DEPARTMENT OF MEDICINE SOLNA,
CLINICAL EPIDEMIOLOGY UNIT
Karolinska Institutet, Stockholm, Sweden

FAMILY HISTORY, GENES AND GENE-ENVIRONMENT INTERACTION IN RISK AND PROGNOSIS OF NON-HODGKIN LYMPHOMA

Fredrik Baecklund

Karolinska Institutet

Stockholm 2017

All previously published papers, illustrations and figures were reproduced with permission from the publisher.
Published by Karolinska Institutet
Printed by E-print AB 2017
© Fredrik Baecklund, 2017
ISBN 978-91-7676-546-3

Family history, genes and gene-environment interaction in risk and prognosis of non-Hodgkin lymphoma

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Fredrik Baecklund

Principal Supervisor:
Associate Professor Karin Ekström Smedby
Karolinska Institutet
Department of Medicine Solna,
Clinical Epidemiology Unit

Co-supervisors:
Professor Johan Askling
Karolinska Institutet
Department of Medicine Solna,
Clinical Epidemiology Unit

Professor Richard Rosenquist
Uppsala University
Department of Immunology, Genetics and Pathology, Science for Life Laboratory

Opponent:
Associate Professor Martin Hutchings
Copenhagen University Hospital
Department of Hematology

Examination Board:
Senior lecturer Fredrik Wiklund
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Associate Professor Olof Akre
Karolinska Institutet
Department of Molecular Medicine and Surgery

Associate Professor Hans Hagberg
Uppsala University
Department of Immunology, Genetics and Pathology
Division of Experimental and Clinical Oncology

To my beloved family and friends. And to all good people in the world.

ABSTRACT

Non-Hodgkin lymphoma (NHL) is a heterogeneous group of malignant diseases arising from lymphocytes. There are about 40 subtypes according to the World Health Organizations classification system, differing in morphology, immunophenotype, genetic and/or clinical features.

Immunodeficiency, family history, autoimmune disease and some infectious agents are established risk factors for NHL. For most incident NHL patients, however, the reason is not known. One of the most common NHL subtypes, follicular lymphoma (FL), has been associated with smoking in most but not all studies. FL risk is strongly associated with germline genetic variation in the human leukocyte antigen (HLA) DRB1 gene, and amino-acid variation in certain positions of the HLA-DRB1 molecule. In Study I, we used a case-control study design to test the hypothesis that smoking is a risk factor for FL among individuals with certain amino-acid combinations in HLA-DRB1, including in positions 70-74 (known as the shared epitope (SE) in rheumatoid arthritis). We found that individuals that carried two SE alleles and were former or current smokers had approximately 2 and 3,5 times increased risk of FL, respectively, as compared to individuals that carried zero SE alleles and never smoked. The interaction was significant when assessed by estimating the attributable proportions of interaction ($0.15 \leq AP_{overall} \leq 1.0$; $0.005 \leq AP_{women} \leq 1.0$). This finding provides further evidence for a role of smoking in follicular lymphomagenesis, and offers a new model to explore FL etiology.

The prognosis of NHL not only varies by subtype, but also within subtypes for reasons that we do not fully understand. Prognostic indices are commonly used in NHL to guide choice of treatment and predict disease course. These include the host factors age and performance status along with indicators of tumor burden, and provide a rough estimation of outcome. Other host factors such as lifestyle factors and medical history as well as host genetic variation are being explored for their potential contribution to better prognostic prediction. This was the focus of Study II-IV.

In Study II, we tested the hypothesis that smoking status, attained education (proxy for socioeconomic status), body mass index, ultraviolet radiation exposure, autoimmune disease or family history of hematopoietic malignancy influenced overall and lymphoma-related survival among NHL patients overall and major subtypes. We found evidence that current smoking, few years of attained education and history of autoimmune disease increased the risk of all-cause death among NHL patients overall. Smoking also increased the risk of death among patients with diffuse large B-cell lymphoma. Attained education was also associated with lymphoma-related death. Further studies are needed to understand mechanisms for these associations and how they should be accounted for in the clinical setting.

In Study III, we tested the hypothesis that single nucleotide polymorphisms (SNPs) have an impact on FL survival or progression by 1) conducting a meta-analysis of two GWAS, and 2) exploring 22 SNPs previously reported to be associated with FL outcome. In the meta-analysis, no SNP was associated with FL survival at genome-wide significance, although one SNP in the ABCA10 gene on chromosome 17 was borderline associated (rs10491178: $P_{random}= 5.24 \times 10^{-8}$). In line with previous studies, two linked SNPs in *IL8* and one in *CD46* were negatively associated with FL progression. Our results mainly provide further evidence of an impact of SNPs involved in immune functions on follicular lymphoma outcome.

In Study IV, we further investigated the hypothesis that inherited factors influence survival among patients with a lymphoid malignancy by testing concordance in survival between two first-degree relatives with lymphoid malignancies. We found that individuals with a first-degree relative with good prognosis had a better survival than individuals with a first-degree relative with expected or poor survival among individuals with the same broad type of lymphoid malignancy, especially indolent lymphomas. Our results support a role of heritability in the outcome primarily of indolent lymphomas.

LIST OF SCIENTIFIC PAPERS

I. **Possible interaction between cigarette smoking and human leukocyte antigen DRB1 variation in risk of follicular lymphoma**

Fredrik Baecklund, Jia-Nee Foo, Johan Askling, Sandra Eloranta, Ingrid Glimelius, Jianjun Liu, Henrik Hjalgrim, Richard Rosenquist, Leonid Padyukov, Karin E. Smedby
Accepted for publication in American Journal of Epidemiology

II. **Lifestyle factors, autoimmune disease and family history in prognosis of non-Hodgkin lymphoma overall and subtypes**

Julia Simard, Fredrik Baecklund, Ellen Chang, Eva Baecklund, Henrik Hjalgrim, Hans-Olov Adami, Bengt Glimelius, Karin E. Smedby
International Journal of Cancer, 2013, 132(11), 2659-66

III. **A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival**

Fredrik Baecklund, Jia-Nee Foo, Paige Bracci, Hatef Darabi, Robert Karlsson, Henrik Hjalgrim, Richard Rosenquist, Hans-Olov Adami, Bengt Glimelius, Mads Melbye, Lucia Conde, Jianjun Liu, Keith Humphreys, Christine Skibola, Karin E. Smedby
BMC Medical Genetics, 2014, 15(1), 113

IV. **Concordance in survival of first-degree relatives diagnosed with a lymphoid malignancy**

Fredrik Baecklund, Sara Ekberg, Richard Rosenquist, Johan Askling, Sandra Eloranta, Karin E. Smedby
Manuscript

CONTENTS

LIST OF ABBREVIATIONS

NHL	Non-Hodgkin lymphoma
WHO	World Health Organization
HLA	Human leukocyte antigen
SNP	Single nucleotide polymorphism
UV	Ultraviolet
GWAS	Genome-wide association study
RA	Rheumatoid arthritis
SLE	Systemic lupus erythematosus
EBV	Epstein-Barr virus
HIV	Human immunodeficiency virus
IPI	International prognostic index
HR	Hazard ratio
FLIPI	Follicular lymphoma international prognostic index
CI	Confidence interval
OR	Odds ratio

1 INTRODUCTION

Lymphoma is a heterogeneous group of malignant diseases arising in cells destined to become lymphocytes. Broadly, they are divided into Hodgkin (10%), B cell (80%) and T/NK cell (10%) lymphomas. B and T/NK cell lymphomas are collectively still often referred to as non-Hodgkin lymphoma (NHL), even though this term is no longer used in the most updated version of the World Health Organization (WHO) classification system of lymphoid malignancies.[1] NHL as a group is among the top 10 most common cancer types worldwide and is estimated to account for approximately 3% of the world's cancer-related deaths.[2] It is a heterogeneous group of 40 subtypes, however, with distinct biology and clinical behavior.[3] These are divided into precursor and mature lymphoid malignancies, distinguished by the absence (precursor) or presence (mature) of completely rearranged immunoglobulin (B cell neoplasms) or T cell receptor genes (T cell neoplasms). Of note, the most recent WHO classifications of lymphoid malignancies also include chronic lymphocytic leukemia and plasma cell tumors (multiple myeloma and plasmocytoma) among mature B cell lymphoma. Although NHL is the main focus of this thesis, in Study IV we include all lymphoid neoplasms, i.e. Hodgkin lymphomas as well as precursor and mature B and T/NK cell lymphomas.

The vast number of different NHL subtypes complicates epidemiological studies of their risk and prognosis, as one factor may be important to one subtype but not another, and each subtype by itself is more or less uncommon.[3] Another complicating factor is that the lymphoma classification system has changed several times throughout history, which makes comparisons over time complicated. For these reasons, epidemiological studies are still done on NHL as a group but increasingly by subtype with as much detail as possible. The relatively small number of patients diagnosed with a certain subtype each year has driven researchers in lymphoma epidemiology to international collaboration. The studies constituting this thesis were performed in this context. Hence, they were both performed among all lymphoid malignancies combined (Study IV), all NHL combined (Study II), as well as among subtypes of these with as much details as possible (Study II and IV). Two studies (Study I and III) focused completely on follicular lymphoma. Two of the studies were done in collaboration with researchers in other countries: Denmark (Study I and III), the United States of America (USA; Study III) and Singapore (Study I and III).

The last two decades, our knowledge have increased considerably about the genetic events and aberrations in the lymphocytes, as well as biological mechanisms in lymphomagenesis. We have also learnt about some host factors (genetic and environmental) that are consistently associated with increased risk of NHL in general and in specific subtypes. However, for most of our patients we still do not know what caused the lymphoma.

The prognosis of NHL is highly variable by subtypes but also within subtypes for reasons that are mostly unknown. Many new prognostic factors have been proposed but so far few have been validated. In clinical practice and clinical trials, international prognostic indexes based on some host (commonly age and performance status) and tumor factors (usually stage and other measures of disease burden) are the most widely used tools to stratify patients into prognostic groups. There is great overlap between these prognostic groups, however, and in consulting our patients we can only make general statements about the prognosis. Information on host factors other than age, sex and performance status would be readily available and provide a simple way to improve prognostic prediction, should they be proven to carry such information. This has led up to researches exploring host factors in lymphoma prognosis in recent years.

In Study I, we tested the hypothesis that smoking is a risk factor for follicular lymphoma in individuals carrying certain amino-acid combinations in the human leukocyte antigen (HLA) DRB1 molecule but not otherwise. In Study II, we investigated the potential prognostic impact of selected host factors among NHL patients and four subtypes. In Study III, we explored the potential prognostic impact of host genetic polymorphisms across the germline genome on follicular lymphoma outcome. In Study IV, we further explored the hypothesis that inherited genetic polymorphisms influence survival in patients with a lymphoid malignancy, by testing if there was concordance in survival time between two first-degree relative with such a malignancy.

2 BACKGROUND

2.1 LYMPHOCYTES

Lymphocytes are white blood cells that defend our bodies against infections, toxins and tumor development.[4] There are about 2 x 10^{12} lymphocytes in the human body, a cell mass comparable to that of the brain or liver.[4] The mature lymphocyte continuously circulate from the blood to the lymph system and back, awaiting contact with a foreign antigen (Figure 1).

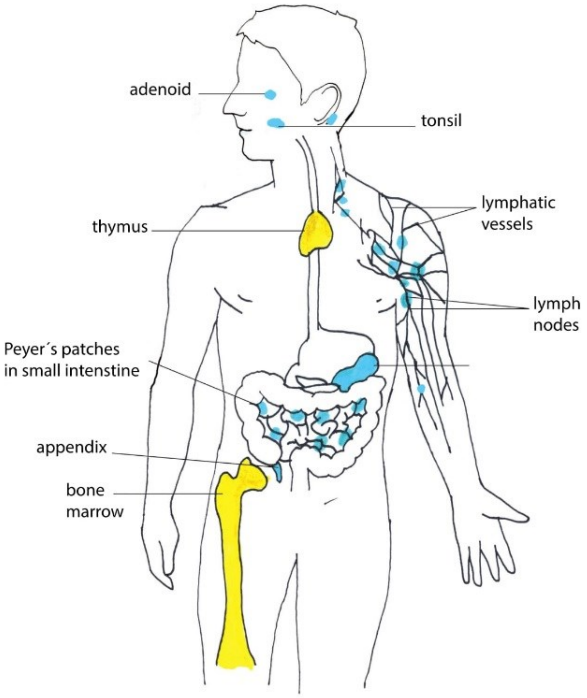


Figure 1. The lymph system and lymphoid organs.

Lymphocytes are divided into three functional types: B (bone marrow-derived) cells, T (thymus-derived) cells and NK (natural killer) cells (Figure 2). B cells make antibodies against extracellular antigens, such as bacteria and toxins.[4] Cytotoxic T cells (CD8+ T cells) attack normal cells infected by viruses or cells that have become neoplastic.[4] Helper T cells (CD4+ T cells) assist in the activation of B and cytotoxic T cells upon exposure to a foreign antigen.[4] While each B and T cell is specific to a certain antigen presented to them by HLA class I and II molecules (adaptive immunity), NK cells attack cells that show signs of stress, such as low expression of HLA molecules due to viral infection or malignant transformation (innate immunity).[4]

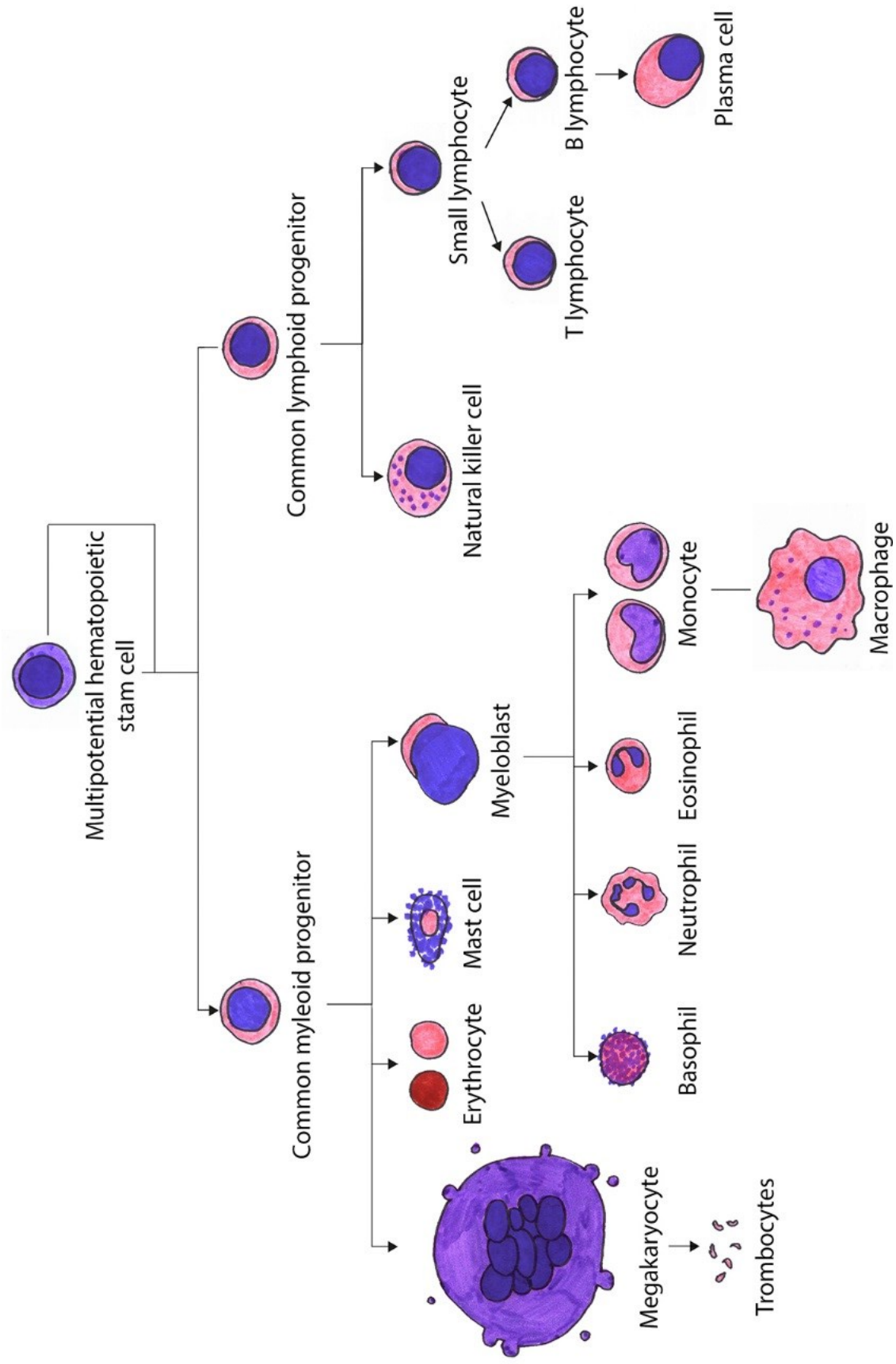


Figure 2. The hematopoiesis. Lymphoid malignancies originate from the lymphoid cell line, which starts with the common lymphoid progenitor.

2.2 ETIOLOGY OF NON-HODGKIN LYMPHOMA

The incidence of NHL increased considerably in many high income countries between 1950 and 1990.[5] The increase now seems to have levelled off but it sparked intense epidemiological research with the aim to explain it. When considering lymphoma risk factors, it is important to remember that lymphoma is a group of heterogeneous diseases that may or may not share risk factors with other lymphoma subtypes. Established risk factors are HIV infection, organ transplant, primary immunodeficiency disorders, autoimmunity and chronic inflammatory disorders, Helicobacter pylori, Hepatitis C virus, Human T cell leukemia/lymphoma virus 1, Human herpes virus 8 and breast implants. Host factors such as family history of lymphoma and germline genetic polymorphisms are also well-established risk factors of NHL subtypes. Lifestyle factors such as cigarette smoking, high body mass index and ultraviolet radiation constitute probable risk factors of NHL. The following sections discuss risk factors with relevance for this thesis, although other factors may also be implicated in NHL risk according to previous studies.[6-10]

2.2.1 Lifestyle factors

Recent pooled analyses from the InterLymph consortium are among the largest studies on lifestyle factors in risk of lymphoma to date.[6-10] Although other, smaller studies (some of which are included in the pooled InterLymph studies), exist, results from the InterLymph studies and the Scandinavian Lymphoma Etiology (SCALE) study (see 4.1 for details) are presented in Table 1 to illustrate the typical association between each factor and NHL risk.

2.2.1.1 Tobacco smoking

The pooled InterLymph analysis demonstrated that smoking was associated with risk of follicular lymphoma among women only but not among men and not with other subtypes or all non-Hodgkin lymphomas combined (Table 1).[7] This is in line with some previous smaller studies, including SCALE (Table 1).[11] The mechanism behind this association and why it differs by sex is unclear. Study I of this thesis addressed this issue by testing the hypothesis that smoking is a risk factor for follicular lymphoma among individuals that carry certain germline genetic variation in the HLA-DRB1 molecule but not otherwise. Given that lymphoma risk is increased in individuals with immune deficiencies, chronic inflammatory and autoimmune disorders (as outlined below), it is noteworthy that smoking has such effects on our immune systems (Figure 3).[12]

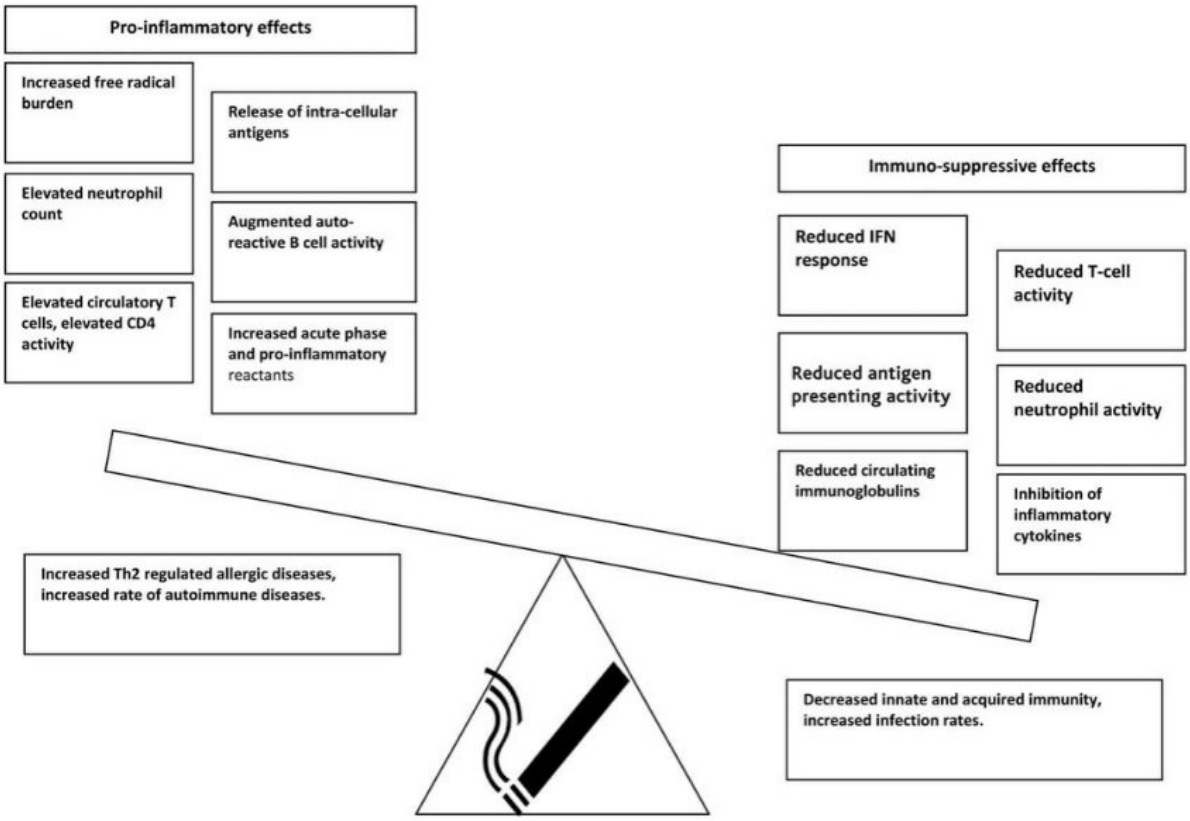


Figure 3. From Arnson, Shoenfeld and Amital, Journal of Autoimmunity 2010,[12] illustrating the effects of cigarette smoking on the immune system. Cigarette smoking both increases inflammatory allergic and autoimmune reactions and decreases activity against infections. Reprinted with permission.

2.2.1.2 Obesity

The parallel increase in NHL incidence and BMI in the populations in the western world led researchers to investigate whether there was an etiological link between the two. Several studies, including SCALE (Table 1),[13] have demonstrated an association between obesity and risk of NHL overall and diffuse large B cell lymphoma, while some studies have found no association (reviewed by Chiu and Hou, 2015[14]). The pooled analyses from InterLymph[6-10] confirms the association between high BMI in adulthood and risk of diffuse large B cell lymphoma but not the risk of all NHL combined (Table 1). High BMI in young adulthood resulted in larger relative risk estimates, however, and was also significantly associated with risk of NHL overall and follicular lymphoma (Table 1). Obesity is associated with low grade chronic inflammation[15] and thus sustained activation of the immune system. Given the well-established increase of lymphoma risk associated with chronic inflammatory disorders, as outlined above, the results from the InterLymph and other studies seem plausible.

2.2.1.3 Ultraviolet light

The incidence of skin cancers (malignant melanoma, squamous cell and basal cell carcinoma) increased in parallel with lymphoma incidence.[16] This fact together with the observation that NHL risk increased after the diagnosis of a skin cancer and vice versa (reviewed by Hu et al 2005[16]) led to the hypothesis that ultraviolet (UV) radiation could contribute to lymphomagenesis. Presumably to many researchers surprise, however, most studies support a protective effect of exposure to UV radiation on risk of NHL overall, diffuse large B cell lymphoma, follicular lymphoma and chronic lymphocytic leukemia (Table 1).[8,17] The mechanisms behind these observations are not clear. UV radiation of the skin has immune modulatory effects, including inhibition of antigen presentation, activation of innate immune responses and suppression of adaptive immune function. Whether these are mediated solely by inducing vitamin D synthesis or through other mechanisms is not completely understood.[18] Interestingly, lower doses of UV light ("suberythemal") seem to inhibit local immune responses, while higher doses ("erythemal") seem to achieve systemic immune effects.[18]

2.2.2 Host factors

2.2.2.1 Familial aggregation

Results from many case-control, cohort and registry-based studies suggest that an individual with a first-degree relative with a lymphoma has about two to three times the risk of lymphoma as compared to individuals without such family history (reviewed by Cerhan and Slager 2015[19]). Although these studies cannot exclude that shared environmental risk factors explain the observations, it is likely that germline polymorphisms contribute to lymphoma risk to some extent. The largest pooled analysis to date supports this notion for several NHL subtypes, including diffuse large B cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia and mantle cell lymphoma (Table 1).[8] Interestingly, the associations are stronger when looking at familiar cases with concordant NHL subtypes, suggesting that inherited genetic factors may be subtype specific.[19]

2.2.2.2 Germline genetic variation

Many candidate gene studies and genome-wide association studies (GWAS) have searched for and found germline genetic polymorphisms associated with risk of NHL overall and/or its subtypes (reviewed by Cerhan and Slager 2015[19]). To date, at least 67 SNPs (where one nucleotide in the DNA sequence is exchanged for another) in 41 loci have been identified in GWAS.[19] Few SNPs identified in candidate gene studies have been replicated in GWAS,[19] indicating many false positive findings with the former study design. Most of the 67 SNPs

	Reference study	Cases	Controls	Exposure	NHL Finding	DLBCL Cases	DLBCL Finding	FL Cases	FL Finding	CLL/SLL Cases	CLL/SLL Finding	MCL Cases	MCL Finding
Tobacco smoking	Schöllkopf et al 2005	3 055	3 187	Never vs. Ever	0.97 (0.87-1.08)	797	0.86 (0.72-1.02)	586	1.38 (1.01-1.88)	752	1.03 (0.85-1.24)	NA	
								307	**women**				
				Never vs. Ever					1.22 (0.94-1.58)				
								279	**men**				
									1.04 (0.64-1.70)				
	Morton et al 2014 InterLymph	17 471	23 096	Never vs. Ever	1.02 (0.97-1.07)	4 667	1.01 (0.94-1.08)	3 530	1.09 (1.00-1.18)	2 440	0.90 (0.81-0.99)	557	1.06 (0.87-1.29)
							women		**women**				
	Cerhan et al 2014 (DLBCL)	4 667	22 639										
	Linet et al 2014 (FL)	3 530	22 639	Never vs. Ever		1 889	1.05 (0.94-4.46)	1 506	1.22 (1.09-1.37)	2 440	0.90 (0.81-0.99)	557	1.06 (0.87-1.29)
	Slager et al 2014 (CLL/SLL)	2 440	15 186				**men**		**men**				
	Smedby et al 2014 (MCL) InterLymph	557	13 766	Never vs. Ever		2 250	0.98 (0.88-1.08)	1 487	0.98 (0.87-1.10)				
Obesity	Chang et al 2005 SCALE study	3 055	3 187	BMI <25.0 BMI ≥25.0	ref 0.9 (0.9-1.0)	795	ref 1.2 (1.0-1.4)	582	ref 0.9 (0.8-1.1)	750	ref 0.8 (0.7-1.0)	148	ref 1.2 (0.9-1.7)
	Morton et al 2014 InterLymph	17 471	23 096	Adult BMI <25 Adult BMI ≥25 Young adult BMI <25 Young adult BMI ≥25	ref 0.97 (0.86-1.10) ref 1.95 (1.51-2.53)	4 667	ref 1.32 (1.11-1.57) ref 3.02 (2.13-4.27)	3 530	ref 1.02 (0.84-1.25) ref 2.13 (1.44-3.14)	2 440	ref 0.80 (0.63-1.03) ref 1.80 (0.93-3.47)	557	ref 1.00 (0.64-1.55) ref 1.17 (0.42-3.22)
Ultraviolett radiation				**Sunbathing at 20 y**									
	Smedby et al 2005 SCALE study	3 055	3 187	Never ≤1 time/w 2-3 times/w ≥4 times/w	ref 0.8 (0.7-0.9) 0.7 (0.6-0.9) 0.7 (0.6-0.9)	796	ref 0.9 (0.7-1.2) 0.8 (0.6-1.0) 0.7 (0.5-0.9)	586	ref 0.6 (0.4-0.8) 0.7 (0.5-0.9) 0.6 (0.5-1.0)	752	ref 0.7 (0.6-0.9) 0.6 (0.5-0.8) 0.7 (0.5-0.9)	NA	
				Recreational sunexposure									
	Morton et al 2014 InterLymph	17 471	23 096	Q1-Q2 hours/w Q3-Q4 hours/w	ref 0.74 (0.66-0.83)	4 667	ref 0.75 (0.64-0.96)	3 530	ref 0.70 (0.58-0.96)	2 440	ref 0.79 (0.64-0.96)	557	ref 0.70 (0.48-1.01)
Socioeconomic status				**Childhood social environment**									
	Smedby et al 2007 SCALE study	3 055	3 187	≤1 individual/room >2-3 >3	ref 1.18 (0.98-1.43) 1.28 (1.02-1.61)	796	ref 1.26 (0.93-1.70) 1.47 (1.04-2.08)	586	ref 1.21 (0.86-1.70) 1.07 (0.70-1.64)	752	ref 1.17 (0.87-1.58) 1.31 (0.93-1.87)	NA	
				Attained education									
	Morton et al 2014 InterLymph	17 471	23 096	high vs. low	0.88 (0.83-0.93)	4 667	0.82 (0.76-0.90)	3 530	0.94 (0.85-1.03)	2 440	0.90 (0.79-1.01)	557	0.84 (0.68-1.06)
Family history of hematopoietic malignancy				**Family history**									
	Chang et al 2005 SCALE study	97	52	no vs. yes	1.8 (1.2-2.5)	24	1.8 (1.1-2.9)	22	2.2 (1.3-3.8)	27	2.0 (1.2-3.3)	NA	
	Morton et al 2014 InterLymph	967 A	795 A	no vs. yes	1.72 (1.54-1.93)	NA	1.57 (1.34-1.83)	NA	1.48 (1.25-2.84)	NA	2.17 (1.77-2.65)	NA	1.99 (1.39-2.84)

Table 1. Risk factors in NHL overall and subtypes in previously published studies.

are specific to subtypes.[19] Each SNP has low impact on NHL risk (relative risk of 0.5-2)[19] but carrying several risk SNPs could possibly lead to larger risk increases. Identifying risk SNPs may also eventually shed light on lymphoma etiology. It is noteworthy that the HLA region at 6p:21 is the most frequently recurring region across GWAS and lymphoma subtypes (Hodgkin, follicular, marginal zone, diffuse large B cell lymphoma and chronic lymphocytic leukemia).[19]

The SNPs most strongly associated with follicular lymphoma risk have been found in the HLA class II region.[19] Further analyses of this region indicate that the observed association may be explained by amino-acid variation in the antigen binding groove of the HLA-DRB1 molecule (Figure 4).[20,21] In particular, amino-acid positions 11, 13, 28 and 30 have been strongly and independently associated with follicular lymphoma risk.[20] These positions are important for antigen binding and presentation by specialized antigen presenting cells to helper T cells, implying importance of these immune functions in follicular lymphomagenesis. Interestingly, amino-acid variation in HLA-DRB1 is a strong risk factor for rheumatoid arthritis too (position 70-74, the shared epitope; Figure 4). Furthermore, this risk factor interacts with smoking in a profound way in rheumatoid arthritis (RA) risk.[22,23] This is the basis for the hypothesis tested in Study I.

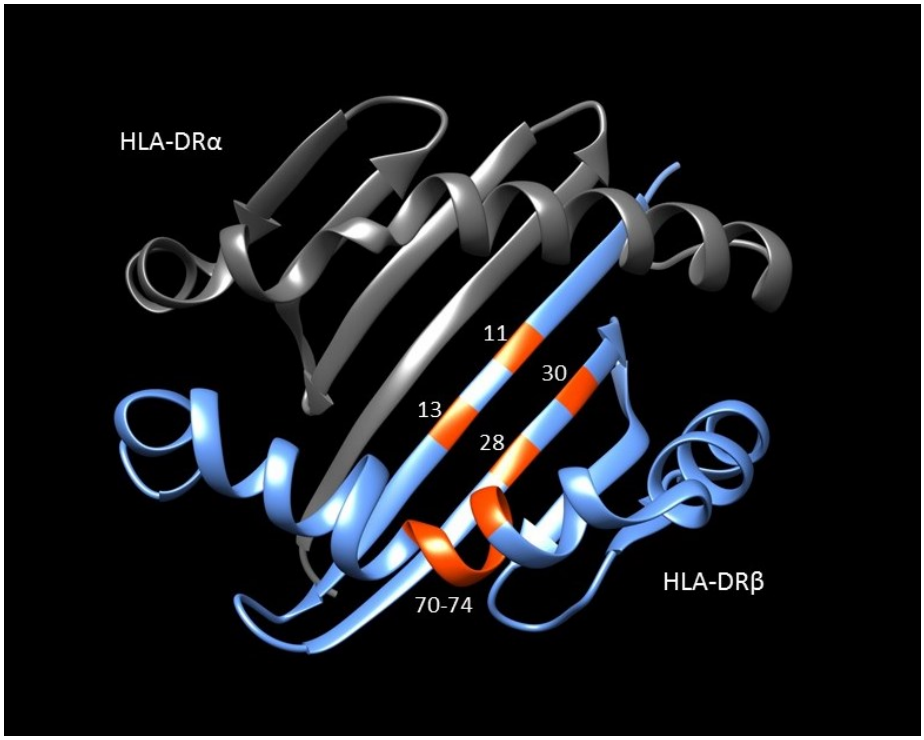


Figure 4. Three dimensional ribbon model of the HLA-DR antigen binding groove. Amino-acid positions 11, 13, 28 and 30 as well as the shared epitope (70-74) are highlighted. The figure was created using UCSF Chimera [24] and based on the Protein Data Bank entry 4GBX [25].

2.2.3 Genetic events

In lymphoma, like other tumors, malignant transformation events include alterations of the genome, such as gene amplifications, deletions and point-mutations.[26-28] In addition, most lymphomas carry chromosomal translocations and B-cell lymphomas in particular also aberrant somatic hypermutations, both of which arise by mechanisms fundamental for the normal B- and T-cell development.[26,27] The germline immunoglobulin and T cell receptor genes contain segments coding for the antigen recognizing part of the final molecule: the variable (V), diversity (D) and joining (J) segments (Figure 5 A).[29,30] During normal lymphocyte development, rearrangement of these gene segments is a prerequisite to produce a large repertoire of B and T cells with distinct immunoglobulin and T cell receptor specificities.[29,30] The downside of this cutting and pasting of genes is the risk of mistakes that sometimes lead to the placement of a promotor and/or enhancer juxtaposed to an oncogene, causing dysregulated gene expression of the latter (Figure 5 B).[26,27] Additionally, in normal B-cell development, somatic hypermutation of the variable region of the immunoglobulin gene further increases antigen affinity.[26] However, aberrant somatic hypermutations in other genes occur in B cell lymphomagenesis, sometimes altering the regulation and/or function of the affected genes or gene products.[26]

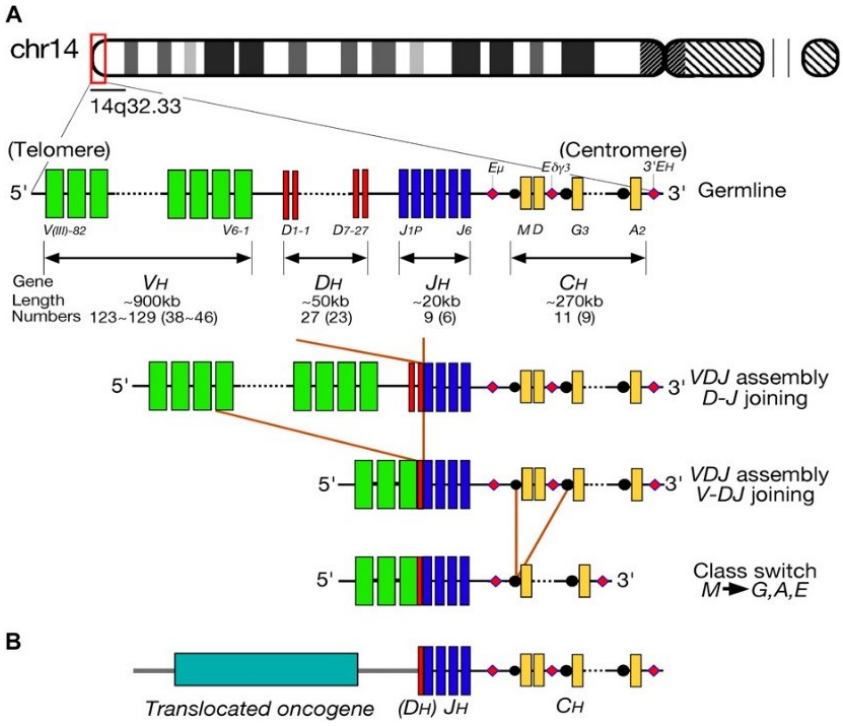


Figure 5, from Dyer et al, Blood 2010.[31] (A) The normal immunoglobulin heavy chain locus at 14q. The V, D and J segments are recombined to produce the gene for specific antigen recognition. The VDJ unit is further combined with constant (C) gene segments. (B) The normal process of cutting and pasting segments of the immunoglobulin genes (or the T cell receptor gene in T cells) involves a risk of errors that sometimes places an oncogene under the control of the promotor or enhancers of an immunoglobulin gene, leading to aberrant expression of the former. Reprinted with permission of ASCO 2017.

2.2.4 Microenvironment

The microenvironment of pre-malignant (and malignant) lymphocytes contains normal inflammatory cells that interact with the neoplastic cells through cell-cell contact and soluble mediators such as cytokines and chemokines (Figure 6).[32] This interaction stimulates cell survival, cell growths and invasion, induce angiogenesis and suppresses anti-tumor immunosurveillance.[32,33] Important oncogenic pathways are activated, including PI3K/AKT, STAT3, Hh and NF-κB.[32] Although the exact mechanisms are not yet known, the link between the microenvironment and lymphomagenesis is clear and is believed to be as important as somatic genetic aberrations for the development of many lymphoma subtypes.[34]

2.2.5 The B cell receptor

The B cell receptor is essential for normal B cell development and maturation.[35] It is also increasingly recognized as essential in the development (and sustainment) of many B cell malignancies.[35] Signals through the B cell receptor pathways stimulate survival, proliferation, migration and resistance to apoptosis.[35] In lymphomagenesis, these pathways may be constitutively activated by chronic antigen simulation by microbial antigens (e.g. Helicobacter pylori in gastric MALT lymphoma) or self-antigens (e.g. myosin in chronic lymphocytic leukemia), or activating mutations in intracellular components of the B cell receptor (e.g. *CARD11* in activated B-cell-like (ABC) diffuse large B cell lymphoma).[35,36] The continuous B cell receptor signaling creates a permissive environment for genetic instability and accumulation of mutations that may lead to tumor initiation and progression.

2.2.6 Autoimmune and chronic inflammatory disorders

Autoimmune and chronic inflammatory disorders such as RA, systemic lupus erythematosus (SLE), primary Sjögren's syndrome and celiac disease increase the risk of lymphoma several times as compared to the general population.[37] Interestingly, the more systemic inflammatory conditions RA and SLE have been associated mainly with diffuse large B cell lymphoma, while the more localized inflammatory conditions primary Sjögrens syndrome

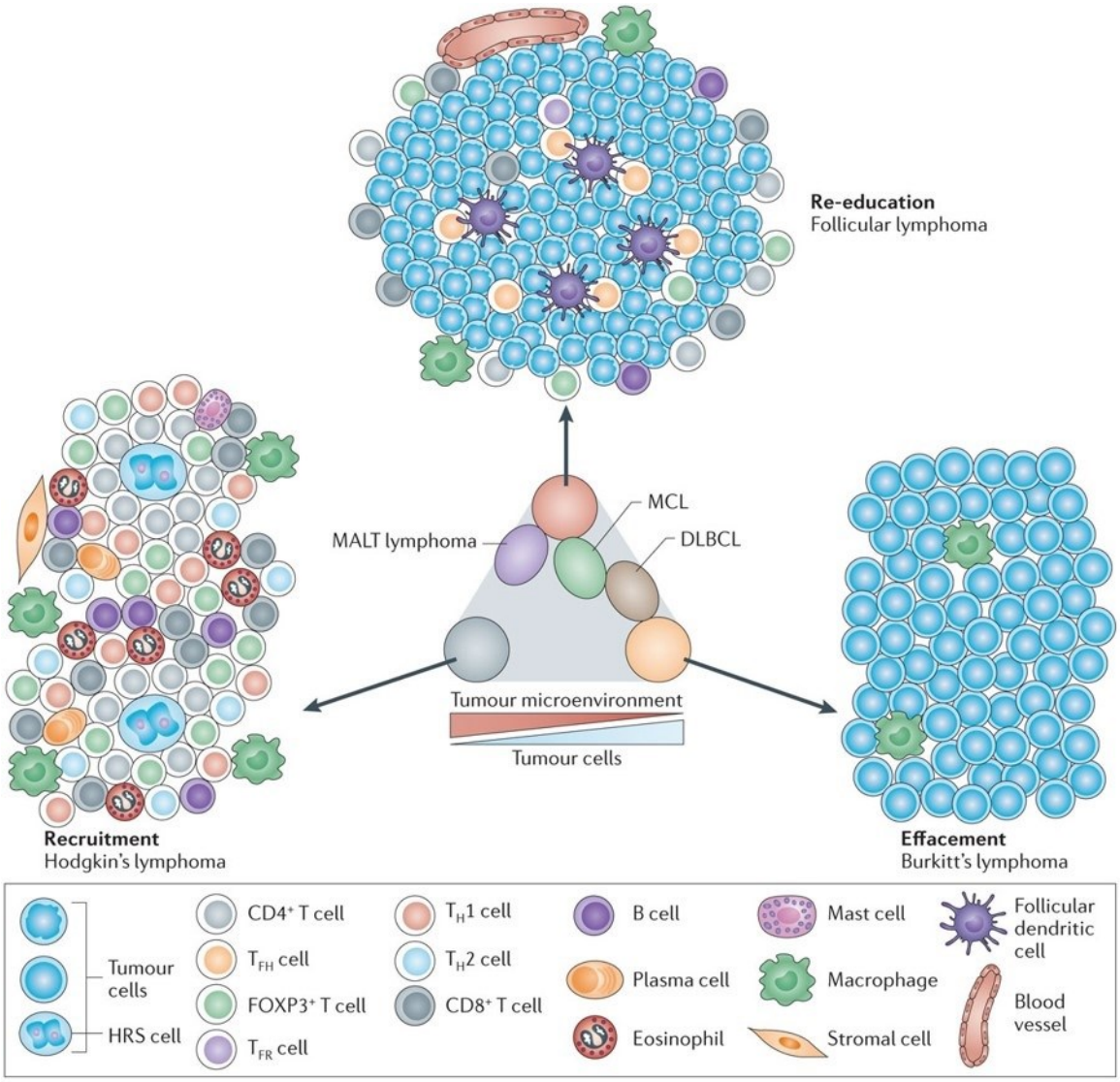


Figure 6, from Scott and Gascoyne 2014,[38] illustrating the typical microenvironment of the three B cell lymphomas that represent the extremes of the spectrum of tumor microenvironment (referred to as "recruitment", "re-education" and "effacement"). The ratio of malignant to normal cells increases across the range from classical Hodgkin lymphoma (approximately 1% tumor cells) to Burkitt lymphoma (>90% tumor cells). The ratio correlates with the dependence of the tumor cells on support from normal cells in the microenvironment. Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: Nature Reviews Cancer,[38] copyright 2017.

and celiac disease are associated with marginal zone lymphoma of MALT type in salivary glands and enteropathy-associated T cell lymphoma in the intestine, respectively.[37] This observation suggests different mechanisms of lymphomagenesis for systemic and local inflammatory conditions. The biological mechanisms behind the increased risks are not known but available evidence suggest a role for sustained activation of B and T cells due to chronic local or systemic inflammation as well as chronic (auto)antigen stimulation, stimulating lymphocyte survival and proliferation.[37]

2.2.7 Viruses

The Epstein-Barr virus (EBV), human T cell leukemia-lymphoma virus 1 and hepatitis C virus are closely associated with several types of lymphoma.[39-41] The best known is EBV, found in close to all endemic Burkitt lymphoma, a subset of classical Hodgkin lymphomas and many post-transplant lymphomas.[40] Orally transmitted ("kissing disease"), EBV causes a latent infection of B cells in most people in the world.[40] In this perspective, EBV associated lymphomas are rare accidents of EBV colonization.[40] The expression of viral proteins provides an anti-apoptotic function contributing to endemic Burkitt and Hodgkin lymphomagenesis, and drive B cell transformation in post-transplant lymphoma.[40]

2.2.8 Immunosuppression and immunodeficiency

The small part of the population affected by immunosuppression or immunodeficiency are at increased risk of malignancies, especially lymphomas. Such conditions include human immunodeficiency virus (HIV) infection, immunosuppressing treatment after organ transplant procedures and primary immunodeficiency syndromes such as primary variable immunodeficiency.[42-44] The risk of lymphoma among HIV infected individuals used to be 75 to 100 times that of the general population but has decreased substantially since the introduction of combined antiretroviral therapy.[43,45] Nevertheless, increased lymphoma risk remains and is still the most common malignancy in this population.[43] In organ transplant recipients, the risk of lymphoma is increased 30-50 times compared to the general population.[45] The risk increase depends on degree and duration of immunosuppression, as illustrated by a higher incidence rate in heart, lung and multi-organ transplants (8-25%) than in kidney and liver transplants (1-5%).[44] Similarly, the risk of lymphoma varies by type of primary immunodeficiency disorder but is generally much higher than in the general population (25 to several 100-fold risk increase).[42] The risk increase is primarily due to increased risk of aggressive lymphoma subtypes.[46,47]

The increased lymphoma risk in the context of immunosuppression and immunodeficiency is believed to be due to reduced immune surveillance, uncontrolled infections of oncogenic viruses such as EBV and chronic antigen stimulation of lymphocytes.[42,43,45] Germline mutations in DNA repair genes may also contribute to lymphomagenesis in primary immunodeficiency syndromes.[42]

2.2.9 Breast implants

Rare cases of anaplastic large cell lymphoma (ALK negative) associated with breast implants have recently been reported.[48] Intriguingly, the vast majority of primary lymphoma of the breast are of B cell origin, while breast implant associated lymphomas are of T cell origin,[49] suggesting a causal role of the implant itself. The pathogenesis is not clear. The implants seem to cause a local chronic inflammation.[48,50,51] It has been speculated that the inflammation is antigen-driven, with antigens derived from the implant itself or possibly from biofilms produced by local infection.[51,52] This would in turn cause sustained T cell activation and in some cases formation of a T cell lymphoma.[51,52]

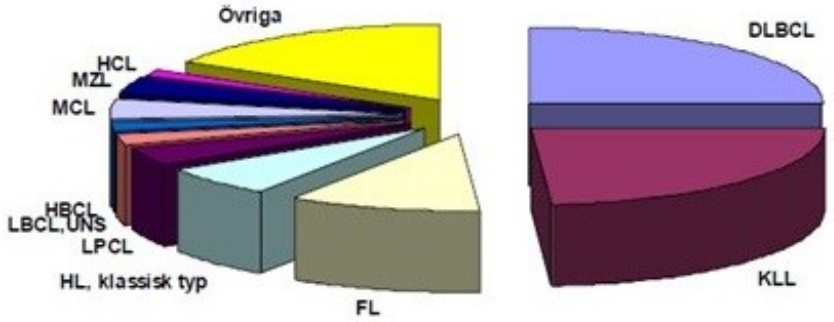
2.3 CLASSIFICATION OF LYMPHOMA SUBTYPES

Numerous classification systems have been used throughout history to define lymphoma subtypes.[53] At the moment, we use the fourth edition of the WHO classification system of lymphoid malignancies published in 2008.[3] A major revision is just being published, which will not be a fifth edition due to technical reasons, however (there are still other volumes pending in the fourth edition of the WHO tumor monograph series).[1]

The cell-of-origin perspective is a major basis for the WHO classification system of lymphomas.[3] It postulates that normal lymphocyte differentiation may get arrested at any stage by transformation events.[29] At each stage of maturation, the normal lymphocyte displays characteristic cell surface markers and changes in the B or T cell receptor genes. These characteristics will in many aspects be retained by the malignant B or T cell counterpart of each stage.[29] However, not all lymphoma subtypes have a clear normal counterpart.[53] Also, a group of lymphomas corresponding to a certain cell-of-origin may demonstrate important differences with respect to clinical features.[53] Hence, the current WHO classification system also takes into account clinical and genetic features when defining lymphoma subtypes.[3]

The WHO classification system identifies more than forty lymphoma subtypes, of which the most common are shown in Figure 7.[1] The subtypes are broadly categorized into Hodgkin lymphoma (10%), B cell lymphoma (80%) and T/NK cell lymphoma (90%), the latter two together often referred to as non-Hodgkin lymphoma. It is often also useful to classify B and T/NK cell lymphomas as indolent or aggressive.[54] Indolent lymphomas grow slowly. In the vast majority of cases, the disease is not curable but survival is nevertheless often counted in many years.[54] Treatment is offered upon symptoms and usually leads to temporary

remission.[54] Aggressive lymphomas grow fast and cause death within months without treatment.[54] With intensive treatment, however, the majority of the patients are cured.[54]



Abbreviation	Subtype	%
DLBCL	Diffuse large B cell lymphomas	25.5
KLL	Chronic lymphocytic leukemia	24.0
FL	Follicular lymphoma	11.0
HL, klassisk typ	Hodgkin lymphoma, classical	6.5
MCL	Mantle cell lymphoma	4.2
LPCL	Lymphoplasmacytic lymphoma	4.0
MZL	Marginal zone lymphoma	3.5
LBCL	Indolent B cell lymphoma	2.3
HBCL	Aggressive B cell lymphoma	2.1
HCL	Hairy cell leukemia	1.5
Övriga	Other subtypes	15.4

Figure 7. The ten most common lymphoma subtypes in the Swedish Lymphoma Register 2003-2005.[55]

2.3.1 Diffuse large B cell lymphoma

The diffuse large B cell lymphoma subtype is investigated in Study II and IV. It is an aggressive lymphoma and the most common NHL subtype (Figure 7). It is a heterogeneous group including a number of entities differing in morphology, genetics and/or biological behavior, such as diffuse large B cell lymphoma not otherwise specified (including germinal center B cell like and activated B cell like types), primary mediastinal large B cell lymphoma, high grade B cell lymphoma with MYC and BCL-2 and/or BCL-6 rearrangements and T cell histiocyte rich large B cell lymphoma.[1] They usually present with fast growing tumors in lymph nodes, spleen, liver, bone marrow or other organs.[3] The tumor cells originate from a mature B cell. Morphologically, the cells are large, grow diffusely and express B cell markers such as CD19, CD20, CD22 and CD79a.[3] Approximately 70% express the B cell leukemia/lymphoma (BCL) 6 oncogene.

2.3.2 Follicular lymphoma

Follicular lymphoma was the focus of Study I and III and was also included among other subtypes in Study II and Study IV. Follicular lymphoma is one of the most common indolent NHL subtypes (Figure 7).[3] It usually presents with enlarged lymph nodes but extranodal involvement, especially of the bone marrow, is common.[3] Clinical behavior is variable, reflecting the heterogeneity of the underlying biology. The tumor cells are derived from germinal center B cells, both centrocytes and centroblasts. They almost always form a nodular growth pattern.[3] The tumor cells are typically positive for CD19, CD20, CD79a, CD21 and CD10 but lack CD5, CD43 and CD11c.[3] Almost all tumor cells express HLA-DR.[3] More than 90% carry the t(14:18) translocation, causing overexpression of the BCL-2 oncogene.[3]

2.3.3 Chronic lymphocytic leukemia and small lymphocytic lymphoma

Chronic lymphocytic leukemia was included in Study IV. Chronic lymphocytic leukemia and small lymphocytic lymphoma are different manifestations of the same disease and together constitute the most common type of indolent lymphoma (Figure 7). They are characterized by the accumulation of mature, clonal B cells in the bone marrow, blood and/or lymphoid organs.[3] A chronic lymphocytic leukemia diagnosis requires $\geq 5.0 \times 10^9$ B lymphocytes/liter in peripheral blood with the characteristic immunophenotype CD19+, CD20+, CD23+ and CD5+.[3] The disease is denoted small lymphocytic leukemia when the tumor cells have a predilection for lymphoid organs and the B lymphocyte count in the blood is $< 5.0 \times 10^9$. In this thesis, for convenience, both manifestations of the disease will be referred to as chronic lymphocytic leukemia.

2.3.4 Mantle cell lymphoma

Mantle cell lymphoma was studied as a separate entity in Study II and included among aggressive lymphomas in Study IV. Mantle cell lymphoma is a fairly uncommon subtype of mature B cell lymphoma, comprising 3-10% of all NHL (Figure 7).[3] It is often viewed upon as intermediate between indolent and aggressive with regard to its biology; like indolent lymphomas it is incurable in the majority of cases, but has a more aggressive biology.[3] Mantle cell lymphoma usually presents with widespread lymph node involvement, often also engaging bone marrow and spleen.[3] The malignant B cells are characterized by the t(11:14) translocation, causing overexpression of cyclin D1.[3]

2.3.5 T cell and NK cell lymphomas

T/NK cell lymphomas were combined into one group in Study II and IV. T cell and NK cell lymphomas constitute approximately 10% of non-Hodgkin lymphomas (7% in the SCALE cohort, see 4.1) and form a heterogeneous group of diseases.[3] Hence, each subtype of T/NK cell lymphoma is a rare entity with the consequence that less is known about them than the more common B cell subtypes. They have in common that they arise from cells of T or NK cell origin and, thus, (variably) express T cell markers CD3, CD2, CD5 and/or CD7 and are negative for B cell markers (CD19, CD20, CD23, CD79a/b).[3] Most cells carry clonal T cell receptor rearrangements.[3] They may present clinically in many different ways, depending on subtype. The spectrum covers everything from indolent disease of the skin, as in mucosis fungoides, to aggressive disease engaging lymphoid tissue as well as extranodal sites as in anaplastic large cell lymphoma.[3]

2.4 PROGNOSTIC FACTORS IN NON-HODGKIN LYMPHOMA AND SUBTYPES

When discussing prognosis of NHL, it is important to remember that it is a group of heterogeneous diseases, differing in biology, clinical course, response to treatment and, hence, prognosis.

Several prognostic markers have been elaborated to help in patient counseling, guide treatment strategy and to stratify patients with NHL in clinical trials. Established prognostic factors differ to some extent by NHL but have in common that they are based on host factors (age and performance status) and disease related factors (stage, number of involved extranodal sites, level of lactate dehydrogenase, hemoglobin, and blood cell count).[56-59] The prognostic factors are combined into prognostic scores that stratify patients into risk groups with different expected survival.[56-59] However, there is still great variability in survival within and overlap between prognostic groups for reasons that we do not fully understand.

Treatment influences NHL survival too. In both aggressive and indolent lymphomas, the treatment backbone is still single or multi-agent chemotherapy. Since the beginning of the millennium, the addition of the monoclonal anti-CD20 antibody rituximab results in important survival benefits for patients with B cell lymphomas.[60-63] Radiotherapy is a local treatment of importance for selected patients. New small molecules targeting tyrosine kinases have become part of the treatment arsenal in chronic lymphocytic leukemia[64] but not yet other subtypes.[65] Immunotherapy, such as anti-CTLA4 or anti-PD1 antibodies, are investigational in NHL at this point.[65] Therapies such as these will change the way we treat NHL in the near future. This may have implications not only on the prognosis but also on what factors are relevant for predicting survival.

2.4.1 Diffuse large B cell lymphoma

First published in 1993, the international prognostic index (IPI) is still the main tool to stratify patients by expected prognosis.[57] Recently, gene expression profile studies and studies of other biological factors have identified other markers of importance for identifying patients with good or poor prognosis.

2.4.1.1 The International Prognostic Index

Since IPI was first proposed in 1993,[57] it has been re-evaluated and shown to be valid also in the rituximab era.[66] It consists of five host and tumor related factors:

- Age >60 years
- Lactate dehydrogenase greater than normal
- Eastern Cooperative Oncology Group performance status ≥2
- Ann Arbour stage III or IV
- Number of extra nodal sites involved ≥2

Each factor gives one point when present. Summing up the points (0 to 5) will stratify the patient into a prognostic group with different expected survival (Table 2).

Table 2. IPI score and corresponding risk groups. The progression-free survival (PFS) and overall survival (OS) estimates are from the rituximab era.[66]

IPI score	Risk group	3 year PFS, %	3 year OS, %
0-1	Low	87	91
2	Low-intermediate	75	81
3	High-intermediate	59	65
4-5	High	50	59

2.4.1.2 Cell of origin – gene expression profiles

Recently, two subtypes of diffuse large B cell lymphoma have been identified that have distinct gene expression profiles that correlate with prognosis.[67] The two gene expression patterns correspond to two different steps of normal B cell differentiation, germinal center B cell (GCB) and activated B cell (ABC).[67] When given similar treatments, patients with the ABC type fare much worse than patients with GCB type (Figure 8).

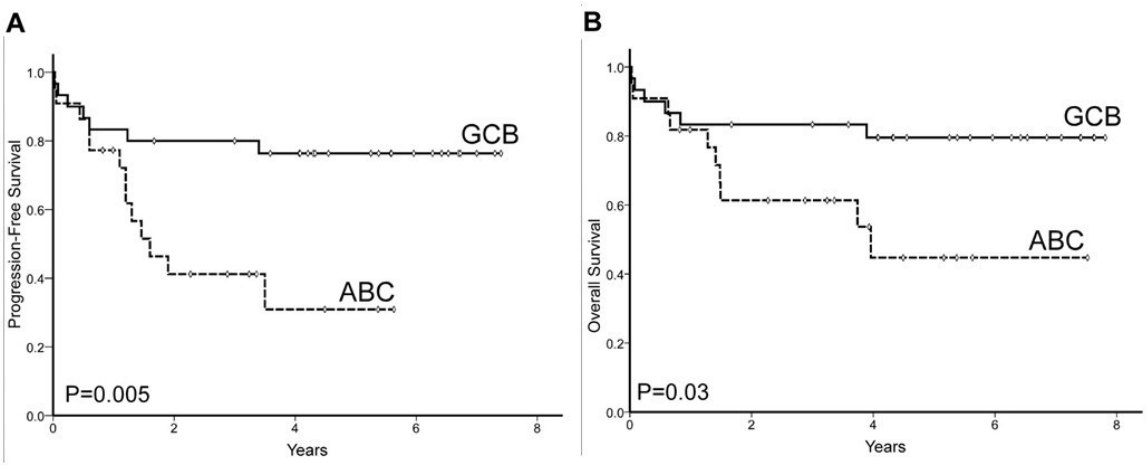


Figure 8, from Gutierrez-Garcia et al, Blood 2011,[67] illustrating how progression-free (A) and overall (B) survival differed between 157 patients with diffuse large B cell lymphomas demonstrating gene expression patterns of either germinal center B (GCB) type or activated B cell (ABC) type. Reprinted with permission of ASCO 2017.

2.4.1.3 MYC translocation and double hit diffuse large B cell lymphoma

Translocations involving the c-MYC oncogene are present in 5-10% of diffuse large B cell lymphomas. It is an independent prognostic factor associated with worse outcome when present (progression-free survival hazard ratio (HR)=3.28, P=0.003, overall survival HR=2.98, P=0.01).[68] Double hit refers to lymphomas that carry a translocation involving the c-MYC gene concurrently with a translocation of BCL-2 or, less commonly, BCL-6.[69] The double hit confers a worse prognosis than a c-MYC translocation alone.[69] A synergistic effect of MYC and BCL-2 overexpression may explain this observation.[69]

2.4.2 Follicular lymphoma

The follicular lymphoma International Prognostic Index (FLIPI) is the main tool used to predict disease course in follicular lymphoma.

2.4.2.1 Follicular Lymphoma International Prognostic Index

FLIPI was first evaluated in the pre-rituximab era[56] but has since been validated in the rituximab era.[70] FLIPI consists of five factors that, when present in a patient, results in one point each:

- Age >60 years
- Ann Arbor stage III or IV
- Hemoglobin level <120 g/L
- Number of involved nodal areas ≥5
- Lactate dehydrogenase greater than normal

Summing up the points (0-5), a patient is attributed to one of three risk groups with different prognosis: Low risk 0-1 points, intermediate risk 2 points, high risk 3-5 points (Figure 9)

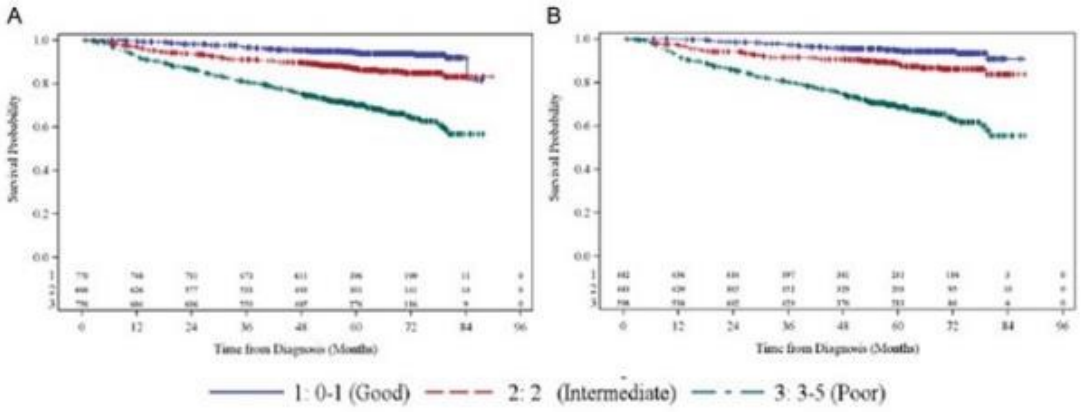


Figure 9, from Nooka et al, Annals of Oncology 2013,[70] illustrating differences in overall survival among (A) all included patients and (B) among patients for whom rituximab was part of the treatment. Reprinted with permission of Oxford University Press.

2.4.2.2 Tumor grade

Follicular lymphomas are graded on a scale from 1 to 3 based on the number of centroblasts per high power field (grade 1: 0-5 centroblasts, grade 2: 6-15 centroblasts, grade 3: >15 centroblasts).[3] Grade 3 has been further subdivided into 3a (centrocytes are present) and 3b (solid sheets of centroblasts).[71] The prognostic importance of tumor grade is limited, however. In practice grade 1 to 3a follicular lymphomas are considered indolent, whereas grade 3b follicular lymphomas are considered aggressive lymphomas and treated as such.

2.4.3 Chronic lymphocytic leukemia and small lymphocytic lymphoma

The Rai and Binet staging systems are the most widely used prognostic tools in the clinical setting.[58,59] Other factors used in the clinical setting are lymphocyte doubling time and certain genetic aberrations: deletion of 17p (del17p), deletion of 11q (del11q), trisomy 12 and deletion of 13q (del13q). Several other factors of less or uncertain importance have been or are under evaluation but will not be covered in this thesis.

2.4.3.1 Rai staging system

The Rai staging system is based on the idea that chronic lymphocytic leukemia progresses along a somewhat predictable path, starting in the blood and bone marrow (stage 0), going on to involving lymph nodes (stage I), spleen and liver (stage II) and eventually lead to bone marrow failure (stage III and IV).[59] The system was later modified to comprise three risk groups (low, intermediate and high) since there was no significant difference in survival between stage I and II and between stage III and IV.

Rai stage	Description	Modified Rai	Median survival in original Rai study
0	Lymphocytosis only	Low	12.5 years
I	Enlarged lymph nodes (> 1 cm)	Intermediate	8.4 years
II	Enlarged liver or spleen		5.9 years
III	Anemia (hemoglobin <110 g/L	High	1.6 years
IV	Thrombocytopenia (<100 x10^9 /L)		

2.4.3.2 Binet staging system

The Binet staging system counts lymphoid sites involved (cervical, axillary and inguinal nodes, liver, spleen) and each site gets equal weight.[58] The presence of anemia (hemoglobin <100 g/L) or thrombocytopenia (platelets <100 x10^9 /L) indicate a worse prognosis, like in the Rai system.

Binet stage	Description	Median survival in original Binet study
A	1 or 2 lymphoid sites involved	Comparable to age-matched controls
B	≥3 lymphoid sites involved	7 years
C	Anemia or thrombocytopenia	2 years

2.4.3.3 Lymphocyte doubling time

A lymphocyte doubling time (actual or estimated) of <12 months indicate a more aggressive disease course and worse prognosis compared to doubling time >12 months (Figure 10).[72]

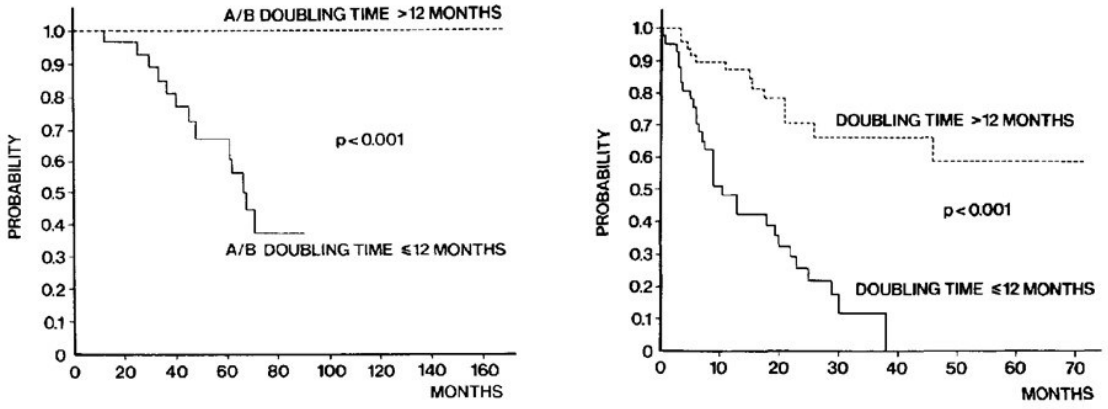


Figure 10, from Montserrat et al, British Journal of Hematology 1986.[72] The graphs shows the estimated survival time (left) and time to treatment (right) in chronic lymphocytic leukemia patient Binet stage A or B, stratified by lymphocyte doubling time. Reprinted with permission of John Wiley and Sons.

2.4.3.4 Genetic aberrations

The four genetic aberrations most commonly used in the clinical setting for prognostic prediction are (in order from worst to best prognosis): deletion of 17p, deletion of 11q, trisomy 12 and deletion of 13q (Figure 11).[73] Mutations in the TP53 gene has similar influence on survival as 17p deletion.[74] The prognostic value of these markers are changing with the arrival of new treatments. The prognosis of patients with 11q deletions has improved with the use of treatments like fludarabine, cyclophosphamide and rituximab.[75] With target therapies such as ibrutinib, the prognosis of patients with 17p deletions or TP53 mutations has improved but still remain significantly worse than that of patients without such aberrations.[76]

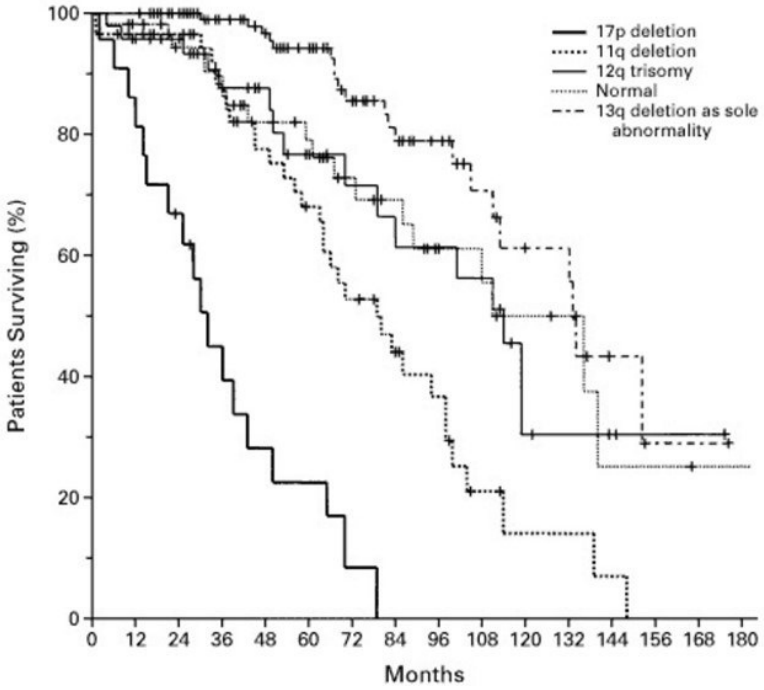


Figure 11, reproduced with permission from New England Journal of Medicine,[73] Copyright Massachusetts Medical Society. The graphs show the estimated survival curves for patients with either of the four genetic abnormalities 17p deletion, 11q deletion, 13q deletion or trisomy 12 or with a normal genotype.

2.4.3.5 Chronic lymphocytic leukemia international prognostic index

In 2016, a new prognostic score was proposed, the chronic lymphocytic leukemia international prognostic index (CLL-IPI).[77] CLL-IPI identifies five individual risk factors of poor survival that are assigned different weights:

- TP53 status (no abnormality versus 17p deletion or TP53 mutation or both)
- IgHV unmutated
- β2-microglobulin concentration >3.5 mg/L
- Binet B-C or Rai I-IV
- Age >65 years

After summing up the points (0-10), a patient is attributed to one of four risk groups with different prognosis: Low risk 0-1 point, intermediate risk 2-3 points, high risk 4-6 points, very high risk 7-10 (Figure 12).

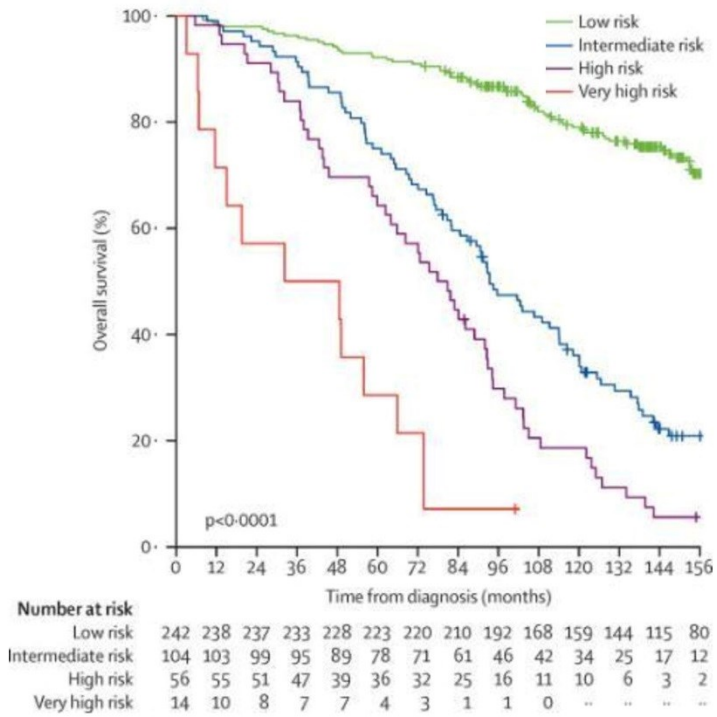


Figure 12, from the CLL-IPI study showing the estimated overall survival by CLL-IPI risk groups in one of the external validation cohorts, which happened to be SCALE (see 4.1).[77] Reprinted from The Lancet Oncology, Copyright 2017, with permission from Elsevier.

2.4.4 Mantle cell lymphoma

2.4.4.1 Mantle cell lymphoma international prognostic index

The mantle cell lymphoma international prognostic index (MIPI) may be used for prognostic prediction of patients with newly diagnosed mantle cell lymphoma.[78] It uses information on age, Eastern Cooperative Oncology Group (ECOG) performance status, lactate dehydrogenase ratio (LDH) and total white blood cell count to define three risk groups.[78]

Points	Age	Performance status	LDH value/upper normal limit ratio	White blood cell count
0	<50	0-1	<0.67	<6.7
1	50-59		0.67-0.99	6.7-9.9
2	60-69	2-4	1.00-1.49	10.0-14.9
3	≥70		≥1.50	≥15.0

MIPI has been validated in a more recent study where patients were given more modern treatments (survival estimates shown in the table below).[79]

MIPI Score	Risk group	5 year OS, %
0-3	Low	83
4-5	Intermediate	63
6-12	High	34

2.4.4.2 High proliferation rate, Ki-67 ≥ 30%

A high proliferation rate among the tumor cells, defined as the expression of the proliferation marker Ki-67 in ≥30% of tumor cells, has been shown to be an independent prognostic marker in mantle cell lymphoma.[80] A recent study suggested combining information on Ki-67 with MIPI for more accurate prognostic prediction (Figure 13).[80]

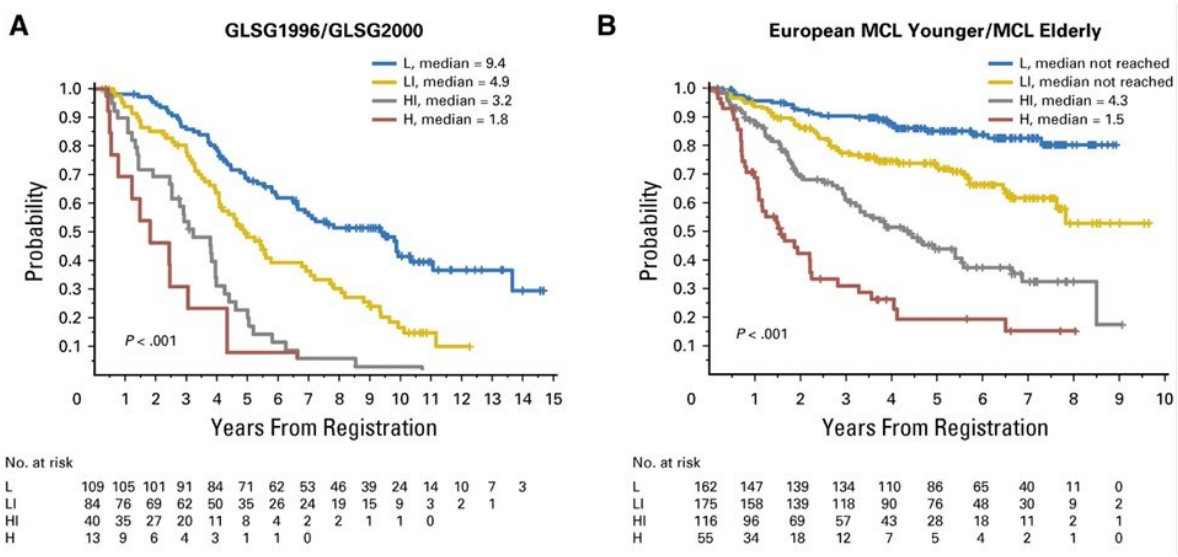


Figure 13, from Hoster el al, Journal of Clinical Oncology 2016.[80] Panels (A) and (B) show the estimated survival curves of two different cohorts of mantle cell lymphoma patients, each stratified by a modified MIPI score that also includes Ki-67 (using a cutoff of 30%). Reprinted with permission. © 2017 American Society of Clinical Oncology. All rights reserved

2.4.5 The B cell lymphoma microenvironment and survival

It is becoming increasingly clear that malignant B cells of different subtypes retain a dependence on normal cells and structures in the tumor microenvironment (reviewed by Scott and Gascoyne, 2014).[38] Cell-cell interactions and soluble mediators such as cytokines and chemokines stimulate survival and proliferation.[38] The ratio between malignant cells and normal cells differs by lymphoma subtype, as well as by degree of dependence of tumor cell on external signals and by the inflammatory response of the host (Figure 6, section 2.2.4).[38] Furthermore, evidence suggest that the composition of the tumor microenvironment also influences the prognosis of patients with B cell lymphoma. In diffuse large B cell lymphoma and follicular lymphoma, studies have shown a strong correlation between patient survival time and gene expression profiles emanating from normal cells in the tumor microenvironment.[81,82] In follicular lymphoma, the proportion of tumor infiltrating T cells has also been associated with prognosis.[83,84]

2.4.6 Host factors and non-Hodgkin lymphoma survival

The host factors age and performance status are included in all or most of the prognostic indexes described above. The last few years, there have been increasing interest in whether other host factors, including lifestyle factors, medical history and host genetic factors, could also hold independent prognostic information. We investigated the association between all-cause or lymphoma-related death and selected host factors (smoking, education, obesity, UV radiation, autoimmune disease and family history of hematopoietic malignancy) among patients with NHL, diffuse large B cell lymphoma, follicular lymphoma, mantle cell lymphoma and T cell lymphoma in Study II.

2.4.6.1 Smoking

Except for Study II, five other studies have investigated the association of smoking with NHL survival (Table 3).[85-89] Among the five studies, at least one of four smoking exposure variables used (smoking status, duration, intensity and pack-years) reached statistical significance in association with survival among NHL patients. Virtually all point estimates indicated increased risk of death irrespective of exposure measure used. One of the studies also performed a meta-analysis including four of the five studies.[87] In the meta-analysis, smoking duration, intensity and pack-years but not smoking status were significantly associated with increased risk of death among NHL patients. The same was observed among follicular lymphoma patients. Leo et al was not included in the meta-analysis, and reported no association between smoking status and follicular lymphoma survival. Among diffuse large B cell lymphoma patients, survival was associated with smoking duration in one study[88] and smoking status in another study[89] but the other three studies and the meta-analysis reported no association. Among patients with chronic lymphocytic leukemia, survival was associated with smoking intensity in one study[85] and smoking status in another study,[89] while a third study reported no association.[88] Regarding lymphoma-specific death, two studies wrote in the text that the estimates for this outcome measure were very similar to those for all-cause death but did not show any data.[85,88] Leo et al. provided data, however, and demonstrated an association with smoking status among patients with NHL overall or chronic lymphocytic leukemia but not with diffuse large B cell lymphoma or follicular lymphoma.[89]

Hence, smoking seems to increase the risk of death among NHL patients and follicular lymphoma patients just like it does in the general population. It is interesting to note that the t(14:18) translocation that occur in >90% of follicular lymphoma may be more common among “healthy” smokers than non-smokers.[90,91] Smoking is well-known to cause deadly lung and cardiovascular diseases in the general population,[92,93] which could account for the

excess risk of death among smoking NHL patients too. Why there would be a differences between the aggressive subtype diffuse large B cell lymphoma and the indolent follicular lymphoma is not obvious. Patients with indolent lymphomas are expected to live for years (>10 years in median) and will have time to die from other causes, such as smoking related morbidities. Possibly, smoking does not get the chance to exert its deadly effect in aggressive lymphoma patients since a good proportion of these (40%) will die from the lymphoma itself. On the other hand, patients with comorbid conditions (caused by smoking) may not tolerate the intensive treatment needed to cure an aggressive lymphoma very well. For this reason, the treating physician may give less intensive treatment, increasing the risk for relapse.[94] Intensive treatment may also aggravate the comorbid conditions and cause increased risk of all-cause death.[94,95]

2.4.6.2 Socioeconomic status

Two previous studies from Denmark and Scotland explored the impact of socioeconomic status on survival among NHL patients (Table 3).[96,97] The Scottish study used residential area as exposure measure, while the Danish study used several measures: attained education, cohabiting status, affiliation to work market and disposable income. Both studies found an increased risk of death among NHL patients with low as compared to high socioeconomic status, irrespective of measure used. With regard to NHL subtypes, the Danish study showed similar associations with all-cause death among patients with diffuse large B cell lymphoma and indolent lymphomas as those observed among NHL overall, although some estimates were not statistically significant. The Scottish study did not explore subtypes.

Both studies also demonstrated that NHL patients with low socioeconomic status had increased risk of having B symptoms and worse performance status compared to NHL patients with high socioeconomic status. In the Danish cohort, patients with low socioeconomic status also had significantly higher risk of other poor prognosis markers, including Ann Arbor stage III or IV, elevated lactate dehydrogenase, and IPI ≥2.[98] These observations indicate more advanced disease stage at diagnosis among patients with low socioeconomic status. A possible explanation for the higher rate of B symptoms, elevated lactate dehydrogenase and IPI ≥2 could be more aggressive disease biology among patients with low socioeconomic status. This hypothesis could make sense if exposure to environmental risk factors differ by residential areas or employment, and that this would have implications for lymphoma biology. Although there is no direct evidence of this, it is notable that several occupations that typically require low educational level have been associated with increased risk of diffuse large B cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia and mantle cell lymphoma.[6,8-10] Another possibility for the worse

performance status and survival observed could be due to more comorbidity due to smoking and/or obesity among individuals with low socioeconomic status.[99]

2.4.6.3 Obesity

Three studies apart from our Study II have investigated the impact of obesity on survival of NHL patients (Table 3).[86,89,100] All three reported an increased risk of all-cause death among obese as compared to normal-weight NHL patients, although the association in one study were just borderline statistically significant.[100] The estimates among patients with diffuse large B cell lymphoma and follicular lymphoma also indicated increased risk of death in association with obesity but none of the estimates reached statistical significance. Two of the three studies also looked at patients with chronic lymphocytic leukemia with conflicting results: a small study of 62 female patients found no association[100] while a larger study including 287 patients of both sexes found an increased risk of death in association with obesity.[89] Only one study investigated the potential impact of obesity on NHL-specific death and found increased risk among patients with NHL or chronic lymphocytic leukemia but not diffuse large B cell lymphoma or follicular lymphoma.[89]

The impact of obesity on risk and prognosis in several cancer forms has been extensively studied, and was recently the focus of a series of articles in Journal of Clinical Oncology.[101] The picture observed in several common cancer forms, like prostate, breast and colon cancer, is similar to the one described for NHL patients above. The reason why obesity would confer poorer survival among patients with lymphoma or other cancers is not clear but several hypotheses exist. Obesity itself contribute to insulin resistance.[101] This leads to hyperinsulinemia which stimulates increased levels of the hormone insulin-like growth factor-1, in turn stimulating cell growth and proliferation and inhibiting apoptosis. In the setting of insulin resistance, adipose tissue inflammation causes production of cytokines and increased levels of leptin and decreased levels of adiponectin. All these factors could potentially stimulate tumor growth and progression.[102,103]

2.4.6.4 Ultraviolet radiation and vitamin D

Two studies have investigated the impact of vitamin D level in the blood and survival among patients with different NHL subtypes (Table 3).[104,105] I was unable to identify studies

Table 3. Host factors in prognosis of NHL and subtypes in previous studies.

	Reference	Cases	Exposure	Outcome	NHL HR (95% CI)	DLBCL Cases	DLBCL HR (95% CI)	FL Cases	FL HR (95% CI)	CLL Cases	CLL HR (95% CI)
Tobacco smoking	Battaglioli 2006	1 138	Never vs. Current	All-cause death	1.04 (0.84-1.29)	338	0.87 (0.57-1.31)	118	0.89 (0.44-1.76)	298	1.22 (0.82-1.83)
			No vs. >39 years		1.26 (0.94-1.70)		1.70 (1.02-2.14)		3.0 (0.82-11.3)		1.19 (0.72-1.95)
			No vs. >31 pack-years		1.60 (1.18-2.18)		1.20 (0.65-2.23)		2.50 (0.82-7.61)		1.37 (0.79-2.40)
	Talamini 2008	268	Never vs. ≥20 cig/day	All-cause death	1.70 (1.06-2.73)	120	1.17 (0.64-23.1)	38	1.28 (0.29-5.68)	29	4.33 (1.09-17.2)
			No vs. ≥30 years		1.32 (0.85-2.05)		1.34 (0.75-2.40)		3.10 (0.74-13.1)		0.73 (0.34-2.90)
	Geyer 2010	471	Never vs. Current	All-cause death	1.50 (0.97-2.29)	170	1.22 (0.58-2.58)	123	2.95 (1.27-6.81)		
			No vs. ≥20 years		1.76 (1.25-2.47)		1.53 (0.87-2.69)		2.48 (1.11-5.52)		
			No vs. ≥17 pack-years		1.88 (1.33-2.66)		1.68 (0.95-2.95)		2.61 (1.19-5.74)		
	Ollberding 2013	523	Never vs. Current	All-cause death	1.4 (1.0-1.9), 0.08	156	1.1 (0.6-2.1)	185	1.5 (0.9-2.4)		
			No vs. ≥30 years		1.3 (1.0-1.7), 0.06		1.2 (0.7-2.0)		1.6 (1.0-2.5), 0.07		
			No vs. ≥30 pack-years		1.2 (0.9-1.7)		1.0 (0.6-1.9)		1.5 (0.9-2.3)		
		meta-analysis above studies	Smoking status	All-cause death	1.24 (0.99-1.56)		0.98 (0.72-1.34)		1.52 (0.85-2.72)		
			Cigarettes per day		1.35 (1.15-1.58)		1.15 (0.82-1.60)		1.70 (1.15-2.52)		
			Years of smoking		1.37 (1.17-1.61)		1.34 (0.97-1.85)		1.93 (1.33-2.78)		
			Pack-years of smoking		1.50 (1.19-1.89)		1.27 (0.91-1.78)		1.81 (1.24-2.62)		
	Leo 2014	1 331	Never vs. Current	All-cause death	1.67 (1.31-2.12)	390	1.53 (1.01-2.32)	214	1.00 (0.48-2.10)	287	2.49 (1.27-4.86)
			Never vs. Current	NHL-spec death	1.47 (1.10-1.97)		1.11 (0.67-1.82)		0.77 (0.31-1.90)		2.88 (1.12-7.44)
Socioeconomic status	Bray 2008	8232	**Residential area**								
			Affluent	All-cause death	Ref						
			Intermediate		1.10 (1.02-1.18)						
			Deprived		1.19 (1.09-1.30)						
	Frederiksen 2012	6234	**Attained education**					**Indolent lymphoma**			
			High (>12 years)	All-cause death	Ref	2670	Ref		Ref		
			Intermediate		1.50 (1.30-1.74)		1.27 (1.11-1.45)		1.18 (1.04-1.35)		
			Low (≤7 or 9 years)		1.63 (1.40-1.90)		1.37 (1.16-1.62)		1.52 (1.37-1.69)		
Body mass Index (kg/m^2)								**FL**			
	Geyer 2010	1 189	Normal (BMI 20-24.9)	All-cause death	Ref	391	Ref	299	Ref		
			Obese (BMI ≥30)		1.32 (1.02-1.70)		1.37 (0.88-2.14)		1.49 (0.81-2.72)		
	Han 2013	Females, 567	Normal (BMI 18.5-24.9)	All-cause death	Reference	182	Ref	130	Ref	62	Ref
			Obese (BMI ≥30)		1.38 (0.99-1.93)		1.77 (0.98-3.21)		1.23 (0.54-2.80)		0.91 (0.22-3.81)
	Leo 2014	1 331	Normal (BMI 22.5-24.9)	All-cause death	Ref	390	Ref	214	Ref	287	Ref
			Obese (BMI ≥30)		1.46 (1.13-1.87)		1.21 (0.78-1.89)		1.96 (0.87-4.42)		2.54 (1.36-4.72)
			Normal (BMI 22.5-24.9)	NHL-spec death	Ref		Ref		Ref		Ref
			Obese (BMI ≥30)		1.77 (1.30-1.75)		1.34 (0.79-2.29)		1.67 (0.64-4.35)		4.99 (1.98-12.6)

(Continued)

	Reference	Cases	Exposure	Outcome	NHL HR (95% CI)	DLBCL Cases	DLBCL HR (95% CI)	FL Cases	FL HR (95% CI)	MCL Cases	MCL HR (95% CI)	T cell lymphoma Cases	T cell lymphoma HR (95% CI)
UV light and vitamin D	Drake 2010	983	Vitamin D sufficient	All-cause death		370	Ref	285	Ref	71	Ref	70	Ref
			Vitamin D insufficient				1.99 (1.27-3.13)		1.52 (0.60-3.88)		1.35 (0.53-3.39)		2.38 (1.04-5.41)
			Vitamin D sufficient	NHL-spec death			Ref		Ref		Ref		Ref
			Vitamin D insufficient				2.16 (1.33-3.51)		0.90 (0.23-3.49)		1.35 (0.53-3.39)		2.26 (0.99-5.17)
	Kelly 2015	(two cohorts)	Vitamin D sufficient	All-cause death				183	Ref				
			Vitamin D insufficient						3.48 (1.52-7.91)				
			Vitamin D sufficient	All-cause death				240	Ref				
			Vitamin D insufficient						1.84 (0.81-4.18)				
			Vitamin D sufficient	PFS				183	Ref				
			Vitamin D insufficient						2.02 (1.18-3.45)				
			Vitamin D sufficient	PFS				240	Ref				
			Vitamin D insufficient						1.66 (1.11-2.48)				
Autoimmune disease	Mikuls 2006	65 with RA	No RA vs. RA	All-cause death	0.95 (0.70-1.30)	28	1.02 (0.66-1.57)						
		1530 w/o RA	No RA vs. RA	NHL-spec death	0.60 (0.37-0.98)		0.63 (0.29-1.36)						
			No RA vs. RA	Non-NHL death	2.16 (1.33-3.50)		1.64 (0.85-3.15)						
	Ji 2011	329 with RA	No RA vs. RA	All-cause death	1.41 (1.26-1.59)								
		56936 w/o RA	No RA vs. RA	NHL-spec death	1.43 (1.23-1.66)								
	Ludvigsson 2013	316 with CD	No CD vs CD	All-cause death	1.35 (0.88-2.08)								
		689 w/o CD											
Family history hematopoietic malignancy						High grade B cell		Low grade B cell		CLL/SLL			Waldenström
	Anderson 2008	25592	No family history	All-cause death	Ref	2284	Ref	4331	Ref	7664	Ref		
		206	History of any lymphoma		0.97 (0.80-1.16)	13	1.02 (0.49-2.16)	28	1.14 (0.66-1.98)	85	1.24 (0.93-1.67)		
	Ji 2009	5036	No family history	All-cause death	Ref								
		98	History of NHL		0.89 (0.65-1.23)								
			No family history	NHL-spec death	Ref								
			History of NHL		0.98 (0.68-1.39)								
			Parents survival time										
		45	<24 months	All-cause death	Ref								
		53	≥24 months		0.62 (0.30-1.26)								
			<24 months	NHL-spec death	Ref								
			≥24 months		0.51 (0.24-1.12)								
	Steingrimsson 2015		No family history	All-cause death								~2100	Ref
			History of lymphoprolif.									93	1.34 (1.03-1.75)
	Mauro 2006		No family history	Survival prob.						1268	67%		
			History of CLL	at 10 years						81	45%, P= 0.08		

NHL= Hodgkin lymphoma, DLBCL= diffuse large B cell lymphoma, FL= follicular lymphoma, CLL= chronic lymphocytic leukemia, MCL= mantle cell lymphoma, PFS= progression-free survival, RA= rheumatoid arthritis, CD= celiac disease

exploring UV radiation, however. Neither of the two studies looked at NHL overall. The first study reported an association of vitamin D insufficiency with increased risk of death among patients with diffuse large B cell lymphoma and T cell lymphomas but not other subtypes, including follicular lymphoma.[104] The second study included two cohorts of follicular lymphoma patients and analyzed them separately.[105] Vitamin D insufficiency was associated with worse progression-free survival in both cohorts and with increased risk of all-cause death in one cohort, although the point estimates were above one in both cohorts.

Vitamin D is despite the name not a vitamin but the precursor of the potent steroid hormone calcitriol.[106] With adequate exposure to UV radiation, the skin produces sufficient amounts of vitamin D. However, for many reasons, a considerable proportion of humans worldwide are not adequately exposed to UV radiation and are consequently vitamin D deficient.[106] Calcitriol exerts several anticancer effects, including inhibition of proliferation and increased sensitivity to apoptotic signals.[106] Aberrantly high expression levels of the enzymes degrading calcitriol (CYP24A1 and CYP27B1) has been demonstrated in some cancers, a resistance mechanism that could be overcome with increased levels of vitamin D in the blood.[106]

2.4.6.5 Autoimmune disease

Three studies apart from Study II have investigated the impact of autoimmune disease on NHL prognosis. Two looked at RA[107,108] and one on celiac disease (Table 3).[109] Regarding RA, Mikuls et al found no association with all-cause or lymphoma-specific death but an increased risk of death from other causes than lymphoma. Ji et al. found a statistically significantly increased risk of both all-cause and lymphoma-specific death among NHL patients with RA compared to those without. There was no association among patients with diffuse large B cell lymphoma. Celiac disease was not associated with survival among NHL patients.[109]

It is thus unclear whether autoimmune disease influences survival of NHL patients. There is data suggesting a more aggressive NHL biology in patients with RA or celiac disease, which would explain an excess mortality in this group compared to NHL patients without autoimmune disease.[109,110] RA patients overall have increased risk of death compared to the general population, primarily due to cardiovascular disease.[111] Comorbidity increases the risk of death in connection with chemotherapy,[95] which could also explain a worse prognosis among RA patients with NHL.

2.4.6.6 Family history of hematopoietic malignancy

Two previous studies have compared survival among lymphoma patients with a first-degree relative with any lymphoma[112] or NHL[113] to survival among sporadic cases and found no difference (Table 3). One of these studies also investigated concordance in survival time among the familial cases and found a statistically non-significant trend.[113]

Just that fact that a NHL patient has a first-degree relative that has had lymphoma does not predict prognosis. A reason for this could be that genetic polymorphism within one family confer a better prognosis than average while genetic polymorphisms within another family confer a worse prognosis than average. Upon combining all familial cases, the average effect of the genetic polymorphisms would go towards zero. The non-significant trend in familial concordance of survival observed in one of the study supports this notion. So does the finding of both protective and deleterious SNPs in candidate gene studies, discussed next.

2.4.6.7 Germline genetic polymorphisms

We explored the potential impact of SNPs on follicular lymphoma survival in Study III. At least 13 previous studies have also investigated selected SNPs in selected genes (candidate gene studies) for their association with follicular lymphoma outcomes, of which a minority were statistically significant (Table 4).[114-125] As shown in Table 4, few of the findings have been validated. Hence, although results are suggestive, there is not yet one SNP convincingly linked to follicular lymphoma prognosis. The rational for exploring genetic polymorphisms is the considerable amount of variability in outcome of patients with the same FLIPI. Also, as described above, gene expression in normal cells in the tumor microenvironment, which are govern by normal DNA, influences follicular lymphoma survival. Genetic polymorphisms could possibly explain these observed differences in gene expression.

Table 4. List of SNPs associated with follicular lymphoma outcome in previous studies.

Chr	Gene	SNP	Minor/ major allele	Study	# FL cases	# SNPs tested	Outcome	Relative risk ratio	# Studies reporting no association [refs]
1	C1qA	rs172378	A/G	Racilia 2008	133	3	PFS	2.5 (2.0, 3.1)	none
1	CD46	rs2466571	C/A	Charbonneau 2012	107	167	EFS	1.49 (0.86, 2.61), p_{trend}<0.05	none
1	CD55	rs2564978	T/C	Charbonneau 2012	107	167	EFS	0.52 (0.30, 0.88)	none
1	CFH	rs1065489	T/G	Charbonneau 2012	107	167	EFS	0.44 (0.24, 0.81)	none
1	CFH	rs1329423	C/T	Charbonneau 2012	107	167	EFS	0.49 (0.29, 0.82)	none
1	CFH	rs3766404	G/A	Charbonneau 2012	107	167	EFS	2.25 (1.31, 3.87)	none
1	CFHR1	rs436719[A]	C/A	Charbonneau 2012	107	167	EFS	0.57 (0.34, 0.96)	none
1	CFHR5	rs6694672	G/T	Charbonneau 2012	107	167	EFS	2.63 (1.41, 4.92)	none
1	FCGR2A	rs1801274	A/G	Weng 2003	87	2	PFS	Decreased rate, $p_{log\text{-}rank}$<0.02	4 [116,117,122,126]
			A/G	Cerhan 2007	278	73	OS	0.58 (0.31, 1.04), p_{trend}=0.04	2 [122,126]
1	FCGR3A	rs396991[A]	G/T	Weng 2003	87	2	PFS	Decreased rate, $p_{log\text{-}rank}$<0.03	5 [116,117,122,125,126]
			G/T	Ghielmini 2005	185	1	EFS	0.5 (0.3, 0.9)	1 [127]
			G/T	Persky 2012	142	2	OS	0.33 (0.11; 0.96)	1 [126]
			G/T	Cartron 2002	49	2	RR	1.5 (1.2, 1.9)	1 [126]
1	MTHFR	rs1801131	C/A	Wang 2009	192	66	OS	2.00 (1.04, 3.84)	none
1	SELE	rs5361		Aschenbrook- Kilfoy 2012	117	82	OS	0.10 (0.02, 0.48)	1 [119]
2	IL1RN	rs454078	T/A	Cerhan 2007	278	73	OS	0.50 (0.28, 0.87)	1 [114]
3	CCR2	rs1799864[A]		Aschenbrook- Kilfoy 2012	117	82	OS	0.27 (0.08, 0.86)	1 [119]
3	FTHFD	rs1127717	G/A	Wang 2009	192	66	OS	1.99 (1.07, 3.7)	none
4	IL2	rs2069762	G/T	Cerhan 2007	278	73	OS	1.81 (1.06, 3.07)	none
4	IL8	rs4073	T/A	Aschenbrook- Kilfoy 2012	117	82	OS	2.60 (1.10, 6.15)	none
			A/T	Cerhan 2007	278	73	OS	0.47 (0.27, 0.79), p_{trend}=0.06	none
4	IL8	rs2227307	T/G	Aschenbrook- Kilfoy 2012	117	82	OS	2.57 (1.07, 6.17)	none
			G/T	Cerhan 2007	278	73	OS	0.53 (0.31, 0.91), p_{trend}=0.11	none
4	IRF2	rs3775567[A]	T/	Gibson 2012[A]	244	6679	OS	3.18 (1.72, 5.87)	none
5	C9	rs1421094	A/G	Charbonneau 2012	107	167	EFS	0.54 (0.32, 0.90)	none
5	IL12B	rs3212227	C/A	Cerhan 2007	278	73	OS	2.01 (1.18, 3.42)	none
6	C6orf15	rs6457327	A/C	Wrench 2011	218	2	TTT	2.25. (1.16, 4.36)	none
				Berglund 2011	102	1	OS	Increased rate, p=0.006	none
6	IFNGR1	rs3799488		Aschenbrook- Kilfoy 2012	117	82	OS	3.19 (1.09, 9.34)	1 [119]
8	GGH	rs719235	T/G	Wang 2009	192	66	OS	2.49 (1.21, 5.14)	none
9	GALNT12	rs10987898[A]	G/	Gibson 2012[A]	244	6679	OS	0.48 (0.27, 0.83)	none
9	GALNT12	rs10819377	T/	Gibson 2012[A]	244	6679	OS	0.62 (0.36, 1.03), p_{trend}<0.001	none
11	CXCR5	rs1790192	G/A	Charbonneau 2013[A]	172	10	EFS	0.64 (0.47; 0.87)	none
16	IL4R	rs1801275		Aschenbrook- Kilfoy 2012	117	82	OS	2.35 (1.07, 5.19)	none
20	BMP7	rs6025446[A]	G/A	Gibson 2012[A]	244	6679	OS	0.41 (0.21, 0.69)	none
22	MIF	rs755622		Aschenbrook- Kilfoy 2012	117	82	OS	2.45 (1.09,5.47)	none

3 HYPOTHESES AND OBJECTIVES

3.1 HYPOTHESES

The overall hypothesis of this thesis is centered around a potential role for host factors in risk as well as prognosis of NHL subtypes. More specifically, we hypothesized that:

I. The contribution of smoking to risk of follicular lymphoma differs by the presence/absence of certain amino-acid variants in the HLA-DRB1 molecule.
II. Host factors (some of which are also risk factors for NHL development) influence lymphoma-related and overall survival of NHL overall and/or subtypes (diffuse large B cell lymphoma, follicular lymphoma, mantle cell lymphoma or T cell lymphoma).
III. Germline genetic polymorphisms in the form of SNPs influence the rate of follicular lymphoma progression.
IV. The prognosis of a lymphoid malignancy and/or subtypes is to some extent genetically inherited.

3.2 OBJECTIVES

By exploring these hypotheses we sought to contribute to the knowledge of what factors influence risk and prognosis of NHL and/or its subtypes. More specifically, the objectives of the four studies were:

I. To investigate if smoking actually is a risk factor of follicular lymphoma, and if so only among indviduals with genetic susceptibility such as germline genetic polymorphisms in HLA-DRB1. This could explain the somewhat conflicting results of previous studies and contribute to our understanding of follicular lymphomagenesis.
II. To identify readily available factors with prognostic importance that could help clinicians in consulting patients with a newly diagnosed NHL.
III. To explore if variation in normal DNA (as opposed to tumor DNA) influences survival of follicular lymphoma patients, which could help us understand observed but unexplained differences in survival in this patient group.
IV. To provide further evidence of a link between germline DNA and survival of lymphoid malignancies. If shown, it would support results from candidate gene and genome-wide associations studies that have found associations between different SNPs and survival of several lymphoid malignancies and encourage further studies of how the host constitution may affect lymphoma survival and how adverse effects could potentially be overcome.

4 DATA SOURCES

4.1 THE SCANDINAVIAN LYMPHOMA ETIOLOGY STUDY

Lymphoma patients and/or control subjects included in the Scandinavian Lymphoma Etiology (SCALE) study were used in Study I, II and III. SCALE was launched with the aim to study risk factors of lymphoma to gain knowledge on lymphoma etiology and understand the increase of lymphoma incidence in the precedent decades. SCALE included adult (≥18 years of age) patients with incident lymphoma between 1999 and 2002 and covered the whole of Sweden and Denmark.[17] Patients were identified through a network of physicians working with lymphoma patients. There was also a backup system that identified new lymphoma diagnoses reported from all pathology and cytology units in Sweden and Denmark. Upon inclusion, the tumor material was reviewed by hematopathologists and classified according to the 2001 WHO classification system.[126] Control subjects were randomly sampled from national population register in Sweden and Denmark every six months between 1999 and 2002, and were frequency matched on the expected number of new lymphoma patients in each country, by age and sex. Study subjects were excluded if they had a history of organ transplantation, HIV infection or other hematopoietic malignancy, or did not speak Swedish or Danish well enough to respond to a questionnaire (see below). Patients and control subjects had to give their informed consent. Of 4506 eligible lymphoma patients, 3740 (83%) were included in the study (case patients). The corresponding numbers for control subjects were 4489 and 3187 (71%). Among the case patients, 3055 (82%) had NHL (796 diffuse large B cell lymphoma, 752 chronic lymphocytic leukemia, 586 follicular lymphoma, 678 other B cell lymphomas, 204 T cell lymphomas). Demographic data and information on potential risk factors were gathered through telephone interviews using computer-aided questionnaires. The majority (80%) were interviewed within three months from diagnosis/study inclusion (range 0-40 months). Of those interviewed, the majority of case patients (2790, 91%) and control subjects (2118, 66%) also gave blood.[127] Genotyping of approximately 300 000 SNPs were performed on a subset of lymphoma cases and controls at the Genome Institute of Singapore (see 5.6 for details).[128]

As years have passed, the case patient cohort has also been possible to use for prognostic studies, with the limitation that most patients have not received primary treatment that included rituximab (about 10% in Swedish SCALE).[129] Clinical data has been collected for Swedish patients from medical records, including information on lymphoma stage, performance status, comorbidity, hemoglobin and lactate dehydrogenase levels, treatments and recurrent disease. For Danish patients, data on lymphoma stage, performance status, treatment and follow-up was obtained from the Danish national lymphoma register LYFO.[130]

4.2 THE EPIDEMIOLOGICAL INVESTIGATION OF RHEUMATOID ARTHRITIS STUDY

For Study III, we used 676 Swedish control subjects included in the population-based Epidemiological Investigation of Rheumatoid Arthritis (EIRA) case-control study between 1999 and 2003.[131] EIRA is an ongoing case-control study of the autoimmune disease rheumatoid arthritis, recruiting cases and controls since 1996. The control subjects were randomly selected from the Swedish national population register and matched with regard to the rheumatoid arthritis patients' age, sex and residential area. All study subjects gave informed consent. Demographic data and information on potential risk factors of rheumatoid arthritis (including smoking) were gathered by questionnaires sent by mail. Blood samples were obtained from 58% of control subjects that had first responded to the questionnaire. As in the SCALE study, SNP genotyping was done at the Genome Institute of Singapore.

4.3 THE UNIVERSITY OF CALIFORNIA SAN FRANCISCO COHORT

We used 213 follicular lymphoma cases with European ancestry from the University of California San Francisco (UCSF) case-control study of lymphoma for Study II. UCSF included incident NHL patients aged 20-84 in the San Francisco Bay area 2001-2006.[132] Patients were identified through a rapid case ascertainment system by the Cancer Prevention Institute of California with case reporting supplemented by Surveillance, Epidemiology and End Results data. All study subjects gave informed consent. Of eligible cases 69% were included. Blood or buccal specimens for DNA preparation were obtained from a majority (87%) of cases a median 27 days after diagnosis. Genotype data of approximately 300 000 SNPs were available for 213 follicular lymphoma patients (see 5.6 for details).

4.4 SWEDISH NATIONAL REGISTERS

4.4.1 The Cancer register

The Swedish Cancer register is kept by the National Board of Health and Welfare. The register includes incident malignant diseases (and some benign tumors) of Swedish residents since 1958.[133] All physicians in Sweden are obliged by National Board of Health and Welfare regulations to report all new malignant tumors to the register. In addition, pathologists and cytologists also have to report every malignant tumor diagnosed at their units. This system ensures completeness and reliability, both of which were estimated to be close to 100% in the late 1970s and in 1998.[134,135] For each case, the register keeps record of

personal identification number, sex, age and date of diagnosis, and type of malignancy stored as International Classification of Disease (ICD) codes, seventh version. Since 1993, histopathologic type is also recorded as Systematized Nomenclature of Medicine (SNOMED) codes.[133]

4.4.2 The Cause of Death register

The Swedish Cause of Death register is kept by Statistics Sweden. The register in its present form keeps record of the date and cause of all deaths of all Swedish residents from 1961.[136] An older version of the register keeps the same information between 1952 and 1960. The date and conditions related to the death are reported to the register on a death certificate issued by the physician that verifies the death. Statistics Sweden assigns ICD codes to the conditions reported on the certificate and selects a single main underlying cause of death as defined by the WHO in the current ICD version (ICD-6 1952-1957, ICD-7 1958-1968, ICD-8 1969-1986, ICD-9 1987-1996, ICD-10 1997- still in use). The register is complete with regard to capturing all deaths along with the date.[136] With regard to the cause of death, it is not as straightforward. The validity of the cause of death reported on the death certificate depends on how well the physician examined the circumstances of the death. This could vary from an educated guess based on information from relatives to a full forensic investigation.[137] A validation study compared the cause of death stated on death certificates with the main diagnosis of hospital discharge records in deaths occurring 1995.[138] The correlation was greater for deaths occurring in hospitals than at home, and for deaths resulting from an emergency condition rather than a chronic one.[138] The best agreement was observed for malignant neoplasms (90-98%).[138] Another problem is that the WHO rules and guidelines on coding and classification of cause of death have developed over a long period of time and have been subject to changes and sometimes open to different interpretations.[137] Furthermore, the underlying cause of death concept has become more difficult to apply as we live longer and often have time to accumulate multiple potentially fatal condition.[137] This makes the validity of comparisons made over long time periods uncertain.

4.4.3 The Total Population register

The Total Population register is maintained by Statistics Sweden and was founded 1968, when a large part of available Swedish population data was computerized [139] At that time, personal records with demographic information were kept by Church of Sweden. This duty was taken over by the Swedish Tax Agency 1 July 1991. The Total Population register includes all individuals that resided (Swedish: folkbokförda) in Sweden at some point since 1968. There is a Total Population register for each year since 1968. The register also includes

selected individual level historical data for longitudinal studies. The Total Population register contains information on life events, including birth, death, emigration and immigration, as well as sex and family relations. The coverage is virtually 100% for births and deaths, 95% for immigration and 91% for emigration. An older version of the Total Population register keeps record of births and deaths from 1961 through 1997.

4.4.4 The Multi-generation register

The Multi-generation register is kept by Statistics Sweden.[140] To be included, an individual must have been born in 1932 or later and registered as resident in Sweden at some point since 1961. These individuals are defined as index persons. The identities of the index persons' biological parents are collected from birth certificates and recorded in the register. Married couples are automatically stated as the biological parents on the birth certificate. For unmarried couples, fatherhood is certified by the couple at a visit to the Social Welfare Committee shortly after birth. The registered parent-index person relations are used to derive biological sisters and brothers, as well as half-siblings. In the case of adoption, this is also recorded in the register. The completeness varies over time and by whether the index person was born in Sweden or not. As illustrated by Figure 14, the coverage of biological mothers and fathers is close to complete for individuals who were alive 1 July 1991 and born in Sweden. For index persons born outside of Sweden, data is much less complete. Also, parental information is missing for a large number of index persons who died 30 June 1991 or before, as illustrated by Figure 15. The reason for this is that the personal records of these individuals were not computerized when the Swedish Tax Agency took over the responsibility from Church of Sweden 1 July 1991. Work of adding data on these individuals is ongoing.

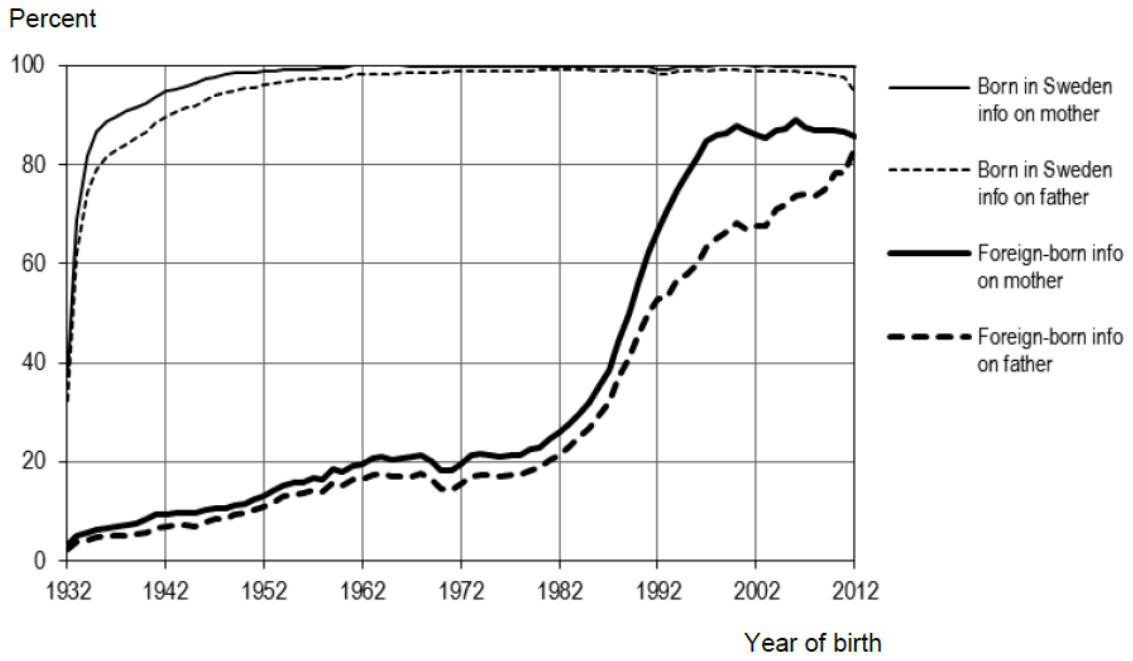


Figure 14, from a report by Statistics Sweden on the Multi-generation register 2012.[140] The graph shows that the coverage of biological parents of index persons alive 1 July 1991 and born in or outside of Sweden, by calendar year of birth of index persons.

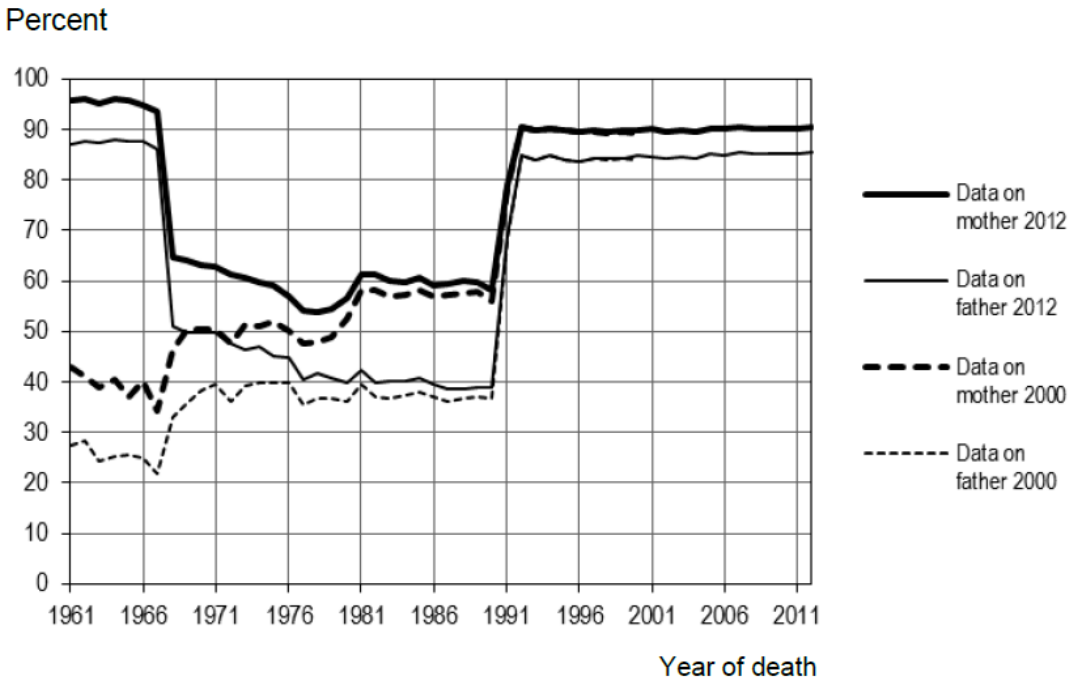


Figure 15, from a report by Statistics Sweden on the Multi-generation register 2012.[140] The graph shows the proportion of individuals with information on their biological parents among persons deceased 1961-2012, when measured in the register the year 2000 and 2012, by calendar year of death of the index persons. The increase in coverage between 2000 and 2012 is thanks to ongoing work to supplement data for index persons deceased 30 June 1991 or before.

4.4.5 Census records

Censuses were performed in Sweden every fifth year between 1960 and 1990, and the records are kept by Statistics Sweden.[141] The censuses covered all individuals registered as residents in Sweden at the time of each census. Information was collected by questionnaires sent by mail and to some extent from available registers. The data collected varied somewhat between censuses but always included individual level information on demographic variables and employment. Based on employment, a socioeconomic index (SEI) was created and included in all censuses except in 1975. Attained educational level was included in the censuses of 1960, 1970 and 1990. Quality controls were performed upon each census and the coverage and completeness were reported to be very good. Detailed information on each census is available at Statistics Sweden (http://www.scb.se/sv_/Hitta-statistik/Historisk-statistik/Digitaliserat---Statistik-efter-serie/Sveriges-officiella-statistik-SOS-utg-1912-/Folk--och-bostadsrakningarna-1860-1990/Folk--och-bostadsrakningen-1965-1990/).

4.4.6 The Longitudinal Integration database for health insurance and labor market studies (LISA)

The Longitudinal Integration database for health insurance and labor market studies (LISA) is maintained by Statistics Sweden.[141] The database holds annual registers since 1990. LISA includes all individuals 15 years of age or older (16 years or older before 2010) that were registered as residents in Sweden 31 December each year. Data is collected from several national registers of high coverage and quality. For each individual, detailed socioeconomic data is available, including employment and level of attained education. The completeness and validity of the data in LISA is dependent on the quality of each register from where the data was obtained. Missing data regarding education level was observed in 1.7% of individuals 25-64 years of age in LISA 2013.[142] The same number in 1990 was 1.9% for individuals 16-64 years of age. The majority of those missing data were born outside of Sweden. In 2002 validation of the Education register, from which data on education level is collected, was done. Misclassification of education level was uncommon (1%) among individuals with high level of education (>12 years), 16% among individuals with intermediate level (10-12 years) and 77% among those low level of education (9 years of less).[143] Typically, misclassified individuals should be moved up one level (Table 5).

Table 5. Proportion (%) of individuals with misclassified education level in LISA 2002 (based on data from the Education register).

Education register	**Validation study**						
	≤9 years	High school ≤2 years	High school 3 years	University <3 years	University ≥3 years	Research level	Total (%)
≤9 years	77	16	5	1	1	0	100
High school ≤2 years	0	88	9	2	1	0	100
High school 3 years	0	0	84	13	3	0	100
University <3 years	0	0	0	89	11	0	100
University ≥3 years	0	0	0	1	99	0	100
Research level	0	0	0	0	0	100	100

5 METHODS AND STATISTICS

5.1 CASE-CONTROL STUDY DESIGN

A case-control study design was used in Study I (Table 6).

The case-control study design starts out from the outcome of interest. Step one is to identify all the individuals that have experienced the outcome of interest in the population under study (the source population). Next, control subjects that have not experienced the outcome are randomly sampled from the source population of the cases. The prevalence of the exposure of interest in the control sample should mirror that of the source population. Information on risk factors of interest and possible confounding factors are collected for cases and control subjects. Ultimately, the relative risk between cases and controls are calculated as odds ratios. What is actually calculated is the odds ratio of being exposed among cases versus controls. However, it can be shown that the odds ratio of being exposed equals the odds ratio of experiencing the outcome.

The sampling of controls can be done in different ways and has implications for how the odds ratio can be interpreted (Figure 16).

a. Case-noncase sampling. The control subjects are randomly sampled from the source population such that they are free from the outcome at the end of follow-up (time point a in Figure 16). An individual can thus not be both a case and control.
b. Incidence density sampling (time-matched sampling). Control subjects are randomly sampled from the source population every time a new case arises (time point b in Figure 16). In this way, an individual selected as control subject may later become a case.
c. Case-cohort sampling. The control subjects are randomly sampled from the whole source populaton, including the case patients, at the beginning of follow-up (time point c in Figure 16). With this design, an individual may work as both control subject and case.

Depending on sampling scheme used, the odds ratio will estimate either a risk ratio (sampling scheme c), odds ratio (sample scheme a) or incidence rate ratio (sampling scheme b).

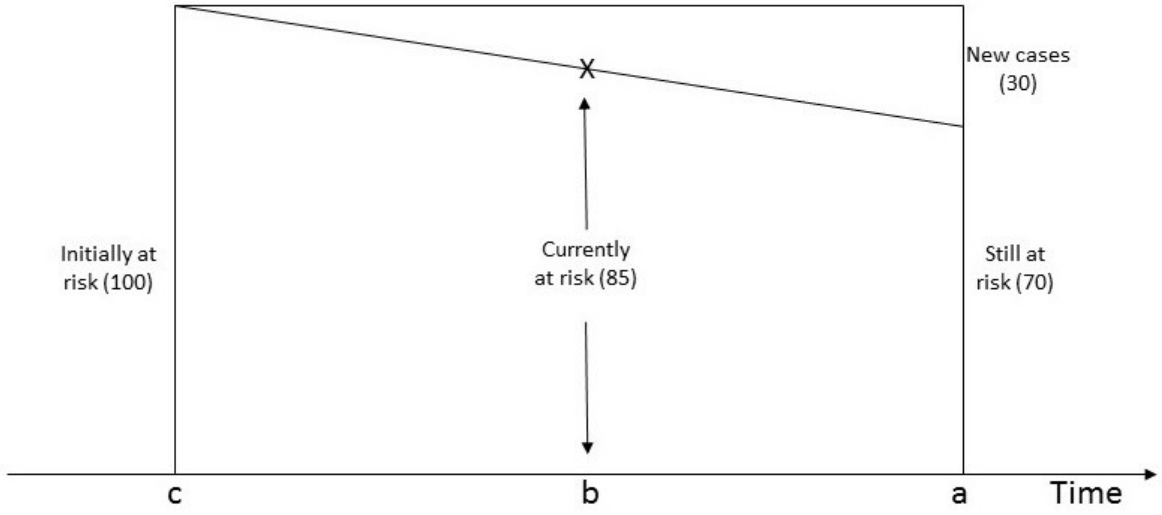


Figure 16 illustrates the different time points at which control sampling can be done in case-control studies.

5.2 COHORT STUDY DESIGN

A cohort study design was used in Study II, III and IV (Table 6).

The cohort study design starts out from the risk factor (exposure) of interest. Individuals that fulfill eligibility criteria and are free of the outcome of interest are selected to the study cohort. The exposures as well as possible confounders are measured for each study subject. Next, the study subjects are followed over time until they have experienced the outcome, are lost to follow-up or until the termination of the study. This design could also be termed internal cohort study as opposed to an external cohort study where an external group is used for comparison. Data for cohort studies can also be obtained from pre-existing national registers, some of which are described in section 4.1.4. This was the case in Study IV. The advantages of using these registers are that they are population based and aim to include all Swedish residents (no selection). Linkage with several national registers may provide a wide range of data. They usually span over a long period of time, allowing for longitudinal studies. Without this kind of data, studies like Study IV would be difficult to do. The major drawback of using national registers is the lack of clinical data.

5.3 STATISTICS

5.3.1 The Kaplan-Meier estimator

We used the Kaplan-Meier estimator[144] in Study II with the aim to illustrate differences in survival by smoking status in patients with NHL. The method uses all available observations, both those that have experienced the outcome and those that have not (censored observations). It is important to keep in mind that the Kaplan-Meier estimator provides univariate estimates. That is, the estimates are not adjusted for possible confounders and will therefore be more or less biased in an observational study.

The Kaplan-Meier estimator makes one assumption to be aware of, namely that the reason for the censoring is independent of the outcome, i.e. that the censoring is non-informative.

5.3.2 Cox proportional hazards regression

We used Cox proportional hazards regression models[145] to estimate the relative mortality rates (hazard ratios) by the risk factors of interest, such as smoking status (Study II), SNP alleles (Study III) or the survival of a first-degree relative (Study IV).

While the Kaplan-Meier method estimates the probability of staying alive over time (the survival function), Cox regression estimates the risk of dying over time (the hazard function). The advantage with Cox regression is that it allows estimation of the hazard ratio for a risk factor of interest while simultaneously accounting for the effect of other variables, such as confounders. Cox regression is also possible to use on censored data. A drawback is that it only estimates hazard ratios and not the baseline hazard for each risk group.

Cox regression make the assumption that hazard ratios, comparing the risk groups under study, are constant during follow-up – the proportional hazards assumption. If the assumption does not hold, the resulting ratio estimates may not be valid. Two ways of testing the assumption are: 1) introducing interaction variables between the risk factor(s) and a variable that represents the underlying time scale in the Cox model (which allows the effect of the risk factor to vary with time), and 2) plotting the Schoenfeld residuals from the Cox regression models against the time scale and test if the magnitude of the residuals depend on time (Grambsch-Therneau test).

5.3.3 Competing risk regression models

In Study II, we assessed the risk of lymphoma-related death in a competing risks regression model.[146] Competing risks are present when two (or more) events are linked such that the occurrence of one event either precludes or alters the probability of the other event to take place. In ordinary Cox regression models, competing events are censored and the observation removed from the risk set when the competing event occurs. It thus treats the competing event as if it was not possible to take place. In competing risks regression, individuals experiencing the competing event remain in the risk set, although their contribution is mitigated by calculated weights (the size determined by the degree of dependence of the two competing events). The resulting hazard of each event (the one of interest and the competing one) is estimated (subdistribution hazard). The subdistribution hazard of the competing event is taken into account when estimating the subdistribution hazard of the outcome of interest.

The interpretation of the HR of the two models become somewhat different. Under certain assumptions, the hazard ratios obtained from Cox regression models provide an estimate of the relative risk of the outcome (given different levels of exposure) in the hypothetical situation where the competing event does not exist. In competing risks regressions, the subdistribution HR has a less intuitive interpretation since individuals who have experienced the competing event remain in the risk set. However, significant differences in the subdistribution hazards suggest that the absolute risk of experiencing the event (in the real world where competing events exist) differs by risk group.

5.3.4 Logistic regression

We used logistic regression models in Study I, when examining the possible interaction between smoking status and HLA-DRB1 variation in risk of follicular lymphoma.

Logistic regression is appropriate for dichotomous outcomes. The risk factor of interest and confounders included in the model on the other hand may be continuous or categorical. Logistic regression estimates the relative risk of the outcome in different risk groups as odds ratios. There is no time scale in logistic regression, making reversed causality a possibility.

5.3.5 Analysis of interaction

We performed an analysis of interaction between smoking status and HLA-DRB1 haplotypes in risk of follicular lymphoma in Study I.

By interaction one could mean statistical interaction between two covariates in a statistical model with the purpose to allow more flexibility in the model. Our purpose was to investigate potential biological interaction between two factors in the pathogenesis of follicular lymphoma, however. An issue that could be debated is how interaction is best measured: on the multiplicative (risk ratio) or additive (risk difference) scale? Most studies report interaction on the multiplicative scale.[147] Some statisticians have argued, however, that its frequent use may be due to convenience rather than consideration about what measure is most preferable, since standard statistical software will automatically give confidence intervals for multiplicative interactions while additional work is needed to obtain measures on the additive scale.[148] The additive scale may be preferred for the following reasons:[148]

- The additive scale is of more value from a public health perspective, since it may help distinguish the groups that would benefit the most from an intervention.
- The additive scale seems to better reflect a biological synergism than does the multiplicative scale.

Furthermore, it can be shown that if each of two exposures under investigation are associated with the outcome, there will be interaction on some scale.[149] It thus seems important to pick scale before starting analysis. Given these arguments and experience from studies of smoking and HLA variation in risk of RA, we chose the additive scale for our analyses of interaction. Consequently, in Study I, interaction was defined as a joint effect of smoking status and amino-acid haplotype combination on follicular lymphoma risk that differed from the sum of the effect of each factor alone (deviation from additivity of effects), as recommended by Rothman et al.[150] Interaction was assessed by estimating the proportion of the joint effect that could be attributed to the interaction itself, the attributable proportion (AP) with 95% confidence intervals, as described by Hosmer and Lemeshow.[151]

AP was calculated by subtracting the relative risk (RR) of follicular lymphoma for current relative to never smoking ($\mathrm{RR}(\mathrm{A}\overline{\mathrm{B}})$) and two amino-acid haplotype alleles relative to no allele ($\mathrm{RR}(\overline{\mathrm{A}}\mathrm{B})$) from the joint effect of the two risk factors relative to having no risk factors ($\mathrm{RR}(\mathrm{AB})$) and then dividing the difference with the joint effect of the two risk factors ($\mathrm{RR}(\mathrm{AB})$):

$$AP = \frac{RR(AB) - RR(A\overline{B}) - RR(\overline{A}B) + 1}{RR(AB)}$$

Where A= presence of current smoking, B= presence of two risk haplotype alleles, $\overline{A}$= absence of current smoking (i.e. never smoking), $\overline{B}$= absence of two risk haplotype alleles (i.e. no risk allele).

5.3.6 The DerSimonian-Laird method for pooling results

In Study III, we wanted to weigh together results from two independent GWAS (SCALE and UCSF, see 4.1 and 4.3 for details regarding these cohorts). The aim was to strengthen the possibility to demonstrate a possible association between investigated SNPs with follicular lymphoma outcome. For this, we used a random effects model described by DerSimonian and Laird in 1986.[152]

The difficulty in combining results from different studies is that they may differ in study design, population characteristics and sampling errors. A fixed effects model assumes that the different studies actually estimate the same true effect of the risk factor of interest on an outcome.[153] Then, different results that may be observed between studies are only due to chance. The random effects model on the other hand assumes that the different studies are in fact estimating different but closely related true effects of the risk factor on the outcome.

The DerSimonian-Laird method considers the difficulties in combining results from different studies by 1) calculating weights for each study based on the amount of information they provide, and 2) including an estimated value of the heterogeneity of effects between the studies in the analysis. Large studies will get more weight than small studies and will thus influence the final estimate more. The greater the estimated heterogeneity between studies, the wider the confidence interval around the estimated average effect will be.

We chose a random effects model for pooling the estimates in the two GWAS because we reasoned that this model may better reflect our lack of knowledge of the possible true nature that the investigated SNPs may have on lymphoma progression as compared to a fix effects model. Although the UCSF cohort was of European ancestry and we accounted for population stratification both in UCSF and SCALE, there may still be a difference in the true effect between the two populations that is not due to chance.

5.4 Study I: Possible interaction between cigarette smoking and human leukocyte antigen DRB1 variation in risk of follicular lymphoma

We used 373 Swedish and Danish incident follicular lymphoma patients and 142 Danish control subjects included in SCALE 1999-2002 (see 4.1; Table 6). Swedish SCALE control subjects could not be used since DNA was not available for this group. Instead, we used 676 Swedish control subjects included in EIRA 1999-2003 (see 4.2). Information on smoking habits was collected by questionnaires by telephone in SCALE and sent by mail in EIRA soon after inclusion in each study. Information on amino-acids at positions 11, 13, 28, 30 and 70-74 in HLA-DRB1 was obtained by imputation based on genome-wide SNP data in both SCALE (see 5.4) and EIRA. First, dense imputation of SNPs across the HLA region was done. In this step, the accuracy of imputation was tested by direct DNA sequencing in a subset of the study subjects and was found to be high (>98% allelic concordance). In the next step, the imputed SNP data was used to derive amino-acid data at each position of the HLA-DRB1 molecule. Ever smoking was defined as daily smoking for at least one year. Smoking status was categorized as current (subject smoked at or had quit <1 year before diagnosis/inclusion), former (subject quit smoking ≥1 year before diagnosis/inclusion) or never smoker. Amino-acid positions 11, 13, 28 and 30 are linked, why we combined them into haplotypes (in total sixteen haplotypes were identified). We further identified the number of shared epitope alleles (amino-acid positions 70-74) carried by each individual. Shared epitope was defined as HLA-DRB1*01, HLA-DRB1*04 or HLA-DRB1*10, excluding HLA-DRB1*0103, HLA-DRB1*0402 and HLA-DRB1*0403 when 4-digit HLA type data was available. The follicular lymphoma diagnosis (outcome) was verified by hematopathologists upon NHL diagnosis (see 4.1 for details). We first estimated the main effects of smoking status and amino-acid haplotypes on follicular lymphoma risk separately using logistic regression models, adjusting for sex, age at diagnosis and country of residence (Sweden or Denmark). Amino-acid haplotypes most statistically significantly (P for trend $<5.0 \times 10^{-4}$) associated with follicular lymphoma risk were brought forward for interaction analysis with smoking status. Interaction between smoking status and amino-acid haplotypes in the risk of follicular lymphoma was assessed on the additive scale, as described above (5.3.6).

	Study I	Study II	Study III	Study IV
Study design	Case-control study	Cohort study	Cohort study	Register-based cohort study
Data sources	SCALE study EIRA study	SCALE study, Swedish part Cause of Death register	SCALE study Cause of Death register Danish national lymphoma register LYFO) UCSF study	Cancer register Total Population register Multi-generation register Censuses of 1960, 70, 80, 90 LISA 1991-2015
Study cohort	373 follicular lymphoma patients (SCALE) 142 Danish control subjects (SCALE) 676 Swedish control subjects (EIRA)	1523 non-Hodgkin lymphoma patients (excluding CLL)	1. 586 follicular lymphoma patients (SCALE and UCSF) 2. 373 follicular lymphoma patients (SCALE)	1. ~39000 patients with a lymphoid malignancy 2. 1701 pairs of first-degree relatives with a lymphoid malignancy 3. 492 pairs of first-degree relatives with the same type of lymphoid malignancy
Year of diagnosis	1999-2002	1999-2002	1999-2002 (SCALE) 2001-2006 (UCSF)	1958-2011
Follow-up time period	Not applicable	1999-2010	1999-2012 (SCALE Sweden) 1999-2009 (SCALE Denmark) 2001-2012 (UCSF)	1958-2013
Exposure	Smoking status at diagnosis/inclusion HLA-DRB1 amino-acid variation	Smoking status Education (Socioeconomic status) Body mass index Sunbathing (UV light exposure) Autoimmune disease Family history of hematopoietic malignancy	Single nucleotide polymorphisms	Survival time of first-degree relative: Good (best quartile of deviance residuals) Expected (two middle quartiles) Poor (worst quartile)
Outcome	Follicular lymphoma	1. All-cause death 2. Lymphoma-related death	1. Lymphoma-specific death 2. Lymphoma progression	All-cause death
Time	Not applicable	Interview → outcome or end-of-follow-up	Diagnosis → outcome, non-lymphoma-specific death or end-of-follow-up	Diagnosis → death, emigration or end-of-follow-up
Co-variates	Sex Age at diagnosis Country of residency	Sex Age at diagnosis Ann Arbor stage Smoking status* Education* Body mass index*	Sex Age at diagnosis Population stratification	Type of lymphoid malignancy Sex Age at diagnosis Year of diagnosis Socioeconomic status
Statistics	Logistic regression models	Cox proportional hazards regression model Kaplan-Meier estimator	Cox proportional hazards regression model DerSimonian-Laird random effects model	Cox proportional hazards regression models

Table 6. Study I-IV at a glance. *****Although exposures of interest, these variables were also considered as confounders for the other exposures and thus included in all Cox models.

5.5 Study II: Lifestyle factors, autoimmune disease and family history in prognosis of non-Hodgkin lymphoma overall and subtypes

We used a cohort of 1523 Swedish NHL patients included in SCALE 1999-2002 (see 4.1; Table 6). Patients with chronic lymphocytic leukemia were excluded from this study. Six factors were investigated for their association with all-cause and lymphoma-related death among all NHL patients or patients with diffuse large B cell lymphoma, follicular lymphoma, mantle cell lymphoma or T cell lymphomas: smoking status, education (as proxy for socioeconomic status), body mass index (BMI), sunbathing habits 5-10 years before diagnosis (as proxy for UV radiation exposure), autoimmune disease and family history of a hematopoietic malignancy.

Information on smoking status, education, BMI, sunbathing habits and autoimmune disease was collected using questionnaires by telephone interviews soon after inclusion in SCALE. Data on family history of a hematopoietic malignancy was obtained from the Multi-generation and Cancer register (see 4.4.4 and 4.4.1). First, first-degree relatives of the NHL patients were identified by linking to the Multi-generation register. Second, first-degree relatives with a hematopoietic malignancy were identified by linkage to the Cancer register. Data on the confounding factors Ann Arbor stage and extranodal disease were obtained for 89% of NHL patients from the national Swedish lymphoma quality registers.[154] Smoking status was defined as described on the previous page (p. 56). We further categorized smoking duration (1-19, 20-39 or ≥40 years), smoking intensity (1-9, 10-19 or ≥20 cigarettes per day) and cigarette-years (none, <200, 200-399 or ≥400 cigarette-years). We also categorized attained education (≤9, 10-12 or >12 years), BMI (<25, 25-29 or ≥30 kg/m^2) and summer sunbathing frequency in the past 5-10 years before diagnosis (never, 1, 2-3 or ≥4 times per week). Autoimmune disease was a composite variable consisting of self-reported physician-made diagnosis of RA (64%), SLE (7%), celiac disease (20%) or primary Sjögrens syndrome (10%) categorized as either present or absent. Ann Arbor stage and extranodal disease were combined in one variable (Ann Arbor stage 1-2, stage 2-3 or extranodal disease).

Outcome data was obtained from the Cause-of death register (see 4.4.2) through 1 October 2010. Lymphoma-related death was defined as having lymphoma as the main underlying or contributory cause of death in the Cause of Death register. Time was counted from diagnosis of NHL until death or end of follow-up (outcome all-cause death) or from diagnosis of NHL until death from lymphoma-related causes or death from other causes or end of follow-up (outcome lymphoma-related death). The relative risk of each potential prognostic factor on the two outcome measures were estimated as hazard ratios (HR) along with 95%

confidence intervals (CI), using Cox proportional hazard models. All Cox models were adjusted for sex, age at diagnosis and stage/extranodal disease. All Cox models also included the potential prognostic factors smoking status, education and BMI, because they were hypothesized to be confounders of each other as well as the other factors under study. The variables sunbathing, autoimmune disease and family history of hematopoietic malignancy were added to the Cox model individually. The proportional hazard assumption was tested by creating interaction variables by multiplying each potential prognostic factor with time and including them in the Cox models. In a sensitivity analysis, the potential bias due to competing risks between lymphoma-related and non-lymphoma-related deaths (when analyzing lymphoma-related death) was assessed in a competing risks regression model.

5.6 Study III: A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival

For this study, we used 373 follicular lymphoma patients from SCALE (both Swedish and Danish part; see 4.1) and 213 from UCSF (see 4.3; Table 6).

In SCALE, 400 follicular lymphoma patients were genotyped for 317 503 SNPs across the genome at the Genome Institute of Singapore. For this, the Illumina HumanHap 300 array (version 1.0; Illumina, Sand Diego, California, USA) was used. Individual SNPs were filtered out on the basis of SNP genotyping call rates (≥95%), sample completion rate (≥90%), minor allele frequency (MAF, ≥0.03) and non-deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$). SNPs were further excluded if there was cluster plot problems or they were located on the X or Y chromosome. Follicular lymphoma patients were excluded if there were gender discrepancies, labeling errors and/or cryptic family relations as indicated by the genome command in PLINK.[155] Finally, population outliers were identified and removed on the basis of the first three principal components, computed using EIGENSTRAT in R.[156] After this filtering process, 373 follicular lymphoma patients and 298 198 SNPs were available for analyses in SCALE.

One part of the study was to test SNPs reported to be associated with follicular lymphoma prognosis in other studies for lymphoma-specific survival and progression in the SCALE cohort. These SNPs were identified by searching PubMed. Few identified SNPs were on the HumanHap 300 array, however. Instead we imputed the values for these SNPs in SCALE, using available SNP genotype data in the dataset and the 1000 Genomes Projects multi-ethnic reference panel and Impute 2.[157,158] A strict threshold was set to genotypes with probabilities >0.9, SNPs with information score >0.8 and call rates >0.9.

In the Swedish part of SCALE, we also collected information on Ann Arbor stage, performance status, number of involved nodal areas, hemoglobin and lactate dehydrogenase levels, treatment and signs of progression from medical records. Follow-up data was obtained from the Cause of Death register (see 4.4.2). In the Danish part of SCALE, information on Ann Arbor stage, performance status and vital status was collected from the Danish National lymphoma register LYFO.[130]

In UCSF, >370 000 SNPs were genotypes using the Illumina HumanCNV370-Duo BeadChip (Illumina, Sand Diego, California, USA). After a similar filtration process as described for SCALE, 213 follicular lymphoma patient and 339 528 SNPs were available for analysis in UCSF. Population stratification was assessed using multidimensional scaling in PLINK. Follow-up data was obtained from the Greater Bay Area Cancer registry.

Lymphoma-specific death was defined as having lymphoma as the main underlying cause of death. Lymphoma progression (Swedish SCALE only) was defined as start of first-line treatment in patients with an initial watch-and-wait strategy >6 months after diagnosis, start of second-line treatment in patients where first-line treatment was initiated ≤6 months after diagnosis, or lymphoma-specific death among these patients if no treatment was given. Time was counted from date of diagnosis until outcome, competing event (which was censored) or end of follow-up (2 February 2012 in SCALE Sweden, 17 December 2009 in SCALE Denmark, 31 August 2012 in UCSF).

We used Cox proportional hazards regression models to estimate the relative risk of the outcomes by number of alleles of each SNP (0, 1 or 2). We assumed an additive model and treated the SNP variable as ordinal in the Cox model. All Cox models were adjusted for sex, age at diagnosis and population stratification (using the first three principal components). The results of the two GWAS (SCALE and UCSF) were pooled using a random effects model (see 5.3.6). Upon assessing lymphoma progression (SCALE Sweden) we tested the proportional hazards assumption by plotting Schoenfeld residuals against time. In complementary analyses of lymphoma progression, models were also adjusted for FLIPI risk groups and first-line rituximab.

5.7 Study IV: Concordance in survival of first-degree relatives diagnosed with a lymphoid malignancy

For Study IV we use 1701 pairs of first-degree relatives with a lymphoid malignancy identified by linkage of the Cancer register (see 4.4.1) and the Multi-generation register (see 4.4.4; Table 6). We first identified all individuals diagnosed with a lymphoid malignancy in the Cancer register 1958-2011 (Figure 17). These individuals were linked to the Multi-generation register to identify their parents and siblings. Ultimately, 39 608 patients with data on first-degree relatives were identified. Among these, we identified families with two (N=1606), three (N=86) of four (N=9) first-degree relatives with a lymphoid malignancy. Among these, we defined index cases as the individual within a family that last was diagnosed with a lymphoid malignancy. Because data on parents are incomplete in the Multi-generation register for individuals that died before 1991 (see 4.4.4), only index patients alive in 1991 were used in the study. In families with three or four members with a lymphoid malignancy, we allowed two or three index cases, respectively. Hence, in a family with three members, two pairs of index case and first-degree relative were formed: pair 1. the individual diagnosed with a lymphoid malignancy last (index) and the one diagnosed second to last (first-degree relative), pair 2. the individual diagnosed second to last (index) and the one diagnosed first (first-degree relative). We formed 1701 pairs of index cases and first-degree relative cases in this way. Among these, 491 pairs had the same type of lymphoid malignancy.

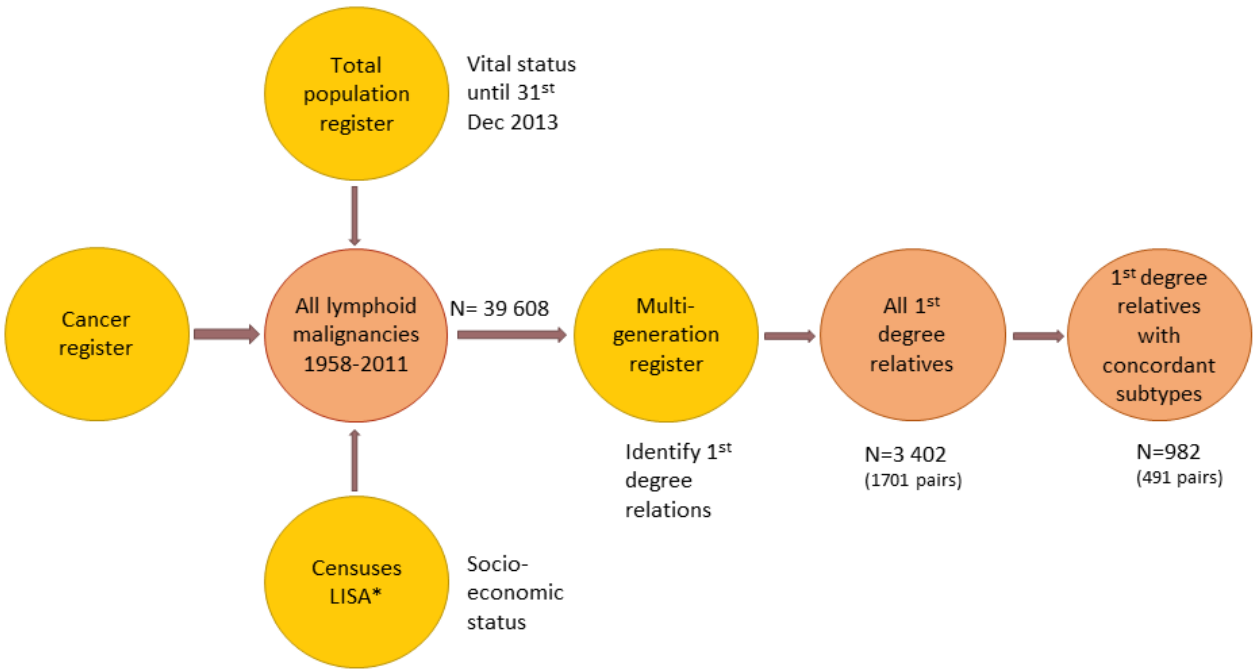


Figure 17. Illustration of the different registers used (yellow circles) and the three cohorts formed (orange circles).

Details on type of lymphoid malignancy increased with time, especially after 1993 when SNOMED codes were included in the Cancer Register. Due to less detail before 1993 and because of few observations for most subtypes, we categorized all patients (39 608) into seven main groups: aggressive lymphoma, indolent lymphoma, Hodgkin lymphoma, chronic lymphocytic leukemia, plasma cell malignancies, acute lymphoblastic leukemia and lymphoma unspecified (see Supplement in connection to the original paper at the end for details). To be able to adjust for socioeconomic status, information on attained education or employment (if education was not available) was obtained from the National Population and Housing censuses 1960, 1970, 1980 and 1990 (see 4.4.5) and the LISA database (see 4.4.6). Statistics Sweden have defined the level of education normally needed for each type of employment. Using this information, we were able to define high (>12 years of education), intermediate (10-12 years) or low (≤9 years) socioeconomic status.

Our outcome measure was all-cause death. This data and migration data was obtained from the Total Population register (see 4.4.3).

The exposure of interest among the index cases was the survival time of their first-degree relative, categorized as good, expected or poor. To do this, we first had to define good, expected and poor survival. Our first step was thus to run a Cox proportional hazard model in which we included all 39 608 patients with a lymphoid malignancy minus the index cases. Covariates in the Cox model were sex, age and year of diagnosis and socioeconomic status, and it was stratified by type of lymphoid malignancy (seven levels). Time was counted from diagnosis until death from any cause, emigration or end of follow-up (17 June 2013). The deviance residuals from this Cox model provided adjusted measures of the survival time of each individual included. Within each type of malignancy (seven levels), we defined good, expected or poor survival as quartiles of deviance residuals: good survival was define as the quartile of deviance residuals corresponding to best survival, intermediate as the two middle quartiles and poor survival as the quartile corresponding to the worst survival.

In the next step, the relative risk of death among the index cases given the survival of their first-degree relative was estimated using the same Cox regression model as in the first step, additionally adjusting for potential family cluster effects. This analysis was performed among all index cases (N=1701) as well as limited to index cases that had the same type of lymphoid malignancy as their first-degree relatives (N=491), and among all lymphoid malignancies as well as among the seven categories hereof.

6 RESULTS AND INTERPRETATION

6.1 Study I: Possible interaction between cigarette smoking and human leukocyte antigen DRB1 in risk of follicular lymphoma

6.1.1 Results

We analyzed three amino-acid haplotypes and the shared epitope (SE) in HLA-DRB1 for interaction with smoking status in risk of follicular lymphoma. Relative to never smoking in combination with no SE alleles, current smoking in combination with 2 SE alleles was associated with an increased risk of follicular lymphoma among all subjects (OR_{former}=2.20, 95%CI 1.10, 4.41, $OR_{current}$=3.56, 95%CI 1.60, 7.92) and among women (OR_{former}=2.95, 95%CI 1.18, 7.37, $OR_{current}$=5.63, 95%CI 2.07, 15.3; Table 7). However, there was no such increase associated with smoking status among all subjects or women carrying no or one SE allele. This was also the case among men irrespective of number of SE alleles carried (Table 7). The attributable proportion (AP) of follicular lymphoma risk due to interaction was 60% among all subjects (AP=0.6, 95%CI 0.15, 1.0) and 50% among women (AP=0.5, 95%CI 0.005, 1.0) who were current smokers and carried 2 SE alleles (Table 7).

Table 7. Number of cases and controls and relative risk[a] of follicular lymphoma by smoking status and number of shared epitope alleles in all study subjects and stratified by sex.

Subgroup and Smoking Status	No. of Cases and Controls by No. of SE alleles and Smoking Status						Relative risk[a] of Follicular Lymphoma							
	0 SE Allele		1 SE Allele		2 SE Alleles		0 SE Allele		1 SE Allele		2 SE Alleles			
	No. of Cases	No. of Controls	No. of Cases	No. of Controls	No. of Cases	No. of Controls	OR	95% CI	OR	95% CI	OR	95% CI	AP[c]	95% CI
Overall														
Never	63	179	70	163	19	36	1.00	Referent	1.28	0.84, 1.94	1.36	0.71, 2.60		
Former	46	135	56	101	20	24	0.71	0.45, 1.14	1.21	0.84, 1.94	2.20	1.10, 4.41		
Current	37	74	42	90	20	13	1.08	0.65, 1.81	1.11	0.68, 1.81	3.56	1.60, 7.92	0.6	0.2-1.0
Men														
Never	28	36	29	39	9	13	1.00	Referent	0.95	0.48, 1.91	0.87	0.32, 2.36		
Former	24	56	37	40	10	8	0.52	0.26, 1.04	1.12	0.57, 2.23	1.61	0.55, 4.67		
Current	18	36	23	25	7	4	0.60	0.28, 1.28	1.02	0.47, 2.20	1.88	0.49, 7.24	NA	
Women														
Never	35	143	41	124	10	23	1.00	Referent	1.57	0.91, 2.71	1.70	0.72, 4.05		
Former	22	79	19	61	10	16	0.94	0.49, 1.78	1.23	0.62, 2.42	2.95	1.18, 7.37		
Current	19	38	19	65	13	9	1.91	0.94, 3.88	1.22	0.62, 2.38	5.63	2.07, 15.3	0.5	0.005-1.0

NA= not appropriate. [a] Relative risk was estimated as odds ratios, OR, with 95% confidence intervals, CI. Regression models were adjusted for country, age and sex (the latter only when men and women were analyzed together). [b] Shared epitope (SE) alleles were defined as HLA-DRB1*04 or HLA-DRB1*10, excluding DRB1*0103, HLA-DRB1*0402 and HLA-DRB1*0403 when four-digit data was available. [c] Attributable proportion (AP) of interaction is presented for individuals exposed to both current smoking and two SE alleles.

6.1.2 Interpretation

There is no other study exploring this or similar interactions in lymphoma risk to relate our results to. Accordingly, it is not possible to draw firm conclusions on the observed associations. If our results can be validated, it means that individuals carrying SE (or some other tightly linked amino-acid haplotype) are especially susceptible to follicular lymphoma upon smoking. From a public health perspective, this would mean that follicular lymphoma could be added to the list of diseases preventable by anti-smoking campaigns. The association could also teach us about mechanisms in lymphomagenesis. The differences in results observed for men and women could be due to considerably less men to analyze, resulting in less power than for the analysis of women. An interaction among men can thus not be excluded. Also, it is hard to imagine a biological explanation to the different results by sex observed in the study. Nevertheless, differences by sex has also been observed in studies of smoking in follicular lymphoma risk, as outlined above (see 2.2.1.1). Hopefully, we will see validation studies exploring this potential interaction in the near future.

6.2 Study II: Life style factors, autoimmune disease and family history in prognosis of non-Hodgkin lymphoma overall and by subtype

6.2.1 Results

6.2.1.1 All-cause death

Current compared to never smoking was associated with a worse survival among all NHL (HR= 1.5, 95% CI 1.2-1.8) and diffuse large B cell lymphoma (HR= 1.8, 95% CI 1.2-2.7) but not among other subtypes (Table 8). Former smoking was not associated with survival among patients with NHL overall or subtypes. Education (proxy for socioeconomic status) ≤9 years compared to >12 years was associated with a worse survival among NHL overall (HR= 1.3, 95% CI 1.1-1.7). Summer sunbathing one or more times per week compared to zero times per week 5-10 years before lymphoma diagnosis was inversely associated with all-cause death among follicular lymphoma patients (linear trend test P= 0.03). In mantle cell lymphoma, this association went in the opposite direction (P= 0.03). There was no association with sunbathing among all NHL combined or other subtype. BMI was not associated with survival among all NHL or subtypes. Autoimmune disease was positively associated with risk of death among all NHL patients (HR= 1.4, 95% CI 1.0-1.8, P< 0.05). Having a family history of hematopoietic malignancy compared to not having such a history was borderline statistically significantly associated with worse survival among all NHL patients (HR= 1.5, 95% CI 1.0-2.0, P> 0.05) and there were no statistically significant associations by subtype.

6.2.1.2 Lymphoma-related death

When exploring lymphoma-related death as the outcome, associations were generally weaker and statistically non-significant (data shown in supplement of published paper).[159] However, low versus high educational attainment was statistically significantly associated with lymphoma-related death among all NHL (HR= 1.3, 95% CI 1.0-1.6, P< 0.05) but not among subtypes. For smoking and autoimmune disease, the point estimates were similar for lymphoma-related death as those observed for all-cause death but did not reach statistical significance.

Table 8. Multivariable-adjusted hazard ratios with 95% confidence intervals for all-cause death among patients with all NHL or subtypes.

	All NHL	DLBCL	Follicular lymphoma	Mantle-cell lymphoma	T-cell lymphoma
Spread and Ann Arbor stage[1]					
Primary extranodal vs. nodal NHL	0.8 (0.6, 1.0)	0.9 (0.6, 1.2)	0.8 (0.4, 1.7)	0.4 (0.1, 1.6)	0.8 (0.4, 1.4)
Stages 1 and 2 vs. 3 and 4	0.5 (0.4, 0.6)	0.4 (0.3, 0.6)	0.4 (0.2, 0.6)	0.9 (0.4, 2.1)	0.5 (0.3, 1.1)
Sex[2]					
Male	1.4 (1.2, 1.7)	1.6 (1.2, 2.1)	1.4 (1.0, 2.0)	1.1 (0.6, 2.2)	1.0 (0.5, 1.8)
Smoking status[2]					
Never	Ref	Ref	Ref	Ref	Ref
Past	1.2 (1.0, 1.4)	1.1 (0.8, 1.4)	1.1 (0.8, 1.7)	0.7 (0.4, 1.3)	1.3 (0.7, 2.6)
Current	1.5 (1.2, 1.8)	1.8 (1.2, 2.7)	1.2 (0.7, 2.0)	0.5 (0.2, 1.2)	1.6 (0.8, 3.3)
Education (years)[2]					
>12	Ref	Ref	Ref	Ref	Ref
10–12	1.1 (0.9, 1.4)	1.1 (0.7, 1.6)	0.9 (0.5, 1.5)	0.7 (0.3, 1.5)	0.7 (0.3, 1.4)
≤9	1.3 (1.1, 1.7)	1.2 (0.8, 1.7)	1.0 (0.6, 1.8)	0.8 (0.3, 1.7)	1.5 (0.7, 3.0)
p-Value for trend	0.01	0.35	0.79	0.78	0.27
BMI (kg/m²)[2]					
<25.0	Ref	Ref	Ref	Ref	Ref
25.0–29.9	1.0 (0.8, 1.1)	0.8 (0.6, 1.0)	0.9 (0.6, 1.3)	1.0 (0.6, 1.8)	1.4 (0.8, 2.4)
≥30.0	1.2 (0.9, 1.5)	1.5 (1.0, 2.2)	0.8 (0.4, 1.5)	2.1 (0.8, 5.3)	0.5 (0.2, 1.6)
p-Value for trend	0.45	0.60	0.49	0.28	0.84
Sunbathing (5–10 years before baseline)[2]					
Never	Ref	Ref	Ref	Ref	Ref
≤1 Time/week	0.8 (0.7, 1.0)	1.1 (0.8, 1.6)	0.5 (0.3, 0.8)	0.9 (0.4, 1.8)	0.4 (0.2, 1.0)
2–3 Times/week	0.9 (0.7, 1.1)	1.2 (0.8, 1.8)	0.6 (0.4, 1.0)	1.3 (0.6, 2.9)	0.7 (0.3, 1.7)
4+ Times/week	0.9 (0.7, 1.1)	1.1 (0.8, 1.6)	0.5 (0.3, 0.9)	2.3 (1.1, 5.1)	0.5 (0.2, 1.2)
p-Value for trend	0.26	0.73	0.03	0.03	0.33
Autoimmune disease[2,3]	1.4 (1.0, 1.8)	1.2 (0.8, 2.0)	1.0 (0.4, 2.6)	0.7 (0.3, 1.7)	1.8 (0.8, 4.2)
Family history of hematopoietic malignancy[2]	1.5 (1.0, 2.0)	1.0 (0.5, 1.9)	1.1 (0.6, 2.3)	1.3 (0.4, 4.2)	0.9 (0.2, 4.0)

[1] Adjusted for sex and age. [2] Stratified by age and spread/stage (primary extranodal; nodal stages 1 and 2; nodal stages 3 and 4; missing), and adjusted for sex, smoking, education and BMI. [3] Self-reported RA, SLE, Sjögrens syndrome or celiac disease.

6.2.2 Interpretation

6.2.2.1 Smoking

Smoking at diagnosis was consistently associated with an increased risk of death among NHL patients in our study as well as in three previous studies, although somewhat different smoking exposure measures were used.[85,86,88] In addition, two more recent studies provide further evidence of this association.[87,89] In a meta-analysis of four of previously published studies, smoking duration, intensity and pack-years but not smoking status were all associated with all-cause death among NHL patients.[87] No study has reported a protective effect of smoking. Based on these six studies, it is reasonable to conclude that there is a link between smoking and survival among NHL patients.

Regarding subtypes, there is less data published and there are inconsistencies between studies. Diffuse large B cell lymphoma survival was associated with smoking in our study and one previous study,[88] while no association was observed in three studies including the meta-analysis.[85-87] Follicular lymphoma survival was associated with smoking in two out of five studies, including the meta-analysis.[85-89] Chronic lymphocytic leukemia survival was also associated with smoking in two out of three studies, although not all exposure measures were significant.[85,88,89] Less power when splitting up NHL into subtypes may be an explanation for the inconsistencies. More data is needed to draw conclusions on the effect of smoking by specific NHL subtypes. This is also the case for lymphoma-related/specific death.

It is well-established that smoking causes life-threatening conditions, such as cancer and lung and cardiovascular diseases, in the general population.[92] Comorbidities like these may cause a poorer performance status, which could make the caring physician to reduce treatment intensity with possible implications for prognosis.[94] Chemotherapy could also aggravate comorbid conditions, also increasing the risk of death.[94,95] Whether smoking directly interacts with cancer treatment is not known but could be a hypothesis.

6.2.2.2 Education – socioeconomic status

Low versus high socioeconomic status was consistently associated with increased risk of death among NHL patients in our study and two previous studies conducted in Denmark[97,98] and Scotland,[96] respectively. The Danish study also used attained education as measure of socioeconomic status (along with some other measures, as described in 2.4.6.2), while the Scottish study used residential area. We also explored and demonstrated the same association with lymphoma-related death. The other two studies did not explore this outcome, nor survival by subtypes. There is no study showing a null or conflicting result. Together, available data suggest that socioeconomic status influences survival after NHL,

possibly through differences in disease stage at diagnosis[96,98] and/or comorbidity (in parallel to the discussion for smoking above).[160] However, no conclusions can be drawn regarding the potential variation of effect of socioeconomic status on lymphoma-specific death, nor the effect on survival among specific subtypes.

6.2.2.3 Summer sunbathing habits 5-10 years before NHL diagnosis – UV radiation

A potential effect of UV radiation might go through vitamin D production, as UV radiation is the most important source of active vitamin D. In our study, we did not observe any clear or consistent associations between sunbathing habits or other UV radiation exposures and outcome of NHL. Vitamin D insufficiency was associated with increased risk of death among patients with diffuse large B cell lymphoma or T cell lymphoma but not among patients with follicular lymphoma or mantle cell lymphoma in one study.[104] Diffuse large B cell lymphoma was also associated with lymphoma-specific death in this study.[104] Among two cohorts of follicular lymphoma patients, vitamin D insufficiency was associated with lymphoma progression, while an association with survival was observed in just one of the two cohorts.[105] Although not conclusive, available data motivates further investigation of the effect of UV radiation and/or vitamin D status on survival of NHL and subtypes, since vitamin D supplementation would be an easy and readily available way to improve prognosis, should the findings be confirmed.

6.2.2.4 Body mass index – obesity

Our results do not support a link between obesity and survival in NHL patients overall or by subtype. Another three studies on the subjects have been published.[86,89,100] Two of these three studies showed that obesity versus normal-weight was associated with increased risk of all-cause and lymphoma-specific death among NHL patients.[86,89] The point estimates of the third study were in agreement with those of the two other studies but were not statistically significant.[100] Regarding subtypes, obesity was associated with increased risk of death among patients with chronic lymphocytic leukemia in one[89] of two studies.[100] Survival of diffuse large B cell lymphoma and follicular lymphoma patients has not been associated with obesity in the three studies, in line with our findings.[86,89,100] Hence, available data is not conclusive still. The prevalence of obesity was considerably lower among our study subjects compared to the three other studies (11% versus 18%[89] 20%[100] and 26%[86]). Possibly, there were also more individuals with extreme obesity in these other studies. The resulting higher power in these three studies could possibly explain differences in results. Further studies are needed to draw conclusions about the impact of obesity on outcome in NHL and subtypes.

As discussed in 2.4.6.3, a possible link between obesity and NHL survival is hypothetical. One proposed hypothesis is that insulin resistance due to obesity increases the level of insulin-like growth factor-1, leptin and cytokines, which all could stimulates cell growth and proliferation.[101-103]

6.2.2.5 Autoimmune disease

In our study, autoimmune disease was a composite variable of four diseases: RA, SLE, primary Sjögren's syndrome and celiac disease. Although different diseases, it seems reasonable to assume that they all might confer worse or unchanged but not improved survival among NHL patients, perhaps through the associated long-standing inflammatory state. Our and other studies provide conflicting results in this area. Mikuls et al compared NHL patients with and without RA and found no association with all-cause death but observed a decreased risk of lymphoma-specific death and an increased risk of deaths due to other causes among NHL patients with concomitant RA.[108] This latter finding could possibly be due to competing risks, since RA patients have an increased risk of death compared with the general population.[161] Ji et al on the other hand found an increased risk of both all-cause and lymphoma-specific death among NHL patients with RA compared to NHL patients without RA.[107] One study has assessed the impact of celiac disease on NHL survival and found no association.[139]

Taken together, there is no clear answer to the hypothesis that having an autoimmune disease confers a worse prognosis when diagnosed with a NHL. The conflicting results could possibly be due to excess risk of death for some NHL subtypes but not others, and the proportion of these could vary between studies. Rituximab has improved the outcome of both B cell NHL and RA since it became available in the beginning of the millennium. Existing studies were conducted during somewhat different time periods (our study 1999-2010, Mikuls et al 1984-2002, Ji et al 1961-2006). Consequently, the proportion of patients that received rituximab as part of their NHL and/or RA treatment is probably different in the three studies, which possibly could influence the association between RA and NHL survival. A few studies have shown that autoimmune diseases such as RA may be linked specifically with the development of diffuse large B-cell lymphoma of the non-germinal center or activated B-cell type, which could have implications for survival.[110]

6.2.2.6 Family history of a hematopoietic malignancy

There are no data that support an association between family history of a hematopoietic malignancy (our study), any lymphoma,[112] or NHL[113] and survival of NHL patients overall, although the association was borderline statistically significant in our study. There is one study of morbus Waldenström patients, which found an increased risk of death among those with a first-degree relative with any lymphoproliferative disorder compared to sporadic cases.[162] One study on survival among patients with familial versus sporadic chronic lymphocytic leukemia found no difference.[163]

One explanation for the general lack of association could of course be that no link exists between family history of a hematopoietic malignancy and NHL survival. Another explanation could be that some familial NHL cases may inherit protective traits while others inherit deleterious traits. Lumping together all familial cases may thus hide potential differences in survival. In line with this reasoning, one of the studies mentioned above demonstrated a non-significant concordance in survival among familiar cases.[113] Furthermore, several candidate gene studies have demonstrated both protective and deleterious single nucleotide polymorphisms in survival of all NHL and subtypes.[114,115,118-121,124] We further explored this further in Study III and IV.

6.3 Study III: A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival

6.3.1 Results

6.3.1.1 Genome-wide association studies and meta-analysis

We found no SNP associated with lymphoma-specific death at the genome-wide significance level of P≤ 5.0 x10^{-8}. We found four SNPs associated at a suggestive level (P≤ 10^{-6}; Table 9). The smallest P value was borderline statistically significant (P= 5.24 x10^{-8}) and was observed for a SNP on chromosome 17q in the ABCA10 gene. The other three were located on chromosome 1 in gene DAB1 (P= 2.22 x10^{-7}), on chromosome 4 in-between genes (P= 2.24 x10^{-6}) and on chromosome 19 in the KLK11 gene (P= 9.38 x10^{-6}).

6.3.1.2 SNPs associated with follicular lymphoma outcome in previous studies

Most SNPs reported to be associated with follicular lymphoma outcome in at least one previous study were not associated with lymphoma-specific survival or lymphoma progression in our study (Table 10). However, the SNP rs2466571 in *CD46* (HR= 1.37, 95% CI

Table 9. The top single nucleotide polymorphisms (SNP) associated with lymphoma-specific death[A] ($P_{random} \leq 10^{-6}$) in the meta-analysis of FL patients in SCALE and UCSF (N=586). SNPs in strong linkage disequilibrium ($r^2 \geq 0.8$) with the top SNPs but with larger P values are not included.

						SCALE		UCSF		Meta-analysis		Heterogeneity				
Chr	SNP	Position	A1	A2	MAF	HR (95% CI)	p_{SCALE}	HR (95% CI)	p_{UCSF}	HR (95% CI)	p_{RANDOM}	p_{HET}	I^2	Gene	Left gene	Right gene
17	rs10491178	64661568	A	G	0.06	3.10 (1.97; 4.89)	1.13E-06	3.50 (1.28; 9.53)	1.36E-02	3.17 (2.09; 4.79)	5.24E-08	0.83	0	ABCA10	ABCA6	LOC100133319
1	rs3131729	57971355	A	G	0.13	2.58 (1.77; 3.76)	9.24E-07	2.00 (0.92; 4.31)	7.58E-02	2.45 (1.75; 3.44)	2.22E-07	0.56	0	DAB1	C8B	LOC729423
4	rs11932201	14304530	C	A	0.17	2.10 (1.48; 2.97)	2.93E-05	2.50 (1.13; 5.55)	2.29E-02	2.16 (1.57; 2.97)	2.24E-06	0.69	0	NA	LOC152742	LOC441009
19	rs2250066	56220931	A	G	0.15	2.07 (1.45; 2.97)	7.31E-05	2.35 (1.01; 5.44)	4.40E-02	2.11 (1.52; 2.94)	9.38E-06	0.79	0	KLK11	KLK10	KLK12

A1= minor allele, A2= major allele, MAF= minor allele frequency, NA= not applicable;
[A] Estimated with hazard ratios, HR, and 95% confidence intervals, CI, adjusting for sex, age at diagnosis and three principal components. The minor allele was investigated for association. [B] MAF was calculated in SCALE.

Table 10. Relative risk[A] of lymphoma-specific death and lymphoma progression for selected SNPs previously associated with any follicular lymphoma outcome in at least one previous study.

							Lymphoma-specific death		Progression[B]	
Chr	Gene	SNP	Position	A1	A2	MAF	HR (95% CI)[A]	p-value	HR (95% CI)[A]	p-value
1	*CD46*	rs2466571	206006669	A	C	0.49	0.98 (0.73; 1.32)	0.90	1.37 (1.10; 1.72)	0.006
1	*CD55*	rs2564978[C]	205561039	T	C	0.31	0.88 (0.62; 1.25)	0.47	1.13 (0.86; 1.49)	0.38
1	*CFH*	rs1065489	194976397	A	C	0.15	0.89 (0.57; 1.39)	0.60	0.92 (0.65; 1.31)	0.66
1	*CFH*	rs1329423[C]	194913010	C	T	0.22	0.94 (0.66; 1.39)	0.81	1.06 (0.79; 1.42)	0.71
1	*CFH*	rs3766404	194918455	G	A	0.16	1.14 (0.77; 1.68)	0.51	1.12 (0.80; 1.55)	0.52
1	*CFHR5*	rs6694672[C]	195212412	G	T	0.09	1.01 (0.61; 1.67)	0.97	1.02 (0.66; 1.57)	0.93
1	*FCGR2A*	rs1801274	159746369	A	G	0.47	0.76 (0.56; 1.03)	0.08	0.86 (0.68; 1.08)	0.19
1	*MTHFR*	rs1801131[C]	11777063	G	T	0.32	0.69 (0.49; 0.97)	0.03	0.59 (0.45; 0.77)	0.0001[E]
1	*SELE*	rs5361	167967684	C	A	0.11	1.04 (0.64; 1.69)	0.89	0.90 (0.61; 1.32)	0.59
2	*IL1RN*	rs454078[C]	113605264	T	A	0.28	0.99 (0.69; 1.42)	0.96	0.94 (0.72; 1.23)	0.66
3	*FTHFD*	rs1127717[C]	127308749	C	T	0.16	0.91 (0.58; 1.43)	0.69	0.98 (0.70; 1.38)	0.92
4	*IL2*	rs2069762[C]	123597430	C	A	0.26	0.63 (0.43; 0.92)	0.02	0.89 (0.69; 1.16)	0.38
4	*IL8*	rs4073[C]	74824888	A	T	0.46	0.93 (0.70; 1.25)	0.64	0.78 (0.62; 0.97)	0.02
4	*IL8*	rs2227307[C]	74825533	G	T	0.45	0.93 (0.69; 1.25)	0.63	0.75 (0.60; 0.94)	0.01
5	*C9*	rs1421094[C]	39391348	A	G	0.37	1.19 (0.88; 1.61)	0.27	1.00 (0.79; 1.27)	0.98
5	*IL12B*	rs3212227[C]	158675528	G	T	0.17	1.21 (0.81; 1.80)	0.36	1.42 (1.03; 1.94)	0.03
6	*C6orf15*	rs6457327[D]	31182009	A	C	0.42	0.90 (0.66; 1.24)	0.53	0.90 (0.71; 1.14)	0.38
6	*IFNGR1*	rs3799488	137561473	G	A	0.12	0.76 (0.46; 1.24)	0.27	1.12 (0.79; 1.57)	0.53
8	*GGH*	rs719235[C]	64114235	A	C	0.32	0.96 (0.69; 1.34)	0.82	1.00 (0.77; 1.29)	0.99
9	*GALNT12*	rs10819377	100643533	A	G	0.46	0.76 (0.56; 1.03)	0.07	0.89 (0.71; 1.12)	0.33
16	*IL4R*	rs1801275	27281901	G	A	0.25	0.91 (0.64; 1.30)	0.62	0.85 (0.65; 1.12)	0.26
22	*MIF*	rs755622[C]	22566392	C	G	0.16	0.76 (0.49; 1.16)	0.20	0.85 (0.61; 1.18)	0.34

[A] Estimated with hazard ratios, HR, and 95% confidence intervals, CI. Adjusted for sex, age at diagnosis and three principal components. [B] Swedish cases only (N=231) for time to progression. [C] Imputed SNP.
[D] rs6457327 has previously been associated with both FL risk and prognosis [E] Opposite direction of association to previous study.[121]
The minor allele (A1) was investigated for association except for rs2466571 (*CD46*), where the major allele (A2) was used as reference for easier comparison with previous study results.

1.10-1.72, P=0.006), two tightly linked SNPs (r^2=0.99) in *IL8* (rs4073 HR= 0.78, 95% CI 0.62-0.97, P= 0.02; rs2227307 HR= 0.75, 95% CI 0.60-0.94, P= 0.01) were associated with lymphoma progression. The smallest P value in our study (P= 0.0001) was observed for a SNP in *MTHFR* in association with lymphoma progression. One SNP in *IL2* was significantly associated with lymphoma-specific death. After Bonferroni correction for multiple testing (2 tests in 27 SNPs= 54 test) only the association with the SNP in *MTHFR* remained significant (corrected P= 0.01).

6.3.2 Interpretation

The new results we provide in the GWAS as well as the validation of previous findings are not enough to draw firm conclusions about whether these SNPs have true effects on follicular lymphoma outcome. None of our top findings (the top SNPs of the GWAS and the three SNPs of the validation study) were in fact statistically significant after taking multiple testing into account (either by setting a conservative significant level of $P \leq 5.0 \times 10^{-8}$ or after Bonferroni correction). Nevertheless, the results are hypothesis generating. The top SNP of the GWAS was very close to being statistically significant and was located in *ABCA10* and in a locus with other ABC transporter genes that theoretically could influence drug resistance.[164,165] It would be interesting to see this finding (and also the other top hits) validated in other populations.

The two SNPs in *IL8* and the one in *CD46* in the validation part of our study were at least nominally statistically significant (P< 0.05) in two or three cohorts, including our own. Hence, these results are the strongest in our investigation. IL8 is a major mediator of the inflammatory response and a potent angiogenic factor. CD46 is a cell membrane protein that inhibits complement components, protecting the cell from damage. Although any potential biological effect of the SNPs is unknown, their location in theoretically relevant genes along with the results from us and others motivate further studies to validate the results in other populations.

Our results regarding the SNPs in *MTHFR* and *IL2* are in conflict with previously reports, since the point estimate went in the opposite direction.[121,166]

6.4 Study IV: Concordance in survival of first-degree relatives diagnosed with a lymphoid malignancy

6.4.1 Results

We show evidence of a familial concordance in survival among first-degree relatives with a lymphoid malignancy (Table 11). A concordant survival was observed only among index cases with first-degree relatives that had the same type of lymphoid malignancy that were affected by any lymphoid malignancy (good vs poor: HR=1.86, 95% CI 1.31-2.63, P for trend= 0.0004), a NHL (good vs poor: HR=1.86, 95% CI 1.21-2.86, P for trend= 0.006) or an indolent lymphoma including chronic lymphocytic leukemia (good vs poor: HR= 2.62, 95% CI 1.14-6.03, P for trend= 0.03). Similar but non-significant trends were also observed for first-degree relatives with chronic lymphocytic leukemia (P for trend= 0.08) or plasma cell malignancies (P for trend= 0.14). There was no association among patients with Hodgkin lymphoma, aggressive lymphoma or acute lymphocytic leukemia, although the number of observations here were small. When analyzing all index cases (both those with concordant and those with non-concordant subtypes as their first-degree relative), there was no statistically significant concordance in survival, although a borderline P value for trend was observed for all subtypes combined (P for trend= 0.06). Also of note, we did not see a familial concordance in survival among index cases with first-degree relatives with non-concordant type of lymphoid malignancy (data not shown).

6.4.2 Interpretation

There is one related, small study that in complementary analyses investigated concordance in survival between 98 parents and 98 offspring with NHL and found a non-significant trend toward better survival in offspring that had parents that survived ≥24 months after NHL diagnosis as compared to offspring with parents that survived <24 months.[113] Our study was much larger and accounted for what was good, expected and poor survival among patients diagnosed in different time periods, ages of diagnosis, sex, socioeconomic status and, importantly, with different types of lymphoid malignancies. The well-known higher heritability of indolent lymphoma risk, especially chronic lymphocytic leukemia, compared to aggressive lymphomas[167,168] gave rise to much higher numbers of indolent lymphoma forms and thus also considerably higher power for analyses for these groups. Perhaps not surprisingly therefore, it was among indolent lymphomas that a statistically significant familial concordance in survival was observed. The vast majority of these were chronic lymphocytic leukemia patients and although the point estimates imply the same trend in this subtype alone it was not statistically significant. Taken together, our data clearly suggest

Table 11. Hazard ratio (HR) of death with 95% confidence intervals (CI) in index cases given the survival of their respective first-degree relative (good, expected or poor). Analyses were performed among all types of lymphoid malignancies together and separated by type among all index cases and limited to index cases with a first-degree relative with a concordant subtype.

		All index cases						Index cases with first-degree relatives with concordant subtype					
Subtype	**Survival of first degree relative**	**# obs.**	**# deaths**	**deaths/1000 person-years**	**HR**	**95% CI**	**P**	**# obs.**	**# deaths**	**deaths/1000 person-years**	**HR**	**95% CI**	**P**
All subtypes	Good	350	157	75	Referent			83	40	73	Referent		
	Expected	866	388	80	1.18	0.99-1.42	0.07	258	130	88	1.53	1.11-2.11	0.009
	Poor	485	230	86	1.23	1.00-1.51	0.050	151	78	93	1.86	1.31-2.63	0.0005
	P for trend				0.06						0.0004		
Hodgkin lymphoma	Good + expected	81	13	19	Referent			30	6	29	Referent		
	Poor	31	6	23	1.21	0.35-4.17	0.73	13	4	30	0.75	0.11-5.21	0.77
Non-Hodgkin Lymphoma	Good	293	135	84	Referent			62	31	82	Referent		
	Expected	772	352	86	1.20	0.99-1.46	0.06	215	110	92	1.53	1.04-2.27	0.03
	Poor	424	204	91	1.20	0.97-1.49	0.09	127	68	108	1.86	1.21-2.86	0.004
	P for trend				0.12						0.006		
Aggressive lymphoma	Good + expected	192	83	104	Referent			21	7	95	Referent		
	Poor	65	33	133	1.05	0.70-1.59	0.81	12	6	132	1.94	0.67-5.60	0.22
Indolent lymphoma including CLL	Good	139	48	57	Referent			27	8	45	Referent		
	Expected	382	124	51	1.21	0.87-1.67	0.25	100	36	55	1.76	0.84-3.71	0.14
	Poor	208	77	62	1.20	0.83-1.75	0.33	67	27	75	2.62	1.14-6.03	0.02
	P for trend				0.38						0.03		
CLL	Good	82	34	67	Referent			23	7	46	Referent		
	Expected	225	87	59	1.15	0.79-1.68	0.48	92	35	57	1.77	0.79-3.97	0.17
	Poor	123	48	65	1.14	0.72-1.80	0.57	52	22	76	2.30	0.95-5.55	0.07
	P for trend				0.61						0.08		
Indolent lymphoma excluding CLL	Good	57	14	39	Referent			4	1	40	Referent		
	Expected	157	37	41	1.19	0.63-2.42	0.59	8	1	27	Unestimable		
	Poor	85	29	57	1.32	0.67-2.58	0.42	15	5	71			
	P for trend				0.43								
ALL	Good + expected	35	25	251	Referent			7	5	244	Referent		
	Poor	8	5	131	0.75	0.30-1.92	0.55	2	2	580	2.31	0.43-12.4	0.33
Plasma cell malignancy	Good	56	37	153	Referent			16	12	153	Referent		
	Expected	146	92	142	1.05	0.72-1.54	0.79	64	41	148	1.12	0.63-2.00	0.69
	Poor	83	55	172	1.30	0.86-1.96	0.22	27	19	184	1.51	0.87-2.61	0.14
	P for trend				0.20						0.14		

CLL= chronic lymphocytic leukemia, ALL= acute lymphoblastic leukemia

some heritability in survival among NHL patients, particularly among indolent lymphoma patients. The fact that first-degree relatives with different types of lymphoid malignancies did not demonstrate a similar familial concordance in survival makes variation in genetic factors related to lymphoma biology a more likely explanation than shared environmental factors or general genetic variation affecting for instance normal ageing. However, considerations regarding mechanisms remain speculative.

6.5 CONSIDERATIONS OF BIAS

6.5.1 Study I: Possible interaction between cigarette smoking and human leukocyte antigen DRB1 variation in risk of follicular lymphoma

The study was of case-control design. It is thus important that the distributions of the risk factors among the control subjects are representative of the distributions in the population that gave rise to the follicular lymphoma cases. Unfortunately, material for DNA isolation was discarded among Swedish control subjects included in SCALE. Therefore, these subjects were not possible to use in this study. Instead, we borrowed Swedish control subjects from the EIRA study. These were randomly selected from the population that gave rise to the rheumatoid arthritis patients of the study, i.e. the population of Sweden (except the most Northern part). Thus far they should be representative also of the vast majority of the Swedish follicular lymphoma cases. The EIRA control subjects were matched against the rheumatoid arthritis patients on age, sex and regions of Sweden. Typical age at rheumatoid arthritis onset is 50-60 years and 70% are women, while these numbers in follicular lymphoma are 65 years and 50%.[169,170] This was reflected in our study in that the EIRA control subjects consisted of more women than the Swedish SCALE control subjects (72% versus 56%), but the age distribution was very similar (interquartile range 48-62 versus 50-62 years). The EIRA study also spanned over a longer time period. These circumstances could raise concern for the EIRA control subjects not being representative of the source population and thus cause biased estimates. HLA-DRB1 variation ought to be fairly constant over the study time period and also by age and sex, however. Smoking habits could vary by these factors, though. To make EIRA control subjects as representative as possible, only those included 1999-2003 were used in the study. Among these, the proportion never, former and current smokers among EIRA and Swedish SCALE controls were very similar (49%, 31%, 20% versus 48%, 33%, 19%). After this exercise we felt reasonably safe that the EIRA control subjects would be valid as comparators in our study.

We used imputed amino-acid data. We started out with genotyped genome-wide SNP data to densely impute approximately 72 000 SNPs across the HLA region, followed by imputation of 372 amino-acid positions and 263 HLA alleles. The accuracy of the imputations was tested in a subset of the samples and was found to be very high (>98% concordance). Misclassifications should thus be very few and also likely be randomly distributed among case and control subjects.

6.5.2 Study II: Life style factors, autoimmune disease and family history in prognosis of non-Hodgkin lymphoma overall and by subtype

Study II was a population-based cohort study (based on a case-control study, SCALE), covering the whole of Sweden. The participation rate was high: 84% for NHL overall, 84% for diffuse large B cell lymphoma and 94% for follicular lymphoma. The most common reason for non-participation among patients with diffuse large B cell lymphoma was early death (6%). This number among follicular lymphoma patients was 1%. Hence, there may have been a non-random selection process among patients with diffuse large B cell lymphoma (and possibly other aggressive NHL) in that the most aggressive cases in particular may have been missed. We can therefore not be sure that our results would be applicable to the most aggressive variants of diffuse large B cell lymphoma. The possible small selection among cases with follicular lymphoma (and perhaps other indolent lymphomas) should likely not have any practical consequence.

Data on autoimmune disease, BMI, smoking, sun exposure and education attained were collected shortly after lymphoma diagnosis through a telephone interview. Retrospective collection of data may always be influenced by the study subjects' ability to remember or willingness to report the actual value of a variable. This may not be a random process with regard to the outcome measure, especially in a case-control study, leading to misclassification that could give biased estimates. In this cohort investigation, however, all study subjects were lymphoma patients and they were unaware of the outcome when data was collected. Any misclassification would thus be randomly distributed among all study subjects, most often leading to a bias toward the null.

To explore UV radiation and socioeconomic status we used proxy variables (summer sunbathing frequency the past 5-10 years and educational level attained). It can always be debated how well these proxies capture the exposure of interest, perhaps in particular our proxy for UV radiation exposure.

Our primary endpoint was all-cause death, which is not easily misclassified. For aggressive lymphomas, it is a fair measure of outcome, since these diseases progress fast and survival is short unless the patient is cured. Indolent lymphomas, however, are more like chronic diseases with a median survival of at least ten years. Since median age at diagnosis is high, many of the patients die from other causes than lymphoma. All-cause death may not be a good measure of prognosis in this group. Instead, lymphoma-related or lymphoma-specific death may better capture lymphoma progression.

We did not have data on level of hemoglobin, lactate dehydrogenase, performance status, molecular prognostic markers or treatment. With the exception of performance status, it is hard to imagine how these established prognostic factors would be associated with the factors we investigated, however. Thus, it is uncertain to what extent not adjusting for them resulted in biased estimates. However, smoking, socioeconomic status, BMI and autoimmune disease are most likely associated with performance status through comorbidity (Figure 18). Comorbidity may in turn influence treatment intensity with possible implications for outcome, in particular among aggressive lymphomas that need intensive treatment to achieve cure. In this way of reasoning, performance status is on the causative path from exposures to outcome. Hence, we would possibly take away the association we were actually interested in measuring by adjusting for performance status. Autoimmune disease and family history of a hematopoietic malignancy may be associated with disease aggressiveness with implications for outcome. Disease aggressiveness may be reflected in level of hemoglobin and lactate dehydrogenase, as well as in disease stage at diagnosis. It could thus be argued that these factors should perhaps not be adjusted for either. Naturally, we were not able to account for unknown confounders.

Nowadays, B cell lymphomas are treated with rituximab in addition to chemotherapy, which has improved survival for this group of patients. Whether the statistically significant results observed in our study would remain in the rituximab era is uncertain.

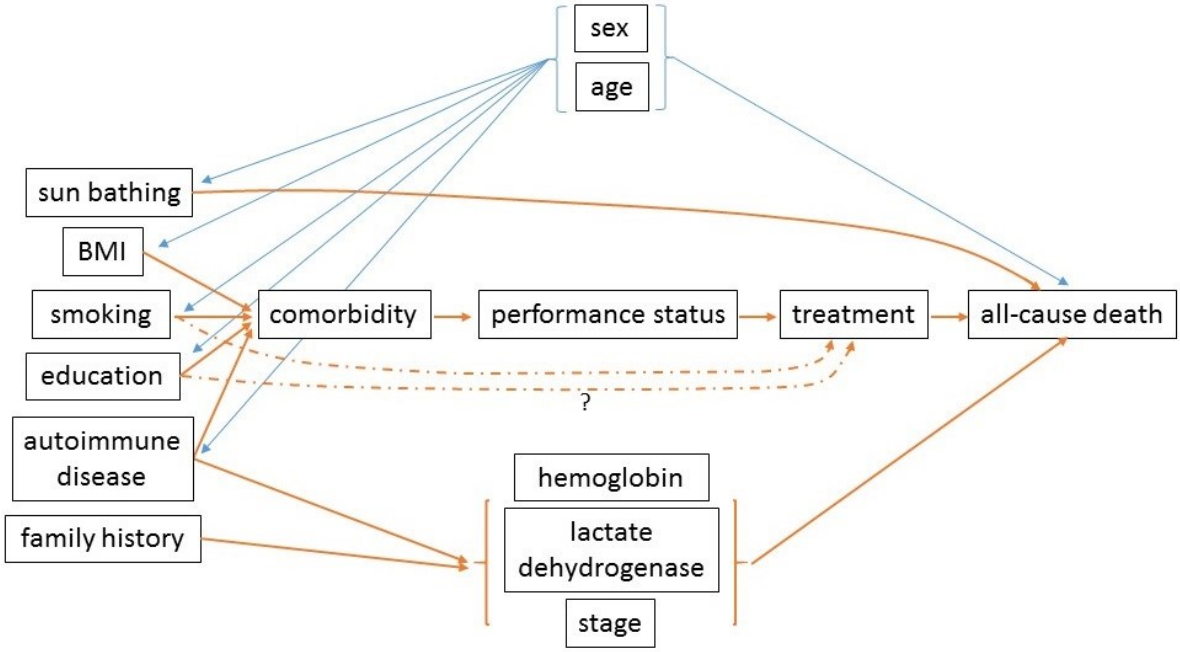


Figure 18. Directed acyclic graph for the associations investigated in Study I. The solid orange lines represent possible paths between investigated prognostic factors and one outcome. The dashed orange lines illustrates another possible path of association. The solid blue lines illustrates confounding associations.

6.5.3 Study III: A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival

Study III was based on two cohorts: 1) SCALE, covering the whole of Sweden and Denmark (see 4.1), and 2) UCSF, covering the San Francisco Bay area (see 4.3). The participation rate was 94% for follicular lymphoma in SCALE and 69% overall in UCSF. Of those included in each study, blood samples were collected from 84% of follicular lymphoma cases in SCALE and 87% of all NHL cases in UCSF. Even though the UCSF cohort was of European ancestry it may be genetically somewhat different from the SCALE cohort. To account for variation in population makeup, both SCALE and UCSF cohorts were adjusted for population stratification and we used a random effects model when pooling the results. Furthermore, we do not know the reasons for nonparticipation in UCSF or not giving blood in SCALE and UCSF. Hence, we cannot excluded a non-random selection process here that could give rise to biased estimates.

We imputed the values for six SNPs in the GWAS and 13 SNPs in the validation part of our study. Even though we used a well-established reference population (1000 Genomes) and only used imputed SNPs with high information score (>0.8), call rate (>0.9) and high

probability of being correct (>0.9) there may be some small misclassification. The misclassification should be random with regard to the outcome, however.

We used lymphoma-specific death and lymphoma-progression as outcome measures in the study, both of which have the potential for misclassification. Any misclassification ought to be random, possibly reducing power. With regard to lymphoma progression, we used time till start of first line or second line treatment or lymphoma-specific death if no second line treatment was given as proxy. This definition may not be accurate. Because indication for treatment is symptoms due to follicular lymphoma, there is not much reason to look for tumor progression before the patient gets symptoms that could be related to that. The time elapsing from establishing tumor progression until start of treatment could also vary depending how fast the tumor grows. Although it may not be an accurate definition it still provides a measure of how fast the tumor grows that should work for our purposes.

We performed multiple testing, indeed. This increases the risk of chance findings. This is why by convention the significance level is set at $P \leq 5.0 \times 10^{-8}$ in GWAS studies these days. This P value corresponds to a significance level of 0.05 corrected for 1 000 000 tests using the Bonferroni method ($0.05/1\,000\,000 = 5.0 \times 10^{-8}$). This significance level is adopted to the number of SNPs normally found on genotyping chips nowadays. Hence, one could argue that this level was a bit too conservative for our GWAS, in which we conducted about 330 000 tests ($0.05/330\,000 = 1.5 \times 10^{-7}$). In the validation part of our study, we used the Bonferroni method to account for multiple testing. We have thus accounted for multiple testing. Chance findings cannot be excluded nevertheless.

It is hard to imagine how factors in the follicular lymphoma international prognostic index (see 2.4.2.1) would affect germline genetic polymorphism. If anything, they may be on the association path from genetic trait to outcome (Figure 19). It is thus not obvious that these should be accounted for in the analysis. We included age and sex in the Cox model although also these variables do not clearly stand to reason. It has been practice in previous studies, however, and to present comparable results, we also accounted for these features. This may have mitigated potential associations under study. Other genetic traits may possibly confound our estimates but we do not know anything about this. Also, at this point we were not primarily trying to estimate the effect of each SNP on the outcome as accurately as possible but rather to find these potential risk SNPs for further validation in other cohorts.

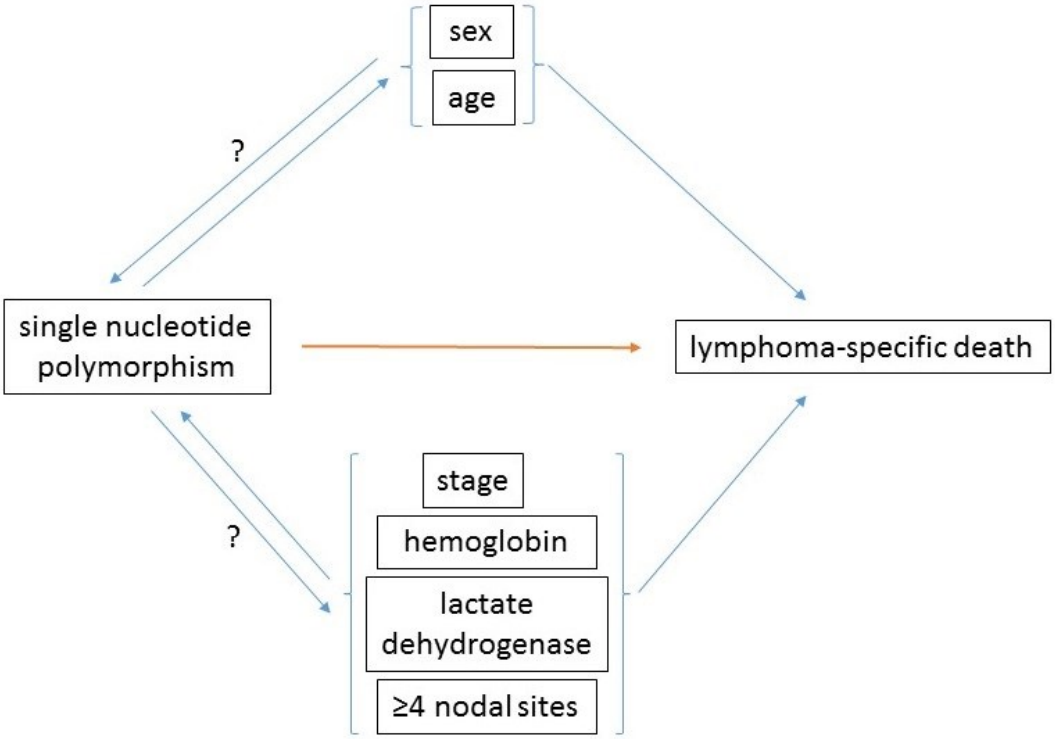


Figure 19. Directed acyclic graph of the associations under study in Study III (solid orange line). The solid blue lines illustrates possible confounding associations.

6.5.4 Study IV: Concordance in survival of first-degree relatives diagnosed with a lymphoid malignancy

This study was a register based cohort study covering the whole of Sweden. Apart from administrative limitations (such as the Multi-generation register being incomplete with regard parents data for individuals diseased before 1 July 1991), the national registers aim to include all eligible residents of Sweden. The registers used and the years covered in Study IV should be close to complete with regard to collected data. The largest proportion of missing data was observed for socioeconomic status (6% overall), collected from the censuses. This did not seem to be a random process, because those with unknown socioeconomic status had significantly worse survival relative to those with high socioeconomic status. In the Cox model, we adjusted for high, intermediate, low as well as having an unknown socioeconomic status.

The drawback with using the national registers is the lack of clinical data. We were unable to do analyses based on detailed information on for example lymphoma subtype and account for prognostic scores or treatments. However, we could often distinguish Hodgkin from non-Hodgkin lymphoma and plasma cell malignancies; aggressive from indolent lymphoma; chronic lymphocytic leukemia from other indolent lymphomas. In a rough way, differences in treatments were accounted for by adjusting for year of diagnosis. Parallel to my reasoning in Study III (see 6.5.2), it is not obvious which prognostic factors, such as prognostic scores, that would bias our estimates and should be accounted for in the regression models (Figure 20). Just like in Study III, these may be on the association path between germline genetic variation and survival.

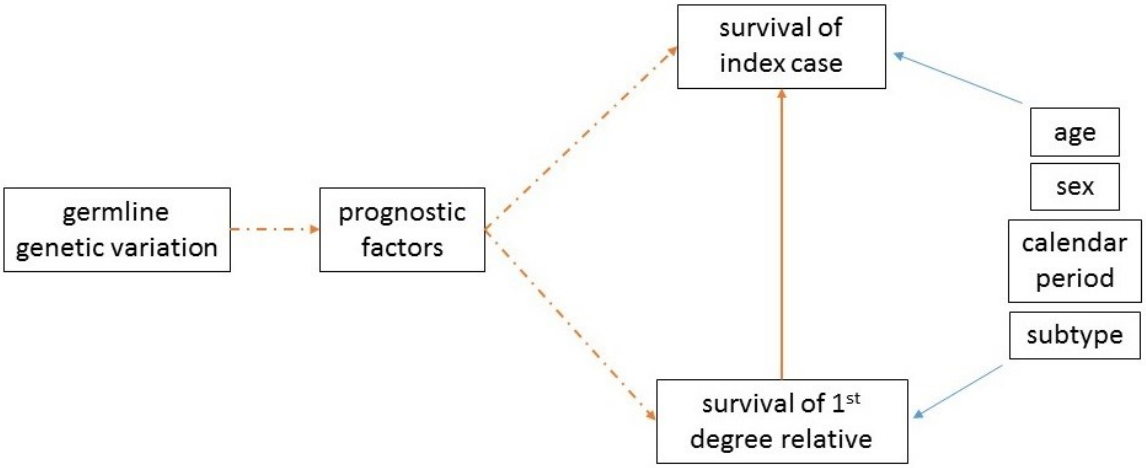


Figure 20. Directed acyclic graph illustrating the association under study in Study IV (solid orange line). The dashed orange line illustrates the hypothetical mechanism that we tried to capture with our study. The solid blue lines represents confounding associations.

7 CONCLUSIONS AND FUTURE PERSPECTIVES

7.1 Study I: Possible interaction between cigarette smoking and human leukocyte antigen DRB1 variation in risk of follicular lymphoma

The result of Study I supports our hypothesis of an interaction between smoking and HLA-DRB1 polymorphisms in risk of follicular lymphoma. The mechanisms behind the putative association is unclear. However, the amino-acid positions in question are located in the antigen-binding groove of the HLA-DRB1 molecule, suggesting a role for differential antigen binding and/or presentation properties. It motivate confirmatory studies in larger datasets. The idea has been introduced in the InterLymph consortium and hopefully we will be able to perform such a validation study soon. The result also inspire studies of interaction between other environmental and genetic factors in follicular lymphoma and other NHL subtypes to help unravel mechanisms of association.

7.2 Study II: Lifestyle factors, autoimmune disease and family history in prognosis of non-Hodgkin lymphoma overall and subtypes

Available data clearly suggests that lifestyle factors, such as smoking and socioeconomic status, have an impact on survival after NHL diagnosis. Both low socioeconomic status and smoking are associated with increased occurrence of poor health due to comorbidities, which may explain the observed associations. Further investigation of these factors in larger datasets to confirm the associations and figure out the mechanisms behind would be interesting and could have clinical implications, e.g. to understand if associations are mediated through comorbidities or if they could be explained by differences in disease characteristics or treatment. Current prognostic scores used in clinical settings do not include host factors other than age and performance status, which is likely to provide only a very crude measurement of the host influence of survival.

Regarding the impact of autoimmune disease on survival of NHL patients, available data motivates further studies, preferably both by specific autoimmune disease and by NHL subtype. To identify a sufficient number of these relatively rare patients, international collaboration will be necessary. InterLymph could be a good forum for this, or other settings such as the Nordic Lymphoma group.

The mere fact that a patient with a lymphoid malignancy has a first-degree relative with a lymphoid malignancy does not seem to help to predict outcome. Risk factors for survival

may be subtype specific. Also, some inherited traits could be protective while others could be deleterious. As a suggestion, future studies are preferably conducted by subtype of lymphoid malignancy and take into consideration the fact that inherited traits could be both protective and deleterious (perhaps by defining survival using deviance residuals or the like, as in Study IV).

7.3 Study III: A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival

Our study and others suggest a link between germline genetic variation and survival of NHL patients. Gene-expression profile studies and studies of the impact of the tumor microenvironment provide a (for the time being) hypothetical link. Although several candidates have been suggested, no specific single nucleotide or other polymorphism has yet been confidently identified. Based on Study III and previous studies, genetic variation in *IL8* and *CD46* are so far the strongest candidates to have prognostic impact in follicular lymphoma, suggesting importance of variation in immune function genes. Much is left to be done to confirm a link between germline genetic polymorphism and survival in NHL and to figure out the biological mechanisms behind.

7.4 Study IV: Concordance in survival of first-degree relatives diagnosed with a lymphoid malignancy

The results of Study IV provide further evidence of the hypothesis that common variants in the constitutional DNA may influence survival of patients with a lymphoid malignancy, although survival concordance in families may also be mediated through other mechanisms. The study supports such a link primarily among indolent lymphomas. Although no significant association was observed among patients with Hodgkin lymphoma, aggressive lymphoma or acute lymphoblastic leukemia, the few patients available for analyses within these groups preclude definitive conclusions. Given high quality register in the Nordic countries, collaboration within the Nordic Lymphoma group might provide a larger dataset that could validate the results of the present study and possibly allow analyses by subtypes.

7.5 FUTURE PERSPECTIVES

- A validation study of the interaction between smoking and HLA-DRB1 variation in the risk of follicular lymphoma has been proposed to InterLymph. We will most likely see such a study in a near future.

- We can expect to see studies of interaction between other environmental and genetic factors in the risk of follicular lymphoma as well as in other NHL subtypes. Although interaction studies are challenging to conduct, it is plausible that genes and environment interact in disease development.

- The role of host factors in the prognosis of NHL subtypes should be further explored within large international collaborations with information about established prognostic factors and ideally among patients following similar treatment schemes. Other host factors than age, sex and performance status may prove valuable when predicting the prognosis of NHL patients in the future. With the introduction of new targeted treatments in NHL, associations between host factors and outcome may also be subject to change.

- We will likely see more comprehensive studies of the prognostic impact of host genetic variation. This area may provide some answers to the variation in clinical trajectories between seemingly identical NHL that we cannot explain today. Ultimately, increased knowledge of this field will hopefully lead to predictive markers of response to a certain treatments and/or new, effective treatment approaches.

8 POPULÄRVETENSKAPLIG SAMMANFATTNING PÅ SVENSKA

Denna avhandling handlar om en grupp elakartade sjukdomar som kan samlas under rubriken non-Hodgkin lymfom (NHL). Det är en något intetsägande rubrik – det enda den egentligen säger är att sjukdomen i fråga inte är ett Hodgkin lymfom. NHL begreppet lever ändå kvar och används även i denna avhandling. NHL består av ett 40-tal olika subtyper, som skiljer sig väsentligt åt i kliniskt förlopp och hur de svarar på våra behandlingar. De har gemensamt att de är elakartade sjukdomar utgångna från lymfocyter, dvs. en sorts vita blodkroppar. Oftast handlar det om B-celler (90%) men ibland T/NK-celler (10%).

Även om NHL som grupp ligger på topp 10-listan över de vanligaste cancersjukdomarna i Sverige (och västvärlden för övrigt) är varje subtyp förhållandevis ovanlig. Pga. detta har många forskningsstudier gjorts på NHL som grupp. Ju mer kunskapen ökat har skillnaderna mellan subtyperna blivit tydligare. Det är nu uppenbart att vissa risk- och prognosfaktorer är gemensamma för NHL som grupp men att många är specifika för enskilda subtyper. I denna avhandling ingår studier både på NHL som grupp och enskilda subtyper.

För de flesta av våra NHL-patienter vet vi inte orsaken till att sjukdomen uppstod. Det är välkänt att ett hämmat immunförsvar, t.ex. pga. HIV eller efter organtransplantation, klart ökar risken för NHL. Vissa infektioner är kopplade till NHL-utveckling, såsom Epstein-Barr virus och Helicobacter pylori. Huruvida livsstilsfaktorer spelar roll för risken att utveckla NHL är inte helt klart, även om flera studier pekar på det. Det är också mycket som talar för att normal variation i vårt DNA spelar roll för risken att insjukna i NHL. Vissa genetiska varianter minskar risken att insjukna medan andra ökar den. Den genetiska variation vi undersökte var av typen där ett baspar i DNA-strängen bytts ut mot ett annat (engelska: single nucleotide polymorphism, SNP). Ofta spelar detta byte ingen roll men i enstaka fall kan det till exempel leda till att en aminosyra i ett protein byts ut mot en annan, vilket skulle kunna påverka proteinets funktion.

För en vanlig NHL-subtyp som heter follikulärt lymfom finns ett antal studier som talar för att rökning ökar risken att insjukna men alla studier är inte helt överens. Det finns också belägg för att SNP:ar i en gen som kodar för molekylen HLA-DRB1 påverkar aminosyrasekvensen i proteinet, vilket har kopplats till risken att insjukna i follikulärt lymfom. I Studie I i denna avhandling undersökte vi detta närmare. Vår hypotes var att rökning ökar risken för follikulärt lymfom särskilt hos personer som har vissa aminosyrakombinationer i HLA-DRB1- molekylen jämfört med personer utan dessa

varianter. I studien kunde vi påvisa ett sådant samband: jämfört med personer som aldrig rökt och inte bar på en variant av HLA-DRB1 kallad shared epitope ökade risken att insjukna gradvis så att personer som var före detta rökare och bar på shared epitope hade 2,2 gånger högre risk medan personer som var rökare och bar på shared epitope hade 3,5 gånger högre risk att insjukna. Detta var tidigare inte undersökt för NHL och behöver verifieras i andra och gärna större studier. Om fyndet står sig kan follikulärt lymfom läggas till listan över sjukdomar där risken ökar pga. rökning. Det kan också lära oss något nytt om mekanismerna som leder till att follikulärt lymfom uppstår.

Vad gäller prognosen för NHL så är den helt beroende av vilken subtyp det handlar om. Högmaligna lymfom, såsom diffust storcelligt B-cellslymfom, leder till döden inom några månader utan behandling men är oftast botbara med medicinsk behandling. Lågmaligna lymfom, såsom follikulärt lymfom, är vanligen inte botbara men växer långsamt. De behandlas vid symtom och hälften av patienterna lever fortfarande 10 år efter diagnos. Prognosen varierar dock påtagligt även inom varje subtyp av NHL, vilket vi inte har kunskap att förstå ännu. För de vanligaste NHL-subtyperna finns prognostiska index, där summan av ett antal riskfaktorer placerar en patient i en prognosgrupp. Patientens ålder och prestationsförmåga (engelska: performance status) ingår nästan alltid. Prognosgrupperna som finns kan dock endast grovt förutsäga förloppet. Därför har studier gjorts av andra värdfaktorer sista åren med hopp om att hitta nya markörer med prognostiskt värde. I detta område ingår även studier om huruvida nedärvd genetisk variation i form av SNP:ar kan förklara skillnaden i överlevnad inom olika NHL subtyper. Betydelsen av värdfaktorer, inklusive nedärvd genetisk variation, för överlevnaden vid lymfom var fokus för Studie II-IV i denna avhandling.

I Studie II undersökte vi om det fanns ett samband mellan överlevnad och rökning, socioekonomiskt status (i vårt fall utbildningsnivå), fetma, exponering för ultraviolett ljus, autoimmun sjukdom eller om man hade en familjemedlem som drabbats av en blodcancer (inklusive lymfom). Vi kunde visa att NHL patienter som rökte, hade låg utbildningsnivå eller var drabbad av en autoimmun sjukdom dog i större utsträckning än NHL patienter utan dessa faktorer. För rökning och lågt socioekonomisk status finns stöd för våra resultat i andra studier. Vi vet inte säkert varför det är såhär men en hypotes med visst stöd är att både rökning och lågt socioekonomisk status är förknippade med ökad sjuklighet (såsom hjärtkärlsjukdom) som kan påverka våra möjligheter att behandla lymfomsjukdomar.

I Studie III och IV undersökte vi om nedärvd genetisk variation påverkade överlevnaden för patienter med follikulärt lymfom (Studie III) eller en lymfoid malignitet (Studie IV). I Studie III undersökte vi dels om någon av cirka 300 000 SNP:ar spridda över DNA-strängarna påverkade risken att dö av lymfomrelaterade orsaker (del 1), dels om 22 SNP:ar som rapporterats påverka prognosen för patienter med follikulärt lymfom i tidigare studier också gjorde det i vårt patientmaterial (del 2). I del 1 kunde vi inte med säkerhet påvisa någon SNP som påverkade överlevnaden, även om effekten av en SNP i en gen med eventuell koppling till läkemedelsresistens stod ut och vara nära att vara statistiskt säkerställd. I del 2 fann vi i likhet med tidigare studier att två SNP:ar i genen för interleukin 8 (som påverkar funktioner i vårt immunförsvar) och en SNP i genen CD46 (som påverkar komplementsystemet i vårt immunförsvar) påverkade prognosen för patienter med follikulärt lymfom. Dessa fynd behöver stöd i ytterligare studier innan man kan dra säkra slutsatser. Om fyndet kan bekräftas innebär det dels en ökad förståelse för de mekanismer som påverkar sjukdomsförloppet och möjligen även nya mål att rikta behandling mot.

I Studie IV sökte vi stöd för att nedärvd genetisk variation påverkar överlevnaden vid lymfoida maligniteter (= Hodgkinlymfom och non-Hodgkinlymfom, inklusive sjukdomar som myelom och akut lymfatisk leukemi). Vår hypotes var att om det finns en nedärvd komponent som påverkar överlevnaden vid lymfoida maligniteter så borde två familjemedlemmar med en lymfoid malignitet ha liknande överlevnadstid. För studien använde vi information från svenska register, som gjorde det möjligt för oss att koppla ihop individer inom en familj (två eller fler) som drabbats av en lymfoid malignitet. I studien kunde vi visa att inom familjer där två eller fler medlemmar hade samma typ av lymfoid malignitet fanns en samstämmighet i överlevnadstid. Med andra ord, personer vars familjemedlem hade en god prognos hade själva en god prognos, medan personer vars familjemedlem hade en dålig prognos själva hade en dålig prognos. Sambandet kunde också påvisas mer specifikt hos personer med lågmalignt NHL. Denna studie motiverar att vi fortsätter studera betydelsen av nedärvd genetisk variation för överlevnaden vid lymfoida maligniteter.

9 ACKNOWLEDGEMENTS

Thank you very much to **all that have contributed** to this thesis by providing your knowledge and support. I could not have done this alone.

I feel especially grateful to:

Karin Ekström Smedby, main supervisor. Thank you for making me part of your research group and for being all one would hope for in a supervisor and more. If it wasn't for you, I would probably not be working with lymphoma now – a horrible thought! Thank you also for suggesting the four interesting projects of my thesis, for always being available to help and support me, for letting me take responsibility and for always being kind and patient. You have done and are still doing a great job in building the Cancer Epidemiology group.

Johan Askling, co-supervisor and also the head of the Clinical Epidemiology Unit. Thank you for supporting and encouraging me during my doctoral studies. Thank you for the clever questions and remarks when I have introduced you to new projects or preliminary results. I have been fortunate to have had you as a co-supervisor. Congratulations for making the Clinical Epidemiology Unit successful and a joyful place to work at.

Richard Rosenquist, co-supervisor. Thank you for supporting me during my doctoral studies. Thank you for engaging in the projects and for encouraging me. Your knowledge and experience in genetics and research have contributed to my learning and the quality of the projects. I feel fortunate to have had you as a co-supervisor.

Sandra Eloranta, statistician. Thank you for always being kind, patient and pedagogic when explaining statistical considerations to me. I am impressed.

Sara Ekberg, statistician. Thank you for your important contribution to the study on familial concordance in survival time! Also, thank you for always being kind and helpful.

Jia-Nee Foo, geneticist at Genome Institute of Singapore. You have helped me immensely in the SNP study and interaction study. Always patient and kind, you have explained tricky concepts of genetic epidemiology and helped me when I got stuck in some analysis.

The co-authors of the four studies of the thesis. *Hans-Olov Adami* and *Bengt Glimelius* for lending your long experience of cancer epidemiology. *Julia Simard* and *Ellen Chang* for your patience and for teaching me basic research concepts when this was a new field to me. Our Danish colleagues *Henrik Hjalgrim* and *Mads Melbye* for great collaboration. Our American colleagues *Paige Bracci*, *Lucia Conde* and *Christine Skibola* for sharing data with us, contributing with knowledge as well as with hands on work with the GWAS. *Jianjun Liu* at the Genome Institute of Singapore for your expert knowledge in genetics and genetic

epidemiology. *Hatef Darabi*, *Keith Humphreys* and *Robert Karlsson* at MEB for your important contribution to the design and performance of the SNP study, as well as for your patience and sharing of knowledge of genetic epidemiology. *Leonid Padyukov* at the Rheumatology Unit for contributing greatly with your experience of interaction studies and knowledge about the HLA region. *Eva Baecklund* for contributing with your knowledge about lymphoma in the context of autoimmune diseases. Needless to say, *Karin Ekström Smedby*, *Johan Askling*, *Richard Rosenquist*, *Sandra Eloranta*, *Sara Ekberg* and *Jia-Nee Foo* contributed greatly to the studies.

The Cancer Epidemiology group, *Karin Ekström Smedby*, *Lingjing Chen*, *Ingrid Glimelius*, *Elsa Brånvall*, *Henrik Benoni*, *Gabriella Frisk*, *Andreas Pettersson*, *Sara Harrysson*, *Åsa Hallqvist Everhov*, *Daniela Gordon*, *Renata Zelic*, *Sara Ekberg* and *Sandra Eloranta*. Thank you for the nice time together so far! You all contribute to the kind, stimulating, fun and allowing atmosphere at our meetings or when we go bowling.

All people working at the Clinical Epidemiology Unit for creating such a nice place to work in. It always feels welcoming in the lunch room. There are no barriers between research groups within the Unit, all are ready to help out when you are in need for assistance with some tricky (or basic) problem. I am happy and proud to be part of the Unit. A special thank you to *Michael Fored*, director of research education as well as a part of the Unit, for your support during the doctoral studies. Also a special thank you to *Anders Ekbom*, for establishing the successful and nice research environment of the Unit.

Lars Klareskog, *Lars Alfredsson* and *Henrik Källeberg* and Lena *Nise* from the **EIRA study group** for sharing data with us.

10 REFERENCES

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375-2390.

2. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Available at http://globocan.iarc.fr/Pages/fact_sheets_population.aspx. : International Agency for Research on Cancer, World Health Organization; 2012:Accessed 27 October 2016.

3. Swerdlow SH, Campo E, Harris NL, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition. Vol. 2. Lyon, France: IARC Press; 2008.

4. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The Molecular Biology of the Cell (ed 6th). New York: Garland Science; 2014.

5. Boffetta P. Epidemiology of adult non-Hodgkin lymphoma. *Annals of Oncology*. 2011;22(Issue suppl 4):iv27-iv31.

6. Cerhan JR, Kricker A, Paltiel O, 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(48):15-25.

7. Linet MS, Vajdic CM, Morton LM, 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(48):26-40.

8. Morton LM, Slager SL, Cerhan JR, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr*. 2014;2014(48):130-144.

9. Smedby KE, Sampson JN, Turner JJ, 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*. 2014;2014(48):76-86.

10. Slager SL, Benavente Y, Blair A, 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(48):41-51.

11. Schollkopf C, Smedby KE, Hjalgrim H, et al. Cigarette smoking and risk of non-Hodgkin's lymphoma--a population-based case-control study. *Cancer Epidemiol Biomarkers Prev*. 2005;14(7):1791-1796.

12. Arnson Y, Shoenfeld Y, Amital H. Effects of tobacco smoke on immunity, inflammation and autoimmunity. *J Autoimmun*. 2010;34(3):J258-265.

13. Chang ET, Hjalgrim H, Smedby KE, et al. Body mass index and risk of malignant lymphoma in Scandinavian men and women. *J Natl Cancer Inst*. 2005;97(3):210-218.

14. Chiu BC, Hou N. Epidemiology and Etiology of Non-Hodgkin Lymphoma. Non-Hodgkin Lymphoma - Pathology, Imagin, and Current Therapy. Switzerland: Springer International Publising AG; 2015.

15. Gregor MF, Hotamisligil GS. Inflammatory mechanisms in obesity. *Annu Rev Immunol*. 2011;29:415-445.

16. Hu S, Federman DG, Ma F, Kirsner RS. Skin cancer and non-Hodgkin's lymphoma: examining the link. *Dermatol Surg*. 2005;31(1):76-82.

17. Smedby KE, Hjalgrim H, Melbye M, et al. Ultraviolet radiation exposure and risk of malignant lymphomas. *J Natl Cancer Inst*. 2005;97(3):199-209.

18. Hart PH, Gorman S, Finlay-Jones JJ. Modulation of the immune system by UV radiation: more than just the effects of vitamin D? *Nat Rev Immunol*. 2011;11(9):584-596.

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

20. Foo JN, Smedby KE, Akers NK, et al. Coding variants at hexa-allelic amino acid 13 of HLA-DRB1 explain independent SNP associations with follicular lymphoma risk. *Am J Hum Genet*. 2013;93(1):167-172.

21. Skibola CF, Berndt SI, Vijai J, et al. Genome-wide Association Study Identifies Five Susceptibility Loci for Follicular Lymphoma outside the HLA Region. *Am J Hum Genet*. 2014;95(4):462-471.

22. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum*. 2004;50(10):3085-3092.

23. Pedersen M, Jacobsen S, Garred P, et al. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. *Arthritis Rheum*. 2007;56(5):1446-1453.

24. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612.

25. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242.

26. Basso K, Dalla-Favera R. Germinal centres and B cell lymphomagenesis. *Nat Rev Immunol*. 2015;15(3):172-184.

27. Leich E, Haralambieva E, Zettl A, et al. Tissue microarray-based screening for chromosomal breakpoints affecting the T-cell receptor gene loci in mature T-cell lymphomas. *J Pathol*. 2007;213(1):99-105.

28. Rossi D, Ciardullo C, Gaidano G. Genetic aberrations of signaling pathways in lymphomagenesis: revelations from next generation sequencing studies. *Semin Cancer Biol*. 2013;23(6):422-430.

29. Kuppers R. Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer*. 2005;5(4):251-262.

30. Krangel MS. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol*. 2009;21(2):133-139.

31. Dyer MJ, Akasaka T, Capasso M, et al. Immunoglobulin heavy chain locus chromosomal translocations in B-cell precursor acute lymphoblastic leukemia: rare clinical curios or potent genetic drivers? *Blood*. 2010;115(8):1490-1499.

32. Blonska M, Agarwal NK, Vega F. Shaping of the tumor microenvironment: Stromal cells and vessels. *Semin Cancer Biol*. 2015;34:3-13.

33. Taylor JG, Gribben JG. Microenvironment abnormalities and lymphomagenesis: Immunological aspects. *Semin Cancer Biol*. 2015;34:36-45.

34. Carbone A, Younes A. Lymphomas and microenvironment: Impact on lymphomagenesis. Foreword. *Semin Cancer Biol*. 2015;34:1-2.

35. Niemann CU, Wiestner A. B-cell receptor signaling as a driver of lymphoma development and evolution. *Semin Cancer Biol*. 2013;23(6):410-421.

36. Sutton LA, Agathangelidis A, Belessi C, et al. Antigen selection in B-cell lymphomas--tracing the evidence. *Semin Cancer Biol*. 2013;23(6):399-409.

37. Baecklund E, Smedby KE, Sutton LA, Askling J, Rosenquist R. Lymphoma development in patients with autoimmune and inflammatory disorders--what are the driving forces? *Semin Cancer Biol*. 2014;24:61-70.

38. Scott DW, Gascoyne RD. The tumour microenvironment in B cell lymphomas. *Nat Rev Cancer*. 2014;14(8):517-534.

39. Jaffe ES, Nicolae A, Pittaluga S. Peripheral T-cell and NK-cell lymphomas in the WHO classification: pearls and pitfalls. *Mod Pathol*. 2013;26 Suppl 1:S71-87.

40. Vockerodt M, Yap LF, Shannon-Lowe C, et al. The Epstein-Barr virus and the pathogenesis of lymphoma. *J Pathol*. 2015;235(2):312-322.

41. Carbone A, Gloghini A. Relationships between lymphomas linked to hepatitis C virus infection and their microenvironment. *World J Gastroenterol*. 2013;19(44):7874-7879.

42. Mortaz E, Tabarsi P, Mansouri D, et al. Cancers Related to Immunodeficiencies: Update and Perspectives. *Front Immunol*. 2016;7:365.

43. Dolcetti R, Gloghini A, Caruso A, Carbone A. A lymphomagenic role for HIV beyond immune suppression? *Blood*. 2016;127(11):1403-1409.

44. Al-Mansour Z, Nelson BP, Evens AM. Post-transplant lymphoproliferative disease (PTLD): risk factors, diagnosis, and current treatment strategies. *Curr Hematol Malig Rep*. 2013;8(3):173-183.

45. Singavi AK, Harrington AM, Fenske TS. Post-transplant lymphoproliferative disorders. *Cancer Treat Res*. 2015;165:305-327.

46. Clarke CA, Morton LM, Lynch C, et al. Risk of lymphoma subtypes after solid organ transplantation in the United States. *Br J Cancer*. 2013;109(1):280-288.

47. Gibson TM, Morton LM, Shiels MS, Clarke CA, Engels EA. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. *AIDS*. 2014;28(15):2313-2318.

48. George EV, Pharm J, Houston C, et al. Breast implant-associated ALK-negative anaplastic large cell lymphoma: a case report and discussion of possible pathogenesis. *Int J Clin Exp Pathol*. 2013;6(8):1631-1642.

49. de Jong D, Vasmel WL, de Boer JP, et al. Anaplastic large-cell lymphoma in women with breast implants. *JAMA*. 2008;300(17):2030-2035.

50. Ferreri AJ, Govi S, Pileri SA, Savage KJ. Anaplastic large cell lymphoma, ALK-negative. *Crit Rev Oncol Hematol*. 2013;85(2):206-215.

51. Wolfram D, Rabensteiner E, Grundtman C, et al. T regulatory cells and TH17 cells in peri-silicone implant capsular fibrosis. *Plast Reconstr Surg*. 2012;129(2):327e-337e.

52. Jarjis RD, Matzen SH. [Breast implant-associated anaplastic large-cell lymphoma]. *Ugeskr Laeger*. 2015;177(48):V07150625.

53. Jaffe ES, Harris NL, Stein H, Isaacson PG. Classification of lymphoid neoplasms: the microscope as a tool for disease discovery. *Blood*. 2008;112(12):4384-4399.

54. Shankland KR, Armitage JO, Hancock BW. Non-Hodgkin lymphoma. *Lancet*. 2012;380(9844):848-857.

55. Svenska Lymfomregistret - rapport för år 2000-2005. Published at http://www.swedishlymphoma.com/rapporter. Accessed 26 October 2016.

56. Solal-Celigny P, Roy P, Colombat P, et al. Follicular lymphoma international prognostic index. *Blood*. 2004;104(5):1258-1265.

57. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med*. 1993;329(14):987-994.

58. Binet JL, Auquier A, Dighiero G, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. 1981;48(1):198-206.

59. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975;46(2):219-234.

60. Coiffier B, Lepage E, Briere J, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(4):235-242.

61. Pfreundschuh M, Trumper L, Osterborg A, et al. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. *Lancet Oncol*. 2006;7(5):379-391.

62. Rummel MJ, Niederle N, Maschmeyer G, et al. Bendamustine plus rituximab versus CHOP plus rituximab as first-line treatment for patients with indolent and mantle-cell lymphomas: an open-label, multicentre, randomised, phase 3 non-inferiority trial. *Lancet*. 2013;381(9873):1203-1210.

63. Hallek M, Fischer K, Fingerle-Rowson G, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet*. 2010;376(9747):1164-1174.

64. Farooqui MZ, Valdez J, Martyr S, et al. Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: a phase 2, single-arm trial. *Lancet Oncol*. 2015;16(2):169-176.

65. Zelenetz AD. Emerging treatment options for B-cell lymphomas. *J Natl Compr Canc Netw*. 2015;13(5 Suppl):666-669.

66. Ziepert M, Hasenclever D, Kuhnt E, et al. Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol*. 2010;28(14):2373-2380.

67. Gutierrez-Garcia G, Cardesa-Salzmann T, Climent F, et al. Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Blood*. 2011;117(18):4836-4843.

68. Savage KJ, Johnson NA, Ben-Neriah S, et al. MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood*. 2009;114(17):3533-3537.

69. Aukema SM, Siebert R, Schuuring E, et al. Double-hit B-cell lymphomas. *Blood*. 2011;117(8):2319-2331.

70. Nooka AK, Nabhan C, Zhou X, et al. Examination of the follicular lymphoma international prognostic index (FLIPI) in the National LymphoCare study (NLCS): a prospective US patient cohort treated predominantly in community practices. *Ann Oncol*. 2013;24(2):441-448.

71. Ott G, Katzenberger T, Lohr A, et al. Cytomorphologic, immunohistochemical, and cytogenetic profiles of follicular lymphoma: 2 types of follicular lymphoma grade 3. *Blood*. 2002;99(10):3806-3812.

72. Montserrat E, Sanchez-Bisono J, Vinolas N, Rozman C. Lymphocyte doubling time in chronic lymphocytic leukaemia: analysis of its prognostic significance. *Br J Haematol*. 1986;62(3):567-575.

73. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000;343(26):1910-1916.

74. Zenz T, Eichhorst B, Busch R, et al. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol*. 2010;28(29):4473-4479.

75. Stilgenbauer S, Schnaiter A, Paschka P, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood*. 2014;123(21):3247-3254.

76. Winqvist M, Asklid A, Andersson PO, et al. Real--world results of ibrutinib in patients with relapsed or refractory chronic lymphocytic leukemia: Data from 95 consecutive patients treated in a compassionate use program. *Haematologica*. 2016.

77. International CLLIPIwg. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol*. 2016;17(6):779-790.

78. Hoster E, Dreyling M, Klapper W, et al. A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Blood*. 2008;111(2):558-565.

79. Hoster E, Klapper W, Hermine O, et al. Confirmation of the mantle-cell lymphoma International Prognostic Index in randomized trials of the European Mantle-Cell Lymphoma Network. *J Clin Oncol*. 2014;32(13):1338-1346.

80. Hoster E, Rosenwald A, Berger F, et al. Prognostic Value of Ki-67 Index, Cytology, and Growth Pattern in Mantle-Cell Lymphoma: Results From Randomized Trials of the European Mantle Cell Lymphoma Network. *J Clin Oncol*. 2016;34(12):1386-1394.

81. Dave SS, Wright G, Tan B, et al. Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med*. 2004;351(21):2159-2169.

82. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med*. 2008;359(22):2313-2323.

83. Wahlin BE, Sander B, Christensson B, Kimby E. CD8+ T-cell content in diagnostic lymph nodes measured by flow cytometry is a predictor of survival in follicular lymphoma. *Clin Cancer Res*. 2007;13(2 Pt 1):388-397.

84. Carreras J, Lopez-Guillermo A, Fox BC, et al. High numbers of tumor-infiltrating FOXP3-positive regulatory T cells are associated with improved overall survival in follicular lymphoma. *Blood*. 2006;108(9):2957-2964.

85. Talamini R, Polesel J, Spina M, et al. The impact of tobacco smoking and alcohol drinking on survival of patients with non-Hodgkin lymphoma. *Int J Cancer*. 2008;122(7):1624-1629.

86. Geyer SM, Morton LM, Habermann TM, et al. Smoking, alcohol use, obesity, and overall survival from non-Hodgkin lymphoma: a population-based study. *Cancer*. 2010;116(12):2993-3000.

87. Ollberding NJ, Evens AM, Aschebrook-Kilfoy B, et al. Pre-diagnosis cigarette smoking and overall survival in non-Hodgkin lymphoma. *Br J Haematol*. 2013;163(3):352-356.

88. Battaglioli T, Gorini G, Costantini AS, et al. Cigarette smoking and alcohol consumption as determinants of survival in non-Hodgkin's lymphoma: a population-based study. *Ann Oncol*. 2006;17(8):1283-1289.

89. Leo QJ, Ollberding NJ, Wilkens LR, et al. Obesity and non-Hodgkin lymphoma survival in an ethnically diverse population: the Multiethnic Cohort study. *Cancer Causes Control*. 2014;25(11):1449-1459.

90. Chiu BC, Dave BJ, Blair A, et al. Cigarette smoking, familial hematopoietic cancer, hair dye use, and risk of t(14;18)-defined subtypes of non-Hodgkin's lymphoma. *Am J Epidemiol*. 2007;165(6):652-659.

91. Bell DA, Liu Y, Cortopassi GA. Occurrence of bcl-2 oncogene translocation with increased frequency in the peripheral blood of heavy smokers. *J Natl Cancer Inst*. 1995;87(3):223-224.

92. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-century hazards of smoking and benefits of cessation in the United States. *N Engl J Med*. 2013;368(4):341-350.

93. Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368(4):351-364.

94. Terret C, Albrand G, Rainfray M, Soubeyran P. Impact of comorbidities on the treatment of non-Hodgkin's lymphoma: a systematic review. *Expert Rev Hematol*. 2015;8(3):329-341.

95. Lee L, Cheung WY, Atkinson E, Krzyzanowska MK. Impact of comorbidity on chemotherapy use and outcomes in solid tumors: a systematic review. *J Clin Oncol*. 2011;29(1):106-117.

96. Bray C, Morrison DS, McKay P. Socio-economic deprivation and survival of non-Hodgkin lymphoma in Scotland. *Leuk Lymphoma*. 2008;49(5):917-923.

97. Frederiksen BL, Dalton SO, Osler M, Steding-Jessen M, de Nully Brown P. Socioeconomic position, treatment, and survival of non-Hodgkin lymphoma in Denmark - a nationwide study. *Br J Cancer*. 2012;106(5):988-995.

98. Frederiksen BL, Brown Pde N, Dalton SO, Steding-Jessen M, Osler M. Socioeconomic inequalities in prognostic markers of non-Hodgkin lymphoma: analysis of a national clinical database. *Eur J Cancer*. 2011;47(6):910-917.

99. Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. 2008;358(23):2468-2481.

100. Han X, Stevens J, Bradshaw PT. Body mass index, weight change, and survival in non-Hodgkin lymphoma patients in Connecticut women. *Nutr Cancer*. 2013;65(1):43-50.

101. Goodwin PJ, Chlebowski RT. Obesity and Cancer: Insights for Clinicians. *J Clin Oncol*. 2016;34(35):4197-4202.

102. Shlomai G, Neel B, LeRoith D, Gallagher EJ. Type 2 Diabetes Mellitus and Cancer: The Role of Pharmacotherapy. *J Clin Oncol*. 2016;34(35):4261-4269.

103. Uddin S, Mohammad RM. Role of leptin and leptin receptors in hematological malignancies. *Leuk Lymphoma*. 2016;57(1):10-16.

104. Drake MT, Maurer MJ, Link BK, et al. Vitamin D insufficiency and prognosis in non-Hodgkin's lymphoma. *J Clin Oncol*. 2010;28(27):4191-4198.

105. Kelly JL, Salles G, Goldman B, et al. Low Serum Vitamin D Levels Are Associated With Inferior Survival in Follicular Lymphoma: A Prospective Evaluation in SWOG and LYSA Studies. *J Clin Oncol*. 2015;33(13):1482-1490.

106. Feldman D, Krishnan AV, Swami S, Giovannucci E, Feldman BJ. The role of vitamin D in reducing cancer risk and progression. *Nat Rev Cancer*. 2014;14(5):342-357.

107. Ji J, Liu X, Sundquist K, Sundquist J. Survival of cancer in patients with rheumatoid arthritis: a follow-up study in Sweden of patients hospitalized with rheumatoid arthritis 1 year before diagnosis of cancer. *Rheumatology (Oxford)*. 2011;50(8):1513-1518.

108. Mikuls TR, Endo JO, Puumala SE, et al. Prospective study of survival outcomes in Non-Hodgkin's lymphoma patients with rheumatoid arthritis. *J Clin Oncol*. 2006;24(10):1597-1602.

109. Ludvigsson JF, Lebwohl B, Rubio-Tapia A, et al. Does celiac disease influence survival in lymphoproliferative malignancy? *Eur J Epidemiol*. 2013;28(6):475-483.

110. Baecklund E, Backlin C, Iliadou A, et al. Characteristics of diffuse large B cell lymphomas in rheumatoid arthritis. *Arthritis Rheum*. 2006;54(12):3774-3781.

111. Kaplan MJ. Cardiovascular complications of rheumatoid arthritis: assessment, prevention, and treatment. *Rheum Dis Clin North Am*. 2010;36(2):405-426.

112. Anderson LA, Pfeiffer RM, Rapkin JS, et al. Survival patterns among lymphoma patients with a family history of lymphoma. *J Clin Oncol*. 2008;26(30):4958-4965.

113. Ji J, Forsti A, Sundquist J, Lenner P, Hemminki K. Survival in non-Hodgkin's lymphoma by histology and family history. *J Cancer Res Clin Oncol*. 2009;135(12):1711-1716.

114. Aschebrook-Kilfoy B, Zheng T, Foss F, et al. Polymorphisms in immune function genes and non-Hodgkin lymphoma survival. *J Cancer Surviv*. 2012;6(1):102-114.

115. Charbonneau B, Maurer MJ, Fredericksen ZS, et al. Germline variation in complement genes and event-free survival in follicular and diffuse large B-cell lymphoma. *Am J Hematol*. 2012;87(9):880-885.

116. Racila E, Link BK, Weng WK, et al. A polymorphism in the complement component C1qA correlates with prolonged response following rituximab therapy of follicular lymphoma. *Clin Cancer Res*. 2008;14(20):6697-6703.

117. Weng WK, Czerwinski D, Timmerman J, Hsu FJ, Levy R. Clinical outcome of lymphoma patients after idiotype vaccination is correlated with humoral immune response and immunoglobulin G Fc receptor genotype. *J Clin Oncol*. 2004;22(23):4717-4724.

118. Weng WK, Levy R. Two immunoglobulin G fragment C receptor polymorphisms independently predict response to rituximab in patients with follicular lymphoma. *J Clin Oncol*. 2003;21(21):3940-3947.

119. Cerhan JR, Wang S, Maurer MJ, et al. Prognostic significance of host immune gene polymorphisms in follicular lymphoma survival. *Blood*. 2007;109(12):5439-5446.

120. Gibson TM, Wang SS, Cerhan JR, et al. Inherited genetic variation and overall survival following follicular lymphoma. *Am J Hematol*. 2012;87(7):724-726.

121. Wang SS, Maurer MJ, Morton LM, et al. Polymorphisms in DNA repair and one-carbon metabolism genes and overall survival in diffuse large B-cell lymphoma and follicular lymphoma. *Leukemia*. 2009;23(3):596-602.

122. Persky DO, Dornan D, Goldman BH, et al. Fc gamma receptor 3a genotype predicts overall survival in follicular lymphoma patients treated on SWOG trials with combined monoclonal antibody plus chemotherapy but not chemotherapy alone. *Haematologica*. 2012;97(6):937-942.

123. Berglund M, Enblad G, Thunberg U. SNP rs6457327 is a predictor for overall survival in follicular lymphoma as well as survival after transformation. *Blood*. 2011;118(16):4489.

124. Wrench D, Leighton P, Skibola CF, et al. SNP rs6457327 in the HLA region on chromosome 6p is predictive of the transformation of follicular lymphoma. *Blood*. 2011;117(11):3147-3150.

125. Cartron G, Dacheux L, Salles G, et al. Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor FcgammaRIIIa gene. *Blood*. 2002;99(3):754-758.

126. Jaffe E. Pathology and genetics of tumours of hematopoietic and lymphoid tissues. Lyon, France: IARC Press, International Agency for Research on Cancer; 2001.

127. Fernberg P, Chang ET, Duvefelt K, et al. Genetic variation in chromosomal translocation breakpoint and immune function genes and risk of non-Hodgkin lymphoma. *Cancer Causes Control*. 2010;21(5):759-769.

128. Smedby KE, Foo JN, Skibola CF, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet*. 2011;7(4):e1001378.

129. Baecklund F, Foo JN, Bracci P, et al. A comprehensive evaluation of the role of genetic variation in follicular lymphoma survival. *BMC Med Genet*. 2014;15(1):113.

130. d'Amore F, Christensen BE, Brincker H, et al. Clinicopathological features and prognostic factors in extranodal non-Hodgkin lymphomas. *Eur J Cancer*. 1991;27(10):1201-1208.

131. Stolt P, Bengtsson C, Nordmark B, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis*. 2003;62(9):835-841.

132. Skibola CF, Bracci PM, Halperin E, et al. Polymorphisms in the estrogen receptor 1 and vitamin C and matrix metalloproteinase gene families are associated with susceptibility to lymphoma. *PLoS One*. 2008;3(7):e2816.

133. Cancer Incidence in Sweden 2011: Sweden National Board of Health and Welfare; 2012.

134. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncol*. 2009;48(1):27-33.

135. Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978. *Acta Radiol Oncol*. 1984;23(5):305-313.

136. Dödsorsaksstatistik - Historik, produktionsmetoder och tillförlitlighet. In: Health SNBo ed. Available at: http://www.socialstyrelsen.se/publikationer2010/2010-4-33; 2010.

137. Johansson LA. Targeting Non-obvious Errors in Death Certificates. Department of Public Health and Caring Sciences. Vol. PhD: Uppsala University; 2008.

138. Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol*. 2000;29(3):495-502.

139. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol*. 2016;31(2):125-136.

140. Multi-generation register 2012 . A description of contents and quality: Statistics Sweden, Background Facts, Population and Welfare Statistics 2013:2.

141. Microdata at Statistics Sweden for research purposes. Örebro: Statistics Sweden Marketing Unit; 2013.

142. Background Facts 2016:1, Integrated database for labour market research (LISA) 1990-2016. Örebro: Statistics Sweden, Population and Welfare Department 2016.

143. Karlsson M. Beskrivning av statistiken - Befolkningens utbildning 2013 (UF0506). Örebro: Statistics Sweden; 2014.

144. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*. 1958(53):457-481.

145. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*. 1972(34):187-220.

146. Fine JGR. A proportional hazards model for the subdistribution of a competing risk. . *Journal of the American Statistical Association*. 1999(94):496-509.

147. Knol MJ, Egger M, Scott P, Geerlings MI, Vandenbroucke JP. When one depends on the other: reporting of interaction in case-control and cohort studies. *Epidemiology*. 2009;20(2):161-166.

148. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. *Epidemiologic Methods*. 2014;Volume 3(3(1)):33-72.

149. Rothman K, Lash T, Greenland S. Modern Epidemiology (ed 3rd Edition). Philadelpiha, PA: Lippincott, Williams & Wilkins; 2008.

150. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol*. 1980;112(4):467-470.

151. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology*. 1992;3(5):452-456.

152. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.

153. Handbook for Systematic Reviews of Interventions. http://handbook.cochrane.org/front_page.htm: The Cochrane Collaboration, edited by Julian Higgins and Sally Green; 2011.

154. Svenska Lymfomregistret: Rapport för år 2000-2006. http://www.swedishlymphoma.com/rapporter: Svenska Lymfomgruppen; 2009.

155. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.

156. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2010:Retrieved from http://www.R-project.org.

157. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.

158. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.

159. Simard JF, Baecklund F, Chang ET, et al. Lifestyle factors, autoimmune disease and family history in prognosis of non-hodgkin lymphoma overall and subtypes. *Int J Cancer*. 2013;132(11):2659-2666.

160. Mackenbach JP, Kunst AE, Cavelaars AE, Groenhof F, Geurts JJ. Socioeconomic inequalities in morbidity and mortality in western Europe. The EU Working Group on Socioeconomic Inequalities in Health. *Lancet*. 1997;349(9066):1655-1659.

161. Myasoedova E, Davis JM, 3rd, Crowson CS, Gabriel SE. Epidemiology of rheumatoid arthritis: rheumatoid arthritis and mortality. *Curr Rheumatol Rep*. 2010;12(5):379-385.

162. Steingrimsson V, Lund SH, Turesson I, et al. Population-based study on the impact of the familial form of Waldenstrom macroglobulinemia on overall survival. *Blood*. 2015;125(13):2174-2175.

163. Mauro FR, Giammartini E, Gentile M, et al. Clinical features and outcome of familial chronic lymphocytic leukemia. *Haematologica*. 2006;91(8):1117-1120.

164. Aung T, Chapuy B, Vogel D, et al. Exosomal evasion of humoral immunotherapy in aggressive B-cell lymphoma modulated by ATP-binding cassette transporter A3. *Proc Natl Acad Sci U S A*. 2011;108(37):15336-15341.

165. Coleman JA, Quazi F, Molday RS. Mammalian P4-ATPases and ABC transporters and their role in phospholipid transport. *Biochim Biophys Acta*. 2013;1831(3):555-574.

166. Cerhan JR, Ansell SM, Fredericksen ZS, et al. Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma. *Blood*. 2007;110(13):4455-4463.

167. Berndt SI, Morton LM, Wang SS, et al. Genetic Heritability of Common Non-Hodgkin Lymphoma Subtypes. 2014 American Society of Hematology Annual Meeting. San Francisco, CA, USA, Abstract available at: http://www.ashg.org/2014meeting/abstracts/fulltext/f140122684.htm.

168. Mitchell JS, Johnson DC, Litchfield K, et al. Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. *Sci Rep*. 2015;5:12473.

169. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. *Autoimmun Rev*. 2005;4(3):130-136.

170. Junlen HR, Peterson S, Kimby E, et al. Follicular lymphoma in Sweden: nationwide improved survival in the rituximab era, particularly in elderly women: a Swedish Lymphoma Registry study. *Leukemia*. 2015;29(3):668-676.