CLL and MM.

Notes: Created using SEER (USA) population data (86); comparable population data is unavailable for Canada.
Abbreviations: SEER, Surveillance, Epidemiology, and End Results.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

1.4.5 Geography

Age-standardized incidence rates of NHL (74,100), HL (9,65,74,83), CLL (91,101) and MM (74,77) are higher in high-human development index/high-income countries compared to low-income regions. Differences between geographic regions are multifactorial and may be partially explained by lifestyle diversity (e.g., diet, smoking, SES) (83,101,102), histological subtype (62,101), viral/bacterial exposures (e.g., Epstein-Barr virus (EBV), HIV) (62,74,83), availability of diagnostic practices (77,83), family history (83), and genetics (77,101,103).

1.4.6 Correlation with other neoplasms

Survivors of lymphoma are at a higher risk for developing other solid tumours, such as lung, bladder, brain and breast cancer, and hematological malignancies (104,105). The risk is altered by the age of diagnosis of lymphoma, type and duration of chemotherapy, effect of radiation, and treatment area (62,104). For example, chemotherapies with alkylating agents have been associated with a higher risk of leukemia (104), and radiation to the neck is linked to a higher risk of thyroid cancer (104). Individuals are at greater risk of secondary cancers if they were diagnosed at a young age (104,106), have a family history of cancer (e.g., breast and colorectal) (62,107,108), or had certain tumour cytogenetic abnormalities (e.g., 17p, 6q or 11q deletion) (109).

1.5 Medical history

Factors that affect the immune system play an important role in the risk of lymphoma (74,110). A personal history of autoimmune conditions and some atopic conditions have been associated with a higher susceptibility to lymphoid cancers. Infectious organisms and medical procedures that affect immune function may also affect lifetime susceptibility to lymphoid cancers.

1.5.1 Immune deficiency

The spectrum of immune-deficiencies includes immune dysregulation through immunosuppression or autoimmune disorders and chronic infections, which enhance the predisposition to lymphoma (74,111,112). The genetic inheritance of immunodeficiency is complex and has an estimated heritability of about 10% (112). Acquired immunodeficiencies can arise as post-transplant lymphoproliferative disorders in 1-2% of organ or allogenic stem cell transplant recipients (113) and are caused by immunosuppressive therapies which are used to prevent donor organ or tissue rejection (33,114–116). Chronic infections associated with immunodeficiencies and elevated risk of lymphoma include HIV (33,62,112,117), human herpesvirus-8 (HHV-8) (112), and *H. pylori* (112). The markedly increased risk of lymphoma is attributed to uncontrolled B-cell stimulation and proliferation, which increase the probability of developing a malignant clone (62,118).

1.5.2 Autoimmune disorders

Autoimmune conditions are a strong established risk factor for lymphoma (110,111,119–123); however, the underlying biology of autoimmune-related lymphomagenesis remains unclear (62,124). There are more than 80 recognized autoimmune conditions in which the immune system fails to recognize itself as self and attacks the body's own tissues (33). Large case-control studies suggest that specific lymphoma subtypes are associated with distinct autoimmune diseases, supporting the idea that there are subtype-specific mechanisms of lymphomagenesis (62,123,124). For example, the clustering of B-cell origin lymphomas (125,126) (e.g., HL (126–129) and DLBCL (125,130)) and autoimmune conditions primarily mediated by B-cell responses, are suggestive of a shared etiology (123,124,130–132). Alternatively, the inflammation and severity of the autoimmune condition may contribute to increased risk of lymphoma due to chronic activation or stimulation of B-cells or T-cells (124). For example, aggressive systemic inflammation among rheumatoid arthritis (RA) cases increases chronic activation of B-cells which increases clonal B-cell populations which may lead to DLBCL (124–126,130). Other frequently observed relationships between autoimmune diseases and elevated risk of lymphoma include: Sjögren's syndrome (SS) and risk of B-cell NHL (111,123,125,126,130–135) and HL (127,128), systemic lupus erythematosus (SLE) and risk of B-cell NHL (111,123,125,126,130–132,136–138) and HL (126–129,136), and RA and risk of B-cell NHL (62,125,126,130,138) and HL (126–129).

1.5.3 Atopic conditions

Established positive associations between immune deficiencies and lymphoma risk, and the origin of lymphoma from cells of the immune system suggest that other forms of immune dysregulation, such as atopic diseases, may also be related to lymphoma. Results of population-based epidemiological studies have been inconsistent; however, large case-control and cohort studies (and meta-analyses) have found a reduced risk of lymphoma in association with a history of atopy, including allergies, eczema and asthma (123,139).

1.5.3.1 Allergies

The relationship between allergies or allergy-related conditions and risk for lymphoma has been widely studied. Allergies are a hyperactive state of the immune system that may lead to increased tumour surveillance, thus decreasing the chance of aberrant cell proliferation (140–

142). Associations between allergies and risk of lymphoma are inconsistent, and may reflect the spectrum of lymphoid cancers and subtype-specific associations, categorization/grouping of allergic conditions (e.g., environmental, hay fever, dust; food, peanuts, shellfish; insect bites, bee strings), selection bias, unadjusted confounders (e.g., age, sex, smoking, race), absence of serological data (e.g., Immunoglobulin E (IgE) concentrations) (141,142) and/or reliance on self-reported data (62,142–144). Despite these limiting factors, a few recurring patterns have emerged: a lower risk of lymphoma has been frequently observed among individuals with hay fever (85,130,132–134,139,145–155), while other environmental allergy associations (e.g., dust, animal dander, bee/wasp stings) are inconsistent (139,143,145,146,154–156). Food allergies are associated with a lower risk of lymphoma (130,133,139,145–148,157), and may be attributed to fruit/vegetable (156), seafood (147), and/or nut allergies (156), while no consistent association between lymphoma and medication allergy has been reported (143,146,147,157–159).

1.5.3.2 Eczema

Eczema is a chronic inflammatory skin disease characterized by patches of red, itchy and inflamed skin (139). Meta-analyses support a modest increase in the risk of lymphoma in patients with eczema (138,160–162). Risk tends to increase among individuals with severe eczema (161,162) and those who use topical corticosteroids (158,161–164), which may reduce immunosurveillance (161). Eczema may also be an artifact of early misdiagnosis of cutaneous T-cell lymphomas (e.g., Mycosis fungoides), which develops slowly over years to decades (139,161), and may be treated with high potency topical corticosteroids (162).

1.5.3.3 Asthma

Asthma is one of the most frequent chronic immune conditions that affects children and adults (150,165). Some prenatal risk factors, such as maternal smoking, are well established, while others such as maternal diet, stress, antibiotic use, and delivery by emergency cesarean section are less clear (165–167). Childhood environment, such as infections and exposure to endotoxins, family size and structure, and SES have been shown to play an important role in asthma risk (150,165–168).

The overactive state of immune response associated with asthma may be a surrogate marker of an increased ability of the immune system to recognize and destroy malignant cells, according to the immune surveillance hypothesis (150,169). A few studies observed a protective relationship between asthma and risk of NHL and HL (139,148,150,158,170,171) (e.g., B-cell

NHL (130,133,139,146,172)); however, most studies observed no association between asthma and independent hematological malignancies (e.g., MCL, MZL, MM, or LPL) (85,131,132,147,148,159,160,170–181).

1.5.4 Infectious organisms

Several hematological malignancies have been associated with preceding infections. Long-term infections that cause chronic immune stimulation and overproduction of lymphocytes increase the likelihood of mutations over time (182). Agents such as *H. pylori* and HCV are associated with a higher risk of MALT lymphoma (33,62,183–191) and splenic MZL (62,131,184–186,192), respectively; treating the infectious agent often eradicates the lymphoma (62,184–186). Infections that weaken the immune system, such as HIV, are a risk factor for BL, immunoblastic, and primary central nervous system (CNS) lymphoma (33,62,182,184). Viruses that directly transform lymphocytes have also been identified. Infection with human T-cell lymphotropic virus increases the risk of adult T-cell leukemia/lymphoma, which is most common in Japan and the Caribbean (62,182,184–187,193). HHV-8 infected lymphocytes may cause primary effusion lymphoma and/or multicentric Castleman's disease, both of which are rare lymphoproliferative disorders (LPDs) (33,62,182,184,185,194,195). EBV is a ubiquitous human herpes virus that has been implicated as a cofactor in the development of several malignancies, including primary CNS lymphoma (33,62,184,196), BL (33,62,184,185,196,197), and HL (9,62,89,196,197).

1.5.5 Medical procedures and medical history/exposures

Medical procedures known to affect lymphoid tissue and/or immune capacity can affect long-term health outcomes and susceptibility to immune-related diseases.

1.5.5.1 Blood transfusions

Allogeneic blood transfusions may induce significant immunosuppression in recipients (198,199), which has been associated with a 20% increase in the risk of lymphoma (199,200). In addition to immunosuppression, plausible mechanisms include transfusion of a chemical carcinogen (e.g., DEHP) (200) or viral transmission (e.g., EBV, HCV, HIV) prior to screening implementation (199,200).

1.5.5.2 Childhood infections and vaccinations

Studies of childhood illnesses provide support for an infectious etiology for lymphoma. A personal history of infectious mononucleosis (153,158,201–205), tuberculosis (153,205–207), or malaria (146,153,208) have been independently associated with an elevated risk of lymphoma. Inconsistencies between studies may be attributed to differences in age at infection, type of infection, duration of exposure, and family size (209–211). The relationship between preventative measures, such as vaccinations, and risk of lymphoma have been inconsistent (110,154,155,159,207,208,212) and vary by histological subtype, type of vaccination (e.g., live attenuated, killed whole virus), and adjuvant (e.g., aluminum hydroxide, aluminum phosphate) (110,155,212).

1.5.5.3 Medication use

A complicating factor in studies of associations with medication use is that the underlying medical conditions that prompt treatment, rather than the medications themselves, may explain the observed associations with lymphoma risk (62). Medications such as antibiotics and aspirin, and a longer duration of use (62,163,164) have been associated with a higher risk of NHL (145,155,156,213), HL (213) and MM (214,215). In contrast, no association between duration and use of medication has also been observed (29,145,163,216). Associations with antibiotic use may reflect an underlying susceptibility to infections, rather than a direct lymphogenic effect of antibiotics themselves.

1.5.5.4 Tonsillectomy

The tonsils are the first lymphoid tissue barrier which defends against foreign pathogens that enter via the mouth (217–219). An altered immune response of the tonsils and/or a tonsillectomy may further impair immune function and increase risk of immune-related diseases (217,218). Recipients of a tonsillectomy have been associated with an increased risk of several autoimmune diseases (217) and hematological malignancies, including NHL (146,155,220) and HL (146,221,222) subtypes. Consistent with the declining immunological role of the tonsils from childhood to adulthood (218), the risk was more pronounced if the procedure occurred during early childhood (146,210). Several of the autoimmune diseases seen among tonsillectomy recipients are also independently associated with an elevated risk of lymphoid cancers, including Graves' disease/hyperthyroidism, Hashimoto disease/hypothyroidism, RA, and SS, suggesting a similar etiology (217).

1.5.5.5 Appendectomy

Pathogenic mechanisms and causative factors involved in appendicitis are not well understood; however, infectious, inflammatory, dietary and genetic factors are thought to be involved (223,224). Associations with cancer and a history of appendicitis or an appendectomy are inconsistent (156,224,225). Some studies report an elevated risk of HL (225–228), but not for MM (177,229), NHL (158,225,229–231), or common NHL subtypes (225,231,232).

1.5.5.6 Splenectomy

The spleen is a reticuloendothelial organ with important hematologic and immunological functions, including clearance of bacteria from the blood and generation of an immune response to foreign pathogens (233–237). Long-term risks following a splenectomy include hematological disorders, such as venous thromboembolism (234–239), overwhelming post-splenectomy infection (234–238,240,241), and hematological cancers, including NHL (145,233,242), HL (233,242) and leukemia (233,242,243).

1.6 Lifestyle and personal factors

Environmental agents and lifestyle exposures during all stages of life may contribute to lymphoid cancer risk. Specifically, childhood exposures to infectious agents may affect immune development and maturation (62). Health and lifestyle factors such as nutrition, smoking, alcohol intake and education capture knowledge-related behaviours and measures of SES. In addition, reproductive and hormonal factors have been shown to lower risk of lymphoma and may partly explain difference in sex distribution of lymphoid cancers (62).

1.6.1 Early childhood

Early life exposures to infections are fundamental to the development and maturation of the immune system (170), and a relative lack of infections may predispose to adult-onset immune-related disorders. Early life factors that stimulate a child's immune system may be protective for such diseases. Measures of SES and family structure that influence age at, and extent of, exposures may be surrogates for early infectious exposures.

1.6.1.1 The hygiene hypothesis

The hygiene hypothesis proposes that early childhood infections may protect against adult-onset immune-related diseases, including allergies, autoimmune conditions and some hematological malignancies (168,244–246). Delayed or lack of infectious exposures during childhood may inhibit a child's immune system from maturing optimally, which increases susceptibility to immune-related disorders (168,247). Measures of family structure (birth order and sibship size), household crowding, day care attendance, and SES relate to the hygiene hypothesis as they affect age, extent and response to infectious exposures (248–251). Number of siblings and birth order can be indicators of age and frequency of exposure to infectious agents during childhood (168,250,251). Specifically, first-born children are exposed to infections at later ages than their younger siblings, who may contract infections from their older siblings (39,168). Likewise, the number of children in a sibship relates to risk of infections during childhood, with larger sibship size correlated with higher infectious burden (39,168,248).

Birth order and sibship size

Findings of studies of lymphoid cancer risk and family structure (birth order and sibship size) have been variable. Several studies found an elevated risk of NHL among individuals with later birth order (147,252–254) or larger sibship sizes (147,252,253), while other studies observed the opposite (157,170,255,256), or no association (158,191,254,257,258). Variability among studies may be attributed in part to differences between heterogeneous subtypes and small sample sizes (147,157,158,191,252–254,256–258). Associations observed between birth order and risk of DLBCL (157,191,252,256,257), FL (157,191,252,253,257), CLL (157,252,253), and T-cell NHL (157,170,191,252,256) have been inconsistent, while no association between birth order and rarer subtypes, such as MM (157,191,258) and BL (170,255), have been consistently observed. Other subtypes, such as MCL, MZL, LPL/WM have not been analyzed. Fewer population-based studies examine sibship size and risk of NHL. Three case-control studies observed an elevated risk of NHL among larger sibship sizes (147,252,253), while 2 cohort and 3 case-control studies observed no association (158,191,254,257,258). Associations observed between sibship size and risk of DLBCL (191,252,257), FL (191,252,253,257), CLL (252,253), T-cell NHL (191,252) and MM (191,258) are inconsistent, while other subtypes (e.g., MCL, MZL, LPL/WM) have not been analyzed.

Several studies observed an inverse relationship between birth order and risk of HL (191,201,259,260), but several others found no association (157,170,202,255,258,261). With the

exception of two small case-control studies (201,262), most population-based studies observed no relationship between sibship size and risk of HL (191,258–260).

Daycare

Daycare attendance (and crowded environments) are indicators of exposure and transmission of pathogenic agents (211,263,264). Children in day care are more likely to experience respiratory tract (264–266) and other infections (266) and less likely to develop adult-onset NHL (252) or acute lymphoblastic leukemia (ALL) (211,245,267).

1.6.1.2 Socioeconomic status: Parental education and family income

Strong indicators of childhood SES include parental education and income, as they capture behaviours that influence the age, extent and response to infectious agents (249,268–270). Generally, lower childhood SES may be indicative of crowded or higher-density living conditions, which increases the likelihood of early and more frequent infectious exposures (179,268). In contrast, individuals of higher SES tend to live in cleaner environments, which may delay infectious exposures. Studies regarding childhood environment are inconsistent. An elevated risk of lymphoma has been reported among individuals with higher maternal (201,203,256,271,272) or paternal education (256). In contrast, individuals with lower parental education (179,249,256,273) has been associated with a higher risk of lymphoma. Other indicators of childhood social environment such as family income (179,274–276), house location (170,252,253,277,278), type of house (e.g., single family home, high density dwellings) (179,201,252,253), and sharing a bed (147,254) or bedroom as a child (147,157,254) were inconsistent or not associated with risk of lymphoma. Discordant findings among studies may be attributed to study design bias, and differences in parental sex, country, participation rates, and subtype of lymphoma (179,201,203,249,252,256).

1.6.1.3 Farm residence

Living on a farm during childhood may lead to earlier and more diverse exposure to pathogens, zoonotic viruses and antigenic agents that encourage an active immune response and strong immune competence (149,166,170,212,277). Consequently, farm residence during childhood is associated with a lower prevalence of hay fever, asthma, allergies (166) and some leukemias and lymphomas (149,152,170,253,277,279,280). These protective effects may be

attributed to animal contact (149,170,212,280,281), frequent farm visits (149,170,212), and early age at exposure (170,212). In contrast, several studies have observed an elevated risk of lymphoma among childhood farm residents, which may be attributed to pesticide/herbicide (281,282), organic solvent (283), or livestock exposures (281,282). Assessing farming-related exposures is complex because these exposures are confounded by other farming-related practices. For example, the type and application of herbicides and insecticides differ between crop and livestock farmers. Discrepancies between studies may be attributed age of exposures (e.g., early childhood, teenager, adult) (170,253,279,281), type of exposures (e.g., pesticides (281,282), herbicides (281,282), livestock (170,279–282,284), crops (278–280,282)) duration of exposures (170,253,281,282), endotoxins (257), and subtype of lymphoid cancer (170,252,253,278,281,284,285), or simply to chance.

1.6.2 Anthropomorphic factors

Anthropomorphic factors which affect adiposity (e.g., undernutrition, overnutrition) are known to suppress immune function and may therefore contribute to lymphoid cancer risk (286). Although the results of independent studies have been inconsistent, meta-analyses suggest that body mass index (BMI) and weight in early adulthood may be more relevant to lymphoma development than current BMI and weight (123,130,133,174,286–288). Larger body size, as measured by BMI (123,130,133,173,286–292), weight (130,286,290,293) and height (130,133,174,175,181,286,288,290,293,294), is generally associated with a higher risk of lymphoma and some subtypes of lymphoid malignancy (e.g., DLBCL, but not CLL). Inconsistent associations between anthropomorphic factors and lymphoid cancer subtypes suggests etiological heterogeneity and that immune dysfunction may be more relevant for lymphomagenic mechanisms in some (but not all) subtypes or that differences in disease etiology, behaviour, and aggressiveness of tumours may explain differences in anthropomorphic factors and subtypes (286,290,291). Increased risk of lymphoma among individuals of taller stature (130,133,174,175,181,286,290,293,294) may reflect cumulative exposure to hormones/growth factors, nutrition, and SES (286,290,293); however, the exact mechanism is not understood.

1.6.3 Education

Education is frequently used as an indicator of SES in epidemiology as education captures knowledge-related assets of a person, is a strong determinant of future employment and income, and indirectly measures environment or occupational risk factors (180,222,269,295). There is limited and contradictory information on the association between education and risk of lymphoma. Higher educational attainment has been associated with an elevated risk of NHL, including T-cell and CLL subtypes (295,296). In contrast, education was inversely associated with risk of DLBCL (295) and MM (180,268), while no association was observed for FL (295). There is no association between HL and educational attainment (179,295,296), with the exception of a small hospital-based case-control study (160 cases, 185 controls), which observed a significant positive association (222).

The relationship between education and cancer risk may be influenced by age of diagnosis, treatment regimes, childhood SES, and sex. Lower educational attainment may be attributed to neurocognitive impairments from standard cancer treatment regimes. Some survivors of childhood HL cancers have poor short- and long-term memory, attention and reduced brain integrity which negatively impacts postsecondary education and employment (297). Reduced brain integrity/function has also been observed among adult breast cancer (298–300) and primary CNS lymphoma (301,302) survivors, although, adult-onset disease are less likely to affect educational attainment than childhood cases.

1.6.4 Tobacco/smoking

A majority of studies on tobacco use do not support a causal association with NHL or MM (29,33,62,180,303); however, several studies report a positive dose-response relationship between smoking/chewing tobacco and FL risk (123,133,303,304).

1.6.5 Alcohol

Current and lifetime alcohol (wine (123,130,131,305), beer (123,130), but not spirits (130,180,305)) use was shown to lower the risk of NHL (33,123), HL (306–308), MM (180,308) and several NHL subtypes (DLBCL (123,130), FL (123,133), CLL (308,309), MZL (123,131), BL (123,174), peripheral T-cell lymphoma (123), and mycosis fungoides (123)). Risk did not vary by

frequency (daily intake) or duration (years) (62,85,130,305) which makes it less likely that these associations are indicative of a causal relationship.

1.6.6 Nutrition/diet

Few reproducible associations between lymphoma and food or micronutrient intake have been observed. Diets high in meat (total (33,62,177,310–313), processed (313–316) or red (292,294,316,317)), dairy (33,62,292,311–314,318,319), and total (62,311,312,315,319,320) or saturated (292,311,319–321) fats have been associated with an elevated risk of lymphoma (177,292,294,310–317,320,321). Consumption of citrus fruits (e.g., oranges, grapefruits) (177,289,313,318,319), cruciferous vegetables (which contain indole-3-carbinol, isothiocyanates, and high Vitamin C) (289,292,313,319,322), and high overall vegetable intake (e.g., cruciferous, dark green, dark yellow, green leafy) (177,289,292,313,318,322) may reduce the risk of some lymphomas (29,33,62,177,289,292,312,313,318,319,322).

1.6.7 Coffee and tea

Few studies report a significant positive association with high daily coffee consumption and risk of lymphoma (313,323), while there is no apparent association between tea consumption and lymphoma (62,180,313,318,324).

1.6.8 Reproductive and hormonal factors

Greater exposure to female reproductive hormones, particularly from multiple pregnancies or exogenous hormones from oral contraceptives or menopausal hormone therapy, have been associated with a lower risk of some lymphoid cancers (130,138,325–327), particularly DLBCL (325,326). Greater estrogen and progesterone levels lead to a reduction in B-cell lymphopoiesis, differentiation and proliferation (325). Younger age at first use and increased duration of oral contraceptive use also saw lower rates of NHL (325–327).

1.7 Occupation and environment

Although no single environmental entity has been convincingly established as a cause of lymphoma, several occupational and environmental exposures have emerged as likely candidates. Exposure to hair dye through occupation or personal use has been associated with an excess risk of lymphoma (130,175,187,328–332). Several studies report positive associations between construction (85,130–132,152,173,231,277,333–337), farming (123,130,173,175,180,181,252,280,282,334,335,338), and medical/ healthcare (132,296,333–335,337,338) occupations and risk of lymphoma; however a specific underlying exposure has not been identified (277). Results of studies of chemical (29,276,339), solvent (29,276), asbestos (29), pesticide (29,276), herbicide (29,276), or livestock (181,212,252,280,282,335,338) exposures have been inconsistent, and no consistent exposure-response pattern has emerged. Moderate to high ultraviolet radiation exposure is inversely associated with risk of lymphoma (62,123,340) and several histological subtypes (85,123,130,131,133,152,175,341–343) (e.g., DLBCL (123,130), FL (123,133), MCL (85,123), MZL (123,131), CLL/SLL (123,175,342,343), T-cell (123,152) and HL (307,341)). Assessing life-time exposure of occupational and environmental agents is challenging; discrepancies between the literature may be attributed in part to limitations of study design, variability between specificity of exposure, assessment/ measurement of exposure, and recall bias (62).

1.8 Genetic susceptibility

Several studies have investigated familial predisposition and germline susceptibility loci in families with lymphoid cancers. Some genetic risk factors appear to be shared among most lymphoid cancers, while other genetic factors appear to be specific to a single lymphoid cancer type, or subtype (41).

1.8.1 Hereditary factors (organized by study design)

1.8.1.1 Twin studies

The concordance of a disease phenotype (e.g., cancer) between monozygotic (MZ) and dizygotic (DZ) twins provides information on hereditary and environmental causality (344). If the

concordance rate of a phenotype in MZ twins (who share all genetic variants) is greater than that for DZ twins (who, on average, share 50% of their genetic variants), then there is evidence for a genetic component (41,344). In contrast, if the concordance is similar among MZ and DZ twins, then shared environmental effects are likely to be more important (41,344).

In a cohort study of 44,788 pairs of twins from Scandinavia (344), there was an excess of concordant MZ twins compared with DZ twins for leukemia (attributed to CLL) (344,345) and MM (344), while no concordant pairs were observed for NHL (344). A study of 179 pairs of MZ twins in which one had HL observed a 100-fold higher risk (standardized incidence ratio) of HL in the second twin, relative to the general population, while none of the 187 pairs of DZ twins became concordant for HL (346). The same study observed a 23-fold higher risk of NHL in the second twin among 110 MZ pairs ascertained because one twin had NHL, compared to population rates (346). Among 164 pairs of DZ twins in which one had NHL, there was a 14-fold higher risk in the second twin, relative to the expected population frequency (346). These observations suggest shared environment and genetic components may be involved (41,346).

1.8.1.2 Case-control and cohort studies

Patterns from case-control and cohort studies are generally consistent and provide strong evidence for familial predisposition to lymphoma (41). An increased risk of NHL was observed among those with a first-degree relative with NHL, HL and several NHL subtypes (e.g., CLL, DLBCL, FL, MZL, MCL, peripheral T-cell lymphoma) (41,123). Similarly, a higher risk of HL was observed among individuals with a first-degree relative with NHL and HL and several NHL subtypes (e.g., CLL, DLBCL, FL, MZL) (41,123). Some lymphoid cancers had a stronger familial component than others, and include: a family history of CLL and risk of CLL (ORs = 2.4-8.5) (123,347), family history of DLBCL and risk of DLBCL (OR, 9.8; 95% CI, 3.1-31) (348), family history of FL and risk of FL (ORs = 3.9-4.0) (348), a family history of LPL/WM and risk of LPL/WM (OR, 20; 95% CI 4.1-98) (349), a family history of HL and risk of HL (ORs = 3.1-8.8) (261,348,350,351), and a family history of CLL and risk of HL (ORs = 2.1-6.3) (348,351,352).

1.8.1.3 Genome-wide association studies (GWAS)

DLBCL

Common SNPs explain ~16% of the variance in DLBCL risk (42,49,50,353). To date, 8 SNPs from 7 loci have been associated with DLBCL in European populations (41,50). Two of

these SNPs are in linkage disequilibrium with a super-enhancer that interacts with the promoters of *CD86* and *AZI2*, which encode proteins required for T-cell activation and survival, and antiviral innate immunity (354). Another locus of interest is located adjacent to the *MYC* oncogene (8q24) (the deregulation of which is observed in BL and some DLBCL cases) (50,355). Two signals reported in 3q27 and 14q32 among East Asian and Japanese GWAS did not replicate among individuals of European descent (41,50). Risk variants for populations of European ancestry did not replicate among small GWAS of Chinese or Japanese populations (356,357). Failure to replicate GWAS across ethnic populations may reflect differences in linkage disequilibrium and not differences in the contributions of the underlying genetic factors.

FL

Common SNPs explain 21% of FL heritability (49,353). Most of the variants associated with FL reside in the human leukocyte antigen (HLA) region on 6p21.33 and 6p21.32 (41,353,358–361). In a meta-analysis the HLA region showed overwhelming association with FL, with 8104 SNPs achieving genome-wide significance (41,359). In addition, five non-HLA loci have been associated with risk of FL. In addition to cell-adhesion and motility, these genes are involved with B-cell migration and apoptosis (e.g., *BCL2, CXCR5*) (360) and the regulation of *MYC* (e.g., *PVT1*) (360), which make them plausible candidates in the etiology of FL (41,359).

MZL

Common SNPs explain approximately 8% of MZL heritability (49). One GWAS study has identified two distinct loci at 6p21.32 (intragenic to *BTNL2*; HLA class II) and 6p21.33 (near HLA-B; HLA class I) (362). *BTNL2* is highly expressed in lymphoid tissues and regulates T-cell proliferation and T-cell mediated responses, whereas *HLA-B* plays a role in immune response (362). Both loci are associated with autoimmune disease (e.g., RA, or SS and RA) (362,363) which suggests possible shared genetic or environmental factors or biological mechanisms that contribute to MZL (41,362,363).

CLL

The estimated contribution of all common variations to the heritability of CLL is 16-34% (42,49,51–54). To date, common genetic variation at 42 loci have been shown to affect risk of CLL (52,364). Many of the SNPs are located in or near genes that are central to B-cell development (e.g., *IRF4*, *RELA*) (41,52,353,365,366), immune response (e.g., *NEDD4, PIAS4*)

(41,52,365), apoptosis (e.g., *BCL2, CASP8*) (41,52,53,365) or maintenance of chromosome integrity (e.g., *TERT, POT1*) (41,52,53,364,365). SNPs at some loci have been associated with autoimmune conditions, such as *BAK1* and SLE (51,52,365). CLL cases diagnosed at a younger age tend to carry a greater number of risk alleles (41,364), suggesting that early-onset CLL is enriched for genetic factors (364).

HL

Common SNPs explain 21-48% of HL heritability (43–45,367). The estimated heritability of CHL, NS HL, and MC HL attributable to common variation is 24%, 25% and 22% (368). To date, 8 GWAS and 1 meta-analysis has been published on HL risk (41,367,369,370). Eighteen independent risk loci for HL have been identified (369), and the strongest findings have been for SNPs mapping to HLA class II (in close proximity to HLA-DRA and HLA-DRB1) (368–370). Several SNPs, including rs6903608 (near HLA-DRA) and rs2281389 (HLA-DPB1) were associated with early onset HL (41,370). Non-HLA associated loci include genes involved in regulation of germinal center B-cells (e.g., *BCL6, MYC*) (368–370), T-cell differentiation and function (e.g., *EOMES, SOSC1, PVT*) (41,367–370), and NF-kB activation (e.g., *AZI2*) (368–370). Variation at several HL risk loci have also been associated with risk of autoimmune disease (367,368), suggesting common genetic susceptibility and/or biological pathways (368). No GWAS has been performed for NLPHL subtypes.

MM

The heritability of MM ascribable to all common variation was approximately 17% in European populations (46–48). To date, 23 independent risk loci for MM have been identified, with an additional locus for t(11;14) translocation MM (47,371). No associations between sex or age at diagnosis were found (47,371). Associated SNPs were enriched in regions of active chromatin and were in or near genes involved in B-cell and plasma cell differentiation (e.g., *ELL2, TNFSF13B*) (47,372), cell cycle and chromosome integrity (e.g., *POT1*) (47,371), chromatin remodeling (e.g., *SP3*) (47,372), and apoptosis/autophagy (e.g., *KLF2*) (47,371,372).

Summary

With two exceptions (*HLA*-6p21 and *MYC*-8q24.21), most genetic loci identified from GWAS are associated with specific subtypes of lymphoid cancers (**Figure 5**). Several risk alleles in the HLA region have been associated with CLL, FL, DLBCL, MZL, and HL disease risk

(41,50,368–370,373). 8q24 has been associated with several B-cell entities, including CLL (*PVT1, MYC*) (52), FL (*PVT1*) (360), DLBCL (*PVT1, MYC*) (41,50,370,373) and HL (*PVT1, MYC*) (41,368–370). Collectively, GWAS-established loci are common variants (minor allele frequency (MAF) > 5%) and have small effect sizes, which is suggestive of a polygenic model for disease susceptibility in the general population (41).

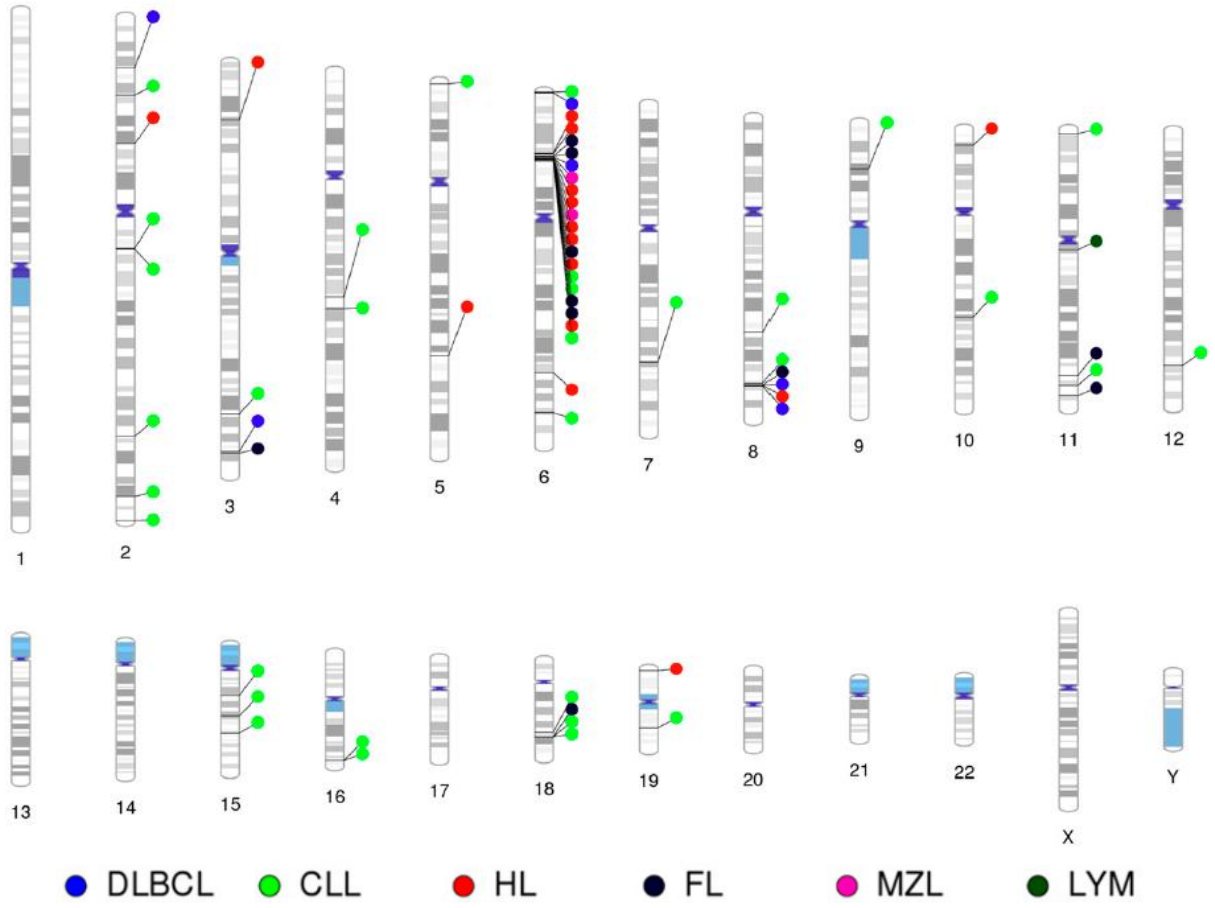


Figure 5: GWAS-discovered loci for lymphoma subtypes mapped to chromosome locations.
Notes: Except for 6p21 and 8q24, there is minimal overlap of loci for subtype-specific susceptibility.
Abbreviations: DLBCL, diffuse large B-cell lymphoma; HL, Hodgkin lymphoma; MZL, marginal zone lymphoma; Lym, lymphoma; CLL, chronic lymphocytic leukemia; FL, follicular lymphoma.
Figure reproduced from: Cerhan and Slager, 2015 (41) (with permission).

1.8.1.4 Family studies (linkage, germline susceptibility studies)

Linkage studies use multiple-case families or sibling pairs to identify chromosomal segments that cosegregate with the disease phenotype. Of the few published linkage studies in lymphoid cancer, several HLA alleles have been associated with risk of HL (374,375), CLL (376) and WM (377). Beyond HLA, linkage studies in CLL, HL and WM families have not definitively identified genes with large effects. For CLL, evidence for linkage was observed on regions of chromosomes 1q, 2q21.2, 3q, 6q, 6p22.1, 11p11, 12q, 13q and 18q21.1 (376,378,379); however, these findings were not replicated, likely due to limited power, small numbers of families/cases or locus heterogeneity (376,378,379). Among 44 high risk HL families, strong evidence for linkage was observed on chromosomes 2, 3, 4, 7, 11 and 17 (380). Among 11 high-risk WM families, evidence for linkage was found on chromosomes 1, 3, 4 and 6 (377). Studies examining high-risk multiple-case families with FL, DLBCL or other homogeneous NHL subtypes has not been published.

1.8.1.5 Candidate gene studies

Candidate gene studies have highlighted several variants in pathway regulators of immune function, cell cycle/proliferation, apoptosis, DNA repair, and carcinogen metabolism (41,381). Some variants are shared between lymphoid cancer subtypes (381–384); however, many associations fail to replicate possibly due to study design, confounding by race/ethnicity, and small sample size (41,381–383,385).

1.8.1.6 Involved pathways

Several studies implicate the role of genetic variants that promote B-cell survival and growth with increased risk of NHL (386). Associations between genes involved in energy regulation and hormone production and metabolism have been identified.

1.8.1.6.1 DNA integrity and methylation patterns

Genes involved in DNA double strand (ds) break repair: Individuals carrying mutations in genes involved in DNA ds break repair are at an elevated risk of LPDs, which underscores the relevance of this pathway in lymphomagenesis (386). SNPs in genes that hinder DNA repair mechanisms increase the likelihood of neoplastic lesions that are relevant to lymphoma. Several polymorphisms in DNA ds break repair and non-homologous end joining genes have been associated with a higher risk of lymphoma, including variants in *ATM, NBS1, RAG1,* and *BRCA1* (381,386–389).

One-carbon metabolism (epigenetic regulators): Genetic variants that affect methylation processes may increase susceptibility to lymphoma by hypo- or hypermethylating proto-oncogenes or tumour suppressor genes, or through viral re-activation (386,390). Genetic variants in folate metabolic pathways (e.g., *MTHFR*) (386,390,391) may also influence DNA methylation, synthesis and repair mechanisms (386). Polymorphisms in epigenetic regulator genes including *TYMS* and *MTR*, have been associated with risk of several lymphomas (386).

1.8.1.6.2 B-cell survival and growth

Pro-inflammatory cytokine genes: SNPs in tumour necrosis factor (*TNF*) and interleukin *(IL)-10* and *-6* genes have been associated with lymphoma and autoimmune diseases (386,392–395). TNF-α is a pro-inflammatory immunoregulatory cytokine and key mediator of lymphocyte responses, NK cell activity and dendritic cell maturation (392). IL-10 and IL-6 are immunoregulatory cytokines (386,389). *TNF* variants have been associated with an increased risk of NHL (386,392), diffuse large-cell lymphoma (DLCL) (386,392), and autoimmune diseases such as RA and SS (386,392,393). *IL-10* variants and haplotypes have been associated with a higher risk of NHL (386,389,392,393) or DLCL (386,392,394), and *IL-6* polymorphisms have been positively associated with T-cell NHL (386,393) and HL (386), as well as type 2 diabetes mellitus (386).

Innate immunity genes, toll-like receptor and caspase recruitment domain family genes: *NOD2* and *TLR4* are vital pro-inflammatory mediators as a first line of defense against viral and bacterial infection, providing non-specific protection against numerous pathogens (386,396). *NOD2* variants are associated with the development of Crohn's disease and MALT lymphomas (396). A premature stop codon in *CARD15* (among other variants) was associated with autoimmune disorders (Crohn's disease and psoriasis) (386,397) and excess risk of lymphoma (386,396,397), and MALT lymphoma subtypes (386,396). Polymorphisms in *CARD15* may contribute to lymphoma susceptibility through the chronic activation of TLRs which activates NF-kB and pro-inflammatory cytokine release/responses (395,397). Variants of *TLR4* have been associated with risk of MALT lymphoma, DLCL, FL, and HL (386,395,397).

Oxidative stress: Reactive oxygen species are implicated in several inflammatory conditions (e.g., inflammatory bowel disease) and in cancer risk (386,398). SNPs in the functional domain of *NOS2A* (a free radical mediator) (386,398) have been associated with a higher risk of NHL, DLCL, FL, and MALT lymphoma, as well as with gastric cancers (386,398). Variants in *SOD2* (which encodes a protein that binds to superoxide by-products) was associated with an

increased risk of B-cell lymphoma (386,398). Other candidate genes with known alterations of functional significance in oxidative stress include: *MPO, SOD2, CYBA, MPO, GPX, OGG1,* and *PPARG* (398).

1.8.1.6.3 Sex hormone production and metabolism

SNPs that regulate sex hormone production and metabolism have also been associated with lymphoma (386,389). Prolactin and estrogens are important in female reproduction and also function as immune modulators that regulate apoptosis and activation and proliferation of immune and B-cells (386,389). SNPs in the *CYP17A1* gene, which encodes an enzyme involved in estrogen and testosterone synthesis, have been associated with an elevated risk of NHL and DLCL subtypes (386,389). SNPs in estrogen and prolactin regulatory genes, such as *COMT* and *PRL*, have also been associated with a higher risk of lymphoma in men and women (386,389).

1.8.1.7 Human leukocyte antigen (HLA)

The search for genetic risk factors in lymphoid neoplasia has continually highlighted the HLA complex (399). HLA molecules initiate the adaptive immune response by presenting pathogen and tumour-derived peptides to T-cells (399). Several case-control studies have documented associations between HLA loci and risk of DLBCL (50,130,392,399), FL (41,353,358–361), CLL (366,399), or HL (368–370). Generally, downregulating the expression of HLA was associated with a higher risk of lymphoid cancers; however, the opposite has also been documented (399). Conflicting results may be attributed to smaller sample size, combined analysis of NHL subtypes or different ethnic groups.

1.8.2 Telomeres

Telomeres are complexes of tandem repeats located at the ends of chromosomes that aid in chromosomal stability. Telomeres progressively shorten with each cell division, leading to chromosomal instability and subsequent cell death with aging. Telomere shortening may be accelerated by various lifestyle factors (e.g., smoking, obesity, lack of exercise, unhealthy diet, stress) that may be associated with age-related illnesses and premature death (400). Shorter telomere length may increase the risk of some leukemias and lymphomas and are of poorer outcomes in patients who develop these malignancies (400–403).

1.9 Thesis hypothesis and objectives

The heterogeneous nature of lymphoma poses a challenge to understand disease etiology. Large population-based epidemiological studies have helped to establish lymphoid cancer risk factors, such as advanced age, male sex, compromised immune function (e.g., immune deficiency diseases, autoimmune conditions, allergies), and a family history of LPDs; however, early childhood lifestyle (e.g., SES, parental education, family income, family structure, etc.), personal (e.g., anthropomorphic factors, education, diet, smoking, etc.), and occupational or environmental agents (e.g., hair dye, pesticides, ultraviolet radiation, etc.) may also affect lymphoid cancer susceptibility. The characterization of lymphoid cancer risk factors is mainly derived from population-based epidemiological studies. However, there remains a gap in our understanding of familial lymphoid cancer etiology. To date, one small study examined clinical and environmental factors in 103 familial WM cases. Characterizing more than 500 familial cases (from over 200 multiple-case families) with heterogeneous lymphoid cancer types, may aid in our understanding of lymphoid cancer etiology, and establish associations in the context of multiple-case families.

Common polymorphisms (allele frequency > 5%) from more than 40 loci have been associated with risk of sporadic lymphoma (41). Clustering of lymphoid cancers in families has been observed (41,123,261,347–352), however, genetic susceptibility factors are largely unknown. To date, a handful of linkage studies in high-risk homogeneous NHL or HL families have identified rare coding variants (374–380). Examining a family with 4 lymphoid cancer cases may identify additional familial lymphoma susceptibility genes. The phenotypes of this family are interesting because they may represent a spectrum of a single disease: 2 NLPHL cases, 1 THRLBCL case (a DLBCL variant with similar clinical features as NLPHL), and 1 DLBCL case (17). Examining germline mutations in the familial context may provide potential insights to the molecular basis of the disease.

Notably, no study has examined extensive environmental risk factors or patterns of disease aggregation, in multiple-case lymphoid cancer families; and no study has examined germline variants among a multiple-case family with heterogeneous rare lymphoid cancers.

The overall objective of my thesis was to characterize the Vancouver collection of multiple-case lymphoid cancer families and to identify patterns and/or etiological factors that may affect lymphoma susceptibility in the familial context. To complete these objectives, I undertook the following studies:

1. Identify patterns of lymphoma co-occurrence in multiple-case lymphoid cancer families that deviate from the population.
2. Compare lymphoid cancer age of onset patterns among multiple-case families and sporadic population cases.
3. Characterize the effects of family structure, early life environment, and immune-related disorders on the risk of lymphoma.
4. Examine shared genetic factors (from SNP array and whole exome sequencing data) in a multigenerational family with 4 lymphoid cancer cases.

1.10 Lymphoid Cancer Family Study (LCFS)

Unique study features:

Our study is limited to families with a history of lymphoid malignancies; however, our collection of multiple-case families differ in important ways from other studies. It is a large, but selected, group of families not limited to one geographic region, specific diagnostic group, twin studies, or analysis of family cancers by death certificates -- methodologies employed by other studies. In the LCF Study, we observe rare subtypes that are described infrequently in peer-reviewed literature. Families have heterogeneous types of lymphoid cancers across multiple generations, which facilitates the elucidation of associations in diverse families. We collected comprehensive environment and lifestyle information in a systematic matter, which allowed for the investigation of established risk factors in the context of multiple-case families.

.

Chapter 2: Methods

The Chapter 2 Methods section contains information regarding eligibility and recruitment for the Lymphoid Cancer Family Study, as well as data collection, quality control, nucleic acid extraction and sequencing information.

A description of statistical methods can be found in the Chapter in which it was performed.

2.1 Research ethics

This study was approved by the British Columbia Cancer (BCC) – University of British Columbia (UBC) joint Clinical Research Ethics Board. All participants provided written informed consent.

2.2 Eligibility and recruitment

Families are eligible for the LCFS if they include 2 or more members (living or deceased) with a lymphoid cancer or a lymphoid cancer precursor. Lymphoid cancers of interest include NHL, HL, CLL and MM, while lymphoid cancer precursors include MGUS. Patients with the disorders of interest, and all of their first-degree relatives, and additional relatives that connect affected family members, were invited to participate. Participation was not geographically limited, though most families were identified through a family member residing in British Columbia (BC), Canada.

Participants were recruited by physician referral (usually oncologists or haematologists), by volunteering through the study website (http://www.bcgsc.ca/faculty/awilson/recruitment-for-lymphoma-study-2013), at BC Cancer's annual Lymphoid Cancer Education Forum, or through other means. These families do not represent a population-based collection. Families were ascertained between 2006 – 2018.

In total, 367 individuals have been contacted to confirm eligibility for the LCF Study. Three hundred and twenty-six (89%) interviews were conducted. Non-response was due to death (11 cases; 3%) or failure to contact (30 cases; 8%). Two hundred and eighty-seven participants (78%) provided a family history that met the LCF Study eligibility requirements, while 30 (8%)

participants refused participation and 6 (2%) participants were ineligible. An additional 78 (21%) individuals were lost due to non-response. There are currently 212 families enrolled in the LCF Study. Once enrolled, family and personal medical history may be collected from multiple family members (cases and unaffected relatives) over time. The approximate response rate is 57.8% and the approximate completion of enrollment rate is 73.9%.

2.3 Data quality control

Individuals with questionable or uncertain diagnosis (e.g., “lymphocytic leukemia”) were excluded from analysis.

2.3.1 Lymphoid cancer subtypes

Lymphoid cancers were classified according to the InterLymph hierarchical classification of lymphoid neoplasms for epidemiological research (404) and include: CLL, NHL excluding CLL, HL, and MM. The InterLymph hierarchical classification of lymphoid neoplasms is based on the 2008 WHO classification of lymphoid neoplasms and the ICD for Oncology (3rd edition). Comparable USA population data (accessed through SEER) were also based on the 2008 or 2016 WHO criteria, and ICD for Oncology (3rd edition).

CLL and SLL are two forms of the same B-cell malignancy and they are treated in the same way. The difference between CLL and SLL is where the malignant clone resides. In SLL cases, most of the cancer cells are found in the lymph nodes, whereas in CLL cases, most of the cancer cells are found in the bone marrow or other lymphoid tissues. In this study, CLL and SLL cases were considered the same entity - we use the term *CLL* to refer to both CLL and SLL cases.

2.3.2 Multiple lymphoid cancer diagnoses

In rare instances where an individual is diagnosed with more than one lymphoid cancer, only the first lymphoid cancer diagnosis would be used to classify the affected individual. A relapsed or recurrent lymphoma may be influenced by cancer treatment methods and/or may have a different genetic architecture than the first lymphoid cancer diagnosis.

Lymphoid cancers that respond favourably to treatment are less likely to relapse. Individuals diagnosed with HL or high-grade NHL subtypes typically do not experience relapse;

however, some high-grade NHL subtypes such as MCL and T-cell NHLs are more likely to relapse.

Of 549 lymphoid cancer cases enrolled in the LCF Study, 25 cases had 2 or more occurrences of lymphoid cancer (**Table 2**). Twelve individuals had a recurrent lymphoid cancer that was the same subtype as their first lymphoma diagnosis. Seven cases were missing subtype information for at least one lymphoid cancer occurrence. Six cases had a second lymphoid cancer diagnosis within 2 years of their first occurrence. There was an average of 12 years between first and second lymphoid cancer diagnosis.

Table 2: Subtype information for cases with more than one occurrence of lymphoma.

	First occurrence		Second occurrence		# cases
	Type	Subtype	Type	Subtype	
Same type and subtype of lymphoma	NHL		NHL		12
		MCL		MCL	1
		DLBCL		DLBCL	3
		FL		FL	2
		NOS		NOS [1]	6
	HL		HL		2
		NLP		NLP	1
		NOS		NOS [1]	1
Same type but different subtype of lymphoma	NHL		NHL		9
		DLBCL		CLL	1
		CLL		DLBCL	3
		MALT		CLL	1
		FL		DLBCL	4
Different types of lymphoma	NHL		HL		1
		MZL		NS	1
	NHL		MM		1
		NOS		MM	1

Notes: [1] One or more lymphoid cancer diagnoses did not have subtype information.
Abbreviations: NHL, non-Hodgkin lymphoma; MCL, mantle cell lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; NOS, not otherwise specified; HL, Hodgkin lymphoma; NLP, nodular lymphocyte predominant; CLL, chronic lymphocytic leukemia; MALT, mucosa-associated lymphoid tissue; MZL, marginal zone lymphoma; NS, nodular sclerosis; MM, multiple myeloma.

2.3.3 SEER, CiNA and BC Cancer data sets

Lymphoid cancer incidence data were available for BC populations from the Cancer in North America (CiNA) (2010-2015) data set and for USA populations from the Surveillance, Epidemiology, and End Results (SEER) (1973-2016) data set. CiNA and SEER data is free to access and requires a Research Data Agreement.

CiNA is a data set released annually by the North American Association of Central Cancer Registries (NAACCR). NAACCR was established in 1987 and serves as a collaborative umbrella organization for 71 population-based cancer registries in North America (405). There are 9 participating registries in Canada (8 provinces and 1 territory) and 55 US registries (49 states, Puerto Rico, and 5 metropolitan registries). CiNA contains comprehensive cancer incidence data for North America, or independent registries, states, provinces or countries. Data published in CiNA contains the most current 5 years of data and is available in SEER*Stat (406).

The SEER Program is the main database that the National Cancer Institute uses to support cancer surveillance in the USA. Data collection began in 1973 with 7 registries and has since expanded to include 18 registries, covering approximately 34.6% of the American population (407). SEER captures more than 400,000 cancer cases annually. Information collected by SEER includes (but is not limited to): patient demographics, tumour characteristics (site, morphology), stage of disease (diagnosis), treatment and outcomes (407).

SEER data were used because comparably comprehensive Canadian data were not available. SEER data contains detailed information on subtype and ethnicity, both of which affect the age at diagnosis distribution of cases. In CiNA data sets, lymphoid cancer information was limited to NHL, HL and MM types; detailed information on CLL cases, as well as, NHL and HL subtypes, and ethnicity data were not available. The following sections compare SEER and CiNA data sets in detail.

2.3.3.1 Descriptive information

SEER contains 100 times more lymphoid cancer cases than comparable BC data available through CiNA data sets (**Table 3**). A larger sample size enables better descriptive values of a sample and avoids errors from testing a small number of possibly atypical samples (e.g., outliers) which may be the circumstance for rare subtypes. The mean and median and age of diagnosis for NHL, HL and MM are comparable between the SEER and BC populations for men and women. A comparison of CLL cases between SEER and CiNA data sets was not possible

because only “lymphocytic leukemia” cases, which includes CLL and ALL subtypes, were available in CiNA data sets.

Table 3: Lymphoid cancer age of diagnosis in SEER (USA) and CiNA (BC) populations.

Type	Sex	Data source	n	**Mean** (y)	**Median** (y)
NHL	Male	SEER, USA	293,911	62.7	66
		CiNA, BC	2662	64.7	65-69
	Female	SEER, USA	231,162	65.4	69
		CiNA, BC	2149	67.1	65-69
HL	Male	SEER, USA	27,006	42.0	39
		CiNA, BC	280	42.6	35-39
	Female	SEER, USA	21,905	41.0	35
		CiNA, BC	236	38.9	30-34
MM	Male	SEER, USA	48,879	68.5	70
		CiNA, BC	803	68.5	65-69
	Female	SEER, USA	39,514	70.0	71
		CiNA, BC	641	70.5	70-74
CLL/LL[1]	Male	SEER, USA	62,513	68.6	70
		CiNA, BC	1124	62.2	65-69
	Female	SEER, USA	43,307	70.9	73
		CiNA, BC	661	63.3	65-69

Notes: [1] A direct comparison of CLL cases between SEER (USA) and CiNA (BC) data sets were not possible due to different lymphoid cancer groupings. SEER (USA) population data were available for CLL cases, whereas CiNA (BC) population data were available for “lymphocytic leukemia” cases which included both CLL and ALL subtypes.
Abbreviations: SEER, Surveillance, Epidemiology, and End Results; CiNA, Cancer in North America.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; MM, multiple myeloma; CLL, chronic lymphocytic leukemia; LL, lymphocytic leukemia; ALL, acute lymphoblastic leukemia.

The age at diagnosis distributions for NHL, HL and MM are comparable between the USA (SEER) and BC (CiNA) populations for men and women (**Figure 6**). Lymphocytic leukemia cases (BC population) were earlier in onset compared to CLL cases (USA population), likely due to the inclusion of ALL cases which are most frequently diagnosed among people aged < 20 years (408).

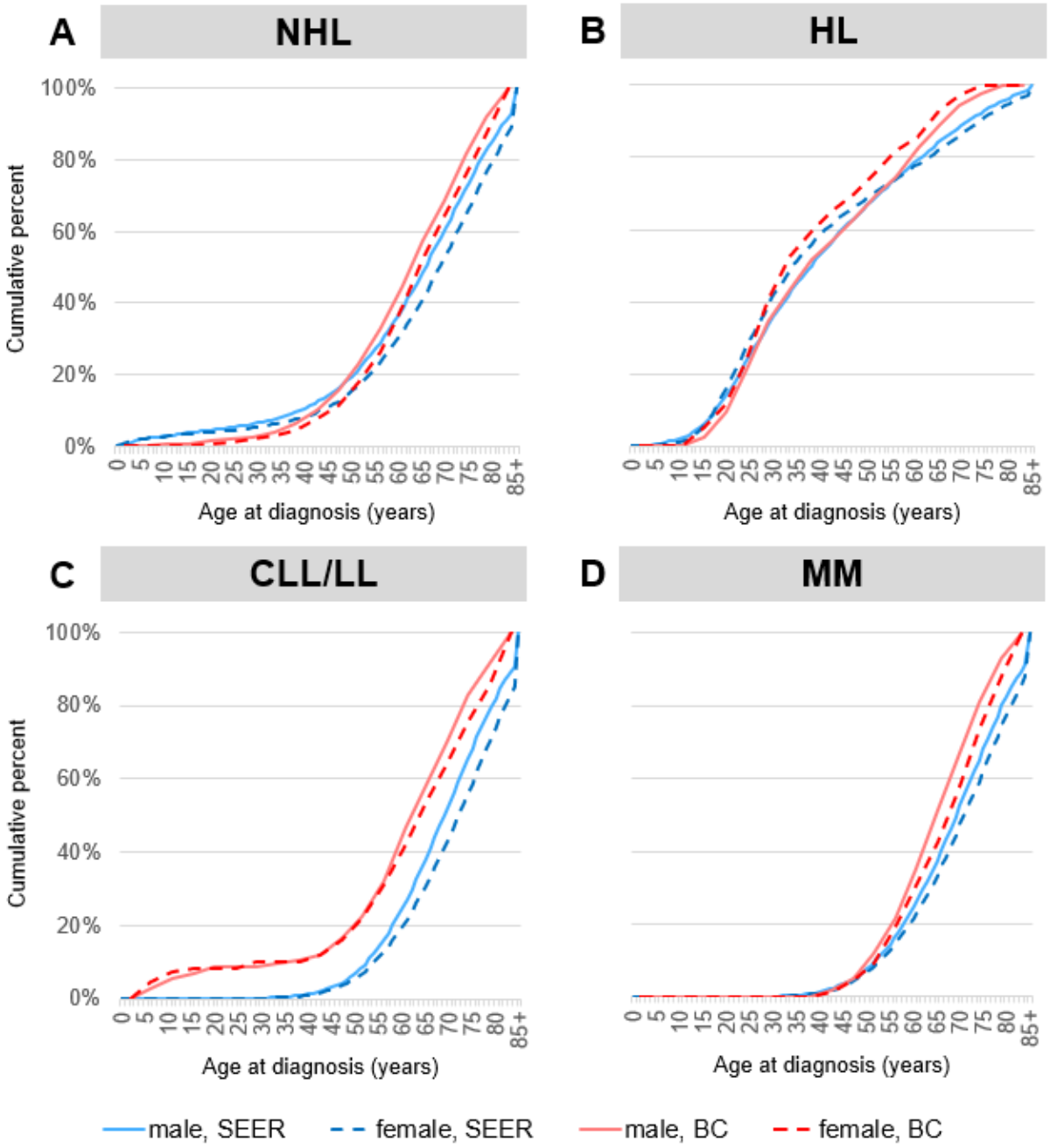


Figure 6: Age at diagnosis distributions for SEER (USA) and CiNA (BC) population data for (**A**) NHL, (**B**) HL, (**C**) CLL/LL, and (**D**) MM cases.

Notes: SEER (USA) population data (blue) is compared to CiNA (BC) population data (red). Male data (solid lines) is compared to female data (dashed lines). Figure C compares CLL cases from SEER (USA) population data to lymphocytic leukemia (CLL and ALL) cases from CiNA (BC) population data.

Abbreviations: SEER, Surveillance, Epidemiology, and End Results; CiNA, Cancer in North America.

Subtype Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; MM, multiple myeloma; CLL, chronic lymphocytic leukemia; LL, lymphocytic leukemia; ALL, acute lymphoblastic leukemia.

2.3.3.2 Subtypes

Lymphoid cancer age at diagnosis differs by type and subtype. For example, NHL, CLL and MM occur frequently in older individuals in their 60s and 70s (29,33,87,88), whereas HL is more frequently a disease of young people between the ages of 15-35 years (9,87). Age at diagnosis distributions also vary within NHL and HL subtypes. For example, the median age of onset for NS (28 years old) and NLP (45 years old) fall between the median age of onset for HL (39 years old). Similarly, the median age of diagnosis for NHL cases is 66 years, however, LPL/WM subtypes are typically earlier in onset (51 years), whereas MZL and DLBCL cases are older (68 and 70 years, respectively) (24,25,409,410). **Figure 7** displays the age at diagnosis distributions for several NHL and HL subtypes. Of the NHL subtypes depicted in **Figure 7A**, BL cases (male) had the earliest median age of diagnosis (40 years), and CLL cases (female) had the oldest median age of diagnosis (71 years). Of the HL subtypes depicted in **Figure 7B**, NS cases (female) had the earliest median age of diagnosis (29 years), and MC cases (female) had the oldest median age of diagnosis (56 years). The median age of diagnosis for NHL and HL subtypes varied by approximately 31 and 27 years, respectively. The variation among age at diagnosis distributions for heterogeneous lymphoid cancer subtypes and sexes suggests it is important to account for these differences when examining the rarity of lymphoid cancer occurrences.

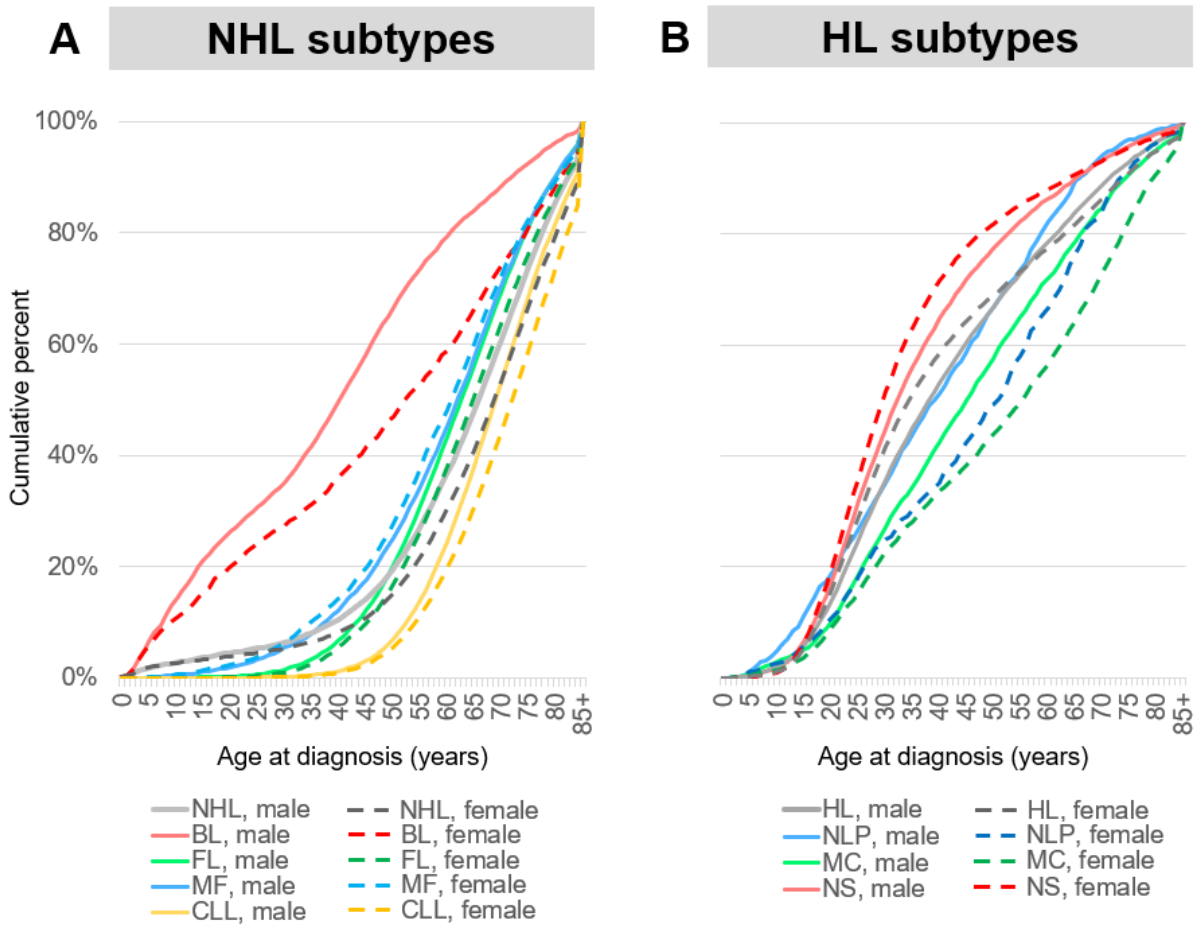


Figure 7: Age at diagnosis distributions for (**A**) NHL and NHL subtypes and (**B**) HL and HL subtypes using SEER (USA) population data.

Notes: SEER data were used because comparably comprehensive Canadian data were not available. Age at diagnosis distributions are compared between male (solid lines) and female (dashed lines) population cases. A) SEER population data are compared for NHL cases considered as a group (grey), and BL (red), FL (green), MF (blue) and CLL (yellow) subtypes. B) SEER population data are compared for HL cases considered as a group (grey), and NLP (blue), MC (green), and NS (red) subtypes.

Abbreviations: SEER, Surveillance, Epidemiology, and End Results.

Subtype abbreviations: NHL, non-Hodgkin lymphoma; BL, Burkitts lymphoma; FL, Follicular lymphoma; MF, mycosis fungoides; CLL, chronic lymphocytic leukemia; HL, Hodgkin lymphoma; NLP, nodular lymphocyte predominant; MC, mixed cellularity; NS, nodular sclerosis.

2.3.3.3 Ethnicity

There are significant racial differences for several characteristics of lymphoid cancer cases, including sex, age at diagnosis, stage and lymph site, and the patterns vary by subtype (411). For example, NHL, HL and CLL incidence is the highest in Caucasian populations, followed by African American and Asian populations (33,62,83,88,91). MM incidence is highest in African

American populations, followed by Caucasian and Asian populations (29,95). Racial differences in the distribution of age of diagnosis have been observed among DLBCL, FL and CLL subtypes, with Caucasian populations having the highest age at diagnosis, followed by Asians and African Americans (411,412). **Figure 8A** and **Figure 8B** exemplify racial differences for DLBCL and MZL cases, respectively, among White, Pakistani and Chinese ethnicities (as defined by SEER). The median age of diagnosis for DLBCL cases varied by 8 years, while the median age of diagnosis for MZL cases varied by 9 years. A majority (> 95%) of LCF Study participants were of White ethnicity, followed by Asian (approximately 3%), and unknown (e.g., ancestry information was unavailable or not classifiable using SEER race codes).

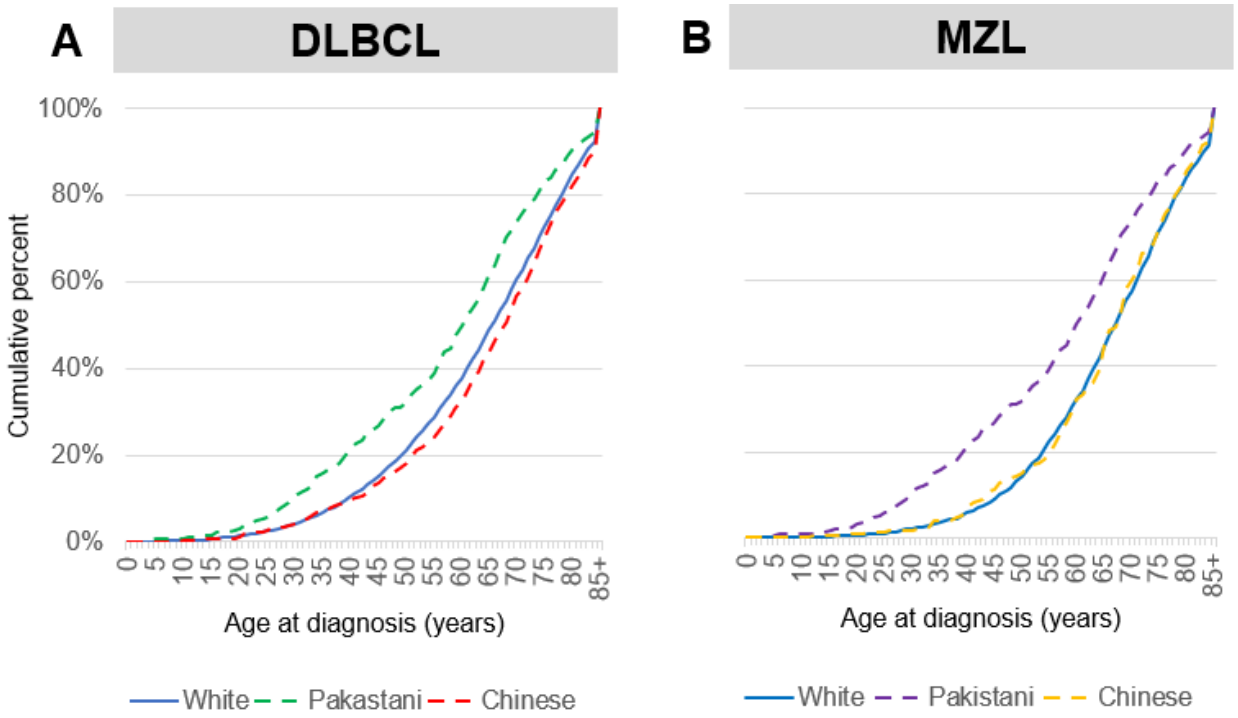


Figure 8: Age at diagnosis distributions for White, Pakistani and Chinese ethnicities for (**A**) DLBCL and (**B**) MZL subtypes using SEER (USA) population data.
Notes: SEER data were used because comparably comprehensive Canadian data were not available. Age at diagnosis distributions are compared for male cases only. Ethnicity groups are defined as per SEER. A) SEER population data are compared for White (blue solid line), Pakistani (green dashed line) and Chinese (red dashed line) DLBCL cases. The median age at diagnosis for DLBCL cases was 60 (Pakistani), 66 (White), and 68 (Chinese). B) SEER population data are compared for White (blue solid line), Pakistani (purple dashed line) and Chinese (yellow dashed line) MZL cases. The median age at diagnosis for MZL cases was 59 (Pakistani), 67 (Chinese), and 68 (White).
Abbreviations: SEER, Surveillance, Epidemiology, and End Results.
Subtype abbreviations: DLBCL, Diffuse large B-cell lymphoma; MZL, marginal zone lymphoma.

The availability and abundance of heterogeneous lymphoid cancer subtypes and diverse ethnicities makes SEER data an important resource for modelling age at diagnosis distributions for lymphoid cancers. For these reasons, SEER data were used as a comparison group for the Lymphoid Cancer Family Study research projects presented in this thesis.

2.3.4 Histopathological confirmation of lymphoma

Living participants provided self-reported lymphoid cancer diagnoses. Probands or a close relative provided lymphoid cancer information for deceased cases. Reported diagnoses were validated by review of original medical, pathology, clinical and/or laboratory records or by referring physician report. When available, histopathology slides from formalin-fixed paraffin-embedded (FFPE) tumours were reviewed by an expert oncology pathologist to confirm the lymphoid cancer diagnosis.

CLL/SLL cases typically do not form solid tumours and so we were unable to use immunohistochemistry to verify the diagnosis; however, in some instances, medical records were available for confirmation of disease.

2.3.5 Familial predisposition genes

Families with a strong history of neoplasms characteristic of Li-Fraumeni syndrome had the *TP53* tumour suppressor gene sequenced to investigate possible mutations contributing to inherited multicancer syndromes. *TP53* exons 5 – 8 were selected for Sanger sequencing due to their high frequency of germline mutations in Li-Fraumeni families (413).

2.4 Data collection

Information about lymphoid and non-lymphoid malignancies, family structure and demographics was obtained systematically using a standardized family history questionnaire and phone interviews with multiple family members.

Age-of-onset and histological subtype were self-reported, or reported by the proband or close relive.

Family members reported on demographic, education of themselves and their parents (less than high school, high school graduate, post-secondary graduate), income during childhood (below average, average, above average), farm history (lived on a farm, did not live on a farm), and location (urban, rural) information. Personal medical history (allergies, autoimmune diseases, surgical procedures) and early lifestyle data were obtained from a self-administered questionnaire. Allergies were classified as drug, environmental or food/diet related. Autoimmune diseases were categorized as systemic conditions with detectable autoantibodies, organ-specific conditions with detectable autoantibodies, or conditions without detectable autoantibodies, according to previous studies (232,414).

Participants were invited to provide a DNA sample through blood, saliva and FFPE tumour blocks as described below. Medical records, pathology/histology reports, and tumour slides were reviewed by an expert oncology pathologist to confirm the self-reported cancer.

2.5 Nucleic acid extraction

2.5.1 Peripheral whole blood

Peripheral whole blood was collected and stored in a glass BD Vacutainer ethylenediaminetetraacetic acid (EDTA) tubes. EDTA is a chelating anticoagulant which is used for routine hematology tests. EDTA preserves the morphology of the cellular elements of blood. DNA extractions are typically performed within 24-48 hours of blood draw. Genomic DNA (gDNA) was extracted from white blood cells using a modified sucrose protocol. Blood from individuals with CLL are separated into granulocyte and lymphocyte fractions using the EasySep™ magnetic bead method (STEMCELL Technologies, Vancouver, BC) before DNA extraction. Plasma and viable frozen cells are collected for immortalization to support later functional studies.

2.5.2 Saliva

Saliva is collected using the DNA Genotek Oragene 500 self-collection kit (415). The Oragene-500 kit contains a stabilizing liquid that can preserve the integrity of DNA in saliva at room temperature for at least 5 years (416). gDNA is extracted using the prepIT-L2P laboratory protocol for manual purification of DNA provided by DNA Genotek (417).

2.5.3 FFPE tissue blocks

FFPE tissue blocks are collected for individuals with lymphoid cancers, when possible. Tumour blocks may also be collected for non-lymphoid tumours from other members of the family that may be of interest. An oncology pathologist obtains and evaluates immunohistochemistry slides to confirm the lymphoid cancer diagnosis. Fresh scrolls (or cores) are obtained for total nucleic acid extraction.

DNA and RNA are extracted from fresh scrolls or cores within 3 weeks using the QIAGEN AllPrep DNA/RNA (Germantown, Maryland, USA) and deparaffinization solution according to the manufacturer's recommended protocol (418).

2.6 Family prioritization, sequencing and variant calling

2.6.1 Pedigree prioritization:

Multiple-case lymphoid cancer families were prioritized for exome sequencing. Pedigrees were prioritized by the number of lymphoid cancer cases and the number of germline and/or somatic samples available for analysis. Individuals with matched tumour samples of sufficient cellularity (≥ 80% malignant cells) were assigned a higher priority than unmatched lymphoid cases. If possible, all lymphoid affected relatives in a single pedigree with available DNA samples were included for sequencing.

Families with three or more lymphoid cancer cases were selected for sequencing. At the time of sequencing, 69 families had 3 or more lymphoid cancer cases and 10 of these families contained at least 3 lymphoid affected germline samples (**Table 4**). Four families with 6 and 11 lymphoid cancer cases were unlikely to receive additional germline samples as they have declined to participate/are unable to contact, or are deceased.

Table 4: Number of germline samples received by 2014, organized by number of lymphoid cancer cases in a family.

	# of germline samples						
# cases per family	0	1	2	3	4	5	Total
2 cases	15 (9%)	51 (30%)	33 (20%)				99 (59%)
3 cases	12 (7%)	18 (11%)	15 (9%)	**3 (2%)**			48 (29%)
4 cases	2 (1%)	6 (4%)	1 (1%)	**2 (1%)**	**2 (1%)**		13 (8%)
5 cases		1 (1%)		**1 (1%)**		**1 (1%)**	3 (2%)
6 cases		2 (1%)	1 (1%)				3 (2%)
8 cases				**1 (1%)**			1 (1%)
11 cases			1 (1%)				1 (1%)
Total	29 (17%)	78 (46%)	51 (30%)	7 (4%)	2 (1%)	1 (1%)	168

Notes: Bold text denotes families of highest priority.

Families with two lymphoid cancer cases that both have a tumour and normal sample were a higher sequencing priority than families with two cases without a matched tumour. Families with cases that had matched tumour-normal samples were assigned a higher priority for sequencing to support future studies and insights into tumourigenesis. Families which have tumour only, or 1 blood and 1 tumour from different or unaffected individuals were assigned a lower priority for sequencing. Families containing a DNA sample from one lymphoid cancer case were the lowest priority.

Unaffected relatives cannot truly be considered unaffected because they may become affected at a later date.

92 individuals were selected from 40 families for whole exome and untranslated region (UTR) sequencing.

2.6.2 Whole exome sequencing (WES)

Library construction and exome capture were performed on DNA from selected blood and saliva samples using Agilent SureSelect v5 + UTR kit, performed by the Library construction and Sequencing core laboratories at the Genome Sciences Centre (Vancouver, BC, Canada). Sequencing was done using the HiSeq 2500 at 125 bp paired-end reads with V4 chemistry at an estimated 100X coverage. Filtered reads were aligned to the hg19 human reference genome using the Burrows-Wheeler Aligner, performed by the Bioinformatics core at the Genome Sciences Centre (Vancouver, BC, Canada).

2.6.3 Joint variant calling

SNPs and indels were jointly called within 40 families (and 92 individuals) with a family history of lymphoma. Variants were jointly called among all families to enable clearer distinction between homozygous reference sites and sites with missing data. Jointly called Variant Call Format files from 4 relatives in Family 133 were used in Chapter 6. Joint variant calling was performed using the HaplotypeCaller in genomic Variant Call Format mode using Genome Analysis Toolkit version 3.0. Read qualities were recalibrated with the Genome Analysis Toolkit with a minimum quality score of 30. Variant calling was performed by the Bioinformatics core at the Genome Sciences Centre (Vancouver, BC, Canada).

Chapter 3: Nonrandom occurrence of lymphoid cancer types in 140 families.

3.1 Introduction

Familial lymphoid cancer is rare; however, it is known that lymphoid cancer risk is elevated among persons with relatives diagnosed with a LPD and depends on the familial relationship, type of lymphoid cancer or precursor lesion, sex, and age of onset (256,419–421). There are abundant reports of familial associations between LPDs, including combinations of NHL (62,108,347,351,352,421–427), HL (62,156,347,351,352,422,423,427–429), CLL (108,347,352,366,422,423,426,430) and MGUS (426), with fewer observations of MM co-occurring in families with other lymphoid cancers (108,424,426,431), possibly due to its relative rarity (352,423). Heritability estimates for susceptibility to individual lymphoid cancers vary by type: 21-48% for HL (43–45), 8% for MZL (49), 21% for FL (49), 9-16% for DLBCL (42,49,50), 16-34% for CLL (42,49,51–54), and 17% for MM (46–48). Clustering of different lymphoid cancers in families is suggestive of inherited genetic factors affecting susceptibility to multiple lymphoid cancers. To date, most reports examined small numbers of families. With the exception of a study evaluating 12 large Australian families with lymphoid and myeloid malignancies across multiple generations (432), our knowledge about lymphoid cancers and their inheritance patterns in large families is limited.

GWASs have identified susceptibility loci for sporadic lymphoid cancers, including genes involved in immune recognition and function (50), particularly the human leukocyte antigen region (359,370) and genes involved in DNA repair (381). Candidate gene studies have also identified putative susceptibility factors for sporadic lymphoid cancers. However, few genes associated with familial lymphoid cancers have been identified: a translocation disrupting *KLHDC8B* was identified in a family with NS HL (433), a rare non-synonymous variant in *KDR* (*VEGFR2*) was identified in two HL families (434), and mutations in *PRF1* have been found in familial hemophagocytic lymphohistiocytosis (435), T- and B-cell lymphomas (436,437), and autoimmune lymphoproliferative syndrome (438). These studies aside, there remains a large gap in our knowledge of the genetics of familial lymphoid cancers.

3.2 Methods

3.2.1 Eligibility and recruitment

At the time the analysis was performed, 140 multiple-case lymphoid cancer families were enrolled in the LCFS. Families were ascertained between 2006 and 2014. Medical records and pathology slides were used to confirm the diagnosis. Lymphoid cancers were classified according to 2008 WHO guidelines and include: CLL, NHL excluding CLL, HL and MM. Subtype-specific analyses were not performed due to small sample size. More information on eligibility and recruitment of families is provided in **Chapter 2.2: Eligibility and recruitment** (Methods), page 37.

3.2.2 Statistical analysis

Families were ascertained largely through referrals; they do not represent a population-based collection. For this reason, we cannot estimate a population size to use as a denominator to calculate the incidence of lymphoid cancers in families vs. in the population. Instead, we test whether specific properties of the familial cases differ from those of sporadic cases. The properties assessed were age of onset, co-occurrence patterns, sex distribution, and inter-generational differences.

Statistical analysis was performed using R version 3.1.3.

For analysis of familial lymphoid cancer co-occurrences, year-of-diagnosis and ethnicity-specific population incidence rates were obtained for each NHL, HL, CLL and MM case. Incidence rates (**Supplementary Table A.1**) were obtained from Surveillance, Epidemiology, and End Results (SEER) databases (439–441), accessed through SEER*Stat software (442). SEER (USA) data were used because comparable Canadian data were not available. Since incidence rates vary by year and by ethnicity, the rates used to calculate the probability of a lymphoid cancer type occurring by chance were weighted by the frequency of observed year-of-diagnosis and ethnicity of cases in the families. The calculation of weighted incidence rates is described in detail in **Appendix A** and **Supplementary Tables A.2**, **A.3** and **A.4.** The weighted incidence rates were used to calculate the expected probabilities of co-occurrence for 10 pairwise relationships (e.g., NHL/NHL, NHL/CLL, NHL/HL, NHL/MM, CLL/CLL, CLL/HL, CLL/MM, HL/HL, HL/MM and MM/MM) (**Supplementary Table A.5**).

A chi-square statistic was used to assess the goodness of fit of observed lymphoid affected pairs to the expected population pairs (R function: chisq.test). Under the null hypothesis, the familial lymphoid cancer co-occurrence rates are consistent with the expected population co-occurrence rates. Under the alternate hypothesis, the familial lymphoid cancer co-occurrence rates are different from the expected population co-occurrence rates. Degrees of freedom is equal to 9. The chi-square goodness-of-fit test was 1-sided with an alpha level of 0.05 as the threshold to reject the null hypothesis. A chi-square goodness of fit test is an appropriate test for categorical data to determine how well the familial proportions (which represent the sample distribution) compare to the expected population distributions. However, some lymphoid cancer pairs have fewer than 5 expected cases, which may inflate the contribution to the chi-square statistic.

Families were separated into those with 2 cases, 3 cases, or 4 or more lymphoid cancer cases. Lymphoid cancer pairs were simulated using population incidence rates, according to the number of pairs by family size. For example, as there were 88 pairs in families with two lymphoid cancer cases, 88 pairs of lymphoid cancers were simulated using population incidence rates and the chi-square statistic was computed. This was repeated 10,000 times, to form a null distribution of chi-square statistics. The *p*-value obtained is the proportion of simulated chi-square statistics that were greater than or equal to the observed chi-square statistic. The signed square root of the contribution to the chi-square statistic can be viewed as a measure of deviation from the expected co-occurrence rate.

Co-occurrence patterns were also examined using 2015 Canadian population incidence rates in replacement of SEER weighted population incidence rates (**Supplementary Figure A.1**). Unlike SEER (USA) population data, ethnicity- and year-specific incidence rates were unavailable for Canadian data (**Supplementary Table A.6**). The expected co-occurrence patterns were comparable using SEER (USA) and Canadian population data.

3.3 Results

Within 140 families, 353 lymphoid cancer cases were identified. Families had 2 to 6 cases spanning 1 to 4 generations. Year of diagnosis was available for 85% of cases, ranging from 1950 to 2014. Demographic variables are displayed in **Table 5**. Medical records were available for 191 (54%) cases, all of which supported their self-reported diagnosis. Eighty-one (23%) cases were

pathologically confirmed. A subtype was reported for 69% and 48% of NHL and HL cases, respectively. Subtype proportions resemble those of the population (data not shown), with the exception of FL and LPL, which were over-represented within NHL (39% and 10% of cases compared to 22% and 1% of NHL cases in Canada).

Table 5: Demographic and medical data for 353 lymphoid cancer cases among 140 multiple-case lymphoid cancer families.

	Type of lymphoma				All cases, n (%)
Characteristic	**NHL, n (%)**	**HL, n (%)**	**CLL, n (%)**	**MM, n (%)**	
Number of cases	178 (50)	54 (15)	101 (29)	20 (6)	353 (100)
Sex					
Male	98 (55)	23 (43)	50 (50)	9 (45)	180 (51)
Female	80 (45)	31 (57)	51 (50)	11 (55)	173 (49)
Age, y					
Mean ± SD	57 ± 17	35 ± 18	61 ± 12	60 ± 14	55 ± 18
Median	57	30	62	63	57
Range	2 - 88	9 - 80	27 - 91	26 - 78	2 - 91
Ethnicity					
White	163 (92)	47 (87)	98 (97)	19 (95)	327 (93)
Asian	10 (6)	1 (2)	-	1 (5)	12 (3)
Unknown[1]	5 (3)	6 (11)	3 (3)	-	14 (4)
Location					
Canada	172 (97)	53 (98)	99 (98)	20 (100)	344 (97)
USA	3 (2)	1 (2)	2 (2)	-	6 (2)
British Isles	2 (1)	-	-	-	2 (1)
Australia	1 (1)	-	-	-	1 (1)
Diagnosis validated					
Medical records	99 (56)	26 (48)	58 (57)	8 (40)	191 (54)
BCC pathologist confirmed	51 (29)	17 (31)	12 (12)	1 (5)	81 (23)

Notes: - represents zero. [1] Unknown ethnicity denotes individuals in which ancestry was unavailable due to missing information, unclassifiable using SEER race codes or unknown among adopted individuals.
Abbreviations: y, years; SD, standard deviation; USA, United States of America; BCC, British Columbia Cancer.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

Lymphoid cancer cases' relatives and relationship types are shown in **Table 6.** Eighty-six families (61%) have lymphoid cancer cases only in first-degree relatives. There are 107 parent-child and 72 sibling pairs. The type of lymphoid cancer was the same for 65% and 60% of parent-child and sibling pairs, respectively.

Table 6: Degree and type of relationship, by size of lymphoid cancer family.

	Two cases	Three cases	≥ Four cases	All cases
No. of families	88	37	15	140
No. of cases	176	111	65	353
Total pairs	88	111	115	314
No. of types of lymphoma				
1	55 (62%)	21 (57%)	3 (20%)	79 (56%)
2	33 (38%)	12 (32%)	8 (53%)	53 (38%)
3	-	4 (11%)	3 (20%)	7 (5%)
4	-	-	1 (7%)	1 (1%)
Relatives				
1°	76 (86%)	61 (55%)	42 (37%)	179 (57%)
2°	9 (10%)	26 (23%)	29 (25%)	64 (20%)
3° or higher	3 (3%)	21 (19%)	33 (29%)	57 (18%)
Case-case relationship				
1° Parent-child	48 (55%)	38 (34%)	21 (18%)	107 (34%)
1° Siblings	28 (32%)	23 (21%)	21 (18%)	72 (23%)
2° Avuncular	5 (6%)	20 (18%)	19 (17%)	44 (14%)
2° Grandparent-grandchild	3 (3%)	6 (5%)	9 (8%)	18 (6%)
2° Half-siblings	1 (1%)	-	1 (1%)	2 (1%)
3° Cousins	2 (2%)	11 (10%)	16 (14%)	29 (9%)
3° or higher, other	1 (1%)	10 (9%)	17 (15%)	28 (9%)

Notes: - represents zero. n pairs (%); ≥ Four Cases includes 10 families with 4 cases (60 pairs), 4 families with 5 cases (50 pairs), and 1 family with 6 cases (15 pairs).
Abbreviations: No., number; 1°, first-degree relative; 2°, second-degree relative; 3°, third-degree relative.

3.3.1 Co-occurrence patterns of lymphoid cancers in families

The co-occurrence of lymphoid cancers within families was investigated to determine whether combinations of lymphoid cancers were distributed according to population incidence rates. For example, because NHL is the most common lymphoid cancer type in the population, we would expect NHL/NHL to be the most frequent pair. **Figure 9** summarizes lymphoid cancer co-occurrence patterns in the LCFS, organized by lymphoid cancer pair categories. Collectively, 140 multiple-case families contained lymphoid cancer pairs that differed from the expected proportions ($p < 0.0001$). Families with 2 cases, 3 cases, or 4 or more cases had lymphoid cancer

	Expected	Two Cases	Three Cases	≥Four Cases
	Families	88	37	15
	Cases	176	111	66
NHL/NHL	32.2	33 (29)	44.4 (48)	19.2 (20)
NHL/CLL	20.5	20.5 (18)	10.2 (11)	15.4 (16)
NHL/HL	11.2	10.2 (9)	6.5 (7)	25 (26)
NHL/MM	17.5	2.3 (2)	8.3 (9)	3.8 (4)
CLL/CLL	3.2	20.5 (18)	16.7 (18)	15.4 (16)
CLL/HL	3.5	1.1 (1)	4.6 (5)	6.7 (7)
CLL/MM	5.5	0 (0)	2.8 (3)	2.9 (3)
HL/HL	1	6.8 (6)	6.5 (7)	9.6 (10)
HL/MM	3	2.3 (2)	0 (0)	1 (1)
MM/MM	2.4	3.4 (3)	0 (0)	1 (1)
	P	<.0001	<.0001	<.0001
	Chi squared	130.07	116.29	169.92

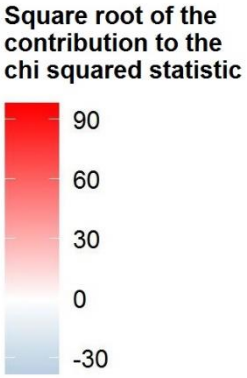


Figure 9: Lymphoid cancer co-occurrence in families with 2 cases, 3 cases or 4 or more lymphoid cancer cases.

Notes: observed % (*n* pairs). Colour is associated with variation from expected frequency of random lymphoid cancer co-occurrences, as determined by the signed square-root of the contribution to the chi-square statistic. For example, according to population rates, CLL/CLL should be 3.2% of pairs. For families with 2 cases, 18 (20.5%) observed pairs were CLL/CLL. This deviates from the expected proportion of 3.2%, and is red because we observed a higher proportion of pairs than expected. Families with 2 cases, 3 cases, or 4 or more cases had lymphoid cancer pairs that differed from the expected population proportions ($p < 0.0001$).

Statistical methods: A chi-square statistic was used to assess the goodness of fit of observed lymphoid affected pairs to the expected population pairs. Lymphoid cancer pairs were simulated using population incidence rates, according to the number of pairs by family size. This was repeated 10,000 times, to form a null distribution of chi-square statistics. The *p*-value obtained is the proportion of simulated chi-square statistics that were greater than or equal to the observed chi-square statistic. The signed square root of the contribution to the chi-square statistic can be viewed as a measure of deviation from the expected co-occurrence rate. Degrees of freedom is equal to 9.

Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

pairs that differed from the expected population proportions ($p < 0.0001$). The test statistic firmly rejects the null hypothesis that there is no difference between the observed familial co-occurrence rates and the expected population co-occurrence rates. Specific lymphoid cancer pairs co-occurred more frequently than expected based on population incidence rates. In families with 2 lymphoid cancer cases CLL/CLL, and HL/HL pairs occurred more frequently by 17% and 6% of pairs. NHL/MM and CLL/MM pairs occurred less frequently than expected by 15% and 6% of pairs. Other pairs occurred similarly to the expected population distribution.

CLL/CLL and HL/HL pairs also occurred more frequently than expected in families with 3 lymphoid cancer cases, but NHL/NHL pairs were also over-represented. NHL/CLL, NHL/HL and NHL/MM pairs occurred less frequently. Other pairs resembled the expected population frequencies.

Families with 4 or more lymphoid cancer cases also had more CLL/CLL and HL/HL cases. Increased NHL/HL and CLL/HL pairs were also observed. NHL/NHL, NHL/CLL and NHL/MM pairs occurred less frequently than expected; Other pairs occurred similarly to the expected population distribution.

Several patterns emerge in CLL and HL families. HL/CLL and HL/HL pairs increased as the number of lymphoid cancer cases per family increased. Interestingly, CLL/CLL occurrences were consistently and substantially above the expected population frequencies for all family sizes. The remaining pairs occurred less frequently than expected (NHL/CLL, NHL/MM, HL/MM, CLL/MM and MM/MM) or did not show a clear pattern (NHL/NHL and NHL/HL) among families with 2, 3 and 4 or more cases.

3.4 Discussion

This collection of 140 multigenerational lymphoid cancer families allows us to examine patterns of co-occurrence and deduce which cases are nonrandom and may have shared susceptibility factors.

The co-occurrence of NHL, HL, CLL and MM in these families demonstrates that some pairs of lymphoid cancers co-occur more often than expected by chance (particularly CLL with CLL, HL with HL), and others occur less often. These trends are more noticeable in families with more cases. This observation is important, as it shows these familial cases are not just random

co-occurrences of sporadic population cases in the same family; they are something different, and are more likely to have a shared genetic basis than sporadic cases.

MM occurrences with other lymphoid cancers were below the expected frequency, suggesting that factors that predispose to NHL, HL and/or CLL may be different from those that affect risk of MM. Our relative lack of MM families may be due in part to ascertainment bias, as our research centre does less research on MM than other lymphoid cancers, and connections with referring oncologists and hematologists may reflect that bias.

Co-occurrence rates were comparable using SEER (USA) and Canadian population data, suggesting that population differences may be small. Canada and USA have comparable age-adjusted lymphoid cancer incidence rates as seen in high-income countries. Ethnicity-specific incidence rates were unavailable for Canadian data; however, 97% of familial cases in this study reside in Canada and may represent the underlying ethnicity distribution of NHL, HL and CLL cancers. MM incidence rates among individuals of African American descent are nearly twice as much as individuals of European descent; however, no MM cases or families in this study were of African American descent.

Co-occurrence patterns are important because families with more affected cases are expected to be enriched for genetic factors, and by examining the trends from small families to larger ones, we can distinguish the patterns resulting from genetic effects from non-genetic effects that may predominate in the general population. Analysis of lymphoid cancer families is likely to result in identification of genetic factors that are less common, but have higher penetrance, than those identified by GWAS of sporadic cases. Families containing unexpected combinations of lymphoid cancers may have the greatest potential for identifying familial lymphoid cancer genes.

A non-genetic explanation could be exposure to a causative environmental agent. However, this would apply to families with fewer cases in this study, as most families with three or more cases reported affected relatives over more distant familial relationships (e.g., second- or third-degree relatives) who are less likely to live together. Furthermore, only 5% of families reported a year of onset within 5 years of onset for other lymphoid affected relatives, suggesting that temporally specific environmental exposures are likely not a major effect in most families.

Limitations of this study include ascertainment bias and recall bias. Recall bias was minimized by taking a systematic detailed family history from multiple family members and cross-checking information regarding cancer diagnosis. In addition, medical records, tumour blocks,

and histopathology slides were requested and reviewed by an expert oncology pathologist to confirm the diagnosis. Of the 353 cases studied, 191 (54%) were verified from medical records and 81 (23%) were verified from review of pathology slides. Every case that was verified matched the reported lymphoid cancer diagnosis, and we expect the remaining self-reported types to have a similar level of accuracy. Familial MM cases were limited, providing inadequate sample size for statistical analysis. Another limitation is survival bias, in which more indolent lymphomas such as FL and LPL are over-represented.

3.5 Conclusion

This work establishes that specific combinations of lymphoid cancers occurring in families are non-random, and are therefore not all chance occurrences of sporadic cases in the same family. These co-occurrences may reflect the effect of genetic factors, some of which may predispose to more than one lymphoid cancer type. Enrichment for some combinations of lymphoid cancer in families support the application of genomic methods to identify genes and genetic variants that underlie familial lymphoid cancers. Awareness of familial lymphoid cancer patterns and the identification of susceptibility genes has the potential to enhance screening methods for affected families in the future.

Chapter 4: Early age of onset of lymphoid cancer in 200 families.

4.1 Introduction

Anticipation is an inheritance pattern where disease severity may increase and/or age of onset decreases in successive generations. This phenomenon has been documented in several diseases, including neurodegenerative disorders (e.g., fragile X syndrome, Huntington's disease, myotonic dystrophy) (443–446), autoimmune diseases (e.g., RA, type 2 diabetes mellitus, Graves' disease) (446,447), cancer predisposition syndromes (e.g., Li-Fraumeni Syndrome) (444,447), and several cancers, including retinoblastoma, breast and ovarian cancer, and some familial leukemia's and lymphomas (45,428,445–452).

Anticipation has been extensively documented in familial lymphoid cancers, including NHL (428,448), HL (45,428), CLL (449–451) and MM (452). In neurodegenerative diseases such as Huntington's disease, the molecular basis of genetic anticipation has been attributed to expansion of trinucleotide repeat sequences (443,444). Although trinucleotide repeat expansion is a widely accepted explanation of the anticipation phenomenon in certain Mendelian diseases, the molecular mechanism remains unknown in cancer. Telomere shortening is a suggested alternative mechanistic explanation, but the evidence is unclear.

Although there is evidence for the existence of anticipation in lymphoid cancer families, this phenomenon is not confidently accepted by the scientific community due to the potential for ascertainment bias, cohort effects, and unknown molecular mechanism.

Ascertainment bias is a systematic misrepresentation in measuring the true frequency of a phenomenon due to data collection methods that target a specific population or group. The resulting study sample may be systematically different from the target population which biases or skews the association (444). Some multiplex families that show anticipation may be a result of ascertainment biases attributed to an overrepresentation of individuals with a younger age-of-onset; however, a different age of onset may reflect generational differences, also known as a cohort effect (444,453).

Cohorts have unique characteristics confounded by age and period effects which may influence disease penetrance (444,453). Health outcomes, according to the year of birth, may coincide with shifts in population exposures to risk factors over time (453). Changing

environmental factors or exposures that are unequally distributed in the population may cause different patterns of disease risk. Improvements in cancer screening programs may also contribute to a cohort effect.

The occurrence of multiple lymphoid cancers in a family and an observed earlier age of cancer onset is suggestive of shared susceptibility factors. Heritability estimates for lymphoid cancers explain a limited proportion of cases, and are complicated by the diversity of hematological subtypes. Reported heritability estimates vary by NHL subtype, with common SNPs explaining approximately 8% of MZL (49), 21% of FL (49), 9-16% of DLBCL (42,49,50), and 16-34% of CLL heritability (42,49,51–54)**.** Common SNPs explained 17% of MM heritability (46–48), and a larger proportion of HL heritability (21-48%) (43–45). Less common histological subtypes such as MCL and T-cell NHLs do not have an estimated SNP-heritability due to lack of GWA-studies.

In a collection of 200 multiple-case lymphoid cancer families, we examined age of onset patterns across generations and lymphoid cancer subtypes while adjusting for several types of ascertainment bias.

4.2 Methods

4.2.1 Study population

Briefly, families were eligible for inclusion if they contain a member diagnosed with lymphoma and at least one additional relative with a lymphoid cancer. Patients with the cancers of interest (NHL, HL, CLL and MM) were invited to participate. More information on eligibility and recruitment of families is provided in **Chapter 2.2**: **Eligibility and recruitment** and **Chapter 2.3: Data quality control** (Methods), page 37.

4.2.2 Data collection

Information regarding data collection is provided in **Chapter 2.4: Data collection** (Methods), page 46. Briefly, information about lymphoid malignancies was obtained systematically. Reported lymphoid cancer diagnoses were confirmed using medical records or

histopathology slides whenever possible. Lymphoid cancers were classified according to the InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research (404).

Each family member was assigned a generation number that depended on their relationship with a lymphoid cancer case (**Figure 10**). Generation 1 represents the first reported lymphoid cancer case or suspected carrier status of lymphoid cancer in each specific family. In instances where the oldest generation contained two or more lymphoid cases (e.g., siblings), a generation label of 2 was assigned as we expect one or both of their parents to be a carrier of a genetic factor that caused lymphoid cancer.

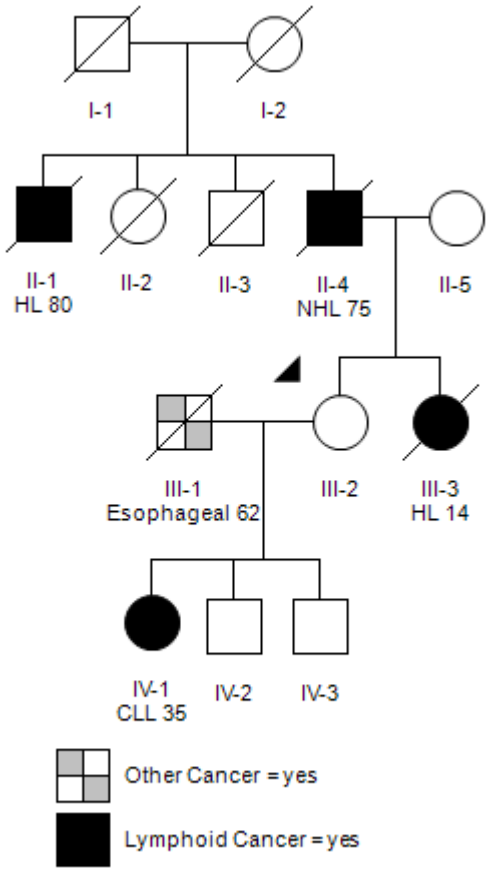


Figure 10: Example pedigree.
Notes: # indicates age of onset (years). Arrow indicates the spokesperson of the family.
Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia.

SEER (USA) population data (86,439,454) were used to control for known covariates that affect lymphoid cancer risk. Incidence data for each combination of sex, age of diagnosis, ethnicity and histological subtype that we observed in our study was converted to a cumulative percentile distribution. Using the cumulative percentile in replacement of age of onset allowed for the uniform comparison of heterogeneous data. In the absence of subtype or ethnicity information, distributions utilizing the more specific information were applied (e.g., B-cell NHL). The percentiles were plotted in replacement of the age of onset variable, and all permutation tests were repeated. SEER data were accessed through SEER*Stat software (442).

4.2.3 Statistical analysis

R version 3.5 was used for analysis.

A chi-square goodness-of-fit test was used to assess if the observed familial sex distribution resembled the population sex distribution (R function: chisq.test). A sign test was used to assess if the observed familial median ages of onset were less than the population median age of onset (R package: signmedian.test).

Median ages of onset were obtained using SEER (USA) population data because comparable Canadian population data were not available.

To examine the effect of lymphoid cancer type on the age of onset, an analysis of variance was performed, assuming the same variance across cancer types. For analysis of anticipation, an F-statistic was used to assess the association between age of onset and generation for each lymphoid subtype. Family members are not independent of each other, so a permutation test was performed. For each cancer type, the age of onset was randomly permuted 10,000 times within each family, for all families, to generate a reference distribution of F-statistics on which to calculate the *p*-value. Permutations were repeated using cumulative percentiles in replacement of age of onset.

Possible ascertainment bias was corrected using 3 methods: 1) removing cases diagnosed before reproductive age (≤ 25 years old), as these individuals may have limited reproductive capabilities which results in oversampling of parent generations (or under-sampling their children); 2) removing probands, as probands are more likely to be self-selected because of an earlier age at diagnosis; and 3) removing cases with a short duration of follow-up, as insufficient time has elapsed for normal or late-disease development among family members in

similar birth cohorts. A short duration was classified as a current age below the population median age of onset. Statistical tests were repeated while controlling for ascertainment bias. Generations with 2 or fewer cases were excluded from the analysis. Subtypes with 10 or fewer cases were not analyzed.

An initial age of onset analyses was performed on 353 lymphoid cancer cases from 140 multiple-case families that were ascertained between 2006 and 2014. Due to smaller sample size, analyses were limited to common lymphoid cancer types (e.g., NHL, HL, CLL and MM). Sample size did not permit corrections for ascertainment biases. This analysis was published in 2017 in *Leukemia & Lymphoma.*

An additional 174 cases from 60 multiple-case lymphoid cancer families were recruited between 2014 and 2018. A larger sample size enabled the analyses of several subtypes, including DLBCL, FL, LPL/WM, MCL, MZL, T-cell NHL, CHL and NS, and allowed for the correction of several types of ascertainment biases. This analysis is presented in this Chapter.

A separate person-time analysis was performed for lymphoid and non-lymphoid cancers in a subset of 140 families. All first-degree relatives of cancer cases who are possible carriers of a putative genetic susceptibility factor were included in the respective person-time analysis. Families with sibling only cases do not contain enough information as to which parent contributes to cancer susceptibility; for such cases only the youngest parent's age was used. Person-years per event was calculated from the number of events and the total relevant person years, for each generation (**Supplementary Table B.1**).

4.3 Results

We report on 527 lymphoid cancer cases in 200 multiple-case lymphoid cancer families (**Table 7**). Most cases were of white ethnicity and resided in Canada or the USA. Cases were predominantly male, with the exception of FL, MZL and NS subtypes, which were 56.1%, 66.7%, and 77.8% female, respectively. The median age of onset was 57 years; 385 cases (73.1%) reported a subtype of lymphoid cancer. Remaining cases were classified as HL, not otherwise specified (NOS) (46, 8.7%) or NHL, NOS (96, 18.2%). Medical records and/or histopathology slides were available for 252 cases (47.8%), all of which supported the self-reported diagnosis.

Table 7: Demographic and medical data.

Characteristic	**All cases**, n (%)
No. cases	527
Age of diagnosis, y	
Mean ± SD	55 ± 17
Median	57
Range	2 - 93
< 30	55 (10.4)
30 - 39	41 (7.8)
40 - 49	62 (11.8)
50 - 59	131 (24.9)
60 - 69	115 (21.8)
70 - 79	86 (16.3)
≥ 80	27 (5.1)
Unknown	10 (1.9)
Sex	
Male	281 (53.3)
Female	246 (46.7)
Ethnicity	
White	507 (96.2)
Asian	17 (3.23)
Mixed	2 (0.38)
Unknown	1 (0.19)
Location[1]	
Canada	469 (88.6)
USA	35 (7.0)
Europe	19 (3.6)
Other	4 (0.8)
Diagnosis validated, n	
Medical records, 144	144 (100)
BCC pathologist, 2	2 (100)
Both, med. & path., 106	106 (100)
No records available, 275	-
Subtypes	
NHL	261 (49.5)
B-cell NHL[2]	304 (57.7)
DLBCL	34 (6.5)
FL	57 (10.8)
LPL/WM	20 (3.8)
MCL	9 (1.7)
MZL	12 (2.3)
CLL	152 (28.8)
B-cell, NOS	91 (17.3)
T-cell	9 (1.8)
HL	79 (15.0)
CHL	30 (5.7)
NS	27 (5.1)
MM	35 (6.6)
Generation	
1	99 (18.8)
2	264 (50.1)
3	131 (24.9)
4	31 (5.9)
5	2 (0.4)

Notes: [1] Europe includes England (n=9), Ireland (n=3), Sweden (n=3), Croatia (n=1), Netherlands (n=1), Scotland (n=1), and Wales (n=1). Other locations include Australia (n=2), El Salvador (n=1), and Iran (n=1). [2] B-cell NHL includes CLL cases, whereas NHL does not include CLL cases.

Abbreviations: y, years; SD, standard deviation; USA, United States of America; BCC, British Columbia Cancer; med, medical records; path, BCC pathologist.

Subtype abbreviations: NHL, non-Hodgkin lymphoma; DLBCL, Diffuse large B-cell lymphoma; FL, Follicular lymphoma; LPL/WM, Lymphoplasmacytic lymphoma/Waldenström macroglobulinemia; MCL, Mantle cell lymphoma; MZL, Marginal zone lymphoma; CLL, chronic lymphocytic leukemia; HL, Hodgkin lymphoma; CHL, Classic HL; NS, nodular sclerosing; MM, multiple myeloma; NOS, not otherwise specified.

4.3.1 Age of onset differs by type of lymphoma

Age of diagnosis distributions for all familial lymphoid cancer cases had a unimodal distribution, with the exception of HL. The mean age of diagnosis varies by type of lymphoid cancer in these families (**Figure 11**.**A**; $p < 0.0001$). Relative to the general population and after controlling for sex, ethnicity and subtype, the mean age of diagnosis percentiles was different for familial lymphoid cancers (**Figure 11**.**B**; $p = 0.0037$).

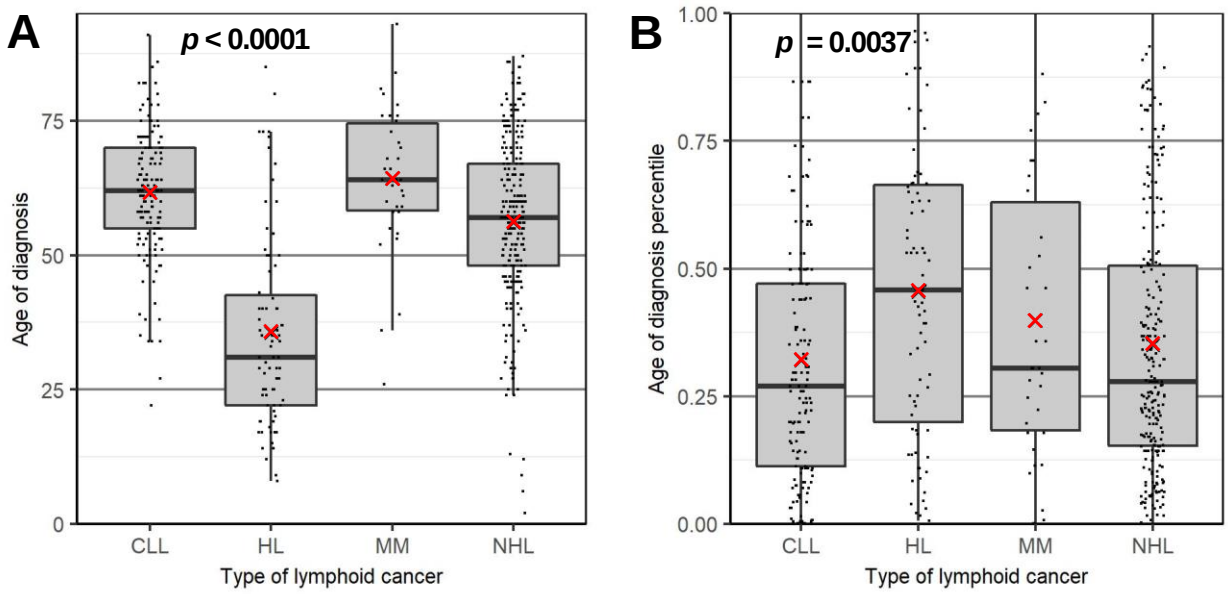


Figure 11: Distributions by (**A**) age of diagnosis for NHL, HL, CLL and MM ($p < 0.0001$) and (**B**) age of diagnosis percentile for NHL, HL, CLL and MM ($p = 0.0037$).
Notes: Red x's indicate the means; the box-plots show the median (black bold lines) and interquartile range (grey box) for each type of lymphoid cancer. Each dot represents one case.
Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

4.3.2 Earlier age of onset in families

Familial NHL, HL and MM cases were diagnosed earlier than population cases, as were DLBCL, FL, CLL, MCL, and MZL subtypes (for age of onset and percentiles). The sign test (**Table 8**) indicates that the median age of onset of lymphoma in multiple-case families was less than the population median age of onset ($p < 0.0001$) for all lymphoid cancers considered as a group. The association was also statistically significant for NHL, HL and major subtypes (DLBCL, FL, CLL and MZL). There was no statistically significant association for LPL/WM or MCL (age or percentile)

Table 8: Age of onset (AoO) and AoO percentiles in multiple-case lymphoid cancer families after controlling for ascertainment bias.

	SEER Population		All cases					Controlling for ascertainment bias: 1. Reproductive (dx <25 yrs) cases removed					Controlling for ascertainment bias: 2. Probands removed					Controlling for ascertainment bias: 3. Short duration of follow-up removed				
	AoO	per	n	Median AoO	*p*	Median per	*p*	n	Median AoO	*p*	Median per	*p*	n	Median AoO	*p*	Median per	*p*	n	Median AoO	*p*	Median per	*p*
Lymphoma	65	50%	483	57	**	30%	**	426	58	**	33%	**	206	58	**	36%	**	228	64		46%	**
NHL[1]	66	50%	254	57	**	28%	**	234	57	**	29%	**	106	60	**	35%	**	110	65		43%	**
B-cell NHL[2]	66	50%	304	58	**	26%	**	287	59	**	27%	**	107	60	*	28%	**	129	67		41%	†
DLBCL	70	50%	34	53.5	**	26%	**	27	56	**	28%	**	8	57.5	†	28%	†	9	60		33%	
FL	60	50%	57	56	**	31%	**	54	55.5	**	32%	**	18	65		57%		19	60		48%	
LPL/WM	51	50%	18	60		22%		18	60		22%	*	9	61		25%		7	67		37%	
MCL	68	50%	9	65		38%		8	64.5		37%	†	4	54		17%		5	68		49%	
MZL	60	50%	12	53	**	17%	**	11	53	**	17%	**	5	52		15%	†	7	53		17%	**
CLL	71	50%	150	62	**	27%	**	149	62	**	27%	**	57	64	**	28%	**	72	68.5	*	44%	**
T-cell NHL	65	50%	9	45	*	28%	*	9	45	*	28%	*	5	45		18%		2	48.5		21%	
HL	39	50%	79	31	**	46%		43	38		63%	**	43	31	*	41%		46	36		53%	
CHL[3]	40	50%	30	32.5	**	53%		18	35	*	63%	*	10	35	*	55%		14	35	†	63%	
NS	28	50%	27	31	†	54%		17	35	**	63%	**	9	35		56%		12	35.5	*	64%	
Myeloma	69	50%	34	64		30%		33	64		30%		17	64		30%		22	68		46%	
All cases		50%	517			30%	**	459	59		32%	**	223	58		35%	**	250	64		46%	**

Significance: ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$.
Notes: [1] NHL excluding CLL cases. [2] B-cell NHL includes CLL cases. [3] CHL includes cases with additional subtype information (e.g., CHL NS, mixed cellularity, lymphocyte-rich or lymphocyte depleted) and does not include CHL, not otherwise specified (NOS) cases.
Abbreviations: SEER, Surveillance, Epidemiology, and End Results; AoO, age of onset; per, percentile; n, sample size; *p*, *p*-value.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; LPL/WM, lymphoplasmacytic lymphoma/Waldenström macroglobulinemia; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; CLL, chronic lymphocytic leukemia; HL, Hodgkin lymphoma; CHL, classic Hodgkin lymphoma; NS, nodular sclerosis; MM, multiple myeloma.

cases, possibly due to small sample size. Relative to the general population, and after controlling for sex, ethnicity, and subtype, the median age of onset percentiles followed a similar pattern as age of onset.

To control for possible ascertainment bias, cases diagnosed before the age of 25 years were excluded. In total, 42 cases with a young age of onset were removed, and an additional 26 cases were excluded from families which no longer had 2 or more lymphoid cancer cases. Familial NHL cases (and B-cell NHL, DLBCL, FL, CLL and T-cell subtypes) were still diagnosed earlier than sporadic lymphoma cases (age and percentiles). Familial MCL ages and percentiles were diagnosed earlier than population cases; however, the relationships were not significant. After excluding young HL cases, familial HL and CHL cases were diagnosed 1 and 5 years earlier than population cases; however, the relationship was only statistically significant for CHL cases. In contrast, excluding young onset HL percentiles cases caused the relationship to reverse, such that familial HL, CHL, and NS percentiles were diagnosed later than population percentiles; these relationships were statistically significant. The relationship was statistically significant for NHL (excluding CLL), DLBCL, FL, MZL and CLL (age and percentiles); while only LPL/WM percentiles were significant and MCL percentiles were marginally significant. Familial MM cases were diagnosed earlier than population ages of onset and percentiles; however, the relationships were not significant.

As another control for ascertainment bias, we removed all 187 probands and also other affected members of 78 families who no longer had 2 or more cases after removal of the probands. No lymphoid cancer cases were removed from 13 families that had an unaffected relative as the spokesperson. With the exception of FL and HL (and subtypes), familial lymphoid cancers were diagnosed earlier than population cases; however, the relationship was statistically significant for all lymphoid cancers considered together, and NHL, B-cell NHL and CLL subtypes (age and percentiles); while DLBCL (age and percentiles) and MZL (percentiles only) were marginally significant. The inverse relationship was not significant for LPL/WM or T-cell lymphoma. With the exception of NS HL, ages of onset for familial HL and CHL cases were diagnosed earlier than population cases, and the relationship was statistically significant. Familial HL percentiles were diagnosed earlier than population percentiles, while familial CHL and NS percentiles were diagnosed later than population percentiles; these relationships were not statistically significant. Familial MM cases were diagnosed earlier than population ages of onset and percentiles; however, the relationships were not significant.

Another correction for ascertainment bias was attempted by excluding 267 cases with a short duration of disease development and other affected members who no longer had 2 or more cases. With the exception of HL (and subtypes), all lymphoid cancer groups and subtypes were diagnosed earlier than the population median percentile age of onset; however, the relationship was only significant for all lymphoid cancers considered as a group, NHL, CLL and MZL. Most familial lymphoid cancer subtypes were diagnosed at or below the population median age of onset; however, the relationship was only significant for familial CLL cases. Familial HL and CHL cases were diagnosed earlier than population cases; however, the relationship was only marginally significant for CHL cases. Familial NS cases were diagnosed later than the population median age of onset; this relationship was statistically significant. Familial HL, CHL and NS percentiles were diagnosed later than population percentiles; however, these relationships were not statistically significant. Familial MM cases were diagnosed earlier than population cases (ages and percentiles); however, neither relationship was statistically significant.

4.3.3 Anticipation

The mean age of onset was significantly younger for later generations for all lymphoid cancers considered together ($p < 0.0001$), and separately for NHL ($p < 0.0001$), HL ($p = 0.0001$), and CLL ($p = 0.0048$) but not MM ($p = 0.2515$; **Figure 12A**; **Table 9**). Anticipation was observed for several lymphoid cancer subtypes including all lymphoid cancers, NHL, HL, B-cell NHL, DLBCL, LPL/WM, and CHL, while NS was marginally significant. FL and MM subtypes were not different from the population, while MCL, MZL and T-cell NHL lacked sufficient sample size for generational analyses. The anticipation effect was largely unchanged for all lymphoid cancers ($p < 0.0001$), NHL ($p < 0.0001$), HL ($p = 0.0003$), CLL ($p = 0.0053$), and MM ($p = 0.2453$) when using percentiles in replacement of age of onset to control for sex, ethnicity and subtype (when necessary) (**Figure 12B**).

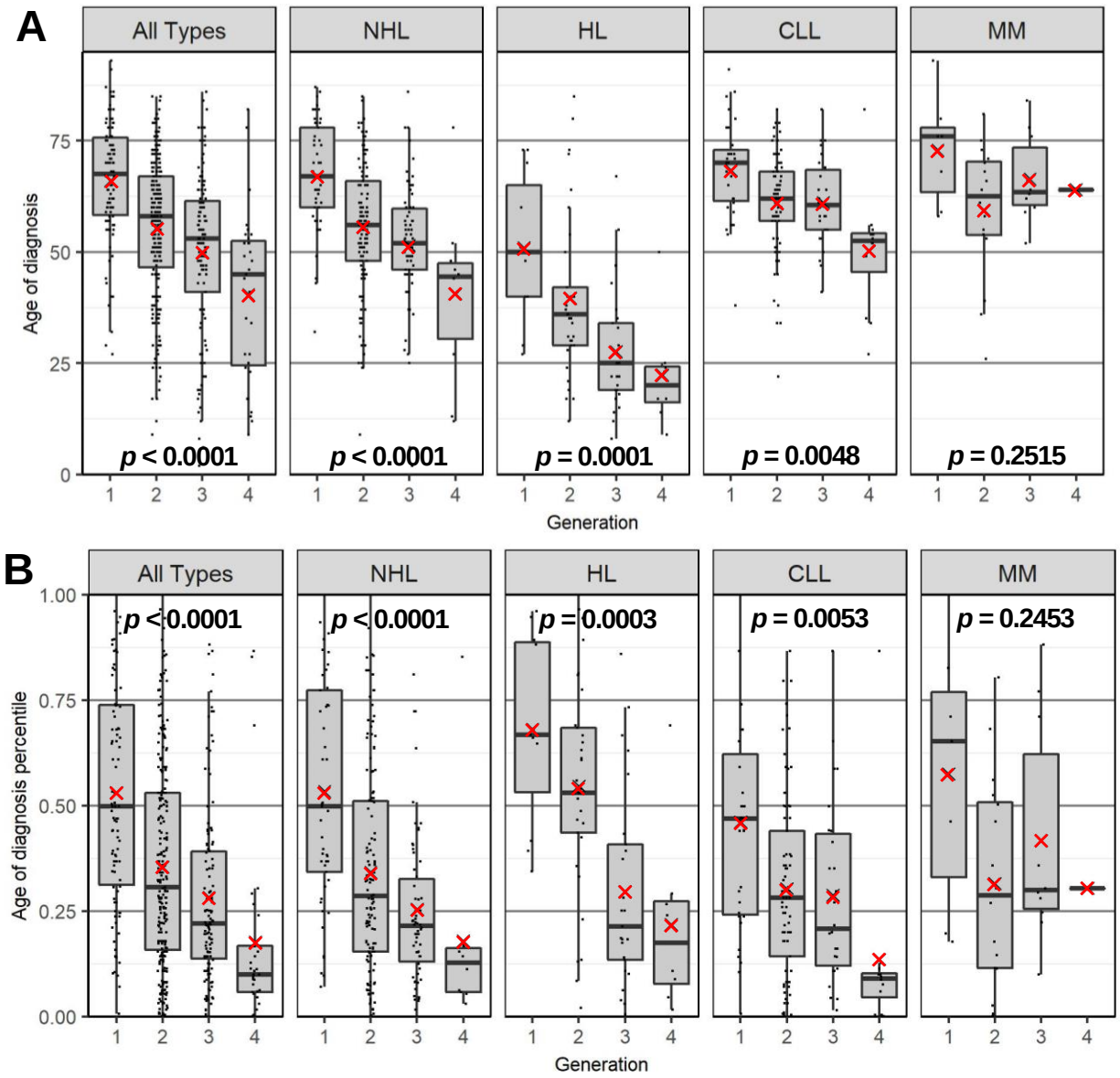


Figure 12: Distributions by generation for (**A**) age of diagnosis for all lymphoid cancers ($p < 0.0001$), NHL ($p < 0.0001$), HL ($p = 0.0001$), CLL ($p = 0.0048$) and MM ($p = 0.2515$), and (**B**) age of diagnosis percentile for all lymphoid cancers ($p < 0.0001$), NHL ($p < 0.0001$), HL ($p = 0.0003$), CLL ($p = 0.0053$) and MM ($p = 0.2453$).
Notes: Red x's indicate the means, the box-plots show the median (black bold line) and interquartile range (grey box) for each type of lymphoid cancer. Each dot represents one case.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

Generational effects are described in detail in the following three tables: all lymphoid caners considered together, NHL, HL, CLL and MM in **Table 9**. NHL and subtypes such as B-cell NHL, DLBCL, FL, LPL/WM, and B-cell NOS in **Table 10**. HL and subtypes such as CHL and NS in **Table 11**. Anticipation effects in the preceding three tables are summarized in **Table 12**.

Anticipation effects were observed among all lymphoid cancers considered as a group, NHL, HL and CLL, but not MM cases. After controlling for 3 types of ascertainment bias, all lymphoid cancers, and NHL cases, were diagnosed earlier in later generations (**Table 9**), and the anticipation effects were statistically significant. Statistically significant anticipation effects were also observed for familial CLL cases after controlling for reproductive ascertainment bias, whereas a marginal trend toward significance was observed after controlling for proband ascertainment bias. Familial HL cases showed evidence of anticipation after controlling for three types of ascertainment biases; however, anticipation effects were only statistically significant after controlling for proband ascertainment bias and short duration of follow-up. MM cases did not show evidence of anticipation before or after controlling for ascertainment biases.

Table 9: Anticipation effects for all lymphoid cancers, NHL, HL, CLL and MM after controlling for ascertainment biases.

				Controlling for ascertainment bias								
	All cases			1. Reproductive (dx < 25 yrs) cases removed			2. Probands removed			3. Short duration of follow-up removed		
	n	Median AoO	Median percentile	n	Median AoO	Median percentile	n	Median AoO	Median percentile	n	Median AoO	Median percentile
All types[1]												
Gen 1	99	67	49.8	88	67.5	49.8	31	67	46.2	44	70	53.2
Gen 2	264	57	22.9	233	58	32.1	79	61	47.0	121	65	49.3
Gen 3	131	53	22.0	102	55	24.2	68	49.5	22.1	59	59	29.5
Gen 4	31	45	9.92	22	51	9.92	16	34.5	9.11	7	52	30.4
Gen 5	2	-	-	2	-	-	2	-	-	1	-	-
p-value		**<0.0001**	**<0.0001**		**<0.0001**	**<0.0001**		**<0.0001**	**<0.0001**		**0.0003**	**0.0064**
NHL[2]												
Gen 1	49	67	49.9	45	67	50.4	15	70	53.3	19	74	63.8
Gen 2	132	56	28.5	117	57	29.9	43	65	42.1	58	65	44.8
Gen 3	70	52	21.8	60	52.5	21.8	31	49	19.5	24	60	32.4
Gen 4	10	44.5	12.7	8	45.5	12.7	7	44	14.2	4	46.5	16.2
p-value		**<0.0001**	**<0.0001**		**<0.0001**	**<0.0001**		**<0.0001**	**0.0003**		**0.0017**	**0.0048**
HL												
Gen 1	11	50	66.8	8	50.5	67.5	3	29	39.3	7	50	66.8
Gen 2	33	36	53.1	24	37.3	55	19	40	54.5	19	40	63.2
Gen 3	25	25	21.4	8	36	50.4	17	25	25.1	18	25	23.2
Gen 4	8	20	17.4	1	-	-	5	24	26.7	1	-	-
Gen 5	2	-	-	2	-	-	2	-	-	1	-	-
p-value		**0.0001**	**0.0003**		0.0951	0.2239		**0.0375**	**0.0458**		0.0817	**0.0485**
CLL												
Gen 1	31	69	45.5	28	69	45.5	13	70	43.7	17	70	44.0
Gen 2	83	62	27.6	76	62	27.6	26	66.5	34.5	43	68	38.5
Gen 3	26	60	20.0	25	60	20.0	16	59	17.9	13	69	43.9
Gen 4	12	52.5	9.02	12	52.5	9.02	4	42	3.2	1	-	-
p-value		**0.0048**	**0.0053**		**0.0043**	**0.0068**		0.0590	0.0532		0.1692	0.2150
MM												
Gen 1	8	76	65.2	7	76	65.2	4	72	55.7	4	72	55.7
Gen 2	16	62.5	28.7	16	63.5	28.7	6	62.5	28.7	10	66	38.3
Gen 3	10	63.5	30.0	9	63	29.5	7	62	29.5	6	71	53.4
Gen 4	1	-	-	1	-	-	-	-	-	1	-	-
p-value		0.2515	0.2453		0.2463	0.2453		0.3306	0.6693		0.4917	0.5032

Notes: [1]All types includes NHL, HL, CLL and MM. [2] NHL (excluding CLL cases).

Abbreviations: AoO, age of onset; dx, diagnosis; yrs, years; n, sample size.

Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

The age of onset was earlier in later generations among familial B-cell NHL (including CLL), DLBCL, FL, LPL/WM and B-cell NOS cases (ages and percentiles; **Table 10**); however, the relationship was not significant among FL cases (ages and percentiles). After controlling for 3 types of ascertainment bias, and when sample size permitted, the age of onset was earlier in later generations for familial B-cell NHL (including CLL), DLBCL, FL, LPL/WM, and B-cell NHL, NOS cases (ages and percentiles). Anticipation was statistically significant among familial B-cell NHL (including CLL) cases after controlling for three types of ascertainment bias. Familial DLBCL and LPL/WM cases also showed anticipation which remained statistically significant after controlling for reproductive ascertainment bias (ages and percentiles).

Sample size did not permit analysis of DLBCL or LPL/WM cases after controlling for proband ascertainment bias or short-duration of follow-up (see summary **Table 12**). Familial FL cases appear to be earlier in onset in later generations after controlling for ascertainment biases; however, the relationships were not statistically significant. MCL, MZL and T-cell NHL were not examined due to insufficient sample size (summary **Table 12**). B-cell NHL, NOS cases showed anticipation which remained statistically significant after controlling for reproductive and proband ascertainment biases (ages and percentiles). B-cell NHL NOS cases were earlier in onset in later generations after controlling for short duration of follow-up; however, the relationship was not statistically significant. The anticipation observations were largely unchanged when using percentiles in replacement of age of onset.

Table 10: Anticipation effects for familial NHL and NHL subtypes after controlling for ascertainment biases.

				Controlling for ascertainment bias								
	All cases			1. Reproductive (dx < 25 yrs) cases removed			2. Probands removed			3. Short duration of follow-up removed		
	n	Median AoO	Median percentile	n	Median AoO	Median percentile	n	Median AoO	Median percentile	n	Median AoO	Median percentile
NHL[1]												
Gen 1	49	67	49.9	45	67	50.4	15	70	53.3	19	74	63.8
Gen 2	132	56	28.5	117	57	29.9	43	65	42.1	58	65	44.8
Gen 3	70	52	21.8	60	52.5	21.8	31	49	19.5	24	60	32.4
Gen 4	10	44.5	12.7	8	45.5	12.7	7	44	14.2	4	46.5	16.2
p-value		**<0.0001**	**<0.0001**		**<0.0001**	**<0.0001**		**<0.0001**	**0.0003**		**0.0017**	**0.0048**
B-cell NHL[2]												
Gen 1	61	68	46.9	57	68	67	21	70	49.8	23	70	49.8
Gen 2	163	58	25.9	153	58	27	51	60	35.9	74	67	40.7
Gen 3	65	55	20.9	60	55.5	20.9	31	58	16.1	31	60	27.8
Gen 4	18	52	9.7	17	52	9.4	8	44.5	6.1	5	52	17.0
p-value		**<0.0001**	**<0.0001**		**<0.0001**	**<0.0001**		**0.0029**	**0.0036**		**0.0041**	**0.0089**
DLBCL												
Gen 1	5	64	46.5	4	71	59.4	-	-	-	1	-	-
Gen 2	22	53.5	25.7	18	55.5	26.8	5	60	27.9	6	58.5	30.1
Gen 3	7	48	21.1	5	48	21.1	3	27	22.7	2	-	-
p-value		**0.0006**	**0.0011**		**0.0031**	**0.0016**						
FL												
Gen 1	12	68.5	62.1	10	68.5	62.1	4	77	85.9	3	79	90.8
Gen 2	29	56	31.8	28	56	32.1	9	58	42.1	11	58	42.1
Gen 3	13	51	20.7	13	51	20.7	3	46	15.2	3	49	20.7
Gen 4	3	46	11.2	3	46	11.2	2	-	-	2	-	-
p-value		0.1241	0.2499		0.5046	0.1243		0.5091	1.00		1.00	1.00
LPL/WM												
Gen 1	7	67	36.8	7	67	36.8	2	-	-	1	-	-
Gen 2	7	57	17.0	7	57	17.0	5	60	22.5	5	60	22.5
Gen 3	6	48.5	7.0	4	48.5	7.0	3	68	39.7	3	68	39.7
p-value		**0.0349**	**0.0328**		**0.0326**	**0.0343**						
B-cell, NOS[3]												
Gen 1	19	70	53.3	19	70	53.3	8	66	45.0	13	70	61.0
Gen 2	48	63	38.9	43	63	41.0	25	67	53.4	33	68	55.9
Gen 3	26	54.5	21.8	23	55	21.8	19	52	20.3	10	64.5	43.4
Gen 4	3	27	5.7	2	-	-	3	27	7.7	-	-	-
p-value		**0.0048**	**0.0014**		**0.0023**	**0.0076**		**0.0008**	**0.0015**		0.6089	0.4113

Notes: [1] NHL (excluding CLL cases). [2] B-cell NHL (includes CLL cases). [3] B-cell, NOS includes B-cell NHL subtypes with no additional subtype information.
Abbreviations: AoO, age of onset; dx, diagnosis; yrs, years; n, sample size.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; LPL/WM, lymphoplasmacytic lymphoma/ Waldenström macroglobulinemia; NOS, not otherwise specified.

Familial ages of onset and percentiles were younger among later generations for individuals with HL, CHL and NS (**Table 11**); however, the relationship was statistically significant for HL (ages and percentiles) and CHL (ages but not percentile) cases, and marginally significant

for NS (ages and percentile) cases. After controlling for short duration of follow-up, CHL cases appeared to be diagnosed earlier; however, the relationship was not statistically significant. We were unable to examine reproductive or proband ascertainment biases due to limited sample size.

Table 11: Anticipation effects for familial HL and HL subtypes after controlling for ascertainment biases.

				Controlling for ascertainment bias								
	All cases			1. Reproductive (dx < 25 yrs) cases removed			2. Probands removed			3. Short duration of follow-up		
	n	Median AoO	Median percentile	n	Median AoO	Median percentile	n	Median AoO	Median percentile	n	Median AoO	Median percentile
HL												
Gen 1	11	50	66.8	8	50.5	67.5	3	29	39.3	7	50	66.8
Gen 2	33	36	53.1	24	37.3	55	14	40	54.5	19	40	63.2
Gen 3	25	25	21.4	8	36	50.4	19	25	25.1	18	25	23.2
Gen 4	8	20	17.4	1	-	-	5	24	26.7	1	-	-
Gen 5	2	-	-	2	-	-	2	-	-	1	-	-
p-value		**0.0001**	**0.0003**		0.0951	0.2239		**0.0375**	**0.0458**		0.0817	**0.0485**
CHL[1]												
Gen 1	4	54	76.4	2	-	-	-	-	-	2	-	-
Gen 2	16	32.5	55.0	12	35.5	64.1	4	35	63.2	8	35.5	64.1
Gen 3	6	26	23.4	2	-	-	4	31	32.8	3	25	18.5
Gen 4	2	-	-	-	-	-	-	-	-	-	-	-
Gen 5	2	-	-	2	-	-	2	-	-	1	-	-
p-value		**0.0380**	0.1383								0.3731	0.5600
NS HL												
Gen 1	3	60	88.1	2	-	-	-	-	-	2	-	-
Gen 2	16	32.5	55.0	12	35	59.6	4	35	59.6	8	35.5	64.1
Gen 3	4	22	24.1	1	-	-	3	27	37.3	1	-	-
Gen 4	2	-	-	-	-	-	-	-	-	-	-	-
Gen 5	2	-	-	2	-	-	2	-	-	1	-	-
p-value,		0.0811	0.0610									

Notes: [1] CHL includes cases with additional subtype information (e.g., CHL NS, mixed cellularity, lymphocyte-rich or lymphocyte depleted) and does not include CHL, NOS cases.
Abbreviations: AoO, age of onset; dx, diagnosis; yrs, years; n, sample size.
Subtype abbreviations: HL, Hodgkin lymphoma; CHL, classic Hodgkin lymphoma; NS, nodular sclerosis.

For lymphoid cancers, person-time analysis supports a reduction in person-year per event from generation 1 through 3, but not generation 4 (**Supplementary Table B.1**). In contrast, person-time analysis of non-lymphoid cancers within these families shows no trend to younger onset in later generations.

Table 12: Summary of anticipation effects in multiple-case lymphoid cancer families after controlling for ascertainment bias.

						Controlling for ascertainment bias														
	All cases					1. Reproductive (dx <25 yrs) cases removed					2. Probands removed					3. Short duration of follow-up removed				
Subtype	n	Median AoO *p*	*p*	Median per *p*	*p*	n	Median AoO *p*	*p*	Median per *p*	*p*	n	Median AoO *p*	*p*	Median per *p*	*p*	n	Median AoO *p*	*p*	Median per *p*	*p*
Lymphoma																				
NHL[1]	261	**<0.0001**	**	**<0.0001**	**	230	**<0.0001**	**	**<0.0001**	**	96	**<0.0001**	**	**0.0003**	**	105	**0.0017**	**	**0.0048**	**
B-cell NHL[2]	307	**<0.0001**	**	**<0.0001**	**	287	**<0.0001**	**	**<0.0001**	**	111	**0.0029**	**	**0.0036**	**	133	**0.0041**	**	**0.0089**	**
DLBCL	34	**0.0006**	**	**0.0011**	**	27	**0.0031**	**	**0.0016**	**	8	-		-		9	-		-	
FL	57	0.1241		0.2499		54	0.5046		0.1243		18	0.5091		1.00		19	1.00		1.00	
LPL/WM	20	**0.0349**	*	**0.0328**	*	18	**0.0326**	*	**0.0343**	*	8	-		-		8	-		-	
MCL	7	-		-		5	-		-		3	-		-		3	-		-	
MZL	11	-		-		10	-		-		3	-		-		6	-		-	
CLL	152	**0.0048**	**	**0.0053**	**	148	**0.0043**	**	**0.0068**	**	59	0.0590	†	0.0532	†	73	0.1692		0.2150	
B-cell, NOS	96	**0.0048**	**	**0.0014**	**	87	**0.0023**	**	**0.0076**	**	55	**0.0008**	**	**0.0015**	**	56	0.6089		0.4113	
T-cell NHL	8	-		-		8	-		-		3	-		-		0	-		-	
HL	77	**0.0001**	**	**0.0003**	**	38	0.0951	†	0.2239	**	41	**0.0375**	*	**0.0458**	*	43	0.0817	†	**0.0485**	*
CHL[3]	30	**0.0380**	*	0.1383		12	-		-		8	-		-		11	0.3731		0.5600	
NS	27	0.0811	†	0.0610	†	12	-		-		7	-		-		8	-		-	
Myeloma	34	0.2515		0.2453		32	0.2463		0.2453		17	0.3306		0.6693		20	0.4917		0.5032	
All cases	517	**<0.0001**	**	**<0.0001**	**	457	**<0.0001**	**	**<0.0001**	**	229	**<0.0001**	**	**<0.0001**	**	258	**0.0003**	**	**0.0064**	**

Significance: ** $p < 0.01$; * $p < 0.05$; † $p < 0.10$

Notes: [1] NHL excluding CLL cases; [2] B-cell NHL including CLL cases; [3] CHL includes cases with additional subtype information (e.g., CHL NS, mixed cellularity, lymphocyte-rich or lymphocyte depleted) and does not include CHL, not otherwise specified (NOS) cases.

Abbreviations: AoO, age of onset; per, percentile; n, sample size; *p*, *p*-value.

Subtype abbreviations: NHL, non-Hodgkin lymphoma; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; LPL/WM, lymphoplasmacytic lymphoma/Waldenström macroglobulinemia; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; CLL, chronic lymphocytic leukemia; HL, Hodgkin lymphoma; CHL, classic Hodgkin lymphoma; NS, nodular sclerosis; MM, multiple myeloma; NOS, not otherwise specified.

4.4 Discussion

This collection of 200 multigenerational lymphoid cancer families allows us to examine age of onset distribution patterns and deduce which cases are most likely to have shared genetic factors.

Most types of familial lymphoid malignancies occurred at a substantially younger age than sporadic cases. The median age of onset for familial NHL (and subtypes), and MM cases was lower than that of the American population medians (454). Controlling for sex, ethnicity and lymphoid cancer subtype using percentiles also supported a younger familial age of onset compared to population cases for NHL (and subtypes) and MM, and is suggestive of a genetic factor shared among cases, regardless of generation. A younger onset of familial cases is consistent with other smaller studies (427,428,450–452).

Anticipation has previously been described for familial NHL, HL and CLL cases (45,428,452,455,456). We observe anticipation across 4 generations of lymphoid cancer families for NHL, DLBCL, LPL/WM, HL and CLL cases. In MM, anticipation was not significant, possibly due to smaller sample size. Adjusting for effects of sex, ethnicity and age of diagnosis yielded results with comparable conclusions; that is, our families show anticipation across 4 generations of lymphoid cancer families. The mechanisms underlying anticipation in LPDs have not been elucidated.

Apparent anticipation may be caused by ascertainment biases, and I used three different approaches to mitigate this. Cases with an earlier age of onset may have limited reproductive opportunities, resulting in oversampling of later-onset cases in parent generations (456). Multiple individuals within each family were interviewed to collect family history, increasing the likelihood of including affected relatives who may have died at an early age. To evaluate the effect of ascertainment bias in younger generations, we removed individuals with a diagnosis at or under the age of 25 years and repeated the statistical tests; anticipation was still observed. Considering the bi-modal distribution of sporadic HL (89,457) the effect of excluding individuals at pre-reproductive age could be different in HL cases than among other lymphoid cancer types.

Cases with a younger age of onset may have earlier or more severe disease symptoms. As a result, these individuals are more likely to be recognized as familial or to be referred for inclusion in this study, thus becoming the proband who brings the family under study. As a

consequence, the ascertainment of these individuals may cause an overrepresentation of earlier age of disease onset in younger generations compared to older generations. Excluding probands from statistical approaches can reduce the impact of this ascertainment bias and allow the true anticipation phenomenon to be observed (444).

Younger generations may be enriched for earlier onset cases, as insufficient time has elapsed for normal or late-disease development among family members in similar birth cohorts. Furthermore, insufficient time has elapsed for younger generations to have children after the disease develops, and for those children to develop disease (if disease progression will occur). To reduce the impact of this ascertainment bias, shorter duration cases (and families who no longer had 2 or more lymphoid cases) were excluded. The results were consistent with other forms of ascertainment bias correction. Ultimately this supports the idea that there could be a real anticipation effect for lymphoid cancers in particular.

Thus, our observations seem unlikely to be caused by ascertainment or other bias, which suggests anticipation exists in familial lymphoid cancers. A possible non-genetic explanation of anticipation is the simultaneous exposure of parents and children or sibling-only cases to a causative environmental agent. However, this would apply to a small percentage of families in this study, as most families reported lymphoid cancers over more than 2 generations. Moreover, this fails to explain families with cases who did not live together, or cases where the parents (or later) generations were diagnosed decades after their children. Furthermore, there is no known environmental factor associated with the development of many different types of lymphoid malignancies. Only 5% of families reported a year of onset within 5 years of onset for other lymphoid affected relatives, suggesting that temporally specific environmental exposures are likely not a major effect in most families. Our observation of anticipation in NHL and HL implies that recent potential increases in surveillance for CLL through routine blood tests are not the sole explanation.

There are several possible explanations for observing anticipation in lymphoid cancer families. The accumulation of germline variants, mutations initiated by a defective DNA repair gene, and inherited telomere abnormalities or shortening could contribute to earlier disease onset across multiple generations. In a non-hematologic setting, anticipation has been explained by expansion of trinucleotide repeats (445). A variation in repeat length at the FRA16A locus has been documented in some cases of familial CLL (450), suggesting that high repeat length could affect lymphoid cancer susceptibility. Repeat expansions (even small increases in repeat size)

can affect gene transcription and cause phenotypic changes. Several leukemia genes are known to contain variable trinucleotide repeats with potential implications for predisposition. It has been hypothesized that expansion of unstable repetitive sequence could be implicated in both familial and sporadic leukemias (450,458); however, studies have failed to detect repeat expansions in CLL, leukemias, and other cancers (450,458,459). It is unlikely that trinucleotide repeat expansion plays a role in the observed anticipation of familial lymphoid cases in this study.

Limitations of this study include other types of ascertainment bias and recall bias. Families were ascertained families through physician referrals, the study website, and an annual patient education event; the study does not have a specific geographic sampling frame. Some NHL and HL cases lacked further subtype information due to absence of medical records or histopathology slides. However, our study contains detailed information on dates of birth, death, and age of onset, verified by medical records and/or an expert oncology pathologist, when possible. Information regarding cancer diagnosis was cross-checked by taking a detailed family history from multiple family members whenever possible. There was no change or misreported lymphoid cancer diagnosis among 252 cases that had medical records and/or pathology slides for examination. Subtype information was available for 63% and 42% of NHL and HL cases, respectively. Familial MM cases were limited, providing inadequate sample size to test for statistical significance.

4.5 Conclusion

The substantially earlier age of onset may reflect underlying susceptibility factors, some of which may predispose to more than one lymphoid cancer type. Earlier age of lymphoid cancer onset in this collection of multiple-case lymphoid cancer families supports the application of genomic methods to identify genes and genetic variants that underlie familial lymphoid cancers. Awareness of familial lymphoid cancer patterns and the identification of susceptibility genes has the potential to enhance screening methods for affected families in the future.

Chapter 5: Family structure, childhood environment and immune-related disorders and risk of lymphoma in lymphoid cancer families.

5.1 Introduction

Lymphoid cancers are a heterogeneous group of neoplasms that arise from immune cells. Collectively, they represent the fifth highest global incidence of cancer (460). Established risk factors include older age, male sex, ethnicity, compromised immune function, and family history of LPDs (86,461). Low-penetrance common genetic polymorphisms that affect pathways related to DNA integrity, B-cell growth and survival and xenobiotic metabolism have been implicated in the development of lymphoid neoplasms (62,386,392). Early-life environment may also modulate risk of immune-related disorders, such as allergies and autoimmune conditions, as well as some lymphoid cancers (244).

The hygiene hypothesis proposes that an early life environment that has a relative lack of exposure to microorganisms and infectious disease inhibits a child's immune system from maturing optimally (244). Consequently, such individuals are more susceptible to adult-onset immune-related disorders. Measures of family structure and crowding relate to the hygiene hypothesis as they may affect age and extent of exposure to infectious diseases, with low birth order and smaller families correlating with higher risk (244).

Associations between early birth order and/or smaller sibship size and increased risk of lymphoma have been reported for lymphoid cancers as a group (157), and separately for NHL (157,158,170,255,256) and HL (157,191,201,259,260,262). However, many other studies report no association between family structure and risk of NHL (170,191,252,255,257), HL (170,191,255,258,259,261), CLL (147,191,252) or MM (157,191,258). A few studies have observed a positive association between later birth order and NHL risk (147,252,253), and larger sibship size and risk of NHL (147,252,253,257), and MM (191). The discordant findings among studies may be partly explained by variations in study design, study population, participant response rate, selection bias, hematological subtypes assessed, or SES (254).

Few studies have examined family structure and environmental factors in the context of multiple-case lymphoid cancer families. Jønsson *et al* (2007) observed a paternal parent-offspring birth order effect with predominance of LPD in the youngest siblings among 24 pairs in 32 families enriched for CLL and B-cell malignancies (462). Royer *et al* (2010) found that familial WM cases

were more likely to have immune-related disorders such as autoimmune diseases, allergies, and specific infections among 103 familial WM and related B-cell disorders (232).

Currently, there remains a large gap in our understanding of the etiology of familial lymphoid cancers (462). We examined family structure, childhood lifestyle, and immune-related disorders among a large cohort of multiple-case lymphoid cancer families, in relation to risk of lymphoid cancer.

5.2 Methods

5.2.1 Study population

Briefly, families are eligible for inclusion if they contain a member diagnosed with lymphoma and at least one additional relative with a lymphoid cancer. For more information, refer to **Chapter 2.2**: **Eligibility and recruitment** (Methods), page 37.

5.2.2 Data collection

Information about lymphoid malignancies, family structure and demographics was obtained systematically using a questionnaire and phone interviews with multiple family members. Family structure and early childhood social environment information, such as parental education, family income, farm residence, and urban/rural residential location was reported by sibship. Personal information regarding education, medical history (allergies, autoimmune diseases, surgical procedures) and early lifestyle for each individual was obtained from a self-administered questionnaire. Allergies were classified as drug, environmental or food/diet. Autoimmune diseases were categorized as systemic, organ-specific, or conditions without detectable autoantibodies (232,414).

We report on 196 families with 524 lymphoid cancer cases among 418 sibships. Of these 418 sibships, 52 lacking family structure (birth order, sibship size) and 17 only-child cases were excluded. The remaining sibships contained 453 cases and 1112 siblings, from which 3 (0.7%) cases and 94 (8%) siblings were removed due to missing age of enrollment or sex. Analyses were conducted on 450 cases and 1018 siblings among 346 sibships.

Lymphoid cancer diagnoses were confirmed histologically or through medical records for 241 of the 450 (54%) cases. Cancers were classified according to the InterLymph hierarchical classification of lymphoid neoplasms for epidemiological research (404).

5.2.3 Statistical methods

Our study examined multiple-case families with a history of hematological malignancies and does not represent a population-based collection.

A chi-square goodness-of-fit test was performed to assess whether the observed sex distribution of the families resembled that of the Canadian population (463). American population data were used in instances where distinct histological subtype information was unavailable (464).

5.2.3.1 Standard logistic regression with generalized estimating equation

The relationship between lifestyle factors and risk of lymphoid cancer was examined using logistic regression with a generalized estimating equation (GEE) to accommodate correlated family data. Odds ratios (ORs) and 95% confidence intervals (CIs) were clustered by family and adjusted for age (continuous) and sex. Potential confounding effects of ethnicity did not change the risk estimates ≥ 10% and were not retained in the final analysis. Independent, exchangeable and autoregressive correlation structures performed similarly; the autoregressive correlation structure was used in subsequent analyses. Covariates assessed include sex, age of enrollment (n=1468), highest level of participant education (n=494), maternal and paternal education (n=759 and n=770, respectively), family income during childhood (n=756), childhood farm residence (n=801), childhood residential location (n=751), allergies (n=354), asthma (n=378), autoimmune diseases (n=378), appendectomy (n=353), and tonsillectomy (n=353). Age of death was used in replacement of age of enrollment for non-living participants. Individuals with missing age, sex, or family structure data were removed from the dataset. Due to their structural dependence, birth order and sibship size were investigated using separate GEE models. Covariates were independently assessed within each GEE model. Statistical analysis were performed using R version 3.5.

Due to variability between age-of-onset patterns, additional analyses were done with HL cases separated into childhood, young-adult or adult onset, according to Cozen *et al* (2009) (age-of-diagnosis ≤ 50, and <40) (226) and Westergaard *et al* (1997) (age-of-diagnosis <15, 15-42 and

> 42) (260). Results from these sensitivity analyses were essentially identical (**Supplementary Table C.1**) to those of all HL cases together; only outcomes based on all HL cases are presented.

5.2.3.2 Stepwise model selection

Stepwise regression (or stepwise model selection) is an automated process that selects the "best" subset of predictors for a particular outcome (e.g., lymphoid cancer). During each iteration, a predictor variable (e.g., age, sex, birth order, maternal education, allergies, etc.) will be removed or added based on the Quasilikelihood under the Independence model Criterion (QIC). The model automatically terminates after finding the best combination of predictors (with the smallest QIC). The QIC is used to compare GEE models and is analogous to the Akaike's Information Criterion (AIC) statistic used for comparing models fit with likelihood-based methods. Stepwise model selection removes unnecessary or redundant predictors that may add noise to the estimation of other variables. Models were built using Rstudio package 'spind' that utilized both backward elimination and forward selection of variables (465).

The availability of early life and disease information varied because not all members of a given sibship completed the self-administered questionnaire. For example, if one member of a sibship reported the mother's level of education, that datum was applied to all members of the sibship. In contrast, allergy information was only available if it was self-reported. The variables had different amounts of missing data and so they were separated into 3 groups of comparable sample size to retain the most information as seen in **Figure 13**.

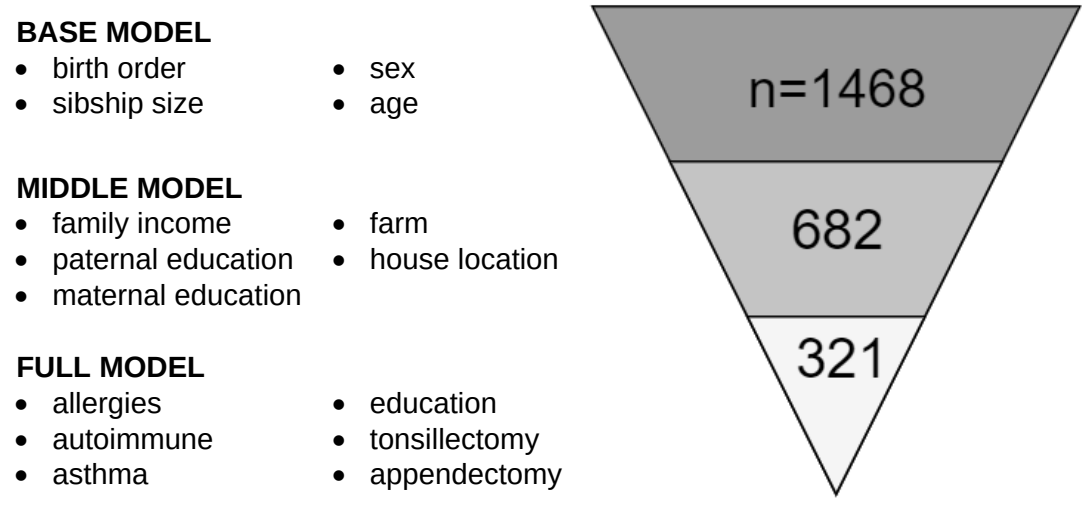


Figure 13: Lifestyle factors grouped by sample size that were used to create three models for stepwise model selection.

Three GEE models were built in a stepwise manner to investigate the relationships between lymphoid cancer risk and lifestyle factors. Complete family structure (birth order and sibship size), age of enrollment, and sex information was available for 1468 participants, which constituted the *base model*. The *middle model* (n=682) contained the base model variables and childhood environment variables: maternal education, paternal education, family income, farm residence and residential location. The *full model* (n=321) was comprised of the middle model variables in addition to personal education, allergies, autoimmune diseases, asthma, appendectomy and tonsillectomy.

5.2.3.3 Sensitivity analysis/permutation tests

A sensitivity analysis was performed by evaluating the association between family structure and lymphoid cancer risk using standard logistic regression and chi-square tests for trend using permuted pairs of independent cases and controls. Each family member was assigned a generation number relative to the founding lymphoid cancer case or presumed carrier (466). One family (of 196) was excluded from permutations due to an unmatched generation variable. The data set supported the random sampling of 95 generation-matched case/control pairs of independent families, such that a maximum of one individual per family was selected (without resampling). Ninety-five pairs were permuted 10,000 times in quadruplicate for use in: 1) logistic regression with birth order, 2) logistic regression with sibship size, 3) a chi-square test for trend in proportion with birth order, and 4) a chi-square test for trend in proportion with sibship size. The distribution of 10,000 *p*-values and coefficient estimates from each permuted analysis were plotted and compared to those observed with the full family data set. The logistic models contained base model variables (age of enrollment, sex, and birth order or sibship size).

5.3 Results

We report on 346 sibships with a lymphoid cancer affected individual within 196 multiple-case families (**Table 13**). Most participants were of white ethnicity (n=1398, 84.6%). The median age of enrollment for cases and unaffected siblings was 62 and 63 years, respectively. Of 450 lymphoid cancer cases, most were NHL (n=221, 49%), 30% were CLL (n=133), 16% were HL (n=70) and the remainder were MM (n=26, 6%). 241 (54%) cases were confirmed histologically or by medical records, all of which supported the reported diagnosis. Most families had 2 cases (n=107, 55%), followed by 3 cases (n=61, 31%) and 4 or more lymphoid cancer cases (n=28, 14%). A majority of NHL (n=96, 43%), CLL (n=62, 47%) and MM (n=14, 54%) cases were from

families that had 2 lymphoid cancer cases, while most HL cases were from families with 4 or more lymphoid cancer cases (n=26, 37%). Familial cases were 54% male; in comparison, Canadian population NHL, HL, CLL and MM cases are 55%, 57%, 61% and 59% male, respectively (463). Of 1018 unaffected siblings, 510 were male and 508 were female. Familial MM (*p*-value 0.0454) and HL (*p*-value 0.0126) cases were significantly less frequently male than population cases.

Table 13: Demographic characteristics and family structure of participants, by lymphoid cancer status

Characteristic	Unaffected, n (%)	Lymphoid affected, n (%)					Total, n
		All types	Subtypes				
			NHL	HL	CLL	MM	
Total	1018 (69.3)	450 (30.7)	221 (49.1)	70 (15.6)	133 (30.0)	26 (5.8)	1468
Sex							
Male	510 (50.1)	242 (53.8)	124 (56.1)	29 (41.4)	79 (59.4)	10 (38.5)	752
Female	508 (49.9)	208 (46.2)	97 (43.9)	41 (58.6)	54 (40.6)	16 (61.5)	716
Age of enrollment[1] (y)							
Mean ± SD	61.7 ± 19.0	61.1 ± 17.3	62.2 ± 16.9	44.8 ± 17.3	67.0 ± 12.8	66.2 ± 13.3	61.5 ± 18.5
Median	63	62	62	42	66	67	63
Range	0.5 - 108	3 - 104	3 - 104	14 - 95	24 - 93	33 - 86	0.5 - 108
< 40	116 (11.4)	55 (12.2)	18 (8.1)	30 (42.9)	5 (3.8)	<5	171
40 - 49	109 (10.7)	42 (9.3)	23 (10.4)	14 (20.0)	<5	<5	151
50 - 59	207 (20.3)	90 (20.0)	51 (23.1)	14 (20.0)	24 (18.0)	<5	297
60 - 69	213 (20.9)	110 (24.4)	51 (23.1)	5 (3.8)	43 (32.3)	11 (42.3)	323
70 - 79	200 (19.6)	94 (20.9)	44 (19.9)	<5	38 (28.6)	8 (30.8)	249
≥ 80	173 (17.0)	59 (13.1)	34 (15.4)	<5	19 (14.3)	<5	232
Birth order							
First born	208 (20.4)	128 (28.4)	66 (29.4)	18 (25.7)	35 (26.3)	9 (34.6)	336
Second born	226 (22.2)	106 (23.6)	54 (24.4)	18 (25.7)	26 (19.5)	8 (30.8)	332
Third born	169 (16.6)	97 (21.6)	51 (23.1)	14 (20.0)	28 (21.1)	<5	266
Fourth born	138 (13.6)	51 (11.3)	22 (10.0)	8 (11.4)	19 (14.3)	<5	189
Fifth or later born	277 (27.2)	68 (15.1)	28 (12.7)	12 (17.1)	25 (18.8)	<5	345
Sibship size							
Two	58 (5.7)	72 (16.0)	36 (16.3)	17 (24.3)	14 (10.5)	<5	130
Three	131 (12.9)	97 (22.6)	47 (21.3)	19 (27.1)	21 (15.8)	10 (38.5)	228
Four	180 (17.7)	90 (20.0)	51 (23.1)	10 (14.3)	27 (20.3)	<5	270
Five or more	649 (63.8)	191 (42.4)	87 (39.4)	24 (34.3)	71 (53.4)	9 (34.6)	840
Ethnicity[2]							
White	967 (95.0)	431 (95.8)	209 (94.6)	68 (97.1)	132 (99.2)	22 (84.6)	1398
Other	51 (5.0)	19 (4.2)	12 (5.4)	<5	<5	<5	70
No. cases per family							
Two	424 (41.7)	193 (42.9)	96 (43.4)	21 (30.0)	62 (46.6)	14 (53.8)	617
Three	331 (32.5)	152 (33.8)	85 (38.5)	23 (32.9)	39 (29.3)	5 (19.2)	483
Four or more	263 (25.8)	105 (23.3)	40 (18.1)	26 (37.1)	32 (24.1)	7 (26.9)	368

Notes: Cells with < 5 cases were suppressed for privacy. [1] Age at death was used for non-living participants. Family members missing age at enrollment (or age at death) or sex were excluded. [2] Ethnicity was classified using SEER race recode groups accessed through SEER*Stat (86).
Abbreviations: y, years; SD, standard deviation.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

5.3.1 Family structure

The associations between birth order position and sibship size with all lymphoid cancers are reported in **Table 14**. Birth order was inversely associated with lymphoid cancer, such that earlier birth order positions had a higher risk of lymphoma. The odds ratios were 0.62 (95% CI: 0.41-0.82) for fourth born compared with first born individuals, and 0.41 (95% CI: 0.30-0.57) for fifth or later born compared to first born. We also observed a strong inverse relationship between sibship size and lymphoid cancer, such that smaller sibships had a higher risk of lymphoma. The odds ratio was 0.58 (95% CI: 0.46-0.72) for sibships of 3, compared with sibships of 2. The odds ratios for sibships of 4 and 5 were 0.39 and 0.23, respectively.

Table 14: Odds ratios for risk of lymphoma according to birth order position and sibship size.

Variable	OR (95% CI)[1,2]
Birth order	
First born	1.00 (Referent)
Second born	0.76 (0.53 - 1.08)
Third born	0.92 (0.65 - 1.27)
Fourth born	**0.62** (0.41 - 0.82)
Fifth or later born	**0.41** (0.30 - 0.57)
Sibship size	
Two	1.00 (Referent)
Three	**0.58** (0.46 - 0.72)
Four	**0.39** (0.31 - 0.48)
Five or more	**0.23** (0.18 - 0.28)

Notes: [1] Adjusted for age at enrollment (continuous) and sex (male/female). Age at death was used for non-living participants. [2] OR and 95% CI estimated by GEE logistic regression (clustered by family) with an autoregressive correlation structure. Bold type, 95% CI does not include 1.00, denoting a significant association.
Abbreviations: OR: odds ratio; CI: confidence interval.

Table 15 shows the effects of family structure on the risk of lymphoid cancer types. Birth order was inversely associated with all lymphoid cancers (OR=0.83, 95% CI: 0.78-0.89) and smaller sibships had a higher risk of all lymphoid cancers (OR=0.82, 95% CI: 0.79-0.85). Larger sibships were significantly associated with a lower risk of lymphoma and several lymphoma subtypes. Birth order was inversely associated with risk of most major lymphoma entities (NHL, B-cell NHL, CLL, FL, MCL, MZL, and MM) but was not significant for HL, DLBCL or T-cell NHL. We observed no differences in the risk patterns associated with childhood, young-adult, or older-

adult onset HL (**Supplementary Table C.1**). Sibship size and birth order effects were similar among families with 2 cases, 3 cases or 4 or more lymphoid cancer cases.

Table 15: Associations between family structure and cancer risk by type and family size.

		Individuals within families, n (%)			**OR** (95% CI)[1,2]	
Variable	**Families**, n (%)	**Unaffected sibs**	**Lymphoid affected**	Total	**Birth order**	**Sibship size**
Entity[3]						
All types	196 (100)	1018 (100)	450 (100)	1468 (100)	**0.83** (0.78 - 0.89)	**0.82** (0.79 - 0.85)
Lymphoid neoplasm	190 (96.9)	991 (97.3)	424 (94.2)	1415 (96.4)	**0.83** (0.77 - 0.89)	**0.82** (0.80 - 0.85)
NHL	175 (89.3)	883 (86.7)	354 (78.7)	1237 (84.3)	**0.80** (0.75 - 0.87)	**0.82** (0.79 - 0.84)
B-cell NHL	162 (82.7)	753 (74.0)	307 (68.2)	1060 (72.2)	**0.80** (0.73 - 0.87)	**0.81** (0.78 - 0.84)
CLL	81 (41.3)	357 (35.1)	133 (29.6)	490 (33.4)	**0.88** (0.78 - 0.98)	**0.84** (0.80 - 0.87)
DLBCL	28 (14.3)	89 (8.7)	30 (6.7)	119 (8.1)	0.93 (0.74 - 1.17)	**0.78** (0.70 - 0.87)
FL	43 (21.9)	155 (15.2)	54 (12.0)	209 (14.2)	**0.76** (0.62 - 0.93)	**0.67** (0.61 - 0.74)
LPL	13 (6.6)	34 (3.3)	17 (3.8)	51 (3.5)	0.88 (0.59 - 1.33)	0.75 (0.46 - 1.21)
MCL	6 (3.1)	34 (3.3)	7 (1.6)	41 (2.8)	**0.51** (0.39 - 0.68)	**0.83** (0.81 - 0.86)
MZL	9 (4.6)	41 (4.0)	10 (2.2)	51 (3.5)	**0.47** (0.30 - 0.74)	**0.82** (0.77 - 0.88)
MALT	<5	25 (2.5)	5 (1.1)	30 (2.0)	**0.56** (0.34 - 0.94)	**0.81** (0.79 - 0.84)
T-cell NHL	9 (4.6)	35 (3.4)	9 (2.0)	44 (3.0)	1.17 (0.90 - 1.51)	**0.76** (0.67 - 0.85)
HL	46 (23.5)	185 (18.2)	70 (15.6)	255 (17.4)	0.93 (0.80 - 1.12)	**0.83** (0.78 - 0.90)
Classic HL	46 (23.5)	173 (17.0)	67 (14.9)	240 (16.3)	1.00 (0.86 - 1.12)	**0.80** (0.73 - 0.89)
NS	22 (11.2)	60 (5.9)	28 (6.2)	88 (6.0)	0.99 (0.76 - 1.30)	**0.71** (0.61 - 0.82)
MM	19 (9.7)	67 (6.6)	26 (5.8)	93 (6.3)	**0.70** (0.52 - 0.94)	**0.78** (0.65 - 0.95)
No. cases per family						
Two	107 (54.6)	426 (68.8)	193 (31.2)	619 (42.2)	**0.83** (0.75 - 0.93)	**0.79** (0.75 - 0.83)
Three	61 (31.1)	332 (68.6)	152 (31.4)	484 (33.0)	**0.73** (0.63 - 0.84)	**0.81** (0.76 - 0.86)
Four or more	28 (14.3)	260 (71.2)	105 (28.8)	365 (24.9)	0.94 (0.85 - 1.05)	**0.85** (0.80 - 0.91)

Notes: Cells with <5 were suppressed for privacy. Lymphomas of unknown lineage that are not otherwise specified (NOS), and entities with fewer than 5 cases and were not analyzed.
[1] Adjusted for age at enrollment (continuous) and sex (male/female). Age at death was used for non-living participants. [2] OR and 95% CI were estimated by GEE logistic regression (clustered by family) with an autoregressive correlation structure. Birth order referent group: first born; Sibship size referent group: two siblings. Bold type, 95% CI does not include 1.00, denoting significant association with risk of lymphoma. [3] Groupings are based on the InterLymph hierarchical classification of lymphoid neoplasm for epidemiologic research (404). Subtype family numbers (n) sums to greater than 196 because some families contain heterogeneous types of lymphoid cancers (e.g., NHL and CLL cases).
Abbreviations: OR: odds ratio; CI: confidence interval.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; LPL, lymphoplasmacytic lymphoma; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; MALT, mucosa-associated lymphoid tissue; HL, Hodgkin lymphoma; NS, nodular sclerosis; MM, multiple myeloma.

To estimate the probability of a chance association with birth order position or sibship size, 40,000 independent permutation tests were performed. Despite a smaller sample size and lower power of permuted data (n=95 case/control pairs), both the regression and chi-square tests for trend supported our findings, with approximately 51% and 93% of *p*-values achieving statistical significance ($p < 0.05$) for birth order and sibship size, respectively (**Supplementary Figure C.1**). Without the family dependence, the odds ratio estimates remained comparable to those from GEE models for sibship size (median OR=0.82) and birth order (median OR=0.82), validating our observations.

5.3.2 Early-life environment and immune-related diseases

Higher maternal education and an above average level of income during childhood was associated with increasing risk of all lymphoid cancers (**Table 16**). Childhood farm residents had a lower risk of lymphoma (OR=0.65, 95% CI: 0.48-0.88), but this was not statistically significant after adjusting for sibship size (data not shown). Cases were less likely than their unaffected siblings to have a post-secondary education (OR=0.62, 95% CI: 0.38-0.99) (**Table 16**), even when adjusting for family structure. There was no relationship between paternal education or childhood house location (urban vs. rural) and lymphoma or subtypes.

Allergies and tonsillectomy were independent risk factors for most major lymphoma entities (**Table 16**) and remained statistically significant after adjusting for family structure (data not shown). Lymphoid cancer risk was increased for individuals with environmental (e.g., hay fever) and drug allergies for several lymphoma entities, whereas food allergies were exclusively associated with risk of NS HL (**Table 16**). History of appendectomy was significantly associated with a 9.7-fold increase in risk of DLBCL. Asthma was not significantly associated with risk of lymphoma with the exception of MM where small sample size makes it unclear. There was no significant association between lymphoma and a personal history of collective autoimmune diseases, systemic autoimmune diseases or autoimmune diseases with no detectable autoantibodies. However, familial lymphoid cancer cases were significantly less likely than their siblings to have had organ-specific autoimmune disease (OR=0.44, 95% CI: 0.20-0.98) after adjusting for sibship size (data not shown).

Table 16: Odds ratios for risk of lymphoma and histological subtypes for childhood lifestyle variables and immune disorders in GEE regression analysis.

InterLymph class	All types	
Category 1		
Category 2		
Category 3		
Category 4 or 6		
Variable[3]	n case/sib	OR (95% CI)[1,2]
Childhood farm residence		
No	201/370	1.00 (Referent)
Yes	65/165	**0.65** (0.48-0.88)
Paternal education		
Less than high school	104/221	1.00 (Referent)
High school graduate	88/171	1.13 (0.84-1.51)
Post-secondary graduate	64/122	1.20 (0.86-1.67)
Maternal education		
Less than high school	94/218	1.00 (Referent)
High school graduate	122/208	**1.35** (1.01-1.79)
Post-secondary graduate	45/72	**1.50** (1.09-2.06)
Childhood family income		
Below average	64/147	1.00 (Referent)
Average	138/287	1.13 (0.84-1.51)
Above average	51/69	**1.75** (1.22-2.50)
Childhood residence		
Rural	106/201	1.00 (Referent)
Urban	145/299	0.97 (0.75-1.25)
Education		
Less than high school	50/28	1.00 (Referent)
High school graduate	105/83	0.79 (0.48-1.30)
Post-secondary graduate	118/110	**0.62** (0.38-0.99)
Asthma		
No	149/170	1.00 (Referent)
Yes	31/28	1.21 (0.73-2.03)
Autoimmune		
No	152/155	1.00 (Referent)
Yes	28/43	0.68 (0.39-1.17)
Organ-specific, No	169/174	1.00 (Referent)
Yes	11/24	0.49 (0.35-1.06)
Systemic, No	167/186	1.00 (Referent)
Yes	13/12	1.11 (0.48-2.57)
No detectable autoAb, No	173/189	1.00 (Referent)
Yes	7/9	0.92 (0.35-2.43)
Allergies		
No	58/99	1.00 (Referent)
Yes	108/89	**2.25** (1.44-3.51)
Drug, No	104/144	1.00 (Referent)
Yes	62/44	**2.30** (1.41-3.73)
Environment, No	94/132	1.00 (Referent)
Yes	72/56	**1.90** (1.21-2.98)
Food, No	130/160	1.00 (Referent)
Yes	36/28	1.69 (0.92-3.11)
Appendectomy		
No	130/160	1.00 (Referent)
Yes	36/27	1.53 (0.80-2.96)
Tonsillectomy		
No	81/117	1.00 (Referent)
Yes	85/70	**1.78** (1.14-2.78)

Table 16: continued

InterLymph class						
Category 1	Lymphoid neoplasms (LN)					cont'd…
Category 2			Non-Hodgkin lymphoma (NHL)			cont'd…
Category 3					B-cell NHL	cont'd…
Category 4 or 6						
Variable[3]	n case/sib	OR (95% CI)[1,2]	n case/sib	OR (95% CI)[1,2]	n case/sib	OR (95% CI)[1,2]
Childhood farm residence						
No	188/360	1.00 (Referent)	157/336	1.00 (Referent)	139/293	1.00 (Referent)
Yes	63/167	**0.70** (0.52-0.94)	58/152	0.76 (0.55-1.03)	53/146	**0.70** (0.51-0.96)
Paternal education						
Less than high school	102/223	1.00 (Referent)	93/212	1.00 (Referent)	88/202	1.00 (Referent)
High school graduate	81/166	1.09 (0.81-1.46)	70/150	1.14 (0.83-1.57)	60/125	1.19 (0.85-1.68)
Post-secondary graduate	61/118	1.19 (0.86-1.66)	46/110	1.06 (0.73-1.52)	41/99	1.07 (0.74-1.53)
Maternal education						
Less than high school	92/220	1.00 (Referent)	81/203	1.00 (Referent)	77/195	1.00 (Referent)
High school graduate	116/204	**1.38** (1.04-1.83)	102/194	**1.42** (1.04-1.96)	85/164	**1.42** (1.03-1.95)
Post-secondary graduate	41/72	**1.42** (1.03-1.96)	28/59	1.23 (0.88-1.73)	24/51	1.23 (0.88-1.73)
Childhood family income						
Below average	60/147	1.00 (Referent)	53/136	1.00 (Referent)	49/128	1.00 (Referent)
Average	131/284	1.13 (0.85-1.50)	114/263	1.16 (0.82-1.65)	100/228	1.20 (0.84-1.72)
Above average	48/70	**1.65** (1.12-2.43)	36/63	1.45 (0.94-2.28)	31/57	1.45 (0.96-2.17)
Childhood residence						
Rural	101/198	1.00 (Referent)	85/181	1.00 (Referent)	78/174	1.00 (Referent)
Urban	136/300	0.92 (0.71-1.19)	116/278	0.93 (0.71-1.22)	102/239	0.98 (0.75-1.30)
Education						
Less than high school	49/27	1.00 (Referent)	45/25	1.00 (Referent)	43/23	1.00 (Referent)
High school graduate	103/81	0.73 (0.45-1.19)	92/76	0.78 (0.46-1.32)	77/77	0.68 (0.40-1.15)
Post-secondary graduate	110/107	**0.56** (0.34-0.90)	92/102	**0.53** (0.21-0.88)	83/94	**0.50** (0.30-0.85)
Asthma						
No	143/165	1.00 (Referent)	123/153	1.00 (Referent)	112/144	1.00 (Referent)
Yes	29/30	1.07 (0.63-1.83)	26/30	1.01 (0.56-1.81)	22/26	1.08 (0.57-2.04)
Autoimmune						
No	146/154	1.00 (Referent)	123/147	1.00 (Referent)	111/134	1.00 (Referent)
Yes	26/41	0.69 (0.40-1.20)	26/36	0.87 (0.49-1.57)	23/36	0.80 (0.44-1.45)
Organ-specific, No	161/172	1.00 (Referent)	138/162	1.00 (Referent)	125/149	1.00 (Referent)
Yes	11/23	0.54 (0.25-1.18)	11/21	0.66 (0.29-1.50)	9/21	0.55 (0.23-1.30)
Systemic, No	159/184	1.00 (Referent)	136/173	1.00 (Referent)	122/160	1.00 (Referent)
Yes	13/11	1.30 (0.55-3.08)	13/10	1.38 (0.59-3.21)	12/10	1.38 (0.58-3.29)
No detectable autoAb, No	167/186	1.00 (Referent)	144/176	1.00 (Referent)	129/163	1.00 (Referent)
Yes	5/9	0.66 (0.23-1.87)	5/7	0.96 (0.29-3.29)	5/7	0.99 (0.29-3.31)
Allergies						
No	144/152	1.00 (Referent)	51/85	1.00 (Referent)	48/79	1.00 (Referent)
Yes	28/43	**2.25** (1.42-3.56)	86/87	**2.23** (1.35-3.66)	76/81	**2.06** (1.23-3.46)
Drug, No	99/140	1.00 (Referent)	124/88	1.00 (Referent)	79/114	1.00 (Referent)
Yes	60/44	**2.30** (1.40-3.79)	48/49	**1.85** (1.15-3.07)	45/46	**1.82** (1.08-3.07)
Environment, No	91/129	1.00 (Referent)	79/119	1.00 (Referent)	75/111	1.00 (Referent)
Yes	68/55	**1.83** (1.13-2.95)	58/53	**1.96** (1.15-3.35)	49/49	1.67 (0.96-2.93)
Food, No	124/156	1.00 (Referent)	107/145	1.00 (Referent)	98/134	1.00 (Referent)
Yes	35/28	1.66 (0.89-3.07)	30/27	1.83 (0.97-3.45)	26/24	1.65 (0.84-3.22)
Appendectomy						
No	127/156	1.00 (Referent)	107/144	1.00 (Referent)	98/134	1.00 (Referent)
Yes	32/27	1.41 (0.73-2.72)	30/27	1.29 (0.64-2.60)	26/25	1.24 (0.60-2.58)
Tonsillectomy						
No	77/111	1.00 (Referent)	61/100	1.00 (Referent)	53/92	1.00 (Referent)
Yes	82/72	**1.68** (1.08-2.63)	76/71	1.53 (0.97-2.41)	71/67	**1.65** (1.02-2.69)

Table 16: continued

InterLymph class						
Category 1	Lymphoid neoplasms (LN) cont'd…					
Category 2	Non-Hodgkin lymphoma (NHL)					
Category 3	B-cell NHL					
Category 4 or 6	Diffuse large B-cell		Follicular lymphoma		CLL	
Variable[3]	n case/sib	OR (95% CI)[1,2]	n case/sib	OR (95% CI)[1,2]	n case/sib	OR (95% CI)[1,2]
Childhood farm residence						
No	20/50	1.00 (Referent)	28/62	1.00 (Referent)	65/160	1.00 (Referent)
Yes	7/29	0.71 (0.43-1.17)	11/42	**0.57** (0.36-0.91)	18/42	1.02 (0.68-1.54)
Paternal education						
Less than high school	11/45	1.00 (Referent)	17/50	1.00 (Referent)	39/101	1.00 (Referent)
High school graduate	9/20	1.40 (0.63-2.09)	14/35	1.19 (0.67-2.12)	23/55	1.13 (0.72-1.77)
Post-secondary graduate	7/14	1.96 (0.80-4.81)	5/11	1.34 (0.82-2.20)	19/46	1.26 (0.74-2.13)
Maternal education						
Less than high school	10/42	1.00 (Referent)	18/48	1.00 (Referent)	33/96	1.00 (Referent)
High school graduate	14/31	1.61 (0.78-3.32)	16/42	1.02 (0.60-1.73)	36/76	**1.55** (1.07-2.26)
Post-secondary graduate	3/5	2.09 (0.44-9.96)	2/5	1.03 (0.61-1.74)	13/30	1.26 (0.76-2.07)
Childhood family income						
Below average	7/37	1.00 (Referent)	8/32	1.00 (Referent)	22/60	1.00 (Referent)
Average	9/19	**2.46** (1.20-5.05)	25/60	**1.67** (1.04-2.69)	50/118	1.35 (0.91-1.99)
Above average	11/23	**2.20** (1.05-4.89)	2/5	**1.75** (1.02-3.01)	7/16	1.17 (0.71-1.92)
Childhood residence						
Rural	12/37	1.00 (Referent)	12/46	1.00 (Referent)	36/76	1.00 (Referent)
Urban	13/38	0.76 (0.46-1.26)	22/52	1.53 (0.92-2.55)	45/121	0.81 (0.55-1.19)
Education						
Less than high school	4/11	1.00 (Referent)	13/5	1.00 (Referent)	13/16	1.00 (Referent)
High school graduate	10/11	**2.44** (1.01-5.92)	19/17	0.47 (0.15-1.44)	31/34	1.18 (0.46-3.06)
Post-secondary graduate	13/22	1.24 (0.42-3.61)	12/19	**0.30** (0.09-0.92)	40/39	1.21 (0.50-2.96)
Asthma						
No	17/33	1.00 (Referent)	26/30	1.00 (Referent)	52/67	1.00 (Referent)
Yes	3/5	1.62 (0.37-7.03)	4/5	1.01 (0.39-2.58)	13/13	1.22 (0.56-2.65)
Autoimmune						
No	14/29	1.00 (Referent)	26/26	1.00 (Referent)	56/65	1.00 (Referent)
Yes	6/9	2.00 (0.62-6.43)	4/9	0.44 (0.11-1.75)	9/15	0.82 (0.33-2.02)
Organ-specific, No	20/34	-	30/39	-	59/71	1.00 (Referent)
Yes	0/4	-	0/6	-	6/9	0.98 (0.27-3.49)
Systemic, No	17/34	1.00 (Referent)	27/32	1.00 (Referent)	61/77	1.00 (Referent)
Yes	3/4	1.89 (0.55-6.52)	3/3	1.18 (0.26-5.28)	4/3	1.44 (0.29-7.17)
No detectable autoAb, No	17/37	1.00 (Referent)	29/34	1.00 (Referent)	65/76	-
Yes	3/1	9.23 (0.72-118)	1/1	0.91 (0.05-17.8)	0/4	-
Allergies						
No	11/19	1.00 (Referent)	8/14	1.00 (Referent)	22/40	1.00 (Referent)
Yes	9/18	1.18 (0.42-3.29)	21/20	2.35 (0.74-7.41)	34/35	**2.52** (1.05-6.07)
Drug, No	14/30	1.00 (Referent)	18/22	1.00 (Referent)	37/57	1.00 (Referent)
Yes	6/7	2.26 (0.62-8.16)	11/12	1.35 (0.52-3.46)	19/18	**2.36** (1.01-5.55)
Environment, No	13/27	1.00 (Referent)	17/20	1.00 (Referent)	35/56	1.00 (Referent)
Yes	7/10	2.78 (0.78-9.96)	12/14	1.14 (0.46-2.78)	21/19	1.77 (0.71-4.41)
Food, No	18/30	1.00 (Referent)	23/29	1.00 (Referent)	42/63	1.00 (Referent)
Yes	2/7	0.57 (0.07-4.39)	6/5	2.03 (0.56-7.41)	14/12	2.39 (0.86-6.65)
Appendectomy						
No	12/33	1.00 (Referent)	23/24	1.00 (Referent)	46/64	1.00 (Referent)
Yes	8/3	**9.72** (3.34-28.3)	6/9	0.80 (0.27-2.40)	10/11	1.30 (0.39-4.37)
Tonsillectomy						
No	8/25	1.00 (Referent)	14/12	1.00 (Referent)	22/37	1.00 (Referent)
Yes	12/11	**5.17** (1.74-15.3)	15/21	0.78 (0.28-2.21)	34/38	1.38 (0.71-2.67)

Table 16: continued

InterLymph class						
Category 1	Lymphoid neoplasms (LN)					
Category 2	Hodgkin lymphoma (HL)					
Category 3			Classic HL			
Category 4 or 6			Nodular sclerosing		Myeloma	
Variable[3]	n case/sib	OR (95% CI)[1,2]	n case/sib	OR (95% CI)[1,2]	n case/sib	OR (95% CI)[1,2]
Childhood farm residence						
No	31/70	1.00 (Referent)	22/41	1.00 (Referent)	13/38	1.00 (Referent)
Yes	5/19	1.08 (0.47-2.38)	3/13	0.45 (0.15-1.31)	2/7	0.90 (0.42-1.91)
Paternal education						
Less than high school	9/31	1.00 (Referent)	5/15	1.00 (Referent)	2/7	1.00 (Referent)
High school graduate	11/20	1.38 (0.67-2.86)	6/12	1.46 (0.73-2.92)	7/16	1.53 (0.75-3.11)
Post-secondary graduate	15/31	1.10 (0.50-2.43)	12/22	1.66 (0.64-4.31)	3/14	0.41 (0.14-1.16)
Maternal education						
Less than high school	11/33	1.00 (Referent)	8/20	1.00 (Referent)	2/7	1.00 (Referent)
High school graduate	14/35	1.03 (0.56-1.91)	7/18	0.94 (0.33-2.67)	6/16	1.35 (0.62-2.95)
Post-secondary graduate	13/26	1.04 (0.45-2.42)	10/16	1.25 (0.32-4.91)	4/9	1.58 (0.53-4.70)
Childhood family income						
Below average	7/23	1.00 (Referent)	4/8	1.00 (Referent)	4/7	1.00 (Referent)
Average	17/43	1.45 (0.84-2.50)	11/30	0.84 (0.39-1.81)	7/23	0.67 (0.30-1.52)
Above average	12/23	1.74 (0.86-3.51)	10/16	1.24 (0.51-3.03)	3/9	0.65 (0.32-1.31)
Childhood residence						
Rural	16/36	1.00 (Referent)	10/21	1.00 (Referent)	5/9	1.00 (Referent)
Urban	20/53	0.82 (0.47-1.45)	15/33	1.07 (0.50-2.30)	9/30	0.66 (0.30-1.45)
Education						
Less than high school	4/4	1.00 (Referent)	4/1	1.00 (Referent)	1/7	1.00 (Referent)
High school graduate	11/18	0.57 (0.10-3.32)	6/10	0.26 (0.04-1.62)	2/16	0.16 (0.01-6.23)
Post-secondary graduate	18/28	0.76 (0.17-3.36)	13/17	0.39 (0.10-1.46)	8/9	0.80 (0.01-46.1)
Asthma						
No	20/39	1.00 (Referent)	4/14	1.00 (Referent)	6/13	1.00 (Referent)
Yes	3/3	1.58 (0.19-13.2)	13/10	1.61 (0.12-20.8)	2/1	**20.8** (1.04-417)
Autoimmune						
No	23/33	-	18/18	-	6/10	1.00 (Referent)
Yes	0/9	-	0/6	-	2/4	1.33 (0.13-13.3)
Organ-specific, No	23/39	-	18/22	-	8/12	-
Yes	0/3	-	0/2	-	0/2	-
Systemic, No	23/40	-	18/23	-	8/12	-
Yes	0/2	-	0/1	-	0/2	-
No detectable autoAb, No	23/39	-	18/21	-	6/13	1.00 (Referent)
Yes	0/3	-	0/3	-	2/1	**26.0** (2.67-253)
Allergies						
No	6/29	1.00 (Referent)	4/14	1.00 (Referent)	1/5	1.00 (Referent)
Yes	16/13	**4.93** (1.77-13.7)	13/10	**6.66** (1.36-32.5)	6/9	1.90 (0.05-68.2)
Drug, No	11/32	1.00 (Referent)	8/20	1.00 (Referent)	5/13	1.00 (Referent)
Yes	11/10	**5.93** (1.37-25.6)	9/4	4.68 (0.80-27.4)	2/1	**50.7** (5.72-449)
Environment, No	12/32	1.00 (Referent)	9/15	1.00 (Referent)	3/8	1.00 (Referent)
Yes	10/10	2.19 (0.69-6.95)	8/9	2.37 (0.53-10.5)	4/6	0.73 (0.05-10.9)
Food, No	17/38	1.00 (Referent)	12/22	1.00 (Referent)	6/10	1.00 (Referent)
Yes	5/4	2.15 (0.53-8.67)	5/2	**5.18** (1.25-21.4)	1/4	0.73 (0.03-18.9)
Appendectomy						
No	20/40	1.00 (Referent)	16/23	1.00 (Referent)	3/11	-
Yes	2/2	8.38 (0.59-118)	1/1	13.4 (0.27-673)	4/3	-
Tonsillectomy						
No	16/32	1.00 (Referent)	13/18	1.00 (Referent)	4/12	1.00 (Referent)
Yes	6/10	**4.51** (1.08-18.9)	4/6	**4.61** (1.07-19.9)	3/2	7.78 (0.47-129)

Notes: Results with fewer than 5 cases should be viewed with caution. Lymphomas of unknown lineage that are not otherwise specified (NOS), and entities with fewer than 5 cases were not analyzed. Groupings are based on the InterLymph hierarchical classification of lymphoid neoplasm for epidemiologic research (404). [1] Adjusted for age at enrollment (continuous) and sex (male/female). Age at death was used for non-living participants. [2] OR and 95% CI were estimated by GEE logistic regression (clustered by family) with an autoregressive correlation structure. Birth order referent group: first born; Sibship size referent group: two siblings. Bold type, 95% CI does not include 1.00, denoting a significant association. [3] Variables are ordered by sample size.
Abbreviations: OR: odds ratio; CI: confidence interval; No detectable autoAb, No detectable autoantibodies.
Subtype abbreviations: NHL, non-Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; HL, Hodgkin lymphoma.

5.3.3 Stepwise model selection

Three GEE models encompassed selection of family structure, early life environment and immune-related disorders. Since the availability of lifestyle and disease information varied among participants, three GEE models were built in a stepwise manner (**Table 17**). The base model contained 1468 individuals, in which birth order and sibship size were independent significant predictors of lymphoid cancer status. The middle model (n=682) retained family income during childhood in addition to birth order and sibship size as significant predictors of lymphoid cancer status. Maternal education, paternal education, childhood house location or farm residence were not significant predictors of familial lymphoid cancer risk in the middle model and were therefore excluded from the stepwise model selection. The full model (n=321) included allergies, autoimmune diseases, tonsillectomy and family structure, with maternal education included in the sibship size (but not birth order) model. Participant education, paternal education, childhood farm residence, house location, family income, asthma or a personal history of an appendectomy were not significant predictors of lymphoid cancer in the full model among multiplex lymphoid cancer families.

Table 17: Odds ratios for risk of lymphoid cancer from stepwise GEE logistic regression models.

Model	Adjusted for **birth order** **OR (95% CI)**[1,2]	Adjusted for **sibship size** **OR (95% CI)**[1,2]
1. Base model, n = 1468		
Family structure	**0.83** (0.78 - 0.89)	**0.82** (0.79 - 0.85)
2. Middle model, n = 682		
Family structure	**0.83** (0.75 - 0.92)	**0.82** (0.78 - 0.85)
Childhood family income		
Below average	1.00 (Referent)	1.00 (Referent)
Average	1.00 (0.75 - 1.32)	**0.78** (0.62 - 0.97)
Above average	1.40 (0.97 - 1.97)	0.99 (0.73 - 1.34)
3. Full model, n = 321		
Family structure	**0.85** (0.75 - 0.98)	**0.82** (0.74 - 0.89)
Allergies	**2.58** (1.59 - 4.20)	**2.46** (1.52 - 3.98)
Autoimmune	0.65 (0.35 - 1.22)	0.58 (0.31 - 1.09)
Tonsillectomy	**1.72** (1.06 - 2.81)	1.51 (0.91 - 2.56)
Maternal education		
Less than high school	(not selected)	1.00 (Referent)
High school graduate	-	**0.53** (0.31 - 0.93)
Post-secondary graduate	-	**0.47** (0.23 - 0.96)

Notes: Family structure represents birth order or sibship size variables. [1] Adjusted for age at enrollment (continuous) and sex (male/female). Age at death was used for non-living participants. [2] OR and 95% CI were estimated by GEE logistic regression (clustered by family) with an autoregressive correlation structure. Birth order referent group: first born; Sibship size referent group: two siblings; Bold type, 95% CI does not include 1.00, denoting a significant association.
Abbreviations: OR: odds ratio; CI: confidence interval.

5.4 Discussion

We assessed associations of family structure and childhood environment with disease in families with multiple lymphoid cancers. We observed an inverse relationship between birth order and cancer risk that was similar for lymphoid cancers collectively and most major subtypes (NHL, CLL, FL and MM). Sibship size was also inversely associated with risk of lymphoma and all subtypes, with the exception of LPL. Higher maternal education, above average income during childhood, allergies and tonsillectomy were independent risk factors for lymphoma, whereas higher participant education and childhood farm residence were protective of lymphoid cancer risk. To our knowledge, this is the largest multiple-case family study to date that supports the hygiene hypothesis contributing to lymphoid cancers.

Familial lymphoid cancer cases were more likely to be male, which is consistent among population studies (157,158,252,254–257,259,467), but not always true among multiple-case family studies (466,468,469). In this study, all lymphoid cancers considered as a group, and NHL (and subtypes) resembled the population sex distribution (463); however, HL and MM cases were significantly less likely to be male. Familial cases may reflect a different and potentially more genetic etiology in comparison to population cases (466). Lower rates of B-cell malignancies among women may be influenced by body size (293) and protective effects of sex hormones during pregnancy (325,470,471). A Swedish study observed a higher risk in familial NHL among same-sex siblings and parent-child pairs (472), suggesting similar environmental or behavioral factors may result from engaging in sex-typed activities during early childhood (473,474).

The observed inverse relationship between family structure and risk of lymphoma in multiple-case lymphoid cancer families is supportive of the hygiene hypothesis and the outcome of several epidemiological population-based studies (157,170,191,201,255,256,259,260,262). Birth order and sibship size are inevitably correlated and distinguishing between their effects is difficult. Generally, eldest siblings are nursed longer, receive more prenatal care and medical surveillance, and may be better nourished than later born siblings (475–479). Children from smaller sibships are typically of higher SES and have an older age at first bacterial or viral disease (210,479), whereas larger families may be subject to crowding that increases the likelihood of sharing infectious agents such as EBV, HHV-8, *H. pylori*, and *Mycoplasma pneumoniae*, all of which are associated with elevated rates of lymphomas and leukemias (184,253,259,261,480–483). Opposite risk patterns for childhood- and adult-onset HL have been documented (260); however, we observed no difference in risk among young-adult or older-adult HL, while childhood-onset cases were too few to analyze fully. Our finding that MCL, MZL and MALT lymphoma were more frequent among earlier born siblings has not been previously reported.

Indicators of infectious exposures that are correlated with childhood SES were also supportive of the hygiene hypothesis, such that individuals with a high childhood SES were at an elevated risk of lymphoid cancer. Strong indicators of childhood SES include parental education and income as they capture knowledge-related behaviours that influence age, extent and response to infectious agents (249,270). More protected or cleaner environments associated with higher SES may delay infectious exposure and increase adult-onset immune-related disease risk (484), which is consistent with population-based associations (201–203,252,485–488) and our observations. In this study, childhood farm residents had a lower risk of lymphoma which is consistent with the hygiene hypothesis and epidemiological population-based studies

(170,212,277). Early and frequent farm visits and animal contact (0-4 years of age) are thought to trigger an early immune response and strong immune competence suggested to prevent childhood lymphomas (170,212,277).

Familial cases were significantly more likely than their unaffected siblings to report a history of allergies and a tonsillectomy, which may indicate defective immune regulation (160,489,490). An elevated risk of allergies has been observed among nonfamilial cases of lymphoma (491), including NHL (160,257,492) (and mature B-cell subtypes (142,257,491)), HL (201), MM (157,491), and familial WM (103 cases) (232); however, some studies observed the opposite effect (146–148,157,253,257,493–495). Explicit correlations between lymphoid cancer subtypes and high molecular weight allergens (493), serum IgE levels (494) and type of allergy (e.g., food, environment) (146,147,157,160,257,491) complicate the elucidation of these relationships.

A positive association between lymphoid cancer and a tonsillectomy has been described among population lymphoid cancer cases (146,218,221,496), but not among multiple-case families (226,232). A tonsillectomy in younger children may indicate severe recurrent tonsillitis (146,218) caused by an altered or impaired immune response which affects lymphogenic mechanisms in adulthood (146,497). Lymphoid cancer risk may be more pronounced in tonsillectomized children because of the declining immunological function of the tonsils from early childhood to adulthood (218,221,498). Viruses such as EBV has been implicated in this role, as it is associated with recurrent bouts of tonsillitis (146,221,499–501).

In this study, a higher risk of DLBCL (but no other subtype) was associated with an appendectomy, which is consistent with some (225,226), but not all epidemiological population-based studies (158,229,232). An appendectomy/appendicitis may reflect susceptibility to infection/inflammation; however, this information was unavailable in for the participants in our study. The removal of the appendix and surrounding lymphoid tissue may alter the natural immune response to pathogenic microorganisms (225,502). With the possible exception of MM, we observed no association between familial lymphoid cancer and asthma, consistent with the literature (142,147,158,160,170,208,257,262,492). Some studies, including ours, were unable to differentiate between allergic and non-allergic asthma, which may explain some of the inconsistency among the studies (62).

There is limited and contradictory information on the associations between education and risk of lymphoma (295). In this study, familial cases were more likely to have lower educational

attainment than their unaffected siblings, which is consistent with sporadic DLBCL (295) and MM (180,268,295) cases, but not all population-based studies (201,202,295). The relationship between education and lymphoid cancer is complex and may be influenced by age of diagnosis, treatment regimens, and childhood SES.

We observed no association among autoimmune disorders and familial lymphoid cancer occurrence. Personal and family history of autoimmune conditions are strong established risk factors for lymphoid cancers (62,124,129,232,414,503–505), so this finding is unexpected. However, we were unable to examine subtype-specific associations among the biologically diverse autoimmune diseases, and a personal history of organ-specific autoimmune disease was associated with lower risk of lymphoma. Among individuals with an organ-specific autoimmune disease, unaffected siblings were on average 11 years younger than lymphoid cancer cases, suggesting that unaffected cases might harbor a predisposition to lymphoid malignancy that has not yet become manifest due to shorter duration of follow-up, and/or ascertainment bias.

Our observations add to the epidemiological support the antigen stimulation hypothesis (142,157,160,201,232,257,491,492), wherein chronic immune stimulation progressively leads to random oncogenic mutations and subsequent cancer development (160,489,490). In contrast, the immune surveillance hypothesis proposes that allergic conditions enhance the ability of the immune system to detect and eliminate malignant cells (160,506), and is also well supported (146–148,157,253,257,493–495). Explicit correlations between lymphoid cancer subtypes and type of immune-related disease (e.g., allergies, asthma, autoimmune condition, etc.) complicate the elucidation of these relationships (62,124,129,146,147,157,160,232,257,414,491,503–505). Inconsistencies among studies may be partially attributable to differences in study designs, reverse causality, gender differences, selection bias, diverse definition and measurement of allergy, hematological subtypes assessed, reliance on self-reported data/recall bias and participant characteristics (e.g., families with a genetic etiology, sporadic cases) (153,157,179,201–203,221,222,249,252,254,256,261,274,466,491,507).

Our study has several strengths, including extensive demographic, family structure and exposure data, and inclusion of unaffected family members. Participation rates of cases and unaffected siblings did not differ by SES or education. Family based studies such as ours do not suffer from response bias due to education and SES disparities among participants as in case-control study designs. Despite the rarity of familial lymphoid malignancies, this study included a relatively large number of families (196). We were able to detect effects of birth order, sibship

size, and childhood environment among familial lymphoid cancer cases while controlling for known lymphoma risk factors (age, sex, ethnicity). Limitations include use of self-reported data, which may be subject to recall/response biases (143,508). We did not have complete atopic disease data or direct markers of infectious exposure, such as number and type of infections, age at infection, or serologic data. Shorter duration of follow-up may have biased some associations because insufficient time elapsed for disease development among siblings and children. Families were not ascertained through a systematic population-based study, which may limit the generalizability of the findings to non-familial lymphoma. However, this study represents the largest, and in terms of demographic and lifestyle information, the most extensively characterized cohort of lymphoid cancer families reported to date.

5.5 Conclusion

This investigation represents the first multiple-case family study to quantify the effects of family structure according to lymphoid cancer type. This is the first study to establish an inverse relationship between family structure (birth order and sibship size) and risk of CLL and MM in the context of families with heterogeneous lymphoid cancers. The observed inverse relationship between family structure and risk of lymphoma is supportive of the hygiene hypothesis, and that childhood exposure to infectious agents may play a role in the risk of multiple types of lymphoid cancers. Our observations indicate that lifestyle factors such as SES and education also correlate with risk of lymphoma. The familial nature of these cancers implies a role of shared genetic and/or environmental factors. Such effects may be modified by lifestyle factors that correlate with birth order and family structure, and could lead to the identification of modifiable factors that protect against lymphoid cancers, even in the context of multiple-case families.

Chapter 6: Allele sharing and identity by descent analyses identify biologically plausible variants in a multiple-case lymphoid cancer family.

6.1 Introduction

The aggregation of lymphoid cancers in several large multiple-case families without mutations in known susceptibility genes suggests the presence of unidentified genomic factors that affect risk of lymphoma. Large multigenerational families are potentially informative for linkage analysis in finding high-penetrance variants. Numerous multiple-case lymphoid cancer families have been the subject of linkage searches and several putative linkage signals were observed.

To date, only a few studies have performed linkage searches in high-risk NHL families (376,377,509,510). A linkage study of 11 families with WM provided evidence for disease loci on chromosomes 1q and 4q (and possibly 3q and 6q); however, no specific gene has been implicated to date (377). Some high-risk CLL families appear to carry susceptibility loci at 2q21 (*CXCR4*) (376,509), 6p22.1 (HLA region) (376), 18q21.1 (*SMAD7)* (376), or 9q21.33 (upstream *DAPK1*) (510). *CXCR4* plays a key role in B lymphopoiesis by regulating proliferation, differentiation, and migration during an immune response. Single nucleotide variants (SNVs) in *CXCR4* have been identified among several families enriched for CLL, including a truncating mutation (W195X) (509) and two missense variants (V139I and G335S) (509) in an evolutionarily conserved domain (509).

Genome-wide SNP genotyping has been performed in several high-risk HL families (380,433,434,511,512). Initial studies found evidence for linkage on chromosome 4p, 2 and 11 (380). Several candidate pathogenic genes (e.g., *AURKA*, *VNN2*, *CYSLTR2*, *EMILIN3*, *SDR42E4*, *FAM107A*, *SLC26A6* and *ACAN*) have been identified in multiple-case HL families (434,512). Most multiplex HL families that have been studied appear to involve different genetic loci. However, few regions linked to or associated with lymphoid cancer and with mutations in more than one family have been observed, and include: the disruption of *KLHDC8B* (3p21.31) by translocation or a 5' UTR SNP which has been observed in 4 HL families (433); a truncating mutation or 3 bp deletion in *NPAT* (11q22.3) found in a family with 4 NLPHL cases and 1 family with an HL and NHL case, respectively (513); a rare (MAF = 0.0007) missense variant in *KDR* (or *VEGFR2*) associated with HL risk in 2 families (434).

With the aim of furthering our understanding of lymphoid cancer susceptibility, we describe a family in which four members were diagnosed with lymphoid cancer (2 NLPHL, 1 T-cell/histiocyte rich diffuse large B-cell lymphoma (THRLBCL), and 1 DLBCL). THRLBCL is a rare subtype of DLBCL that can morphologically resemble NLPHL. NLPHL is generally an indolent lymphoma, whereas THRLBCL is typically diagnosed in advanced clinical stages with a poor prognosis (17,514). THRLBCL and NLPHL cases share a number of diagnostic features, including cellular histology, morphology, and gene expression, suggesting that these entities represent a spectrum of a single disease (514). The aggregation of 2 NLPHL and a THRLBCL case in a sibship is suggestive of a shared genetic etiology. Seeking to identify major risk locus, we conducted a genome-wide SNP based identity by descent (IBD) analysis of the entire pedigree, and whole exome sequencing (WES) of lymphoid cancer affected family members. Variants that segregated with lymphoid cancer among all 4 affected family members or the 3 brothers were prioritized.

6.2 Methods

6.2.1 Recruitment of Family 133

We recruited a four-generation Canadian family of European origin, comprising 52 members. Written informed consent was obtained from all study subjects. A detailed family history revealed four members of the family were diagnosed with lymphoid cancers (**Figure 14**). Three out of twelve siblings were diagnosed with a lymphoid cancer, and a fourth lymphoid cancer case was in a maternal aunt. The proband, III-1, is a male diagnosed with NLPHL at the age of 19 years. A brother of the proband, III-5, was diagnosed with THRLBCL at 48 years of age, and another brother of the proband, III-8, was diagnosed with NLPHL at 33 years of age. The maternal aunt, II-6, was diagnosed with DLBCL at 60 years of age. The diagnoses were confirmed through medical records and histopathology slides reviewed by an expert oncology pathologist. None had a significant medical history or past history of cancer with one exception: II-6 (maternal aunt) had been diagnosed with breast cancer 1 year prior to a DLBCL diagnosis. NLPHL relapsed in III-1 at 44 years old.

Additional personal health information was available for 25 members of Family 133. Other immune-phenotypes include a personal history of a tonsillectomy for III-1 (affected proband), III-5 (affected brother), II-6 (affected maternal aunt), but not for III-8 (affected brother), nor 7 other

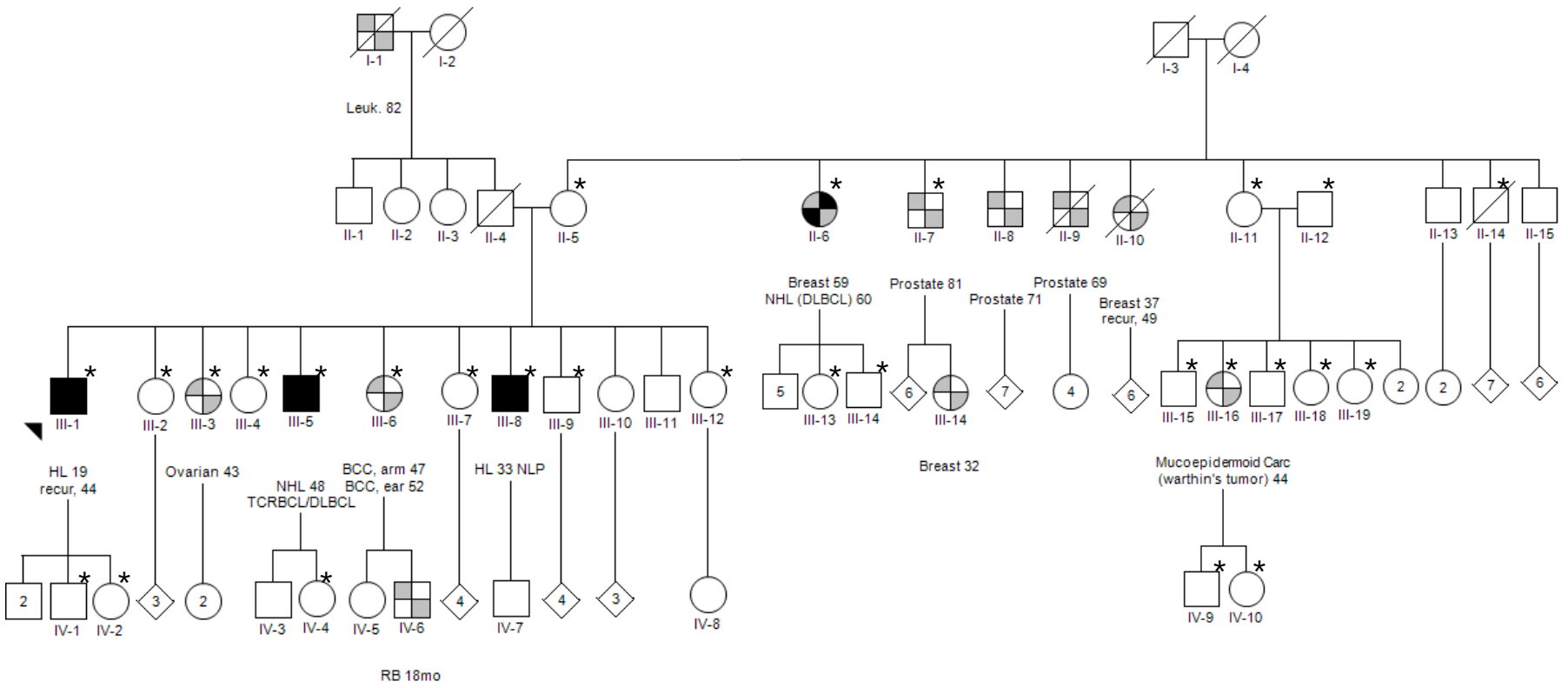


Figure 14: Pedigree of a European-ancestry family with multiple lymphoid cancers.

Notes: * genomic DNA available for genotyping. # corresponds to the age of diagnosis. Squares represent males, circles represent females. Arrow indicates proband. Black shading designates lymphoid cancer cases that were exome sequenced.

Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; NLP, nodular lymphocyte predominant; DLBCL, diffuse large B-cell lymphoma; THRLBCL, T-cell/histiocyte-rich large B-cell lymphoma; RB, retinoblastoma; Mucoepidermoid carc, mucoepidermoid carcinoma; BCC, basal cell carcinoma; Breast, breast cancer; Prostate, prostate cancer; Leuk, leukemia; recur, relapse lymphoma.

unaffected siblings in which personal medical information was available. Six additional unaffected relatives reported having a tonsillectomy. Three unaffected relatives and one presumed carrier reported having an autoimmune condition, including II-5 (mother and presumed carrier) diagnosed with RA, III-2 (unaffected female sibling) diagnosed with hypothyroidism, III-4 (unaffected female sibling), diagnosed with RA, and III-12 (unaffected female siblings), diagnosed with psoriatic arthritis.

6.2.2 Sample collection

Phenotype data and peripheral blood or saliva were collected from 30 members of Family 133, 4 of whom were affected by a lymphoid malignancy. FFPE tissue blocks were obtained for 5 lymphoid tumours (1 recurrent lymphoma) and 3 non-lymphoid cancer tumours (breast, prostate, mucoepidermoid carcinoma). gDNA was extracted from peripheral blood or saliva samples according to the manufacturer's protocol (415). Tumour DNA and RNA were extracted from FFPE tissue blocks using the QIAGEN AllPrep DNA/RNA (Germantown, Maryland, USA) and the recommended deparaffinization solution according to manufacturer's protocol (418).

6.2.3 SNP genotyping and quality control

Family members who are currently unaffected may go on to develop a lymphoid cancer and are therefore considered 'unknown' instead of unaffected. Constitutional DNA samples from 30 family members (4 affected, 26 unaffected/unknown) and 4 tumour DNA samples (1 lymphoid, 3 non-lymphoid) were subject to genotyping of 4,559,465 markers using a custom Infinium Omni5Exome-4 v1.3 BeadChip (Illumina, San Diego, California, USA) (515) at the McGill University/Genome Quebec Innovation Centre (Montreal, Quebec, Canada). Markers had a mean and median physical spacing on 0.64 kb and 0.33 kb respectively (515).

For purposes of SNP typing quality control, the data from Family 133 were analyzed in conjunction with 1162 other samples genotyped in the same batch. The additional 1162 samples are from the Healthy Aging Study which is also under the supervision of Dr. Angela Brooks-Wilson.

Illumina GenomeStudio v 2.0 software was utilized with the GenTrain 2.0 clustering algorithm. To maximize the number of informative markers, per-sample QC was performed before per-marker QC, when possible. In total, 1196 samples from two studies (34 LCFS, 1162 Healthy

Aging Study) were genotyped at 4,559,465 markers and QC measures applied. SNPs were clustered using high quality samples (call rate ≥ 99%). 140,863 SNPs were excluded from 1192 germline samples according to the following criteria: GenCall scores < 0.15 (per Illumina recommendations), GenTrain scores < 0.15, cluster separation < 0.40, call frequency < 99%, 1 or more genotype discrepancies between 25 pairs of replicate samples, and normalized intensity cluster mean (“R mean”; AA, AB and BB) ≤ 0.20. At this point, batched per-marker QC steps were completed and 30 LCFS germline samples from Family 133 were separated from 1162 Healthy Aging Study samples.

Additional per-sample QC was performed on 30 germline samples from Family 133. Samples were examined for discordance with ascertained sex to identify plating errors or sample mix-ups. Gender mismatch was checked using PLINK (516,517) and by examining the X chromosome homozygosity and Y chromosome hemizygosity rates for constitutional DNA samples. One LCFS sample was removed due to discordant sex. One LCFS sample was removed due to low marker call rate (~66%).

Additional per-marker filtering was performed on 28 LCFS germline samples. 2,608,199 (59%) markers that were duplicated, monomorphic, or had Mendelian inheritance errors in Family 133 were excluded. Approximately 1,813,047 markers in 28 samples were used in downstream analysis.

6.2.4 Exome sequencing and joint variant calling

Four individuals with lymphoid cancer were selected for exome and UTR sequencing as described in **Chapter 2.6.2: Whole exome sequencing** on page 49. Joint variant calling was performed in a cohort of 40 families (92 samples) with familial lymphoid cancers as described in **Chapter 2.6.3: Joint variant calling** on page 50.

6.2.5 Confirming pedigree relationships

6.2.5.1 Kinship coefficient

Kinship coefficients were estimated using the KING-Robust kinship coefficient algorithm (518), which measures the relatedness between two samples. Specifically, it is the probability that an allele selected randomly from individual 1 and an allele selected randomly from the same autosomal locus in individual 2 are IBD. Relationships are inferred reliably for close relatives (up

to the third-degree relationships). Relationship inferences for each first-, second- and third-degree pair-wise comparison performed as expected (data not shown).

6.2.5.2 Proportion of the genome shared IBD among relatives

The proportion of the genome shared IBD among all pairwise relationships was estimated using: 1) KING (518,519) and 2) SNP and Variation Suite (SVS), Golden Helix (520).

KING

IBD segments were estimated using KING (518,519) and plotted for all pair-wise relationships in the family. KING uses the estimated the proportion of the genome shared identical by descent on one and two chromosomes (IBD1, and IBD2, respectively) to infer relatedness. KING can infer relationships that are duplicate/MZ twin, parent-offspring, first-degree relative (siblings), second-degree relative, third-degree relative and fourth-degree relatives.

Inferred relationships were visualized by plotting the proportion of IBD1 and IBD2 segments for each pairwise relationship.

SNP and Variation Suite (Golden Helix)

An identity-by-descent estimation was calculated using SVS (Golden Helix) (520). Pairwise IBD estimates were plotted as a heat map.

6.2.5.3 Comparison of shared IBD segments among pair-wise relationships

IBD segments were estimated using KING, Refined IBD, and FastIBD (as described in **6.3.6**: **Identity-by-descent**). IBD segments from Refined IBD and FastIBD were plotted using the PhenoGram visualization tool (521). All IBD segments (IBD0, IBD1, and IBD2) from KING relationship inferences were plotted for each pair of relatives using the king_segments_plot.R script (519).

6.2.6 Identity-by-descent

IBD mapping provides an alternative to linkage analysis in the presence of allelic and locus heterogeneity by detecting clusters of individuals who share a common allele (522). IBD analysis can lead to the identification and fine mapping of the critical regions in which likely causal variants may be located (522).

Approximately 1.8 million markers in 28 samples were used for IBD segment detection using the following tools: 1) Refined IBD (523,524), 2) FastIBD (525), 3) KING (518,519), and 4) SNP and Variation Suite (SVS) (Golden Helix) (520).

IBD segment detection and quality assurance measures:

Refined IBD detects phased haplotype segments of a specified length threshold and calculates the likelihood of an IBD model (a haplotype shared IBD) vs. non-IBD model (no haplotype shared). Candidate segments above a specified threshold are reported as IBD segments (523,524). Refined IBD incorporates modeling of linkage disequilibrium (LD) and is able to make use of all high-density markers without increasing false positive rates (523,524). The probabilistic approach better accounts for haplotype phase uncertainty, relative to other IBD detection methods (e.g., FastIBD). Refined IBD does not overestimate IBD segment endpoints, but may miss some parts of an IBD segment (523). As a result, Refined IBD has a better true discovery rate (than FastIBD).

FastIBD (Beagle version 3.3.1) uses a non-probabilistic approach to estimate haplotypes using unphased data and is less accurate than Refined IBD (525). FastIBD phasing and haplotype accuracy is improved by combining the FastIBD output (candidate IBD segments) from 10 runs with random seed parameters. The recommended FastIBD threshold $< 10^{-10}$ is used to identify segments greater than 1 centimorgan (cM) in length. FastIBD tends to overestimate IBD segment endpoints but misses very little of the true underlying IBD segment. As a result, FastIBD can have a higher false discovery rate and a higher sensitivity than Refined IBD (525).

KING estimates IBD segments and can check family relationships and flag pedigree errors. Close relatives (e.g., first- and second-degree) can be identified reliably based on the estimated kinship coefficient algorithm, whereas relationship inferences for more distant relationships (e.g., third- and fourth-degree relatives) are more challenging (518). Inferred relationships are visualized by KING which can be used to confirm pedigree errors. IBD segments are rapidly and accurately inferred between all pairs of individuals. Pair-wise IBD segments of close relatives were plotted using an R script (king_segments_plot.R script (519)).

Genome-wide IBD was also measured between all pairs of samples using a probabilistic approach through Golden Helix, SVS (520).

IBD segment detection was performed according to the following workflow:

Beagle v5 & Refined IBD: Beagle v5.0 (526) was used to phase and impute missing alleles or genotypes using a reference panel of phased genotypes obtained from 1000 Genomes Project Phase 3 data release (version 5a). Markers that are not present in the reference panel were not retained. Short breaks and gaps between IBD segments were merged using merge-ibd-segments utility. Consensus IBD regions were reconstructed from 8 variations of phasing and IBD detection as described above.

Beagle v5.0 & Refined IBD: Beagle v5.0 (526) was used to phase SNP data. Refined IBD (2018 release) (523) used non-missing phased genotypes to detect IBD segments. IBD segments were merged using merge-ibd-segments utility to remove short breaks and gaps in IBD segments that have at most one discordant homozygote and that are less than 0.6 cM in length per author guidelines (Refined IBD v5 documentation) (523). Consensus IBD regions were reconstructed from 8 variations of phasing and IBD detection using the following parameters: 1) 12 (default), 30, and 50 phasing iterations, 2) 40 (default) and 60 IBD segment window size, and 3) 0.5 and 1.5 (default) IBD segment length.

FastIBD (Beagle v3.3.2): FastIBD is incorporated in Beagle v3.3.2 (527). The FastIBD algorithm was used to impute missing data, infer haplotype phase and detect genetic regions that are identical-by-descent. Haplotypes were sampled in 10 runs using random seeds as per author recommendations (527,528). Haplotypes shared by pairs of samples with a FastIBD score less than the default threshold were retained (per author recommendations) (528). A FastIBD score < 10^{-10} provides strong evidence that the shared haplotype is IBD if the length of the shared haplotype length is ≥ 1 cM. Short breaks and gaps between IBD segments were removed prior to merging IBD segments across all runs (according to author guidelines) (527).

KING v2.2.3: The KING statistical approach is designed to analyze unphased SNP data (518). IBD segments were detected for all pairwise relationships. Relationship inference and IBD detection are not impacted by the LD structure, or high-density genotype data (518).

SVS (Golden Helix): SVS estimates IBD between all pairs of samples using LD pruned data per author guidelines (520). This function is used as a quality control measure.

6.2.7 Variant extraction and filtering

Candidate IBD segments from Refined IBD, FastIBD and KING were used to identify regions that are shared IBD in 4 lymphoid cancer cases and 3 brothers. IBD segments from Refined IBD were smaller than (and confined to) IBD segments identified from FastIBD and KING. The IBD segment endpoints from Refined IBD were within the candidate regions identified by FastIBD or KING, and so only FastIBD and KING outputs were used for analysis.

All genotypes in IBD regions were extracted from WES data for filtering and prioritization.

Variants were annotated using the SVS Software (Golden Helix, Inc.) (**Table 18**). Germline variants were filtered to identify missense, nonsense, frameshift, UTR and splice-site variants. Synonymous SNVs were systematically evaluated using the regSNPs-splicing tool which prioritizes synonymous variants based on their impact of mRNA splicing and protein function (529). Uncommon non-silent variants (MAF < 5%) catalogued in Trans-Omics for Precision Medicine (TOPMed) and the Genome Aggregation Database (gnomAD) (European and global ancestry) were retained. Variants were annotated using the database for nonsynonymous SNPs' functional predictions (dbNSFP) (530,531) which provides annotations of 6 functional prediction tools (SIFT, PolyPhen2, FATHMM, MutationTaster, MutationAssessor, and FATHMM MKL); variants flagged as "benign" with 4 or more prediction tools were removed. Scaled C-scores from the combined annotation-dependent depletion (CADD) method were applied to further prioritize the variants; C-scores of 10 or lower were excluded, thus retaining all variants predicted to be in the top 10% of the most deleterious in the human genome, respectively (532). Variants that were not annotated using the preceding tools (e.g., TOPMed, gnomAD, dbNSFP, or CADD) were passed through to the next step in the filtering pipeline. After implementing filtering, Genome Browse (Golden Helix, Inc.) was used to visually confirm the potential candidate variants by rechecking raw binary alignment map (BAM) file data.

Candidate IBD segments from FastIBD and KING were also used to identify regions that segregate with disease status among the 3 brothers with NLPHL and THRLBCL subtypes. Variants were filtered using the same workflow described above and in **Table 18** with the exception of a lower MAF threshold (MAF <1%).

Table 18: Filtering pipeline for the identification of candidate variants.

Filtering criteria	Description	4 cases	3 brothers
1. Extract exome variants from IBD loci	• Extract all exome variants from IBD segments predicted by KING and FastIBD.		
2. Variant segregation	• Remove variants that do not segregate with disease (4 cases or 3 brothers).		
3. Variant QC and classification	• Retain nonsynonymous, splicing, frame shift, UTR, etc. predicted by RefSeq and Ensemble. • Retain unannotated variants.		
4. Variant frequency	• Annotate with TOPMed and gnomAD. • Retain MAF in European or global population**:** • Retain unannotated variants.	<0.05	<0.01
5. Functional predictions, dbNSFP	• Remove variants predicted as "benign" from 4 or more prediction tools. • Retain unannotated variants.		
6. Functional prediction, CADD	• Remove variants with a scaled C-score < 10. • Retain unannotated variants.		
7. Variant verification	• Remove variants with no support in BAM file or a read depth <10. • Remove variants with a European MAF**:**	<0.05	<0.01
8. Candidate variants	• Biologically plausible.		

Abbreviations: IBD, identity-by-descent; QC, quality control; UTR, untranslated region; TOPMed, Trans-Omics for Precision Medicine; gnomAD, Genome Aggregation Database; MAF, minor allele frequency; dbNSFP, database for nonsynonymous SNPs' functional predictions; CADD, combined annotation dependent depletion; BAM, binary alignment map.

6.3 Results

6.3.1 Confirming pedigree relationships

We report on a family in which 28 members were genotyped to identify candidate risk genes that may affect susceptibility to lymphoid cancer. Examining the affected status in the family's pedigree indicated that an autosomal dominant mode of inheritance was more likely.

Pairwise IBD analysis were performed for all individuals in Family 133. Calculated Kinship coefficients confirmed the reported family relationships (data not shown). The proportion of the genome shared IBD confirmed the reported family relationships using KING and SVS (Golden Helix) (data not shown). The KING inferred pairwise relationships were visualized by plotting the proportion of IBD1 and IBD2 segments (**Supplementary Figure D.1**). SVS (Golden Helix) pairwise IBD estimations were plotted as a heat map (**Supplementary Figure D.2**). IBD segment plots were manually inspected among pairwise relationships (data not shown).

6.3.2 Inferred IBD tracts

IBD segments shared by 4 cases were 342.1 Mb and 359.8 Mb as detected by KING and FastIBD, respectively (**Table 19**). Segments from FastIBD and KING were combined to yield 385 Mb of candidate IBD regions across 14 chromosomes. 14,485 SNPs were extracted from the inferred IBD segments.

Table 19: Identity-by-descent segments shared in 4 lymphoid cancer cases in Family 133.

	FastIBD			KING		
Chr #	Region start, bp	Region end, bp	Size, Mbp	Region start, bp	Region end, bp	Size, bp
1	84702301	86572817		84222374	86575279	
	104992206	112913252		105033164	113278651	
	150860055	168341409		150482255	168341409	
	204247388	240209042	63.23	204412538	239285428	63.33
2	184078964	190452586		184122339	190452586	
	213463406	214527050		220117458	227959973	
	220063783	227959973				
	231544844	243172645	26.96			14.17
3	4982356	10941732		4924153	10941732	
	151262987	175616591	30.31	152643911	175678538	29.05
4	24931019	45276356	20.34	24871338	46071039	21.20
5	82083323	96912369	14.83	82149437	97314428	15.16
7	22158627	25676893		30896440	57934758	
	30860218	67814731		61071103	67849501	
	138742775	150554584	52.28	138744679	151478809	46.55
8	128236439	146296414	18.06	128251854	146296414	18.04
9	77764836	81064897	3.30	77770756	81064897	3.29
10	90736213	115594725	24.86	90585138	115782271	25.20
14	34632233	76697490	42.07	34586025	76697490	42.11
15	37539807	53606685		37566761	53606685	
	65224676	67257792		65224676	67332467	
	69250142	88914580		69118161	88914580	
	91585401	102400037	48.58	91586237	102400037	48.76
17	11225528	13676445	2.45	11238992	13676445	2.44
18	57376365	65316557	7.94	57123916	65317012	8.19
20	63244	4683295	4.62	63244	4683295	4.62
Total			359.8			342.1

Abbreviations: Chr, chromosome; bp, base pair; Mbp, mega base pair.

IBD segments shared by the 3 affected brothers were 1063.3 Mb and 1071.5 Mb as detected by KING and FastIBD, respectively (**Table 20**). Segments from FastIBD and KING were combined to yield 1119.3 Mb of candidate IBD regions across 21 chromosomes. 69,793 SNPs were extracted from the inferred IBD segments.

Table 20: Identity-by-descent segments shared by 3 affected brothers in Family 133.

	FastIBD			KING		
Chr #	Region start, bp	Region end, bp	Size, Mbp	Region start, bp	Region end, bp	Size, Mbp
1	84552579	86575279		84222374	86575279	
	104992206	175386849		105033164	175361190	
	204081892	240209042	108.54	204412538	239285428	107.55
2	2420556	114992849		2428847	115208681	
	134864344	137616646		134590598	137744509	
	139469472	190452586		138319054	190452586	
	220063783	243172645	189.42	220117458	243172645	191.13
3	60799	10941732		60799	10941732	
	151262987	195026078	54.64	152643911	192315287	50.55
4	10823723	82666782	71.84	10952541	49091782	38.14
5	79394840	129492766	50.10	79395829	131873073	52.48
6	151865	22378919		151865	22378919	
	52144221	112972697		51859957	58770624	
			83.06	61880512	122346851	83.51
7	12509325	113719614		14436385	57934758	
	137571931	151497926		61071103	113702000	
			115.14	136085428	151504945	111.55
8	128236439	146296414	18.06	128251854	146296414	18.04
9	77764836	81064897	3.3	77770756	81064897	3.29
10	21021079	65463074		20795744	39076581	
	90590746	115594725		42427074	65749663	
			69.45	90108614	115782271	67.28
11	13997399	36897737		14005139	36897403	
	119383083	125878883	29.40	119376672	125861219	29.38
13	19121950	26538585		19121950	26576949	
	64106818	112091771	55.42	63930037	111152191	54.68
14	34632233	101664495	67.03	34586025	101670530	67.08
15	29733283	67257792		29584194	67332467	
	69250142	102400037	70.67	69118161	102400037	71.03
16	84870	12153374		84870	12138068	
	59907431	90157057	42.32	59781029	88274442	40.55
17	1389	13676445		1489668	13676445	
	69668280	81052105	20.06	70200743	81052105	23.04
18	57376365	65316557	7.94	57123916	65317012	8.19
19	53609534	59095126	5.49	54916601	59095126	4.17
20	63244	4683295	4.62	63244	4683295	4.62
21	37999782	48094803	10.10	38074253	48094803	10.02
22	16054713	48988668		19643104	21988774	
			32.93	23932181	48904896	27.32
Total			1071.47			1063.31

Abbreviations: Chr, chromosome; bp, base pair; Mbp, mega base pair.

6.3.3 Gene prioritization

Of 14,484 exome SNVs located in the 385 Mb of inferred IBD segments, the stepwise filtering strategy employed in **Figure 15** identified 32 potential candidate variants listed in **Table 21**. Of 14,453 excluded variants, half did not segregate with disease status (n=7346, 50.7%), 28.0% were common in the population (European ancestry), 18.9% were synonymous or intronic, 231 variants were predicted benign/not deleterious and 84 variants lacked adequate support (read depth, ethnicity-specific MAF, biological relevance, etc.).

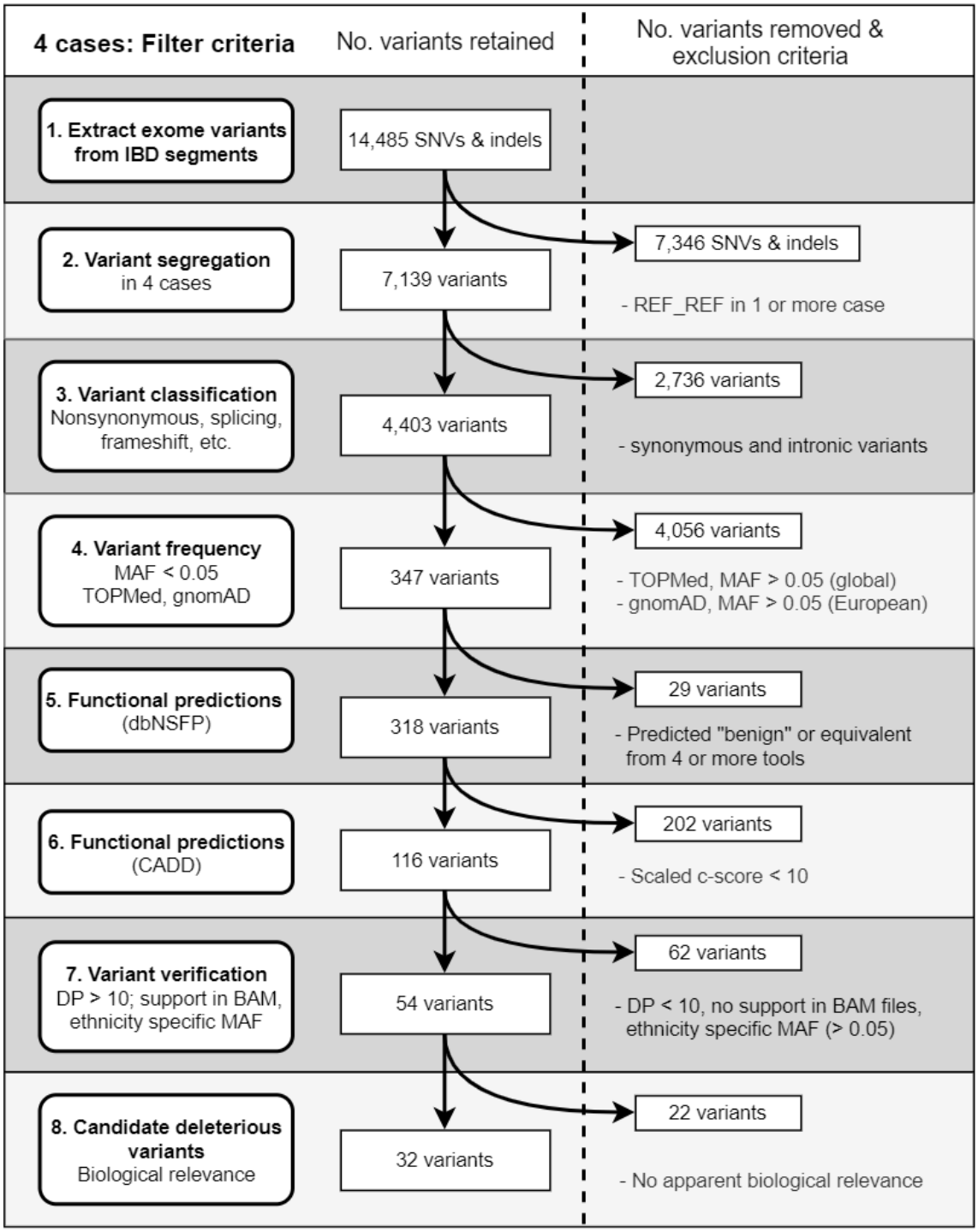


Figure 15: Filtering steps for the identification of candidate variants in IBD segments that segregate with 4 cases of lymphoid cancer in Family 133.

Abbreviations: IBD, identical by descent; TOPMed, Trans-Omics for Precision Medicine; gnomAD, Genome Aggregation Database; MAF, minor allele frequency; dbNSFP, database for nonsynonymous SNPs' functional predictions; CADD, combined annotation dependent depletion; DP, depth of coverage; BAM, binary alignment map.

Table 21: List of 32 candidate germline variants in IBD regions in 4 affected members of Family 133.

Region	Marker	Gene	Class	SNV/Indel	Ref/Alt	rsID	MAF	CADD
1q21.3	1:154318761-SNV	*ATP8B2*	Missense	p.Leu978Met	C/A	rs139958998	0.0002/0.0001	22.4
1q22	1:155257818-SNV	*HCN3*	Missense	p.Pro630Leu	C/T	rs35001694	0.0266/0.0212	23.3
1q23.2	1:160134012-SNV	*ATP1A4*	Missense	p.Thr282Met	C/T	rs144463520	0.0016/0.0011	25.1
1q24.1	1:166908811-SNV	*ILDR2*	Splice site	g.splice acceptor	G/A	rs41269698	0.0239/0.0136	15.4
1q24.2	1:167394479-DEL	*POU2F1*	UTR3	c.*9433delG	G/-	rs1354339916	0.0020/NA	13.81
1q32.2	1:210010524-SNV	*UTP25*	Missense	p.Asp344Asn	G/A	rs41274840	0.0171/0.0149	27
1q42.13	1:228506912-SNV	*OBSCN*	Missense	p.Leu4820Pro	T/C	rs188302055	0.0076/0.0062	26.6
1q42.13	1:229787055-SNV	*URB2*	Missense	p.Arg1408Gln	G/A	rs41310553	0.0149/0.0142	28.8
1q42.2	1:233802497-SNV	*KCNK1*	Missense	p.Arg171His	G/A	rs143945189	0.0087/0.0089	26.1
7p13	7:43978084-SNV	*UBE2D4*	Missense	p.Gly27Ser	G/A	rs61751727	0.0382/0.0169	22.9
7q34	7:139246917-SNV	*HIPK2*	UTR3	c.*10756G>A	G/A	rs116922249	0.0220/0.0121	12.63
7q34-35	7:143092269-SNV	*EPHA1*	Missense	p.Pro697Leu	G/A	rs34372369	0.0736/0.0590	26.9
8q24.21	8:128750540-SNV	*MYC*	Missense	p.Asn26Ser	A/G	rs4645959	0.0241/0.0252	25.2
8q24.3	8:145623265-DEL	*CPSF1*	Inframe deletion	p.Met659delATG	ATG/-	rs781863095	NA/NA	22.1
9q21.13	9:78938198-SNV	*PCSK5*	Missense	p.Cys1418Arg	T/C	Novel	NA/NA	25.1
10q23.33	10:97050708-SNV	*PDLIM1*	UTR5	c.-36G>T	G/T	rs1401258146	NA/<0.0001	11.28
10q24.1	10:99225645-SNV	*MMS19*	Missense	p.Ala558Val	G/A	rs12360068	0.0440/0.0278	23.9
10q24.2	10:101578952-SNV	*ABCC2*	Missense	p.Leu849Arg	T/G	rs17222617	0.0243/0.0108	18.86
10q24.2	10:103908969-SNV	*ABCC2*	Missense	p.Glu1592Lys	G/A	rs779672747	NA/<0.0001	24.4
10q25.1-2	10:111894999-SNV	*ADD3*	UTR3	c.*1623C>T	C/T	rs41291896	0.0237/0.0186	16.05
14q22.1	14:51370849-SNV	*ABHD12B*	Missense	p.Phe227Leu	T/C	rs7154732	0.0452/0.0595	27.6
14q22.1	14:52508948-SNV	*NID2*	Missense	p.Thr594Met	G/A	rs150406341	0.0043/0.0038	24.6
14q22.3	14:57947421-SNV	*CCDC198*	Stop gained	p.Gln182Ter	G/A	rs34960436	0.0322/0.0210	36
14q23.1	14:60433392-SNV	*LRRC9*	Missense	p.Arg433Cys	C/T	rs35533709	0.0095/0.0085	29
14q24.1	14:67939508-SNV	*TMEM229B*	UTR3	c.*629G>A	G/A	rs554492666	NA/<0.0001	15.36
14q24.3	14:74970041-SNV	*LTBP2*	Missense	p.Val1590Ala	A/G	rs139932140	0.0122/0.0060	24.1
15q14-15	15:40093446-DEL	*GPR176*	Frameshift	p.Pro433Profx114	G/-	Novel	NA/NA	26.3
15q15.1	15:42162467-SNV	*SPTBN5*	Stop gained	p.Arg1883Ter	G/A	rs61750839	0.0560/0.0283	35
15q15.2	15:43552663-SNV	*TGM5*	Missense	p.Thr42Asn	G/T	rs148913728	0.0026/0.0017	16.43
15q21.2	15:51740235-SNV	*DMXL2*	UTR3	c.*946T>C	T/C	rs76328997	0.0308/0.0258	16.84
15q25.3	15:85383070-SNV	*ALPK3*	Missense	p.Arg389Leu	G/T	rs1314564648	<0.0001/<0.0001	24.8
15q25.3	15:86087095-SNV	*AKAP13*	Missense	p.Gly191Arg	G/A	rs74502151	0.0171/0.0117	24.8

Notes: MAF corresponds to gnomAD (European ancestry)/TOPMed (European ancestry) allele frequencies.

Abbreviations: SNV, single nucleotide variant; MAF, minor allele frequency; CADD, combined annotation-dependent depletion; UTR, untranslated region.

Most (97%) of the 32 candidate variants were nonsynonymous SNPs. Seven of the variants were located in highly functional sites (splice acceptor, 5' UTR, or 3' UTR), and two SNVs were novel.

Three variants were identified as potentially deleterious and plausibly biologically relevant. The first was a heterozygous c.77A>G substitution in the second exon of the *MYC* oncogene, resulting in a p.Asn26Ser substitution. The MAF of this variant is 0.02524 from TOPMed (European ancestry), and 0.0241 from gnomAD databases (European ancestry). The mutation is predicted to be protein damaging by SIFT (score: 0.01), MutationTaster, FATHMM MKL, and possibly damaging by PolyPhen2 (score: 0.984). This variant has a high scaled CADD C-score of 25.2, which indicates that Asn26 is predicted to be in the top 1% of potentially deleterious substitutions in the human genome (532,533). *MYC* spans 6,001 bp (hg 19 assembly), has 3 exons, and 439 amino acids (AAs). The p.Asn26Ser variant is located at the N-terminus and does not affect the Leucine-zipper region or helix-loop-helix domain (protein ID: P01106).

The second interesting variant was a heterozygous missense SNV, p.Pro697Leu, in the erythropoietin-producing hepatoma receptor-A1 (*EPHA1*) gene. *EPHA1* encodes a kinase implicated in cell development, particularly in the nervous and immune systems. The variant is predicted to be protein damaging by 6 functional prediction tools (SIFT score = 0.03, PolyPhen2 score = 0.855, MutationTaster, MutationAssessor, FATHMM and FATHMM MKL), and had a scaled CADD C-score of 26.9 (top 1% of potentially deleterious substitutions) (532,533). No tool predicted the variant as tolerated. The European ancestry MAF is 0.0736 and 0.0590 from gnomAD and TOPMed, respectively; however, the global MAF is 0.0359 (TOPMed). *EPHA1* spans 18,604 bp, has 18 exons, and 976 AAs (protein ID: P21709). The variant is located in a kinase domain, 7 bp from a splice acceptor site in exon 13. This germline mutation has been observed in 14 cases of non-lymphoid cancer (534).

The third candidate was a heterozygous missense p.Ala558Val SNV in cytosolic iron-sulfur assembly component homolog (*MMS19*). The European ancestry MAF is 0.0440 and 0.0278 for gnomAD and TOPMed, respectively. The variant was predicted to be protein damaging by 4 functional prediction tools (SIFT score = 0.02, PolyPhen2, MutationTaster, and FATHMM MKL), and had a scaled CADD C-score of 23.9 (top 1% of potentially deleterious substitutions) (532,533). *MMS19* spans 40,471 bp, has 33 exons, and 1,030 AAs (protein ID: Q96T76). MMS19 is a key component of the cytosolic iron-sulfur protein assembly complex, a multiprotein complex

involved in nucleotide excision repair (NER) and transcription. It is also a part of the mitotic spindle-associated MMXD complex, which plays a role in chromosome segregation.

Two highly deleterious candidate variants in coiled-coil domain containing 198 (*CCDC198*, also known as *C14orf105*) and spectrin beta non-erythrocytic 5, (*SPTBN5*) were predicted to cause loss of function. A heterozygous nonsense mutation (p.Gln183Ter) in exon 5 of *CCDC198*, an uncharacterized protein, resulted in a premature stop codon. The allele is less common in the European and global population (MAF = 0.0322 and 0.0210 for gnomAD and TOPMed, respectively). The variant was predicted damaging by MutationTaster and FATHMM MKL and had a scaled CADD C-score of 36 (top 0.1% of potentially deleterious variants in the human genome) (532,533). No tool predicted the variant as tolerated. *CCDC198* spans 24,567 bp, has 6 exons and 9,296 AAs (protein ID: Q9NVL8).

A nonsense mutation (p.Arg1883Ter) in exon 10 of *SPTBN5* was observed as a heterozygous variant in 3 cases, and homozygous in 1 case, in Family 133. *SPTBN5* is involved in RET signaling and cytokine signaling in the immune system. The allele is less common in individuals of European ancestry (MAF = 0.0560, gnomAD and MAF = 0.0283, TOPMed). *SPTBN5* spans 45,932 bp, has 68 exons and 3,674 AA (protein ID: Q9NRC6). The variant was predicted damaging by MutationTaster and had a scaled CADD C-score of 35 (top 0.1% of potentially deleterious variants in the human genome) (532,533). The variant was predicted to be tolerated by FATHMM MKL. The variant is located in exon 31 in the Spectrin 12 repeat unit.

The remaining 28 regional candidate variants that segregated with disease status in Family 133 were in 27 genes.

The stepwise filtering employed in **Figure 16** identified 16 rare biologically plausible candidate variants that were shared among the 3 affected brothers (**Table 22**). Biologically plausible variants include nonsense variants and genes with a known involvement in cancer or immune processes. Five of the variants were located in highly functional sites (splice donor, 3' UTR) and 11 variants were nonsynonymous.

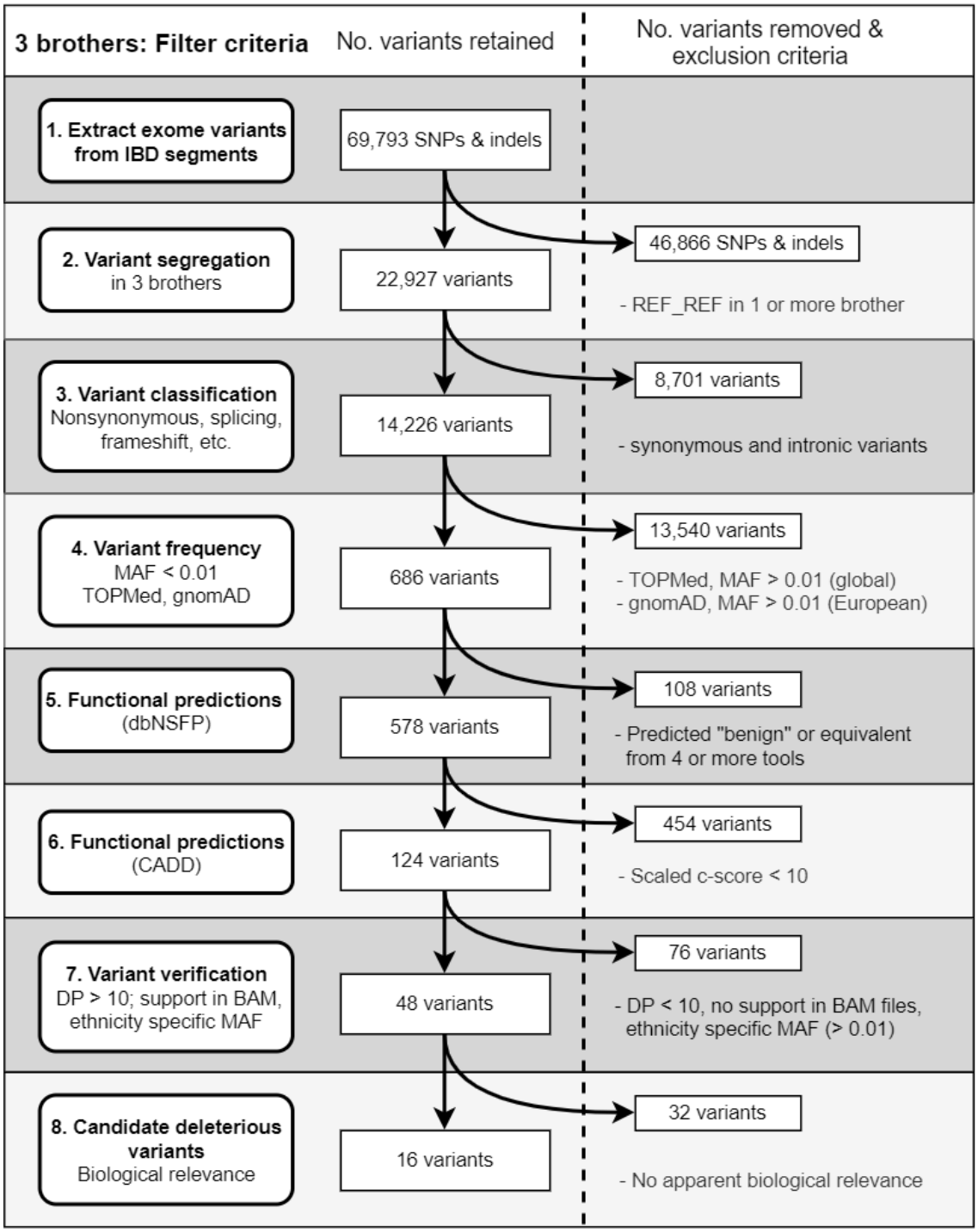


Figure 16: Filtering steps for the identification of candidate variants in IBD segments that segregate with 3 brothers with lymphoid cancer in Family 133.

Abbreviations: IBD, identical by descent; TOPMed, Trans-Omics for Precision Medicine; gnomAD, Genome Aggregation Database; MAF, minor allele frequency; dbNSFP, database for nonsynonymous SNPs' functional predictions; CADD, combined annotation dependent depletion; DP, depth of coverage; BAM, binary alignment map.

Table 22: List of 16 candidate germline variants in IBD regions shared by 3 brothers with lymphoid cancer in Family 133.

Region	Marker	Gene	Class	SNV/Indel	Ref/Alt	rsID	MAF	CADD
1q23.1	1:156770583-Del	*PRCC*	UTR3	c.331delG	G/-	rs201707266	0.0124/0.0110	
2p23.2	2:28826915-SNV	*PLB1*	Splice site	g.splice donor	G/A	rs139396774	0.0021/0.0005	28.1
2p22.3	2:33788730-Ins	*RASGRP3*	UTR3	c.*926_928insAA	-/AA	rs35972837	0.0013/0.0080	
3p26.2	3:3190253-Del	*TRNT1*	UTR3	c.415delTTC	TTC/-	rs567907193	0.0081/0.0058	
3q27.1	3:183217542-SNV	*KLHL6*	Missense	p.Thr328Lys	G/T	rs372412048	0.0002/0.0001	24.1
4p15.32	4:15739419-SNV	*BST1*	Stop gained	p.Trp284Ter	G/A	rs144539516	0.0047/0.0053	38
4p14	4:38051515-SNV	*TBC1D1*	Stop gained	p.Arg636Ter	C/T	rs1234254986	<0.0001/NA	38
4p14	4:38800282-SNV	*TLR1*	Missense	p.Ile57Met	T/C	rs145135062	0.0046/0.0030	22.3
6q16.3	6:102134167-SNV	*GRIK2*	Missense	p.Ser297Trp	C/G	rs61996330	0.0010/0.0010	32
10p11.21	10:35927592-SNV	*FZD8*	UTR3	c.2771C>T	C/T	rs188557882	0.0067/0.0048	19.87
11p15.2	11:14865399-SNV	*PDE3B*	Stop gained	p.Arg783Ter	C/T	rs150090666	0.0006/0.0007	35
15q26.1	15:90167839-Del	*TICRR*	Inframe deletion	p.Pro1435delCTC	CTC/-	rs533002408	0.0038/0.0004	
16p13.3	16:231021-SNV	*HBQ1*	Missense	p.Pro115Ala	C/G	rs144961211	0.0072/0.0071	19.27
16p13.3	16:2158680-SNV	*PKD1*	Missense	p.Arg2163Gln	C/T	rs145217118	0.0046/0.0012	33
17q25.3	17:78222401-SNV	*SLC26A11*	Stop gained	p.Gln484Ter	C/T	rs369853134	<0.0001/NA	42
22q11.22	22:22681849-Del	*IGLV1-50*	Frameshift deletion	g.305838_305839TC	TC/-	rs561565707	0.0032/0.0029	

Notes: MAF corresponds to gnomAD (European ancestry)/TOPMed (European ancestry) allele frequencies.

Abbreviations: SNV, single nucleotide variant; MAF, minor allele frequency; CADD, combined annotation-dependent depletion; UTR, untranslated region.

A two-nucleotide insertion in the 3' UTR of Ras guanyl-releasing protein 3 (*RASGRP3*) was observed as a heterozygous variant in the 3 brothers. *RASGRP3* is a guanine nucleotide exchange factor for Ras (oncogene) and Rap1 (tumour suppressor). The European ancestry MAF is 0.0013 and 0.0080 for gnomAD and TOPMed, respectively. *RASGRP3* spans 128,427 bp, has 32 exons, and 690 AAs (protein ID: Q8IV61). Functional prediction tools (dbNSFP) and CADD scores were unavailable for this variant.

The second interesting variant was a heterozygous missense p.Thr328Lys in the fourth exon of Kelch-like family member 6 (*KLHL6*). The missense variant was in one of 6 conserved Kelch repeat units. *KLHL6* is involved in B-lymphocyte antigen receptor signaling and germinal-centre B-cell formation. Polymorphisms in *KLHL6* have been observed in B-cell cancers (DLBCL, CLL and FL) and rarely in non-B-cell malignancies. The European ancestry MAF is 0.0002 and 0.00014, for gnomAD and TOPMed, respectively. The missense variant was predicted damaging by 3 tools (PolyPhen2, MutationTaster, and FATHMM MKL) and had a scaled CADD C-score of 24.1 (top 1% of potentially deleterious variants in the human genome) (532,533). Three tools (SIFT, MutationAssessor and FATTHMM) also predicted the variant as tolerated. *KLHL6* spans 68,182 bp, has 7 exons and 621 AAs (protein ID: Q8WZ60).

The third candidate was a heterozygous missense p.Ile57Met in the fourth exon of toll-like receptor 1 (*TLR1*). The missense variant is located in the sixteenth leucine-rich repeat unit (of 19). TLR1 participates in the innate immune response to microbial agents. TLRs are important for B-cell activation, maturation and memory. The European MAF is 0.0046 and 0.0030 for gnomAD and TOPMed, respectively. The variant was predicted damaging by 6 tools (SIFT, PolyPhen2, MutationTaster, MutationAssessor, FATHMM, and FATHMM MKL) and had a scaled CADD C-score of 22.3 (top 1% of deleterious variants in the human genome) (532,533). No tool predicted the functional variant as benign. *TLR1* spans 69,263 bp, has 4 exons and 786 AAs (protein ID: Q15399).

The fourth interesting candidate shared among the 3 brothers was a heterozygous nonsynonymous p.Arg783Ter that caused a premature stop codon in exon 12 of cGMP-inhibited 3',5'-cyclic phosphodiesterase B (*PDE3B*). *PDE3B* is involved in angiogenesis. Diseases associated with polymorphisms in *PDE3B* include CLL, type 2 diabetes mellitus and hypercholesterolemia. The premature stop codon was inserted in the phosphodiesterase (PDEase) domain which plays a role in signal transduction. The RAPGEF3 and PIK3R6 interaction domains were retained in the truncated protein; however, a poly-glu repeat unit was

lost. p.Arg783Ter was predicted damaging by 2 tools (MutationTaster and FATHMM) and had a scaled CADD C-score of 35 (top 0.1% of deleterious variants in the human genome (532,533). No tool considered the substitution to be benign. The European ancestry allele frequency is 0.0006 and 0.0.0007 for gnomAD and TOPMed, respectively. *PDE3B* spans 228,494 bp, has 16 exons and 1,112 AAs (protein ID: Q13370).

The fifth interesting candidate was a heterozygous nonsense p.Trp284Ter in bone marrow stromal antigen 1 (*BST1*). *BST1* facilitates pre-B-cell growth. Diseases associated with polymorphic variants of *BST1* include paroxysmal nocturnal hemoglobinuria and RA. The premature stop codon was within 10 bp of a splice acceptor site and caused the loss of a lipidation pro-peptide sequence. The variant was predicted to be in the top 0.1% of potentially deleterious variants in the human genome with a scaled CADD C-score of 38. Functional predictions using other tools (dbNSFP) were not available for this variant. The European ancestry allele frequency is 0.0047 and 0.00534 for gnomAD and TOPMed, respectively. *BST1* spans 71,241 bp, has exons 5 and 318 AAs (protein ID: Q10588).

The remaining 11 regional candidate variants that segregated with disease status in the 3 brothers were in 11 genes, and include the following:

- A 1 bp deletion in the 3'UTR of Papillary Renal Cell Carcinoma (*PRCC*). *PRCC* is involved in pre-mRNA splicing and the regulation of cell cycle progression. Chromosomal translocations that cause gene fusions are associated with papillary renal cell carcinoma.

- A 1 bp substitution at the splice donor site of intron 40 (of 57) in Membrane-associated Phospholipase B1 (*PLB1*), which facilitates the absorption of digested lipids. This variant may affect regions downstream of the splice site which includes 1 repeat unit and the membrane localization signal. *PLB1* may be a candidate gene for RA risk (535).

- A 3 bp deletion in the 3' UTR of TRNA Nucleotidyl Transferase 1 (*TRNT1*), which catalyzes the addition of CCA to the 3' terminus of tRNA molecules. There are two autosomal recessive diseases associated with *TRNT1*: retinitis pigmentosa and erythrocytic microcytosis, and sideroblastic anemia with B-cell immunodeficiency, periodic fevers and developmental delay.

- TBC1 Domain Family Member 1 (*TBC1D1*) is involved in regulating cell growth and differentiation. A premature stop codon was observed in exon 11 (of 20) within 5 bp of a splice donor site and caused the loss of the 200 AA Rab-GTPase-activating protein

domain. Diseases associated with *TBC1D1* include colloid adenoma and congenital anomalies of the kidneys and urinary tract.

- Glutamate Ionotropic Receptor Kainate Type Subunit 2 (*GRIK2*) forms a four-subunit excitatory neurotransmitter receptor in the mammalian brain. A missense variant in exon 6 (of 16) was downstream of the protein kinase domain and homopolymeric Poly-Glu stretch. Diseases associated with *GRIK2* include autosomal recessive non-syndromic intellectual disability and temporal lobe epilepsy.

- A 1 bp substitution variant in the 3' UTR of Frizzled Class Receptor 8 (*FZD8*), which is an intronless gene that encodes a transmembrane receptor for Wnt proteins. Diseases associated with *FZD8* include interstitial cystitis and exudative vitreoretinopathy.

- A TOPBP1 Interacting Checkpoint And Replication Regulator (TICRR) is a regulator of DNA replication and S/M and G2/M checkpoints. An out-of-frame 3 bp deletion resulted in the loss of 1 AA (proline) in exon 20 of 22; the variant does not affect any known domains or regions. *TICRR* is highly expressed in several solid cancers and may be involved in tumourigenesis (536).

- Hemoglobin Subunit Theta 1 (*HBQ1*) is a member of the human alpha-globin gene cluster which is expressed during early embryonic life. The missense variant occurs in the 3rd (of 3 exons) and does not disrupt the iron binding sites. Diseases associated with *HBQ1* include hemoglobin D and hemoglobin E disease.

- Polycystin 1, Transient Receptor Potential Channel Interacting (*PKD1*) encodes a member of the polycystin protein family which functions as a regulator of calcium permeable cation channels. A missense variant in exon 15 (of 46) occurred in the REJ domain which is of unknown function. Disorders associated with *PKD1* include polycystic kidney disease.

- Sodium-independent sulfate anion transporter (*SLC26A11*) is one of 26 anion exchangers in the SLC family that maintain homeostasis and intracellular electrolyte balance. A premature stop codon was inserted in exon 15 of 18 which falls on the C-terminal region of the highly conserved Sulphate Transporter and AntiSigma factor antagonist domain that possesses general NTP-binding activity. Diseases associated with *SLC26A11* include Pendred syndrome (early hearing loss).

- A 2 bp frameshift deletion was observed in the probable non-functional Immunoglobulin Lambda Variable 1-50 (*IGLV1-50*) gene. The non-functional open reading frame generally cannot participate in the synthesis of a productive immunoglobulin chair due to altered V-(D)-J or class switch recombination (537).

6.4 Discussion

Despite previous family-based linkage studies, population-based GWAS, and several candidate gene studies, a large proportion of the heritability of familial lymphoid cancers remains unexplained. We describe the first exome sequencing approach to identity lymphoid cancer predisposition genes in a multi-generational family with 4 heterogeneous lymphoid cancer cases. We identified two uncommon and one common potentially disease-causing alleles on 8q24.21 (*MYC* p.Asn26Ser), 10q24.1 (*MMS19* p.Ala558Val) and 7q34 (*EPHA1* p.Pro697Leu), and two loss of function variants on 14q22.3 (*CCDC198* p.Gln182Ter) and 15q15.1 (*SPTBN5* p.Arg1883Ter). We also identified 3 rare alleles on 2p22.3 (*RASGRP3*), 3q27.1 (*KLHL6*) and 4p14 (*TLR1*), which may affect susceptibility to NLPHL and THRLBCL subtypes observed among the 3 brothers.

MYC

MYC is a proto-oncogene that encodes a transcription factor involved in the regulation of 10-15% of all human genes (538,539). A recurrent *MYC* translocation t(8;14)(q24;q32) is found in most BL cases and 10% of DLBCL cases (50,355,539,540). *MYC* alterations causing gene amplification or overexpression have been implicated in mature B-cell lymphoid cancer susceptibility as well as in other non-lymphoid cancers (538,541–543); however, *MYC* dysregulation alone is not enough to cause lymphoma (538,543,544). The missense variant found within Family 133 lies within the N-terminus region, downstream from the active regulator elements and homopolymer stretches. An earlier study demonstrated that mutants missing the N-terminus end of Myc (AAs 1-100) are less able to induce apoptosis and growth, and less able to repress cell cycle check-points than wild type (540,541), suggesting that N-terminus amino acids are critical for function. However, the functional effects of p.Asn26Ser substitution is unknown and merits functional analysis (540,541).

The 8q24 locus is one of two genomic loci that has been associated with multiple lymphoid cancer subtypes through GWAS and candidate gene studies (41). The 128-130 Mb genomic interval at 8q24.21 harbours several independent loci with different cancer specificities, including B-cell subtypes, CLL (373), FL (360,370), DLBCL (41,50,355,370,373), and HL (41,369,370), as well as breast, prostate, colorectal, and bladder cancer (370). A GWAS of 3,857 DLBCL cases of European ancestry identified two risk alleles (rs1325592 and rs4733601) near *PVT1* and *MYC* (which displayed minimal LD) (50). Similarly, a GWAS of 589 CHL cases of European ancestry identified two independent risk alleles at 8q24.21 (rs2019960 and rs2608053) that localized to intron 6 of *PVT1* and an 82-kb region between the telomere and *PVT1*, respectively (370). *PVT1* and *MYC* are in close proximity on chromosome 8 and there is evidence that *PVT1-MYC* interactions are established regulatory networks (545,546). For example, the *PVT1* locus encodes several microRNAs that interact with *MYC* during T-lymphomagenesis and T-cell activation (370,384); and *PVT1* has two non-canonical MYC-binding sites (E-box CACGCG) in the promoter region proximal to the transcription start site (546,547). The close proximity of *PVT1* and the *MYC* oncogene (which is commonly deregulated in BL and some DLBCL cases (50,355)) and the identification of *PVT1* and *MYC* lymphoid cancer risk alleles suggests that a germline variant in the 8q24.21 region could contribute to lymphoid cancer risk in Family 133. Risk alleles at 8q24.21 have also been associated with a higher risk of breast and prostate cancers (538,543). Interestingly, the sibship of the maternal aunt (II-6) had 5 individuals who developed non-lymphoid cancer; 2 were breast and 3 were prostate cancers. Ultimately, this suggests the 8q24.21 loci may be associated with susceptibility to several cancer types, and variation in *MYC* may contribute to lymphoid cancer susceptibility in this family.

EPHA1

Another finding was the disruptive common variant on 7q34 *EPHA1*, p.Pro697Leu which segregated with the 4 lymphoid cases. Erythropoietin-producing human hepatocellular (Eph) receptors are the largest subfamily of receptor tyrosine kinases, which are further categorized into 2 subfamilies, EphA and EphB, based on their sequence homology and preferential binding to ephrin-A and ephrin-B ligands (548,549). Once activated, Eph receptors promote immune cell development, tumourigenesis, and tumour-associated angiogenesis and tumour progression (549,550). Accumulating evidence suggests that *EPHA1* expression is correlated with tumour malignancy and prognosis (548,549). Overexpression of *EPHA1* has been associated with colorectal, breast, ovarian, and prostate cancer, while the down-regulation is associated with non-

melanoma skin cancer and glioblastoma (549). Similarly, aberrant expression of other EphA and EphB proteins has been observed in leukemias and lymphomas (550,551). CRISPR/Cas9 knockdown of *EPHA1* was shown to down-regulate *MYC* expression in ovarian cancer cells (549), suggesting that *EPHA1* variants could affect penetrance of a *MYC* variant, or synergize with it, to affect cancer susceptibility.

MMS19

MMS19 has a critical role in the biogenesis of iron-sulfur proteins which function in genome stability (NER), RNA polymerase II function, and telomere length regulation (552–554). SNPs in *MMS19* have been associated with risk of developing pancreatic cancer (555) and breast cancer in non-*BRCA* families (552). Point mutations in *MMS19* can cause altered DNA sensitivity to ultraviolet and alkylating cross-linking agents, which negatively impact the ability of NER machinery to remove DNA lesions (555). Mutants in eukaryotic *MMS19* show phenotypic defects including sensitivity to genotoxic stress and extended telomere length (554). Complete *Mms19* knockout in mice caused protein instability and early embryonic death (554,556).

SPTBN5 and CCDC198

SPTBN5 plays a functional role in RET signaling and cytokine signaling in the immune system. Diseases associated with atypical *SPTBN5* include macular holes, usher syndrome and pre-eclampsia, but there is no known association with cancer (557). CCDC198 is an uncharacterized protein with a known protein interaction with AMPD2 which plays a role in energy metabolism. *CCDC198* variants have been observed in individuals with fibromyalgia syndrome (558). *CCDC198* hypomethylation and transcriptional repression has been observed in hepatocellular carcinoma (559) and clear-cell carcinoma (560). Otherwise, the functional effects of CCDC198 and SPTBN5 proteins are not well characterized.

We identified several potentially deleterious germline variants that segregate with disease status in Family 133, but no compelling high-penetrance variant. The most intriguing variant shared among 4 cases lies in the N-terminus region of the *MYC* proto-oncogene. MYC is a global regulator of fundamental cellular processes, the deregulation of which leads to tumourigenesis (50,355,538–541). Prior work has identified *MYC* (and 8q24 locus) variants associated with an increased risk of hematological malignancies (41,50,355,360,370,373,539,540), and other cancer

(370,538,541–543). However, disease development may also be influenced by additional genomic lesions which are the basis for the heterogeneity of cancers observed in this family.

The morphological resemblance of NLPHL and THRLBCL subtypes warranted the investigation of shared variants among the three brothers. We identified several potentially deleterious germline variants that are highly penetrant and biologically plausible.

RASGRP3

RASGRP3 encodes guanine nucleotide exchange factors for Ras and Rap1, both of which have a known involvement in cancer. *RASGRP3* is required for B-cell receptor signaling and development, and is important in antigen responses in mature B-cells (561). Variants in *RASGRP3* have been associated with innate and adaptive immune responses related with autoimmune conditions, including SLE and SS (562–564). Mice that lack *RasGRP1* expression develop late onset LPDs and autoimmune syndromes (564). *RASGRP* family members have been implicated in the development of B-cell lymphomas, acute myeloid leukemia, and T-cell ALL, as well as other cancers such as prostate cancer and melanoma (561). However, the functional effects of c.*926_928insAA are unknown. The 3' UTR is involved in regulatory processes such as RNA stability, mRNA translation and localization (565). Variants that alter the 3' UTR may disrupt the polyadenylation signal or binding sites for miRNA or RNA-binding proteins which influences mRNA stability, expression, translation efficiency, and localization (565).

KLHL6

KLHL6 is a lymphoid-tissue specific BTB-kelch protein that is highly expressed in germinal centre B-cells and involved in B-cell receptor signaling (566–571). Mutations in *KLHL6* have been observed in individuals with DLBCL, CLL, FL, and MM (566–568,572). The KLHL6 protein contains a BTB-domain and six highly conserved Kelch repeats (567). Most cancer-associated variants occur in the BTB-domain (567) which prevent substrate binding and catalytic activity (566,567). However, low-frequency variants in the Kelch-domain have been observed among some DLBCL, CLL and FL cases (567,568,572), and among the 3 brothers in Family 133. Variants in the Kelch-domain may partially disrupt substrate binding which may inhibit B-cell maturation and subsequent adaptive immune responses (567). During an immune response to a pathogen, *KLHL6* expression is upregulated in the germinal centre which enables mature B-cells to proliferate, differentiate and undergo somatic hypermutation (570). *Klhl6* knockout mice have an

overabundance of immature B-cells attributed to lack of B-cell differentiation (567,569). The origin of LP cells (the malignant cell of NLPHL) is of germinal B-cell origin (571), suggesting that aberrant *KLHL6* expression and LP cell formation may be correlated and contribute to NLPHL development. The diversity of B-cell lymphoid cancer subtypes associated with genetic variation in *KLHL6* (566–568,572) suggests the deregulation of *KLHL6* may be an early step in B-cell lymphomagenesis.

TLR1

Another finding was the disruptive rare variant p.Ile57Met in *TLR1*. TRLs are essential to the regulation of the innate immune system and important regulators of acquire immune system (573). Specifically, TLRs are involved in the proliferation, somatic hypermutation, induction of class switch recombination and formation of germinal centres required for the transformation of antibody-secreting plasma cells or memory B-cells (573). TLRs are structurally characterized by an extracellular leucine-rich repeat (LRR) motif which is important for ligand binding (573–575). Each LRR domain is 27-29 AA in length and each TRL typically contains 19-25 LRRs (574). The extracellular LRR domains interact with pathogen-associated exogenous factors initiate the adaptive immune response (573). Functional studies established that LLR12-17 domains of TRL1 were critical for cell activation while LRR1-12 domains had minor effects on immune response (575,576). The rare missense variant shared by the 3 affected brothers was located in LRR-16, suggesting it could have functional consequences for immune response. Several low-penetrance polymorphisms in TLRs have been associated with an increased risk of NHL, HL and B-cell subtypes (MALT lymphoma, DLCL, FL) (362,375,386,395,397,573,577,578).

BST1

BST1 (also known as *CD157*) is involved in neuronal development and maintenance as well as pre-B-cell growth (579,580). *BST1* encodes a cell surface receptor on bone marrow stroma cells where it promotes the proliferation of hematopoietic progenitor cells (579). It is highly expressed in primary epithelial ovarian cancer and malignant pleural mesothelioma (579), as well as in the CNS (579). GWA-studies have consistently associated *BST1* polymorphisms with susceptibility for Parkinson's-disease (579,581,582). The premature stop codon observed in the 3 brothers lies within the ADP-ribosyl cyclase chain and is predicted to affect enzyme function. However, additional functional studies are required to confirm the relationship between *BST1* and lymphoid cancer susceptibility.

Other potentially deleterious genes with nonsense variants lacked evidence of an association with cancer or immune-related disorder, or a biologically plausible pathway, including: *PDE3B, TBC1D1*, and *SLC26A11*. *PDE3B* regulates energy homeostasis of adipocytes, hepatocytes, hypothalamic cells and β cells (583). *PDE3B* is mainly implicated in lipolysis and may be associated with obesity and diabetes (583). *TBC1D1* regulates cell growth and differentiation and is a risk loci for familial obesity (584,585). *SLC26A11* encodes a sulfate transporter which regulates homeostasis and intracellular electrolyte balance (586). Additional rare susceptibility variants that were shared among the 3 brothers include: *TRNT1*, variations in which may cause sideroblastic anemia with B-cell immunodeficiency (587); *TICRR*, which is involved in DNA replication (536), and *IGVL1-50*, which is a non-functional immunoglobulin protein (588).

We identified several potentially deleterious germline variants that segregate with disease status in Family 133. The most intriguing variant that is shared among 4 cases lies in the N-terminus region of the *MYC* proto-oncogene. MYC is a global regulator of fundamental cellular processes, the deregulation of which leads to tumourigenesis (50,355,538–541). Prior work has identified *MYC* (and 8q24 locus) variants associated with an increased risk of hematological malignancies (41,50,355,360,370,373,539,540), and other cancer (370,538,541–543). *EPHA1* and *MMS19* are also likely candidates for lymphoid cancer susceptibility; however, we observed no compelling high to moderately penetrant susceptibility variant that was shared among 4 lymphoid cancer cases in Family 133.

One possibility is that a large multiplex family may have one or more relative that has a sporadic form the same cancer observed in the family. Phenocopies have been described among non-carriers of families with high-risk susceptibility alleles for breast cancer (589), multiple endocrine neoplasia Type 1 tumours (590) and Huntington's disease (591). The shared morphology between NLPHL and THRLBCL subtypes also supports the possibility that the 3 brothers may have different factors contributing to lymphoid susceptibility than other members of the family with lymphoma (e.g., II-6).

We observed several rare deleterious biologically plausible variants shared among the 3 brothers with NLPHL and THRLBCL subtypes. The most intriguing variant lies within the Kelch-domain of *KLHL6*. *KLHL6* is involved in B-cell receptor signaling and polymorphisms in this gene are observed almost exclusively in B-cell malignancies (566–568,572). However, disease

development may also be influenced by additional genomic lesions which are the basis for the heterogeneity of cancers observed in this family.

It is also possible that the risk variant lies within an intergenic region and therefore was not captured by WES. Nevertheless, we were able to identify few rare and uncommon variants that have a potential role in lymphomagenesis in Family 133.

Consistent with other multiple-case family studies, the most likely mode of inheritance in Family 133 is autosomal dominant with incomplete penetrance. However, the possibility of autosomal recessive or compound heterozygous variants shared among the three brothers were also explored. In the event that the maternal aunt is a phenocopy, other possible modes of inheritance include X-linked recessive or Y-linked. Lymphoid cancers are more prevalent in men than women; however, support for sex-linked genetic factors is limited. Horwitz & Wiernik (1999) found evidence that a gene conferring risk for sex-concordant HL siblings resides on the short arm of the pseudoautosomal region of the X and Y chromosome (592). To date, 6 studies comprised of high-risk WM or HL families have found no evidence of X (377,380,434,511–513) or Y (434,511,512) chromosome variants or loci in lymphoid cancer susceptibility, while other high-risk family studies did not examine the X (376,433,510–512) or Y (376,377,380,433,510–513) chromosomes.

Multifactorial inheritance

Lymphoid cancer inheritance patterns are complicated by phenotypic variability, age-related penetrance, and gender-specific cancer risk (593,594). Family-based studies have been unsuccessful in identifying high-risk or rare alleles causing Mendelian disease, which suggests that familial lymphoid cancers may not be monogenic disorders. To date, most lymphoid cancer susceptibility loci include common (MAF > 5%) or low-frequency variants, supporting a multifactorial or polygenic model for susceptibility (41). A polygenetic disease model may include a combination of low- and moderate-risk variants that contribute to a range of phenotypes. Autism Spectrum Disorder (ASD) is an example of a polygenic disease with heterogeneous phenotypes (594–599), however, monogenic forms account for 5-10% of all ASD cases (595,596,600). Some ASD cases that are polygenic may have a moderate group of low-risk variants in combination with a moderate-risk variant, or a large group of low-risk variants (596–600). Relatives of polygenic ASD cases also display autistic features, which suggests that a portion of low-frequency variants is sufficient for the occurrence of the endophenotype (594,596). The low-risk variants observed

in Family 133 (and other multiple-case lymphoid cancer families) are consistent with a polygenic disease model with variable immune-phenotypes including heterogeneous lymphoid cancers, autoimmune diseases, allergies and a personal history of a tonsillectomy. Some monogenic forms of disease, such as ASD, Parkinson's disease (601) and primary familial hypertrophic cardiomyopathy (602,603), show incomplete penetrance and variable expressivity (596,598,599), which suggests that an additional factor (genetic, epigenetic or environmental) may also be required for disease development (594,596,599).

Tumour analysis

Paired tumour-normal analysis enables the identification of somatic biallelic inactivation through loss of heterozygosity or a second mutation (or second hit) on the opposite allele, which are common cancer-initiating mechanisms in individuals with germline pathogenic variants in tumour suppressor genes (604–606). The assessment of loss of heterozygosity and second hits in tumours may be informative of the molecular mechanism of tumourigenesis and aid in the identification of germline cancer predisposition genes (604–606).

In family 133, 2 lymphoid tumours (II-6, DLBCL and III-5, THRLBCL) and 2 non-lymphoid tumours (III-3, ovarian and II-7, prostate cancer) were genotyped at ~4.6 million markers. Two additional NLPHL tumours (III-1 and III-8) were not genotyped due to limited malignant cellularity (<5%). Tumour information is limited to SNP array data only (not WES), and is further complicated by aneuploidy, non-aberrant cell admixture, and intratumoural heterogeneity (607); however, tools that mitigate these effects (e.g., PennCNV-tumor, GIANT, SOMATICs) may improve performance in identifying aberrations from tumour SNP array data (608).

Tumour samples from III-5 (THRBCL) and II-5 (DLBCL) are likely to provide insights into the genetic landscape of lymphoid cancers in Family 133. Although tumour samples from III-3 and II-7 are of non-lymphoid origin (ovarian and prostate cancer, respectively), they may have similar cancer-initiating mechanisms as lymphoid cancer cases within Family 133.

Family studies

Although next generation sequencing (NGS) represents a powerful approach to decipher the genetic predisposition to hereditary diseases, it comes with several challenges (552). In many cases, lists of genes with shared potentially deleterious variants from familial lymphoma WES

studies are different, which may be attributable to genetic heterogeneity and use of different bioinformatics pipelines and tools used to analyze the datasets. In addition, some of the filters used to prioritize variants (e.g., in silico prediction tools) may mis-classify SNPs causing erroneous inclusion or exclusion of some variants. In this study, variant prioritization followed similar procedures and thresholds as recent familial WES projects, with the exception of a higher MAF threshold (0.05, as opposed to 0.01) to retain uncommon SNPs in this analysis. When possible, multiple tools were used for phasing, IBD segment detection, and variant functional prediction. WES is a well-recognized strategy to identify rare disease-causing variants; however, we were unable to identify non-exonic abnormalities beyond the 5' and 3' UTR, such as regulatory variants. We cannot exclude the possibility that susceptibility to lymphoma in this family may be mediated through non-exonic genetic variation. Furthermore, the FastIBD algorithm has a tendency to overestimate IBD endpoints, which increases the rate of false discovery. Given the extensive IBD detection runs applied in this family, false positives cannot be ruled out.

Most genome-wide scans of multiple-case lymphoid cancer families study one disease entity (e.g., all CLL cases, or all HL cases). GWA-studies have identified two susceptibility loci that are shared among diverse lymphoma subtypes (e.g., 6p21 and 8q24.21); however, a majority of risk variants and genes do not overlap (41). To date, most linkage studies have identified low-frequency variants with small-to-medium effects (41). We were able to identify germline variants in Family 133 which may affect susceptibility to diverse lymphoid subtypes. *MYC* is an intriguing candidate to harbour germline variants predisposing to lymphoid cancers because translocations and somatic mutations are frequently detected in lymphoid cancers (50,355,538–540); however, the familial nature of these cancers is suggestive of a highly-penetrant variant which was not observed among all lymphoid cancer cases in Family 133. A rare biologically plausible variant was observed in *KLHL6*; however, p.Thr328Lys (*KLHL6*) was only shared among the 3 brothers and not the maternal aunt (II-6), suggesting DLBCL may of a different combination of etiological factors (or sporadic).

This is the first study in which candidate germline variants have been identified in a multiple-case family with several different types of lymphoid cancer (NLPHL, THRLBCL and DLBCL). We identified several rare and uncommon plausible variants in biologically relevant genes for further investigation. However, we cannot rule out the possibility that results were observed by chance.

In the future, we plan to perform linkage analyses to attempt to exclude some of these variants of interest. Variants of interest will be also be examined in 87 familial lymphoma cases in the LCFS cohort in which WES data are available.

6.5 Conclusion

The diversity of lymphoid cancer subtypes complicates the identification of genetic factors that predispose to familial lymphoid cancers. Genome-wide scans and exome sequencing studies have identified risk alleles with small effects, suggesting that, at the population level, lymphoid cancer susceptibility may be polygenic. With the exception of 6p21-HLA and 8q24.21, familial susceptibility factors appear to be associated with distinct histological subtypes rather than a combination of subtypes (466). Further studies are required to validate possible susceptibility factors.

Chapter 7: Discussion, conclusion and significance.

7.1 Summary

The body of work presented in this dissertation highlights the strengths and challenges related to identifying susceptibility factors in familial lymphoid cancers. This dissertation provides a basic framework for methodological and biological considerations in the design, analysis and interpretation of family-based studies.

Chapter 3 showed that familial lymphoid cancer co-occurrence patterns are different from the expected population patterns, suggesting that some combinations may have a shared genetic basis. Specifically, families enriched for HL cases or CLL cases may have a stronger underlying genetic basis than other combinations of lymphoma. These observations support the application of genomic methods to identify gene variants that affect lymphoid cancer susceptibility in the familial context.

Age of onset (Chapter 4) in familial lymphoid cancer cases is substantially earlier than comparable population data, even after controlling for 3 types of ascertainment bias. The familial age of onset was earlier in later generations of families, a phenomenon known as anticipation. Apparent anticipation may be caused by ascertainment biases, and I used three approaches to mitigate this. Multigenerational families that display earlier age of onset across generations are candidates for the application of genomic methods to identify susceptibility factors.

There are several well-characterized risk factors for lymphoma, including factors that affect the immune system, male sex, and a first-degree relative with an LPD. However, population-based association studies have yielded conflicting results for other potential risk factors, such as lifestyle factors, childhood SES, medical procedures, and family structure (birth order and sibship size). Currently, only 2 studies have examined family structure and immune-related characteristics in the familial context (232,462). Jønsson *et al* (2007) observed a paternal parent-offspring birth order effect with predominance of LPD in the youngest siblings among 24 pairs in 32 families enriched for CLL and B-cell malignancies (462). Royer *et al* (2010) found that familial WM cases were more likely to have immune-related disorders (autoimmune diseases, allergies, and some infections) among 103 familial WM and related B-cell disorders (232).

In Chapter 5, I conducted the largest multiple-case family-based study to quantify the effects of early life variables and immune-related diseases on the risk of distinct histological subtypes. We report on 450 lymphoid cancer cases and 1018 unaffected siblings in 196 families with multiple-cases of lymphoid cancer. The risk of lymphoma tended to decrease with later birth order and larger sibship sizes. Childhood SES variables, such as high maternal education and family income, were associated with an elevated risk of lymphoma in the familial setting. Factors that affect the immune system, such as allergies and tonsillectomy were also independent risk factors for several lymphoid cancer subtypes. Some of these are also established risk factors for sporadic lymphoid cancer although, the association occurs in the opposite direction in some instances. This chapter represents the largest multiple-case family-based study to evaluate early life environment and immune-related diseases in the familial context that has been reported. Furthermore, this is the first family-based study to examine etiological factors among several lymphoid cancer subtypes.

In Chapter 6, I conducted a genome-wide IBD analysis and WES to identify susceptibility factors in a family with 4 lymphoid cancer cases. Our study identified uncommon (MAF < 0.05) and rare (MAF < 0.01) biologically plausible variants that segregated with lymphoid cancers in 4 family members or 3 brothers. We identified three deleterious biologically relevant variants of low-penetrance and two loss of function variants in relatively uncharacterized genes that were shared among 4 lymphoid cancer cases. Notably, all 32 candidate variants had high scaled CADD C-scores, which suggests these variants are in the top 1% or 0.1% of potentially deleterious substitutions in the human genome (532).

Few candidate genes belonged to a known pathway involved in cancer etiology. One interesting candidate variant was an uncommon (MAF < 0.05, European ancestry) heterozygous c.77A>G substitution in the second exon of the *MYC* oncogene resulting in a p.Asn26Ser substitution. Other variants in the 8q24.21 loci (near *MYC* and *PVT1*) have been demonstrated to affect *MYC* regulation and are associated with a higher risk lymphoid (CLL, FL, DLBCL, HL) and other cancers (breast, prostate), which are observed within this family. *MYC* is an intriguing candidate to harbour germline variants predisposing to lymphoid cancers because translocations and somatic mutations are frequently detected in lymphoid cancers (50,355,538–540); however, we would expect the susceptibility allele to be highly penetrant.

Among the 3 brothers, we observed 5 deleterious biologically relevant rare variants which could be highly penetrant. Three of these variants (*KLHL6, RASGRP3* and *TLR1*) have

established associations with lymphoid cancer susceptibility (386,395,397,561,566–568,572). TLR susceptibility alleles have been implicated in the development of MALT lymphoma, DLCL, FL and HL (386,395,397). *RASGRP3* has been implicated in the development of B-cell lymphomas, acute myeloid leukemia and T-cell acute lymphoblast lymphomas (561). High- and low-penetrant *KLHL6* variants have been observed in B-cell malignancies (566–568,572). These genes are intriguing candidates to harbour germline variants which predispose to lymphoid cancers.

7.2 Strengths and limitations

7.2.1 Family ascertainment

Families were not ascertained by means of a systematic population-based study; families were collected largely through oncologist referrals. We cannot estimate a population size to use as a denominator to calculate the incidence of lymphoid cancers in families compared to the population as a whole. For this reason, we tested whether specific properties of familial lymphoid cancers differed from those of sporadic cases in regards to co-occurrence patterns, age of onset and sex distribution. The value of these associations may be limited by the use of SEER (USA) population-data as comparable Canadian data were unavailable. Furthermore, the observations in this project may be limited to multiple-case families (and not sporadic cases).

7.2.2 Controlling for known risk factors (percentiles) and ascertainment bias

Strong established lymphoid cancer risk factors include family history, compromised immune function, older age, male sex, and Caucasian ethnicity for most histological subtypes. For this reason, it is important to consider the rarity of each lymphoid cancer occurrence to best determine if cases in a family are earlier onset or of rarer subtypes, which may be suggestive of underlying genetic factors. For example, a DLBCL and NS HL case both diagnosed at 30 years of age would be prioritized differently. Considering that the median age of DLBCL diagnosis is 65 and the median age of NS HL diagnosis is 28 years, we can recognize that a young age of NS HL onset is often seen in sporadic cases, while a young age of DLBCL onset is rare. Similarly, the effects of sex, ethnicity and subtype can be weighted within families. Controlling for these factors (using population-based percentiles) allows for more uniform comparison of

heterogeneous lymphoid cancer data. However, the value of these adjustments may be limited by use of SEER (USA) population-based data, as comparable Canadian population data were unavailable.

The diversity of age of onset distributions for histological subtypes has been challenging to address among families with heterogeneous multiple-case lymphoid cancer families. As a result, most publications examine age of onset patterns among families with one histological subtype. This limits our understanding of disease patterns to homogeneous lymphoid cancer families. In this project, we were able to control for several known factors that affect disease onset, allowing for the uniform comparison of heterogeneous families.

Furthermore, we were able to evaluate the anticipation phenomenon in lymphoid cancer families by accounting for ascertainment bias. Although anticipation has been suggested for numerous familial cancers (including lymphoma), its occurrence has not been confidently established because of concerns about ascertainment bias and available statistical methods. In this study, we were able to adjust for ascertainment bias and several known risk factors (age, ethnicity, sex, subtype) to uniformly examine heterogeneous lymphoma families for evidence of anticipation. However, our anticipation analysis was restricted to subtypes with a larger sample size (e.g., NHL, B-cell NHL, FL, CLL, HL, CHL, and MM; but not DLBCL, LPL/WM or NS HL).

We also examined lymphoid cancer co-occurrence patterns in multiple-case families to determine which combinations were enriched in families and may therefore have a different combination of etiological factors contributing to lymphoid cancer susceptibility. Co-occurrence patterns were examined using weighted SEER (USA) population data and Canadian population data. Ethnicity- and year-specific incidence rates were unavailable for Canadian population data; however, the expected population rates were similar when using USA or Canadian population data, suggesting lymphoid cancer susceptibility factors may be comparable between both high-income countries.

7.2.3 Family-based risk factors

Most currently established lymphoid cancer risk factors stem from population-based studies of sporadic (non-familial) cases. To date, one published study examined clinical and environmental factors in a cohort of 84 multiple-case WM and related B-cell disorder families (103 cases) (232). My thesis represents the largest well-characterized cohort of multiple-case lymphoid

cancer families reported to date. We quantified the effects of childhood environment and SES and of immune-related phenotypes and medical conditions (Chapter 5) that are commonly associated with lymphoid cancer risk among sporadic (non-familial) cases. Our data support the importance of early life environment in the susceptibility of lymphoid cancers among multiple-case families, a group who may also have a greater underlying genetic susceptibility than most sporadic cases. Moreover, the lymphoid cancer families we studied are heterogeneous, allowing us to evaluate risk of several distinct histological entities.

The assessment of sporadic lymphoid cancer risk factors has yielded conflicting results. Our cohort provides support for some established risk factors in the familial setting and also identifies new relationships. For example, we found that MCL, MZL and MALT lymphomas were more frequent among earlier born siblings, whereas reported population-based studies have lacked sufficient sample size to analyze these entities. We established that certain environmental factors affect risk of lymphoma in the context of multiple-case families, who may also have underlying genetic susceptibility factor.

Lymphoid cancers are multifactorial diseases that may be partially explained by genetics and lifestyle or environmental factors. Twins and multiplex families with an underlying genetic predisposition are especially valuable to identify environmental influences that may increase or decrease the risk of disease (609). Individuals that develop a lymphoid cancer are said to have surpassed the “liability threshold” and therefore show the disease phenotype. The liability required to exceed the threshold level is the same in all individuals; however, individuals with affected relatives (especially first-degree relatives) will have a higher chance of exceeding the threshold level and developing lymphoma due to shared genetic and environmental factors. Generally, the later in life a multifactorial disease develops, the more dependent it is on environmental factors (and the lower the heritability); this is supported by the observed earlier age of lymphoma onset among multiple-case families. Gene-environment interactions may also play an important role in familial lymphoid cancers. For example, 5 SNPs (rs1800893, rs4251961, rs1800630, rs13306698, rs1799931) and tobacco smoking were significantly associated with the risk of NHL (610). Sporadic cases may require a greater combination of environmental risk factors to develop lymphoma, whereas multiplex families may have an underlying predisposition and may therefore be more susceptible to environmental factors and thus demonstrate different patterns of risk.

7.2.4 Phenotypic heterogeneity

Lymphoid cancer subtypes have distinct phenotypic characteristics and clinical symptoms. Medical records, pathology reports, and histopathology tumours/slides were used to confirm the self-reported lymphoid cancer diagnoses in this study. All histopathology tumours and slides were reviewed by an expert oncology pathologist. All diagnostic cases were reviewed and questionable or uncertain diagnoses (e.g., "lymphocytic leukemia") were removed from the analysis. When possible, statistical analyses were restricted to distinct histological subtypes to aid interpretability. This cohort of families contained well-defined histological subtypes and other phenotypic traits.

7.2.5 Data quality and statistical methods

Family-based study designs are unique in that they use relatives to assess the genetic and epidemiology of disease. Families with homogeneous diseases can be used to evaluate the potential genetic basis of a condition of interest by examining phenotypic patterns and disease co-occurrence, without the required collection of DNA. Historically, family-based studies have been the primary approach for detecting disease-causing genes through segregation and linkage methods. Relative to case-control studies, family-based studies may not suffer from response bias, SES disparities or population stratification. In this study, we collected family history information from multiple family members, whenever possible, to reduce the possibility of response bias. Smaller families tend to have more homogeneous environmental exposures that may be associated with disease etiology. As a result, family studies are a natural control for genetic background and environmental factors that may be difficult to measure or control in other study designs. Another potential benefit is quality assurance measures because the same data (e.g., phenotype) are collected from multiple family members. The main disadvantage of family-based studies is the difficulty in accumulating numerous (and large) families that are well characterized. Since family recruitment began 14 years ago (in 2006) for the data set used in this study, only 218 multiplex lymphoid cancer families have been identified.

Many statistical tests have an underlying assumption of independent samples. Using such tests on data from family studies that violate this assumption can inflate type I error, producing false inferences. In this study, lymphoid cancer family members are not independent of each other, and so permutation tests were used to generate a reference distribution on which to compare the observed *p*-value or test-statistic. In Chapter 5 in which we examined early lifestyle variables and the associated risk of lymphoma, logistic regression with a generalized estimating

equation was used to accommodate correlated family data. The odds ratios and confidence intervals were clustered by family, which caused wider confidence intervals; nonetheless, several tests achieved statistical significance. Other statistical methods, such as the bootstrap procedure can be used to create an independent subsample of unrelated individuals (1 per family), but this greatly reduces power. Statistical methods employed in this thesis may also be applied in other multiple-case family studies.

7.2.5.1 Statistical methods

This dissertation uses diverse methods and perspectives to examine determinants of lymphoid cancer risk. The following section describes advantages and disadvantages of statistical methods and alternative methodologies that can be used for familial studies.

Person-time analysis

Person-time analysis estimates the time-at-risk that all participants contribute to a study. It permits study enrollment at different time periods and accounts for participants that leave the study, are lost to follow-up or die during the study period. A disadvantage of a person-time estimate is the assumption that the probability of disease development during the study period (and over time) is constant, such that 10 persons followed for one year is equivalent to one person followed for 10 years. For this reason, the interpretability of chronic or late-onset diseases (such as cancer) may be limited. In Chapter 4 (age of onset), a person-time analysis (**Supplementary Table B.1**) revealed a reduction in person-years per lymphoid-event between generations 1 through 3, while no trend was observed for non-lymphoid cancer cases. A reduction in person-years per lymphoid-cancer event supported the observed anticipation among multiple-case families. In Chapter 5 (hygiene hypothesis), a person-time analysis could be used to measure the person-years per event for allergies, asthma, autoimmune disorders, tonsillectomy, or appendectomy within multiple-case lymphoid cancer families; however, it would have a smaller sample size than the methods used, as not all participants provided age of diagnosis for immune-related diseases (e.g., allergies, autoimmune, asthma).

Survival analysis

Survival analysis estimates the expected duration of time until an event happens (611). A survival analysis can incorporate multiple covariates and use censored or incomplete data (e.g.,

no observed failure) (611). Nonparametric methods such as the Kaplan-Meier estimate and the log-rank test are very flexible; however, it is difficult to incorporate covariates within these models, and is therefore challenging to describe how subpopulations differ in survival functions. In contrast, the Cox Proportional Hazards model is most commonly used to assess the effect of factors (e.g., treatment) while controlling for the effects of other covariates (611); it is a semiparametric technique that makes no assumption about the distribution of survival time (611). The hazards ratio can be interpreted in terms of the exponentiated logistic regression coefficient (611) and may be applied to familial data (612). A Cox Proportional Hazards model could be used in Chapter 5 to measure the ratio of the hazard rates (or failure rate) for lymphoid cancer and other disorders such as allergies, autoimmune diseases and asthma. A survival analysis may also control for censored data and truncation bias in Chapter 4 (age of onset).

Logistic regression

A logistic regression models the probabilities for discrete outcomes and can control the effects of multiple covariates (e.g., age, sex, ethnicity) as demonstrated in Chapter 5. Logistic regression measures the relevance of a predictor (coefficient size) and also provides the direction of association (positive or negative). Use of longitudinal or family-based data are problematic for logistic regression because the data are highly correlated, and so a GEE extension may be used (613–615). GEE enables the production of reasonably accurate standard errors by estimating the within-cluster similarity of the residuals (613–615). However, the GEE approach does not contain explicit terms for between-cluster variation (613).

A logistic regression may struggle with expressiveness (e.g., interactions terms), which require manual addition, whereas other models may have better predictive performance. In Chapter 5, interaction terms between early lifestyle variables, surgical procedures and immune-related diseases were explored, however, no significant interaction terms were identified (with the exception of family structure variables). A main disadvantage of logistic regression is the assumption of linearity between the dependent variable and the independent variables, which is rarely linearly separable in real world data. A logistic regression analysis may not be appropriate when the research question involves the length of time until the end point occurs (611).

7.2.5.2 Multiple comparisons corrections

In statistics, a multiple testing problem may occur in a set of simultaneous statistical inferences. A large number of statistical tests will have some *p*-values less than α (usually 0.05) purely by chance, even if the null hypothesis is true. Several statistical techniques have been developed to mitigate multiple testing problems (616,617). These techniques require the adjustment of α (also known as familywise error rate or FWER) or the false discovery rate (FDR) (616).

Adjusting the α controls the overall probability of making at least one false discovery (or type 1 error) (616). The Bonferroni correction method is one of the most commonly used approaches for multiple comparisons. It is very conservative, and with many tests, the adjusted α will become very small which reduces power and decreases the possibility of making any true discoveries (616). The Bonferroni correction is alternatively called Bonferroni inequality, Boole's inequality, or Dunn's approximation (618). Another method, the Sidak correction, assumes the tests are independent, and it is less conservative but (slightly) more powerful than the Bonferroni correction (616–618). A benefit of Sidak's equation is that it may be used for categorical and ordinal data in addition to continuous data (618).

Where FWER controls for the probability of making a type 1 error *at all*, FDR procedures allow for type 1 error (false positives) but control for the proportion of these false positives in relation to true positives (616). This is done by adjusting the decision made for the *p*-value associated with each individual test to decide rejection or not (616). Although this will result in a higher type 1 error rate, it has higher power and therefore affords a greater probability of true discoveries (616). Holm's Step-Down procedure is calculated after conducting all hypothesis tests within a family of statistical tests (616). Holm's procedure is more powerful than Bonferroni's inequality (618). Holm's Step-Down procedure makes no distributional assumptions or logical assumptions about the hierarchy of the hypothesis to be tested, and does not assume independence of comparisons (618). Hochberg's Step-Up procedure controls the FDR in a similar (but more complex) process as Holm's Step-Down procedure, but it is more powerful than Holm's Step-Down procedure (616,618).

The power of multiple comparison procedures may be affected by several factors, including design and statistical assumptions. The level to which alpha is set affects a procedure's power to detect significant associations. Smaller values of alpha make it more difficult for a multiple comparison procedure to reject the null hypothesis. Multiple comparison procedures

assume equal group sizes, and therefore the presence of unequal sample sizes may affect the procedure's ability to maintain the specified Type 1 error rate. Several correction procedures have statistical assumptions (e.g., independence of observations, homogeneity of variance, normality), the violation of such may affect the accuracy of the *p*-values, which in turn affects the Type 1 error control and power (618).

Multiple test corrections are challenging to apply in Chapter 5 due to the hierarchical classification of types and subtypes of lymphoid cancers. For example, in Chapter 5, FL and DLBCL cases would also be grouped in B-cell, NHL and lymphoma subgroups.

Chapter 3 (co-occurrence), Chapter 4 (age of onset) and Chapter 5 (hygiene hypothesis) seek to answer unique research questions which have not been addressed in the context of multiple-case lymphoid cancer families. Specifically, Chapter 3 examines patterns of familial lymphoid cancer co-occurrence compared to the expected population rates; Chapter 4 compares age of onset patterns between sporadic population cases and multiple-case families while controlling for ascertainment bias; and Chapter 5 examines patterns of early lifestyle variables and immune-related diseases on the risk of lymphoma in multiple-case lymphoid cancer families. Because these Chapters seek to address unique research questions, these analyses may be referred to as *hypothesis-generating studies* rather than hypothesis testing or confirmatory studies. Hypothesis-generating studies differ methodologically from confirmatory studies in several ways.

Hypothesis testing research has one or more a priori hypothesis based on existing theory or data (619), whereas hypothesis-generating research explores a set of data searching for relationships and patterns, and then proposes a hypothesis which may be tested in a subsequent study. Confirmatory studies are used to corroborate (replicate) a hypothesis from a hypothesis-generating study. An observational study cannot provide unequivocal results, and therefore an experiment, clinical trial or different study cohort may be required for confirmation. Confirmatory studies also require control of inflated false-positive error risk that is caused by testing multiple null hypotheses (620).

In regards to Chapter 3 (co-occurrence), population-based studies have examined lymphoid cancer co-occurrence patterns among sporadic cases, while no study has examined the familial co-occurrence patterns. In regards to Chapter 4 (age of onset), several familial studies have examined lymphoid cancer age of onset patterns with incorrect statistical methods, or inadequate control of ascertainment or other biases (45,428,448–452). In regards to Chapter 5

(hygiene hypothesis), these relationships have been explored among sporadic population cases in case-control or cohort studies, with the exception of two small family-based studies: Jønsson *et al* (2007) (462) and Royer *et al* (2010) (232).

7.2.6 Identification of genetic factors

Earlier chapters established that some families are likely to have an underlying genetic predisposition to lymphoma. The exome sequencing and genome-wide detection strategies outlined in Chapter 6 are well-validated methods for identifying rare high-penetrance variants of major effects in families and have been previously used by other multi-generational families with hematological malignancies and other cancers.

A significant limitation is the number of families sequenced for variant identification and replication; however, it is unlikely that diverse lymphoid cancer families will have the same susceptibility loci. IBD segment detection produced numerous candidate chromosomal regions in 4 affected relatives. Furthermore, interpretation of data is limited by the assumption that family members who are currently unaffected are truly unaffected and will not develop the disease. In this family, 2 lymphoid cancer cases were diagnosed with HL and 2 cases were diagnosed with NHL. Because HL onset is typically much earlier than other NHL subtypes, we were unable to identify other relatives as truly unaffected as they were still too young (in their 5^{th} decade of life).

A benefit of family-based studies is the identification and removal of Mendelian errors that might otherwise be considered true variants in non-familial genomic studies.

7.2.7 Association vs causation:

Although the data presented in this dissertation implies that some patterns of disease co-occurrence, age of onset, and early childhood variables may increase the risk of some types of lymphoma, these findings are associations that do not necessarily imply a causal mechanism. Similarly, the identification of putative genetic variants that segregate with disease requires functional validation. It is impossible to resolve the functional implications of these associations without replication and validation.

7.3 Future directions

7.3.1 Family recruitment and collaborations

This dissertation includes the largest cohort of lymphoid cancer families in which lymphoid cancer co-occurrence, age of onset patterns, childhood environment and immune-related risk factors have been examined. We had sufficient sample sizes to investigate associations for distinct histological subtypes (e.g., DLBCL, FL, CLL). However, we had insufficient sample sizes to analyze rarer phenotypes (e.g., T-cell NHL, autoimmune conditions such as SLE, and SS) in these families. Future research would benefit from a larger cohort of families. Family recruitment is ongoing in Vancouver, and additional multiple-case lymphoid cancer families can be obtained from collaborative research partners through the Lymphoid Cancer Families Consortium (LCFC), an international collaboration of scientists who have ongoing familial lymphoma studies. A joint-collaborative research study would increase sample sizes and could aid in the identification of additional risk factors in multiple-case lymphoid cancer families. A larger sample size would also enable the analysis of rarer lymphoid subtypes. In addition, increased sample sizes would allow for the analysis of individual autoimmune diseases (e.g., SLE, SS, type 1 diabetes mellitus), allergies (e.g., hay fever, dust) and atopic conditions. The establishment of explicit associations could result in better characterized risk factors and preventative measures.

7.3.2 Sequencing

The candidate variants observed in Family 133 and several published studies of multiple-case lymphoid cancer families are heterogeneous, and no unequivocal lymphoid cancer-causing gene of major effect has yet been identified (376,377,380,433,434,509–513,621). We can improve our understanding of candidate genes by sequencing additional multiple-case families with interesting phenotypes. WES allows for a cost-effective acquisition of genetic information to identify putative disease-causing genetic variants. Interesting candidate genetic factors can be screened in additional multiple-case families (e.g., Vancouver LCF Study, or LCFC cohort) by NGS, Sanger sequencing or a TaqMan assay. We plan to use linkage analysis to try to exclude some parts of the genome shared by Family 133 and simplify interpretation of variants. In addition, candidate variants will be genotyped in additional family members and in other families from the Lymphoid Cancer Families Study.

This collaborative effort is exemplified by a recently accepted publication in *Leukemia* "In search of genetic factors predisposing to familial hairy cell leukemia (HCL): exome-sequencing of four multiplex HCL pedigrees." by Alexander Pemov, and associates [Paper #19-LEU-1150RR] (621). Here, a *CASP9* p.H237P variant shared by 4 relatives with HCL was identified through WES. Additional genotyping in 129 multiplex lymphoid cancer pedigrees revealed that the variant was observed in two additional unrelated families with heterogeneous lymphoid cancers.

Coding variants of small effects size may explain a limited proportion of multiple-case lymphoid cancer families. Whole genome sequencing would increase the search space for susceptibility factors, as it has the potential to uncover larger somatic and germline structural variants, copy number alterations, and cis-regulatory mutations. Mutations in epigenetic modifiers have been associated with lymphomagenesis by deregulating B- and T-cell differentiation during immune response (622). Bisulfite sequencing would elucidate altered patterns of methylation or genomic stability and identify possible candidate genes for follow-up. Gene-environment interactions have been demonstrated to affect risk of lymphoma and may play a role in familial etiology. Some identified gene-environment interactions include hair dye usage and elevated risk of NHL due to genetic variation in *NAT1* and *NAT2* genes (610,623); and sun exposure and sensitivity increased the risk of NHL which was mediated by *IRF4* (610,624). Integrating somatic data from tumours would also provide insights into driver mutations and pathway dysregulation implicated in lymphoid cancer development.

7.3.3 Screening for infectious agents

Infections capable of directly remodeling the epigenome are known to influence lymphomagenesis; the detection of these infections agents in whole blood or tumour cells may be informative for disease etiology (622,625). For example, EBV-related lymphomas are heterogeneous but frequently harbor latent EBV within tumour cells. Some subtypes, such as BL or HIV-associated primary CNS lymphoma are EBV-positive in nearly 100% of cases (625), whereas other subtypes such as HL and DLBCL cases are predicted to be EBV-positive in 30% or very rare instances (625). Viruses such as EBV and HIV prevent recognition by host immunosurveillance by remaining latent inside host cells through epigenetic modification of their genome to mimic their host genome (622,626). EBV may induce tumourigenesis by promoting epigenetic alterations of histones which causes chromatin accessibility for two established latency oncogenes (622,627). Overexpression of these oncogenes eventually results in the development

of EBV-associated tumours (33,62,184,185,196,197,622). Cell-free EBV-DNA can be detected in peripheral blood from dying tumour cells, and is commonly used as a surrogate marker for EBV-positive tumours (625) Although the treatment for EBV-positive and -negative tumours is largely the same, the etiology of lymphomagenesis is different. Therefore, EBV-positive tumours in the familial setting may not harbour genetic susceptibility factors that predispose to cancer.

EBV detection in EBV-associated tumours can be performed through serological tests (e.g., heterophile antibody test) or other methods, such as whole genome sequencing. For FFPE human tissues, an immunohistochemical stain is available, but requires the evaluation by a qualified pathologist.

7.3.4 Functional studies

Candidate genes with sufficient evidence for association with lymphoma risk in multiple-case lymphoid cancer families may not be well characterized (e.g., variants in *CCDC198* and *SPTBN5* identified in Chapter 6). Higher priority variants would be exceptional candidates for functional studies. A practical experiment would be to screen immortalized cell-lines of mutation carriers (and other key family members) for gene expression (the Lymphoma Cancer Families Study collects a lymphocyte fraction during DNA extraction from whole blood samples). Co-immunoprecipitation assays may be used to identify binding partners for uncharacterized genes as described earlier. CRISPR-based genome-editing could also be used to introduce variants of interest into a model cell line, followed by measurements of activity.

7.3.5 Targeted treatment

The identification of susceptibility genes may encourage the development of novel and/or more targeted therapies for treatment of lymphoma. For example, individuals that harbour germline *BRCA1/2*-mutations (and some somatic *BRCA1/2* variants) can be treated with a poly-ADP ribose polymerase inhibitor, which prevent the repair of nicked or ds DNA breaks causing subsequent tumour cell death (628). Given the diverse clinical presentation of lymphoma subtypes and the relative lack of shared susceptibility loci among diverse lymphoid cancer types (with the exception of HLA-6p21.3 and 8q24.21), it is unlikely that lymphomagenesis in multiplex lymphoid cancer families is attributable to one rare highly penetrant variant that has a modest-to-large effect size.

Current treatment options reflect the somatic heterogeneity of disease by providing a combination therapy with several pharmaceutical targets. Pharmaceutical agents typically target key components of disease pathogenesis, which may overlap between histological subtypes (e.g., Rituximab targets CD20-positive B-cells) (629). Precision medicine (or precision oncology) may be used to identify cancer-driving genomic alterations and inform treatment (630). For example, more than 30 mutations in *BRAF* have been observed in several cancers, including NHL, hairy cell leukemia, breast, colorectal cancer and melanoma (630). Individuals with a *BRAF* mutation may be prescribed the same treatment irrespective of their cancer type and location (630). In the future, tumour-normal and variant-pathway analysis may elucidate novel mutation-guided therapies.

7.3.6 Susceptibility patterns and genetic screening

This research establishes that patterns of lymphoid cancers that are unlikely to occur in the general population are more likely to have an underlying genetic predisposition. Similarly, multi-generational families that display evidence of anticipation are of interest to identify or screen for susceptibility factors. For research purposes, identifying families that display these phenotypes can lead to molecular testing to identify putative candidate variants. Several genes have been associated with risk of HL and CLL in the familial context, which could inform future screening of disease in families.

Higher risk families often have homogeneous lymphoid cancers (e.g., all CLL cases) and may be of greater priority for genetic testing and screening methods. Members of these families may be screened for pathogenic variants in CLL-associated genes to determine if they have a genetic predisposition. Family members who carry a putative genetic variant may seek early screening with their doctor ("watch and wait"). The watch-and-wait approach is characteristic of CLL as this disease can be detected from standard physical exams and routine blood draws.

The relatively low incidence of lymphoma, moderate familial risk, and lack of screening tests and associated therapeutic interventions, argue against active clinical surveillance for lymphoma in affected families at this time (41). A combination of genetic risk scores and other factors may improve lymphoid cancer prediction ability over time.

7.4 Significance and contribution to the field

In Chapter 3, I showed that familial lymphoid cancer co-occurrence patterns were different from expected population patterns. These observations are important because it shows that these familial cases are not just random co-occurrence of sporadic population cases in the same family; they are something different, and are likely to have a different combination of etiological factors than sporadic cases. Multiple epidemiological studies have provided clear evidence of CLL (344,345,347,352,467,468,631) and HL (261,345,346,348,350–352,429,632,633) clustering in families, which was also observed in this published chapter. However, current epidemiological literature has not compared lymphoid cancer clustering within families to the expected population frequencies.

Chapter 4 showed that familial ages of onset were substantially earlier than comparable population cases for lymphoid cancers considered as a group, and for NHL, B-cell NHL, DLBCL, FL, MZL, T-cell NHL, CLL, HL and CHL cases. To date, three studies have compared age of lymphoid cancer onset distributions of parent or child generations to that of SEER population data (428,448,449). These studies were limited to a small number of families (24 to 42 families) with parent-child pairs only and did not include families with sibling only cases, or families with cases that skip a generation, which we observe in our collection of lymphoid cancer families. Furthermore, ascertainment biases were not corrected, nor were ethnicity, or age of onset differences between heterogeneous lymphoid cancers (e.g., NHL onset is typically two to three decades later than HL onset). The methods used in Chapter 4 address many of the preceding issues of sample size, ascertainment biases, and potential confounders (e.g., ethnicity), while including *all* multigenerational families and heterogeneous lymphoid cancer subtypes.

Although anticipation has been documented in familial lymphoid cancers, its occurrence has not been widely accepted because of concerns about the potential for ascertainment bias and available statistical methods. To date, 8 studies have reported anticipation effects among lymphoid cancer families (45,428,448–452). However, the interpretability of these findings is limited for several reasons including: use of parent-child pairs only (45,428,450–452), invalid statistical methods or violation of statistical assumptions (45,428,450–452), or uncontrolled ascertainment biases (45,428,448–452). Furthermore, some studies that examined anticipation among heterogeneous lymphoid cancer types did not report subtype information (45,428,448) and did not account for different age of onset patterns among heterogeneous lymphoma subtypes (428,448,451).

In Chapter 4, ascertainment biases and known covariates (age, sex, ethnicity, subtype) were adjusted for while maintaining a moderate to large sample size to examine anticipation effects for heterogeneous lymphoid cancer subtypes. This chapter is also the first study to document anticipation effects among homogeneous lymphoid cancer subtypes (e.g., CLL, DLBCL, LPL/WM).

Chapter 5 examines the role that childhood environment and infectious exposures play on the risk of familial lymphoid cancers. To date, a small number of cohort studies and several case-control studies have examined associations between infectious exposures and risk of sporadic lymphoid cancers; however, there is very limited insight in the familial context. Two studies have examined family structure and immune-related characteristics in multiple-case lymphoid cancer families (232,462).

Jønsson *et al* (2007) observed a paternal parent-offspring birth order effect with predominance of LPD in the youngest siblings among 24 pairs in 32 families enriched for CLL and B-cell malignancies (462). Associations between sibship size and risk of familial CLL and B-cell disorders were not investigated (462). Other immune-related and infectious exposure variables, such SES, parental education, allergies and autoimmune conditions were not available.

Royer *et al* (2010) examined clinical, environmental and occupational differences among 81 families with multiple-cases of WM and related B-cell disorders. One hundred and three familial WM cases were more likely than 272 unaffected relatives to report a history of autoimmune diseases, infections, and allergies. These associations support a role for chronic immune stimulation in the etiology of familial WM (232). Familial WM cases were also more likely to report exposures to farming, pesticides, wood dust and organic solvents compared to their unaffected family members, suggesting a possible role for environmental factors in the development of familial WM (232). Although exploratory, this was the first multiple-case family-based study to examine the effects of environmental factors and the risk of lymphoid cancers. Other factors such as chronic inflammation, surgical procedures (tonsillectomy/adenoidectomy and appendectomy), smoking, and alcohol were not associated with risk of lymphoma among WM families. Measures of childhood lifestyle such as birth order, sibship size, SES, parental education, and family income were not examined.

These two studies aside, there remains a large gap in the understanding of familial lymphoid cancers and its etiology.

In Chapter 5, a collection of 196 heterogeneous lymphoid cancer families were used to examine early childhood environment and personal medical history differences among multiple-case families. This was the first study to establish an inverse relationship between family structure (birth order and sibship size) and risk of cancer for several lymphoid cancer subtypes among multiple-case families. Our finding that MCL, MZL and MALT lymphoma were more frequent among earlier born siblings has not been previously investigated, likely due to small sample size. Strong indicators of high SES during childhood, such as high maternal education and above-average parental income, were associated with an elevated risk of NHL and B-cell subtypes including DLBCL, CLL, FL, but not HL or MM. Living on a farm during childhood was associated with a lower risk of lymphoid and B-cell disorders such as FL. These associations between early childhood environment and risk of lymphoma have not been investigated among other multiple-case family studies. Several population-based epidemiological studies and one WM familial study have explored farm-related exposures; however, most of these studies focus on adult farm-exposures rather than early childhood exposures (which was the focus in our study).

The relationship between autoimmune disorders and allergies has been explored in one multiple-case family-based study; however, these associations are limited to a small number of families, and WM subtypes (232). My thesis (Chapter 5) includes more than 4 times as many cases, and expands the relationship to heterogenous subtypes, such as NHL, B-cell, CLL, HL and CHL. In addition, this thesis is the first to provide evidence that an appendectomy may be associated with the risk of DLBCL, and a tonsillectomy is associated with the risk of lymphoma, including B-cell NHL, DLBCL, HL and CHL subtypes, in the familial context. To date, one multiple-case family study observed no association between tonsillectomy or appendectomy and risk of WM (232).

Numerous high-risk multiple-case lymphoid cancer families have been the subject of whole-genome linkage searches (376,377,380,511,513), or NGS and variant filtering approaches (434,512,634). To date, most other studies have conducted genome-wide searches on homogeneous families with CLL, WM or HL cases (376,377,380,511,513) with additional families (with fewer cases) used for targeted variant sequencing and validation. Linkage studies and NGS studies have not been successful in identifying rare alleles for Mendelian forms of these cancers. The evaluation of low-frequency variants with intermediate effects is still in early research phases, but will continue to be challenged by sample size issues (41).

Chapter 6 examines genetic factors that are shared IBD among a family with 4 heterogeneous lymphoid cancer cases. Several likely deleterious biologically relevant variants were identified. These gene variants will be further characterized in extended sets of multiple-case lymphoid cancer families. Characterization of loss of heterozygosity and somatic mutations in tumours from affected individuals within this family will also be done. Susceptibility loci identified in high-risk lymphoid cancer families may also play a role in the development of sporadic lymphoid cancers.

7.5 Conclusion

Lymphoid cancers are comprised of numerous biologically and clinically heterogeneous subtypes which poses a challenge to understand the etiology of distinct disease entities. Large population-based epidemiological studies have helped to establish lymphoid cancer risk factors, such as advanced age, male sex, comprised immune function and a family history of LPDs; however, there remains a gap in our understanding of familial lymphoid cancer etiology.

The familial nature of heterogeneous lymphoid cancers in multiple-case families suggests a role of shared genetic and/or environmental factors. Relative to the population, our collection of multiple-case lymphoid cancer families showed different co-occurrence and age of onset patterns, suggesting that these families have a different combination of etiological factors. Our analyses showed for the first time that family structure and early life factors affect the risk of multiple types of lymphoid cancers in families. Our observations also provide evidence implicating chronic immune stimulation in the development of heterogeneous lymphoid cancers. These findings and others suggest that familial lymphoid cancers may be influenced by a complex interplay between medical, lifestyle, environmental and host genetic factors. With this comprehensive assessment of multiple-case lymphoid cancer family risk factors, we are beginning to learn more about potential mechanisms underlying lymphomagenesis. In the future, this knowledge may translate into screening or preventative methods and therapeutic targets for lymphoid cancers.

References

1. Olsen M. Overview of hematologic malignancies. Clin Adv Hematol Oncol. 2008;6:4–6.
2. American Society of Hematology . Blood cancers [Internet]. Am. Soc. Hematol. 2020. Available from: https://www.hematology.org/Patients/Cancers/
3. American Society of Hematology . Leukemia [Internet]. Am. Soc. Hematol. 2020. Available from: https://www.hematology.org/Patients/Cancers/Leukemia.aspx
4. American Society of Hematology . Lymphoma [Internet]. Am. Soc. Hematol. 2020. Available from: https://www.hematology.org/Patients/Cancers/Lymphoma.aspx
5. American Society of Hematology . Myeloma [Internet]. Am. Soc. Hematol. 2020. Available from: https://www.hematology.org/Patients/Cancers/Myeloma.aspx
6. Canadian Cancer Society. What is multiple myeloma? [Internet]. Can. Cancer Soc. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/multiple-myeloma/multiple-myeloma/?region=on
7. Foundation LR. About lymphoma [Internet]. 2020. Available from: https://lymphoma.org/aboutlymphoma/
8. Canadian Cancer Society . What is Hodgkin lymphoma? [Internet]. Can. Cancer Soc. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/hodgkin-lymphoma/hodgkin-lymphoma/?region=on
9. Union for International Cancer Control. Hodgkin lymphoma (adult). Union Int. Cancer Control. 2014.
10. IARC. GLOBOCAN 2012 v1.0: Estimated cancer incidence, mortality and prevalence worldwide in 2012. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al., editors. Int. Agency Res. Cancer. Lyon, France; 2012.
11. Wang HW, Balakrishna JP, Pittaluga S, Jaffe ES. Diagnosis of Hodgkin lymphoma in the modern era. Br J Haematol. 2019;184:45–59.
12. Eichenauer DA, Engert A, Andre M, Federico M, Illidge T, Hutchings M, et al. Hodgkin's lymphoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014;25:iii70–5.
13. American Cancer Society . What Is non-Hodgkin lymphoma? [Internet]. 2019. Available from: https://www.cancer.org/cancer/non-hodgkin-lymphoma/about/what-is-non-hodgkin-lymphoma.html
14. Canadian Cancer Society . Burkitt lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/more-types-of-nhl/burkitt-lymphoma/?region=on
15. Canadian Cancer Society. Chronic lymphocytic leukemia statistics [Internet]. 2013. Available from: http://www.cancer.ca/en/cancer-information/cancer-type/leukemia-chronic-lymphocytic-cll/statistics/?region=on

16. Canadian Cancer Society . Diffuse large B-cell lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/diffuse-large-b-cell-lymphoma/?region=on

17. Hartmann S, Döring C, Jakobus C, Rengstl B, Newrzela S, Tousseyn T, et al. Nodular lymphocyte predominant Hodgkin lymphoma and T-cell/histiocyte rich large B-cell lymphoma - Endpoints of a spectrum of one disease? PLoS One. 2013;8:1–10.

18. Canadian Cancer Society . Follicular lymphoma [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/follicular-lymphoma/?region=on

19. Canadian Cancer Society . Hairy cell luekemia [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/leukemia/leukemia/hairy-cell-leukemia/?region=on

20. Canadian Cancer Society . Lymphoplasmacytic lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/more-types-of-nhl/lymphoplasmacytic-lymphoma/?region=on

21. Steingrimsson V, Landgren O, Kristinsson SY. Epidemiology of Waldenström Macroglobulinemia. In: Leblond V, Treon S, Dimoploulos M, editors. Waldenström's Macroglobulinemia. AG, Switzerland: Springer; 2016. page 97–109.

22. Canadian Cancer Society . Mantle cell lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/mantle-cell-lymphoma/?region=on

23. Canadian Cancer Society . MALT lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/malt-lymphoma/?region=on

24. Canadian Cancer Society . Nodal marginal zone lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/more-types-of-nhl/nodal-marginal-zone-lymphoma/?region=on

25. Canadian Cancer Society . Splenic marginal zone lymphoma [Internet]. Can. Cancer Soc. 2020. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/more-types-of-nhl/splenic-marginal-zone-lymphoma/?region=on

26. Greenberg AJ, Rajkumar SV, Vachon CM. Familial monoclonal gammopathy of undetermined significance and multiple myeloma: Epidemiology, risk factors, and biological characteristics. Blood. 2012;119:5359–66.

27. Mcmaster ML, Landgren O. Prevalence, clinical aspects, and natural history of IgM MGUS. Clin Cytom. 2010;97:91–7.

28. Goldin LR, Mcmaster ML, Caporaso NE. Precursors to lymphoproliferative malignancies. Cancer Epidemiol Biomarkers Prev. 2013;22:533–9.

29. Alexander DD, Mink PJ, Adami H-OO, Cole P, Mandel JS, Oken MM, et al. Multiple myeloma: A review of the epidemiologic literature. Int J Cancer. 2007;120:40–61.

30. Malcolm TIM, Hodson DJ, Macintyre EA, Turner SD. Challenging perspectives on the cellular origins of lymphoma. Open Biol. 2016;6:1–12.

31. Pratap S, Scordino TS. Molecular and cellular genetics of non-Hodgkin lymphoma: Diagnostic and prognostic implications. Exp Mol Pathol. 2019;106:44–51.

32. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127:2375–91.

33. Grulich AE, Vajdic CM. The epidemiology of non-Hodgkin lymphoma. Pathology. 2005;37:409–19.

34. Shanafelt TD, Kay NE, Jenkins G, Call TG, Zent CS, Jelinek DF, et al. B-cell count and survival: Differentiating chronic lymphocytic leukemia from monoclonal B-cell lymphocytosis based on clinical outcome. Blood. 2009;113:4188–96.

35. Cheson BD, Bennett JM, Grever M, Kay N, Keating MJ, O'Brien S, et al. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: Revised guidelines for diagnosis and treatment. Blood. 1996;87:4990–7.

36. Strati P, Shanafelt TD. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: Diagnosis, natural history, and risk stratification. Blood. 2015;126:454–62.

37. Call TG, Norman AD, Hanson CA, Achenbach SJ, Kay NE, Zent CS, et al. Incidence of chronic lymphocytic leukemia and high-count monoclonal B-cell lymphocytosis using the 2008 guidelines. Cancer. 2014;120:2000–5.

38. Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Lentini M, et al. Changes in the incidence, pattern of presentation and clinical outcome of early chronic lymphocytic leukemia patients using the 2008 International Workshop on CLL guidelines. Expert Rev Hematol. 2014;7:691–5.

39. Healey R, Patel JL, de Koning L, Naugler C. Incidence of chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis in Calgary, Alberta, Canada. Leuk Res. 2015;39:429–34.

40. Lenartova A, Johannesen TB, Tjønnfjord GE. National trends in incidence and survival of chronic lymphocytic leukemia in Norway for 1953–2012: A systematic analysis of population-based data. Cancer Med. 2016;5:3588–95.

41. Cerhan JR, Slager SL. Familial predisposition and genetic risk factors for lymphoma. Blood. 2015;126:2265–73.

42. Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, et al. Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. J Natl Cancer Inst. 2015;107:1–11.

43. Sud A, Mitchell J, Cooke R, Swerdlow A, Houlston R. A genomic approach to estimating the heritability of Hodgkin lymphoma. Clin Lymphoma Myeloma Leuk. 2015. page S39–

40.

44. Thomsen H, da Silva Filho MI, Försti A, Fuchs M, Ponader S, von Strandmann EP, et al. Heritability estimates on Hodgkin's lymphoma: A genomic- versus population-based approach. Eur J Hum Genet. 2014;23:824–30.

45. Shugart YY, Hemminki K, Vaittinen P, Kingman A, Dong C. A genetic study of Hodgkin's lymphoma: An estimate of heritability and anticipation based on the familial cancer database in Sweden. Hum Genet. 2000;106:553–6.

46. Mitchell JS, Johnson DC, Litchfield K, Broderick P, Weinhold N, Davies FE, et al. Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. Sci Rep. 2015;5:12473.

47. Went M, Sud A, Försti A, Halvarsson B-M, Weinhold N. Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. Nat Commun. 2018;9:1–10.

48. Chattopadhyay S, Thomsen H, Yadav P, da Silva Filho MI, Weinhold N, Nöthen MM, et al. Genome-wide interaction and pathway-based identification of key regulators in multiple myeloma. Commun Biol. 2019;2:1–10.

49. Berndt SI, Morton LM, Wang SS, Teras LR, Slager SL, Vijai J, et al. Genetic heritability of common non-Hodgkin lymphoma subtypes. ASHG 2014 Abstr. 2014.

50. Cerhan JR, Berndt SI, Vijai J, Ghesquières H, McKay J, Wang SS, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B-cell lymphoma. Nat Genet. 2014;46:1233–8.

51. Berndt SI, Camp NJ, Skibola CF, Vijai J, Wang Z, Gu J, et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. Nat Commun. 2016;7:1–9.

52. Law PJ, Berndt SI, Speedy HE, Camp NJ, Sava GP, Skibola CF, et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. Nat Commun. 2017;8:1–12.

53. Berndt SI, Skibola CF, Joseph V, Camp NJ, Nieters A, Wang Z, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat Genet. 2013;45:868–76.

54. Di Bernardo MC, Broderick P, Catovsky D, Houlston RS. Common genetic variation contributes significantly to the risk of developing chronic lymphocytic leukemia. Haematologica. 2013;98:23–4.

55. Canadian Cancer Society . Non-Hodgkin lymphoma statistics [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/statistics/?region=on

56. Canadian Cancer Society . Hodgkin lymphoma statistics [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/hodgkin-lymphoma/statistics/?region=on

57. Kriachok IA. Chronic lymphocytic leukemia [Internet]. Exp. Oncol. 2012. page 378–80.

Available from: https://www.cancer.ca/en/cancer-information/cancer-type/leukemia-chronic-lymphocytic-cll/statistics/?region=on

58. Canadian Cancer Society. Multiple myeloma statistics [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/multiple-myeloma/statistics/?region=on

59. Canadian Cancer Society. Chronic lymphocytic leukemia statistics [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/leukemia-chronic-lymphocytic-cll/statistics/?region=on

60. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, et al. Projected estimates of cancer in Canada in 2020. Can Med Assoc J. 2020;192:E199–205.

61. Canadian Cancer Statistics Advisory Committee. Canadian Cancer Statistics 2019. Can. Cancer Stat. 2019. Toronto, ON; 2019.

62. Alexander DD, Mink PJ, Adami H-O, Chang ET, Cole P, Mandel JS, et al. The non-Hodgkin lymphomas: A review of the epidemiologic literature. Int J Cancer. 2007;120:1–39.

63. Glaser SL, Clarke CA, Keegan THM, Chang ET, Weisenburger DD. Time trends in rates of Hodgkin lymphoma histologic subtypes: True incidence changes or evolving diagnostic practice? Cancer Epidemiol Biomarkers Prev. 2015;24:1474–88.

64. Chihara D, Ito H, Matsuda T, Shibata A, Katsumi A, Nakamura S, et al. Differences in incidence and trends of haematological malignancies in Japan and the United States. Br J Haematol. 2014;164:536–45.

65. Adamson P, Bray F, Costantini AS, Tao M-H, Weiderpass E, Roman E. Time trends in the registration of Hodgkin and non-Hodgkin lymphomas in Europe. Eur J Cancer. 2007;43:391–401.

66. Ries L, Kosary C, Hankey B, Miller B, Clegg L, Edwards B. SEER Cancer Statistcs Review, 1973-1996. Bethesda, MD; 1997.

67. Devesa SS, Fears T. Non-Hodgkin's lymphoma time trends: United States and international data. Cancer Res. 1992;52:5432–40.

68. Sant M, Minicozzi P, Mounier M, Anderson LA, Brenner H, Holleczek B, et al. Survival for haematological malignancies in Europe between 1997 and 2008 by region and age: Results of EUROCARE-5, a population-based study. Lancet Oncol. 2014;15:931–42.

69. Ye X, Mahmud S, Skrabek P, Lix L, Johnston JB. Long-term time trends in incidence, survival and mortality of lymphomas by subtype among adults in Manitoba, Canada: A population-based study using cancer registry data. BMJ Open. 2017;7:1–17.

70. Bosetti C, Levi F, Ferlay J, Lucchini F, Negri E, La Vecchia C. Incidence and mortality from non-Hodgkin lymphoma in Europe: The end of an epidemic? Int J Cancer. 2008;123:1917–23.

71. Ekström-Smedby K. Epidemiology and etiology of non-Hodgkin lymphoma - A review. Acta Oncol (Madr). 2006;45:258–71.

72. Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. Blood. 2006;107:265–76.

73. Cancer Research UK. Hodgkin lymphoma statistics [Internet]. Cancer Res. UK. 2019. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/hodgkin-lymphoma#heading-One

74. American Cancer Society ., Barnes C, Bray F, Drope J, Fedewa S, Ferlay J, et al. Global cancer facts & figures 4th Edition. 4th Editio. Am. Cancer Soc. Atlanta, GA; 2018.

75. Biggar R. AIDS-related cancers in the era of highly active antiretroviral therapy. Oncol - Willist Park. 2001;15:439–48.

76. Negri E, Little D, Boiocchi M, La Vecchia C, Franceschi S. B-cell non-Hodgkin's lymphoma and hepatitis C virus infection: A systematic review. Int. J. Cancer. 2004. page 1–8.

77. Cowan AJ, Allen C, Barac A, Basaleem H, Bensenor I, Curado MP, et al. Global burden of multiple myeloma: A systematic analysis for the global burden of disease study 2016. JAMA Oncol. 2018;4:1221–7.

78. Cancer Research UK . Myeloma mortality statistics [Internet]. Cancer Res. UK. 2019. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/mortality

79. Vélez R, Turesson I, Landgren O, Kristinsson SY, Cuzick J. Incidence of multiple myeloma in Great Britain, Sweden, and Malmö, Sweden: The impact of differences in case ascertainment on observed incidence trends. BMJ Open. 2016;6:1–5.

80. Turesson I, Velez R, Kristinsson SY, Landgren OLA. Patterns of multiple myeloma during the past 5 decades: Stable incidence rates for all age groups in the population but rapidly changing age distribution in the clinic. Mayo Clin Proc. 2010;85:225–30.

81. Velez R, Beral V, Cuzick J. Increasing trends of multiple myeloma mortality in England and Wales; 1950-79: Are the changes real? J Natl Cancer Inst. 1982;69:387–92.

82. Devesa SS, Silverman DT. Cancer incidence and mortality trends in the United States: 1935-74. J Natl Cancer Inst. 1978;60:545–71.

83. Salati M, Cesaretti M, Macchia M, El Mistiri M, Federico M. Epidemiological overview of Hodgkin lymphoma across the Mediterranean basin. Mediterr J Hematol Infect Dis. 2014;6:1–10.

84. American Cancer Society . What are the risk factors for chronic lymphocytic leukemia? [Internet]. 2019. Available from: https://www.cancer.org/cancer/chronic-lymphocytic-leukemia/causes-risks-prevention/risk-factors.html

85. Smedby KE, Sampson JN, Turner JJ, Slager SL, Maynadié M, Roman E, et al. Medical history, lifestyle, family history, and occupational risk factors for mantle cell lymphoma: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. Oxford University Press; 2014;2014:76–86.

86. National Cancer Institute, DCCPS, Surveillance Research Program . Surveillance,

Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2018 Sub (1975-2016 varying). 2019.

87. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin K (eds). SEER Cancer Statistics Review 1975-2013; Non-Hodgkin Lymphoma. SEER Cancer Stat. Rev. 1975-2013, Natl. Cancer Institute. Bethesda, MD, http//seer.cancer.gov/csr/1975_2013/, based Novemb. 2015 SEER data submission, posted to SEER web site, April 2016. Bethesda, MD; 2016.

88. Canadian Cancer Society . Risk factors for chronic lymphocytic leukemia (CLL) [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/leukemia-chronic-lymphocytic-cll/risks/?region=on

89. Thomas RK, Re D, Zander T, Wolf J, Diehl V. Epidemiology and etiology of Hodgkin's lymphoma. Ann Oncol. Oxford University Press; 2002;13:147–52.

90. Evens AM, Antillón M, Aschebrook-Kilfoy B, Chiu BC-H. Racial disparities in Hodgkin's lymphoma: A comprehensive population-based analysis. Ann Oncol. 2012;23:2128–37.

91. Ruchlemer R, Polliack A. Geography, ethnicity and "roots" in chronic lymphocytic leukemia. Leuk Lymphoma. 2013;54:1142–50.

92. Scherr P, Mueller N. Non-Hodgkin's lymphomas. 2nd ed. Schottenfeld D, Fraumeni J, editors. New York: Oxford University Press; 1996.

93. Shvidel L, Shtarlid M, Klepfish A, Sigler E, Berrebi A. Epidemiology and ethnic aspects of B-cell chronic lymphocytic leukemia in Israel. Leukemia. 1998;12:1612–7.

94. Yavorkovsky LI, Terebkova ZF, Nikulshin SV, Yavorkovsky LL. B-chronic lymphocytic leukaemia in Latvia: Epidemiological aspects. Eur J Haematol. 1993;51:214–7.

95. McNally RJQ, Cairns DP, Eden OB, Kelsey AM, Taylor GM, Birch JM. Examination of temporal trends in the incidence of childhood leukemias and lymphomas provides etiological clues. Leukemia. 2001;15:1612–8.

96. Gale RP, Cozen W, Goodman MT, Wang FF, Bernstein L. Decreased chronic lymphocytic leukemia incidence in asians in Los Angeles County. Leuk Res. 2000;24:665–9.

97. Cuttner J. Increased incidence of hematologic malignancies in first-degree relatives of patients with chronic lymphocytic leukemia. Cancer Invest. 1992;10:103–9.

98. Herrinton L, Goldoft M, Schwartz S, Weiss N. The incidence of non-Hodgkin's lymphoma and its histologic subtypes in Asian migrants to the United States and their descendants. Cancer Causes Control. 1996;7:224–30.

99. Au WY, Gascoyne RD, Gallagher RE, Le N, Klasa RD, Liang RHS, et al. Hodgkin's lymphoma in Chinese migrants to British Columbia: A 25-year survey. Ann Oncol. 2004;15:626–30.

100. Fitzmaurice C, Hamavid H, Dicker D, Allen C, Naghavi M. The global burden of non-Hodgkin lymphoma. Clin Lymphoma Myeloma Leuk. 2015. page S70.

101. Perry AM, Diebold J, Nathwani BN, Maclennan KA, Müller-Hermelink HK, Bast M, et al. Non-Hodgkin lymphoma in the developing world: Review of 4539 cases from the International Non-Hodgkin Lymphoma Classification Project. Haematologica. 2016;101:1244–50.

102. Akinyemiju T, Ogunsina K, Okwali M, Sakhuja S, Braithwaite D. Lifecourse socioeconomic status and cancer-related risk factors: Analysis of the WHO study on global ageing and adult health (SAGE). Int J Cancer. 2017;140:777–87.

103. Wei X, Calvo-Vidal MN, Chen S, Wu G, Revuelta M V., Sun J, et al. Germline lysine-specific demethylase 1 (lsd1/kdm1a) mutations confer susceptibility to multiple myeloma. Cancer Res. 2018;78:2747–59.

104. Tward JD, Wendland MMM, Shrieve DC, Szabo A, Gaffney DK. The risk of secondary malignancies over 30 years after the treatment of non-Hodgkin lymphoma. Cancer. 2006;107:108–15.

105. Pirani M, Marcheselli R, Marcheselli L, Bari A, Federico M, Sacchi S. Risk for second malignancies in non-Hodgkin's lymphoma survivors: A meta-analysis. Ann Oncol. 2011;22:1845–58.

106. Mudie NY, Swerdlow AJ, Higgins CD, Smith P, Qiao Z, Hancock BW, et al. Risk of second malignancy after non-Hodgkin's lymphoma: A British cohort study. J Clin Oncol. 2006;24:1568–74.

107. Sud A, Hemminki K, Houlston RS. Second cancer risk following Hodgkin lymphoma. Oncotarget. 2017;8:78261–2.

108. Frank C, Fallah M, Chen T, Mai EK, Sundquist J, Försti A, et al. Search for familial clustering of multiple myeloma with any cancer. Leukemia. 2016;30:627–32.

109. Tsimberidou AM, Wen S, McLaughlin P, O'Brien S, Wierda WG, Lerner S, et al. Other malignancies in chronic lymphocytic leukemia/small lymphocytic lymphoma. J Clin Oncol. 2009;27:904–10.

110. Lankes HA, Fought AJ, Evens AM, Weisenburger DD, Chiu BC-H. Vaccination history and risk of non-Hodgkin lymphoma: A population-based, case-control study. Cancer Causes Control. 2009;20:517–23.

111. Smedby KE, Vajdic CM, Falster M, Engels EA, Martinex-Maza O, Turner J, et al. Autoimmune disorders and risk of non-Hodgkin lymphoma subtypes: A pooled analysis within the InterLymph Consortium. Blood. 2008;111:4029–38.

112. Chua I, Quinti I, Grimbacher B. Lymphoma in common variable immunodeficiency: Interplay between immune dysregulation, infection and genetics. Curr Opin Hematol. 2008;15:368–74.

113. Canadian Cancer Society . Post-transplant lymphoproliferative disorder [Internet]. 2019. Available from: https://www.cancer.ca/en/cancer-information/cancer-type/non-hodgkin-lymphoma/non-hodgkin-lymphoma/more-types-of-nhl/post-transplant-lymphoproliferative-disorder/?region=ab

114. Opelz G, Henderson R. Incidence of non-Hodgkin lymphoma in kidney and heart

transplant recipients. Lancet. 1993;342:1514–6.

115. Clarke CA, Morton LM, Lynch C, Pfeiffer RM, Hall EC, Gibson TM, et al. Risk of lymphoma subtypes after solid organ transplantation in the United States. Br J Cancer. 2013;109:280–8.

116. Opelz G, Döhler B. Lymphomas after solid organ transplantation: A Collaborative Transplant Study report. Am J Transplant. 2004;4:222–30.

117. Franceschi S, Dal Maso LD, La Vecchia CL. Advances in the epidemiology of HIV-associated non-Hodgkin's lymphoma and other lymphoid neoplasms. Int J Cancer. 1999;83:481–5.

118. Grulich AE, Wan X, Law MG, Milliken ST, Lewis CR, Garsia RJ, et al. B-cell stimulation and prolonged immune deficiency are risk factors for non-Hodgkin's lymphoma in people with AIDS. AIDS. 2000;14:133–40.

119. Oku K, Atsumi T. Systemic lupus erythematosus: Nothing stale her infinite variety. Mod Rheumatol. Taylor & Francis; 2018;28:758–65.

120. Patel R, Shahane A. The epidemiology of Sjögren's syndrome. Clin Epidemiol. 2014;6:247–55.

121. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. Bull NYU Hosp Jt Dis. 2006;64:12–5.

122. Scott DL, Wolfe F, Huizinga TWJ. Rheumatoid arthritis. Comorbidity Rheum Dis. 2010;376:1094–108.

123. Morton LM, Slager SL, Cerhan JR, Wang SS, Vajdic CM, Christine F, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:130–44.

124. Kleinstern G, Maurer MJ, Liebow M, Habermann TM, Koff JL, Allmer C, et al. History of autoimmune conditions and lymphoma prognosis. Blood Cancer J. 2018;8:1–10.

125. Smedby KE, Hjalgrim H, Askling J, Chang ET, Gregersen H, Porwit-MacDonald A, et al. Autoimmune and chronic inflammatory disorders and risk of non-Hodgkin lymphoma by subtype. J Natl Cancer Inst. 2006;98:51–60.

126. Anderson LA, Gadalla S, Morton LM, Landgren O, Pfeiffer R, Warren JL, et al. Population-based study of autoimmune conditions and the risk of specific lymphoid malignancies. Int J Cancer. 2009;125:398–405.

127. Kristinsson SY, Landgren O, Sjöberg J, Turesson I, Björkholm M, Goldin LR. Autoimmunity and risk for Hodgkin's lymphoma by subtype. Haematologica. 2009;94:1468–9.

128. Landgren O, Engels EA, Pfeiffer RM, Gridley G, Mellemkjaer L, Olsen JH, et al. Autoimmunity and susceptibility to Hodgkin lymphoma: A population-based case-control study in Scandinavia. J Natl Cancer Inst. 2006;98:1321–30.

129. Hemminki K, Försti A, Sundquist K, Sundquist J, Li X. Familial associations of lymphoma and myeloma with autoimmune diseases. Blood Cancer J. 2017;7:1–5.

130. Cerhan JR, Kricker A, Paltiel O, Flowers CR, Wang SS, Monnereau A, et al. Medical history, lifestyle, family history, and occupational risk factors for diffuse large B-cell lymphoma: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:15–25.

131. Bracci PM, Benavente Y, Turner JJ, Paltiel O, Slager SL, Vajdic CM, et al. Medical history, lifestyle, family history, and occupational risk factors for marginal zone lymphoma: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:52–65.

132. Vajdic CM, Landgren O, McMaster ML, Slager SL, Brooks-Wilson A, Smith A, et al. Medical history, lifestyle, family history, and occupational risk factors for lymphoplasmacytic lymphoma/Waldenström's macroglobulinemia: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:87–97.

133. Linet MS, Vajdic CM, Morton LM, De Roos AJ, Skibola CF, Boffetta P, et al. Medical history, lifestyle, family history, and occupational risk factors for follicular lymphoma: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:26–40.

134. Söderberg KC, Jonsson F, Winqvist O, Hagmar L, Feychting M. Autoimmune diseases, asthma and risk of haematological malignancies: A nationwide case-control study in Sweden. Eur J Cancer. 2006;42:3028–33.

135. Landgren O, Engels EA, Caporaso NE, Gridley G, Mellemkjaer L, Hemminki K, et al. Patterns of autoimmunity and subsequent chronic lymphocytic leukemia in Nordic countries. Blood. 2006;108:292–6.

136. Björnådal L, Löfström B, Yin L, Lundberg IE, Ekbom A. Increased cancer incidence in a Swedish cohort of patients with systemic lupus erythematosus. Scand J Rheumatol. 2002;31:66–71.

137. Bernatsky S, Boivin JF, Joseph L, Rajan R, Zoma A, Manzi S, et al. An international cohort study of cancer in systemic lupus erythematosus. Arthritis Rheum. 2005;52:1481–90.

138. Zhang Y, Holford TR, Leaderer B, Zahm SH, Boyle P, Morton LM, et al. Prior medical conditions and medication use and risk of non-Hodgkin lymphoma in Connecticut United States women. Cancer Causes Control. 2004;15:419–28.

139. Vajdic CM, Falster MO, Sanjose S De, Martínez-Maza O, Becker N, Bracci PM, et al. Atopic disease and risk of non-Hodgkin lymphoma: An InterLymph pooled analysis. Cancer Res. 2009;69:6482–9.

140. Gupta R, Sheikh A, Strachan DP, Anderson HR. Burden of allergic disease in the UK: Secondary analyses of national databases. Clin Exp Allergy. 2004;34:520–6.

141. Helby J, Bojesen SE, Nielsen SF, Nordestgaard BG. IgE and risk of cancer in 37, 747 individuals from the general population. Ann Oncol. 2015;26:1784–90.

142. Melbye M, Smedby KE, Lehtinen T, Rostgaard K, Glimelius B, Munksgaard L, et al. Atopy and risk of non-Hodgkin lymphoma. J Natl Cancer Inst. 2007;99:158–66.

143. Briggs NC, Levine RS, Brann EA. Allergies and risk of non-Hodgkin's lymphoma by subtype. Cancer Epidemiol Biomarkers Prev. 2002;11:401–7.

144. Musolino C, Allegra A, Minciullo PL, Gangemi S. Allergy and risk of hematologic malignancies: Associations and mechanisms. Leuk Res. 2014;38:1137–44.

145. Holly EA, Lele C, Bracci PM, McGrath MS. Case-control study of non-Hodgkin's lymphoma among women and heterosexual men in the San Francisco Bay Area, California. Am J Epidemiol. 1999;150:375–89.

146. Becker N, Deeg E, Rüdiger T, Nieters A. Medical history and risk for lymphoma: Results of a population-based case-control study in Germany. Eur J Cancer. 2005;41:133–42.

147. Grulich AE, Vajdic CM, Kaldor JM, Hughes AM, Kricker A, Fritschi L, et al. Birth order, atopy, and risk of non-Hodgkin lymphoma. J Natl Cancer Inst. 2005;97:587–94.

148. Yang J, Hong'en X, Xiaodong L, Shiliang L, Baihua L, Yongshi J. Allergic conditions are not associated with the risk of non-Hodgkin's lymphoma or Hodgkin's lymphoma: A systematic review and meta-analysis. Onco Targets Ther. 2017;10:2189–98.

149. Rudant J, Orsi L, Menegaux F, Petit A, Baruchel A, Bertrand Y, et al. Childhood acute leukemia, early common infections, and allergy: The ESCALE study. Am J Epidemiol. 2010;172:1015–27.

150. Zhou M-H, Yang Q-M. Association of asthma with the risk of acute leukemia and non-Hodgkin lymphoma. Mol Clin Oncol. 2015;3:859–64.

151. Merrill RM, Isakson RT, Beck RE. The association between allergies and cancer: What is currently known? Ann Allergy, Asthma Immunol. American College of Allergy, Asthma & Immunology; 2007;99:102–17.

152. Wang SS, Flowers CR, Kadin ME, Chang ET, Hughes AM, Ansell SM, et al. Medical history, lifestyle, family history, and occupational risk factors for peripheral T-cell lymphomas: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:66–75.

153. Vineis P, Crosignani P, Sacerdote C, Fontana A, Masala G, Miligi L, et al. Haematopoietic cancer and medical history: A multicentre case-control study. J Epidemiol Community Health. 2000;54:431–6.

154. Holly E, Lele C. Non-Hodgkin's lymphoma in HIV-positive and HIV-negative homosexual men in the San Francisco Bay Area: Allergies, prior medication use, and sexual practices. J Acquir Immune Defic Syndr. 1997;15:211–22.

155. Holly EA, Bracci PM. Population-based study of non-Hodgkin lymphoma, histology, and medical history among human immunodeficiency virus-negative participants in San Francisco. Am J Epidemiol. 2003;158:316–27.

156. Bernstein L, Ross RK. Prior medication use and health history as risk factors for non-Hodgkin's lymphoma: Preliminary results from a case-control study in Los Angeles County. Cancer Res. 1992;52:5510s-5516s.

157. Becker N, de Sanjose S, Nieters A, Maynadié M, Foretova L, Cocco PL, et al. Birth order, allergies and lymphoma risk: Results of the European collaborative research project

Epilymph. Leuk Res. 2007;31:1365–72.

158. Cartwright RA, McKinney PA, O'Brien C, Richards IDG, Roberts B, Lauder I, et al. Non-Hodgkin's lymphoma: Case-control epidemiological study in Yorkshire. Leuk Res. 1988;12:81–8.

159. Lewis DR, Pottern LM, Brown LM, Silverman DT, Haves RB, Schoenberg JB, et al. Multiple myeloma among Blacks and Whites in the United States: The role of chronic antigenic stimulation. Cancer Causes Control. 1994;5:529–39.

160. Söderberg KC, Hagmar L, Schwartzbaum J, Feychting M. Allergic conditions and risk of hematological malignancies in adults: A cohort study. BMC Public Health. 2004;4:1–6.

161. Legendre L, Barnetche T, Mazereeuw-Hautier J, Meyer N, Murrell D, Paul C. Risk of lymphoma in patients with atopic dermatitis and the role of topical treatment: A systematic review and meta-analysis. J Am Acad Dermatol. 2015;72:992–1002.

162. Arellano FM, Arana A, Wentworth CE, Fernández-Vidaurre C, Schlienger RG, Conde E. Lymphoma among patients with atopic dermatitis and/or treated with topical immunosuppressants in the United Kingdom. J Allergy Clin Immunol. 2009;123:1111–6.

163. Kato I, Koenig K, Shore R, Baptiste M, Lillquist P, Frizzera G, et al. Use of anti-inflammatory and non-narcotic analgesic drugs and risk of non-Hodgkin's lymphoma (NHL) (United States). Cancer Causes Control. 2002;13:965–74.

164. Chang ET, Smedby KE, Hjalgrim H, Schöllkopf C, Porwit-MacDonald A, Sundström C, et al. Medication use and risk of non-Hodgkin's lymphoma. Am J Epidemiol. 2005;162:965–74.

165. Subbarao P, Mandhane PJ, Sears MR. Asthma: Epidemiology, etiology and risk factors. Can Med Assoc J. 2009;181:E181–90.

166. Von Ehrenstein OS, Von Mutius E, Illi S, Baumann L, Böhm O, Von Kries R. Reduced risk of hay fever and asthma among children of farmers. Clin Exp Allergy. 2000;30:187–93.

167. Ober C, Yao T-C. The genetics of asthma and allergic disease: A 21st century perspective. Immunol Rev. 2011;242:10–30.

168. Almqvist C, Olsson H, Fall T, Lundholm C. Sibship and risk of asthma in a total population: A disease comparative approach. J Allergy Clin Immunol. 2016;138:1219–22.

169. Linabery AM, Jurek AM, Duval S, Ross JA. The association between atopy and childhood/adolescent leukemia: A meta-analysis. Am J Epidemiol. 2010;171:749–64.

170. Rudant J, Orsi L, Monnereau A, Patte C, Pacquement H, Landman-Parker J, et al. Childhood Hodgkin's lymphoma, non-Hodgkin's lymphoma and factors related to the immune system: The Escale Study (SFCE). Int J Cancer. 2011;129:2236–47.

171. Linabery AM, Erhardt EB, Fonstad RK, Ambinder RF, Bunin GR, Ross JA, et al. Infectious, autoimmune and allergic diseases and risk of Hodgkin lymphoma in children and adolescents: A Children's Oncology Group study. Int J Cancer. 2014;1469:1454–69.

172. Dikalioti SK, Chang ET, Dessypris N, Papadopoulou C, Skenderis N, Pourtsidis A, et al.

Allergy-associated symptoms in relation to childhood non-Hodgkin's as contrasted to Hodgkin's lymphomas: A case-control study in Greece and meta-analysis. Eur J Cancer. 2012;48:1860–6.

173. Aschebrook-Kilfoy B, Cocco P, La Vecchia C, Chang ET, Vajdic CM, Kadin ME, et al. Medical history, lifestyle, family history, and occupational risk factors for mycosis fungoides and sézary syndrome: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:98–105.

174. Mbulaiteye SM, Morton LM, Sampson JN, Chang ET, Costas L, de Sanjosé S, et al. Medical history, lifestyle, family history, and occupational risk factors for sporadic Burkitt lymphoma/leukemia: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:106–14.

175. Slager SL, Benavente Y, Blair A, Vermeulen R, Cerhan JR, Costantini AS, et al. Medical history, lifestyle, family history, and occupational risk factors for chronic lymphocytic leukemia/small lymphocytic lymphoma: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:41–51.

176. Brown LM, Gridley G, Check D, Landgren O. Risk of multiple myeloma and monoclonal gammopathy of undetermined significance among white and black male United States veterans with prior autoimmune, infectious, inflammatory, and allergic disorders. Blood. 2008;111:3388–94.

177. Vlajinac HD, Pekmezović TD, Adanja BJ, Marinković JM, Kanazir MS, Suvajdžć ND, et al. Case-control study of multiple myeloma with special reference to diet as risk factor. Neoplasma. 2003;50:79–83.

178. Mills P, Beeson W, Fraser G, Phillips R. Allergy and cancer: Organ site-specific results from the Adventist Health Study. Am J Epidemiol. 1992;136:287–95.

179. Chang ET, Zheng T, Weir EG, Borowitz M, Mann RB, Spiegelman D, et al. Childhood social environment and Hodgkin's lymphoma: New findings from a population-based case-control study. Cancer Epidemiol Biomarkers Prev. 2004;13:1361–70.

180. Boffetta P, Stellman SD, Garfinkel L. A case-control study of multiple myeloma nested in the American Cancer Society prospective study. Int J Cancer. 1989;43:554–9.

181. Monnereau A, Slager SL, Hughes AM, Smith A, Glimelius B, Habermann TM, et al. Medical history, lifestyle, and occupational risk factors for hairy cell leukemia: The InterLymph non-Hodgkin lymphoma subtypes project. J Natl Cancer Inst - Monogr. 2014;2014:115–24.

182. American Cancer Society . Non-Hodgkin lymphoma risk factors [Internet]. Am. Cancer Soc. 2019. Available from: https://www.cancer.org/cancer/non-hodgkin-lymphoma/causes-risks-prevention/risk-factors.html

183. Iqbal T, Mahale P, Turturro F, Kyvernitakis A, Torres HA. Prevalence and association of hepatitis C virus infection with different types of lymphoma. Int J Cancer. 2016;138:1035–7.

184. Engels EA. Infectious agents as causes of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev. 2007;16:401–4.

185. Bassig BA, Lan Q, Rothman N, Zhang Y, Zheng T. Current understanding of lifestyle and environmental factors and risk of non-Hodgkin lymphoma: An epidemiological update. J Cancer Epidemiol. 2012;2012:1–27.

186. Levine AM. Lymphomas and leukemias due to infectious organisms. Hematology. 2012;17:S87–9.

187. Fisher SG, Fisher RI. The epidemiology of non-Hodgkin's lymphoma. Oncogene. 2004;23:6524–34.

188. de Sanjose S, Dickie A, Alvaro T, Romagosa V, Villanueva MG, Domingo-Domenech E, et al. Helicobacter pylori and malignant lymphoma in Spain. Cancer Epidemiol Biomarkers Prev. 2004;13:944–8.

189. Wotherspoon AC, Doglioni C, Diss TC, Pan L, Moschini A, de Boni M, et al. Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of Helicobacter pylori. Lancet. 1993;342:575–7.

190. Hjalgrim H, Engels EA. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: A review of the epidemiological evidence. J Intern Med. 2008;264:537–48.

191. Altieri A, Castro F, Bermejo JL, Hemminki K. Number of siblings and the risk of lymphoma, leukemia, and myeloma by histopathology. Cancer Epidemiol Biomarkers Prev. 2006;15:1281–6.

192. de Sanjose S, Benavente Y, Vajdic CM, Engels EA, Morton LM, Bracci PM, et al. Hepatitis C and non-Hodgkin lymphoma among 4784 cases and 6269 controls from the International Lymphoma Epidemiology Consortium. Clin Gastroenterol Hepatol. 2008;6:451–8.

193. Murphy EL, Figueroa JP, Gibbs WN, Holding-Cobham M, Cranston B, Malley K, et al. Human T-lymphotropic virus type I (HTLV-I) seroprevalence in Jamaica I. Demographic determinants. Am J Epidemiol. 1991;133:1114–24.

194. Borie R, Cadranel J, Guihot A, Marcelin AG, Galicier L, Couderc LJ. Pulmonary manifestations of human herpesvirus-8 during HIV infection. Eur Respir J. 2013;42:1105–18.

195. de Sanjosé S, Goedert JJ, Marshall V, Bellas C, Benavente Y, Bosch R, et al. Risk of malignant lymphoma associated with human herpesvirus-8: A case-control study in Spain. Br J Cancer. 2004;90:2145–8.

196. IARC . Epstein-Barr virus and Kaposi's sarcoma herpesvirus/human herpesvirus-8. Proc. IARC Work. Gr. Eval. Carcinog. Risks to Humans. Lyon, France; 1997.

197. Thorley-Lawson DA, Gross A. Persistence of the Epstein–Barr virus and the origins of associated lymphomas. N Engl J Med. 2004;350:1328–37.

198. Blajchman MA. Transfusion immunomodulation or TRIM: What does it mean clinically? Hematology. 2005;10:208–14.

199. Chang CM, Quinlan SC, Warren JL, Engels EA. Blood transfusions and the subsequent risk of hematologic malignancies. Transfusion. 2010;50:2249–57.

200. Cerhan JR, Wallace RB, Dick F, Kemp J, Parker AS, Zheng W, et al. Blood transfusions and risk of non-Hodgkin's lymphoma subtypes and chronic lymphocytic leukemia. Cancer Epidemiol Biomarkers Prev. 2001;10:361–8.

201. Gutensohn N, Cole P. Childhood social and environment and Hodgkin's disease. N Engl J Med. 1981;304:135–40.

202. Serraino D, Francheschi S, Talamini R, Barra S, Negri E, Carbone A, et al. Socio-economic indicators, infectious diseases and Hodgkin's disease. Int J Cancer. 1991;47:352–7.

203. Hjalgrim H, Smedby KE, Rostgaard K, Molin D, Hamilton-Dutoit S, Chang ET, et al. Infectious mononucleosis, childhood social environment, and risk of Hodgkin lymphoma. Cancer Res. 2007;67:2382–8.

204. Becker N, Fortuny J, Alvaro T, Nieters A, Maynadié M, Foretova L, et al. Medical history and risk of lymphoma: Results of a European case-control study (EPILYMPH). J Cancer Res Clin Oncol. 2009;135:1099–107.

205. Monnereau A, Orsi L, Troussard X, Berthou C, Fenaux P, Marit G, et al. History of infections and vaccinations and risk of lymphoid neoplasms: Does influenza immunization reduce the risk? Leukemia. 2007;21:2075–9.

206. Doody MM, Linet MS, Glass AG, Friedman GG, Pottern LM, Boice Jr JD, et al. Leukemia, lymphoma, and multiple myeloma following selected medical conditions. Cancer Causes Control. 1992;3:449–56.

207. Gramenzi A, Buttino I, D'Avanzo B, Negri E, Franceschi S, La Vecchia C. Medical history and the risk of multiple myeloma. Br J Cancer. 1991;63:769–72.

208. Tavani A, La Vecchia C, Franceschi S, Serraino D, Carbone A, Franceshia S, et al. Medical history and risk of Hodgkin's and non-Hodgkin's lymphomas. Eur J Cancer Prev. 2000;9:59–64.

209. MacArthur AC, McBride ML, Spinelli JJ, Tamaro S, Gallagher RP, Theriault GP. Risk of childhood leukemia associated with vaccination, infection, and medication use in childhood: The Cross-Canada Childhood Leukemia Study. Am J Epidemiol. 2008;167:598–606.

210. Vineis P, Miligi L, Crosignani P, Fontana A, Masala G, Nanni O, et al. Delayed infection, family size and malignant lymphomas. J Epidemiol Community Health. 2000;54:907–11.

211. Jourdan-Da Silva N, Perel Y, Méchinaud F, Plouvier E, Gandemer V, Lutz P, et al. Infectious diseases in the first year of life, perinatal characteristics and childhood acute leukaemia. Br J Cancer. 2004;90:139–45.

212. Becker N, Deeg E, Nieters A. Population-based study of lymphoma in Germany: Rationale, study design and first results. Leuk Res. 2004;28:713–24.

213. Chang ET, Zheng T, Weir EG, Borowitz M, Mann RB, Spiegelman D, et al. Aspirin and the risk of Hodgkin's lymphoma in a population-based case-control study. J Natl Cancer Inst. 2004;96:305–15.

214. Landgren O, Zhang Y, Zahm SH, Inskip P, Zheng T, Baris D. Risk of multiple myeloma

following medication use and medical conditions: A case-control study in Connecticut women. Cancer Epidemiol Biomarkers Prev. 2006;15:2342–7.

215. Nuyujukian DS, Voutsinas J, Bernstein L, Wang SS. Medication use and multiple myeloma risk in Los Angeles County. Cancer Causes Control. 2014;25:1233–7.

216. Walter RB, Milano F, Brasky TM, White E. Long-term use of acetaminophen, aspirin, and other nonsteroidal anti-inflammatory drugs and risk of hematologic malignancies: Results from the prospective vitamins and lifestyle (VITAL) study. J Clin Oncol. 2011;29:2424–31.

217. Ji J, Sundquist J, Sundquist K. Tonsillectomy associated with an increased risk of autoimmune diseases: A national cohort study. J Autoimmun. 2016;72:1–7.

218. Vestergaard H, Westergaard T, Wohlfahrt J, Hjalgrim H, Melbye M. Tonsillitis, tonsillectomy and Hodgkin's lymphoma. Int J Cancer. 2010;127:633–7.

219. Van Kempen MJP, Rijkers GT, Van Cauwenberge PB. The immune response in adenoids and tonsils. Int Arch Allergy Immunol. 2000;122:8–19.

220. La Vecchia C, Negri E, Franceschi S. Medical history and the risk of non-Hodgkin's lymphomas. Cancer Epidemiol Biomarkers Prev. 1992;1:533–6.

221. Liaw K-L, Adami J, Gridley G, Nyren O, Linet MS. Risk of Hodgkin's disease subsequent to tonsillectomy: A population-based cohort study in Sweden. Int J Cancer. 1997;72:711–3.

222. Bonelli L, Vitale V, Bistolfi F, Landucci M, Bruzzi P. Hodgkin's disease in adults: Association with social factors and age at tonsillectomy. A case-control study. Int J Cancer. 1990;45:423–7.

223. Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy is followed by increased risk of Crohn's disease. Gastroenterology. 2003;124:40–6.

224. Cope JU, Askling J, Gridley G, Mohr A, Ekbom A, Nyren O, et al. Appendectomy during childhood and adolescence and the subsequent risk of cancer in Sweden. Pediatrics. 2003;111:1343–50.

225. Mohammadi M, Song H, Cao Y, Glimelius I, Ekbom A, Ye W, et al. Risk of lymphoid neoplasms in a Swedish population-based cohort of 337,437 patients undergoing appendectomy. Scand J Gastroenterol. 2016;51:583–9.

226. Cozen W, Hamilton AS, Zhao P, Salam MT, Deapen DM, Nathwani BN, et al. A protective role for early oral exposures in the etiology of young adult Hodgkin lymphoma. Blood. 2009;114:4014–20.

227. Hyams L, Wynder EL. Appendectomy and cancer risk: An epidemiological evaluation. J Chronic Dis. 1967;21:391–415.

228. Bierman HR. Human appendix and neoplasia. Cancer. 1968;21:109–18.

229. Mellemkjær L, Johansen C, Linet MS, Gridley G, Olsen JH. Cancer risk following appendectomy for acute appendicitis (Denmark). Cancer Causes Control. 1998;9:183–7.

230. Silingardi V, Venezia L, Tampieri A, Gramolini C. Tonsillectomy, appendectomy and

malignant lymphomas. Scand J Haematol. 1982;28:59–64.

231. Bernard SM, Cartwright RA, Bird CC, Gerald Richards ID, Lauder I, Roberts BE. Aetiologic factors in lymphoid malignancies: A case-control epidemiological study. Leuk Res. 1984;8:681–9.

232. Royer RH, Koshiol J, Giambarresi TR, Vasquez LG, Pfeiffer RM, McMaster ML. Differential characteristics of Waldenström macroglobulinemia according to patterns of familial aggregation. Blood. 2010;115:4464–71.

233. Kristinsson SY, Gridley G, Hoover RN, Check D, Landgren O. Long-term risks after splenectomy among 8,149 cancer-free American veterans: A cohort study with up to 27 years follow-up. Haematologica. 2014;99:392–8.

234. Thai LH, Mahévas M, Roudot-Thoraval F, Limal N, Languille L, Dumas G, et al. Long-term complications of splenectomy in adult immune thrombocytopenia. Medicine (Baltimore). 2016;95:1–7.

235. Rodeghiero F, Ruggeri M. Short- and long-term risks of splenectomy for benign haematological disorders: Should we revisit the indications? Br J Haematol. 2012;158:16–29.

236. Buzelé R, Barbier L, Sauvanet A, Fantin B. Medical complications following splenectomy. J Visc Surg. 2016;153:277–86.

237. Weledji EP. Benefits and risks of splenectomy. Int J Surg. 2014;12:113–9.

238. Thomsen RW, Schoonen WM, Farkas DK, Riis A, Jacobsen J, Fryzek JP, et al. Risk for hospital contact with infection in patients with splenectomy. Ann Intern Med. 2009;151:546–56.

239. Thomsen RW, Schoonen WM, Farkas DK, Riis A, Fryzek JP, Sørensen HT. Risk of venous thromboembolism in splenectomized patients compared with the general population and appendectomized patients: A 10-year nationwide cohort study. J Thromb Haemost. 2010;8:1413–6.

240. Schwartz PE, Sterioff S, Mucha P, Melton LJ, Offord KP. Postsplenectomy sepsis and mortality in adults. J Am Med Assoc. 1982;248:2279–83.

241. Bisharat N, Omari H, Lavi I, Raz R. Risk of infection and death among post-splenectomy patients. J Infect. 2001;43:182–6.

242. Sun LM, Chen HJ, Jeng L Bin, Li TC, Wu SC, Kao CH. Splenectomy and increased subsequent cancer risk: A nationwide population-based cohort study. Am J Surg. 2015;210:243–51.

243. Mellemkjaer L, Olsen JH, Linet MS, Gridley G, McLaughlin JK. Cancer risk after splenectomy. Ugeskr Laeger. 1995;157:5097–100.

244. Strachan DP. Hay fever, hygiene, and household size. BMJ. 1989;299:1259–60.

245. Oikonomopoulou K, Brinc D, Kyriacou K, Diamandis EP. Infection and cancer: Revaluation of the hygiene hypothesis. Clin Cancer Res. 2013;19:2834–41.

246. Bach JF. The hygiene hypothesis in autoimmunity: The role of pathogens and commensals. Nat Rev Immunol. 2018;18:105–20.

247. Zuckerman MK, Armelagos GJ. The hygiene hypothesis and the second epidemiologic transition (Ch 16). In: Zuckerman MK, editor. Mod Environ Hum Heal Revising Second Epidemiol transition, First Ed. First Edit. John Wiley & Sons, Inc.; 2014. page 301–20.

248. Marcotte EL, Ritz B, Cockburn M, Yu F, Heck JE. Exposure to infections and risk of leukemia in young children. Cancer Epidemiol Biomarkers Prev. 2014;23:1195–203.

249. Carozza SE, Puumala SE, Chow EJ, Fox EE, Horel S, Johnson KJ, et al. Parental educational attainment as an indicator of socioeconomic status and risk of childhood cancers. Br J Cancer. 2010;103:136–42.

250. House T, Keeling MJ. Household structure and infectious disease transmission. Epidemiol Infect. 2009;137:654–61.

251. Kuper H, Hsieh CC, Stuver SO, Mucci LA, Tzonou A, Zavitsanos X, et al. Birth order, as a proxy for age at infection, in the etiology of hepatocellular carcinoma. Epidemiology. 2000;11:680–3.

252. Smedby KE, Hjalgrim H, Chang ET, Rostgaard K, Glimelius B, Adami HO, et al. Childhood social environment and risk of non-Hodgkin lymphoma in adults. Cancer Res. 2007;67:11074–82.

253. Bracci PM, Dalvi TB, Holly EA. Residential history, family characteristics and non-Hodgkin lymphoma, a population-based case-control study in the San Francisco Bay Area. Cancer Epidemiol Biomarkers Prev. 2006;15:1287–94.

254. Grulich AE, Vajdic CM, Falster MO, Kane E, Smedby KE, Bracci PM, et al. Birth order and risk of non-Hodgkin lymphoma - True association or bias? Am J Epidemiol. 2010;172:621–30.

255. Von Behren J, Spector LG, Mueller BA, Carozza SE, Chow EJ, Fox EE, et al. Birth order and risk of childhood cancer: A pooled analysis from five U.S. states. Int J Cancer. 2012;128:2709–16.

256. Crump C, Sundquist K, Sieh W, Winkleby MA, Sundquist J. Perinatal and family risk factors for non-Hodgkin lymphoma in early life: A Swedish national cohort study. J Natl Cancer Inst. 2012;104:923–30.

257. Cozen W, Cerhan JR, Martinez-Maza O, Ward MH, Linet M, Colt JS, et al. The effect of atopy, childhood crowding, and other immune-related factors on non-Hodgkin lymphoma risk. Cancer Causes Control. 2007;18:821–31.

258. Bevier M, Weires M, Thomsen H, Sundquist J, Hemminki K. Influence of family size and birth order on risk of cancer: A population-based study. BMC Cancer. 2011;11:163–73.

259. Chang ET, Montgomery SM, Richiardi L, Ehlin A, Ekbom A, Lambe M. Number of siblings and risk of Hodgkin's lymphoma. Cancer Epidemiol Biomarkers Prev. 2004;13:1236–43.

260. Westergaard T, Melbye M, Pedersen JB, Frisch M, Olsen JH, Andersen PK. Birth order, sibship size and risk of Hodgkin's disease in children and young adults: A population-based study of 31 million person-years. Int J Cancer. 1997;72:977–81.

261. Crump C, Sundquist K, Sieh W, Winkleby MA, Sundquist J. Perinatal and family risk factors for Hodgkin lymphoma in childhood through young adulthood. Am J Epidemiol. 2012;176:1147–58.

262. Bernard SM, Cartwright RA, Darwin CM, Richards I, Roberts B, O'Brien C, et al. Hodgkin's disease: Case-control epidemiological study in Yorkshire. Br J Cancer. 1987;55:85–90.

263. Ma X, Buffler PA, Selvin S, Matthay KK, Wiencke JK, Wiemels JL, et al. Daycare attendance and risk of childhood acute lymphoblastic leukaemia. Br J Cancer. 2002;86:1419–24.

264. Oliveira PD de, Bertoldi AD, da Silva BGC, Rodrigues Domingues M, Neumann NA, da Silveira MF. Day care attendance during the first 12 months of life and occurrence of infectious morbidities and symptoms. J Pediatr (Rio J). 2018;95:657–66.

265. Schuez-Havupalo L, Toivonen L, Karppinen S, Kaljonen A, Peltola V. Daycare attendance and respiratory tract infections: A prospective birth cohort study. BMJ Open. 2017;7:1–8.

266. Collet JP, Ducruet T, Floret D, Cogan-Collet J, Honneger D, Boissel JP. Daycare attendance and risk of first infectious disease. Eur J Pediatr. 1991;150:214–6.

267. Gilham C, Peto J, Simpson J, Roman E, Eden TOB, Greaves MF, et al. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: Findings from UK case-control study. Br Med J. 2005;330:1294–7.

268. Baris D, Brown LM, Silverman DT, Hayes R, Hoover RN, Swanson GM, et al. Socioeconomic status and multiple myeloma among US Blacks and Whites. Am J Public Health. 2000;90:1277–81.

269. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Smith GD. Indicators of socioeconomic position (part 1). J Epidemiol Community Health. 2006;60:7–12.

270. Zill N. Parental schooling and children's health. Public Health Rep. 1996;111:34–43.

271. Rios L, Vásquez L, Oscanoa M, Maza I, Gerónimo J. Advancing parental age and risk of solid tumors in children: A case-control study in Peru. J Oncol. 2018;2018:1–9.

272. Glaser SL, Clarke CA, Nugent RA, Stearns CB, Dorfman RF. Social class and risk of Hodgkin's disease in young-adult women in 1988-94. Int J Cancer. 2002;98:110–7.

273. Marcotte EL, Ritz B, Cockburn M, Clarke CA, Heck JE. Birth characteristics and risk of lymphoma in young children. Cancer Epidemiol. 2014;38:48–55.

274. Gutensohn NM, Shapiro DS. Social class risk factors among children with Hodgkin's disease. Int J Cancer. 1982;30:433–5.

275. Johnston JM, Grufferman S, Bourguet CC, Delzell E, Delong ER, Cohen HJ. Socioeconomic status and risk of multiple myeloma. J Epidemiol Community Health. 1985;39:175–8.

276. Mao Y1, Hu J, Ugnat AM WK. Non-Hodgkin's lymphoma and occupational exposure to chemicals in Canada. Canadian Cancer Registries Epidemiology Research Group. Ann

Oncol. 2000;11:69–73.

277. Pearce N, Bethwaite P. Increasing incidence of non-Hodgkin's lymphoma: Occupational and environmental factors. Cancer Res. 1992;52:5496–501.

278. Jones RR, Yu C-L, Nuckols JR, Cerhan JR, Airola M, Ross JA, et al. Farm residence and lymphohematopoietic cancers in the Iowa Women's Health Study. Environ Res. 2014;133:353–61.

279. Hofmann JN, Hoppin JA, Lynch CF, Poole JA, Purdue MP, Blair A, et al. Farm characteristics, allergy symptoms, and risk of non-Hodgkin lymphoid neoplasms in the agricultural health study. Cancer Epidemiol Biomarkers Prev. 2015;24:587–94.

280. 'T Mannetje A, Eng A, Pearce N. Farming, growing up on a farm, and haematological cancer mortality. Occup Environ Med. 2012;69:126–32.

281. Tranah GJ, Bracci PM, Holly EA. Domestic and farm-animal exposures and risk of non-Hodgkin's lymphoma in a population-based study in the San Francisco Bay Area. Cancer Epidemiol Biomarkers Prev. 2008;17:2382–7.

282. Pahwa P, McDuffie HH, Dosman JA, Robson D, McLaughlin JR, Spinelli JJ, et al. Exposure to animals and selected risk factors among Canadian farm residents with Hodgkin's disease, multiple myeloma, or soft tissue sarcoma. J Occup Environ Med. 2003;45:857–68.

283. Hardell L, Eriksson M, Degerman A. Exposure to phenoxyacetic acids, chiorophenols, or organic solvents in relation to histopathology, stage, and anatomical localization of non-Hodgkin's lymphoma. Cancer Res. 1994;54:2386–9.

284. Cocco P, Satta G, D'Andrea I, Nonne T, Udas G, Zucca M, et al. Lymphoma risk in livestock farmers: Results of the Epilymph study. Int J Cancer. 2013;132:2613–8.

285. Carozza SE, Li B, Wang Q, Horel S, Cooper S. Agricultural pesticides and risk of childhood cancers. Int J Hyg Environ Health. 2009;212:186–95.

286. Hidayat K, Li HJ, Shi BM. Anthropometric factors and non-Hodgkin's lymphoma risk: Systematic review and meta-analysis of prospective studies. Crit Rev Oncol Hematol. 2018;129:113–23.

287. Bertrand KA, Giovannucci E, Zhang SM, Laden F, Rosner B, Birmann BM. A prospective analysis of body size during childhood, adolescence, and adulthood and risk of non-Hodgkin lymphoma. Cancer Prev Res. 2013;6:864–73.

288. Maskarinec G, Erber E, Gill J, Cozen W, Kolonel LN. Overweight and obesity at different times in life as risk factors for non-Hodgkin's lymphoma: The multiethnic cohort. Cancer Epidemiol Biomarkers Prev. 2008;17:196–203.

289. Brown LM, Gridley G, Pottern LM, Baris D, Swanson CA, Silverman DT, et al. Diet and nutrition as risk factors for multiple myeloma among blacks and whites in the United States. Cancer Causes Control. 2001;12:117–25.

290. Cerhan JR, Bernstein L, Severson RK, Davis S, Colt JS, Blair A, et al. Anthropometrics, physical activity, related medical conditions, and the risk of non-Hodgkin lymphoma. Cancer Causes Control. 2005;16:1203–14.

291. Pan SY, Mao Y, Ugnat AM, Canadian Cancer Registeries Epidemiology Research Group. Physical activity, obesity, energy intake, and the risk of non-Hodgkin's lymphoma: A population-based case-control study. Am J Epidemiol. 2005;162:1162–73.

292. Skibola CF. Obesity, diet and risk of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev. 2007;16:392–5.

293. Li Q, Chang ET, Bassig BA, Dai M, Qin Q, Gao Y, et al. Body size and risk of Hodgkin's lymphoma by age and gender: A population-based case-control study in Connecticut and Massachusetts. Cancer Causes Control. 2013;24:287–95.

294. Zhang S, Hunter DJ, Rosner BA, Colditz GA, Fuchs CS, Speizer FE, et al. Dietary fat and protein in relation to risk of non-Hodgkin's lymphoma among women. J Natl Cancer Inst. 1999;91:1751–8.

295. Hermann S, Rohrmann S, Linseisen J, Nieters A, Khan A, Gallo V, et al. Level of education and the risk of lymphoma in the European prospective investigation into cancer and nutrition. J Cancer Res Clin Oncol. 2010;136:71–7.

296. Hemminki K, Li X. University and medical education and the risk of cancer in Sweden. Eur J Cancer Prev. 2004;13:199–205.

297. Krull KR, Sabin ND, Reddick WE, Zhu L, Armstrong GT, Green DM, et al. Neurocognitive function and CNS integrity in adult survivors of childhood Hodgkin lymphoma. J Clin Oncol. 2012;30:3618–24.

298. Wefel JS, Lenzi R, Theriault RL, Davis RN, Meyers CA. The cognitive sequelae of standard-dose adjuvant chemotherapy in women with breast carcinoma: Results of a prospective, randomized, longitudinal trial. Cancer. 2004;100:2292–9.

299. Jansen CE, Dodd MJ, Miaskowski CA, Dowling GA, Kramer J. Preliminary results of a longitudinal study of changes in cognitive function in breast cancer patients undergoing chemotherapy with doxorubicin and cyclophosphamide. Psychooncology. 2008;17:1189–95.

300. Mandelblatt JS, Hurria A, Mcdonald BC, Saykin AJ, Stern RA, VanMeter JW, et al. Cognitive effects of cancer and its treatments at the intersection of aging: What do we know; What do we need to know? Semin Oncol. 2013;40:1–26.

301. Freilich RJ, Delattre J-Y, Monjour A, DeAngelis LM. Chemotherapy without radiation therapy as initial treatment for primary CNS lymphoma in older patients. Neurology. 1996;42:435–9.

302. Fliessbach K, Helmstaedter C, Urbach H, Althaus A, Pels H, Linnebank M, et al. Neuropsychological outcome after chemotherapy for primary CNS lymphoma: A prospective study. Neurology. 2005;64:1184–8.

303. Morton LM, Hartge P, Holford TR, Holly EA, Chiu BCH, Vineis P, et al. Cigarette smoking and risk of non-Hodgkin lymphoma: A pooled analysis from the International Lymphoma Epidemiology Consortium (InterLymph). Cancer Epidemiol Biomarkers Prev. 2005;14:925–33.

304. Herrinton LJ, Friedman GD. Cigarette smoking and risk of non-Hodgkin's lymphoma

subtypes. Cancer Epidemiol Biomarkers Prev. 1998;7:25–8.

305. Briggs NC, Levine RS, Bobo LD, Haliburton WP, Brann EA, Hennekens CH. Wine drinking and risk of non-Hodgkin's lymphoma among men in the United States: A population-based case-control study. Am J Epidemiol. 2002;156:454–62.

306. Monnereau A, Orsi L, Troussard X, Berthou C, Fenaux P, Soubeyran P, et al. Cigarette smoking, alcohol drinking, and risk of lymphoid neoplasms: Results of a French case-control study. Cancer Causes Control. 2008;19:1147–60.

307. Maggioncalda A, Malik N, Shenoy P, Smith M, Sinha R, Flowers CR. Clinical, molecular, and environmental risk factors for Hodgkin lymphoma. Adv Hematol. 2011;2011:1–10.

308. Gorini G, Stagnaro E, Fontana V, Miligi L, Ramazzotti V, Amadori D, et al. Alcohol consumption and risk of Hodgkin's lymphoma and multiple myeloma: A multicentre case-control study. Ann Oncol. 2007;18:143–8.

309. Rota M, Porta L, Pelucchi C, Negri E, Bagnardi V, Bellocco R, et al. Alcohol drinking and risk of leukemia - A systematic review and meta-analysis of the dose-risk relation. Cancer Epidemiol. 2014;38:339–45.

310. Epstein MM, Chang ET, Zhang Y, Fung TT, Batista JL, Ambinder RF, et al. Dietary pattern and risk of Hodgkin lymphoma in a population-based case-control study. Am J Epidemiol. 2015;182:405–16.

311. Purdue MP, Bassani DG, Klar NS, Sloan M, Kreiger N, Paulse B, et al. Dietary factors and risk of non-Hodgkin lymphoma by histologic subtype: A case-control analysis. Cancer Epidemiol Biomarkers Prev. 2004;13:1665–76.

312. Tavani A, Pregnolato A, Negri E, Franceschi S, Serraino D, Carbone A, et al. Diet and risk of lymphoid neoplasms and soft tissue sarcomas. Nutr Cancer. 1997;27:256–60.

313. Chang ET, Smedby KE, Zhang SM, Hjalgrim H, Melbye M, Öst Å, et al. Dietary factors and risk of non-Hodgkin lymphoma in men and women. Cancer Epidemiol Biomarkers Prev. 2005;14:512–20.

314. Rohrmann S, Linseisen J, Jakobsen MU, Overvad K, Raaschou-Nielsen O, Tjonneland A, et al. Consumption of meat and dairy and lymphoma risk in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2011;128:623–34.

315. Solans M, Castelló A, Benavente Y, Marcos-Gragera R, Amiano P, Gracia-Lavedan E, et al. Adherence to the Western, Prudent, and Mediterranean dietary patterns and chronic lymphocytic leukemia in the MCC-Spain study. Haematologica. 2018;103:1881–8.

316. Matsuo K, Hamajima N, Hirose K, Inoue M, Takezaki T, Kuroishi T, et al. Alcohol, smoking, and dietary status and susceptibility to malignant lymphoma in Japan: Results of a hospital-based case-control study at Aichi Cancer Center. Japanese J Cancer Res. 2001;92:1011–7.

317. Tavani A, La Vecchia C, Gallus S, Lagiou P, Trichopoulos D, Levi F, et al. Red meat intake and cancer risk: A study in Italy. Int J Cancer. 2000;86:425–8.

318. Talamini R, Polesel J, Montella M, Dal Maso L, Crovatto M, Crispo A, et al. Food groups and risk of non-Hodgkin lymphoma: A multicenter, case-control study in Italy. Int J

Cancer. 2006;118:2871–6.

319. Zheng T, Holford TR, Leaderer B, Zhang Y, Zahm SH, Flynn S, et al. Diet and nutrient intakes and risk of non-Hodgkin's lymphoma in Connecticut women. Am J Epidemiol. 2004;159:454–66.

320. Bertrand KA, Giovannucci E, Rosner BA, Zhang SM, Laden F, Birmann BM. Dietary fat intake and risk of non-Hodgkin lymphoma in 2 large prospective cohorts. Am J Clin Nutr. 2017;106:650–6.

321. Gao Y, Li Q, Bassig BA, Chang ET, Dai M, Qin Q, et al. Subtype of dietary fat in relation to risk of Hodgkin lymphoma: A population-based case-control study in Connecticut and Massachusetts. Cancer Causes Control. 2013;24:485–94.

322. Kelemen LE, Cerhan JR, Lim U, Davis S, Cozen W, Schenk M, et al. Vegetables, fruit, and antioxidant-related nutrients and risk of non-Hodgkin lymphoma: A National Cancer Institute-Surveillance, Epidemiology, and End Results population-based case-control study. Am J Clin Nutr. 2006;83:1401–10.

323. Parodi S, Merlo FD, Stagnaro E, Working Group for the Epidemiology of Hematolymphopoietic Malignancies in Italy. Coffee consumption and risk of non-Hodgkin's lymphoma: Evidence from the Italian multicentre case–control study. Cancer Causes Control. 2017;28:867–76.

324. Ugai T, Matsuo K, Sawada N, Iwasaki M, Yamaji T, Shimazu T, et al. Coffee and green tea consumption and subsequent risk of malignant lymphoma and multiple myeloma in Japan: The Japan Public Health Center-based Prospective Study. Cancer Epidemiol Biomarkers Prev. 2017;26:1352–6.

325. Lee JS, Bracci PM, Holly EA. Non-Hodgkin lymphoma in women: Reproductive factors and exogenous hormone use. Am J Epidemiol. 2008;168:278–88.

326. Zhang Y, Holford TR, Leaderer B, Boyle P, Zahm SH, Zhang B, et al. Menstrual and reproductive factors and risk of non-Hodgkin's lymphoma among Connecticut women. Am J Epidemiol. 2004;160:766–73.

327. Nelson RA, Levine AM, Bernstein L. Reproductive factors and risk of intermediate- or high-grade B-cell non-Hodgkin's lymphoma in women. J Clin Oncol. 2001;19:1381–7.

328. Zhang Y, Holford TR, Leaderer B, Boyle P, Zahm SH, Flynn S, et al. Hair-coloring product use and risk of non-Hodgkin's lymphoma: A population-based case-control study in Connecticut. Am J Epidemiol. 2004;159:148–54.

329. Rauscher GH, Shore D, Sandler DP. Hair dye use and risk of adult acute leukemia. Am J Epidemiol. 2004;160:19–25.

330. Zhang Y, de Sanjose S, Bracci PM, Morton LM, Wang R, Brennan P, et al. Personal use of hair dye and the risk of certain subtypes of non-Hodgkin lymphoma. Am J Epidemiol. 2008;167:1321–31.

331. Zahm SHSH, Weisenburger DD, Babbitt PA, Saal RC, Vaught JB, Blair A. Use of hair coloring products and the risk of lymphoma, multiple myeloma, and chronic lymphocytic leukemia. Am J Public Health. 1992;82:990–7.

332. Bolt HM, Golka K. The debate on carcinogenicity of permanent hair dyes: New insights. Crit Rev Toxicol. 2007;37:521–36.

333. Schenk M, Purdue MP, Colt JS, Hartge P, Blair A, Stewart P, et al. Occupation/industry and risk of non-Hodgkin's lymphoma in the United States. Occup Environ Med. 2009;66:23–31.

334. Zheng T, Blair A, Zhang Y, Weisenburger DD, Zahm SH. Occupation and risk of non-Hodgkin's lymphoma and chronic lymphocytic leukemia. J Occup Environ Med. 2002;44:469–74.

335. Baris D, Dt S, Lm B, Gm S, Rb H, Ag S, et al. Occupation, pesticide exposure and risk of multiple myeloma. Scand J Work Environ Heal. 2004;30:215–22.

336. Hall NEL, Rosenman KD. Cancer by industry: Analysis of a population-based cancer registry with an emphasis on blue-collar workers. Am J Ind Med. 1991;19:145–59.

337. Ji J, Hemminki K. Socioeconomic/Occupational risk factors for lymphoproliferative diseases in Sweden. Ann Epidemiol. 2006;16:370–6.

338. Svec MA, Ward MH, Dosemeci M, Checkoway H, De Roos AJ. Risk of lymphatic or haematopoietic cancer mortality with occupational exposure to animals or the public. Occup Environ Med. 2005;62:726–35.

339. Hartge P, Smith MT. Environmental and behavioral factors and the risk of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev. 2007;16:367–8.

340. Boffetta P, van der Hel O, Kricker A, Nieters A, de Sanjosé S, Maynadié M, et al. Exposure to ultraviolet radiation and risk of malignant lymphoma and multiple myeloma - A multicentre European case-control study. Int J Epidemiol. 2008;37:1080–94.

341. Karunanayake CP, Singh G V., Spinelli JJ, McLaughlin JR, Dosman JA, McDuffie HH, et al. Occupational exposures and Hodgkin lymphoma: Canadian case-control study. J Occup Environ Med. 2009;51:1447–54.

342. Linet MS, Schubauer-Berigan MK, Weisenburger DD, Richardson DB, Landgren O, Blair A, et al. Chronic lymphocytic leukaemia: An overview of aetiology in light of recent developments in classification and pathogenesis. Br J Haematol. 2007;139:672–86.

343. Smedby KE, Hjalgrim H, Melbye M, Torrång A, Rostgaard K, Munksgaard L, et al. Ultraviolet radiation exposure and risk of malignant lymphomas. J Natl Cancer Inst. 2005;97:199–209.

344. Lichtenstein P, Holm NVN V, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer. N Engl J Med. 2000;343:78–85.

345. Albright F, Teerlink C, Werner TL, Cannon-Albright LA. Significant evidence for a heritable contribution to cancer predisposition: A review of cancer familiality by site. BMC Cancer. 2012;12:1471–2407.

346. Mack TM, Cozen W, Shibata DK, Weiss LM, Nathwani BN, Hernandez AM, et al. Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. N Engl J Med. 1995;332:788–92.

347. Goldin LR, Björkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. Haematologica. 2009;94:647–53.

348. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Highly increased familial risks for specific lymphoma subtypes. Br J Haematol. 2009;146:91–4.

349. Kristinsson SY, Björkholm M, Goldin LR, McMaster ML, Turesson I, Landgren O. Risk of lymphoproliferative disorders among first-degree relatives of lymphoplasmacytic lymphoma/Waldenstrom macroglobulinemia patients: A population-based study in Sweden. Blood. 2008;112:3052–6.

350. Friedman DL, Kadan-Lottick NS, Whitton J, Mertens AC, Yasui Y, Liu Y, et al. Increased risk of cancer among siblings of long-term childhood cancer survivors: A report from the childhood cancer survivor study. Cancer Epidemiol Biomarkers Prev. 2005;14:1922–7.

351. Goldin LR, Pfeiffer RM, Gridley G, Gail MH, Li X, Mellemkjaer L, et al. Familial aggregation of Hodgkin lymphoma and related tumors. Cancer. 2004;100:1902–8.

352. Chang ET, Smedby KE, Hjalgrim H, Porwit-MacDonald A, Roos G, Glimelius B, et al. Family history of hematopoietic malignancy and risk of lymphoma. J Natl Cancer Inst. 2005;97:1466–74.

353. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009;41:873–5.

354. Kleinstern G, Yan H, Hildebrandt MAT, Vijai J, Berndt SI, Ghesquières H, et al. Inherited variants at 3q13.33 and 3p24.1 are associated with risk of diffuse large B-cell lymphoma and implicate immune pathways. Hum Mol Genet. 2019;ddz228:1–10.

355. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet. 2011;43:830–7.

356. Tan DEK, Foo JN, Bei J-X, Chang J, Peng R, Zheng X, et al. Genome-wide association study of B-cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. Nat Genet. 2013;45:804–7.

357. Kumar V, Matsuo K, Takahashi A, Hosono N, Tsunoda T, Kamatani N, et al. Common variants on 14q32 and 13q12 are associated with DLBCL susceptibility. J Hum Genet. 2011;56:436–9.

358. Skibola CF, Conde L, Foo JN, Riby J, Humphreys K, Sillé FCM, et al. A meta-analysis of genome-wide association studies of follicular lymphoma. BMC Genomics. 2012;13:1–6.

359. Skibola CF, Berndt SI, Vijai J, Conde L, Wang Z, Yeager M, et al. Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. Am J Hum Genet. 2014;95:462–71.

360. Conde L, Halperin E, Akers NK, Brown KM, Smedby KE, Rothman N, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. Nat Genet. 2010;42:661–4.

361. Smedby KE, Foo JN, Skibola CF, Darabi H, Conde L, Hjalgrim H, 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;7:1–11.

362. Vijai J, Wang Z, Berndt SI, Skibola CF, Slager SL, de Sanjose S, et al. A genome-wide association study of marginal zone lymphoma shows association to the HLA region. Nat Commun. 2015;6:1–7.

363. Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AHM, et al. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet. 2007;80:867–75.

364. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJSS, Holroyd A, Wang Y, et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. Nat Publ Gr. 2014;46:56–60.

365. Slager SL, Skibola CF, Di Bernardo MC, Conde L, Broderick P, McDonnell SK, et al. Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. Blood. 2012;120:843–6.

366. Slager SL, Rabe KG, Achenbach SJ, Vachon CM, Goldin LR, Strom SS, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. Blood. 2011;117:1911–7.

367. Frampton M, Inacio M, Broderick P, Thomsen H, Forsti A, Vijayakrishnan J, et al. Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. Nat Commun. 2013;4:1–7.

368. Sud A, Thomsen H, Law PJ, Försti A, da Silva Filho MI, Holroyd A, et al. Genome-wide association study of classical Hodgkin lymphoma identifies key regulators of disease susceptibility. Nat Commun. 2017;8:1–11.

369. Sud A, Thomsen H, Orlando G, Försti A, Law PJ, Broderick P, et al. Genome-wide association study implicates immune dysfunction in the development of Hodgkin lymphoma. Blood. 2018;132:2040–52.

370. Enciso-Mora V, Broderick P, Ma Y, Jarrett RF, Hjalgrim H, Hemminki K, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet. 2010;42:1126–30.

371. Broderick P, Chubb D, Johnson DC, Weinhold N, Försti A, Lloyd A, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. Nat Genet. 2016;44:58–61.

372. Thomsen H, Chattopadhyay S, Weinhold N, Vodicka P, Vodickova L, Hoffmann P, et al. Genome-wide association study of monoclonal gammopathy of unknown significance (MGUS): Comparison with multiple myeloma. Leukemia. 2019;33:1817–21.

373. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet. 2010;42:132–6.

374. Diepstra A, Niens M, Meerman G, Poppema S, van den Berg A. Genetic susceptibility to

Hodgkin's lymphoma associated with the human leukocyte antigen region. Eur J Haematol. 2005;75:34–41.

375. Kushekhar K, Van Den Berg A, Nolte I, Hepkema B, Visser L, Diepstra A. Genetic associations in classical Hodgkin lymphoma: A systematic review and insights into susceptibility mechanisms. Cancer Epidemiol Biomarkers Prev. 2014;23:2737–47.

376. Sellick GS, Wild RW, Houlston RS, Goldin LR, Caporaso N, Slager SL, et al. A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia. Blood. 2007;110:3326–33.

377. McMaster ML, Goldin LR, Bai Y, Ter-Minassian M, Boehringer S, Giambarresi TR, et al. Genomewide linkage screen for Waldenström macroglobulinemia susceptibility loci in high-risk families. Am J Hum Genet. 2006;79:695–701.

378. Goldin LR, Ishibe N, Sgambati M, Marti GE, Fontaine L, Lee MP, et al. A genome scan of 18 families with chronic lymphocytic leukaemia. Br J Haematol. 2003;121:866–73.

379. Goldin LR, Slager SL. Familial CLL: Genes and environment. Hematology. 2007;2007:339–45.

380. Goldin LR, McMaster ML, Ter-Minassian M, Saddlemire S, Harmsen B, Lalonde G, et al. A genome screen of families at high risk for Hodgkin lymphoma: Evidence for a susceptibility gene on chromosome 4. J Med Genet. 2005;42:595–601.

381. Rendleman J, Antipin Y, Reva B, Adaniel C, Przybylo JA, Dutra-Clarke A, et al. Genetic variation in DNA repair pathways and risk of non-Hodgkin's lymphoma. PLoS One. 2014;9:e101685.

382. Skibola CF, Bracci PM, Nieters A, Brooks-Wilson A, De Sanjosé S, Hughes AM, et al. Tumor necrosis factor (TNF) and lymphotoxin-α (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph consortium. Am J Epidemiol. 2010;171:267–76.

383. Fernberg P, Chang ET, Duvefelt K, Hjalgrim H, Eloranta S, Sørensen KM, et al. Genetic variation in chromosomal translocation breakpoint and immune function genes and risk of non-Hodgkin lymphoma. Cancer Causes Control. 2010;21:759–69.

384. Wang SS, Purdue MP, Cerhan JR, Zheng T, Menashe I, Bruce K, et al. Common gene variants in the tumor necrosis factor (TNF) and TNF receptor superfamilies and NF-kB transcription factors and non-Hodgkin lymphoma risk. PLoS One. 2009;4:1–9.

385. Smedby KE, Hjalgrim H. Epidemiology and etiology of mantle cell lymphoma and other non-Hodgkin lymphoma subtypes. Semin Cancer Biol. 2011;21:293–8.

386. Skibola CF, Curry JD, Nieters A. Genetic susceptibility to lymphoma. Haematologica. 2007;92:960–9.

387. Ekstrom Smedby K, Lindgren CM, Hjalgrim H, Humphreys K, Schollkopf C, Chang ET, et al. Variation in DNA repair genes ERCC2, XRCC1, and XRCC3 and risk of follicular lymphoma. Cancer Epidemiol Biomarkers Prev. 2006;15:258–65.

388. Leeksma OC, De Miranda NF, Veelken H. Germline mutations predisposing to diffuse large B-cell lymphoma. Blood Cancer J. 2017;7:1–9.

389. Siddiqui R, Onel K, Facio F, Offit K. The genetics of familial lymphomas. Curr Oncol Rep. 2004;6:380–7.

390. Ehrlich M. DNA methylation in cancer: Too much, but also too little. Oncogene. 2002;21:5400–13.

391. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet. 1995;10:111–3.

392. Rothman N, Skibola CF, Wang SS, Morgan G, Lan Q, Smith MT, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: A report from the InterLymph Consortium. Lancet Oncol. 2005;7:27–38.

393. Lan Q, Zheng T, Rothman N, Zhang Y, Wang SS, Shen M, et al. Cytokine polymorphisms in the Th1/Th2 pathway and susceptibility to non-Hodgkin lymphoma. Blood. 2006;107:4101–8.

394. Wang SS, Cerhan JR, Hartge P, Davis S, Cozen W, Severson RK, et al. Common genetic variants in proinflammatory and other immunoregulatory genes and risk for non-Hodgkin lymphoma. Cancer Res. 2006;66:9771–81.

395. Nieters A, Beckmann L, Deeg E, Becker N. Gene polymorphisms in Toll-like receptors, interleukin-10, and interleukin-10 receptor alpha and lymphoma risk. Genes Immun. 2006;7:615–24.

396. Rosenstiel P, Hellmig S, Hampe J, Ott S, Till A, Fischbach W, et al. Influence of polymorphisms in the NOD1/CARD4 and NOD2/CARD15 genes on the clinical outcome of Helicobacter pylori infection. Cell Microbiol. 2006;8:1188–98.

397. Forrest MS, Christine F, Lightfoot TJ, Paige M, Willett E V, Martyn T, et al. Polymorphisms in innate immunity genes and risk of non-Hodgkin lymphoma. Br J Haematol. 2006;134:180–3.

398. Wang SS, Scott F, Cerhan JR, Hartge P, Severson RK, Cozen W, et al. Polymorphisms in oxidative stress genes and risk for non-Hodgkin lymphoma. Carcinogenesis. 2006;27:1828–34.

399. Mcaulay KA, Jarrett RF. Human leukocyte antigens and genetic susceptibility to lymphoma. Tissue Antigens. 2015;86:98–113.

400. Shammas MA. Telomeres, lifestyle, cancer and aging. Curr Opin Clin Nutr Metab Care. 2012;14:28–34.

401. Noy A. Telomeres: The long and short of developing non-Hodgkin lymphoma. Clin Cancer Res. 2009;15:7114–5.

402. Blackburn NB, Charlesworth JC, Marthick JR, Tegg EM, Marsden K a, Srikanth V, et al. A retrospective examination of mean relative telomere length in the Tasmanian Familial Hematological Malignancies Study. Oncol Rep. 2015;33:25–32.

403. Lan Q, Cawthon R, Shen M, Weinstein SJ, Virtamo J, Lim U, et al. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of non-Hodgkin lymphoma. Clin Cancer Res. 2009;15:7429–33.

404. Turner JJ, Morton LM, Linet MS, Clarke CA, Kadin ME, Vajdic CM, et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): Update and future directions. Blood. 2010;116:90–9.

405. North American Association of Central Cancer Registries . About NAACCR [Internet]. North Am. Assoc. Cent. Cancer Regist. Inc. (NAACCR, Inc.). 2019. Available from: https://www.naaccr.org/about-naaccr/

406. North American Association of Central Cancer Registries . CiNA public use data set [Internet]. North Am. Assoc. Cent. Cancer Regist. 2020. Available from: https://www.naaccr.org/cina-public-use-data-set/

407. Surveillance, Epidemiology, and End Results (SEER) Program . Overview of the SEER Program [Internet]. Surveillance, Epidemiol. End Results Progr. 2019. Available from: https://seer.cancer.gov/about/overview.html

408. National Cancer Institute . Cancer Stat Facts: Leukemia — Acute Lymphocytic Leukemia (ALL) [Internet]. Surveillance, Epidemiol. End Results Progr. 2019. Available from: https://seer.cancer.gov/statfacts/html/alyl.html

409. Khalil MO, Morton LM, Devesa SS, Check DP, Curtis RE, Weisenburger DD, et al. Incidence of marginal zone lymphoma in the United States 2001 – 2009 with a focus on primary anatomic site. Br J Haematol. 2014;165:67–77.

410. Smith A, Crouch S, Lax S, Li J, Painter D, Howell D, et al. Lymphoma incidence, survival and prevalence 2004-2014: Sub-type analyses from the UK's Haematological Malignancy Research Network. Br J Cancer. 2015;112:1575–84.

411. Li Y, Wang Y, Wang Z, Yi D, Ma S. Racial differences in three major NHL subtypes: Descriptive epidemiology. Cancer Epidemiol. 2015;39:8–13.

412. Flowers CR, Shenoy PJ, Borate U, Bumpers K, Douglas-Holland T, King N, et al. Examining racial differences in diffuse large B-cell lymphoma presentation and survival. Leuk Lymphoma. 2013;54:268–76.

413. Frebourg T, Barbier N, Yan Y, Garber JE, Dreyfus M, Fraumeni J, et al. Germ-line p53 mutations in 15 families with Li-Fraumeni syndrome. Am J Hum Genet. 1995;56:608–15.

414. Mellemkjaer L, Pfeiffer RM, Engels EA, Gridley G, Wheeler W, Hemminki K, et al. Autoimmune disease in individuals and close family members and susceptibility to non-Hodgkin's lymphoma. Arthritis Rheum. 2008;58:657–66.

415. DNA Genotek. Oragene Discover OGR-500. Ottawa, ON; 2017.

416. Iwasiow RM, Desbois A, Birnboim HC. Long-term stability of DNA from saliva samples stored in the Oragene self-collection kit. DNA Genotek. Ottawa, ON, Canada; 2019.

417. Genotek D. Laboratory protocol for manual purification of DNA from whole sample. DNA Genotek. Ottawa, ON: DNA Genotek; 2018. page 1–8.

418. Qiagen. AllPrep DNA/RNA FFPE handbook. Germantown, MD; 2017.

419. Grufferman S, Barton J, Eby N. Increased sex concordance of sibling pairs with Behcet's disease, Hodgkin's disease, multiple sclerosis, and sarcoidosis. Am J Epidemiol.

1987;126:365–9.

420. Lan Q, Zheng T, Chanock S, Zhang Y, Shen M, Wang SS, et al. Genetic variants in caspase genes and susceptibility to non-Hodgkin lymphoma. Carcinogenesis. 2007;28:823–7.

421. Chatterjee N, Hartge P, Cerhan JR, Cozen W, Davis S, Ishibe N, et al. Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. Cancer Epidemiol Biomarkers Prev. 2004;13:1415–22.

422. Goldin LR, Pfeiffer RM, Li X, Hemminki K. Familial risk of lymphoproliferative tumors in families of patients with chronic lymphocytic leukemia: Results from the Swedish Family-Cancer Database. Blood. 2004;104:1850–5.

423. Goldin LR, Landgren O, McMaster ML, Gridley G, Hemminki K, Li X, et al. Familial aggregation and heterogeneity of non-Hodgkin lymphoma in population-based samples. Cancer Epidemiol Biomarkers Prev. 2005;14:2402–6.

424. Altieri A, Chen B, Bermejo JL, Castro F, Hemminki K. Familial risks and temporal incidence trends of multiple myeloma. Eur J Cancer. 2006;42:1661–70.

425. Fallah M, Kharazmi E, Pukkala E, Tretli S, Olsen JH, Tryggvadottir L, et al. Familial risk of non-Hodgkin lymphoma by sex, relationship, age at diagnosis and histology: A joint study from five Nordic countries. Leukemia. 2016;30:373–8.

426. Landgren O, Kristinsson SY, Goldin LR, Caporaso NE, Blimark C, Mellqvist U, et al. Risk of plasma cell and lymphoproliferative disorders among 14 621 first-degree relatives of 4458 patients with monoclonal gammopathy of undetermined significance in Sweden. Blood. 2009;114:791–6.

427. Paltiel O, Schmit T, Adler B, Rachmilevitz EA, Polliack A, Cohen A, et al. The incidence of lymphoma in first-degree relatives of patients with Hodgkin disease and non-Hodgkin lymphoma: Results and limitations of a registry-linked study. Cancer. 2000;88:2357–66.

428. Alexandrescu DT, Garino A, Brown-Balem KA, Wiernik PH. Anticipation in families with Hodgkin's and non-Hodgkin's lymphoma in their pedigree. Leuk Lymphoma. 2006;47:2115–27.

429. Kharazmi E, Fallah M, Pukkala E, Olsen JHJH, Tryggvadottir L, Sundquist K, et al. Risk of familial classical Hodgkin lymphoma by relationship, histology, age, and sex: A joint study from five Nordic countries. Blood. 2015;126:1990–5.

430. Lindqvist EK, Goldin LR, Landgren O, Blimark C, Mellqvist U, Turesson I, et al. Personal and family history of immune-related conditions increase the risk of plasma cell disorders: A population-based study. Blood. 2011;118:6284–92.

431. Wiernik PH, Wickramasinghe D, Dutcher JP. Families with both Hodgkin lymphoma and multiple myeloma in their pedigrees. Clin Adv Hematol Oncol. 2015;13:257–60.

432. Tegg EM, Thomson RJ, Stankovich J, Banks A, Flowers C, McWhirter R, et al. Evidence for a common genetic aetiology in high-risk families with multiple haematological malignancy subtypes. Br J Haematol. 2010;150:456–62.

433. Salipante SJ, Mealiffe ME, Wechsler J, Krem MM, Liu Y, Namkoong S, et al. Mutations in

a gene encoding a midbody kelch protein in familial and sporadic classical Hodgkin lymphoma lead to binucleated cells. Proc Natl Acad Sci. 2009;106:14920–5.

434. Rotunno M, McMaster ML, Boland J, Bass S, Zhang X, Burdette L, et al. Whole exome sequencing in families at high risk for Hodgkin lymphoma: Identification of a predisposing mutation in the KDR gene. Haematologica. 2016;101:853–60.

435. Stepp SE, Dufourcq-Lagelouse R, Le Deist F, Bhawan S, Certain S, Mathew PA, et al. Perforin gene defects in familial hemophagocytic lymphohistiocytosis. Science (80-). 1999;286:1957–60.

436. Clementi R, Emmi L, Maccario R, Liotta F, Moretta L, Danesino C, et al. Adult onset and atypical presentation of hemophagocytic lymphohistocytosis in siblings carrying PRF1 mutations. Blood. 2002;100:2266–8.

437. Clementi R, Locatelli F, Garaventa A, Emmi L, Bregni M, Cefalo G, et al. A proportion of patients with lymphoma may harbor mutations of the perforin gene. Blood. 2005;105:4424–9.

438. Clementi R, Dagna L, Dianzani U, Cupre L, Dianzani I, Ponzoni M, et al. Inherited perforin and Fas mutations in a patient with autoimmune lymphoproliferative syndrome and lymphoma. N Engl J Med. 2004;351:1419–24.

439. SEER. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2014 Sub (1973-2012) - National Cancer Institute, Surveillance Re. 2014.

440. SEER. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 13 Regs Research Data, Nov 2014 Sub (1992-2012) - National Cancer Institute, Surveillance Research Program. 2014.

441. SEER. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2014 Sub (1973-2012) - National Cancer Institute, Surveillance Research Program. 2014.

442. SEER. Surveillance Research Program, National Cancer Institute SEER*Stat software, Version 8.2.1. Bethesda, MD; 2017.

443. Ranen NG, Stine OC, Abbott MH, Sherr M, Codori A-M, Franz ML, et al. Anticipation and instability of IT-15 (CAG)N repeats in parent-offspring pairs with Huntington disease. Am J Hum Genet. 1995;57:593–602.

444. Guindalini RSC, Song A, Fackenthal JD, Olopade OI, Huo D. Genetic anticipation in BRCA1/BRCA2 families after controlling for ascertainment bias and cohort effect. Cancer. John Wiley and Sons Inc.; 2016;122:1913–20.

445. Ashley CT, Warren ST. Trinucleotide repeat expansion and human disease. Annu Rev Genet. 1995;29:703–28.

446. Martinez-Delgado B, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M, Osorio A, et

al. Genetic anticipation is associated with telomere shortening in hereditary breast cancer. PLoS Genet. 2011;7:e1002182.

447. Brown BW, Costello TJ, Hwang SJ, Strong LC. Generation or birth cohort effect on cancer risk in Li-Fraumeni syndrome. Hum Genet. 2005;118:489–98.

448. Wiernik P, Wang S, Hu X-P, Marino P, Paietta E. Age of onset evidence for anticipation in familial non-Hodgkin's lymphoma. Br J Haematol. 2000;108:72–9.

449. Wiernik PH, Ashwin M, Hu X, Paietta E, Brown K. Anticipation in familial chronic lymphocytic leukaemia. Br J Haematol. 2001;113:407–14.

450. Auer RL, Dighiero G, Goldin LR, Syndercombe-Court D, Jones C, McElwaine S, et al. Trinucleotide repeat dynamic mutation identifying susceptibility in familial and sporadic chronic lymphocytic leukaemia. Br J Haematol. 2007;136:73–9.

451. Awan H, Jønsson V, Johannesen TB, Ly B, Tjønnfjord GE. Anticipation in families with chronic lymphocytic leukemia and other lymphoproliferative disorders. Transl Oncogenomics. 2010;4:1–9.

452. Lynch HT, Watson P, Tarantolo S, Wiernik PH, Quinn-Laquer B, Isgur Bergsagel K, et al. Phenotypic heterogeneity in multiple myeloma families. J Clin Oncol. 2005;23:685–93.

453. Keyes KM, Utz RL, Robinson W, Li G. What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971–2006. Soc Sci Med. 2010;70:1100–8.

454. Howlader N, Noone A, Krapcho M, Garshell J, Miller D, Altekruse S, et al. SEER Cancer Statistics Review 1975-2011. Bethesda, MD; 2011.

455. Deshpande H, Hu X-P, Marino P, Jan N, Wiernik P. Anticipation in familial plasma cell dyscrasias. Br J Haematol. 1998;103:696–703.

456. Shugart YY, Hemminki K, Vaittinen P, Kingman A. Apparent anticipation and heterogeneous transmission patterns in familial Hodgkin's and non-Hodgkin's lymphoma: Report from a study based on Swedish cancer database. Leuk Lymphoma. 2001;42:407–15.

457. Cartwright RA, Watkins G. Epidemiology of Hodgkin's disease: A review. Hematol Oncol. 2004;22:11–26.

458. Horwitz M. The genetics of familial leukemia. Leukemia. 1997;11:1347–59.

459. King BL, Peng H-Q, Goss P, Huan S, Bronson D, Kacinski BM, et al. Repeat expansion detection analysis of (CAG)n tracts in tumor cell lines, testicular tumors, and testicular cancer families. Cancer Res. 1997;57:209–14.

460. The Leukemia & Lymphoma Society of Canada. NHL Subtypes [Internet]. Leuk. Lymphoma Soc. Canada. Toronto, ON; 2017. Available from: http://www.llscanada.org/lymphoma/non-hodgkin-lymphoma/diagnosis/nhl-subtypes

461. American Cancer Society ., Alteri R, Brooks D, Chambers W, Chang E, Cotter J, et al. Cancer facts & figures 2018. Am. Cancer Soc. Atlanta, GA; 2018.

462. Jønsson V, Tjønnfjord G, Samuelsen SO, Johannesen T, Olsen J, Sellick G, et al. Birth order pattern in the inheritance of chronic lymphocytic leukaemia and related lymphoproliferative disease. Leuk Lymphoma. 2007;48:2387–96.

463. Canadian Cancer Society's Advisory Committee on Cancer Statistics . Canadian Cancer Statistics 2017 Special topic: Pancreatic cancer. Can. Cancer Soc. Toronto, ON: Canadian Cancer Society, Statistics Canada, Public Health Agency of Canada; 2017.

464. Howlader N, Noone A, Krapcho M, Garshell J, Miller D, Altekruse S, et al. National Cancer Institute SEER Cancer Statistics Review 1975-2012. Natl Cancer Inst. 2015;103:1975–2012.

465. Carl G, Kuehn I, Levin S. spind: Spatial methods and indices. 2018.

466. Jones SJ, Voong J, Thomas R, English A, Schuetz J, Slack GW, et al. Nonrandom occurrence of lymphoid cancer types in 140 families. Leuk Lymphoma. 2017;58:2134–43.

467. Altieri A, Bermejo JL, Hemminki K. Familial risk for non-Hodgkin lymphoma and other lymphoproliferative malignancies by histopathologic subtype: The Swedish Family-Cancer Database. Blood. 2005;106:668–72.

468. Mauro FR, Giammartini E, Gentile M, Sperduti I, Valle V, Pizzuti A, et al. Clinical features and outcome of familial chronic lymphocytic leukemia. Haematologica. 2006;91:1117–20.

469. Crowther-Swanepoel D, Wild R, Sellick G, Dyer MJS, Mauro FR, Cuthbert RJG, et al. Insight into the pathogenesis of chronic lymphocytic leukemia (CLL) through analysis of IgVH gene usage and mutation status in familial CLL. Neoplasia. 2008;111:5691–3.

470. Kane E V., Roman E, Becker N, Bernstein L, Boffetta P, Bracci PM, et al. Menstrual and reproductive factors, and hormonal contraception use: Associations with non-Hodgkin lymphoma in a pooled analysis of InterLymph case-control studies. Ann Oncol. 2012;23:2362–74.

471. Kravdal O, Hansen S. Hodgkin's disease: The protective effect of childbearing. Int J Cancer. 1993;55:909–14.

472. Czene K, Adami H-O, Chang ET. Sex- and kindred-specific familial risk of non-Hodgkin's lymphoma. Cancer Epidemiol Biomarkers Prev. 2007;16:2496–9.

473. Rust J, Golombok S, Hines M, Johnston K, Golding J, The ALSPAC Study Team . The role of brothers and sisters in the gender development of preschool children. J Exp Child Psychol. 2000;77:292–303.

474. McHale SM, Crouter AC, Tucker CJ. Family context and gender role socialization in middle childhood: Comparing girls to boys and sisters to brothers. Child Dev. 1999;70:990–1004.

475. Davis JN. Birth order, sibship size, and status in modern Canada. Hum Nat. 1997;8:205–30.

476. Hrdy SB. Fitness tradeoffs in the history and evolution of delegated mothering with special reference to wet-nursing, abandonment, and infanticide. Ethol Sociobiol. 1992;13:409–42.

477. Brenøe AA, Molitor R. Birth order and health of newborns: What can we learn from Danish registry data? J Popul Econ. Journal of Population Economics; 2018;31:363–95.

478. Kaplan BA, Mascie-Taylor CG. Mother's age, birth order and health status in a British national sample. Med Anthropol. 1992;13:353–67.

479. Horton S. Birth order and child nutritional status: Evidence from the Philippines. Econ Dev Cult Change. 1988;36:341–54.

480. McNally RJQ, Eden TOB. An infectious aetiology for childhood acute leukaemia: A review of the evidence. Br J Haematol. 2004;127:243–63.

481. Pica F, Volpi A. Transmission of human herpesvirus 8: An update. Curr Opin Infect Dis. 2007;20:152–6.

482. Lehtinen M, Ögmundsdottir HM, Bloigu A, Hakulinen T, Hemminki E, Gudnadottir M, et al. Associations between three types of maternal bacterial infection and risk of leukemia in the offspring. Am J Epidemiol. 2005;162:662–7.

483. Alexander FE. Is Mycoplasma pneumonia associated with childhood acute lymphoblastic leukemia? Cancer Causes Control. 1997;8:803–11.

484. Stiller CA. What causes Hodgkin's disease in children? Eur J Cancer. 1998;34:523–8.

485. Alexander FE, McKinney PA, Williams J, Ricketts TJ, Cartwright RA. Epidemiological evidence for the "two-disease hypothesis" in Hodgkin's disease. Int J Epidemiol. 1991;20:354–61.

486. Alexander FE, Richetts TJ, McKinney PA, Cartwright RA. Community lifestyle characteristics and incidence of Hodgkin's disease in young people. Int J Cancer. 1991;48:10–4.

487. Clarke CA, Glaser SL, Keegan THM, Stroup A. Neighborhood socioeconomic status and Hodgkin's lymphoma incidence in California. Cancer Epidemiol Biomarkers Prev. 2005;14:1441–7.

488. LeShan L, Marvin S, Lyerly O. Some evidence of a relationship between Hodgkin's disease and intelligence. Arch Gen Psychiatry. 1959;1:477–9.

489. Gallagher RP, Spinelli JJ, Elwood JM, Skippen DH. Allergies and agricultural exposure as risk factors for multiple myeloma. Br J Cancer. 1983;48:853–7.

490. McWhorter WP. Allergy and risk of cancer. Cancer. 1988;62:451–5.

491. Shadman M, White E, De Roos AJ, Walter RB. Associations between allergies and risk of hematologic malignancies: Results from the VITamins and lifestyle cohort study. Am J Hematol. 2013;88:1050–4.

492. Koshiol J, Lam TK, Gridley G, Check D, Brown LM, Landgren O. Racial differences in chronic immune stimulatory conditions and risk of non-Hodgkin's lymphoma in veterans from the United States. J Clin Oncol. 2011;29:378–85.

493. Mirabelli MC, Zock J-P, D'Errico A, Kogevinas M, de Sanjosé S, Miligi L, et al. Occupational exposure to high molecular weight allergens and lymphoma risk among

Italian adults. Cancer Epidemiol Biomarkers Prev. 2009;18:2650–4.

494. Ellison-Loschmann L, Benavente Y, Douwes J, Buendia E, Font R, Alvaro T, et al. Immunoglobulin E levels and risk of lymphoma in a case-control study in Spain. Cancer Epidemiol Biomarkers Prev. 2007;16:1492–8.

495. Erber E, Lim U, Maskarinec G, Kolonel LN. Common immune-related risk factors and incident non-Hodgkin lymphoma: The multiethnic cohort. Int J Cancer. 2009;125:1440–5.

496. Vianna NJ, Davies JNP, Harris S, Lawrence CE, Arbuckle J, Marani W, et al. Tonsillectomy and childhood Hodgkin's disease. Lancet. 1980;316:338–40.

497. Ceran Ö, Aka S, Öztemel D, Uyanik B, Özkozaci T. The relationship of tonsillar hyperplasia and asthma in a group of asthmatic children. Int J Pediatr Otorhinolaryngol. 2004;68:775–8.

498. Scadding GK. Immunology of the tonsil: A review. J R Soc Med. 1990;83:104–7.

499. Endo LH, Ferreira D, Montenegro MCS, Pinto GA, Altemani A, Bortoleto AE, et al. Detection of Epstein-Barr virus in tonsillar tissue of children and the relationship with recurrent tonsillitis. Int J Pediatr Otorhinolaryngol. 2001;58:9–15.

500. Yamanaka N, Kataura A. Viral infections associated with recurrent tonsillitis. Acta Otolaryngol Suppl. 1984;416:30–7.

501. Grady D, McClung JE, Veltri RW, Sprinkle PM, Veach JS. Association of Epstein-Barr virus with acute exudative tonsillitis. Otolaryngol Neck Surg. 1982;90:11–5.

502. Tzeng Y-M, Kao L-T, Kao S, Lin H-C, Tsai M-C, Lee C-Z. An appendectomy increases the risk of rheumatoid arthritis: A five-year follow-up study. PLoS One. 2015;10:1–9.

503. Davidson WF, Mukhopadhyay P, Williams MS, Naghashfar Z, Zhou JX, Morse HC. A model system for studying mechanisms of B-cell transformation in systemic autoimmunity. In: St. Georgiev V, Western K, McGowan JJ, editors. Natl Inst Allergy Infect Dis NIH. Volume 1. Springer Science & Business Media; 2008. page 1–530.

504. Mackay IR, Rose NR. Autoimmunity and lymphoma: Tribulations of B-cells. Nat Immunol. 2001;2:793–5.

505. Cárdenas-Roldán J, Rojas-Villarraga A, Anaya J-M. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC Med. 2013;11:73–95.

506. Burnet M. Cancer - A biological approach. III. Viruses associated with neoplastic conditions. IV. Practival applications. Br Med J. 1957;1:841–7.

507. Hardell L, Bengtsson NO. Epidemiological study of socioeconomic factors and clinical findings in Hodgkin's disease, and reanalysis of previous data regarding chemical exposure. Br J Cancer. 1983;48:217–25.

508. Chang ET, Smedby KE, Hjalgrim H, Glimelius B, Adami H-O. Reliability of self-reported family history of cancer in a large case–control study of lymphoma. J Natl Cancer Inst. 2006;98:61–8.

509. Crowther-Swanepoel D, Qureshi M, Dyer MJS, Matutes E, Dearden C, Catovsky D, et al.

Genetic variation in CXCR4 and risk of chronic lymphocytic leukemia. Blood. 2009;114:4843–6.

510. Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen S-S, et al. Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. Cell. 2007;129:879–90.

511. Lawrie A, Han S, Sud A, Hosking F, Cezard T, Turner D, et al. Combined linkage and association analysis of classical Hodgkin lymphoma. Oncotarget. 2018;9:20377–85.

512. Ristolainen H, Outi Kilpivaara, Kamper P, Taskinen M, Saarinen S, Leppa S, et al. Identification of homozygous deletion in ACAN and other candidate variants in familial classical Hodgkin lymphoma by exome sequencing. Br J Haematol. 2015;170:425–39.

513. Saarinen S, Aavikko M, Aittoma K, Launonen V, Lehtonen R, Franssila K, et al. Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. Blood. 2011;118:493–8.

514. Jiang M, Bennani NN, Feldman AL. Lymphoma classification update: T-cell lymphomas, Hodgkin lymphomas, and histiocytic/dendritic cell neoplasms. Expert Rev Hematol. 2017;10:239–49.

515. Illumina. Infinium Omni5Exome-4 v1.3 BeadChip. 2016;3–6.

516. Purcell S. PLINK 1.9 [Internet]. 2019. Available from: http://pngu.mgh.harvard.edu/purcell/plink/

517. Purcell S, Neale B, Todd-brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

518. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26:2867–73.

519. Chen W-M. KING tutorial: Relationship inference [Internet]. 2019. Available from: http://people.virginia.edu/~wc9c/KING/manual.html

520. Golden Helix Inc. SNP & Variation Suite v8. Bozeman, MT; 2019.

521. Wolfe D, Dudek S, Ritchie MD, Pendergrass SA. Visualizing genomic information across chromosomes with PhenoGram. BioData Min. 2013;6:1–12.

522. Henden L, Freytag S, Afawi Z, Baldassari S, Berkovic SF, Bisulli F, et al. Identity by descent fine mapping of familial adult myoclonus epilepsy (FAME) to 2p11.2–2q11.2. Hum Genet. 2016;135:1117–25.

523. Browning BL. Refined IBD documentation. 2018;

524. Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics. 2013;194:459–71.

525. Huang L, Bercovici S, Rodriguez JM, Batzoglou S. An effective filter for IBD detection in large data sets. PLoS One. 2014;9.

526. Browning BL. Beagle 5.0 Documentation. 2018.

527. Browning BL. Beagle 3.3. Genetics. 2010;

528. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. Am J Hum Genet. The American Society of Human Genetics; 2011;88:173–82.

529. Zhang X, Li M, Lin H, Rao X, Feng W, Yang Y, et al. regSNPs-splicing: A tool for prioritizing synonymous single-nucleotide substitution. Hum Genet. 2017;136:1279–89.

530. Liu X, Jian X, Boerwinkle E. dbNSFP: A lightweight atabase of human nonsynonymous SNPs and their functional predictions. Hum Mutat. 2011;32:894–8.

531. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat. 2013;34:E2393–402.

532. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.

533. Liu Y, Kheradmand F, Davis CF, Scheurer ME, Wheeler D, Tsavachidis S, et al. Focused analysis of exome sequencing data for rare germline mutations in familial and sporadic lung cancer. J Thorac Oncol. 2016;11:52–61.

534. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2009;446:153–8.

535. Okada Y, Diogo D, Greenberg JD, Mouassess F, Achkar WAL, Fulton RS, et al. Integration of sequence data from a consanguineous family with genetic data from an outbred population identifies PLB1 as a candidate rheumatoid arthritis risk gene. PLoS One. 2014;9:1–12.

536. Yu Q, Pu S-Y, Wu H, Chen X-Q, Jiang J-J, Gu K-S, et al. TICRR contributes to tumorigenesis through accelerating DNA replication in cancers. Front Oncol. 2019;9:1–14.

537. Lefranc M. IMGT (ImMunoGeneTics) locus on focus. A new section of Experimental and Clinical Immunogenetics. Exp Clin Immunogenet. 1998;15:1–7.

538. Ott G, Rosenwald A, Campo E. Understanding MYC-driven aggressive B-cell lymphomas: Pathogenesis and classification. Blood. 2013;122:3884–91.

539. Xu-Monette ZY, Deng Q, Manyam GC, Tzankov A, Li L, Xia Y, et al. Clinical and biologic significance of MYC genetic mutations in de novo diffuse large B-cell lymphoma. Clin Cancer Res. 2016;22:3593–605.

540. Oster SK, Ho CSW, Soucie EL, Penn LZ. The myc oncogene: Marvelously complex. Adv Cancer Res. 2002;84:81–154.

541. Figueiredo JC, Knight JA, Cho S, Savas S, Onay V, Briollais L, et al. Polymorphisms cMyc-N11S and p27-V109G and breast cancer risk and prognosis. BMC Cancer. 2007;7:1–8.

542. Cole MD. MYC association with cancer risk and a new model of MYC-mediated

repression. Cold Spring Harb Perspect Med. 2014;4:1–13.

543. Chen H, Liu W, Roberts W, Hooker S, Fedor H, DeMarzo A, et al. 8q24 allelic imbalance and MYC gene copy number in primary prostate cancer. Prostate Cancer Prostatic Dis. 2010;13:238–43.

544. Janz S, Potter M, Rabkin CS. Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. Genes Chromosomes Cancer. 2003;36:211–23.

545. Tseng Y-Y, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, et al. PVT1 dependence in cancer with MYC copy-number increase. Nature. 2014;512:82–6.

546. Onagoruwa OT, Pal G, Ochu C, Ogunwobi OO. Oncogenic role of PVT1 and therapeutic implications. Front Oncol. 2020;10:1–11.

547. Carramusa L, Contino F, Ferro A, Minafra L, Perconti G, Giallongo A, et al. The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. J Cell Physiol. 2007;213:511–8.

548. Ieguchi K, Maru Y. Roles of EphA1/A2 and ephrin-A1 in cancer. Cancer Sci. 2019;110:841–8.

549. Cui Y, Wu B, Flamini V, Evans BAJ, Zhou D, Jiang WG. Knockdown of EPHA1 using CRISPR/CAS9 suppresses aggressive properties of ovarian cancer cells. Anticancer Res. 2017;37:4415–24.

550. Charmsaz S, Scott AM, Boyd AW. Targeted therapies in hematological malignancies using therapeutic monoclonal antibodies against Eph family receptors. Exp Hematol. 2017;54:31–9.

551. Oricchio E, Nanjangud G, Wolfe AL, Schatz JH, Mavrakis KJ, Jiang M, et al. The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. Cell. 2011;147:554–64.

552. Hamdi Y, Boujemaa M, Ben Rekaya M, Ben Hamda C, Mighri N, El Benna H, et al. Family specific genetic predisposition to breast cancer: Results from Tunisian whole exome sequenced breast cancer cases. J Transl Med. 2018;16:1–13.

553. Han YF, Huang HW, Li L, Cai T, Chen S, He XJ. The cytosolic iron-sulfur cluster assembly protein MMS19 regulates transcriptional gene silencing, DNA repair, and flowering time in Arabidopsis. PLoS One. 2015;10:1–21.

554. Stehling O, Vashisht A a, Mascarenhas J, Jonsson ZO, Sharma T, Netz DJ a, et al. MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity. Science (80-). 2012;337:195–9.

555. Moxley KM, Benbrook DM, Queimado L, Zuna RE, Thompson D, McCumber M, et al. The role of single nucleotide polymorphisms of the ERCC1 and MMS19 genes in predicting platinum-sensitivity, progression-free and overall survival in advanced epithelial ovarian cancer. Gynecol Oncol. 2013;130:377–82.

556. Gari K, Ortiz AML, Borel V, Flynn H, Skehel JM, Boulton SJ. MMS19 links cytoplasmic iron-sulfur cluster assembly to DNA metabolism. Science (80-). 2012;337:243–345.

557. Papal S, Cortese M, Legendre K, Sorusch N, Dragavon J, Sahly I, et al. The giant spectrin BV couples the molecular motors to phototransduction and Usher syndrome type I proteins along their trafficking route. Hum Mol Genet. 2013;22:3773–88.

558. Feng J, Zhang Z, Wu X, Mao A, Chang F, Deng X, et al. Discovery of potential new gene variants and inflammatory cytokine associations with fibromyalgia syndrome by whole exome sequencing. PLoS One. 2013;8:e65033.

559. Udali S, Guarini P, Ruzzenente A, Ferrarini A, Guglielmi A, Lotto V, et al. DNA methylation and gene expression profiles show novel regulatory pathways in hepatocellular carcinoma. Clin Epigenetics. ???; 2015;7:1–13.

560. Yamaguchi K, Huang Z, Matsumura N, Mandai M, Okamoto T, Baba T, et al. Epigenetic determinants of ovarian clear cell carcinoma biology. Physiol Behav. 2017;176:139–48.

561. Ksionda O, Limnander A, Roose JP. RasGRP Ras guanine nucleotide exchange factors in cancer. Front Biol (Beijing). 2013;8:508–32.

562. Lessard CJ, Li H, Adrianto I, Ice JA, Rasmussen A, Grundahl KM, et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. Nat Genet. 2013;45:1284–94.

563. Morris DL, Sheng Y, Zhang Y, Wang Y, Zhu Z, Tombleson P, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet. 2016;48:940–9.

564. Coughlin JJ, Stang SL, Dower NA, Stone JC, Coughlin JJ, Stang SL, et al. RasGRP1 and RasGRP3 regulate B-cell proliferation by facilitating B-cell receptor-Ras signaling. J Immunol. 2005;175:7179–84.

565. Steri M, Idda ML, Whalen MB, Orru V. Genetic variants in mRNA untranslated regions. Wiley Interdiscip Rev RNA. 2018;9:1–36.

566. Choi J, Lee K, Ingvarsdottir K, Bonasio R, Saraf A, Florens L, et al. Loss of KLHL6 promotes diffuse large B-cell lymphoma growth and survival by stabilizing the mRNA decay factor roquin2. Nat Cell Biol. 2018;20:586–96.

567. Choi J, Zhou N, Busino L. KLHL6 is a tumor suppressor gene in diffuse large B-cell lymphoma. Cell Cycle. Taylor & Francis; 2019;18:249–56.

568. Weigert O, Kopp N, Lane AA, Yoda A, Dahlberg SE, Neuberg D, et al. Molecular ontogeny of donor-derived follicular lymphomas occurring after hematopoietic cell transplantation. Cancer Discov. 2012;2:47–55.

569. Kunder CA, Roncador G, Advani RH, Gualco G, Bacchi CE, Sabile JM, et al. KLHL6 is preferentially expressed in germinal center – derived B-cell lymphomas. Am J Clin Pathol. 2017;148:465–76.

570. Kroll J, Shi X, Caprioli A, Liu H-H, Waskow C, Lin K-M, et al. The BTB-kelch protein KLHL6 is involved in B-lymphocyte antigen receptor signaling and germinal genter formation. Mol Cell Biol. 2005;25:8531–40.

571. Schmitz R, Stanelle J, Hansmann M-L, Küppers R. Pathogenesis of classical and lymphocyte-predominant Hodgkin lymphoma. Annu Rev Pathol Mech Dis. 2009;4:151–

74.

572. Kim J, Hwang B, Park SN, Huh S, Im K, Choi S, et al. Genomic profile of chronic lymphocytic leukemia in Korea identified by targeted sequencing. PLoS One. 2016;11:1–24.

573. Isaza-Correa JM, Liang Z, van den Berg A, Diepstra A, Visser L. Toll-like receptors in the pathogenesis of human B-cell malignancies. J Hematol Oncol. 2014;7:1–10.

574. Botos I, Segal DM, Davies DR. The structural biology of toll-like receptors. Structure. 2012;19:447–59.

575. Omueti KO, Beyer JM, Johnson CM, Lyle EA, Tapping RI. Domain exchange between human toll-like receptors 1 and 6 reveals a region required for lipopeptide discrimination. J Biol Chem. 2005;280:36616–25.

576. Liang S, Hosur KB, Lu S, Nawar HF, Weber BR, Tapping RI, et al. Mapping of a microbial protein domain involved in binding and activation of the TLR2/TLR1 heterodimer. J Immunol. 2010;182:2978–85.

577. Purdue MP, Lan Q, Wang SS, Menashe I, Zheng T-Z, Grulich AE, et al. A pooled investigation of Toll-like receptor gene variants and risk of non-Hodgkin lymphoma. Carcinogenesis. 2009;30:275–81.

578. Monroy CM, Cortes AC, Lopez MS, D'Amelio AJM, Etzel CJ, Younes A, et al. Hodgkin disease risk: Role of genetic polymorphisms and gene – gene interactions in inflammation pathway genes. Mol Carcinog. 2011;50:36–46.

579. Ferrero E, Lo Buono N, Morone S, Parrott R, Mancini C, Brusco A, et al. Human canonical CD157/Bst1 is an alternatively spliced isoform masking a previously unidentified primate-specific exon included in a novel transcript. Sci Rep. 2017;7:1–14.

580. Higashida H, Liang M, Yoshihara T, Akther S, Fakhrul A, Stanislav C, et al. An immunohistochemical, enzymatic and behavioral study of CD157/BST-1 as a neuroregulator. BMC Neurosci. 2017;18:1–12.

581. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet. 2009;41:1303–7.

582. Wang S, Xu Y-F, Ding X-Y, Liu Z-R, Ding Y, Jin B, et al. Association between bone marrow stromal cell antigen 1 gene polymorphisms and the susceptibility to Parkinson's disease: A meta-analysis. Neurosci Lett. 2015;599:120–4.

583. Degerman E, Ahmad F, Chung YW, Guirguis E, Omar B, Stenson L, et al. From PDE3B to the regulation of energy homeostasis. Curr Opin Pharmacol. 2011;11:676–82.

584. Meyre D, Farge M, Lecoeur C, Proenca C, Durand E, Allegaert F, et al. R125W coding variant in TBC1D1 confers risk for familial obesity and contributes to linkage on chromosome 4p14 in the French population. Hum Mol Genet. 2008;17:1798–802.

585. An D, Toyoda T, Taylor EB, Yu H, Fujii N, Hirshman MF, et al. TBC1D1 regulates insulin- and contraction-induced glucose transport in mouse skeletal muscle. Diabetes. 2010;59:1358–65.

586. Vincourt JB, Jullien D, Amalric F, Girard JP. Molecular and functional characterization of SLC26A11, a sodium-independent sulfate transporter from high endothelial venules. FASEB J. 2003;17:890–2.

587. Lougaris V, Chou J, Baronio M, Gazzurelli T, Moratto D, Badolato R, et al. Novel biallelic TRNT1 mutations resulting in sideroblastic anemia, combined B- and T-cell defects, hypogammaglobulinemia, recurrent infections, hypertrophic cardiomyopathy and developmental delay. Clin Immunol. 2018;188:20–2.

588. Uniprot. UniProtKB - A0A075B6I6 (LV150_HUMAN): Probable non-functional immunoglobulin lambda variable 1-50 [Internet]. 2020. Available from: https://www.uniprot.org/uniprot/A0A075B6I6

589. Smith A, Moran A, Boyd MC, Bulman M, Shenton A, Smith L, et al. Phenocopies in BRCA1 and BRCA2 families: Evidence for modifier genes and implications for screening. J Med Genet. 2007;44:10–5.

590. Burgess JR, Nord B, David R, Greenaway TM, Parameswaran V, Larsson C, et al. Phenotype and phenocopy: The relationship between genotype and clinical phenotype in a single large family with multiple endocrine neoplasia type 1 (MEN 1). Clin Endocrinol (Oxf). 2000;53:205–11.

591. Moore RC, Xiang F, Monaghan J, Han D, Zhang Z, Edström L, et al. Huntington disease phenocopy is a familial prion disease. Am J Hum Genet. 2001;69:1385–8.

592. Horwitz M, Wiernik PH. Pseudoautosomal linkage of Hodgkin disease. Am J Hum Genet. 1999;65:1413–22.

593. Kuhlen M, Taeubner J, Brozou T, Wieczorek D, Siebert R, Borkhardt A. Family-based germline sequencing in children with cancer. Oncogene. 2019;38:1367–80.

594. Geschwind DH. Autism: Many genes, common pathways? Cell. 2008;135:391–5.

595. Guo H, Wang T, Wu H, Long M, Coe BP, Li H, et al. Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. Mol Autism. Molecular Autism; 2018;9:1–12.

596. Wiśniowiecka-Kowalnik B, Nowakowska BA. Genetics and epigenetics of autism spectrum disorder — current evidence in the field. J Appl Genet. 2019;60:37–47.

597. Ben-David E, Shifman S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. PLoS Genet. 2012;8:1–13.

598. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet. 2010;19:176–87.

599. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. 2012;367:1321–31.

600. Griesi-Oliveira K, Laurato Sertié A. Autism spectrum disorders: An updated guide for genetic counseling. Einstein (Sao Paulo). 2017;15:233–8.

601. Reed X, Bandrés-Ciga S, Blauwendraat C, Cookson MR. The role of monogenic genes in

idiopathic Parkinson’s disease. Neurobiol Dis. 2019;124:230–9.

602. Wright CF, West B, Tuke M, Jones SE, Patel K, Laver TW, et al. Assessing the pathogenicity, penetrance, and expressivity of putative disease-causing variants in a population setting. Am J Hum Genet. 2019;104:275–86.

603. Roma-Rodrigues C, Fernandes AR. Genetics of hypertrophic cardiomyopathy: Advances and pitfalls in molecular diagnosis and therapy. Appl Clin Genet. 2014;7:195–208.

604. Mandelker D, Ceyhan-Birsoy O. Evolving significance of tumor-normal sequencing in cancer care. Trends in Cancer. 2020;6:31–9.

605. Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer. 2001;1:157–62.

606. Thiagalingam S, Foy RL, Cheng K, Lee HJ, Thiagalingam A, Ponte JF. Loss of heterozygosity as a predictor to map tumor suppressor genes in cancer: Molecular basis of its occurrence. Cancer Biol. 2002;14:65–72.

607. Van Loo P, Nilsen G, Nordgard SH, Vollan HKM, Borresen-Dale A-L, Kristensen VN, et al. Analyzing cancer samples with SNP arrays. In: Wang J, Tan AC, Tian T, Walker JM, editors. Next Gener microarray Bioinforma Methods Protoc. New York, NY: Springer Protocols; 2012. page 57–72.

608. Li A, Liu Y, Zhao Q, Feng H, Harris L, Wang M. Genome-wide identification of somatic aberrations from paired normal-tumor samples. PLoS One. 2014;9:e87212.

609. Lobo I. Multifactorial inheritance and genetic disease. Nat Educ. 2008;1:1–5.

610. Zhang J, Ye X, Wu C, Fu H, Xu W, Hu P. Modeling gene-environment interaction for the risk of non-Hodgkin lymphoma. Front Oncol. 2019;9:1–12.

611. Schober P, Vetter TR. Survival analysis and interpretation of time-to-event data: The tortoise and the hare. Anesth Analg. 2018;127:792–8.

612. Keiley MK, Martin NC. Survival analysis in family research. J Fam Psychol. 2005;19:142–56.

613. Hanley JA, Negassa A, Edwardes MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: An orientation. Am J Epidemiol. 2003;157:364–75.

614. Stillman S. Review of generalized estimating equations by Hardin and Hilbe. Stata J. 2003;3:208–10.

615. Hardin JW, Hilbe JM. Generalized estimating equations. First. Hardin JW, Hilbe JM, editors. Stat. Ref. Online. Boca Raton, FL: Chapman & Hall/CRC; 2003.

616. Zheng P. An overview of methods to address the multiple comparison problem. Towar Data Sci. Toronto, ON; 2018;

617. Abdi H. The Bonferonni and Šidák corrections for multiple comparisons. Cogn Sci - UC San Diego. San Diego, CA; 2006;1–9.

618. Zweifel MJ. The power and type I error rate of Holm’s procedure when the assumptions

of normality and variance homogeneity are violated. University of Nebraska-Lincoln; 2014.

619. Hartwick J, Barki H. Research report — Hypothesis testing and hypothesis generating research: An example from the user participation literature. Inf Syst Res. 1994;5:446–9.

620. Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. Acta Orthop. 2019;90:297.

621. Pemov A, Pathak A, Jones S, Dewan R, Merberg J, Karra S, et al. In search of genetic factors predisposing to familial hairy cell leukemia (HCL): Exome-sequencing of four multiplex HCL pedigrees. Leukemia. 2020;

622. Di Pietro A, Good-Jacobson KL. Disrupting the code: Epigenetic dysregulation of lymphocyte function during infectious disease and lymphoma development. J Immunol. 2018;201:1109–18.

623. Morton LM, Bernstein L, Wang SS, Hein DW, Rothman N, Colt JS, et al. Hair dye use, genetic variation in N-acetyltransferase 1 (NAT1) and 2 (NAT2), and risk of non-Hodgkin lymphoma. Carcinogenesis. 2007;28:1759–64.

624. Gathany AH, Hartge P, Davis S, Cerhan JR, Severson RK, Cozen W, et al. Relationship between interferon regulatory factor 4 genetic polymorphisms, measures of sun sensitivity and risk for non-Hodgkin lymphoma. Cancer Causes Control. 2009;20:1291–302.

625. Kanakry JA, Ambinder RF. EBV-related lymphomas: New approaches to treatment. Curr Treat Options Oncol. 2013;14:224–36.

626. Paschos K, Allday MJ. Epigenetic reprogramming of host genes in viral and microbial pathogenesis. Trends Microbiol. 2010;18:439–47.

627. Martin KA, Lupey LN, Tempera I. Epstein-Barr virus oncoprotein LMP1 mediates epigenetic changes in host gene expression through PARP1. J Virol. 2016;90:8520–30.

628. Bougie O, Weberpals JI. Clinical considerations of BRCA1- and BRCA2-mutation carriers: A review. Int J Surg Oncol. 2011;2011:1–11.

629. Smith MR. Rituximab (monoclonal anti-CD20 antibody): Mechanisms of action and resistance. Oncogene. 2003;22:7359–68.

630. Nussinov R, Jang H, Tsai C-J, Feixiong C. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. PLoS Comput Biol. 2019;15:1–54.

631. Wang SS, Slager SL, Brennan P, Holly EA, De Silvia S, Bernstein L, et al. Family history of hematopoietic malignancies and risk of non-Hodgkin Lymphoma: A pooled analysis of 10,214 cases and 11,906 controls from the InterLymph Consortium. Blood. 2007;109:3479–88.

632. Altieri A, Hemminki K. The familial risk of Hodgkin's lymphoma ranks among the highest in the Swedish Family-Cancer Database. Leukemia. 2006;20:2062–3.

633. Siebert R, Louie D, Lacher M, Schluger A, Offit K. Familial Hodgkin's and non-Hodgkin's

lymphoma: Different patterns in first-degree relatives. Leuk Lymphoma. 1997;27:503–7.

634. Srivastava A, Giangiobbe S, Kumar A, Paramasivam N, Dymerska D, Behnisch W, et al. Identification of familial Hodgkin Lymphoma predisposing genes using whole genome sequencing. Front Bioeng Biotechnol. 2020;8:1–11.

Appendices

Appendix A - Supplementary materials for Chapter 3

A.1 Supplementary tables

Supplementary Table A.1: Ethnicity and year of diagnosis incidence rates for NHL, HL, CLL and MM cases using USA (SEER) population data.

	White ethnicity				Asian ethnicity			
Year	**NHL**	**HL**	**CLL**	**MM**	**NHL**	**HL**	**CLL**	**MM**
1973	8.1	3.7	3.7	4.2	7	0.9	0.9	3
1974	8.8	3.4	3.4	4.3	7.4	1.4	1.4	4.6
1975	9.6	3.2	3.2	4.6	6.7	1.6	1.6	4.2
1976	10.1	2.9	2.9	4.8	8.9	1	1	3.9
1977	10.9	3.3	3.3	4.8	8.8	0.8	0.8	2.9
1978	11.7	3.1	3.1	4.4	8.8	0.7	0.7	3.6
1979	11.8	3.1	3.1	4.5	9.6	0.8	0.8	3.4
1980	11.9	3	3	4.4	10.5	1.2	1.2	3.6
1981	13	3.2	3.2	4.5	13	1	1	3.4
1982	13.4	3.1	3.1	4.8	7.2	0.8	0.8	3.1
1983	13.9	3.3	3.3	4.7	8	0.8	0.8	3.5
1984	14.8	3.4	3.4	5	10.2	0.8	0.8	3.6
1985	14.5	3.3	3.3	4.8	10.6	1	1	3.3
1986	15.7	2.9	2.9	4.6	9	1.4	1.4	4.2
1987	15.9	3.3	3.3	5.4	10.7	1.3	1.3	3.4
1988	16.7	3.4	3.4	4.7	10.4	1.2	1.2	2.7
1989	16.8	3.4	3.4	4.8	11.1	1.2	1.2	3.5
1990	17.5	3.3	3.3	4.8	11.8	0.9	0.9	3.7
1991	17.5	3.3	3.3	5.3	11.5	0.9	0.9	4
1992	17.5	2.9	6.6	5.2	11.6	1.3	1.6	3.1
1993	17.6	3	6.1	4.9	11.5	0.8	1.4	3.5
1994	18.1	2.9	6.4	5	13.4	1.1	1.6	3.4
1995	18.7	2.9	6.2	5	13.2	0.9	1.5	2.7
1996	18.7	3	6.3	5.2	12.1	0.7	1.6	3.5
1997	18.8	2.9	6	5.4	12.5	1	1.4	4
1998	19	3.1	5.9	5.1	12.4	0.9	1.5	3.2
1999	19.3	3.1	5.8	4.9	14	1.1	1.2	3.3
2000	18.7	2.8	6.1	5.2	12.3	1.1	1.5	3.8
2001	19.3	2.9	6.4	5	14.3	1.3	1.4	3.4
2002	19.5	3	6.4	5.1	13.7	1.4	1.6	3.5
2003	19.7	2.9	6.5	5	13.7	1.2	1.6	3.3
2004	20.2	2.9	6.7	5.2	13.8	1.3	1.5	3.2
2005	20.1	3.1	6.5	5.3	13.2	1.1	1.4	3.6
2006	20	3	6.6	5	12.2	1.5	1.6	3.3
2007	20.3	3.1	6.4	5.1	13.3	1.4	1.4	3.5
2008	20.4	3	6.4	5.3	13.9	1.3	1.5	3.7
2009	20.1	2.9	6.1	5.3	13.5	1.2	1.6	3.5
2010	20.3	2.8	6	5.5	13.5	1.5	1.3	3.9
2011	19.6	2.8	5.8	5.4	13.4	1.3	1.5	3.6
2012	19.9	2.7	5.4	5.5	13.2	1.3	1.3	3

Notes: SEER population data obtained using Site and Morphology variable "Lymphoma subtype recode/WHO 2008."

- NHL variable defined as 2.Non-Hodgkin lymphoma - [2(a)2.8.2.Multiple myeloma/plasma-cell leuk + 2(a)2.1.1 Chronic/Small lymphocytic leuk/lymph].
- HL variable defined as 1.Hodgkin Lymphoma.
- CLL variable defined as 2(a)2.1.1.Chronic/Small lymphocytic leuk/lymph.
- MM variable defined as 2(a)2.8.2.Multiple myeloma/plasma-cell leuk.
- SEER ethnicities defined as "White" for 1973-2012, "Other (American Indian/AK Native, Asian/Pacific Islander) for 1973-1991, and "Asian or Pacific Islander" for 1992-2012.

Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

Supplementary methods:

Calculating weighted incidence rates

For analysis of familial lymphoid cancer co-occurrence patterns, we obtained year-of-diagnosis and ethnicity-specific population incidence rates for each NHL, HL, CLL and MM case. Incidence rates (**Supplementary Table A.1**) were obtained from Surveillance, Epidemiology, and End Results (SEER) databases (439) (440) (441), accessed through SEER*Stat software (442). SEER data were used because comparable Canadian data were not available. Since incidence rates vary by year and by ethnicity, the rates used to calculate the probability of a lymphoid cancer type occurring by chance were weighted by the frequency of observed year-of-diagnosis and ethnicity of cases in the families.

Example:

For example, 1 of 160 NHL cases (or 0.625 % of NHL cases) in this study were diagnosed in 1993; this case was of European (or white) ancestry (count data is available in **Supplementary Table A.2;** frequency data is available in **Supplementary Table A.3**). The frequency of observed cases was multiplied by the year- and ethnicity-specific incidence rate (in **Supplementary Table A.1**) to obtain a weighted incidence rate (**Supplementary Table A.4**). For example, the frequency of observed white NHL cases diagnosed in 1993 was 0.625%, and the 1993 white NHL incidence rate was 17.6 (per 100,000 people); multiplying the frequency (0.625% or 0.00625) by the incidence rate (17.6) produced an incidence rate that was weighted by the frequency of our observed familial data; 0.110 (**Supplementary Table A.4**). The weighted incidence rates for all observed combinations of ethnicity, year-of-diagnosis, and type of lymphoid cancer, were totaled and normalized to 1.00, representing the probability of occurrence (**Supplementary Table A.5**). The probabilities of observing a lymphoid cancer type were cross multiplied for all possible combinations of lymphoid cancer pairs. For example, the expected probability of observing an

NHL case is 0.568. The probability of observing an NHL paired with an NHL is therefore 0.586 x 0.568, or 0.323. All pair combinations total to 1 or 100%.

Year and ethnicity specific incidence rates were matched to the same/closest year of diagnosis for each lymphoid cancer case. A diagnosis prior to 1973 would have a 1973 incidence rate applied.

Supplementary Table A.2: Observed NHL, HL, CLL and MM Lymphoid Cancer Family Study cases by ethnicity and year of diagnosis.

	White ethnicity				Asian ethnicity			
Year	**NHL**	**HL**	**CLL**	**MM**	**NHL**	**HL**	**CLL**	**MM**
1973	7	8	7	3	0	0	0	0
1974	0	0	0	0	0	0	0	0
1975	1	0	0	0	0	0	0	0
1976	3	0	0	0	0	0	0	0
1977	3	0	0	0	0	0	0	0
1978	2	0	1	0	0	0	0	0
1979	0	0	0	0	0	0	0	0
1980	1	1	1	0	0	0	0	0
1981	0	0	0	2	0	0	0	0
1982	1	1	2	0	0	0	0	0
1983	1	0	1	0	0	0	0	0
1984	1	1	1	2	0	0	0	0
1985	1	2	0	0	0	0	0	0
1986	1	0	0	0	0	0	0	0
1987	1	1	2	0	0	0	0	0
1988	3	0	0	0	0	0	0	0
1989	1	0	0	0	0	0	0	0
1990	2	3	3	0	0	0	0	0
1991	1	2	0	0	0	0	0	0
1992	4	1	2	0	0	0	0	0
1993	1	0	0	0	0	0	0	0
1994	3	0	1	0	0	0	0	0
1995	2	1	1	1	0	0	0	0
1996	0	0	3	0	1	0	0	0
1997	6	0	0	0	0	0	0	0
1998	4	3	3	0	0	0	0	0
1999	5	0	2	0	1	0	0	0
2000	4	2	2	0	0	0	0	0
2001	4	3	4	2	0	0	0	0
2002	7	1	2	0	1	0	0	0
2003	8	1	7	0	0	0	0	1
2004	7	0	2	1	0	1	0	0
2005	9	2	5	2	1	0	0	0
2006	5	2	6	0	0	0	0	0
2007	8	3	7	0	1	0	0	0
2008	7	1	9	1	1	0	0	0
2009	9	3	3	0	0	0	0	0
2010	8	0	3	1	1	0	0	0
2011	8	4	6	2	0	0	0	0
2012	11	2	8	0	3	0	0	0
Total	150	48	94	17	10	1	0	1

Supplementary Table A.3: Frequency of observed NHL, HL, CLL and MM cases by ethnicity and year of diagnosis.

	White ethnicity				Asian ethnicity			
Year	**NHL**	**HL**	**CLL**	**MM**	**NHL**	**HL**	**CLL**	**MM**
1973	0.04375	0.163265	0.074468	0.166667	0	0	0	0
1974	0	0	0	0	0	0	0	0
1975	0.00625	0	0	0	0	0	0	0
1976	0.01875	0	0	0	0	0	0	0
1977	0.01875	0	0	0	0	0	0	0
1978	0.0125	0	0.010638	0	0	0	0	0
1979	0	0	0	0	0	0	0	0
1980	0.00625	0.020408	0.010638	0	0	0	0	0
1981	0	0	0	0.111111	0	0	0	0
1982	0.00625	0.020408	0.021277	0	0	0	0	0
1983	0.00625	0	0.010638	0	0	0	0	0
1984	0.00625	0.020408	0.010638	0.111111	0	0	0	0
1985	0.00625	0.040816	0	0	0	0	0	0
1986	0.00625	0	0	0	0	0	0	0
1987	0.00625	0.020408	0.021277	0	0	0	0	0
1988	0.01875	0	0	0	0	0	0	0
1989	0.00625	0	0	0	0	0	0	0
1990	0.0125	0.061224	0.031915	0	0	0	0	0
1991	0.00625	0.040816	0	0	0	0	0	0
1992	0.025	0.020408	0.021277	0	0	0	0	0
1993	0.00625	0	0	0	0	0	0	0
1994	0.01875	0	0.010638	0	0	0	0	0
1995	0.0125	0.020408	0.010638	0.055556	0	0	0	0
1996	0	0	0.031915	0	0.00625	0	0	0
1997	0.0375	0	0	0	0	0	0	0
1998	0.025	0.061224	0.031915	0	0	0	0	0
1999	0.03125	0	0.021277	0	0.00625	0	0	0
2000	0.025	0.040816	0.021277	0	0	0	0	0
2001	0.025	0.061224	0.042553	0.111111	0	0	0	0
2002	0.04375	0.020408	0.021277	0	0.00625	0	0	0
2003	0.05	0.020408	0.074468	0	0	0	0	0.055556
2004	0.04375	0	0.021277	0.055556	0	0.020408	0	0
2005	0.05625	0.040816	0.053191	0.111111	0.00625	0	0	0
2006	0.03125	0.040816	0.06383	0	0	0	0	0
2007	0.05	0.061224	0.074468	0	0.00625	0	0	0
2008	0.04375	0.020408	0.095745	0.055556	0.00625	0	0	0
2009	0.05625	0.061224	0.031915	0	0	0	0	0
2010	0.05	0	0.031915	0.055556	0.00625	0	0	0
2011	0.05	0.081633	0.06383	0.111111	0	0	0	0
2012	0.06875	0.040816	0.085106	0	0.01875	0	0	0
	- - - - - - - -	- - - - - - - -	- - - - - - - -	- - - - - - - -	- - - - - - - -	- - - - - - - -	- - - - - - - -	- - - - - - - -

Supplementary Table A.4: Weighted incidence rates of observed NHL, HL, CLL and MM cases by ethnicity and year of diagnosis.

	White ethnicity				Asian ethnicity				Total ethnicities			
Year	**NHL**	**HL**	**CLL**	**MM**	**NHL**	**HL**	**CLL**	**MM**	**NHL**	**HL**	**CLL**	**MM**
1973	0.354	0.604	0.276	0.700					0.354	0.604	0.276	0.700
1974												
1975	0.060								0.060			
1976	0.189								0.189			
1977	0.204								0.204			
1978	0.146		0.033						0.146		0.033	
1979												
1980	0.074	0.061	0.032						0.074	0.061	0.032	
1981				0.500								0.500
1982	0.084	0.063	0.066						0.084	0.063	0.066	
1983	0.087		0.035						0.087		0.035	
1984	0.093	0.069	0.036	0.556					0.093	0.069	0.036	0.556
1985	0.091	0.135							0.091	0.135		
1986	0.098								0.098			
1987	0.099	0.067	0.070						0.099	0.067	0.070	
1988	0.313								0.313			
1989	0.105								0.105			
1990	0.219	0.202	0.105						0.219	0.202	0.105	
1991	0.109	0.135							0.109	0.135		
1992	0.438	0.059	0.140						0.438	0.059	0.140	
1993	0.110								0.110			
1994	0.339		0.068						0.339		0.068	
1995	0.234	0.059	0.066	0.278					0.234	0.059	0.066	0.278
1996			0.201		0.076				0.076		0.201	
1997	0.705								0.705			
1998	0.475	0.190	0.188						0.475	0.190	0.188	
1999	0.603		0.123		0.088				0.691		0.123	
2000	0.468	0.114	0.130						0.468	0.114	0.130	
2001	0.483	0.178	0.272	0.556					0.483	0.178	0.272	0.556
2002	0.853	0.061	0.136		0.086				0.939	0.061	0.136	
2003	0.985	0.059	0.484					0.183	0.985	0.059	0.484	0.183
2004	0.884		0.143	0.289		0.027			0.884	0.027	0.143	0.289
2005	1.131	0.127	0.346	0.589	0.083				1.213	0.127	0.346	0.589
2006	0.625	0.122	0.421						0.625	0.122	0.421	
2007	1.015	0.190	0.477		0.083				1.098	0.190	0.477	
2008	0.893	0.061	0.613	0.294	0.087				0.979	0.061	0.613	0.294
2009	1.131	0.178	0.195						1.131	0.178	0.195	
2010	1.015		0.191	0.306	0.084				1.099		0.191	0.306
2011	0.980	0.229	0.370	0.600					0.980	0.229	0.370	0.600
2012	1.368	0.110	0.460		0.248				1.616	0.110	0.460	
Sum	**17.06**	**3.073**	**5.678**	**4.667**	**0.833**	**0.027**	**0**	**0.183**	**17.89**	**3.1**	**5.678**	**4.85**
Total	**30.48**				**1.04**				**31.52**			

	NHL	**HL**	**CLL**	**MM**
Probability of occurrence	0.568	0.098	0.180	0.154

Supplementary Table A.5: Estimated probability of NHL, HL, CLL and MM occurring in the population.

Type of cancer	Probability of occurrence
NHL	0.568
HL	0.098
CLL	0.180
MM	0.154

Notes: The probabilities of observing a lymphoid cancer type (above) are cross multiplied to calculate the probability of observing all possible combinations of lymphoid cancer pairs. E.g., the expected probability of observing an NHL case is 0.568. The probability of observing an NHL paired with an NHL is therefore 0.568 x 0.568, or 0.323. All pair combinations total to 1 or 100%. Population incidence rates were obtained using SEER (USA) year- and ethnicity-specific incidence rates.

Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

Supplementary Table A.6: Estimated probability of NHL, HL, CLL and MM occurring in the Canadian population based on 2015 incidence rates.

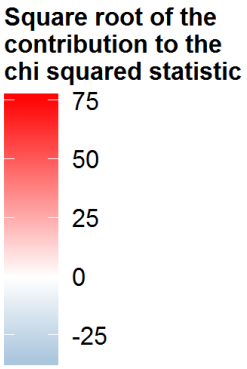
Type of cancer	Probability of occurrence
NHL	0.501
HL	0.085
CLL	0.189
MM	0.224

Notes: 2015 Incidence rates for NHL, HL, CLL and MM were obtained from the Canadian Cancer Society. Population incidence rates were normalized to 1.00 or 100%. Year- and ethnicity- specific incidence rates were unavailable for Canadian data.

Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

Supplementary Figure A.1: Lymphoid cancer co-occurrence in families with two, three and four of more lymphoid cancer cases based on 2015 Canadian incidence rates.

	Expected	Two Cases	Three Cases	≥Four Cases
	Families	88	37	15
	Cases	176	111	66
NHL/NHL	37.3	33 (29)	43.5 (47)	16.5 (18)
NHL/CLL	14.1	20.5 (18)	16.7 (18)	15.6 (17)
NHL/HL	6.4	10.2 (9)	8.3 (9)	22 (24)
NHL/MM	16.7	2.3 (2)	2.8 (3)	6.4 (4)
CLL/CLL	5.3	20.5 (18)	15.7 (17)	15.6 (17)
CLL/HL	2.4	1.1 (1)	3.7 (4)	8.3 (9)
CLL/MM	6.3	0 (0)	2.8 (3)	2.8 (3)
HL/HL	1.1	6.8 (6)	6.5 (7)	9.2 (10)
HL/MM	2.9	2.2 (2)	0 (0)	0.8 (1)
MM/MM	7.5	3.4 (3)	0 (0)	2.5 (3)
	P	<.0001	<.0001	<.0001
	Chi squared	146.72	122.02	275.25



Notes: observed % (*n* pairs). Colour is associated with variation from expected frequency of random lymphoid cancer co-occurrences, as determined by the signed square-root of the contribution to the chi-square statistic.
Abbreviations: NHL, non-Hodgkin lymphoma; HL, Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; MM, multiple myeloma.

Appendix B - Supplementary materials for Chapter 4

B.1 Supplementary tables

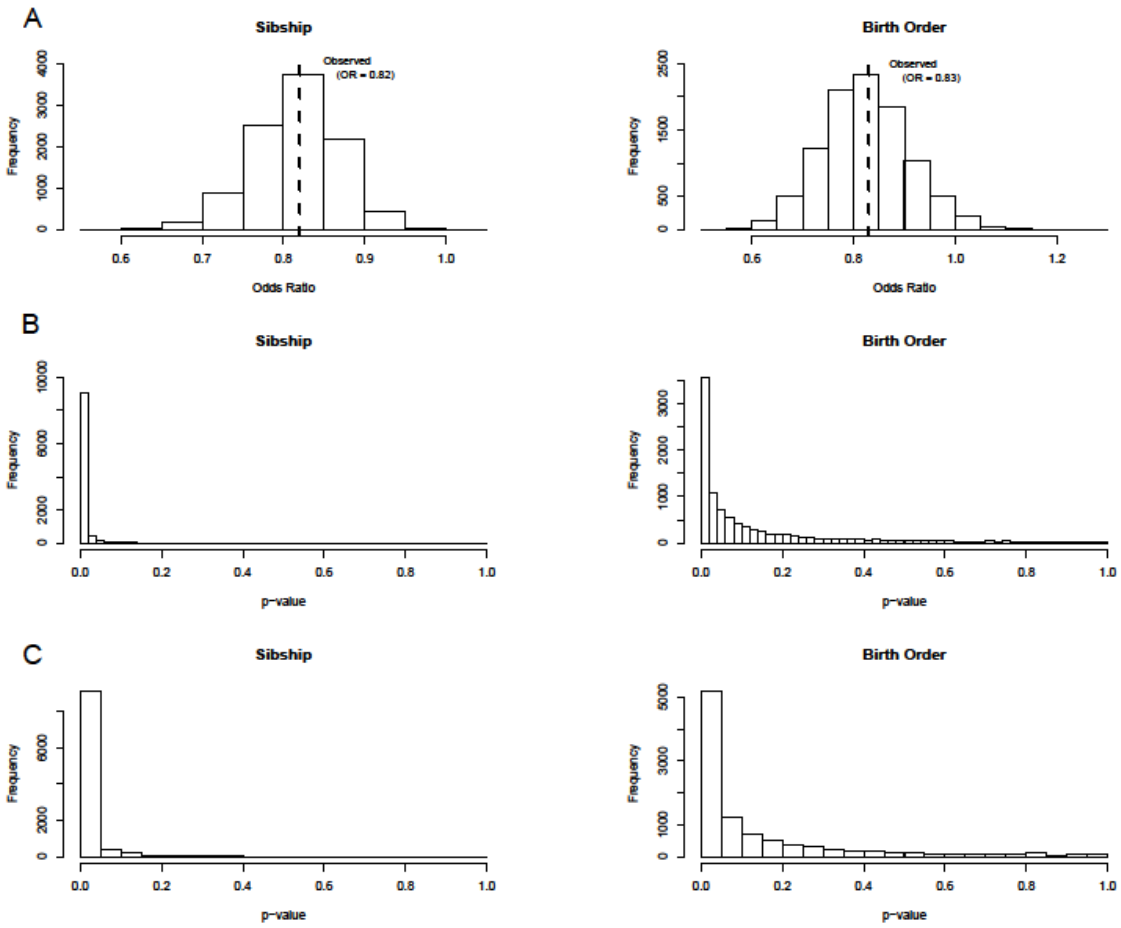
Supplementary Table B.1: Person-time and events for lymphoid cancer cases and non-lymphoid cancer cases.

Generation	Lymphoid cancer cases			Non-lymphoid cancer cases		
	Number of events	Total years	Person-years per event	Number of events	Total years	Person-years per event
0	0	78	***/***	0	103	***/***
1	12	3150	263	5	2521	504
2	51	10498	206	33	10497	318
3	52	10527	202	16	10673	667
4	10	3143	314	3	3210	1070
5	0	126	***/***	0	126	***/***

Notes: Person-time analysis was calculated for 140 families only. This analysis was completed during an earlier stage of analysis (in 2016), whereas corrections for ascertainment bias were performed in 2019 with a larger sample size.

Appendix C - Supplementary materials for Chapter 5

C.1 Supplementary figures



Supplementary Figure C.1: Results of logistic regression and chi-squared tests for trend of permuted family data. (**A**) Distributions of 10,000 odds ratio (OR) estimates and (**B**) corresponding *p*-values obtained from logistic regression with sibship size and logistic regression with birth order. The dashed line represents the observed OR from the GEE logistic regression with the full family data set. (**C**) Distributions of the *p*-values obtained from chi-square tests for trend in proportion of cases with sibship size and birth order.

C.2 Supplementary tables

Supplementary Table C.1: Associations between risk of HL and family structure, stratified by patient age at diagnosis.

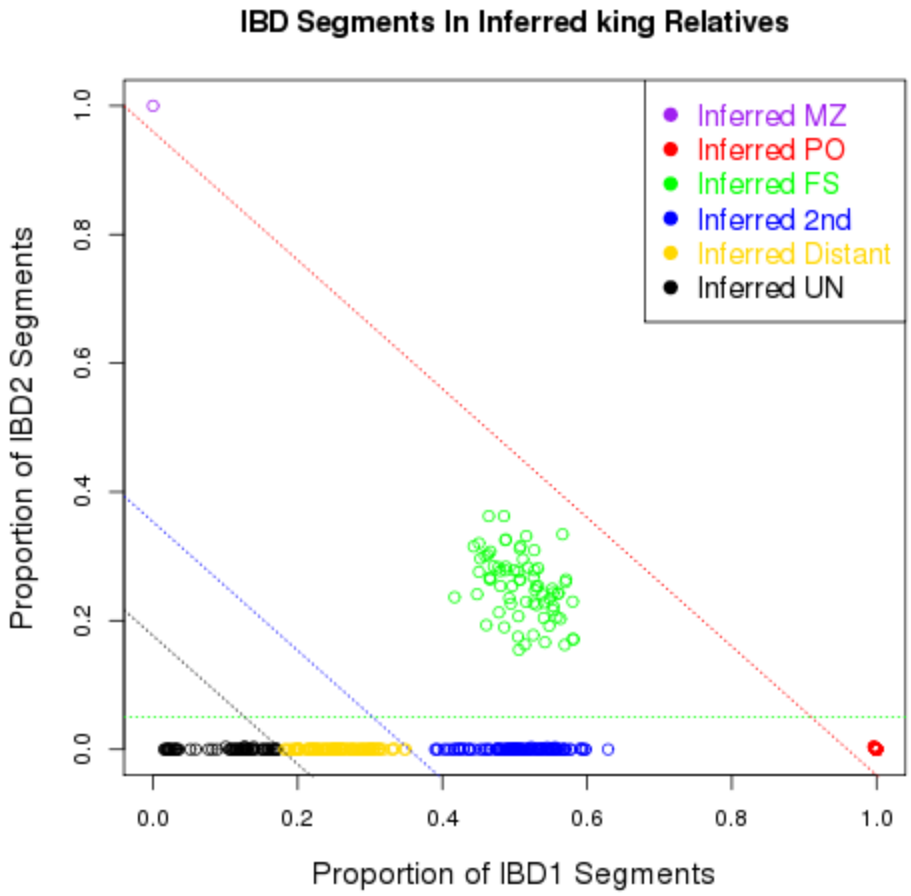
			Individuals within families, n			**OR (95% CI)**[1,2]	
Type, dx age	Age cutoff criteria	**Families**, n	**Unaffected sibs**	**Lymphoid affected**	Total	**Birth order**	**Sibship size**
HL, all	N/A	46	185	70	255	0.94 (0.80 - 1.12)	**0.83** (0.78 - 0.90)
HL, ≤ 50	Cozen *et al*, (2009). (226)	41	149	58	207	0.96 (0.77 - 1.20)	**0.88** (0.83 - 0.94)
HL, < 40		38	103	47	150	0.98 (0.76 - 1.28)	**0.85** (0.77 - 0.95)
HL, > 50		12	39	12	51	1.07 (0.70 - 1.62)	**0.67** (0.56 - 0.79)
HL, ≥ 40		21	85	23	108	0.94 (0.70 - 1.27)	**0.84** (0.78 - 0.91)
HL, <15	Westergaard *et al* (1997). (260)	5	9	5	14	1.13 (0.62 - 2.06)	1.09 (0.65 - 1.83)
HL, 15-42		36	128	48	176	1.00 (0.77 - 1.29)	**0.89** (0.83 - 0.94)
HL, > 42		16	54	17	71	1.01 (0.67 - 1.52)	**0.74** (0.64 - 0.86)

Notes: [1] Adjusted for age at enrollment (continuous) and sex (male/female). Age at death was used for non-living participants. [2] OR and 95% CI were estimated by GEE logistic regression (clustered by family) with an autoregressive correlation structure. Birth order referent group: first born; Sibship size referent group: two siblings; Bold type, 95% CI does not include 1.00, denoting a significant association.

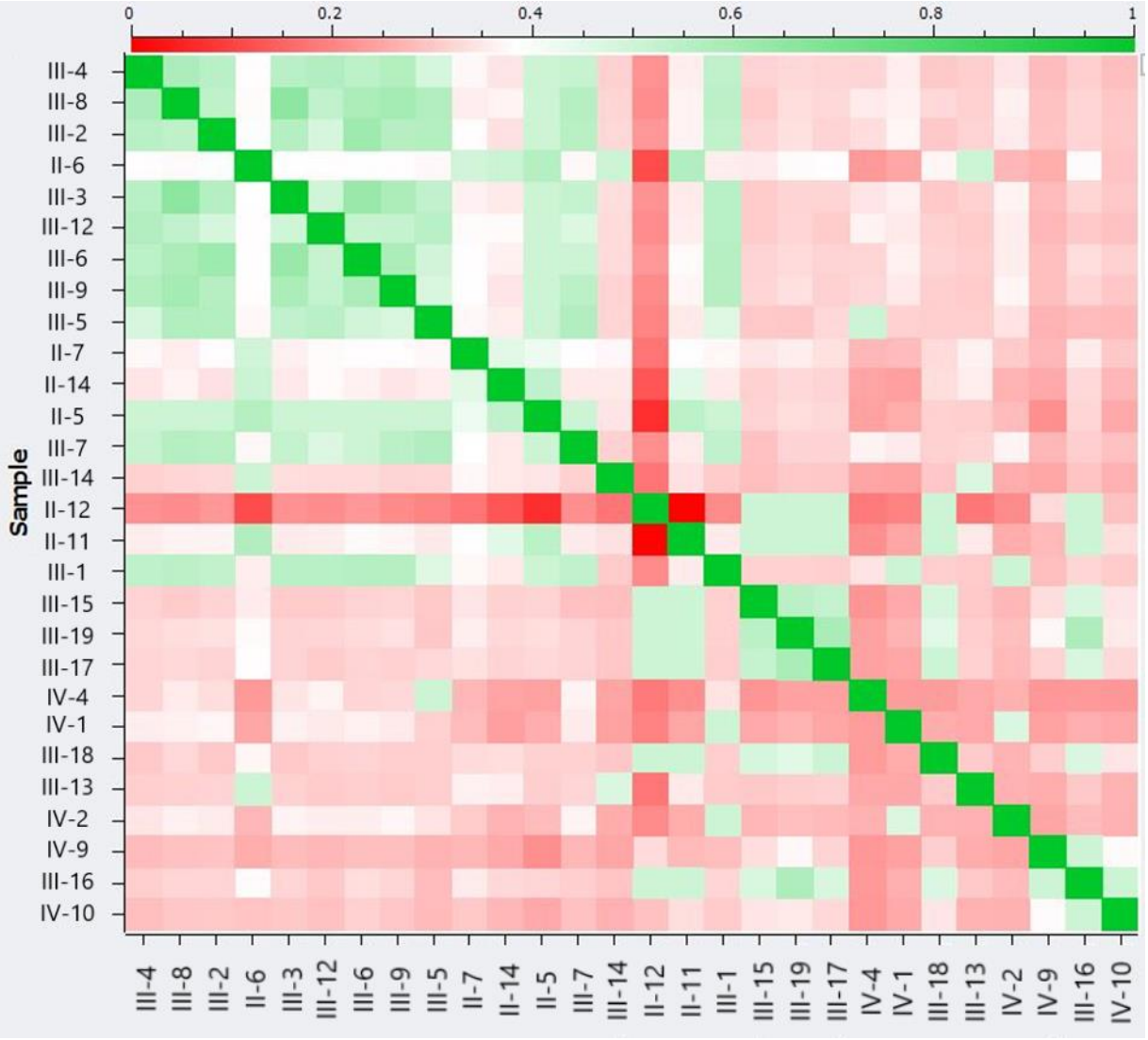
Abbreviations: OR, odds ratio; CI, confidence interval; dx, age at diagnosis; HL, Hodgkin lymphoma.

Appendix D - Supplementary materials for Chapter 6

D.1 Supplementary figures



Supplementary Figure D.1: Proportion of IBD1 and IBD2 segments among all pair-wise relationships.
Abbreviations: IBD1, identity by descent on one chromosome; IBD2, identity by descent on two chromosomes; MZ, monozygotic twins; PO, parent-offspring; FS, first degree relative (siblings); 2nd, second degree relative; distant, a relative that is 3rd or 4th degree; UN, unrelated.



Supplementary Figure D.2: A heat-map of IBD showing the relatedness level colour comparison between samples.