Title
B-Cell Activation and Non-Hodgkin Lymphoma Risk in an HIV Positive Population

Author
Chang, Po-Yin

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

B-Cell Activation and Non-Hodgkin Lymphoma Risk in an HIV Positive Population

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Po-Yin Chang

2013

ABSTRACT OF THE DISSERTATION

B-Cell Activation and Non-Hodgkin Lymphoma Risk

in an HIV Positive Population

by

Po-Yin Chang

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2013

Professor Zuo-Feng Zhang, Chair

Background: B-cell non-Hodgkin lymphoma (NHL) in HIV populations (AIDS-NHL) has become the leading cause of AIDS-defining cancers. Studies suggested that genetic or serum markers of B-cell activation are related to AIDS-NHL. However, associations between HIV viral load and AIDS-NHL risk have not been explicitly explored with consideration of B-cell activation markers. Furthermore, associations of hepatitis C virus (HCV) infection to AIDS-NHL risk are inconclusive.

Methods: We used two nested case-control studies within the Multicenter AIDS Cohort Study – the genetic association study including 183 AIDS-NHL cases and 533 matched HIV-positive controls, and the serum cytokine study of 179 AIDS-NHL cases and 179 HIV-infected controls – to explore associations between HIV viral load or HCV infection and AIDS-NHL risk. To screen out possible B-cell activation biomarkers, we implemented weighted gene co-expression network analysis (WGCNA) and visANT. We explored three HIV viral load measurements: (1) viral load at set point, indicating a balance between HIV viral replication and clearance, (2) pre-HAART viral load closest and prior to AIDS-NHL diagnosis, and (3) slope and intercept for linear regression on viral load by year. HCV infection was measured at multiple time-points retrospectively using stored serum specimen or prospectively since 2001. We calculated odds ratios (ORs) and 95% confidence interval (CIs), using conditional logistic regression models including potential confounders and additional single nucleotide polymorphisms (SNPs) or serum cytokines levels suggested by the WGCNA.

Results: The WGCNA screened out sCD27, sCD30, and CXCL13 from 328 SNPs and 31 cytokines measured longer than 3-year preceeding AIDS-NHL diagnosis. One-unit increase of $\log_{10}$-HIV RNA at set point was associated with an increased AIDS-NHL risk (SNPs adjusted OR: 2.01; 95% CI: 1.35-2.99, sCD27 and sCD30 adjusted OR: 3.65; 95% CI: 1.52-8.72). Ever HCV infection was possibly associated with AIDS-NHL susceptibility (adjusted OR: 1.77; 95% CI: 0.89-3.50). Association remained after SNP adjustment (OR: 1.75, 95% CI: 0.97-3.13).

Conclusion: Our results suggested a higher viral-load set point was associated with elevated AIDS-NHL risk from models with or without biomarkers. Point-estimates consistently indicated associations between HCV and AIDS-NHL risk.

The dissertation of Po-Yin Chang is approved.

Roger Detels

Otoniel Martínez-Maza

Janet Sinsheimer

Zuo-Feng Zhang, Committee Chair

University of California, Los Angeles

2013

DEDICATION

For Tien, Yu-Chiao, and J.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES viii

ACKNOWLEDGEMENTS x

VITA xii

CHAPTER 1: INTRODUCTION AND BACKGROUND 1

1.1 Introduction 1

1.2 Epidemiology of Non-Hodgkin's Lymphoma in HIV Positive Populations 2

1.3 Risk Factors for HIV-related NHL 3

1.4 Lymphomagenic Mechanisms for HIV-related NHL 3

1.5 Chronic B-cell Hyperactivation 4

1.6 DNA Modification and Lymphomagenesis in B-cell Development 5

1.7 B-cell Lymphoma and Genetic Variations in AICDA Pathway and DNA Repair Mechanism 6

1.8 Hepatitis C Virus and B-cell Lymphoma 9

1.9 Gaps in AIDS-NHL Research 10

CHAPTER 2: RESEARCH OBJECTIVES AND METHODOLOGY 12

2.1 Research Objectives 12

2.2 Specific Aims and Hypotheses 12

2.3 Study Design and Methods 14

2.4 Study Population 15

2.5 Determination of Serum Levels of Inflammation-related Markers (Cytokines) 17

2.6 SNP Selection and Genotype Determination 21

2.7 Statistical Analysis 24

CHAPTER 3: RESULTS 37

3.1 Baseline Demographics 37

3.2 Network Analysis 45

3.3 Network Visualization 46

3.4 HIV Viral Burden and AIDS-NHL Risk 54

3.5 HCV Infection and AIDS-NHL Risk 61

CHAPTER 4: DISCUSSION 65

4.1 Network Analysis 65

4.2 HIV Viral Burden and AIDS-NHL Risk 70

4.3 HCV Infection and AIDS-NHL Risk 74

CHAPTER 5: CONCLUSION AND PUBLIC HEALTH IMPLICATIONS 77

CHAPTER 6: FUTURE WORKS 79

Network Analysis 79

Shrinkage Regression 81

APPENDIX I. SUPPLEMENTARY TABLES 83

REFERENCES 84

LIST OF FIGURES AND TABLES

Figure 1. Transcriptional regulation of *AICDA* expression by several molecules and receptors, including toll-like receptors (TLRs), B-cell receptor (BCR) complex, cytokine receptors, and TNF receptor family. IL-4, TGF-β, and CD40L are known enhancers for AICDA expression. 7

Figure 3.1. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.001. IL-10 had most links (21 links) to other biomarkers, followed by sCD30, sCD27, IP-10, IL-2, and λ free light chain. 49

Figure 3.2. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.005. sCD27 had most links (14 links) to other biomarkers, followed by CXCL13, IP-10, IL-6, and neopterin..... 50

Figure 3.3. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.01. IP-10 had the most links (13 links) to other biomarkers, followed by CXCL13, IL-6, sCD27, sCD30, and neopterin. 51

Figure 3.4. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.05. sCD27 had most links (9 links) to other biomarkers, followed by IP-10, sCD30, neopterin, and κ free light chain. 52

Figure 3.5. Network of 31 biomarkers with topological overlap matrix (TOM) larger than 0.1. sCD27 was the potential hub biomarker that linked to neopterin, IP-10, sCD30, CXCL13, and κ free light chain..... 53

Figure 3.6. Summarized and strata-specific adjusted OR relating HIV VL at set point and B-cell lymphoma susceptibility in HIV+ MACS participants. Forty-two cases and thirty-five controls had missing cytokines levels. The analysis was stratified by serum medium levels of sCD27, sCD30, neopterin, IP10, and CXCL13, after natural log transformation. These molecules are possible cytokines representing B-cell activation identified in the network analysis..... 60

Figure 3.7. Summarized and strata-specific aOR relating HCV and lymphoma susceptibility in HIV-infected MACS participants. The analysis was stratified by serum medium levels of sCD27, sCD30, neopterin, IP10, and CXCL13, after natural log-transformation. These molecules are possible cytokines representing B-cell activation identified in the WGCNA. 64

Table 3.1. Distribution of demographic factors of cases and controls from genetic (n = 716) and serum marker (n = 358) association study 40

Table 3.2. Distribution of HIV RNA measurements during MACS follow-up..... 44

Table 3.3. Listing of module-specific biomarkers measured larger than 3 year prior to reference date, by correlation coefficient with its module eigenvector (ME) 45

Table 3.4. Listing of markers for network visualization (13 SNPs and 18 serum cytokines measured longer than 3 year preceding reference date) and their correlations with the NHL status 47

Table 3.5. Associations between different measures of HIV viral burden and AIDS-NHL risk, adjusted for B-cell activation-related genetic markers in the genetic association study (n = 716) 56

Table 3.6. Associations between measures of HIV viral burden and AIDS-NHL risk, adjusted for serum levels of sCD27 and sCD30 in the serum marker study (n = 358) 59

Table 3.7. Associations between HCV infection and AIDS-NHL risk in genetic association study (n = 716) and serum marker study (n = 358) 63

Table S4.1. Spearman correlation coefficients among 10 serum inflammation-related markers (cytokines) (> 3 years preceding AIDS-NHL diagnosis) in the first module from the WGCNA, by case and control status 83

ACKNOWLEDGEMENTS

We go throughout life with our memories, and those memories travel with us as we move forward, influencing our path. This dissertation summarizes a huge part of my life, and my memories so far in the United States. This journey is full of acknowledgements, both profession and recreational, yet all very personal.

First of all, I want to express my deepest gratitude to my academic advisor and committee chair, Dr. Zuo-Feng Zhang. Dr. Zhang constantly guided and trained me in the art of scientific research though both his passion and insights on cancer epidemiology and tremendous morals. He has taught me more than just science, he has taught me about life and philosophy.

I also want to extend my great appreciation to Dr. Shehnaz Hussain for generously providing guidance and research materials. In particular, I enjoyed every meeting with her and always felt fully-charged from her radiating personality. I would also like to thank Dr. Roger Detels for his warmest care of me as an international student, and guidance on my research of the Multicenter AIDS Cohort Study. I am grateful to Dr. Otoniel Martínez-Maza for his expertise and generous nature, as well as for providing research materials. Lastly, Dr. Janet Sinsheimer has provided tremendous and insightful advice, as well as kindness.

I especially acknowledge all MACS participants and staff for their contributions. This research could never have been finished without them.

I particularly thank Dr. Shen-Chih Chang for her patience, helpful suggestions, and adorable personality. For Gina and Erin, I cannot emphasize more on how much I appreciated your help and support during our Ph.D. training; nor can I emphasize enough how lucky I am to

have had you as a group, taking care of one another. In addition, I want to thank my colleagues in the Cancer Epidemiology group: Travis Meyers, Ying Yang, Claire Kim, Fayth Miles, Heather Tarleton, and Fong Liu. We have shared our laughs, ideas, complaints, stress, snacks, and coffee. All of you have helped transform my life at UCLA into the thoughts and words of my dissertation.

My volleyball fellows and roommates at UCLA are my family in the U.S.: Jason, Vincent, Mariah, En-wei, Riva, Janie, Ben, Sean, Shelly, Jackie, Dennis, Brady, Charles, Brian, Jay, Hao, Liza, Terryl, and Nicole. Each one of you has a special place in my life. Every volleyball game we played, meal we shared, trail we hiked, and every snowfall we saw fulfilled and completed my life the past six years. I thank you all for being my family and leaving me such precious memories.

Finally, I would like to thank my parents and siblings. No words can express my love toward them. Their unconditional love has supported my walking this far away from home, and yet remaining so closely bound. I want to say thank you and I love you all.

VITA

September 14, 1981	Born, Taoyuan, Taiwan
1999 – 2003	National Taiwan University (NTU), Taipei, Taiwan BS in Pharmacy Presidential Award Student Service Award, Taipei Pharmacists Association, Taiwan
2003 – 2005	Graduate Institute of Clinical Pharmacy, NTU, Taipei, Taiwan MS in Clinical Pharmacy Thesis title: Analysis of lifetime medical cost and establishment of chemotherapy cost model in patients with stage IV non-small cell lung cancer
2003 – 2004	Intern Pharmacist National Taiwan University Hospital, Taipei, Taiwan Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan
2004 – 2005	Assistant Taiwan Drug Relief Foundation, Taipei, Taiwan Adverse Drug Reaction Evaluation team
2005	Research Assistant NTU Graduate Institute of Clinical Pharmacy
2005 – 2007	2nd Lieutenant, Military Police, Presidential Office, Taipei, Taiwan
2007 – 2008	Research Assistant National Yang-Ming University, Taipei, Taiwan Institute of Health and Welfare Policy
2008	Admission to MPH program in Epidemiology University of California, Los Angeles (UCLA) Fielding School of Public Health
2009	Ph.D. program in UCLA Epidemiology
2009 – 2012	UCLA Graduate Division Unrestricted Fellowship Award
2010	Research Assistant UCLA Epidemiology
2011 – 2012	Research Assistant

	UCLA Epidemiology
2012 – 2013	Special Reader II UCLA Epidemiology EPID 200C "Methods III: Analysis"
2012	UCLA Graduate Division Dissertation Year Fellowship
2013	Staff Research Associate II UCLA Epidemiology

PUBLICATIONS AND PRESENTATIONS

*Yang T, ***Chang PY**, Park SL, Bastani D, Chang SC, Morgenstern H, Tashkin DP, Mao JT, Papp JC, Rao JY, Cozen W, Mack TM, Greenland S, Zhang ZF. Tobacco smoking, *NBS1* polymorphisms, and survival of lung and upper aerodigestive tract cancers with semi-Bayes adjustment for hazard-ratio variation. *Cancer Causes Control*. 2013 Oct 29 [PubMed PMID: 24166361] (*same contributions)

Hsieh CF, **Chang PY**, Lu PY, Huang WF. The patterns of outpatient off-label carbamazepine use and the potential impact of regulatory labelling process in Taiwan. *J Pharm Health Services Res.* 2011; 2:165-73.

Hsiao FY, Tsai YW, Huang WF, Wen YW, Chen PF, **Chang PY**, Kuo KN. A comparison of aspirin and clopidogrel with or without proton pump inhibitors for the secondary prevention of cardiovascular events in patients at high risk for gastrointestinal bleeding. *Clin Ther* 2009; 3: 2038-47

Chang PY, Hussain SK, Martinez-Maza O, Detels R, Zhang ZF. HIV viral load and non-Hodgkin lymphoma risk. Center for AIDS Research (CFAR) Cross-disciplinary Conference, February 5, 2013, UCLA.

Hussain SK, **Chang PY**, Martinez-Maza O, Zhang ZF, Jacobson L, Detels R. Risk factors for squamous cell skin cancer in HIV. *Infect Agent Cancer*. 2012;7:O26.

Zhang ZF, Chang SC, **Chang PY**, Yang Y, Rao JY. Single nucleotide polymorphisms in genes from multiple pathways and survival of bladder cancer. 5th International Consortium of Bladder Cancer (ICBC) Meeting, 2012, Madrid, Spain.

Chang PY, Zhang ZF, and the Lung and UADT cancer study team. Associations between variants of the 8q24 chromosome and survival in smoking-related cancer sites. 35th Annual Meeting of the American Society of Preventive Oncology (ASPO), March 5-8, 2011, Las Vegas, NV.

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 Introduction

Non-Hodgkin Lymphoma (NHL) is a heterogeneous disease with different histological, pathological, and molecular pathogenic profiles. Due to its high incidence among HIV-infected individuals, NHL was defined as one of the AIDS-defining illnesses by the U.S. CDC in 1985.[1] The introduction of the highly active anti-retroviral therapy (HAART) in 1996 has led to reduced incidence of some, but not all,[2, 3] subtypes of NHL. The risk of NHL in HIV seropositive (HIV+) populations in the HAART era is still elevated with a sex-, age-, and race-standardized incidence ratio (SIR) of 6.5 (95% CI: 5.4 – 7.7), using the pre-AIDS epidemic general population as the reference group.[3] Several epidemiological and histologic profiles of HIV-related NHL (AIDS-NHL) are distinct from those in immunocompetent or general populations. First, almost all AIDS-NHL is B-cell lymphoma of germinal center (GC) or post-GC origin.[4, 5] Second, some subtypes of B-cell lymphomas occur exclusively in HIV+ populations, including primary effusion lymphoma (PEL) and plasmablastic lymphoma (PBL) of the oral cavity.[6] Third, while the primary central nervous system lymphoma (PCNSL), one subtype specifically observed in HIV+ populations, is mainly Epstein-Barr virus (EBV) dominant, NHL in general populations and most Burkitt lymphoma (BL) in HIV+ populations are EBV negative.[4, 7]

While incidence of EBV-related lymphoma has been decreased in the HAART era, incidences of AIDS-NHL subtypes that are not necessarily EBV positive remain similar.[2] It has been proposed that chronic B-cell hyperactivation plays essential role in the lymphomagenesis of AIDS-NHL. A number of studies supported associations between AIDS-NHL risk and biomarkers that are indicative of B-cell activation or immune function.[7-19] However, associations between AIDS-NHL and its risk factors, such as HIV viral burden and HCV infection, with

adjustment for B-cell activation-related biomarkers are unclear. The objectives of this study are to identify potential biomarkers from existing abundant genetic and serum biomarkers, and to explore associations between HIV viral load or HCV infection and AIDS-NHL risk, in models that contain these potential biomarkers.

1.2 Epidemiology of Non-Hodgkin's Lymphoma in HIV Positive Populations

About four percent of AIDS patients have NHL as their first AIDS-defining illness and 60-70% of AIDS-NHL patients show advanced stage with high or intermediate grade at diagnosis.[20] According to World Health Organization (WHO) NHL classification,[21] several subtypes are common in AIDS-NHL, including Burkitt lymphoma (BL), diffuse large B cell lymphoma (DLBCL) with centroblastic and immunoblastic morphologies, and primary central nervous system lymphoma (PCNSL). Approximately 80% of AIDS-NHL occur at extranodal sites and central nerves system (CNS) is the most common site. PCNSL accounts for 15-30% of HIV-associated NHL.[6, 22] Primary effusion lymphoma (PEL) and plasmablastic lymphoma (PBL) of the oral cavity are two rare, HIV-specific lymphomas (5% of all AIDS-NHL).[6]

Incidence of PCNSL and DLBCL has decreased in the HAART era. The SIR for PCNSL reduced from 8.4 (95% confidence interval, CI: 6.9-10.0/1,000 person-years) to 1.1 (95% CI: 0.7-1.7/1,000 person-years) in the post-HAART era.[23] It has been noticed that patients with PCNSL or DLBCL have EBV positive lymphoma and relative low $CD4^+$ T-cell counts at lymphoma diagnosis. On the contrary, compared to the pre-HAART era, BL incidence is stable or even marginally increased. BL patients usually have normal to mildly reduced $CD4^+$ T-cell count at diagnosis. Engels *et al*[3] reported higher SIR in the post-HAART era (SIR_{HAART}: 17; 95%

CI: 8.6-31/100,000 person-years) compared to the pre-HAART era ($SIR_{pre\text{-}HAART}$: 7.1, 95% CI: 0.2-40/100,000 person-years) with an adjusted RR of 3.8 (95% CI: 0.5-29.4).

HIV infected people have longer time at risk for lymphoma in the HAART era, research on predictive markers or risk factors for NHL are needed, particularly on B-cell hyperactivation in NHL subtypes that are not EBV positive.

1.3 Risk Factors for HIV-related NHL

Decreased immunity, represented by $CD4^+$ T-cell count, is the most important risk factor for AIDS-NHL.[3, 24-26] Other risk factors include older age,[26, 27] higher HIV viral load or higher cumulative viremia,[26, 28-30] not-receiving HAART,[25, 26] longer time since seroconversion,[24] and EBV infection.[31] Some oncogenic viruses related to B-cell lymphomagenesis are potential risk factors for AIDS-NHL, including EBV,[6, 20, 22, 31] Kaposi sarcoma-associated herpes virus/human herpes virus 8 (KSHV/HHV8),[32-34] and hepatitis C virus[35-37]. While several studies indicated HCV infection is positively associated with moderate NHL risk in general populations,[38-40] associations between HCV and AIDS-NHL are inconclusive.[41-46]

1.4 Lymphomagenic Mechanisms for HIV-related NHL

At least two lymphomagenic mechanisms have been proposed for AIDS-NHL: (1) aberrant immunoregulation of EBV-infected B-cells and (2) chronic B-cell hyperactivation.[4, 47] With gradually dysfunctional cellular immunity in HIV+ individuals, EBV can consistently affect B lymphocytes without eradication and lead to B-cell lymphomas.[47] Piriou *et al* has

observed that EBV+ NHL patients lose $CD8^+$ T cells specific to EBV and $CD4^+$ T cells specific to EBV nuclear antigen 1 (EBNA 1) prior to lymphoma, compared with patients with EBV- NHL.[48]

Not all AIDS-NHL tumors are EBV+, in particular systemic lymphomas.[6, 22, 49] The lymphomagenesis due to chronic B-cell hyper-activation related to HIV infection or other oncogenic viruses, has been postulated.[47, 50]

1.5 Chronic B-cell Hyperactivation

HIV itself can induce B-cell activation through CD40-dependent or independent mechanism. Elevated level of serum immunoglobulin (Ig) and cytokines have been noticed in HIV positive individuals,[4, 51, 52] indicating a hyperactivated B-cell production by HIV infection. These molecules include IL-4, IL-6, IL-10, B cell-activating factor of the TNF family (BAFF), type I Interferon-α (INFα), tumor necrosis factor α (TNF-α), and Ig level (IgG, IgA and IgE).[53-57] Some cytokines are able to regulate *AICDA* transcription, others indicate B-cell stimulation or immune activation. Epeldegui *et al* presented a direct induction of *AICDA* expression by HIV virions in human B lymphocytes,[58] suggesting a potential direct lymphomagenesis by HIV virus, a non-oncovirus, through CD40 ligand-dependent mechanism.

As most AIDS-NHL patients have lymphoma of B-cell origin, observational studies also suggest associations of AIDS-NHL risk with abnormal expression of B-cell activation molecules, such as cytokines, chemokines, Ig free light chains (FLC), and *AICDA* product.[10, 12, 59-64] Grulich *et al* reported positive association of NHL risk with elevated concentrations of serum globulin (a surrogate marker for serum Ig) and IgG, supporting the B-cell stimulation mechanism in HIV-

related NHL.[24] Some cytokines such as IL-1β, IL-4, IL-10, and TNF-α could be induced by HIV or HCV and enhanced hypermutation or class switch recombination of Ig in B-lymphocytes.[36, 54] Some cytokines including IL-6, IL-10, sCD23, sCD27,sCD30, sCD44, CXCL13, CXCR5, IP10, neopterin, and Ig FLC show increased levels in HIV-related lymphoma cases, but not in HIV positive controls.[7-11, 13-15, 17, 18] These molecules all relate to B-cell stimulation, interact with one another, and support the hypothesis of B-cell hyperactivation in AIDS-NHL.

1.6 DNA Modification and Lymphomagenesis in B-cell Development

During B-cell maturation and differentiation, immature B cells migrate from bone marrow to germinal centers (GC) in peripheral lymph nodes and undergo class switch recombination (CSR) and somatic hypermutation (SHM). The DNA strand breaks and repairs in both CSR and SHM are necessary to increase the diversity (CSR) and affinity (SHM) of Ig.[65]

The activation-induced cytidine deaminase (AICDA) is indispensable for CSR and SHM.[66, 67] *AICDA*, along with genes in DNA repair pathway, might potentially contribute to lymphomagenesis, if mutations and translocations accumulate during CSR and SHM. AICDA sometimes targets proto-oncogenes[68] and several lines of evidence supported that AICDA leads to the *c-MYC*/*IgH* translocation in BL or various *BCL6* translocations in DLBCL.[69-71] The *AICDA* deficiency has been shown to prevent the *bcl6* dependent, GC-derived B-cell lymphoma in mouse models.[50] Several viruses, including HIV,[58] HCV,[36, 72] and EBV,[73, 74] can induce *AICDA* expression. Higher *AICDA* expression was observed in peripheral mononuclear blood cells in AIDS-NHL cases prior to diagnosis, compared with HIV positive and negative controls.[12]

Additionally, a number of studies reported increased NHL risk associated to genetic variations in DNA repair pathway, including non-homologous end joining (*RAG1*[75] and *LIG4*[76]), homologous recombination (*XRCC3*[76] and *BRCA2*[75]), nucleotide excision repair (*ERCC5*[77]), or base excision repair (*XRCC1*[76]). Observational and experimental study results suggest *AICDA* pathway and DNA repair pathway relate to B-cell activation. Their associations with NHL in HIV+ populations remain to be elucidated.

1.7 B-cell Lymphoma and Genetic Variations in AICDA Pathway and DNA Repair Mechanism

Several molecules and receptors, as shown in Figure 1,[78] are thought to regulate transcription of *AICDA*. These molecules belong to toll-like receptors (TLRs), B-cell receptor (BCR) complex, cytokine receptors, and TNF receptor family. IL-4, TGF-β, and CD40L are known enhancers for *AICDA* expression.[68] During the process of maturation and differentiation at the GC, B cells undergo DNA-modifying processes to increase diversity and affinity of Ig. AICDA is the most important enzyme to these error prone processes. Therefore, genetic variations on *AICDA* pathway and DNA repair mechanisms might be important in B-cell lymphomagenesis and NHL susceptibility.

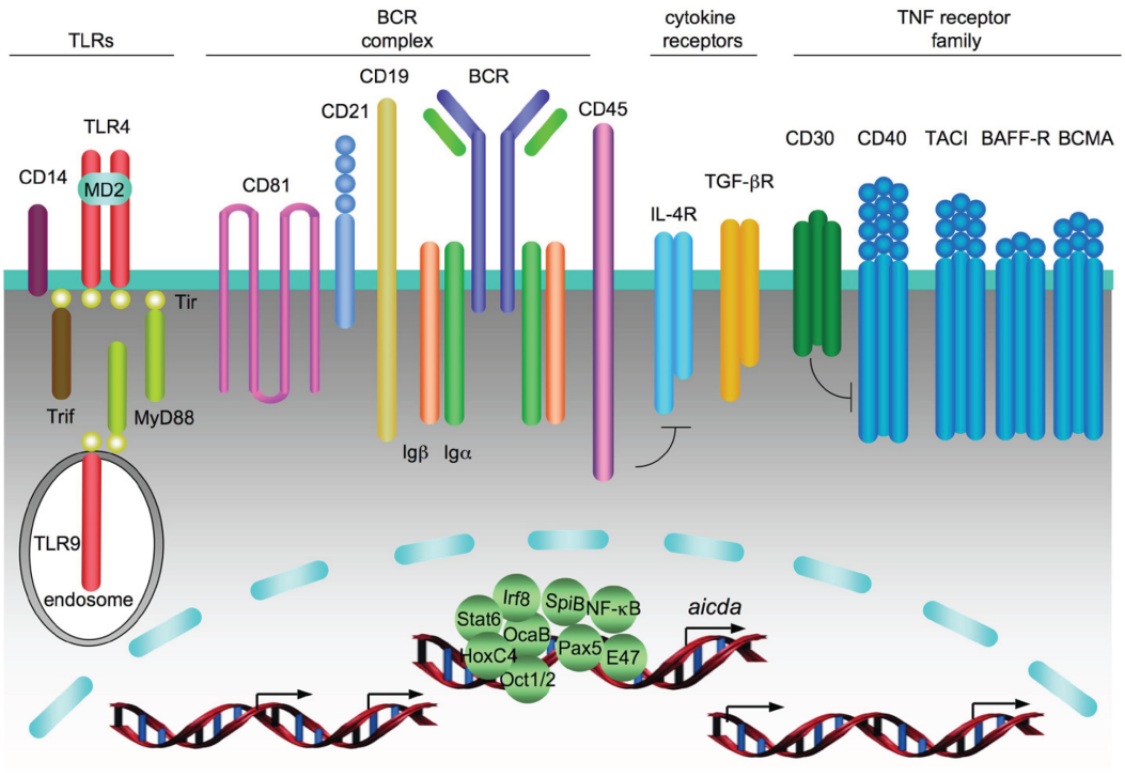


Figure 1. Transcriptional regulation of *AICDA* expression by several molecules and receptors, including toll-like receptors (TLRs), B-cell receptor (BCR) complex, cytokine receptors, and TNF receptor family. IL-4, TGF-β, and CD40L are known enhancers for AICDA expression.[68]

Studies have reported positive relationships between NHL and various single nucleotide polymorphisms (SNPs) in the TNF super family pathway, in some populations.[79-84] SNPs in tumor necrosis factor (*TNF* rs1800629),[84] TNF receptor super families (*TNFSF13B* rs2582869, *TNFSF13C* rs6002551, *TNFSF7* rs16994592, and *FAS* rs4934436),[83] interferon regulatory factor 4 (*IRF4* rs12211228),[83] and *IL-10* (rs1800890, rs1800896)[84] were found to be associated with non-HIV-related NHL. On the other hand, no clear associations were found between NHL and SNPs on DNA repair genes in base excision repair (BER), mismatch repair (MMR), and DNA polymerases.[76, 77, 85-90] Only two *MGMT* SNPs showed borderline positive association with NHL risk (*MGMT* rs2308321, AG/GG vs AA: OR =1.33, 95% CI: 1.00-1.78; *MGMT* rs2308327, AG/GG vs AA: OR = 1.31, 95% CI: 0.99-1.75).[76]

Limited research on genetic polymorphisms has been done in the HIV positive population.[9, 16, 60] The *IL-10* rs1800871 SNP showed an inverse association with AIDS-NHL risk in the CNS subtype (CC vs. CT/TT: OR = 0.3, 95% CI: 0.1-0.7), but not systemic lymphoma (CC vs. CT/TT: OR = 1.0) in .[16] No other associations for SNPs on *IL10RA*, *CXCL12*, *IL-13*, *IL-4*, *IL-4R*, *CCL5*, and *BCL6* were found in the same study.[16] Associations between AIDS-NHL risk and tagSNPs in *CXCL13* or *CXCR5* have been explored.[9] An inverse association was observed for AIDS-NHL and the minor allele of *CXCL13* rs355689 (TC vs TT, OR = 0.65, 95% CI: 0.45-0.96). *CXCL13* rs355689 was also associated with reduced serum CXCL13 level. These results indicated potential associations between lymphomagenesis and CXCL13 genotype that was supported at CXCL13 phenotype.

1.8 Hepatitis C Virus and B-cell Lymphoma

It has been consistently reported in several observational studies that HCV infection is associated with increased risk of B-cell lymphoproliferative disorders such as mixed cryoglobulinemia[91] and NHL[38-40] in general populations. The moderate to strong associations between NHL and HCV infection are noticed, in particular with B-cell NHL (OR=5.04, 95% CI: 3.59-7.06) and extranodal NHL (RR=3.72, 95% CI: 2.37, 5.86).[38-40] The chronic B-cell activation driven by HCV is implicated by higher *Ig* hypermutation, as well as increased expression of AICDA and DNA polymerases in B lymphocytes after HCV infection.[36, 37] Furthermore, Ito *et al* observed enhanced expression of B-cell lymphoma genes, including *BAL*, *STK13*, *CCND1*, and *CCND2*, in peripheral blood B cells of patients with chronic HCV infection.[92]

Such associations, however, could not be clearly replicated in HIV positive populations from case-control studies nested within large AIDS or HIV cohorts[42, 43] or a cohort study using registry database.[44] Unlike studies in the general population, studies in HIV populations usually had far smaller sample size and were subject to bias from uncontrolled confounders ($CD4^+$ T-cell count, HIV viral load) and potential HCV measurement errors.

Alternatively, HCV could have smaller oncogenic effect on NHL than does EBV or HIV.[42, 43] Preliminary analyses from a case-control study nested within the multicenter AIDS cohort study (MACS), however, found increased risk for AIDS-NHL (OR = 1.88, 95% CI: 1.09-3.21) and systemic AIDS-NHL (OR = 1.86, 95% CI: 1.04-3.31).[41] Increased AIDS-NHL risk was also associated with ever HCV infection in a sub-population of men who have sex with

men.[42] Further studies with larger sample size and prospective HCV exposure measurement are required to elucidate inconclusive results and differences between general and HIV+ populations.

1.9 Gaps in AIDS-NHL Research

Chronic B-cell hyperactivation in HIV+ populations has long been recognized with elevated serum biomarkers in the HIV infected individuals. Previous studies suggested associations between serum or genetic biomarkers and AIDS-NHL risk.[7-9] As all biomarkers are related to B-cell stimulation, studies were subject to collinearity among covariates and potential multiple-comparison issues. In addition, while measures of hundreds of biomarkers are accessible, pathway-based or systems biology approaches are needed to explore biomarkers indicative for B-cell stimulation and/or predictive for AIDS-NHL.

Epidemiological studies have indicated that HIV viral load and cumulative HIV viremia are predictive risk factors for NHL in HIV positive populations, independent to $CD4^+$ T cell count.[26, 28-30] Cumulative HIV viremia (area under the curve of HIV viral load by time), HIV viral load at baseline or the time of lymphoma diagnosis, are main exposure of interest. However, clinical interpretation is not clear. For example, the HIV viral load at lymphoma diagnosis shows little practical value to 'predict' the NHL occurrence. Also, high HIV viral loads at baseline do not necessarily indicate sub-optimal HIV control in the future. Additionally, B-cell activation biomarkers are not included in the statistical models. Although uncontrolled or sub-optimal HIV infection is a risk factor for NHL, the strength and extent of associations of HIV viral burden with AIDS-NHL risk, while adjusting for B-cell activation-related biomarkers, is not known.

As HCV infection is a consistently suggested risk factor for NHL in general populations, results for associations between HCV infection and AIDS-NHL risk are inconclusive. Although

several lines of evidence support the hypothesis of HCV-related B-cell lymphomagenesis, there are limited studies focusing on HCV and AIDS-NHL risk. Furthermore, most studies were relatively small and did not adjust for confounding factors.

CHAPTER 2: RESEARCH OBJECTIVES AND METHODOLOGY

2.1 Research Objectives

Study objectives are to explore and confirm associations between risk factors and NHL susceptibility in an HIV-positive population with consideration of genetic and/or serum markers that are related to B-cell activation.

2.2 Specific Aims and Hypotheses

Objective 1: Potential biomarkers essential to B-cell activation

Hypothesis 1

We hypothesize that genetic susceptibility and/or serum inflammation-related markers (cytokines) may be correlated to B-cell activation.

Specific Aim 1

To screen out potential biomarkers (genetic and/or serum cytokines) from correlated markers, including 32 serum cytokines measured at three time-points prior to AIDS-NHL diagnosis and 329 genetic polymorphisms at 41 genes on the DNA repair pathway, using network analysis. Potential biomarkers suggested by network analysis will be further included in HIV and HCV models.

Objective 2: HIV viral load and AIDS-NHL

Hypothesis 2

It is known that a higher HIV viral burden may increase the risk of AIDS-NHL. We hypothesize that this relationship will remain when adjusting for genetic and/or serum biomarkers identified by network analysis that are related to B-cell activation.

Specific Aim 2

To evaluate potential associations between measures of HIV viral load and AIDS-NHL risk in an HIV-infected population adjusting for B-cell activation related markers. Measures of HIV viral load include (a) viral load at natural steady-state in early infection (the set point) before the initiation of antiretroviral therapy, (b) viral load prior and closest to AIDS-NHL diagnosis, and/or (c) intercept and slope for linear regression on viral load by time. HIV models will further include potential biomarkers identified from network analysis.

Objective 3: HCV infection and AIDS-NHL

Hypothesis 3

Studies show that HCV infection is positively associated with NHL risk in general populations and may be correlated to B-cell activation. We hypothesize that HCV infection prior to NHL diagnosis is positively associated with AIDS-NHL risk in HIV+ individuals, with or without adjustment for B-cell activation-related biomarkers.

Specific Aim 3

To evaluate potential associations between ever/never HCV infection, as well as chronic HCV infection, and AIDS-NHL risk in an HIV-positive population. Models will further include biomarkers that are identified from specific aim 1.

2.3 Study Design and Methods

Study Overview

This study used data from two case-control studies nested within an ongoing prospective Multicenter AIDS Cohort Study (MACS) which are conducted in Baltimore, Chicago, Pittsburgh, and Los Angeles. The MACS study has a total enrollment of 6,972 homosexual and bisexual male participants from three recruitment periods: (1) April 1984 to March 1985 (4,954 participants); (2) April 1987 to September 1991 (668 participants); (3) October 2001 and August 2003 (1,350 participants).[93, 94] In addition to the baseline questionnaire, the MACS participants biannually receive in-depth follow-up interviews regarding demographics, psychosocial field, behavioral and attitude field, medical history, medication use, antiviral medication adherence, and clinical outcomes. [95] They are also asked to provide blood samples at each visit. Abundant bio-information is collected through each visit, including HIV viral loads, T-cell counts, EBV infection profile, and HCV infection status, etc. Biological specimens are stored in the central repository in Baltimore.

As of May 2010, 666 of 4,088 (16.3%) sero-negative MACS participants became HIV positive (sero-converters) and 311 of them developed AIDS. Among 2,884 sero-prevalent MACS participants, 1,597 participants developed AIDS.[96] The follow-up rate of HIV positive participants is up to 86.7% in MACS cohort.[96]

We used two existing nested case-control studies within the MACS cohort to examine our specific aims: the genetic association study[9] and the serum marker study.[7, 9] Both studies included HIV positive participants. The genetic association study included 183 NHL cases and 533 HIV positive controls to explore associations between AIDS-NHL risk and genetic

polymorphisms located at genes related to DNA repair pathway during somatic hypermutation (SHM) or class switch recombination (CSR) processes. The serum marker association study included 179 NHL cases and 179 HIV positive controls and focused on associations between AIDS-NHL risk and 32 serum inflammation-related markers (cytokines) measured at three time-points prior to lymphoma diagnosis.

2.4 Study Population

Case Definition

We selected the MACS participants who developed pathologically-confirmed NHL during the follow-up before April 2003 (the serum marker study)[7, 9] and July 2010 (the genetic association study).[9] The AIDS-NHL diagnosis was obtained at autopsy, or confirmed with both pathology reports and state cancer registries. The center for analysis and management of MACS data (CAMACS) provided information of tumor morphology, topology, and NHL diagnosis dates. We used the WHO 2008 classification and InterLymph hierarchical classification for epidemiologic research to determine NHL subtypes and locations.[97] Due to different inclusion period and serum/DNA sample availability, two studies did not contain the same AIDS-NHL cases. There were 172 overlapped cases in both studies. Eleven and seven cases were included specifically in the genetic and serum studies.[9]

Control Definition and Matching Criteria

Table 2.1 lists the number of participants and matching factors in the genetic association (3-to-1 matching) and serum marker (1-to-1 matching) studies. We implemented individual matching in both studies. HIV positive controls in both studies were matched on duration of HIV infection and HIV sero-status (HIV prevalent participants or sero-converters during the MACS

follow-up). These matching factors were chosen to decrease the strong confounding effect of HIV infection period on AIDS-NHL risk. In the genetic association study, the $CD4^+$ T-cell count at the time of lymphoma diagnosis was matched. In the serum marker study, the HIV positive controls were matched on the biospecimen availability and the time of sampling. One hundred and seventy-two cases overlapped in both studies and had information of both serum inflammation-related markers (cytokines) and single nucleotide polymorphisms (SNP).[9] There were only 50 overlapped controls in both studies. We additionally genotyped controls from serum marker study. In total, 174 controls in serum marker study had genotyping data.

Table 2.1. Matching factors and available biomarkers in two nested case-control studies

	Case-control study nested within the MACS	
	Genetic association study	Serum marker study
Case number[1]	183	179
Control number[2]	533 (3-to-1 matching)	179 (1-to-1 matching)
Matching factors	1. MACS enrollment period 2. Date of HIV infection 3. Duration of follow-up 4. Race 5. CD4 categories (0-49, 50-59, 100-199, 200-349, 350-499, >500 counts/mL) 6. Sero-status at enrollment	1. Duration of HIV infection based on the known date of HIV seroconversion (21 cases), or date of entry (± 1 year) into MACS as HIV-seroprevalent (158 cases) 2. Serum sample availability at 3 sampling time points (± 1 year): >3 years pre-NHL (closest to 4 years), 1-3 years pre-NHL (closest to 2 years), 0-1 year pre-NHL (closest to 0.5 years).
Available biomarkers	SNP	Serum inflammation-related markers (cytokines); SNP

SNP single nucleotide polymorphisms
[1]172 cases were included in both studies. [2]50 controls were included in both studies.

2.5 Determination of Serum Levels of Inflammation-related Markers (Cytokines)

A number of studies have shown that AIDS-NHL individuals have elevated serum or plasma levels of B-cell stimulatory cytokines, compared to HIV+ controls.[7-15, 17, 24, 58, 60, 62, 98] These cytokines include interleukins, chemokines, soluble forms of cytokine receptors, free light chains (FLC) of immunoglobulins (Ig), and other molecules associated with immune system activation or inflammation. These molecules stimulate, indicate, or participate in B-cell maturation, differentiation, and/or Ig production. Table 2.2 lists cytokines included in our research, their functional classification, and putative immune-activating functions in the first two columns.

The enzyme-linked immunosorbent assay (ELISA) method was used to measure serum cytokines levels at three time points. We followed manufacturers' protocols to perform these tests. Detailed methods for measurement can be found in previous publications.[7-9] Serum levels of some interleukins (IL-1β, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12), CXCL13, INF-γ, TNF-α, GM-CSF, and VEGF were determined by the Fluorokine® MAP Human Inflammation Kit (R&D Systems, Minneapolis, MN, USA). The levels of sCD40L, other interleukins (IL-17A, IL-17F, IL-20, IL-23, IL-27, IL-31, LIF), IP-10 and SDF-1α were measured using the Procarta™ Cytokine Assay (Affymetrix, Santa Clara, CA, USA). Both Fluorokine® and Procarta™ assays use ELISA-based, high-throughput Luminex platform. The fluorescence intensity of micro-particles for each marker was quantified using a Bioplex 200™ (Luminex) System Analyzer (Biorad, Hercules, CA, USA). Data was analyzed using BioPlex Manager (v 4.1.1) software. We set detection limit at either the lowest value that the BioPlex Manager software could calculate (using the standard curve), or the lowest value of the standard curve, whichever was lower. Table 2.2 lists the assays, manufacturers, and detection limits for inflammation-related markers.

Serum levels of soluble CD23 (sCD23), sCD27, sCD30, C-reactive protein (CRP), κ and λ free light chains of immunoglobulin, and neopterin were determined by ELISA, following protocols from manufacturers. Described previously,[7] the total serum IgE level was measured by CIA-7.12 and CIA-4.15 monoclonal antibodies following published process with some modifications.[99, 100]

Table 2.2. Putative function, detection limit, and assay (manufacturer) for 32 inflammation-related serum markers

Inflammation-related serum markers	Putative Functions	Detection limits	Assay (Manufacturer)
Interleukins (IL)			
IL-1β	B-cell maturation and proliferation	0.06 pg/mL	Fluorokine® MAP Human Inflammation Kit (R&D Systems, MN)
IL-2	Ig production	0.02 pg/mL	
IL-4	B-cell stimulation, Ig expression/secretion	0.82 pg/mL	
IL-5	B-cell stimulation and Ig secretion	0.02 pg/mL	
IL-6	B-cell stimulation and maturation	0.06 pg/mL	
IL-10	Enhancement of B cell survival, proliferation, and Ig production	0.04 pg/mL	
IL-12	Immune system stimulation	0.12 pg/mL	
IL-17A	Stimulation of cytokines production	2.44 pg/mL	Procarta Cytokine Assay (Affymetrix, Santa Clara, CA)
IL-17F	Stimulation of TNFα, IL-1β production	0.16 pg/mL	
IL-20	B-cell stimulation (IL-10 subfamily)	0.08 pg/mL	
IL-23	Immune system stimulation	0.1 pg/mL	
IL-27	Induction of *IgH* CSR activity	2.44 pg/mL	
IL-31	Immune system stimulation	0.1 pg/mL	
ILF	B-cell differentiation and maturation	0.08 pg/mL	
CXC chemokine family			
CXCL13	B-cell homing regulation	7.8 pg/mL	Fluorokine® MAP Human Inflammation Kit (R&D Systems, MN)
IL-8 (CXCL8)	Immune system activation	0.58 pg/mL	
IP-10 (CXCL10)	Immune system activation	2.44 pg/mL	Procarta Cytokine Assay (Affymetrix, Santa Clara, CA)
SDF-1α (CXCL12)	B-cell homing regulation	2.70 pg/mL	
Cytokines with hormone-like functions			
GM-CSF	Growth factor for B cells	0.04 pg/mL	Fluorokine® MAP Human Inflammation Kit (R&D Systems, MN)
VEGF	Affect immune function through lymphangiogenesis	0.84 pg/mL	

(Table 2.2 continued)

Inflammation-related serum markers	Putative Functions	Detection limits	Assay (Manufacturer)
Interferons (IFN)			
IFNγ	Immune system activation	0.02 pg/mL	Fluorokine® MAP Human Inflammation Kit (R&D Systems, MN)
IgE-related markers			
IgE	Indication of *IgH* CSR activity	8 ng/mL	CIA-7.12 and CIA-4.15 monoclonal antibodies
sCD23	Induction of *IgH* CSR activity	0.1 pg/mL	ELISA (Bender MedSystems USA, CA)
Ig free light chains(FLC)			
κ	Indicator of B-cell hyper-activation	6 μg/L	ELISA (Biovendor, Modrice, Czech Republic)
λ	Indicator of B-cell hyper-activation	5 μg/L	
κ:λ	Indicator of B-cell clonal expansion	Sum of κ + λ	
Tumor Necrosis Factor (TNF) and TNF receptor superfamily			
TNFα	Immunostimulant	0.16 pg/mL	Fluorokine® MAP Human Inflammation Kit (R&D Systems, MN)
sCD27	Immune system activation	1 U/mL	PeliKine-compact ELISA Kit and Toolset (CLB/Sanquin, Netherlands)
sCD30	Immune system activation	6 units/mL	ELISA (Bender MedSystems USA, CA)
sCD40L	Immune system activation	1.62 pg/mL	Procarta Cytokine Assay (Affymetrix, Santa Clara, CA)
Others			
C-reactive protein (CRP)	Indication of pro-inflammatory cytokines activity	0.25 μg/mL	ELISA (Virgo CRP 150, Hemagen, MD, USA)
Neopterin	Indication of pro-inflammatory immune status; marker of immune activation	0.7 nmol/L	ELISA (IBL International GMBH, Hamburg, Germany)

ELISA enzyme-linked immunosorbent assay; *GM-CSF* granulocyte-macrophage colony-stimulating factor; *Ig* immunoglobulin; *IP-10* IFNγ-induced protein 10; *SDF* stromal cell-derived factor; *VEGF* vascular endothelial growth factor

2.6 SNP Selection and Genotype Determination

We selected SNPs on genes that are related to the AICDA pathway, B-cell activation, or immune stimulation. Table 2.3 lists 41 candidate genes. Details on selection process for SNPs at *CXCL13* and *CXCR5* have been published.[9] We applied similar method for SNPs at other 39 genes.

We selected candidate SNPs with minor allele frequency (MAF) of at least 5% from the International HapMap project (http://hapmap.ncbi.nlm.nih.gov) or the National Institute of Environmental Health Sciences (NIEHS) SNPs Program (http://egp.gs.washington.edu).[101] We used the Genome Variation Server (GVS, http://gvs.gs.washington.edu/GVS/) to group any two correlated SNPs on one gene (correlation coefficient $r^2 \geq 0.80$) into one bin.[102, 103] One SNP for each group was selected as a tagSNP, given that it showed high probability of functional changes (e.g. in exons or other functional areas) and high Illumina design scores.[9]

We used the high-throughput Illumina GoldenGate® assay (San Diego, CA) to determine genotypes of SNPs, according to manufacturer's protocol. DNA for genotyping was extracted from peripheral blood mononuclear cells using QIAmp DNA Blood Mini Kit (Qiagen, CA). The positive controls (DNA samples with known genotypes) and negative controls (reagent without DNA) were used in each genotyping plate. For quality control, we genotyped all SNPs in 44 duplicate study participants and got a duplicate concordance rate of 99.97%.

The selection and quality control process resulted in 329 SNPs. We removed 55 of 384 SNPs for analysis due to monomorphic genotype (n = 14), missing values in more than 85% study participants (n = 1), genotyping failure (n = 27), or violation of Hardy-Weinberg Equilibrium (*P*-value < 0.001 in HIV-positive controls, n = 13). The average call rate was 98.6% (94%-100%) in 329 SNPs.

Table 2.3. Listing of 41 genes related to *AICDA* pathway, DNA repair, or B-cell activation

AICDA pathway/ B-cell activation	Gene	Full name
AICDA pathway		
Transcriptional regulation for *AICDA*	*PAX5*	Paired box 5
	TNFSF13B	Tumor necrosis factor (ligand) superfamily, member 13b
	IL4	Interleukin 4
	IL21	Interleukin 21
	sCD40L	CD40 ligand
AICDA	*AICDA*	Activation-induced cytidine deaminase
DNA repair mechanism		
BER pathway	*UNG*	Uracil-DNA glycosylase
	APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1
	APEX2	APEX nuclease (apurinic/apyrimidinic endonuclease) 2
	FEN1	Flap structure-specific endonuclease 1
MMR pathway	*EXO1*	Exonuclease 1
	MLH1	MutL homolog 1, colon cancer, nonpolyposis type 2
	MLH3	MutL homolog 3
	MSH2	MutS homolog 2, colon cancer, nonpolyposis type 1
	MSH3	MutS homolog 3
	MSH6	MutS homolog 6
	PCNA	Proliferating cell nuclear antigen
	PMS1	PMS1 postmeiotic segregation increased 1
	PMS2	PMS2 postmeiotic segregation increased 2
	RPA1	Replication protein A1, 70kDa
	RPA2	Replication protein A2, 32kDa
	RPA3	Replication protein A3, 14kDa
DNA polymerases	*POLH*	Polymerase (DNA directed), eta
	POLI	Polymerase (DNA directed), iota
	POLK	Polymerase (DNA directed), kappa
	POLL	Polymerase (DNA directed), lambda
	POLM	Polymerase (DNA directed), mu
	POLQ	Polymerase (DNA directed), theta
NHEJ	*DCLRE1C*	DNA cross-link repair 1C
	XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5
	XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6

(Table 2.3, continued)

AICDA pathway/ B-cell activation	Gene	Full name
DNA Repair	*ATM*	Ataxia telangiectasia mutated
	H2AFX	H2A histone family, member X
	RAD52	RAD52 homolog (S. cerevisiae)
	REV1	REV1 homolog (S. cerevisiae)
	RE3L	REV3-like, catalytic subunit of DNA polymerase zeta (yeast)
	TP53BP1	Tumor protein p53 binding protein 1
	MGMT	O-6-methylguanine-DNA methyltransferase
B-cell activation chemokine signaling pathway	*CXCL13*	Chemokine (C-X-C motif) ligand 13
	CXCR5	Chemokine (C-X-C motif) receptor 5
	PRKACA	cAMP-dependent protein kinase catalytic subunit alpha

BER base excision repair; *MMR* mismatch repair; *NHEJ* non-homologus end joining

2.7 Statistical Analysis

The demographic distributions between AIDS-NHL cases and HIV-positive controls were described for the genetic and serum marker studies, respectively. The χ^2 test or Fisher's exact test was performed for categorical variables. For continuous variables, we presented median and interquartile ranges (IQR). If normal distribution assumption is not violated, we performed student's t test for mean values between cases and controls. We defined the reference date as the NHL diagnosis date for AIDS-NHL cases. For HIV-positive controls, the reference date was the date of first reporting HIV infection plus time period (days) between first known HIV-infected date and NHL diagnosis date of corresponding cases. Based on our study design and matching criteria, controls should have the same HIV infection period as their corresponding cases at the reference date.

Serum cytokines levels were natural logarithm-transformed and treated as continuous variables. The measurement values of serum cytokines below the detection limit were set to the value half of the detection limit. Some cytokines (IL-1β, IL-2, IL-4, IL-12, IL-17A, IL-23, IL-31, GM-CSF, and SDF-1α) were treated as binary variables given that too many participants showed undetectable levels. We did not observe a serious violation of normality assumption for all natural log-transformed serum levels. We analyzed genetic polymorphisms assuming additive genetic model. To identify potential biomarkers from hundreds of genetic and serum markers that are all correlated to the B-cell activation pathway, we applied weighted gene co-expression network analysis (WGCNA).[104] The visANT was used for network visualization. We performed the network analysis on the serum marker study as participants (179 AIDS-NHL cases and 179 controls) had both serum and genetic markers.

Weighted Gene Co-expression Network Analysis

We used the WGCNA as a screening tool to identify potential biomarkers that are important in the B-cell activation pathway for further adjustment in HIV and HCV models. Three steps are needed for the WGCNA.[104] First, we defined a power adjacency function, AF(S) = S^b, such that the network in our study approximates to a scale-free topology network with a soft threshold to connect two dots (biomarkers). Each biomarker corresponds to a dot (or node) in a network. The WGCNA defines AF(S) the absolute value of Pearson correlation coefficient between two markers. By raising the AF(S) to a power of b, any two dots connect to each other while the connectivity is penalized by a power of b. This power adjacency function forms a weighted network.[104] The soft thresholding power in our study was set to be 3, considering the trade-off between scale independence (given a scale-free topology model fit of $r^2 = 0.8$) and mean connectivity (the highest mean number of connections).[104] Figure 2.1 illustrates scale independence, as well as mean connectivity, by soft thresholding powers.

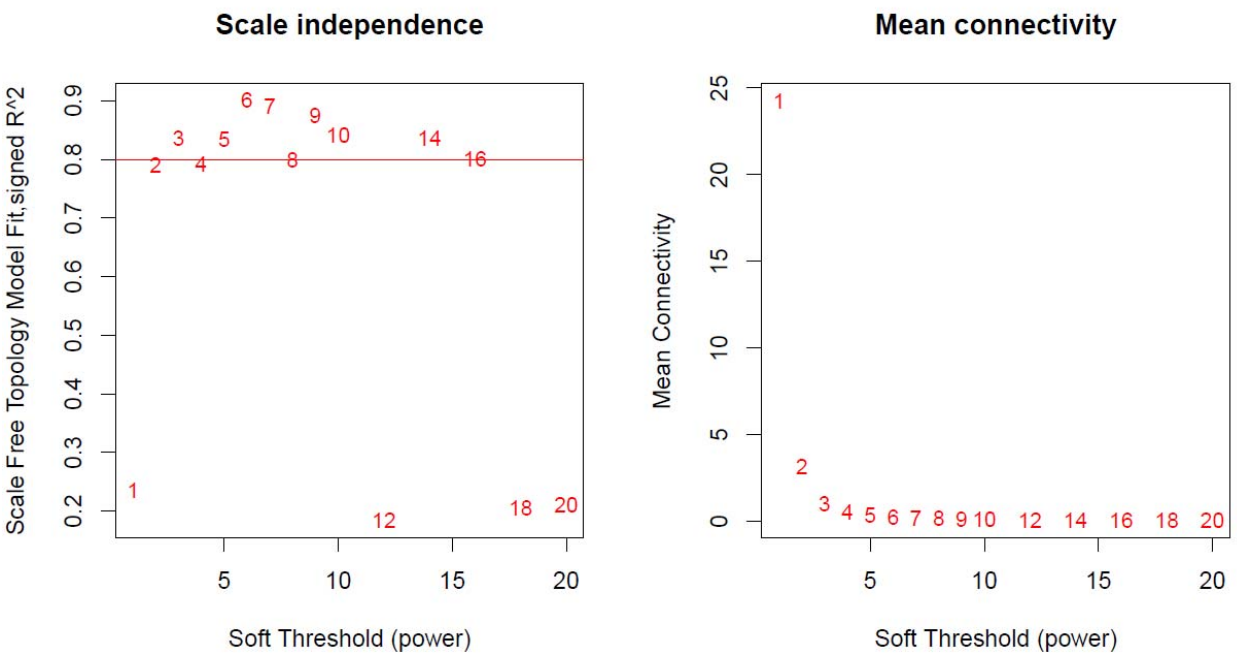


Figure 2. 1. Scale independence (left panel) and mean connectivity (right panel) under different soft thresholds for adjacency function in the serum marker study. The red line in the left panel indicates $r^2 = 0.8$. The soft thresholding of 3 would have a good fit of scale-free topology model (left) and similar mean connectivity under a soft thresholding of 2 (right).

Second, the WGCNA grouped biomarkers into several modules, using the average linkage hierarchical clustering. The WGCNA transforms adjacency into topological overlap matrix (TOM) to reduce noise and spurious associations.

$\text{TOM}_{ij} = \frac{\sum_u a_{iu}a_{uj}+a_{ij}}{\min(k_i,k_j)+1-a_{ij}}$, where $k_i = \sum_j a_{ij}$, is the connectivity, which equals to the sum of adjacency matrix of biomarkers i (a_{ij}).[104]

The WGCNA applies average linkage hierarchical clustering using dissimilarity of TOM (1-TOM). We performed network analysis on serum markers measured greater than 3 years prior to reference date to avoid potential reverse causality for serum markers measured at 1-3 or 0-1 years preceding the reference date. We set the minimum membership in each module to be 1, meaning that at least one marker is included per module. Table 2.4 shows preliminary results of WGCNA from the network containing 329 SNPs and 32 cytokines. Figure 2.2 illustrates the clustering dendrogram of 74 modules identified in the network. Each color represented one module. We used biomarkers identified from the network analysis for further adjustment in HIV and HCV models.

Table 2.4. Preliminary WGCNA results for the network containing SNPs and cytokines measured greater than 3 years prior to reference date in the serum marker association study, with soft-thresholding power of 3, minimum module size of 1.

	Biomarkers including SNPs and cytokines measured larger than 3 years preceding reference date
Study participants	347
Participants excluded[1]	11
Serum markers (cytokines) [2]	31
SNP[3]	328
Module number	74
Number of biomarkers in the largest module	10
Number of biomarkers in the smallest module	2
Number of biomarkers in the grey module[4]	44

[1]Participants were excluded due to outliers (n = 3 in all three time points) or having too many miss values (n = 11), according to the criteria in WGCNA R Package.
[2]The SDF-1α was excluded.
[3]*POLL* rs4919556 was excluded.
[4]Biomarkers that could not be grouped into any module were clustered into the grey module.

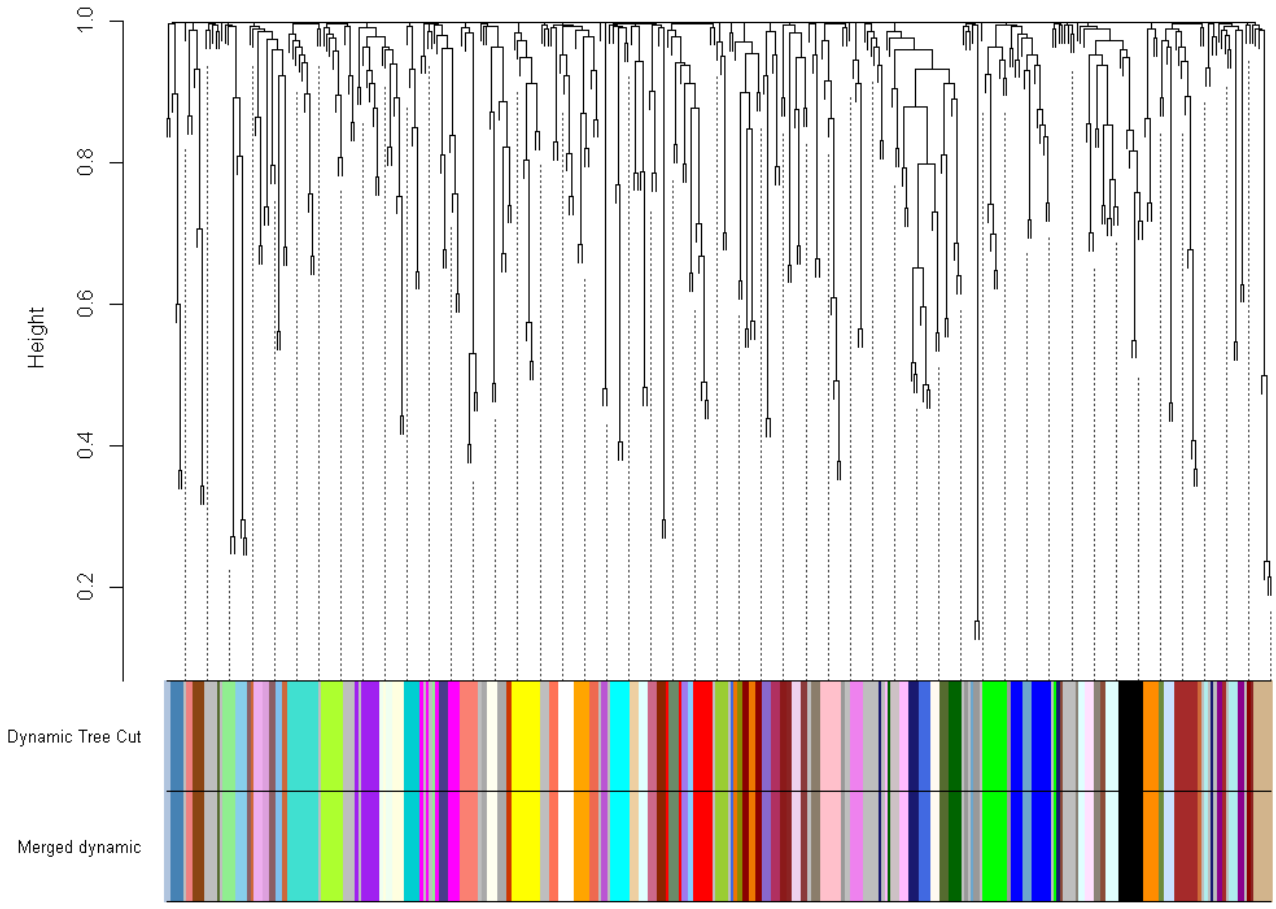


Figure 2.2. Clustering dendrogram of 74 modules in the serum marker study, containing 31 serum cytokines measured longer than 3 years prior to reference date and 328 SNPs. The WGCNA used dissimilarity of topological overlap matrix (TOM) to perform the average linkage hierarchical clustering.

Third, the WGCNA identified module-specific hub biomarkers. The WGCNA uses the module eigenvector (ME) from the first principal component to represent the module. It calculates Pearson correlation coefficients between ME and outcome (AIDS-NHL case/control status), as well as the Pearson correlation of ME with every node (biomarkers) within the module. We selected top six modules showing high correlations between ME and AIDS-NHL status (*P*-values < 0.05) for further network analysis, as listed in Table 2.5.

We identified members (serum or genetic markers) in the first module in Table 2.5 whose correlations with the ME were larger than 0.6. The HIV and HCV models included these biomarkers for B-cell activation adjustment.

Table 2.5. Pearson correlation coefficients of module eigenvector with the AIDS-NHL case/control status from the WGCNA

Module	Serum cytokines measured > 3 years preceding reference date	
	r for ME and AIDS-NHL risk[1]	*P*-value
1	0.335	$1.51*10^{-10}$
2	0.135	0.0117
3	-0.132	0.0139
4	0.122	0.0236
5	0.115	0.0329
6	0.112	0.0376

[1]Pearson correlation coefficients between module eigenvector and case/control status with *P*-values less than 0.05

Network visualization: visANT

We used the visANT, a free online source (http://visant.bu.edu/) developed at Boston University, to visualize networks that contains roughly 500 reciprocal associations.[105] Based on their correlations with the outcome (case/control status), we chose first 31 markers (18 cytokines, 13 SNPs) to capture influential biomarkers. There were (31*30)/2 = 465 reciprocal associations. visANT could visualize a network using several entries of association, such as topological overlap, correlation, adjacency, or connectivity (number of interactions). We used the topological overlap measure (TOM) and input 465 reciprocal TOMs into the visANT. Any two nodes are connected in the visANT if their association measurements are larger than a hard threshold. Since there is no common cut-off to connect two nodes, our study explored and illustrated several visANT networks by different thresholds.

HIV Models

Distribution of HIV viral load measurement times was described in both study populations. We used VL denote $\log_{10}$ transformed HIV RNA level (copies/mL): VL ≡ HIV viral load $\log_{10}$ copies per mL. Three approaches were explored for HIV viremia in HIV models: (1) VL at set point, (2) pre-HAART VL prior and closest to the reference date, (3) the slope and intercept for linear regressions of VL by year in participants who have more than 2 HIV RNA measurements preceding reference date. The intercept is on $\log_{10}$ scale and the slope represented annual change of VL (ΔVL/year).

The virologic set point of HIV infection is a steady-state of HIV viral load prior to any antiretroviral therapy, indicating a balance between HIV viral replication and its clearance.[106, 107] In the first several weeks of primary infection, the HIV viral load could burst as high as 10^7 copies/mL in blood since there are numerous susceptible T cells and antiviral immune responses have not yet fully reacted.[108] Following a period of fluctuation, usually several months, the HIV viral load stabilizes and reaches a steady-state or the "set point". The period from HIV sero-conversion to the steady-state varies across individuals. Sero-converters and seroprevalent MACS participants followed different algorithms for viral load estimation. For sero-converters, the HIV viral-load set point referred to the mean HIV RNA level obtained 12 to 24.5 months after sero-conversion such that the interim slope of viral load by time approximated zero.[109, 110] For sero-prevalent men, HIV viral-load set point was approximated from study visit 3 or 4 to avoid potential viral load measurement error in the early phase of MACS study.[109, 110]

Conditional logistic regression was used to calculate odds ratio (OR) and 95% confidence intervals (CIs) relating HIV viral load to AIDS-NHL risk. Multivariable regression models

included $CD4^+$ T-cell count (continuous), age at reference (continuous), prior AIDS diagnosis (yes/no), HCV infection (yes/no), smoking (ever/never), recreational amphetamine use (ever/never),[111] and time receiving HAART (years) and ART (years). Some covariates were not included in the serum marker study due to a relatively small sample size and missing data.

HCV Models

According to the HCV testing algorithm in the MACS, the HCV status was measured retrospectively using archival serum samples from the baseline visit until 2001, and prospectively every six-month since 2001 (Multicenter AIDS Cohort Study, Data Working Group Meeting, CAMACS Report, October 8, 2010).

HCV status was determined by both testing results: the antibody to HCV (anti-HCV) using enzyme immunoassay (ADVIA Centaur HCV assay), and the serum HCV RNA levels by PCR (COBAS AmpliPrep TaqMan HCV Assay).[112] A participant was considered HCV negative if all available HCV antibody tests prior to NHL diagnosis date were negative. A participant was considered to have cleared his HCV infection if he had a negative HCV RNA test result prior to reference date preceded by (1) at least two positive anti-HCV test results, or (2) at least two results of detectable HCV RNA, or (3) a detectable HCV RNA level at study baseline. A participant was considered chronically HCV infected if he had a positive HCV RNA test result at the reference date preceded by either a positive anti-HCV or HCV RNA result. At baseline visit, 38 (0.6%) of 6,972 participants had no HCV status: 19 (0.3%) did not receive any HCV test, and 19 had inconclusive results.

Conditional logistic regression was used to calculate OR (95% CI) relating HCV infection (ever/never) and AIDS-NHL susceptibility. We also explored associations between

chronic HCV infection (ever/never) and AIDS-NHL risk. Multivariable models included $CD4^+$ T-cell count (continuous), age at reference date (continuous), prior AIDS diagnosis (yes/no), VL at set point (continuous), smoking (yes/no), and time receiving HAART (years) and ART (years).

Adjustment for B-cell Activation Markers

HIV and HCV models additionally included SNPs data in the genetic association study. We observed 20 of 329 SNPs showing associations with AIDS-NHL in preliminary analysis. These SNPs were candidates to be included in HIV and HCV models for further B-cell activation adjustment. Table 2.6 lists these 20 SNPs and crude associations. No published studies focus on associations between these SNPs and AIDS-NHL risk. We performed stepwise significance-test on confounder effect to select SNPs into our HIV or HCV models with α level of 0.20 (enter threshold) and 0.25 (stay threshold).[113]

HIV and HCV models included the hub biomarkers identified from network analysis for B-cell activation adjustment in the serum marker study.

Table 2.6. Candidate SNPs for adjustments in HIV or HCV models

Number	SNP	Gene	Allele change	Minor allele frequency	Crude OR[1] (95% CI)
1	rs1760945	*OSGEP (APEX)*	C→T	11.55	0.53 (0.34-0.83)
2	rs28382694	*APEX2*	C→A	21.13	3.15 (1.33-7.45)
3	rs1800889	*ATM*	C→T	5.37	0.43 (0.22-0.85)
4	rs611646	*ATM*	A→T	41.63	0.72 (0.57-0.91)
5	rs851777	*EXO1*	C→T	16.22	1.45 (1.03-2.06)
6	rs1784304	*HMBS (H2AFX)*	C→A	24.75	0.76 (0.56-1.03)
7	rs1800734	*MLH1*	G→A	23.26	0.75 (0.55-1.02)
8	rs1382539	*MSH3*	G→A	28.43	0.76 (0.58-1.00)
9	rs1677645	*MSH3*	T→C	28.94	0.77 (0.59-1.00)
10	rs1805355	*MSH3*	G→A	8.84	1.54 (1.04-2.29)
11	rs979182	*POLK*	C→T	10.11	1.45 (0.98-2.14)
12	rs4919556	*POLL*	T→C	48.36	1.24 (1.01-1.52)
13	rs7778745	*POLM*	A→G	23.91	0.71 (0.52-0.96)
14	rs2582869	*TNFSF13B*	G→A	36.90	1.32 (1.04-1.67)
15	rs1041569	*TNFSF13B*	A→T	22.80	1.34 (1.05-1.71)
16	rs1224166	*TNFSF13B*	C→T	11.90	1.65 (1.17-2.32)
17	rs1059262	*ALKBH2*	T→G	18.33	0.65 (0.46-0.90)
18	rs668844	*XRCC5*	C→A	45.58	0.80 (0.64-1.00)
19	rs41296769	*XRCC5*	C→G	10.35	9.0 (1.82-44.59)
20	rs207884	*XRCC5*	C→G	37.14	1.37 (1.07-1.75)

SNP single nucleotide polymorphism; *cOR* crude odds ratio

[1]OR relating SNPs (additive genetic model) to AIDS-NHL susceptibility. SNPs showed borderline associations with AIDS-NHL were also included.

Multiple Imputation for Missing Data

We used multiple imputations to impute missing values of covariates in HIV and HCV models.[114, 115] The imputed values were based on the observed data and the missing-data pattern, assuming a missing at random.[115] We assumed that the probability of a variable that is missing depends only on the observed data, and there are no unmeasurable common causes of the missing data and NHL risk in our study.[116] We ran 10 rounds of multiple imputation to predict missing covariates values. The final imputed values were summarized from the 10 imputed results. We included every covariate in HIV or HCV models (viral-load set point, $CD4^+$ T-cell count at reference date, prior AIDS diagnosis, HCV infection, age, HIV infection year, smoking status, recreational amphetamine use, time in years from the first HAART or ART use to reference date). For the genetic association study, 20 candidate SNPs were also included. For the serum marker study, biomarkers suggested by the WGCNA (sCD27, sCD30, IP-10, CXCL13, neopterin) were included. The case/control status was not included in the imputation.

We used SAS 9.1v (SAS Institute, Cary, NC) for data analyses and multiple imputation, the comprehensive package for WGCNA in R environment, and the visANT program for network visualization.

CHAPTER 3: RESULTS

3.1 Baseline Demographics

Table 3.1 shows demographic factors distributions between cases and controls from two studies. Cases had larger median age at reference date than controls in both studies and *P*-values from Student's t test were both smaller than 0.05. Both studies showed different ethnic distributions (*P*-values for Fisher's exact test: 0.022 [genetic study], 0.049 [serum marker study]). Cases consisted of less non-Hispanic Whites and more Hispanic Whites than did controls in both studies. We did not observe different HIV status (sero-prevalent and converter) at diagnosis nor distinct years of known HIV infection in both studies (median: 6.8 years [genetic study], 6.9 years [serum marker study]). These were expected as HIV infection duration was one of the matching factors in both studies.

Less than 10% of participants ever received HAART while most participants were under ART prior to reference date. Cases had longer period under the HAART in the genetic study (3.34 [1.52-5.57] *vs.* 1.77 years [0.57-4.32]) and shorter period under the HAART in the serum study (1.93 [1.01-3.50] *vs.* 2.84 years [0.40-3.53]) than controls. The HAART receiving cases were not fully overlapped in two studies: 7 HAART recipients who later developed NHL were included in both studies. Five of 12 and 2 of 9 cases who ever received HAART were specific to the genetic and serum association studies, respectively. Conversely, both studies showed that cases had longer period between first ART date to reference date than controls (*P*-values were both larger than 0.05) although it appeared that more controls in the genetic study received ART (62 *vs.* 71%, *P*-value = 0.018).

$CD4^+$ T-cell count in category was similar between groups in the genetic association study as it was a matching factor (median cells/mm^3 in cases *vs.* controls: 79 [IQR: 23-293] *vs.* 87 [IQR: 28-270]). CD4 slope in the pre-HAART era decreased more rapidly among cases in both studies.

More cases in both studies than controls had prior AIDS diagnosis other than lymphoma (55 *vs.* 42%, $P = 0.003$ and 55 *vs.* 11%, $P < 0.001$, for genetic and serum studies, respectively). Among these cases in both studies, pneumocystis pneumonia (PCP) and Kaposi sarcoma (KS) accounted for one-third and 22% of the first AIDS diagnosis, respectively. In controls, 42% (n = 224) in the genetic study and 11% (n = 20) in the serum marker study had AIDS diagnosis other than lymphoma preceding reference date. The PCP and KS were also two most prevalent diagnoses in controls who had prior AIDS diagnosis. The distributions of lymphoma subtypes were similar in two case groups as they were highly overlapped. Thirty percent of cases had primary central nervous system lymphoma. Diffuse large B-cell lymphoma accounted for more than half of the systemic lymphoma in both studies.

Table 3.1 also shows distributions of smoking, drinking, and recreational amphetamine use. In the genetic association study, 68% of controls ever smoked with a median cumulative smoking pack-year of 17.0 (IQR: 3.1-32.3), higher than 67% of cases who ever smoked with 12.6 (IQR: 3.0-30.9) pack-years. In the serum marker study, more controls than cases were ever smokers (75 vs. 64%, $P = 0.049$) with higher cumulative pack-years (median: 18.6 [IQR: 2.9-34.5] vs. 11.2 [IQR: 2.8-30.8] pack-years). The drinking status was similar between groups – only less than 4% of cases and controls in both studies were non-drinkers. The recreational

amphetamine use was also high in these two populations. More than 90% of participants in both studies used them at least once.

Table 3.1. Distribution of demographic factors of cases and controls from genetic (n = 716) and serum marker (n = 358) association study

	Genetic association study (n = 716)			Serum marker association study (n = 358)		
	AIDS-NHL cases	HIV+ controls	*P*-value	AIDS-NHL cases	HIV+ controls	P-value
Participants, n	183	533		179	179	
Median age at time of NHL diagnosis, years (IQR)	41.8 (36.1, 46.5)	40.1 (35.4, 44.7)	0.046[1]	41.4 (35.8, 46.1)	39.7 (35.4, 43.5)	0.006[1]
Ethnicity, n (%)			0.022[2]			0.049[2]
White, non-Hispanic	153 (83)	483 (91)		149 (83)	156 (87)	
White, Hispanic	18 (10)	25 (4)		19 (11)	8 (4)	
Black, non-Hispanic	12 (7)	21 (4)		11 (6)	12 (7)	
Other	0 (0)	4 (1)		0	3 (2)	
MACS cohort, n (%)			0.761			0.737
1984 recruits	163 (89)	479 (90)		158 (88)	160 (89)	
1987 recruits	20 (11)	54 (10)		21 (12)	19 (11)	
HIV status, n (%)			0.599			1.000
Sero-prevalent[3]	163 (89)	482 (90)		158 (88)	158 (88)	
Converter[4]	20 (11)	51 (10)		21 (12)	21 (12)	
Years of known HIV infection period prior to ref. date, median (IQR)	6.8 (4.7, 9.6)	6.8 (4.7, 9.7)	0.937[1]	6.7 (4.6, 9.3)	6.7 (4.6, 9.3)	1.000[1]

IQR interquartile range; *ref.* reference

[1] Student's t-test for means; [2] Fisher's exact test;

[3] MACS participants who were HIV+ at entry; [4] MACS participants who were HIV- at entry and became HIV+ during follow-up;

(*Table 3.1, continued)*

	Genetic association study (n = 716)			Serum marker association study (n = 358)		
	AIDS-NHL cases	HIV+ controls	*P*-value	AIDS-NHL cases	HIV+ controls	*P*-value
HAART prior to ref. date, n (%)			0.258			1.000
Never	171 (94)	484 (91)		170 (95)	170 (95)	
Yes	12 (6)	49 (9)		9 (5)	9 (5)	
Years from first HAART date to ref. date, median (IQR)	3.34 (1.52, 5.57)	1.77 (0.57, 4.32)	0.285[1]	1.93 (1.01, 3.50)	2.84 (0.40, 3.53)	0.701[1]
ART prior to ref. date, n (%)			0.018			0.134
Never	71 (38)	156 (29)		68 (38)	82 (46)	
Yes	112 (62)	377 (71)		111 (62)	97 (54)	
Years from first ART date to ref. date, median (IQR)	3.2 (1.6, 4.7)	2.5 (1.2, 4.6)	0.562[1]	3.1 (1.5, 4.5)	2.4 (1.2, 3.7)	0.181[1]
CD4 slope pre-HAART, median (IQR)[5]	-67.2 (-108.4, -29.0)	-59.6 (-95.8, -34.6)		-73.6 (-108.7, -34.7)	-43.6 (-70.7, -17.4)	
missing, n (%)	23 (13)	4 (1)		21 (12)	5 (3)	
$CD4^+$ T-cell count at matching[5], median (IQR), cells/mm^3	79 (23, 293)	87 (28, 270)				
$CD4^+$ T-cell count (cells/mm^3) at matching, n (%)						
0 - 49	74 (41)	214 (40)				
50 - 99	25 (14)	74 (14)				
100 - 199	28 (15)	83 (16)				
200 - 349	22 (12)	66 (12)				
350 - 499	13 (7)	39 (7)				
>500	21 (11)	57 (11)				

HAART highly active anti-retroviral therapy; *ref.* reference; *ART* anti-retroviral therapy; *IQR* interquartile range;
[1] Student's t-test for means; [5] Not normal distribution

(*Table 3.1, continued)*

	Genetic association study (n = 716)			Serum marker association study (n = 358)		
	AIDS-NHL cases	HIV+ controls	*P*-value	AIDS-NHL cases	HIV+ controls	*P*-value
First AIDS diagnosis prior to ref. date, n (%)						
No	83 (45)	309 (58)	0.003	80 (45)	159 (89)	<.001
Yes	100 (55)	224 (42)		99 (55)	20 (11)	
Kaposi sarcoma (KS)	22 (22)	46 (21)	0.294[2]	22 (22)	3 (15)	0.164[2]
Pneumocystis pneumonia	35 (34)	83 (37)		33 (33)	9 (45)	
Cytomegalovirus infection	7 (7)	8 (4)		7 (7)	1 (5)	
Dementia	2 (2)	13 (6)		2 (2)	3 (15)	
Candida esophagitis	13 (13)	18 (8)		11 (11)	1 (5)	
Others	23 (22)	54 (24)		24 (24)	3 (15)	
NHL subtypes, n (%)						
PCNSL	56 (31)			57 (32)		
Systemic NHL	127 (69)			122 (68)		
DLBCL	70 (55)			65 (53)		
BL, BL-like	19 (15)			19 (16)		
Other/Non-specified	38 (30)			38 (31)		
Recreational amphetamine use, n (%)			0.439			0.389
Never	9 (5)	34 (7)		9 (5)	6 (3)	
Ever	176 (95)	495 (93)		164 (92)	173 (97)	
Smoking, n (%)			0.627			0.049
Never	62 (33)	167 (32)		59 (33)	44 (25)	
Ever	123 (67)	362 (68)		114 (64)	135 (75)	
Cumulative pack-years[5], median (IQR)	12.6 (3.0, 30.9)	17.0 (3.1, 32.3)		11.2 (2.8, 30.8)	18.6 (2.9, 34.5)	
Drinking, n (%)			0.823			0.516
Never	7 (4)	22 (4)		7 (4)	5 (3)	
Ever	178 (96)	507 (96)		166 (93)	174 (97)	

PCNSL primary central nervous system lymphoma; *DLBCL* diffuse large B-cell lymphoma; *BL* Burkitt's lymphoma; *IQR* interquartile range; [5] Not normal distribution

Table 3.2 lists the distribution of HIV viral load measurements. In the genetic association study, 5% of all participants did not have any HIV RNA measurement and 23% of cases and 20% of controls had only 1 measurement of HIV viral load prior to reference date. These participants with one or none measurement were excluded in HIV models that explored individual slopes and intercepts for linear regressions of VL by time. We presented the distribution of HIV VL among participants with one measurement. Cases had higher median viral load ($\log_{10}$ copies/mL) than controls (5.02 [4.25, 5.39] *vs.* 4.57 [4.07, 5.05]). More cases (52%) than controls (28%) had VL larger than 5 (100,000 HIV RNA copies/mL) in the genetic association study. The *P*-value for viral load in category between groups was 0.047. In the serum marker study, 22% of AIDS-NHL cases and 8% of HIV-positive controls had only one HIV RNA measurement (*P*-value < 0.001). The median viral load ($\log_{10}$ copies/mL) was 5.04 (4.55-5.42) in cases, larger than that in controls (4.48, 3.96-4.92). Among participants with one measurement, 56% of AIDS-NHL cases and 21% of controls had $\log_{10}$ RNA copies/mL larger than 5. More controls (50%) than cases (31%) had VL between 4 and 5.

In the genetic association study, 72% of cases and 74% of controls had at least 2 measurements of HIV RNA level ($P = 0.394$). In the serum marker study, more controls (87%) than cases (69%) had at least 2 HIV RNA measurements ($P < 0.001$). The median measurement counts was 4 (IQR: 2-8) and 14 (IQR: 4-34) for cases and controls who had at least 2 HIV measurements.

Table 3.2. Distribution of HIV RNA measurements during MACS follow-up

HIV RNA measurements	Genetic association study (n = 716)			Serum marker study (n = 358)		
	AIDS-NHL cases (n=183)	HIV+ controls (n=533)	*P*-value	AIDS-NHL cases (n = 179)	HIV+ controls (n = 179)	*P*-value
• Participants with at least 2 measurements, n (%)	132 (72)	396 (74)	0.394	124 (69)	155 (87)	<.001
Median cumulative times of measurement, (IQR)	5 (3, 8)	5 (3, 10)		4 (2, 8)	14 (4, 34)	
> 2 measurements	94 (51)	306 (58)	0.108	86 (48)	132 (74)	<.001
2 measurements	38 (28)	90 (17)	0.278	38 (21)	23 (13)	0.034
• One measurement, n (%)	42 (23)	106 (20)	0.431	39 (22)	14 (8)	<.001
VL[1], median (IQR)	5.02 (4.25, 5.39)	4.57 (4.07, 5.05)		5.04 (4.55, 5.42)	4.48 (3.96, 4.92)	
VL[1], n (%)			0.047[2]			0.024[2]
< 3	2 (5)	6 (6)		0	2 (14)	
3 - 4	6 (14)	18 (17)		5 (13)	2 (14)	
4 - 5	12 (29)	52 (49)		12 (31)	7 (50)	
> 5	22 (52)	30 (28)		22 (56)	3 (21)	
• None measurement, n (%)	9 (5)	29 (5)	0.431	16 (9)	10 (5)	<.001

IQR interquartile range; *VL* viral load
[1]VL: $\log_{10}$-HIV RNA level; [2] Fisher's exact test

3.2 Network Analysis

We preformed the weighted gene co-expression network analysis using the serum marker association study (n = 358). Table 2.5 lists 6 modules from network analysis that the ME correlated with the AIDS-NHL case/control status at an alpha level of 0.05. Table 3.3 further lists biomarkers within each module. In the first module, its ME was correlated to the NHL status with a correlation coefficient of 0.34 (*P*-value = $1.51*10^{-10}$, Table 2.5). Within the first module, cytokines with a high to low correlations to ME included sCD27 (r = 0.93), sCD30 (r = 0.85), IP-10 (r = 0.72), CXCL13 (r = 0.71), and neopterin (r = 0.68). These cytokines all had correlation coefficients to ME larger than 0.6. Other cytokines in the first module included sCD23 (r = 0.48), IL-6 (r = 0.40), TNFα (r = 0.38), IL-10 (r = 0.27), and CRP (r = 0.21). Biomarkers in other modules were SNPs on the same gene, including *XRCC5*, *TP53BP1*, *MSH3*, and *VDR*.

Table 3.3. Listing of module-specific biomarkers measured larger than 3 year prior to reference date, by correlation coefficient with its module eigenvector (ME)

Module[1]	ME correlation with NHL[2]	Genetic or serum markers within module	
		Correlation with ME ≥ 0.6	Correlation with ME < 0.6
1	0.34	sCD27, sCD30, IP-10, CXCL13, Neopterin,	sCD23, IL-6, TNFα, IL-10, CRP
2	0.14	*XRCC5* rs41296769, *XRCC5* rs41257924	*CXCR5* rs2230321
3	-0.13	*TP53BP1* rs2602141, *TP53BP1* rs689647, *TP53BP1* rs2242069	
4	0.12	*MSH3* rs1805355, *MSH3* rs1592878, *MSH3* rs6151935	
5	0.12	*VDR* rs1544410, *VDR* rs739837	*VDR* rs3847987
6	0.11	*MSH3* rs40139 *MSH3* rs33013 *MSH3* rs32985	*MSH3* rs6151662

[1] Total modules = 74;
[2] Significant Pearson correlation coefficients of module eigenvector (ME) with NHL status at α = 0.05;

3.3 Network Visualization

We used the VisANT to visualize a network that included SNPs and serum cytokines measured longer than 3 years preceding reference date. This time-point was chosen to consider the lag time of lymphomagenesis and avoid potential reverse causality.

Table 3.7 lists 31 biomarkers in the visANT including 18 inflammation-related cytokines and 13 SNPs. Their correlations with the AIDS-NHL status (*P*-values) and assigned modules by the WGCNA are also shown. All 10 serum markers in the first WGCNA module were selected according to the criteria. Some SNPs in modules in Table 3.6 were also selected into the visANT, including *XRCC5* rs41296769 (module 2), *TP53BP1* rs2602141 and rs689647 (module 3), *MSH3* rs1805355 and rs1592878 (module 4), *VDR* rs1544410 (module 5), and *MSH3* rs40139 and rs33013 (module 6).

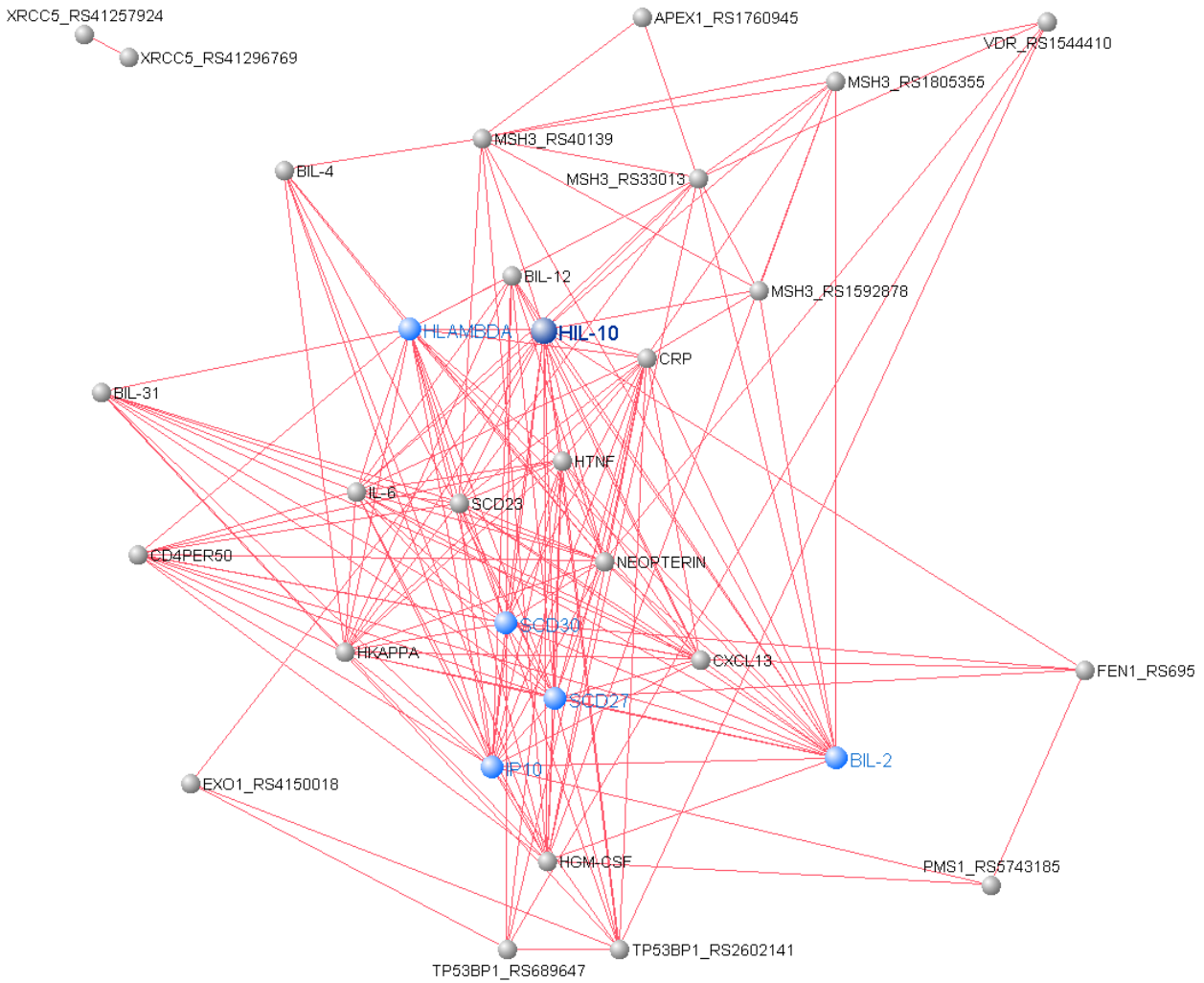
Table 3.4. Listing of markers for network visualization (13 SNPs and 18 serum cytokines measured longer than 3 year preceding reference date) and their correlations with the NHL status

Biomarkers	WGCNA module[1]	Pearson correlation coefficient	*P*-value
IP-10	1	0.372	$7.5*10^{-13}$
CXCL13	1	0.372	$8.0*10^{-13}$
CD4 slope	12	-0.366	$1.8*10^{-12}$
Neopterin	1	0.309	$4.1*10^{-9}$
IL-6	1	0.304	$7.8*10^{-9}$
sCD30	1	0.287	$5.3*10^{-8}$
sCD27	1	0.278	$1.5*10^{-7}$
λ	61	0.252	$1.9*10^{-6}$
IL-10	1	0.245	$4.1*10^{-6}$
κ	61	0.219	$3.9*10^{-5}$
TNFα	1	0.206	0.0001
GM-CSF	37	0.202	0.0002
sCD23	1	0.189	0.0004
CRP	1	0.171	0.0014
MSH3 rs1805355	4	0.166	0.0020
IL-2	15	0.152	0.0044
VDR rs1544410	5	0.152	0.0045
TP53BP1 rs2602141	3	-0.150	0.0050
IL-31	36	0.139	0.0097
XRCC5 rs41296769	2	0.133	0.0134
MSH3 rs40139	6	0.127	0.0177
EXO1 rs4150018	11	-0.125	0.0196
XRCC5 rs41257924	2	0.121	0.0241
IL-12	15	-0.118	0.0280
FEN1 rs695867	15	-0.118	0.0286
IL-4	15	-0.117	0.0291
PMS1 rs5743185	15	-0.117	0.0293
MSH3 rs1592878	4	0.116	0.0305
APEX1 rs1760945	8	-0.116	0.0309
TP53BP1 rs689647	3	-0.115	0.0326
MSH3 rs33013	6	0.114	0.0334

[1] First 6 WGCNA modules are shown in Table 3.3. Other modules are not shown as their ME have weaker correlations with the case/control status with *P*-values larger than 0.05.

Figure 3.1 to Figure 3.5 illustrate our targeting network with several cut-off points to connect two nodes, from TOM ≥ 0.001 (Figure 3.1) to TOM ≥ 0.1 (Figure 3.5) using visANT. Biomarkers that had highest links to other markers are also listed. Serum inflammation-related molecules suggested by the visANT were similar to those in the first WGCNA module (Table 3.3). sCD27, sCD30, IP-10, IL-6, CXCL13, and neopterin were molecules identified in both visANT and WGCNA. These molecules showed most links with other biomarkers given different threshold of TOM in the visANT, and showed high correlations with the module eigenvector ($r > 0.6$) in the WGCNA.

Setting the TOM to a level larger than 0.005, the visANT illustrated that SNPs on *XRCC5*, *MSH3*, or *TP53BP1* connected to other SNPs on the same gene.



Navy blue: 21 links (n = 1)

IL-10

Light blue: 18-19 links (n = 5)

sCD30, sCD27, IP-10,

IL-2, λ.

Figure 3.1. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.001. IL-10 had most links (21 links) to other biomarkers, followed by sCD30, sCD27, IP-10, IL-2, and λ free light chain.

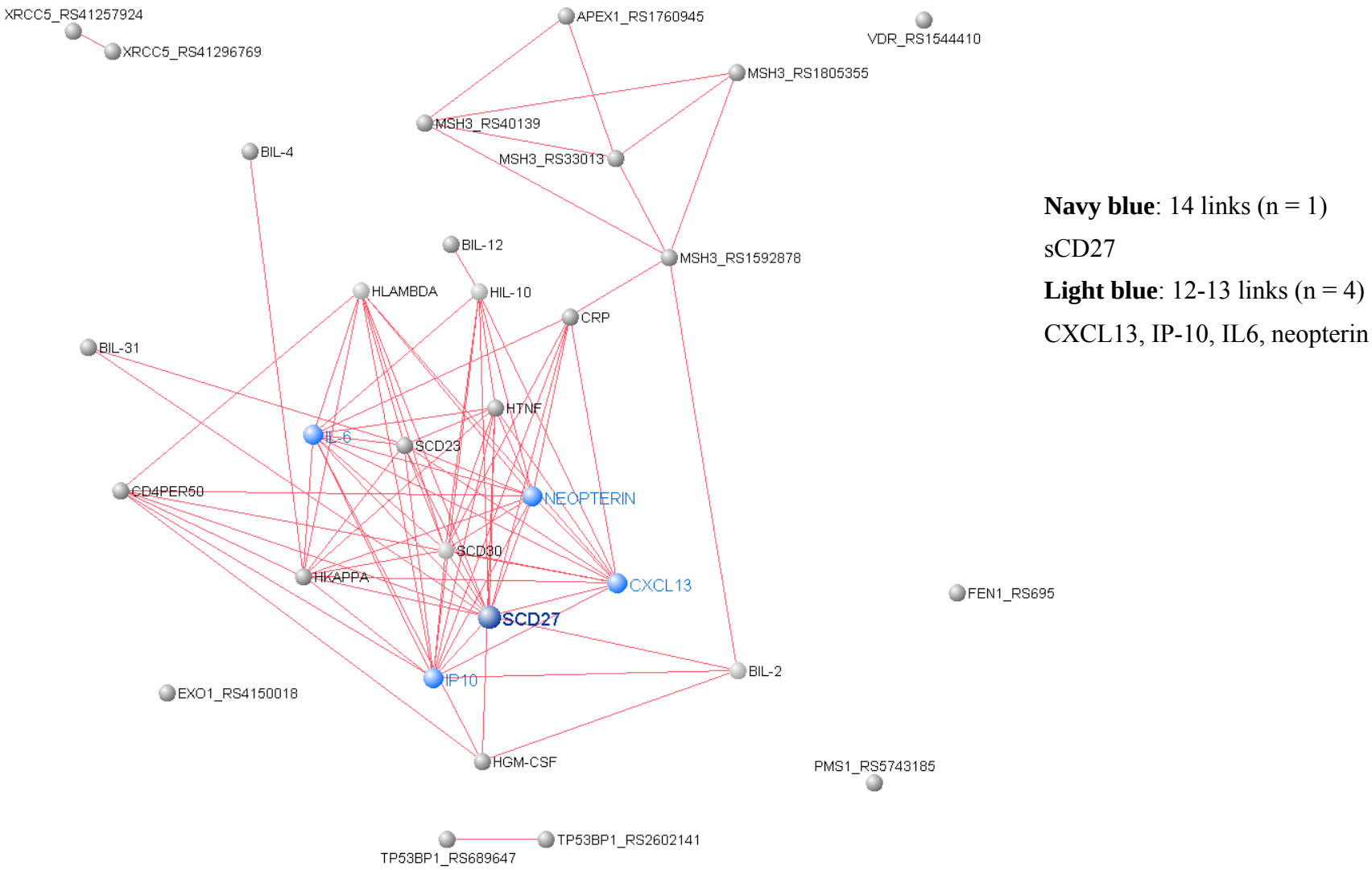
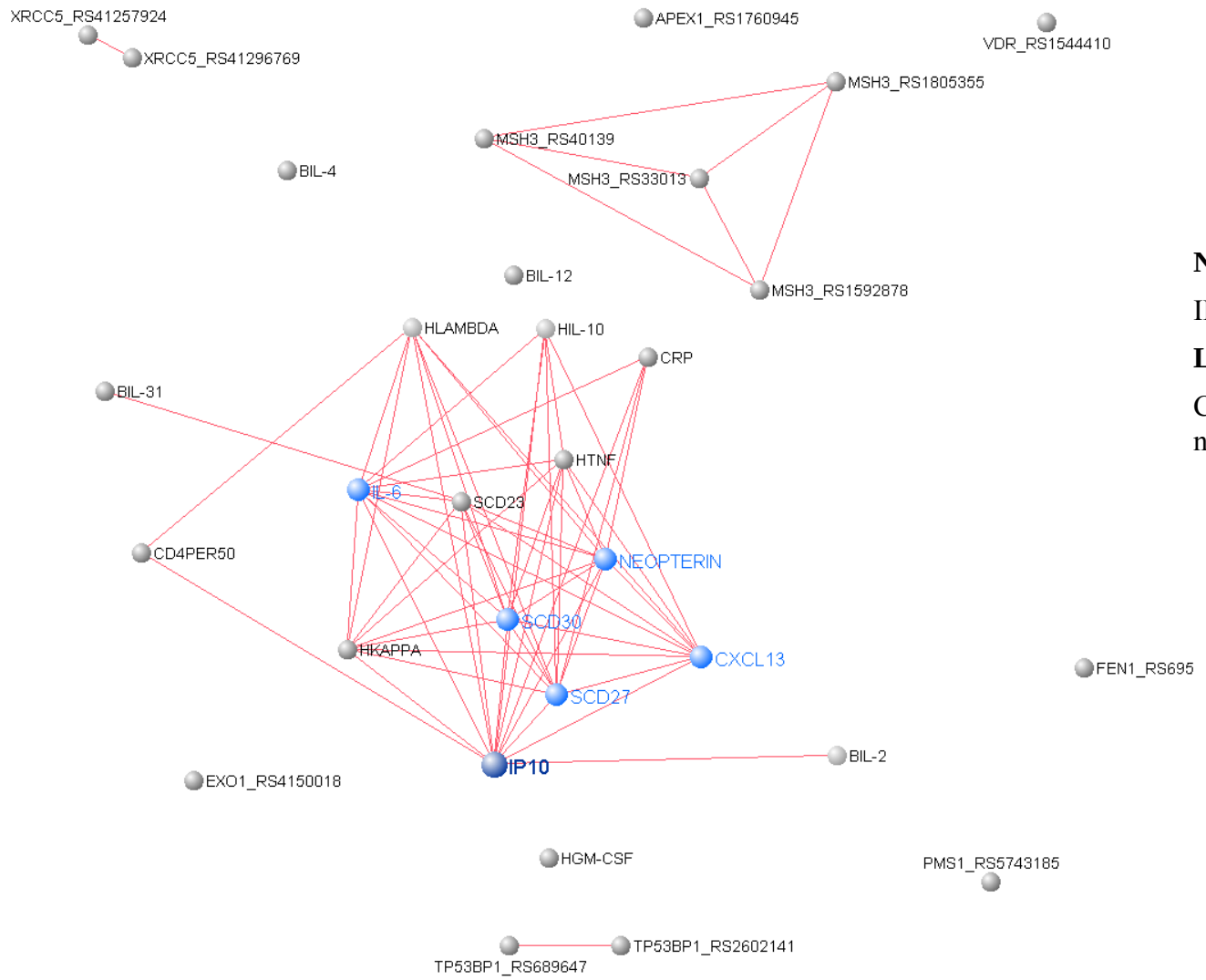


Figure 3.2. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.005. sCD27 had most links (14 links) to other biomarkers, followed by CXCL13, IP-10, IL-6, and neopterin.



Navy blue: 13 links (n = 1)

IP-10

Light blue: 10-11 links (n = 5)

CXCL13, IL6, sCD27, sCD30, neopterin,

Figure 3.3. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.01. IP-10 had the most links (13 links) to other biomarkers, followed by CXCL13, IL-6, sCD27, sCD30, and neopterin.

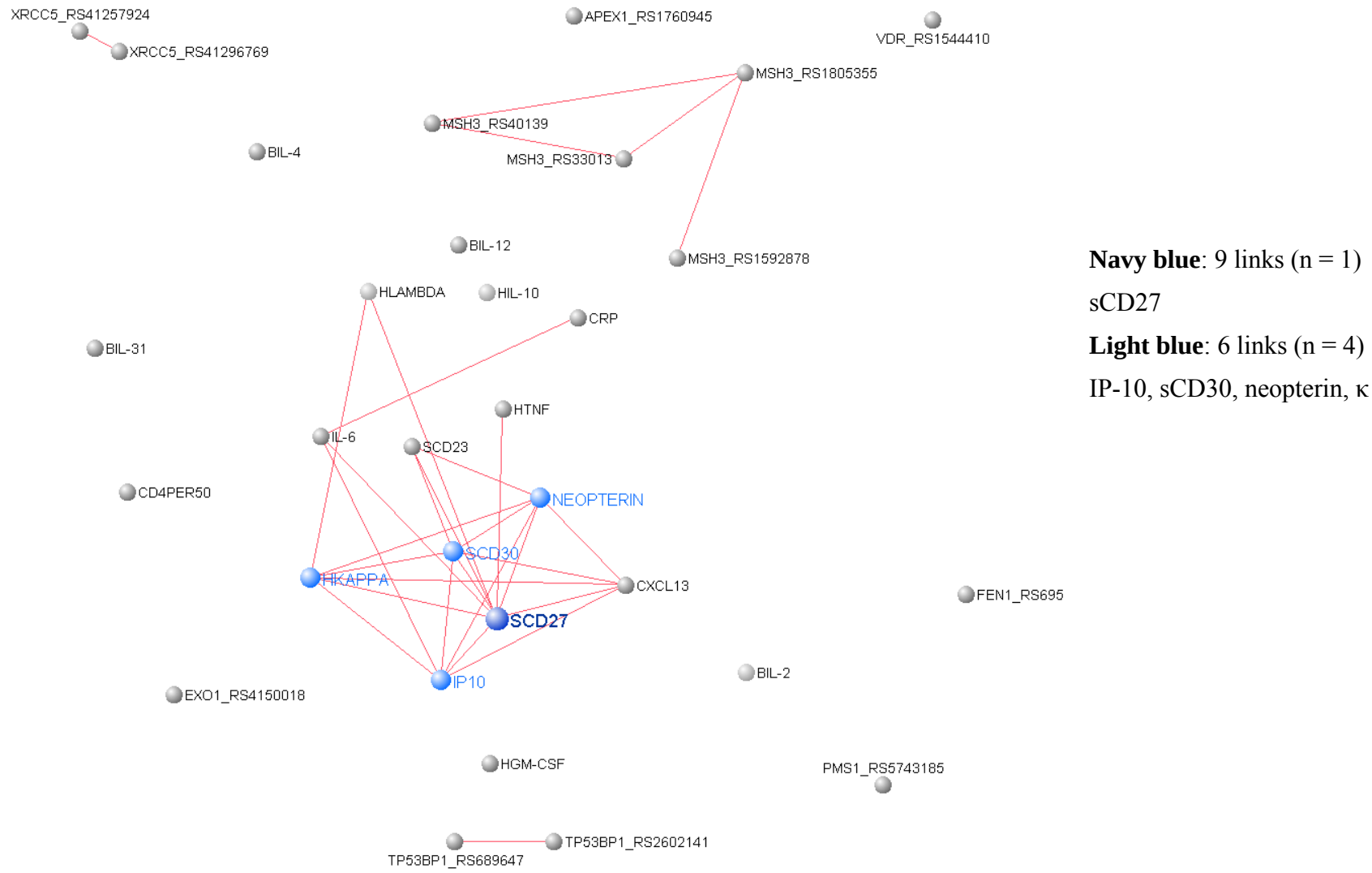


Figure 3.4. Network of 31 biomarkers (18 serum markers and 13 genetic polymorphisms) with topological overlap matrix (TOM) larger than 0.05. sCD27 had most links (9 links) to other biomarkers, followed by IP-10, sCD30, neopterin, and κ free light chain.

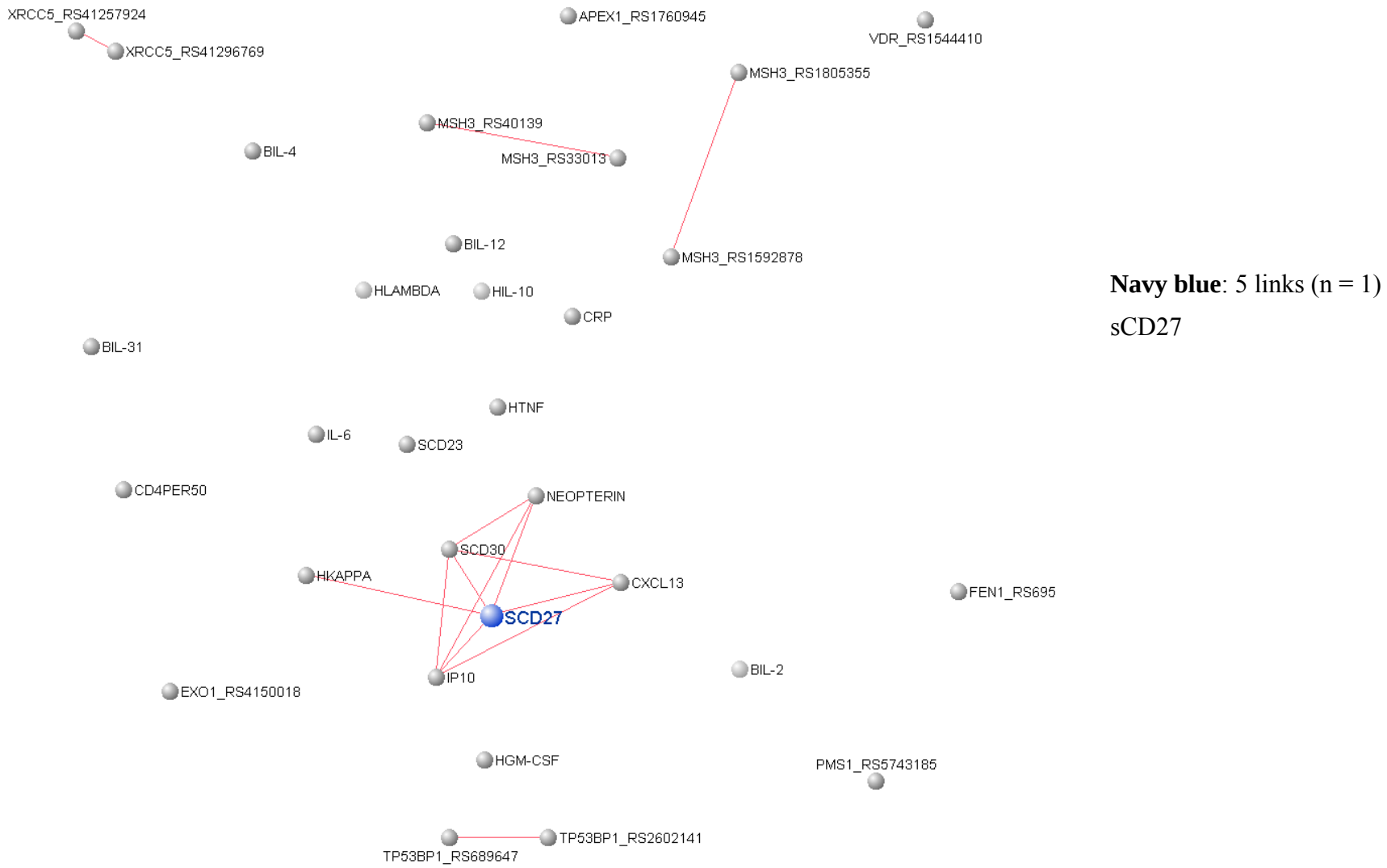


Figure 3.5. Network of 31 biomarkers with topological overlap matrix (TOM) larger than 0.1. sCD27 was the potential hub biomarker that linked to neopterin, IP-10, sCD30, CXCL13, and κ free light chain.

3.4 HIV Viral Burden and AIDS-NHL Risk

The HIV or HCV models in the genetic association study included SNPs (listed in Table 2.3) for adjustments, using confounder-selection strategy of stepwise regression method. Results from conditional logistic regressions on different measures of HIV viral burden and AIDS-NHL risk in the genetic association (Table 3.5) and serum marker study (Table 3.6) are shown respectively. Table 3.5 shows in model 1 associations between $\log_{10}$-transformed HIV RNA copies per mL (VL) and the AIDS-NHL susceptibility with adjustment for potential confounders including $CD4^+$ T-cell count at reference, age, prior AIDS diagnosis, HCV infection, smoking, recreational amphetamine use, HIV infection years, and years of receiving HAART and ART. The complete-data and multiple-imputation analyses returned similar associations. We presented summarized associations from 10 rounds of multiple-imputation as they were more conservative than those from complete data analysis. With one unit increase of $\log_{10}$-HIV RNA at set point, or 10-fold increase of HIV RNA copies per mL measured at steady-state, the odds of having AIDS-NHL among HIV-infected participants in the MACS increased 1.75 folds (adjusted OR: 1.75; 95% CI: 1.20-2.53, model 1, multiple imputation). The 95% CI not including 1 indicated a positive, monotonic trend for this association. Compared with participants with VL at steady-state smaller than 4 (HIV RNA level less than 10,000 copies/mL), HIV+ participants with VL set point between 4 and 5 (HIV RNA stabilized between 10,000 and 100,000 copies/mL) showed 1.75 odds (95% CI: 1.01-2.90), and participants with VL > 5 (HIV RNA >100,000 copies/mL) had 2.06 odds (95% CI: 1.00-4.24) of developing AIDS-NHL.

We also observed a positive association between one-unit increase of VL closest, prior to the reference date in the pre-HAART era and AIDS-NHL risk (adjusted OR: 2.16, 95% CI: 1.43-3.24, Model 1, multiple imputation). Models included potential confounders. Table 3.5 also

shows associations between AIDS-NHL risk and individual-specific slope and intercept for linear regression on pre-HAART VL by time. Holding covariates constant in model 1, including the slope for linear regression, one-unit increase of the intercept was positively associated with an increased risk of AIDS-NHL (adjusted OR: 1.49, 95% CI: 1.01-2.21). However, imprecise association was indicated between the slope (VL change per year) and AIDS-NHL risk with an OR of 1.48 and 95% CI covering one (0.22-9.78).

Model 2 further included several SNPs chosen from the stepwise regression method. Shown in Table 3.8, these SNPs locate at *OSGEP* (*APEX)*, *ATM*, *EXO1, HMB2 (H2AFX)*, *MLH1*, *MSH3*, *POLL*, *POLM*, *TNFSF13B*, and *XRCC5*. Results from both model 1 and 2 with multiple imputation were consistently suggesting moderate and positive associations between HIV viral burden and AIDS-NHL susceptibility.

Table 3.5. Associations between different measures of HIV viral burden and AIDS-NHL risk, adjusted for B-cell activation-related genetic markers in the genetic association study (n = 716)

Genetic association study	Crude association		Model 1				Model 2			
			Complete data analysis		Multiple imputation		Complete data analysis		Multiple imputation	
	cOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
[1]VL at set point[2]										
Continuous	2.26	(1.51-3.39)	2.07	(1.32-3.25)	1.75	(1.20-2.53)	2.73	(1.60-4.64)	2.01	(1.35-2.99)
Categorical										
VL < 4	1		1		1		1		1	
4 ≤ VL < 5	1.98	(1.16-3.39)	2.03	(1.09-3.80)	1.72	(1.01-2.90)	2.68	(1.29-5.56)	1.98	(1.13-3.47)
VL ≥ 5	2.85	(1.33-6.11)	2.53	(1.07-6.00)	2.06	(1.00-4.24)	4.83	(1.73-13.46)	2.76	(1.26-6.05)
Pre-HAART VL[1] closest to ref. date	2.32	(1.57-3.42)	2.08	(1.35-3.19)	2.16	(1.43-3.24)	2.12	(1.31-3.42)	2.22	(1.46-3.38)
Any pre-HAART VL[1] prior to ref. date										
[2]Intercept (VL[1])	1.57	(1.08-2.29)	1.92	(1.21-3.04)	1.49	(1.01-2.21)	2.35	(1.36-4.05)	1.57	(1.05-2.36)
[2]Slope (ΔVL/year)	1.46	(0.23-9.16)	2.47	(0.34-18.1)	1.48	(0.22-9.78)	4.15	(0.42-41.29)	1.69	(0.22-12.99)

cOR crude odds ratio; *aOR* adjusted odds ratio; *CI* confidence interval; *VL* viral load; *HAART* highly active anti-retroviral therapy; *ref.* reference.

Model 1 included $CD4^+$ T-cell count at reference (continuous), age at reference, prior AIDS diagnosis (yes/no), HCV infection (ever/never), smoking (yes/no), recreational amphetamine use (yes/no), HIV infection years, time (years) receiving HAART and ART.

Model 2 included covariates in model 1 and additional SNPs that were chosen by stepwise significance-test of confounder effect with α level of 0.2 (to enter) and 0.25 (to stay). These SNPs were *APEX* rs1760945, *ATM* rs1800889, *EXO1* rs851777, *H2AFX* rs1784304, *MLH1* rs1800734, *MSH3* rs1677645, *POLL* rs4919556, *POLM* rs7778745, *TNFSF13B* rs1041569, and *XRCC5* rs207884.

[1] VL was $\log_{10}$ HIV viral load; 1 unit VL increase represents 10-fold increase of HIV RNA copies/ml.

[2] HIV viral load at set point is a steady-state of HIV viral load, indicating a balance between HIV infection and its clearance within infected individuals.

[3] Participant-specific intercept and slope for linear regression on VL measurements by time.

Table 3.6 shows associations between HIV viral burden and AIDS-NHL risk in the serum marker study. After controlling for potential confounders, summarized associations from multiple-imputation analysis were similar to complete-data analysis. One-unit increase of VL set point, or 10-fold increase of HIV RNA copies per mL, was correlated with a 3.77-fold increased odds of having AIDS-NHL in the serum marker study (adjusted OR: 3.77, 95% CI: 1.48-9.65, multiple-imputation). For VL set point in category, model 1 suggested that participants with VL set point larger than 4 (or HIV RNA level greater than 10,000 copies/mL) would have 2.24-time odds of AIDS-NHL susceptibility compared with those having VL less than 4 (adjusted OR: 2.24, 95% CI: 0.85-5.88) although the 95% CI covered the null. The VL in the pre-HAART era measured at time point closest to the reference date, was also found to be associated with increased risk of AIDS-NHL in HIV-infected MACS participants (adjusted OR: 3.71, 95% CI: 1.27-10.89).

Model 2 further included serum levels of sCD27 and sCD30 measured longer than 3 years preceding AIDS-NHL diagnosis. Both cytokines were identified from the WGCNA and visANT (sCD27), highly correlated with the module eigenvector (r for ME: 0.93and 0.85 for sCD27 and sCD30, respectively) and case/control status (r for NHL status: 0.28 and 0.29 for rCD27 and sCD30, respectively). Results from conditional logistic regression indicated strong, positive associations between HIV viral burden and AIDS-NHL risk, after controlling for B-cell activation-related cytokines. One-unit increase of VL at set point was associated with 3.65-fold increased odds of NHL susceptibility in HIV-positive populations (adjusted OR = 3.65, 95% CI: 1.52-8.72, multiple imputation). Compared with HIV-infected participants who had VL < 4, participants with VL larger than 4 showed 2.11-fold odds of NHL development (adjusted OR= 2.11, 95% CI: 0.87-5.10, multiple imputation). We also observed positive association between

pre-HAART VL and AIDS-NHL risk. One-unit increase of VL prior to HAART-era was associated with 5-fold increased odds of having AIDS-NHL in HIV+ individuals. However, the wide confidence interval implied unstable associations or an issue with sparse data, or both.

We also explored the potential heterogeneity of odds ratios relating HIV VL set point to NHL susceptibility by serum levels of cytokines in the serum study. We performed stratified analysis on serum levels of five molecules, including sCD27, sCD30, IP-10, CXCL13, and neopterin. These molecules all showed high correlations ($r > 0.6$) with the module eigenvector in the first WGCNA module. Figure 3.6 illustrates summarized and strata-specific ORs, adjusted for $CD4^+$ T-cell count. There might be an odds-ratio variation by serum neopterin levels as the summarized point estimate (OR = 3.91) located within strata-specific point estimates (OR = 13.06 and 1.92). However, our study might not have enough power to detect the heterogeneity due to smaller sample size (*P* for heterogeneity = 0.990), or there might be no heterogeneity.

Table 3.6. Associations between measures of HIV viral burden and AIDS-NHL risk, adjusted for serum levels of sCD27 and sCD30 in the serum marker study (n = 358)

Serum marker study	Crude association		Model 1				Model 2			
			Complete data analysis		Multiple imputation		Complete data analysis		Multiple imputation	
	cOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
[1]VL at set point[2]										
Continuous	6.31	(3.10-12.88)	3.98	(1.19-13.28)	3.77	(1.48-9.65)	3.54	(1.26-9.90)	3.65	(1.52-8.72)
Categorical										
VL < 4	1		1		1		1		1	
VL ≥ 4	3.25	(1.70-6.21)	1.25	(0.33-4.80)	2.24	(0.85-5.88)	1.65	(0.55-4.95)	2.11	(0.87-5.10)
Pre-HAART VL[1] closest to ref. date	5.51	(2.59-11.7)	2.41	(0.76-7.66)	3.71	(1.27-10.89)	3.46	(1.49-8.07)	5.03	(1.40-18.13)

cOR crude odds ratio; *aOR* adjusted odds ratio; *CI* confidence interval; *VL* viral load; *HAART* highly active anti-retroviral therapy; *ref.* reference.

Model 1 included $CD4^+$ T-cell count closest and prior to reference (continuous), age at reference, prior AIDS diagnosis (yes/no), smoking (yes/no), and HAART (yes/no).

Model 2 included $CD4^+$ T-cell count at the third serum sampling time (continuous), age at reference, prior AIDS diagnosis (yes/no), HAART (yes/no), and serum levels of sCD27 and sCD30 measured greater than 3-year preceding AIDS diagnosis.

[1] VL is $\log_{10}$ HIV RNA level.

[2] HIV viral load at set point is a steady-state of HIV RNA level, indicating a balance between HIV infection and its clearance within infected individuals.

	AIDS-NHL cases		HIV+ controls	
	n	VL at set point, median (IQR)	n	VL at set point, median (IQR)
sCD27 (log_e U/mL)				
< 6.0	54	4.55 (4.22, 5.07)	94	4.13 (3.56, 4.44)
≥ 6.0	83	4.48 (4.14, 4.82)	50	4.12 (3.96, 4.43)
P for heterogeneity: 0.893				
sCD30 (log_e U/mL)				
< 4.52	61	4.49 (4.23, 4.92)	91	4.03 (3.55, 4.41)
≥ 4.52	76	4.51 (4.14, 4.93)	53	4.31 (4.07, 4.48)
P for heterogeneity: 0.256				
Neopterin (log_e nmol/L)				
< 2.67	58	4.73 (4.37, 5.13)	100	4.17 (3.67, 4.42)
≥ 2.67	79	4.43 (4.09, 4.76)	44	4.16 (3.85, 4.50)
P for heterogeneity: 0.990				
IP10 (log_e pg/mL)				
< 7.55	65	4.56 (4.25, 5.06)	105	4.14 (3.56, 4.45)
≥ 7.55	72	4.48 (4.12, 4.77)	39	4.16 (3.91, 4.43)
P for heterogeneity: 0.795				
CXCL13 (log_e pg/mL)				
< 4.55	60	4.51 (4.15, 5.03)	101	4.16 (3.65, 4.44)
≥ 4.55	77	4.49 (4.18, 4.84)	43	4.16 (3.82, 4.52)
P for heterogeneity: 0.703				

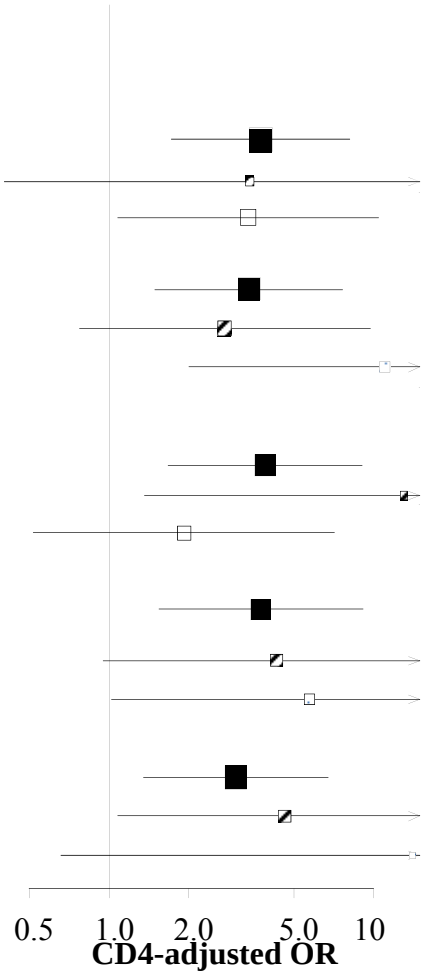


Figure 3.6. Summarized and strata-specific adjusted OR relating HIV VL at set point and B-cell lymphoma susceptibility in HIV+ MACS participants. Forty-two cases and thirty-five controls had missing cytokines levels. The analysis was stratified by serum medium levels of sCD27, sCD30, neopterin, IP10, and CXCL13, after natural log transformation. These molecules are possible cytokines representing B-cell activation identified in the network analysis.

3.5 HCV Infection and AIDS-NHL Risk

Table 3.7 shows associations between HCV infection and AIDS-NHL risk in both studies. Results from model 1 in both studies suggested increased odds of lymphoma susceptibility among HCV/HIV co-infected participants compared with HIV mono-infected participants. HCV models included continuous $CD4^+$ T-cell count, age, prior AIDS diagnosis, smoking, HAART, and years receiving ART. In HCV models without adjustment for HIV viral load, point estimates from multiple-imputation analysis indicated 1.59-fold odds increase (aOR: 1.59 [0.93-2.76] and 1.59 [0.50-4.99] for the genetic and serum marker studies, respectively), although 95% CIs included the null. Point-estimates for odds ratio relating chronic HCV infection to lymphoma susceptibility consistently implied moderate yet imprecise associations, with or without adjustment for HIV viral load.

Model 2 shows adjusted ORs relating HCV infection to AIDS-NHL risk with adjustment for B-cell activation-related SNPs or serum cytokines levels. In genetic association study, complete-data analysis suggested positive associations relating ever or chronic HCV infection to AIDS-NHL risk, with or without HIV viral load adjustment. HCV models with multiple-imputation regressed associations toward the null, indicating borderline positive associations. In serum marker study, no apparent associations between HCV infection and AIDS-NHL risk were observed after controlling for serum levels of cytokines (sCD27, sCD30, and CXCL13). These cytokines were identified from the network analysis.

Figure 3.7 illustrates $CD4^+$ T-cell count adjusted ORs relating HCV infection to NHL risk in HIV-infected MACS participants, stratified by serum medium levels of cytokines. The network analysis indicated these cytokines were possibly related to B-cell activation. Odds ratios

relating HCV infection to AIDS-NHL seem to be heterogeneous by serum CXCL13 or neopterin levels. However, our study might not have enough power to detect the heterogeneity, or there might be no heterogeneity (both *P* for heterogeneity were larger than 0.05).

Table 3.7. Associations between HCV infection and AIDS-NHL risk in genetic association study (n = 716) and serum marker study (n = 358)

HCV infection	Case/Control, n (%)	Crude association		Model 1				Model 2			
				Complete data analysis		Multiple imputation		Complete data analysis		Multiple imputation	
		cOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)	aOR	(95% CI)
Genetic association study (n = 716)											
Never	158 (86)/484 (91)	1		1		1		1		1	
Ever	25 (14)/49 (9)	1.57	(0.94-2.64)	1.97	(1.09-3.58)	1.59	(0.93-2.76)	2.85	(1.45-5.70)	1.75	(0.97-3.13)
Ever + VL at set point				2.08	(1.02-4.28)	1.53	(0.88-2.66)	3.79	(1.57-9.14)	1.70	(0.94-3.09)
Chronic	18 (10)/34 (6)	1.59	(0.87-2.89)	1.85	(0.93-3.70)	1.58	(0.84-2.96)	2.41	(1.09-5.35)	1.77	(0.89-3.50)
Chronic + VL at set point				2.29	(0.98-5.36)	1.53	(0.80-2.90)	3.48	(1.28-9.45)	1.75	(0.87-3.54)
Serum marker study (n = 358)											
Never	154 (86)/163 (91)	1		1		1		1		1	
Ever	25 (14)/16 (9)	1.75	(0.86-3.56)	1.52	(0.42-5.57)	1.59	(0.50-4.99)	2.04	(0.50-8.34)	1.44	(0.46-4.50)
Ever + VL at set point				1.36	(0.18-10.5)	2.05	(0.49-8.48)	2.38	(0.34-16.75)	1.59	(0.43-5.87)
Chronic	18 (10)/10 (6)	1.89	(0.84-4.24)	1.58	(0.40-7.36)	1.96	(0.47-8.16)	1.63	(0.37-7.14)	1.30	(0.38-4.51)
Chronic + VL at set point				1.90	(0.24-14.9)	2.29	(0.50-10.35)	2.01	(0.33-12.36)	1.51	(0.38-6.01)

cOR crude odds ratio; *aOR* adjusted odds ratio; *CI* confidence interval;

Model 1 included $CD4^+$ T-cell count closest and prior to reference (continuous), age, prior AIDS diagnosis (yes/no), smoking (ever/never), HAART (yes/no), and time (years) receiving ART.

Model 2 in genetic association study additionally included additional SNPs chosen from stepwise significance-test of confounder effect with α level of 0.2 (to enter), 0.25 (to stay), as listed in Table 3.7. Model 2 in serum marker study included covariates in model 1 and additional serum inflammation-related markers (sCD27, sCD30, and CXCL13) measured longer than 3-year preceding AIDS-NHL diagnosis.

	HCV/HIV co-infection, n (%)	HIV mono-infection, n (%)	aOR	95% CI
	Cases/controls	Cases/controls		
sCD27 (log_e U/mL)				
< 6.0	9 (11)/ 10 (8)	73 (89)/ 113 (92)	**1.72**	**0.65 - 4.55**
≥ 6.0	16 (16)/ 6 (11)	81 (84)/ 50 (89)	4.15	0.34 - 49.57
P for heterogeneity: 0.925			2.73	0.45 -16.68
sCD30 (log_e U/mL)				
< 4.52	11 (12)/ 9 (7)	78 (88)/ 111 (93)	**1.56**	**0.58 - 4.18**
≥ 4.52	14 (16)/ 7 (12)	76 (84)/ 52 (88)	5.20	0.29 - 93.01
P for heterogeneity: 0.546			3.24	0.35 - 30.20
Neopterin (log_e nmol/L)				
< 2.67	10 (11)/ 8 (6)	81 (89)/122 (94)	**2.23**	**0.81 - 6.15**
≥ 2.67	15 (17)/ 8 (16)	73 (83)/ 41 (84)	0.62	0.05 - 7.84
P for heterogeneity: 0.190			3.46	0.65 - 18.49
IP10 (log_e pg/mL)				
< 7.55	11 (12)/ 12 (9)	81 (88)/ 119 (91)	**1.84**	**0.68 - 4.97**
≥ 7.55	14 (16)/ 4 (8)	73 (84)/ 44 (92)	2.23	0.52 - 9.51
P for heterogeneity: 0.770			2.10	0.18 - 24.36
CXCL13 (log_e pg/mL)				
< 4.55	15 (18)/ 12 (9)	70 (82)/ 115 (91)	**2.51**	**0.90 - 6.99**
≥ 4.55	10 (11)/ 4 (8)	84 (89)/ 48 (92)	9.15	1.07 - 78.19
P for heterogeneity: 0.978			0.50	0.03 - 8.15

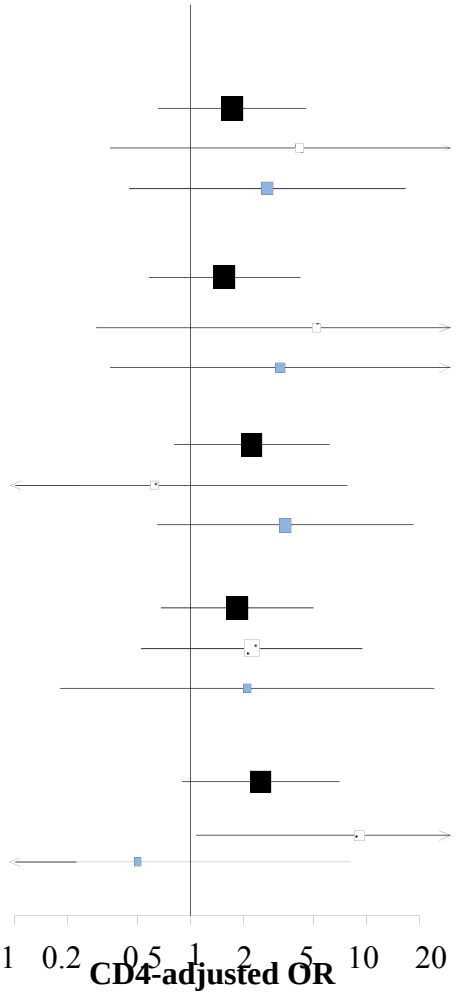


Figure 3.7. Summarized and strata-specific aOR relating HCV and lymphoma susceptibility in HIV-infected MACS participants. The analysis was stratified by serum medium levels of sCD27, sCD30, neopterin, IP10, and CXCL13, after natural log-transformation. These molecules are possible cytokines representing B-cell activation identified in the WGCNA.

CHAPTER 4: DISCUSSION

4.1 Network Analysis

In this dissertation, we screened out potential biomarkers from 328 SNPs and 32 serum cytokines in the serum marker study using the weighted gene co-expression network analysis (WGCNA). These biomarkers included sCD27, sCD30, neopterin, CXCL13, and IP-10, as well as SNPs on *XRCC5*, *TP53BP1*, *MSH3*, *VDR*, and *REV3L*. visANT was used to visualize networks including 13 SNPs and 18 serum cytokines. Using the TOM as an association entry, visANT showed that sCD27 had the highest links among the 31 biomarkers chosen for visANT. Other essential cytokines identified from visANT included IP-10, sCD30, and neopterin. The WGCNA and visANT returned similar results.

WGCNA is a newly developed analysis and gradually adopted in several cancer researches, including predictive biomarkers in prostate cancer,[117] breast cancer,[118] glioma,[119] and glioblastoma.[120] It uses soft thresholding to build up a weighted, scale-free network and generalizes topological overlap matrix (TOM) to cluster nodes into modules. The WGCNA further calculates module eigenvector (ME) to 'represent' a specific module. It uses correlations between ME and research outcome (AIDS-NHL case/control status in our study), and between ME and individual nodes (biomarkers) as well, to capture essential role-players in a network.

Genetic and serum markers in our study are expected to correlate to one another as they were selected into our study based on their putative effects during B-cell development, B-cell maturation, and immune function. traditional analyses would not be able to deal with hundreds of covariates that are highly correlated. Furthermore, the multiple-comparisons issue was also a concern. A network analysis, such as the WGCNA,[120] is one developed approach for statistical

dimension reduction, specially focusing on system-level functionality of genes from microarray data. The WGCNA builds up a weighted, co-expression network that every node (biomarkers) connects to all other nodes based on soft-thresholding, which distinguishes the WGCNA from other network modeling. The connectivity (strength of connection) of any two nodes in the WGCNA depends on the adjacency function, which is the absolute value of Pearson correlation coefficient raised to a power of β. Every node connects to all other dotes in the WGCNA network and less correlated connectivity is penalized accordingly by the power β.

In the WGCNA, nodes (or biomarkers) are categorized into modules using the topological overlap matrix. We could identify essential biomarkers within modules through several correlations – markers' correlations with module eigenvector or outcome. The eigenvector is one commonly used statistical approach to reduce number of covariates (dimensionality reduction) although it lacks practical and meaningful interpretations. We used the WGCNA to identify potential 'hub biomarkers' that showed high correlations with module eigenvector. These hub biomarkers give more practical usage and practical interpretations than eigenvectors. In this dissertation, our HIV and HCV models included hub biomarkers from the WGCNA to adjust for B-cell activation. On the other hand, HIV or HCV models could otherwise include the module eigenvector for B-cell activation adjustment.

The WGCNA resulted in several potential biomarkers, including sCD27, sCD30, neopterin, CXCL13, and IP-10, from 33 serum molecules and 328 genetic polymorphisms. The visANT also indicated that sCD27 had most links to other biomarkers given TOM > 0.1. Other visANT suggested biomarkers included IP-10, sCD30, neopterin, and κ free light chain if TOM was larger than 0.05. These molecules are all indicative of or involved in B-cell stimulation,

differentiation, or immune activation. Increased serum markers' levels suggest an immune stimulatory environment including B-cell activation. While HIV infection has long been acknowledged to induce B-cell hyperactivation, a number of studies support the hypothesis of B-cell hyperactivation contributing to lymphomagenesis in HIV+ individuals.[7-10, 17-19] These studies reported positive associations between elevated serum levels and AIDS-NHL risk, even in years preceding lymphoma diagnosis.

Some etiologic explanations in lymphogenesis have been proposed for chemokines (CXCL13, IP-10) or tumor necrosis factor (TNF) receptor superfamily cytokines (CD27, CD30).

The chemokines (CXCL13 and IP-10) share similar chemo-attracting effects on B lymphocytes to the germinal center. In germinal centers, immature B cells differentiate and increase the affinity and diversity of immunoglobulin through two DNA-modifying processes (somatic hypermutation, SHM, and class switch recombination, CSR). Chemokines may potentially contribute to the formation of ectopic germinal center-like structures, contributing to chronic B-cell activation and consequent accumulation of mutations and translocations.[9, 121] Furthermore, some studies have shown increased CXCL13 levels and decreased CXCR5 expression (receptor of CXCL13) on circulating B cells in HIV+ populations.[13, 122, 123] Aberrant expression of chemokines and their receptors might interfere with B-cells homing, leading to inappropriate homing to other tissues and potential extra-nodal lymphomagenesis.[9, 17]

CD27 and CD30 are both cell surface receptors of TNF receptor superfamily and have proteolytic products -- soluble CD27 (sCD27) and soluble CD30 (sCD30). CD70 and CD30L (CD153), both TNF-like cytokines, are ligands for CD27 and CD30, respectively. CD27 is found on resting T cells, subset of B cells, and natural killer (NK) cells, and directly stimulates IgG and

IgM synthesis and B-cell differentiation after CD70 ligation.[124, 125] CD30 is prominently expressed on malignant B lymphocytes of Hodgkin lymphoma, anaplastic large cell lymphoma, and a small portion of normal T and B lymphocytes activated by antigens. CD30 and CD30L co-stimulate T-cell and B-cell differentiation and cytokine production.[126] Higher serum levels of sCD27 or sCD30 have been found to be predictive of disease progression in HIV+ populations [127, 128] and indicative of B-cell lymphoma susceptibility in general population.[129-131] Positive associations between AIDS-NHL risk and elevated levels of sCD27 and sCD30 still evident in HIV+ populations after controlling for HIV infection period and effect of immune deterioration ($CD4^+$ T cell count slope).[7, 10, 14] These findings support the hypothesis of chronic B-cell hyperactivation in AIDS-NHL patients.

The prognostic value of neopterin has long been recognized in cancer susceptibility[132, 133] and HIV infection progression.[134] Elevated serum level of neopterin have also been observed in acute EBV infection or hepatitis. Macrophages and monocytes are the source of neopterin in humans, presumably after being activated by IFN-γ.[134] Other macrophage stimulating cytokines, such as IFN-β, GM-CSF, have no inducing effect on neopterin secretion. B lymphocytes, granulocytes, NK cells do not produce neopterin upon activation. Increased neopterin level is negatively associated with $CD4^+$ T-cell count, and positively with HIV viral load, disease progression, and overall mortality in HIV+ populations.[134, 135] The association between neopterin and AIDS-NHL risk also been reported using specimens collected several years prior to lymphoma diagnosis.[8, 9]

These previous studies, however, faced common challenges of multiple-comparison issues, collinearity of highly correlated serum inflammation-related markers, and unstable results

from sparse data. The WGCNA network summarized cytokines into one module (Table 3.6, sCD27, sCD30, IP-10, CXCL13, neopterin, sCD23, IL-6, TNFα, IL-10, and CRP) and further suggested that sCD27, sCD30, and IP-10 had high correlations ($r > 0.70$) with the module eigenvector. The Spearman correlations among these 10 biomarkers support this finding (Supplementary Table S4.1). sCD27 was most correlated ($r > 0.5$) with other biomarkers in the same module. This finding is unexpected, given several prior studies focusing on associations of AIDS-NHL risk with these serum molecules in the same population. Further studies are needed to explore biological mechanisms.

4.2 HIV Viral Burden and AIDS-NHL Risk

HIV viral load has long been recognized as one important prognostic marker for HIV infection progression and AIDS defining illness,[109, 110] in addition to $CD4^+$ T-cell count or CD4 slope.[109] Cancer risk in HIV positive populations has been decreased in the HAART era, in particular Kaposis sarcoma (KS) and non-Hodgkin lymphoma (NHL). However, some subtypes of AIDS-NHL show no evidence of a decline in the HAART era, such as Burkitt's lymphoma. Risk of Burkitt's lymphoma in the HAART era has not been decreased in several studies (post-HAART *vs.* pre-HAART, relative risk, RR = 1.18 [0.48-2.88][2] and RR = 3.8, [0.50-29.4][3]), although the number of patients with Burkitt's lymphoma was small and confidence intervals were wide. Several studies indicated that HIV RNA level at diagnosis[136-138], or cumulative HIV viremia[28, 138, 139], is positively associated with AIDS-NHL risk or incidence. Our results using two nested case-control studies within the MACS cohort support these findings. Noticeably, cases and controls in both studies were matched on the HIV infection period to reduce impact of HIV progression (Table 2.1). After accounting for HIV sero-converting status and HIV infection years (median years of HIV infection period prior to reference date: 6.8 and 6.7 for genetic and serum marker study, respectively), we still observed positive associations between AIDS-NHL risk and HIV viral burden. Controlling for genetic or serum biomarkers related to B-cell activation in the model, our study still indicated positive associations between HIV viral burden and AIDS-NHL risk. These suggested associations support the hypothesis that HIV viral burden could increase AIDS-NHL risk, with or without adjusting for B-cell activation-related biomarkers.

In our research, HIV viral load at several time points was assessed, including viral load at set point or at time closest and prior to the reference date. We observed similar associations

before and after B-cell activation markers adjustment. In the genetic study, HIV and AIDS-NHL risk associations are similar between model 1 and 2 (Table 3.5). Model 2 included 10 of 20 SNPs using step-wise significance-test of covariate selection. SNPs were selected into HIV models if their associations with the outcome were significant at α level of 0.20 and retained if their associations with the outcome were significant at α level of 0.25. Nine SNPs located at DNA repair genes (*APEX* rs1760945, *ATM* rs1800889, *EXO1* rs851777, *H2AFX* rs1784304, *MLH1* rs1800734, *MSH3* rs1677645, *POLL* rs4919556, *POLM* rs7778745, and *XRCC5* rs207884) and one at *TNFSF13B* (rs1041569). Preliminary analysis showed weak to moderate strength of associations relating individual SNP to AIDS-NHL risk in our population. Our results indicated that additional SNPs in multivariable HIV models did not drastically change associations. Moreover, models also included continuous $CD4^+$ T-cell count at reference date to adjust for potential residual effect even though cases and controls were matched on $CD4^+$ T-cell count. Our study results support the predictive role of HIV viral load and raise a possibility that B-cell activation related SNPs might not as important as the HIV viral load to AIDS-NHL susceptibility.

In the serum marker study, point estimates indicated strong associations between HIV RNA level and AIDS-NHL risk although wide confidence intervals indicated imprecise associations. This may be due to small sample size or unstable models, or both. Associations might also be due to chance alone. HIV models included serum levels of sCD27 and sCD30 measured longer than 3 years preceding reference date. Table 3.8 shows similar associations between viral load at set point and AIDS-NHL risk between model 1 and 2 (adjusted OR: 3.77 [95% CI: 1.48-9.65] and 3.65 [95% CI: 1.2-8.72] for model 1 and 2, respectively). The point estimate for pre-HAART viral load closest to reference date however increased from 3.71 (model 1) to 5.03 (model 2) after adjustment for B-cell activation-related cytokines. These results

implied unstable modeling or potential over-adjustment. Given these findings, HIV viral load at set point might be a better indicator than pre-HAART viral load prior to reference date to predict AIDS-NHL development.

The HIV viral load set point is a steady-state of HIV infection. At set point, a balance is reached between HIV viral replication and clearance within HIV-infected subjects naïve to antiretroviral treatment.[106, 107] The period from HIV sero-conversion to this steady-state, and the viral load at set point, vary greatly across individuals. Previous studies in the pre-HAART era showed that the HIV viral load set point is an independent predictive marker for AIDS development or death from AIDS.[109, 110, 140] Mellors JW, et al. reported a better predictor of HIV viral load set point for HIV progression than $CD4^+$ T-cell count and neopterin level at baseline of the MACS follow-up.[109] Our study further indicated HIV viral load set point might be a better marker than other measures of HIV viremia to predict the risk of AIDS-NHL. AIDS-NHL models could consider including HIV viral load set point and $CD4^+$ T-cell count for better model-fitting.

We also used slope and intercept for simple linear regression on $\log_{10}$ viral load by year for each participant who had at least two HIV RNA measurements. This approach was inspired by $CD4^+$ T-cell count slope[140] and constant/slope in HIV VL virologic decay pattern,[141] both of which are predictive of AIDS malignancies. However, this was only an exploratory approach since missing values are of concern in our study. Less than 75% of study participants had at least 3 measurements of HIV RNA from baseline to reference date (at least 3 HIV RNA measurements in cases and controls: 51% and 57% for the genetic study; 48% and 74% for the serum marker study, Table 3.2). We still observed elevated risk of AIDS-NHL with one unit

increase of intercept (log_{10} viral RNA copies/mL) and slope (ΔVL/year) for linear regression. These associations are expected, as we only included pre-HAART HIV RNA data and most participants showed an upward trend of viral load during the follow-up. Nonetheless, this approach is far from accurate, as a complex mathematic formula is needed to better depict HIV viral burden or HIV viremia[139, 141]. With limited practical use or public health relevance, the approach still underlies the importance of long-term treatment to reduce HIV viral load given that sub-optimal HIV control is possibly associated with increased risk of AIDS-NHL risk.

4.3 HCV Infection and AIDS-NHL Risk

HCV infection has been reported to be associated with increased risk of B-cell lymphoma in the general population,[40, 142, 143] particularly of diffuse large B-cell lymphoma subtype.[40, 142] Its association with NHL in the HIV population, however, is inconclusive.[41-46] Using two nested case-control studies within the MACS cohort, HCV infection appeared to be marginally associated with increased AIDS-NHL risk before and after adjustment for B-cell activation biomarkers. Marginal associations might be due to insufficient sample size as point-estimates for odds ratios consistently indicated a higher AIDS-NHL risk among HCV/HIV co-infected participants than HIV mono-infected participants.

Few epidemiological studies examined HCV-NHL association in HIV populations. Some,[43-46] but not all,[41, 42] of them showed null associations. It is worth pointing out that none of these studies specifically focused on HIV-infected men who have sex with men. Interestingly, an increased OR relating anti-HCV antibody to AIDS-NHL risk was suggested in the subgroup of men who have sex with men in a Swiss cohort ($CD4^+$ T-cell count at baseline and age adjusted OR = 2.37; 95% CI: 1.03-5.43).[42]

Participants in our studies are all men who have sex with men. Point-estimates for adjusted OR indicated marginal associations relating ever HCV infection to AIDS-NHL risk without adjusting for HIV viral load (adjusted OR: 1.59, [0.96-2.76] for the genetic study, 1.59 [0.50-4.99] for the serum marker study, respectively). In addition to $CD4^+$ T-cell count, our models further included the HIV RNA level set point and other risk factors for AIDS-NHL. The HIV RNA level set point is a strong predictive factor for AIDS-NHL. As HCV prevalence is low in our study, further replicate researches with larger sample size are needed.

HCV infection was usually measured retrospectively in previous studies with only one-time anti-HCV measurement. The MACS implements measurements of HCV antibody and HCV RNA at multiple time-points, one at baseline visit, to confirm HCV infection status. These measurements were performed retrospectively on stored serum from baseline visit through 2001, and prospectively since 2001 in every six-month. Chronic HCV infection is based on detectable HCV RNA level with preceding positive anti-HCV. The measurement error would be non-differential between NHL cases and HIV controls as the determination of NHL status is independent to HCV measurement. The bias due to HCV measurement error would be limited or toward the null.

Multivariable HCV models included continuous $CD4^+$ T-cell count prior to reference date, HIV RNA level at set point, prior AIDS diagnosis, age, smoking, HAART and years of ART. These factors were either potential confounding factors or strong predictive factors for HCV-NHL risk. We further included B-cell activation markers in HCV models. Different multivariable regressions returned similar point-estimates across models, indicating associations between HCV infection and AIDS-NHL risk. Although there is a possibility of over-adjustment, we still observed similar point estimates for crude and adjusted odds ratios.

Suggested associations between chronic HCV infection and AIDS-NHL risk, however, did not appear to be stronger than ever HCV infection and NHL risk association. According to the HCV test algorithm, a participant with chronic HCV infection had at least one detectable HCV RNA level and at least one prior anti-HCV positive result, meaning they would have detectable HCV RNA levels for at least 6 months. Similar associations from chronic and ever HCV infection models are expected as large portions of ever HCV-infected participants in our

population had chronic HCV infection (72% of cases and 67% of controls who were ever HCV-infected had chronic HCV infection.) HIV-infected individuals who had chronic HCV infection or higher HCV viral load for longer time might result in increased AIDS-NHL susceptibility through B-cell hyperactivation. Preliminary results support this hypothesis. Subjects with chronic HCV/HIV co-infection showed higher serum levels of cytokines indicative of B-cell activation compared to subjects with HIV or HCV mono-infection.[144] These cytokines include IFNα2, IL-2, IL-3, IL-6, IL-8, IL12, and IL15. Further studies are needed to elucidate possible correlations between chronic HCV infection, HCV viral load, and/or HCV infection period and AIDS-NHL susceptibility.

As there is a global HCV outbreak among HIV-infected men who have sex with men, an increased HCV prevalence in this population is expected.[145-147] Our research enrolled the largest number of AIDS-NHL cases in an MSM populations to date and had prospective HCV infection measurement with little or non-differential measurement error. Positive associations relating HIV/HCV co-infection to AIDS-NHL risk were consistently suggested or implied after controlling for stronger confounders and predictors. Further studies in the HAART era are needed to examine potential association of HCV with AIDS-NHL susceptibility among HIV-infected individuals.

CHAPTER 5: CONCLUSION AND PUBLIC HEALTH IMPLICATIONS

In summary, this dissertation performed network analysis to identify potential biomarkers from 32 serum inflammation markers and 328 SNPs that are putatively related to B-cell activations or increased immune function. Both WGCNA and network visualization of visANT distinguished several serum inflammation markers, such as cytokines that showed high correlation ($r > 0.6$) with the eigenvector for a cluster most correlated to AIDS-NHL case/control status. These serum inflammation-related markers included sCD27, sCD30, CXCL13, IP10, and neopterin. All were measured longer than 3 years preceding the reference date. The visANT also indicated that sCD27 has the highest links with other serum markers (neopterin, IP-10, sCD30, CXCL13, and κ free light chain), given a topological overlap matrix larger than 0.1. For B-cell activation adjustment, sCD27, CD30, and CXCL13 were candidates to be included in HIV and HCV models.

The associations between HIV viral burden and AIDS-NHL risk were explored in this dissertation by several approaches in the genetic association study (n = 716) or the serum marker study (n = 358). Three exploratory measures included HIV RNA level at set point, pre-HAART HIV viral load that are closest and prior to reference date, and slope and intercept for linear regression on $\log_{10}$-HIV RNA copies/mL by year. These measures indicate either cumulative HIV viral burden or one time-point marker. The HIV RNA at set point, a steady-state of HIV viral load observed in several months after sero-conversion, was positively associated with the AIDS-NHL risk in HIV models with and without adjustment for SNPs or serum cytokines (adjusted OR = 2.01 [1.35-2.99] for the genetic association study with adjustment for SNPs; adjusted OR = 3.65 [1.52-8.72] for the serum marker study with adjustment for sCD27 and sCD30). HIV RNA level at set point well predicts the AIDS-NHL risk with public health

relevance. Reduction of HIV level in the early phase of HIV infection may decrease AIDS-NHL risk, suggesting that early detection of sero-conversion and regular anti-retroviral therapy along HIV infection are important to decrease AIDS-NHL risk.

This dissertation also explored associations between ever *vs.* never HCV infection, as well as chronic HCV infection, and AIDS-NHL risk. Associations were consistently implied by point estimates in HCV models that included $CD4^+$ T-cell count. ORs in HCV models before and after adjustment for B-cell activation implied HIV/HCV co-infection is associated with increased susceptibility of AIDS-NHL (adjusted OR = 1.75 [0.97-3.13] for the genetic association study with SNPs adjustment; adjusted OR = 1.44, [0.46-4.50] for the serum marker study including serum levels of sCD27, sCD30, CXCL13). We only observed imprecise associations in a MSM population that contains the most NHL cases to date. While a global HCV outbreak among HIV-infected MSM populations has been noticed since 2000, our study tries to fill up the gap of AIDS-NHL risk in HIV/HCV co-infected populations. Further relationship between HIV/HCV co-infection and AIDS-NHL risk among HIV participants remains to be elucidated in the HAART era.

CHAPTER 6: FUTURE WORKS

Network Analysis

The WGCNA is one useful approach to construct networks under different settings. We applied the WGCNA to screen out potential biomarkers for further HIV and HCV models adjustment, using a network with minimum module membership of 1. While this setting may build up a network that is based on both correlations between any two markers and between a marker and the outcome, some findings are worth discussion. First, there are 74 modules (Figure 2.2) from 348 markers (32 cytokines and 328 SNPs), indicating too many modules with few members. Each module conveys relatively little information. Second, few modules simultaneously contain serum cytokines and genetic polymorphisms. This phenomenon is expected, as the minimum module membership of 1 would potentially cluster markers into continuous serum cytokines or categorical SNPs. Third, the SNPs clustered into one module are usually located on the same gene, such as *TP53BP1, MSH3,* and *VDR* in module 3 to 6 (Table 3.3). Few modules contain polymorphisms on different genes, suggesting a possibility of linkage disequilibrium of module-specific SNPs or a problematic grouping on SNPs.

Networks are worth exploring with several reasonable WGCNA settings. There is no guideline or rule for minimum module membership in the WGCNA. Different minimum module membership might result in distinct networks for relatively fewer inputs (hundreds of markers) like ours. With an adjacency function raised to a power of 3, we constructed networks with minimum module membership of 5, 7, and 8, respectively. Table 6.1 shows preliminary results of module number, number of markers in grey module, and number of markers in the largest module. We observed a trade-off between total module number and number of markers in the grey module by different minimum module membership. For example, given a minimum module

membership of 7, there are 15 modules in total. However, 193 of 360 (54%) markers could not be categorized into any module and were lumped in the grey module. When we set the minimum module membership to 5, the network contained a total of 26 modules and put 114 of 360 (32%) markers in the grey module.

Table 6. 1. Three networks from different minimum module memberships by the WGCNA for 32 cytokines and 328 SNPs

Minimum module membership	Module number	Number of markers in the grey module (%)	Number of markers in the largest module
5	26	114 (32)	20
7	15	193 (54)	20
8	10	237 (66)	25

These exploratory results demonstrate the flexibility and instability of WGCNA networks in our research. These networks largely depend on the minimum module membership. Relatively few markers and two types of data (continuous and categorical markers) might explain the network instability.

We could construct a network including only one large module containing 328 SNPs data without cytokines information. This approach was not adopted in present research since we aimed to screen out potential markers from both genetic and serum biomarkers. The small sample size of serum markers prevents us to construct networks specific to serum markers. The WGCNA is a useful tool to analyze thousands of gene expression data on a genome-wide scale. However, its application for networks with only hundreds of markers is unclear. In addition, its

robustness for networks containing two types of data (binary and continuous) needs further research.

Shrinkage Regression

Several shrinkage approaches are available for model selection, such as semi-Bayes shrinkage for coefficients,[148] the LASSO (least absolute shrinkage and selection operator),[149] and the elastic net.[150] These methods improve coefficients estimation or model selection through introducing some bias or setting a constraint on ordinary least squares estimates. The decreased variance outweighs introduced bias and thus theoretically improves model prediction.

The LASSO is developed to improve coefficients estimation in linear models. The positive constant imposed to the sum of the absolute values of coefficients, could shrink or render 0 some coefficients and thus select covariates into models.[149] The LASSO is useful for model selection when there are many predictors (p), such as our study where there are 329 genetic and 32 serum markers ($p = 361$). However, the LASSO does not perform superiorly when covariates are highly correlated[149] or are way more than study sample sizes ($p >> n$).[150] Specifically, the LASSO tends to select only one out of a group of highly correlated variables into model and discard others.

In our study, genetic or serum markers are presumably highly correlated as they are related to inflammation pathway, immune function, or B-cell activation. Our study also contains longitudinal measurements of 32 serum markers at three time-points (0-1, 1-3, and >3 years prior to reference date). In this situation, Zuo and Hastie in 2005 proposed a generalized form of the LASSO – the elastic net – for model fitting.[150] The elastic net preserves the characters of the LASSO and further includes the 'grouping effect' for a set of highly correlated covariates in a p

>> n situation. The elastic net is suitable for analysis in systems biology research, including analysis in pathway-based approach or microarray gene-expression data.[151, 152]

Our study implemented the network analysis to screen out potential role players among several hundreds of biomarkers and further adjusted for them in HIV or HCV models. However, several approaches are available and suitable to execute shrinkage and model selection at the same time, such as shrinkage regression including LASSO or the elastic net. Further research is worthy applying these statistical methods for future study.

APPENDIX I. SUPPLEMENTARY TABLES

Table S4.1. Spearman correlation coefficients among 10 serum inflammation-related markers (cytokines) (> 3 years preceding AIDS-NHL diagnosis) in the first module from the WGCNA, by case and control status

Cytokines	sCD27	sCD30	IP10	CXCL13	Neopterin	κ	κ: λ	IL6	IL10	SCD23	TNFα	CRP
Correlation coefficient among HIV+ controls												
sCD27	1	0.643	0.557	0.507	0.549	0.527	0.32	0.308	0.218	0.393	0.322	0.082
sCD30	0.643	1	0.43	0.458	0.386	0.252	0.156	0.239	0.253	0.407	0.302	0.11
IP10	0.557	0.43	1	0.471	0.371	0.363	0.204	0.296	0.154	0.227	0.248	0.081
CXCL13	0.507	0.458	0.471	1	0.324	0.375	0.294	0.177	0.181	0.369	0.167	0.068
Neopterin	0.549	0.386	0.371	0.324	1	0.341	0.229	0.243	0.057	0.294	0.152	0.205
κ	0.527	0.252	0.363	0.375	0.341	1	0.704	0.143	-0.035	0.257	0.016	-0.006
κ: λ	0.32	0.156	0.204	0.294	0.229	0.704	1	0.039	0.065	0.215	0.089	-0.122
IL6	0.308	0.239	0.296	0.177	0.243	0.143	0.039	1	0.168	0.188	0.147	0.442
IL10	0.218	0.253	0.154	0.181	0.057	-0.035	0.065	0.168	1	0.041	0.395	0.152
sCD23	0.393	0.407	0.227	0.369	0.294	0.257	0.215	0.188	0.041	1	0.111	-0.03
TNFα	0.322	0.302	0.248	0.167	0.152	0.016	0.089	0.147	0.395	0.111	1	0.063
CRP	0.082	0.11	0.081	0.068	0.205	-0.006	-0.122	0.442	0.152	-0.03	0.063	1
Correlation coefficient among AIDS-NHL cases												
sCD27	1	0.728	0.623	0.546	0.599	0.647	0.445	0.316	0.209	0.402	0.335	0.146
sCD30	0.728	1	0.427	0.543	0.472	0.428	0.251	0.15	0.151	0.36	0.286	0.114
IP10	0.623	0.427	1	0.451	0.556	0.452	0.263	0.319	0.066	0.25	0.277	0.249
CXCL13	0.546	0.543	0.451	1	0.407	0.295	0.12	0.223	0.201	0.24	0.142	0.182
Neopterin	0.599	0.472	0.556	0.407	1	0.419	0.214	0.24	0.149	0.398	0.279	0.255
κ	0.647	0.428	0.452	0.295	0.419	1	0.722	0.298	0.144	0.179	0.255	0.142
κ: λ	0.445	0.251	0.263	0.12	0.214	0.722	1	0.157	0.041	0.104	0.146	0.127
IL6	0.316	0.15	0.319	0.223	0.24	0.298	0.157	1	0.236	0.197	0.331	0.375
IL10	0.209	0.151	0.066	0.201	0.149	0.144	0.041	0.236	1	-0.017	0.053	0.056
sCD23	0.402	0.36	0.25	0.24	0.398	0.179	0.104	0.197	-0.017	1	0.193	0.101
TNFα	0.335	0.286	0.277	0.142	0.279	0.255	0.146	0.331	0.053	0.193	1	0.182
CRP	0.146	0.114	0.249	0.182	0.255	0.142	0.127	0.375	0.056	0.101	0.182	1

REFERENCES

1. Revision of the case definition of acquired immunodeficiency syndrome for national reporting--United States. Centers for Disease Control, Department of Health and Human Services. Ann Intern Med 1985;103(3):402-3.

2. International Collaboration on HIV and Cancer. Highly active antiretroviral therapy and incidence of cancer in human immunodeficiency virus-infected adults. J Natl Cancer Inst 2000;92(22):1823-30.

3. Engels EA, Biggar RJ, Hall HI, et al. Cancer risk in people infected with human immunodeficiency virus in the United States. Int J Cancer 2008;123(1):187-94.

4. Martinez-Maza O, Breen EC. B-cell activation and lymphoma in patients with HIV. Curr Opin Oncol 2002;14(5):528-32.

5. Gaidano G, Capello D, Carbone A. The molecular basis of acquired immunodeficiency syndrome-related lymphomagenesis. Semin Oncol 2000;27(4):431-41.

6. Grogg KL, Miller RF, Dogan A. HIV infection and lymphoma. J Clin Pathol 2007;60(12):1365-72.

7. Breen EC, Magpantay L, Jacobson LP, et al. B-cell stimulatory cytokines and markers of immune activation are elevated several years prior to the diagnosis of AIDS-associated non-Hodgkin's B cell lymphoma. Cancer Epidemiol Biomarkers Prev 2011;20(7):1303-14.

8. Vendrame E, Hussain S, Breen E, C., et al. Serum Levels of Cytokines, and Biomarkers for Inflammation and Immune Activation, and HIV-Associated Non-Hodgkin B cell Lymphoma Risk. In. Los Angeles; 2013.

9. Hussain SK, Zhu W, Chang SC, et al. Serum levels of the chemokine CXCL13, genetic variation in CXCL13 and its receptor CXCR5, and HIV-associated non-hodgkin B-cell lymphoma risk. Cancer Epidemiol Biomarkers Prev 2013;22(2):295-307.

10. Breen EC, Fatahi S, Epeldegui M, Boscardin WJ, Detels R, Martinez-Maza O. Elevated serum soluble CD30 precedes the development of AIDS-associated non-Hodgkin's B cell lymphoma. Tumour Biol 2006;27(4):187-94.

11. Purdue MP, Lan Q, Martinez-Maza O, et al. A prospective study of serum soluble CD30 concentration and risk of non-Hodgkin lymphoma. Blood 2009;114(13):2730-2.

12. Epeldegui M, Breen EC, Hung YP, Boscardin WJ, Detels R, Martinez-Maza O. Elevated expression of activation induced cytidine deaminase in peripheral blood mononuclear cells precedes AIDS-NHL diagnosis. AIDS 2007;21(17):2265-70.

13. Widney DP, Breen EC, Boscardin WJ, et al. Serum levels of the homeostatic B cell chemokine, CXCL13, are elevated during HIV infection. J Interferon Cytokine Res 2005;25(11):702-6.

14. Widney D, Gundapp G, Said JW, et al. Aberrant Expression of CD27 and Soluble CD27 (sCD27) in HIV Infection and in AIDS-Associated Lymphoma. Clin Immunol 1999;93(2):114-23.

15. Yawetz S, Cumberland WG, van der Meyden M, Martinez-Maza O. Elevated serum levels of soluble CD23 (sCD23) precede the appearance ofacquired immunodeficiency syndrome--associated non-Hodgkin's lymphoma. Blood 1995;85(7):1843-9.

16. Wong HL, Breen EC, Pfeiffer RM, et al. Cytokine signaling pathway polymorphisms and AIDS-related non-Hodgkin lymphoma risk in the multicenter AIDS cohort study. AIDS 2010;24(7):1025-33.

17. Widney DP, Gui D, Popoviciu LM, et al. Expression and Function of the Chemokine, CXCL13, and Its Receptor, CXCR5, in Aids-Associated Non-Hodgkin's Lymphoma. AIDS Res Treat 2010;2010:164586.

18. Ouedraogo DE, Makinson A, Kuster N, et al. Increased T-cell activation and Th1 cytokine concentrations prior to the diagnosis of B-cell lymphoma in HIV infected patients. J Clin Immunol 2013;33(1):22-9.

19. Rabkin CS, Engels EA, Landgren O, et al. Circulating cytokine levels, Epstein-Barr viremia, and risk of acquired immunodeficiency syndrome-related non-Hodgkin lymphoma. Am J Hematol 2011;86(10):875-8.

20. Fisher SG, Fisher RI. The epidemiology of non-Hodgkin's lymphoma. Oncogene 2004;23(38):6524-34.

21. Raphael M, Borisch B, Jaffe E. Lymphomas associated with infection by the human immune deficiency virus (HIV). In: Jaffe E, Harris N, Stein H, Vardiman J, eds. World Health Organization Classification of Tumours, Pathology, and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. Leon, France: IARC Press; 2001:260-3.

22. Carbone A, Gloghini A. AIDS-related lymphomas: from pathogenesis to pathology. Br J Haematol 2005;130(5):662-70.

23. Diamond C, Taylor TH, Aboumrad T, Anton-Culver H. Changes in acquired immunodeficiency syndrome-related non-Hodgkin lymphoma in the era of highly active antiretroviral therapy. Cancer 2006;106(1):128-35.

24. Grulich AE, Wan X, Law MG, et al. B-cell stimulation and prolonged immune deficiency are risk factors for non-Hodgkin's lymphoma in people with AIDS. AIDS 2000;14(2):133-40.

25. Stebbing J, Gazzard B, Mandalia S, et al. Antiretroviral treatment regimens and immune parameters in the prevention of systemic AIDS-related non-Hodgkin's lymphoma. J Clin Oncol 2004;22(11):2177-83.

26. Bower M, Fisher M, Hill T, et al. CD4 counts and the risk of systemic non-Hodgkin's lymphoma in individuals with HIV in the UK. Haematologica 2009;94(6):875-80.

27. Bohlius J, Schmidlin K, Costagliola D, et al. Incidence and risk factors of HIV-related non-Hodgkin's lymphoma in the era of combination antiretroviral therapy: a European multicohort study. Antivir Ther 2009;14(8):1065-74.

28. Zoufaly A, Stellbrink HJ, Heiden MA, et al. Cumulative HIV viremia during highly active antiretroviral therapy is a strong predictor of AIDS-related lymphoma. J Infect Dis 2009;200(1):79-87.

29. Kirk O, Pedersen C, Cozzi-Lepri A, et al. Non-Hodgkin lymphoma in HIV-infected patients in the era of highly active antiretroviral therapy. Blood 2001;98(12):3406-12.

30. Bonnet F, Balestre E, Thiebaut R, et al. Factors associated with the occurrence of AIDS-related non-Hodgkin lymphoma in the era of highly active antiretroviral therapy: Aquitaine Cohort, France. Clin Infect Dis 2006;42(3):411-7.

31. Thorley-Lawson DA, Allday MJ. The curious case of the tumour virus: 50 years of Burkitt's lymphoma. Nat Rev Micro 2008;6(12):913-24.

32. Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. N Engl J Med 1995;332(18):1186-91.

33. Deloose ST, Smit LA, Pals FT, Kersten MJ, van Noesel CJ, Pals ST. High incidence of Kaposi sarcoma-associated herpesvirus infection in HIV-related solid immunoblastic/plasmablastic diffuse large B-cell lymphoma. Leukemia 2005;19(5):851-5.

34. Oksenhendler E, Boulanger E, Galicier L, et al. High incidence of Kaposi sarcoma-associated herpesvirus-related non-Hodgkin lymphoma in patients with HIV infection and multicentric Castleman disease. Blood 2002;99(7):2331-6.

35. Engels EA. Infectious agents as causes of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev 2007;16(3):401-4.

36. Machida K, Cheng KT, Pavio N, Sung VM, Lai MM. Hepatitis C virus E2-CD81 interaction induces hypermutation of the immunoglobulin gene in B cells. J Virol 2005;79(13):8079-89.

37. Machida K, Cheng KT, Sung VM, et al. Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and protooncogenes. Proc Natl Acad Sci U S A 2004;101(12):4262-7.

38. Matsuo K, Kusano A, Sugumar A, Nakamura S, Tajima K, Mueller NE. Effect of hepatitis C virus infection on the risk of non-Hodgkin's lymphoma: a meta-analysis of epidemiological studies. Cancer Sci 2004;95(9):745-52.

39. Dal Maso L, Franceschi S. Hepatitis C virus and risk of lymphoma and other lymphoid neoplasms: a meta-analysis of epidemiologic studies. Cancer Epidemiol Biomarkers Prev 2006;15(11):2078-85.

40. de Sanjose S, Benavente Y, Vajdic CM, et al. Hepatitis C and non-Hodgkin lymphoma among 4784 cases and 6269 controls from the International Lymphoma Epidemiology Consortium. Clin Gastroenterol Hepatol 2008;6(4):451-8.

41. Ng LJ. Hepatitis B and C infections, single nucleotide polymorphisms of the inflammatory pathway, and the risk of non-Hodgkin’s lymphoma. Los Angeles: Univeristy of California; 2008.

42. Franceschi S, Polesel J, Rickenbach M, et al. Hepatitis C virus and non-Hodgkin's lymphoma: findings from the Swiss HIV Cohort Study. Br J Cancer 2006;95(11):1598-602.

43. Waters L, Stebbing J, Mandalia S, et al. Hepatitis C infection is not associated with systemic HIV-associated non-hodgkin's lymphoma: A cohort study. Int J Cancer 2005;116(1):161-3.

44. Engels EA, Frisch M, Lubin JH, Gail MH, Biggar RJ, Goedert JJ. Prevalence of Hepatitis C Virus Infection and Risk for Hepatocellular Carcinoma and Non-Hodgkin Lymphoma in AIDS. J Acquir Immune Defic Syndr 2002;31(5):536-41.

45. Levine AM, Nelson R, Zuckerman E, et al. Lack of association between hepatitis C infection and development of AIDS-related lymphoma. J Acquir Immune Defic Syndr Hum Retrovirol 1999;20(3):255-8.

46. Besson C, Pialoux G, Dupont B, et al. Lack of association between HCV infection and HIV-related lymphoma. Br J Haematol 1999;105(2):568-9.

47. Epeldegui M, Widney DP, Martinez-Maza O. Pathogenesis of AIDS lymphoma: role of oncogenic viruses and B cell activation-associated molecular lesions. Curr Opin Oncol 2006;18(5):444-8.

48. Piriou E, van Dort K, Nanlohy NM, van Oers MH, Miedema F, van Baarle D. Loss of EBNA1-specific memory CD4+ and CD8+ T cells in HIV-infected patients progressing to AIDS-related non-Hodgkin lymphoma. Blood 2005;106(9):3166-74.

49. Carbone A. Emerging pathways in the development of AIDS-related lymphomas. Lancet Oncol 2003;4(1):22-9.

50. Pasqualucci L, Bhagat G, Jankovic M, et al. AID is required for germinal center-derived lymphomagenesis. Nat Genet 2008;40(1):108-12.

51. Lane HC, Masur H, Edgar LC, Whalen G, Rook AH, Fauci AS. Abnormalities of B-Cell Activation and Immunoregulation in Patients with the Acquired Immunodeficiency Syndrome. N Engl J Med 1983;309(8):453-8.

52. Lane HC, Fauci AS. Immunologic abnormalities in the acquired immunodeficiency syndrome. Annu Rev Immunol 1985;3:477-500.

53. Martin G, Roy J, Barat C, Ouellet M, Gilbert C, Tremblay MJ. Human Immunodeficiency Virus Type 1-Associated CD40 Ligand Transactivates B Lymphocytes and Promotes Infection of CD4+ T Cells. J Virol 2007;81(11):5872-81.

54. He B, Qiao X, Klasse PJ, et al. HIV-1 Envelope Triggers Polyclonal Ig Class Switch Recombination through a CD40-Independent Mechanism Involving BAFF and C-Type Lectin Receptors. J Immunol 2006;176(7):3931-41.

55. Kundu RK, Sangiorgi F, Wu LY, et al. Expression of the human immunodeficiency virus-Tat gene in lymphoid tissues of transgenic mice is associated with B-cell lymphoma. Blood 1999;94(1):275-82.

56. Chirmule N, Kalyanaraman VS, Lederman S, et al. HIV-gp 160-induced T cell-dependent B cell differentiation. Role of T cell-B cell activation molecule and IL-6. J Immunol 1993;150(6):2478-86.

57. Rieckmann P, Poli G, Fox CH, Kehrl JH, Fauci AS. Recombinant gp120 specifically enhances tumor necrosis factor-alpha production and Ig secretion in B lymphocytes from HIV-infected individuals but not from seronegative donors. J Immunol 1991;147(9):2922-7.

58. Epeldegui M, Thapa DR, De la Cruz J, Kitchen S, Zack JA, Martinez-Maza O. CD40 ligand (CD154) incorporated into HIV virions induces activation-induced cytidine deaminase (AID) expression in human B lymphocytes. PLoS One 2010;5(7):e11448.

59. Crabb Breen E, van der Meijden M, Cumberland W, Kishimoto T, Detels R, Martínez-Maza O. The Development of AIDS-Associated Burkitt's/Small Noncleaved Cell Lymphoma Is Preceded by Elevated Serum Levels of Interleukin 6. Clinical Immunology 1999;92(3):293-9.

60. Breen EC, Boscardin WJ, Detels R, et al. Non-Hodgkin's B cell lymphoma in persons with acquired immunodeficiency syndrome is associated with increased serum levels of IL10, or the IL10 promoter -592 C/C genotype. Clin Immunol 2003;109(2):119-29.

61. Yawetz S, Cumberland WG, van der Meyden M, Martinez-Maza O. Elevated serum levels of soluble CD23 (sCD23) precede the appearance ofacquired immunodeficiency syndrome--associated non-Hodgkin's lymphoma. Blood 1995;85(7):1843-9.

62. Schroeder JR, Saah AJ, Hoover DR, et al. Serum soluble CD23 level correlates with subsequent development of AIDS-related non-Hodgkin's lymphoma. Cancer Epidemiol Biomarkers Prev 1999;8(11):979-84.

63. Widney D, Gundapp G, Said JW, et al. Aberrant expression of CD27 and soluble CD27 (sCD27) in HIV infection and in AIDS-associated lymphoma. Clin Immunol 1999;93(2):114-23.

64. Breen EC, Epeldegui M, Boscardin WJ, Widney DP, Detels R, MartÃ-nez-Maza O. Elevated levels of soluble CD44 precede the development of AIDS-associated non-Hodgkin's B-cell lymphoma. AIDS 2005;19(15):1711-2.

65. Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. Annu Rev Immunol 2008;26:261-92.

66. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 2000;102(5):553-63.

67. Arakawa H, Hauschild J, Buerstedde J-M. Requirement of the Activation-Induced Deaminase (AID) Gene for Immunoglobulin Gene Conversion. Science 2002;295(5558):1301-6.

68. Nagaoka H, Tran TH, Kobayashi M, Aida M, Honjo T. Preventing AID, a physiological mutator, from deleterious activation: regulation of the genomic instability that is associated with antibody diversity. Int Immunol 2010;22(4):227-35.

69. Ramiro AR, Jankovic M, Eisenreich T, et al. AID Is Required for c-myc/IgH Chromosome Translocations In Vivo. Cell 2004;118(4):431-8.

70. Schwartz RS, Lenz G, Staudt LM. Aggressive Lymphomas. N Engl J Med 2010;362(15):1417-29.

71. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature 2001;412(6844):341-6.

72. Endo Y, Marusawa H, Kinoshita K, et al. Expression of activation-induced cytidine deaminase in human hepatocytes via NF-[kappa]B signaling. Oncogene 2007;26(38):5587-95.

73. He B, Raab-Traub N, Casali P, Cerutti A. EBV-encoded latent membrane protein 1 cooperates with BAFF/BLyS and APRIL to induce T cell-independent Ig heavy chain class switching. J Immunol 2003;171(10):5215-24.

74. Epeldegui M, Hung YP, McQuay A, Ambinder RF, Martinez-Maza O. Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. Mol Immunol 2007;44(5):934-42.

75. Scott K, Adamson PJ, Barrans SL, Worrillow LJ, Willett EV, Allan JM. RAG1 and BRCA2 polymorphisms in non-Hodgkin lymphoma. Blood 2007;109(12):5522-3.

76. Shen M, Purdue MP, Kricker A, et al. Polymorphisms in DNA repair genes and risk of non-Hodgkin's lymphoma in New South Wales, Australia. Haematologica 2007;92(9):1180-5.

77. Shen M, Zheng T, Lan Q, et al. Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma among women in Connecticut. Hum Genet 2006;119(6):659-68.

78. Xu Z, Pone EJ, Al-Qahtani A, Park SR, Zan H, Casali P. Regulation of aicda expression and AID activity: relevance to somatic hypermutation and class switch DNA recombination. Crit Rev Immunol 2007;27(4):367-97.

79. Rothman N, Skibola CF, Wang SS, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. Lancet Oncol 2006;7(1):27-38.

80. Morton LM, Purdue MP, Zheng T, et al. Risk of non-Hodgkin lymphoma associated with germline variation in genes that regulate the cell cycle, apoptosis, and lymphocyte development. Cancer Epidemiol Biomarkers Prev 2009;18(4):1259-70.

81. Purdue MP, Lan Q, Kricker A, et al. Polymorphisms in immune function genes and risk of non-Hodgkin lymphoma: findings from the New South Wales non-Hodgkin Lymphoma Study. Carcinogenesis 2007;28(3):704-12.

82. Saberi Hosnijeh F, Krop EJM, Scoccianti C, et al. Plasma Cytokines and Future Risk of Non-Hodgkin Lymphoma (NHL): A Case-Control Study Nested in the Italian European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev 2010;19(6):1577-84.

83. Wang SS, Purdue MP, Cerhan JR, et al. Common gene variants in the tumor necrosis factor (TNF) and TNF receptor superfamilies and NF-kB transcription factors and non-Hodgkin lymphoma risk. PLoS One 2009;4(4):e5360.

84. Skibola CF, Bracci PM, Nieters A, et al. Tumor necrosis factor (TNF) and lymphotoxin-alpha (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph Consortium. Am J Epidemiol 2009;171(3):267-76.

85. Worrillow L, Roman E, Adamson PJ, Kane E, Allan JM, Lightfoot TJ. Polymorphisms in the nucleotide excision repair gene ERCC2/XPD and risk of non-Hodgkin lymphoma. Cancer Epidemiol 2009;33(3-4):257-60.

86. Schuetz JM, MaCarthur AC, Leach S, et al. Genetic variation in the NBS1, MRE11, RAD50 and BLM genes and susceptibility to non-Hodgkin lymphoma. BMC Med Genet 2009;10:117.

87. Baris S, Celkan T, Batar B, et al. Association between genetic polymorphism in DNA repair genes and risk of B-cell lymphoma. Pediatr Hematol Oncol 2009;26(6):467-72.

88. Liu J, Song B, Wang Z, et al. DNA repair gene XRCC1 polymorphisms and non-Hodgkin lymphoma risk in a Chinese population. Cancer Genet Cytogenet 2009;191(2):67-72.

89. 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.

90. Paz-y-Mino C, Perez JC, Fiallo BF, Leone PE. A polymorphism in the hMSH2 gene (gIVS12-6T>C) associated with non-Hodgkin lymphomas. Cancer Genet Cytogenet 2002;133(1):29-33.

91. Agnello V. The etiology and pathophysiology of mixed cryoglobulinemia secondary to hepatitis C virus infection. Springer Semin Immunopathol 1997;19(1):111-29.

92. Ito M, Murakami K, Suzuki T, et al. Enhanced expression of lymphomagenesis-related genes in peripheral blood B cells of chronic hepatitis C patients. Clinical Immunology 2010;135(3):459-65.

93. Detels R, Phair JP, Saah AJ, et al. Recent Scientific Contributions to Understanding HIV/AIDS from the Multicenter AIDS Cohort Study. J Epidemiol 1992;2(2):Suppl: 11-9.

94. Multicenter AIDS Cohort Study. (Accessed July 8, 2013, at http://www.statepi.jhsph.edu/macs/macs.html.)

95. MACS Forms. Multicenter AIDS Cohort Study. (Accessed July 8, 2013, at http://www.statepi.jhsph.edu/macs/forms.html.)

96. Detels R. Multicenter AIDS Cohort Study Dossier updates. Lecture Note. Los Angeles: University of California, Los Angeles; 2010 June 2.

97. Turner JJ, Morton LM, Linet MS, et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. Blood 2010;116(20):e90-8.

98. Breen EC, van der Meijden M, Cumberland W, Kishimoto T, Detels R, Martínez-Maza O. The Development of AIDS-Associated Burkitt's/Small Noncleaved Cell Lymphoma Is Preceded by Elevated Serum Levels of Interleukin 6. Clin Immunol 1999;92(3):293-9.

99. Kanowith-Klein S, Saxon A. Fc epsilon receptors on human cell lines and peripheral blood lymphocytes detected by binding of IgE immune complexes. J Clin Immunol 1985;5(1):38-45.

100. Zhang K, Clark EA, Saxon A. CD40 stimulation provides an IFN-gamma-independent and IL-4-dependent differentiation signal directly to human B cells for IgE production. J Immunol 1991;146(6):1836-42.

101. Altshuler D, Brooks L, Chakravarti A, Collins F, Daly M, Donnelly P. A haplotype map of the human genome. Nature 2005;437(7063):1299-320.

102. International HapMap C, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851-61.

103. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a Maximally Informative Set of Single-Nucleotide Polymorphisms for Association Analyses Using Linkage Disequilibrium. Am J Hum Genet 2004;74(1):106-20.

104. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4:Article17.

105. Hu Z, Mellor J, Wu J, Yamada T, Holloway D, DeLisi C. VisANT: data-integrating visual framework for biological networks and modules. Nucleic Acids Research 2005;33(suppl 2):W352-W7.

106. Henrard DR, Phillips JF, Muenz LR, et al. Natural history of HIV-1 cell-free viremia. JAMA 1995;274(7):554-8.

107. Pedersen C, Katzenstein T, Nielsen C, Lundgren JD, Gerstoft J. Prognostic Value of Serum HIV-RNA Levels at Virologic Steady State After Seroconversion: Relation to CD4 Cell Count and Clinical Course of Primary Infection. JAIDS Journal of Acquired Immune Deficiency Syndromes 1997;16(2):93-9.

108. Feinberg MB. Changing the natural history of HIV disease. The Lancet 1996;348(9022):239-46.

109. Mellors JW, Munoz A, Giorgi JV, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med 1997;126(12):946-54.

110. Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 1996;272(5265):1167-70.

111. Chao C, Jacobson L, Tashkin D, et al. Recreational amphetamine use and risk of HIV-related non-Hodgkin lymphoma. Cancer Causes Control 2009;20(5):509-16.

112. Falade-Nwulia O, Seaberg E, Rinaldo C, Badri S, Witt M, Thio C. Comparative risk of liver-related mortality from chronic hepatitis B versus chronic hepatitis C virus infection. Clin Infect Dis 2012;55(4):507-13.

113. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. Am J Epidemiol 1993;138(11):923-36.

114. Desai M, Kubo J, Esserman D, Terry MB. The handling of missing data in molecular epidemiology studies. Cancer Epidemiol Biomarkers Prev 2011;20(8):1571-9.

115. Little R, Rubin D. Statistical analysis with missing data. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2002.

116. Greenland S, Rothman K. Fundamentals of epidemiologic data analysis. In: Rothman K, Greenland S, Lash T, eds. Modern Epidemiology. 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008:219.

117. Wang L, Tang H, Thayanithy V, et al. Gene networks and microRNAs implicated in aggressive prostate cancer. Cancer Res 2009;69(24):9490-7.

118. Clarke C, Madden SF, Doolan P, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. Carcinogenesis 2013.

119. Ivliev AE, t Hoen PA, Sergeeva MG. Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. Cancer Res 2010;70(24):10060-70.

120. Horvath S, Zhang B, Carlson M, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc Natl Acad Sci U S A 2006;103(46):17402-7.

121. Amft N, Curnow SJ, Scheel-Toellner D, et al. Ectopic expression of the B cell-attracting chemokine BCA-1 (CXCL13) on endothelial cells and within lymphoid follicles contributes to the establishment of germinal center-like structures in Sjogren's syndrome. Arthritis Rheum 2001;44(11):2633-41.

122. Forster R, Schweigard G, Johann S, et al. Abnormal expression of the B-cell homing chemokine receptor BLR1 during the progression of acquired immunodeficiency syndrome. Blood 1997;90(2):520-5.

123. Cagigi A, Mowafi F, Phuong Dang LV, et al. Altered expression of the receptor-ligand pair CXCR5/CXCL13 in B cells during chronic HIV-1 infection. Blood 2008;112(12):4401-10.

124. Agematsu K, Kobata T, Yang FC, et al. CD27/CD70 interaction directly drives B cell IgG and IgM synthesis. Eur J Immunol 1995;25(10):2825-9.

125. Nagumo H, Agematsu K, Shinozaki K, et al. CD27/CD70 interaction augments IgE secretion by promoting the differentiation of memory B cells into plasma cells. J Immunol 1998;161(12):6496-502.

126. Croft M. Co-stimulatory members of the TNFR family: keys to effective T-cell immunity? Nat Rev Immunol 2003;3(8):609-20.

127. De Milito A, Aleman S, Marenzi R, et al. Plasma levels of soluble CD27: a simple marker to monitor immune activation during potent antiretroviral therapy in HIV-1-infected subjects. Clin Exp Immunol 2002;127(3):486-94.

128. Pizzolo G, Vinante F, Morosato L, et al. High serum level of the soluble form of CD30 molecule in the early phase of HIV-1 infection as an independent predictor of progression to AIDS. AIDS 1994;8(6):741-5.

129. De Roos AJ, Mirick DK, Edlefsen KL, et al. Markers of B-cell activation in relation to risk of non-Hodgkin lymphoma. Cancer Res 2012;72(18):4733-43.

130. van Oers MH, Pals ST, Evers LM, et al. Expression and release of CD27 in human B-cell malignancies. Blood 1993;82(11):3430-6.

131. Kersten MJ, Evers LM, Dellemijn PL, et al. Elevation of cerebrospinal fluid soluble CD27 levels in patients with meningeal localization of lymphoid malignancies. Blood 1996;87(5):1985-9.

132. Reibnegger G, Fuchs D, Fuith LC, et al. Neopterin as a marker for activated cell-mediated immunity: application in malignant disease. Cancer Detect Prev 1991;15(6):483-90.

133. Sucher R, Schroecksnadel K, Weiss G, Margreiter R, Fuchs D, Brandacher G. Neopterin, a prognostic marker in human malignancies. Cancer Letters 2010;287(1):13-22.

134. Fuchs D, Hausen A, Reibnegger G, Werner ER, Dierich MP, Wachter H. Neopterin as a marker for activated cell-mediated immunity: Application in HIV infection. Immunology Today 1988;9(5):150-5.

135. Nyamweya S, Townend J, Zaman A, et al. Are plasma biomarkers of immune activation predictive of HIV progression: a longitudinal comparison and analyses in HIV-1 and HIV-2 infections? PLoS One 2012;7(9):e44411.

136. Guiguet M, Boué F, Cadranel J, Lang J-M, Rosenthal E, Costagliola D. Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. Lancet Oncol 2009;10(12):1152-9.

137. Engels EA, Pfeiffer RM, Landgren O, Moore RD. Immunologic and virologic predictors of AIDS-related non-hodgkin lymphoma in the highly active antiretroviral therapy era. J Acquir Immune Defic Syndr 2010;54(1):78-84.

138. Bruyand M, Thiebaut R, Lawson-Ayayi S, et al. Role of uncontrolled HIV RNA level and immunodeficiency in the occurrence of malignancy in HIV-infected patients during the combination antiretroviral therapy era: Agence Nationale de Recherche sur le Sida (ANRS) CO3 Aquitaine Cohort. Clin Infect Dis 2009;49(7):1109-16.

139. Cole SR, Napravnik S, Mugavero MJ, Lau B, Eron JJ, Saag MS. Copy-Years Viremia as a Measure of Cumulative Human Immunodeficiency Virus Viral Burden. American Journal of Epidemiology 2010;171(2):198-205.

140. Mellors JW, Margolick JB, Phair JP, et al. Prognostic value of HIV-1 RNA, CD4 cell count, and CD4 Cell count slope for progression to AIDS and death in untreated HIV-1 infection. JAMA 2007;297(21):2349-50.

141. Marconi VC, Grandits G, Okulicz JF, et al. Cumulative viral load and virologic decay patterns after antiretroviral therapy in HIV-infected subjects influence CD4 recovery and AIDS. PLoS One 2011;6(5):e17956.

142. Nieters A, Kallinowski B, Brennan P, et al. Hepatitis C and risk of lymphoma: results of the European multicenter case-control study EPILYMPH. Gastroenterology 2006;131(6):1879-86.

143. Mele A, Pulsoni A, Bianco E, et al. Hepatitis C virus and B-cell non-Hodgkin lymphomas: an Italian multicenter case-control study. Blood 2003;102(3):996-9.

144. Rahman S, Connolly JE, Manuel SL, Chehimi J, Montaner LJ, Jain P. Unique Cytokine/Chemokine Signatures for HIV-1 and HCV Mono-infection versus Co-infection as Determined by the Luminex® Analyses. J Clin Cell Immunol 2011;2(1):104.

145. van der Helm JJ, Prins M, del Amo J, et al. The hepatitis C epidemic among HIV-positive MSM: incidence estimates from 1990 to 2007. AIDS 2011;25(8):1083-91.

146. van de Laar T, Pybus O, Bruisten S, et al. Evidence of a large, international network of HCV transmission in HIV-positive men who have sex with men. Gastroenterology 2009;136(5):1609-17.

147. van de Laar TJW, Matthews GV, Prins M, Danta M. Acute hepatitis C in HIV-infected men who have sex with men: an emerging sexually transmitted infection. AIDS 2010;24(12):1799-812.

148. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. International Journal of Epidemiology 2007;36(1):195-202.

149. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological) 1996;58(1):267-88.

150. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005;67(2):301-20.

151. Garnett MJ, Edelman EJ, Heidorn SJ, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 2012;483(7391):570-5.

152. Yoshihara K, Tsunoda T, Shigemizu D, et al. High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. Clin Cancer Res 2012;18(5):1374-85.