Title

MicroRNA-related Polymorphisms and Non-Hodgkin Lymphoma Susceptibility in the Multicenter AIDS Cohort Study

Author

Peckham, Erin Christine

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

MicroRNA-related Polymorphisms and Non-Hodgkin Lymphoma Susceptibility in the Multicenter AIDS Cohort Study

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Erin Christine Peckham

2014

ABSTRACT OF THE DISSERTATION

MicroRNA-related Polymorphisms and Non-Hodgkin Lymphoma Susceptibility in the Multicenter AIDS Cohort Study

by

Erin Christine Peckham

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2014

Professor Shehnaz K. Hussain, Co-Chair

Professor Zuo-Feng Zhang, Co-Chair

Background: MicroRNAs (miRNA) regulate gene expression through binding to the 3' untranslated region of target messenger RNA (mRNA) and have been implicated in lymphomagenesis. Only a few studies have been conducted to evaluate the effects of miRNA-related genes. In this study, we will test hypotheses to determine whether single nucleotide polymorphisms (SNPs) located within miRNA binding sites, miRNA coding regions, or miRNA processing genes, (collectively referred to as miRNA-related SNPs), are associated with susceptibility of AIDS-associated non-Hodgkin lymphoma (AIDS-NHL) among men from the Multicenter AIDS Cohort Study (MACS).

Methods: We genotyped 25 miRNA-related SNPs in 180 AIDS-NHL cases and 529 matched HIV-infected controls. Adjusted odds ratios (OR_{adj}) and 95% confidence intervals (CIs) were

calculated using conditional logistic regression, controlling for age, $CD4^+$ T-cell count, HIV viral load, prior AIDS diagnosis, history of anti-retroviral therapy, race, and history of HCV infection, to estimate the associations between miRNA-associated SNPs and overall, systemic and central nervous system (CNS) AIDS-NHL. The semi-Bayes (SB) approach was used to correct for multiple comparisons. Through the incorporation of a normal prior of null-association (prior odds ratio, OR_{prior} = 1.00; 95% prior limits: 0.25, 4.00), we proposed to pull observed associations toward the null, emphasizing our uncertainty. To give a comprehensive overview of our data and results, we presented both suggested (alpha=0.10) and statistically significant results (alpha=0.05), while noting those that remained evident after SB correction. We did place highlighted emphasis on significant results which remained evident after the SB approach as we were more confident in the report of these results (*GEMIN3* rs197412, *miR-196a* rs11614913, and *HIF1A* rs2057482). Further, results were focused to those from analyses imputed for HIV viral load at set point and $CD4^+$ T-cell count at reference date. Last, we estimated the association between SNPs within miRNA processing genes and mean $\log_e$ miRNA serum levels among a subgroup of 77 MACS participants, comprised of 61 AIDS-NHL cases and 16 HIV-infected controls. Mean ratios (MR_{adj}) and 95% CIs were calculated using linear regression, controlling for AIDS-NHL case-control status, $CD4^+$ T-cell count at date of serum sample, and race.

Results: *GEMIN3* rs197412, a non-synonymous SNP within a gene involved in miRNA processing and maturation, was associated with AIDS-NHL overall (OR_{adj}= 1.35 per variant allele; 95% CI: 1.03-1.78). In subgroup analyses, an inverse association was observed between *miR-196a* rs11614913 and CNS HIV/AIDS NHL (CT/TT vs. CC OR_{adj}= 0.43; 95% CI: 0.22-0.87). The variant allele of *HIF1A* rs2057482 was associated with increased risk of systemic AIDS-NHL (OR_{adj}= 1.83 per variant allele; 95% CI: 1.16-2.90), whereas the same allele was

associated with decreased risk of CNS AIDS-NHL (OR_{adj} = 0.38 per variant allele; 95% CI: 0.17-0.87). These miRNA-related SNP associations with AIDS-NHL remained evident after SB correction. In the evaluation of genotype-phenotype associations, there was a suggested association between carriership of the variant allele (C) of *GEMIN3* rs197412 with higher miRNA-222 serum levels compared to those with the referent genotype (MR_{adj} = 1.21 per variant allele; 95% CI: 0.98-1.49), suggesting a potential functional effect of *GEMIN3* rs197412 in relation to miRNA-222 expression.

Conclusion: These results suggest that miRNA-related SNPs are associated with AIDS-NHL susceptibility. Specifically, our study demonstrated that *GEMIN3* rs197412 was associated with AIDS-NHL susceptibility, and that genetic variation within this miRNA processing gene may influence subsequent miRNA expression. Further, *miR-196a* rs11614913 and *HIF1A* rs2057482 AIDS-NHL were also associated with AIDS-NHL, highlighting the diverse array of biologic mechanisms underlying AIDS-lymphomagenesis.

The dissertation of Erin Christine Peckham is approved.

Abdelmonem Afifi

Roger Detels

Otoniel Martínez-Maza

Shehnaz K. Hussain, Committee Co-Chair

Zuo-Feng Zhang, Committee Co-Chair

University of California, Los Angeles

2014

DEDICATION

In loving memory of my father Samuel Charles Peckham, my step-father David Scott Pickard, and my first public health mentor, Dr. E. Richard Brown.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES XII

ACKNOWLEDGEMENTS XV

VITA XVIII

CHAPTER 1: INTRODUCTION AND BACKGROUND 1

1.1. Non-Hodgkin Lymphoma 1

1.1.1. Definition and WHO Categorization 1

1.1.2. Epidemiology of Non-Hodgkin Lymphoma 2

1.1.3. Risk Factors for Non-Hodgkin Lymphoma 6

1.1.4. Pathogenesis Overview 10

1.1.5. Pathogenesis Specific to the HIV/AIDS Population 11

1.2. microRNA 13

1.2.1. Overview 13

1.2.2. miRNA Biogenesis, Target Regulation, and SNP-Interaction 14

1.2.3. miRNA and Cancer 17

1.2.4. microRNA and Immune Function and B-cell development 20

1.2.5. microRNA and Non-Hodgkin Lymphoma 21

1.2.6. microRNAs and miRNA-related SNPs as Potential Biomarkers of Cancer Development 23

1.3. Gaps in the Literature 24

CHAPTER 2: RESEARCH OBJECTIVES AND METHODOLOGY 26

2.1. Research Objectives 26

2.2. Specific Aims and Hypotheses 26

2.3. Study Design and Methods 28

2.3.1. Study Overview 28

2.4. Study Population 29

2.4.1. Case Definition and Reference Date 29

2.4.2. Control Definition, Reference Date and Matching Criteria 30

2.4.3. Data Collection 30

2.5. miRNA-related SNP Selection (Specific Aim 1) 31

2.5.1. Detailed miRNA Selection (Step 1) 32

2.5.2. Detailed Target Gene Selection (Step 2) 33

2.5.3. SNP Selection 36

2.6. Genotyping 37

2.7. Quality Control Measures 38

2.7.1. Post-genotyping SNP and Sample Inclusion and Exclusion Criteria 38

2.7.2. Initial Measures 39

2.7.3. Coriell Sample Concordance 40

2.7.4. Call Rate Determination 40

2.7.5. Sample Coverage 40
2.7.6. Hardy-Weinberg Equilibrium 41
2.7.7. Linkage Disequilibrium 41
2.7.8. Duplicate Sample Concordance 41
2.7.9. QC Conclusions and Summary of Data for Analysis 42

2.8. Statistical Analysis 42
2.8.1. SNP Frequencies 42
2.8.2. Description of Study Variables 43
2.8.3. Longitudinal Covariate Data and Variable Creation 43
2.8.4. Covariate Selection 45
2.8.5. Statistical Analyses and Model Decisions 47
2.8.6. Sensitivity Analyses 47
2.8.7. Missing Data Issues 49
2.8.8. Interpretation of Results Given Missing Data 51
2.8.9. Semi-Bayesian Approach for Multiple Comparisons 53

2.9. Investigation of the relationship between miRNA-related SNPs and mean $\log_e$ serum levels of four miRNAs established as biomarkers of HIV progression and AIDS-NHL susceptibility (Specific Aim 2) 59
2.9.1. Study Background 60
2.9.2. Analytic Plan 61

CHAPTER 3: RESULTS 70

3.1. Genotyped SNPs 70

3.2. AIDS-NHL Case and HIV Positive Control Demographics 70

3.3. microRNA-related SNPs and AIDS-NHL Susceptibility 72

3.3.1. SNPs and Overall, Systemic and CNS AIDS-NHL Susceptibility 72

3.3.2. Association of microRNA-196a rs11614913 and HIF1A rs2057482 on Overall, Systemic and CNS AIDS-NHL 74

3.3.3. Subgroup analyses: SNPs and Overall AIDS-NHL among Caucasians and those without a Primary Kaposi's Sarcoma Diagnosis 75

3.3.4. Investigation of Study Result Robustness 76

3.4. Associations between SNPs in miRNA Processing Genes and miRNA-21, miRNA-122, miRNA-222, and miRNA-223 serum levels (Specific Aim 2) 77

CHAPTER 4: DISCUSSION 116

4.1 Highlighted Result 116

4.2. miRNA-Related SNPs and AIDS-NHL Susceptibility 117

4.3. SNPs within miRNA processing genes and Differential miRNA Serum Levels 125

4.4. Strengths and Limitations 128

CHAPTER 5: CONCLUSION AND PUBLIC HEALTH IMPLICATIONS 133

CHAPTER 6: FUTURE RESEARCH 135

APPENDIX 137

REFERENCES 144

LIST OF FIGURES AND TABLES

Figure 2.1 The Process from miRNA and Candidate Gene Selection, to Target Gene Determination and SNP Identification 63

Figure 2.2 Comparison of miRNA serum levels to natural log transformed miRNA serum levels to investigate normality 66

Figure 3.1 Overview of log-additive p-values for AIDS-NHL Overall 89

Figure 3.2 Overview of log-additive p-values for Systemic AIDS-NHL 90

Figure 3.3 Overview of log-additive p-values for CNS AIDS-NHL 91

Figure 3.4 Mean miR-21 serum levels by Gemin3 rs197412 111

Figure 3.5 Mean miR-222 serum levels by Gemin3 rs197412 112

Figure 3.6 Mean miR-223 serum levels by Gemin3 rs197412 113

Figure 3.7 Mean miR-21 serum levels by RAN rs14035 114

Figure 3.8 Mean miR-223 serum levels by XPO5 rs11077 115

Table 1.1 Characteristics of NHLs that occur in the general and HIV-infected populations 25

Table 2.1 SNPs genotyped with post-genotyping study MAF and Genotype Distributions 64

Table 3.1 Distribution of SNPs among HIV-NHL Cases and HIV+ Controls 81

Table 3.2 Demographic Characteristics of AIDS-NHL Cases and HIV+ Controls in the MACS 84

Table 3.3 Comparison of Demographic Characteristics of Cohort 1 versus Cohort 2 to Inform Result Interpretation..... 88

Table 3.4 Association of miRNA-related SNPs and overall, systemic or CNS AIDS-NHL 92

Table 3.5 Association of microRNA-196a rs11614913 and HIF1A rs2057482 on Overall, Systemic and CNS AIDS-NHL, under a dominant model using data imputed for HIV Viral Load and CD4+ T-cell count at reference date 97

Table 3.6 Association of miRNA-related SNPs and Overall AIDS-NHL Among Those without a Primary KS diagnosis, and Among Caucasian-Only Participants 98

Table 3.7 Display of original statistically significant estimates alongside semi-bayes corrected posterior estimates to highlight results that remained evident after SB correction, listed by table number, data source (imputed dataset or complete case data), and NHL site 101

Table 3.8 Association of miRNA-related SNPs and Overall AIDS-NHL, using Unconditional Logistic Regression..... 103

Table 3.9. Investigating Model and Highlighted Result Robustness, miR-related SNPs and Overall AIDS-NHL..... 106

Table 3.10. Association of SNPs within miRNA Processing Genes and Pre-Reference Date miRNA serum levels among subgroup of 77 study participants (n=61 AIDS-NHL and 16 controls), Using Linear Regression..... 107

Table 3.11. Association of SNPs within miRNA Processing Genes and Pre-Reference Date miRNA serum levels among subgroup of 77 study participants (n=61 AIDS-NHL and 16 controls), Using Linear Regression, stratified by case-control status..... 109

ACKNOWLEDGEMENTS

This academic journey would not have been possible without the continued support of my committee members, all long-standing innovators and inspiring contributors to the field of public health—Dr. Abdelmonem Afifi, Dr. Roger Detels, and Dr. Otoniel Martínez-Maza. I am grateful for their gracious natures, time, fruitful conversation, expertise and guidance throughout this process.

To say that I am forever indebted to my Co-Chairs, Drs. Zuo-Feng Zhang and Shehnaz K. Hussain, is an understatement. Their guidance, intellectual support, and expertise have been the key factors in the execution of this dissertation. Working with them both has truly been a gift for which I am exceedingly appreciative, and their instrumental role in my professional and personal transition throughout this period will not be forgotten.

Dr. Zhang guided my preparation and supported my acceptance into this program; a program which has truly expanded my skill-set as an epidemiologist. Throughout the entirety of this program, Dr. Zhang guided me with extreme wisdom, always fostered the growth of my personal and academic excellence, and taught me many things about both personal and professional life. I am forever thankful for his intellectual support, for his thoughtful remarks, and for the care to which he gives all his students.

Dr. Hussain has truly inspired both professional and personal growth. On top of devoting countless hours to the development of this project, she has taken the time to guide my academic progress and personal development with grace. Most importantly, I am forever thankful for the kind way in which she always challenged me to pursue my academic best. Years from now, I

will continue be thankful for the extreme privilege to learn from, and work with, such an innovative, sharp and exemplary molecular epidemiologist.

Dr. Marc Strassburg gave me the opportunity to get my feet wet as a future educator, and I never could have imagined how much I would learn from him, or how dear to my heart he would become.

The exceptional faculty and staff at the UCLA Center for Health Policy Research gave me my start in public health eleven years ago and helped me develop personally and professionally during my first eight years at UCLA. Working with Dr. E. Richard Brown will forever be a highlight in my academic career. However, getting to know such an incredible human being, and inspirational and insightful man on a personal level was more than a gift. Dr. Brown was a man who truly lived and breathed public health promotion, and it is my hope that his mentorship will continue to shine through in my future work. I thank Celeste Maglan for giving a seventeen-year old girl from Big Bear the amazing opportunity that was my first job in public health—little did we know how far the position would take me.

Eternal thanks to my colleagues for their contributions, thoughtful conversations, shared triumphs and daily support: Shen-Chih Chang, Po-yin Chang, Solomon Makgoeng, Ying Yang, Travis Meyers, Fayth Miles, Esther Chang, Claire Kim, Dan Keebler, Aileen Baecker, Gina Wallar, Heather Tarleton, Fong Liu and Weiming Zhu.

Whenever I have needed them, they have been there! The administrative staff of the UCLA department of Epidemiology, especially Josie Wei, Kathy Bonfire, Joy Miller, and Barbara Cooke, all have been exceptionally supportive.

A special thank you to the MACS participants whose contribution and dedication made this scientific work possible, and extreme gratitude to our funding sources for their continued

support of our scientific endeavors and educational pursuits: NIH/NCI T32 Cancer Epidemiology Training Program (T32CA009142), and the UCLA AIDS Institute (FDP-NIH AI028697).

Finally, there is no way to ever fully thank the small army of amazing individuals who provided continual support of both professional and personal endeavors over the course of this academic journey: my mother Teresa, sister Anna, brother Micah, Aunt Connie, Uncle Larry, Grandmother Angie, Grandfather Elwyn, the entire Gregory family, Tom and Kathy Dell, Marty, Jami, Kyndall and the Bond family, Dan, the Hughes, Charlene, Cindi, Puspen, Babu, and Laura. To those whom I have lost along the way, especially my Dads, Nana, and mentors Rick and Lana, I am forever thankful for how they taught me to celebrate life and the importance of small and random acts of kindness—lessons which have paved a smooth path for my time here at UCLA. Last, I am forever thankful for the joy and love that my fiancé Shaun has infused into my life and especially for his immense support, unwavering confidence and encouragement, and peaceful nature during this process.

VITA

2003 – 2007	University of California at Los Angeles, USA Bachelor of Science, Ecology and Evolutionary Biology
2003 – 2010	Research Assistant UCLA Center for Health Policy Research
2007 – 2009	Graduate School of Public Health, UCLA, Ca, USA MPH in Epidemiology
2009	PhD program in Epidemiology at UCLA School of Public Health, Department of Epidemiology
2012 – 2014	Teaching Assistant UCLA Executive Master's in Public Health Program EPID100 "Introduction to Epidemiology"
2013 – 2014	UCLA Graduate Division Dissertation Year Fellowship
2014	American Society of Preventive Oncology Annual Meeting, New Investigators Workshop Invitee

PUBLICATIONS AND PRESENTATIONS

Peckham E (2014). "microRNA-related polymorphisms and Non-Hodgkin Lymphoma Susceptibility in an HIV/AIDS Cohort." Oral presentation at The American Society of Preventive Oncology Annual Meeting, New Investigators Workshop. Arlington, Virginia, March.

Wyn R and **Peckham E** (2009). "Health Disparities Among California's Nearly Four Million Low-Income Nonelderly Adult Women." Oral presentation at The California Wellness Foundation's Conference on Women's Health. San Francisco, CA, October.

Wyn R and **Peckham E** (2008). "Women's Health in California: Health Status, Health Behaviors, Health Insurance Coverage and Use of Services Among California Women Ages 18-64." Oral presentation at The California Wellness Foundation's Conference on Women's Health. Los Angeles, CA, October.

Lavarreda SA and **Peckham EC** (2008). "Discontinuous Medi-Cal and Healthy Families and its Impact on Low-Income Children's Access to Care." Oral presentation at the American Public Health Association Annual Meeting. San Diego, CA, October.

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1. Non-Hodgkin Lymphoma

1.1.1. Definition and WHO Categorization

Non-Hodgkin Lymphoma is the most common hematological cancer in adults worldwide, and includes a heterogeneous group of malignancies that are overall defined as cancers of the lymphocytes.[1] These malignancies differ in several components, including etiologies, aggressiveness, treatment responsiveness, and patient survival.[1] Non-Hodgkin lymphomas arise from either T- or B-cells, with approximately 90% of tumors being of B-cell origin and 6.8% being of T-cell origin.[2] Common subtypes of B-cell lymphoma in the general population include diffuse large B-cell lymphoma (DLBCL; ~23.1% of all NHLs) and follicular lymphoma (~12% of all NHLs), accounting for nearly 35% of all non-Hodgkin lymphoma cases.[2]

The World Health Organization (WHO) categorizes non-Hodgkin lymphomas according to their incidence in immunocompetent and immunocompromised individuals, such as those with HIV/AIDS.[2] The first category includes NHLs that occur in both immunocompetent and immunocompromised individuals, such as AIDS-related non-Hodgkin lymphomas (AIDS-NHLs). These AIDS-NHLs are primarily composed of three high-grade B-cell malignancies: Burkitt lymphoma (30% of AIDS-NHLs), DLBCL with immunoblastic features (~10% of AIDS-NHLs), and DLBCL with centroblastic features (~25% of AIDS-NHLs; Table 1.1).[3, 4] Furthermore, in HIV-infected patients, the majority of these lymphomas occur systemically (80%), while the remainder occur at extranodal sites, predominantly the primary central nervous system (CNS).[5] The second category includes NHLs that occur specifically in immunocompromised, HIV-positive individuals, such as plasmablastic lymphoma of the oral

cavity and primary effusion lymphoma.[6] The final WHO category includes polymorphic B-cell lymphoma which occurs in individuals with other immunocompromising situations.[7] WHO recognizes that the genetic profiles of these subcategories of lymphomas are vastly different, thus influencing diagnostic, prognostic and treatment options (Table 1.1).

1.1.2. Epidemiology of Non-Hodgkin Lymphoma

Incidence rates of non-Hodgkin lymphoma are higher in the United States compared to other countries, with strikingly the highest rate for men worldwide.[8] According to the American Cancer Society's Cancer Facts and Figures 2014, of the 1,665,540 new cancer cases to be diagnosed in the United States in 2014, 4.3% will be attributable to non-Hodgkin lymphoma in both sexes (70,800). Among both men and women separately, non-Hodgkin lymphoma is the sixth most common cancer, with 38,270 new NHL cases in men and 32,530 new NHL cases in women predicted to occur during 2014.[9] In regards to mortality, 3.2% of all cancer deaths in each sex are attributable to NHL. Nearly 20,000 NHL related deaths are expected to occur in 2014, with 10,470 deaths occurring in men and 8,520 deaths occurring in women. According to the American Cancer Society, overall incidence rates in men have remained stable since 1991, however in women NHL has steadily increased by 1.1% per year over the past twenty years. In regards to survival, rates for non-Hodgkin lymphoma include an 81% one-year survival rate, 69% five-year survival rate and a 58% ten-year survival rate.[9] However, these survival rates, like those of other cancers, are dependent on NHL subtype and stage at diagnosis.

Worldwide, the incidence of non-Hodgkin lymphoma has been rising steadily over the past decades.[10] According to Globocan 2012, the age-standardized incidence rates of non-Hodgkin lymphoma in men and women in developed countries were 10.3 and 7.1 (per 100,000 per year), while in less developed countries the age-standardized incidence rates were 4.3 and 2.8

(per 100,000 per year), indicating that NHL occurs more frequently in developed counties.[11] Furthermore, the United States has the highest incidence rate of non-Hodgkin lymphoma worldwide, with European countries close behind. In Asia, incidence rates are much less and most cases are associated with Epstein-Barr virus (EBV) infection, with an age-standardized incidence rate of nearly 4.3 per 100,000 men per year according to Globocan 2012. In Africa, reliable estimates of NHL incidence are difficult to obtain due to varying diagnostic procedures and lower-quality record keeping. However, in 2012 it was estimated that the age-standardized incidence rate was nearly 5.5 per 100,000 per year in Southern African men, with a 2.3-3.0 per 100,000 per year age-standardized incidence rate among men in the Ivory Coast, Mali, Gambia and Guinea regions.

In regards to age, non-Hodgkin lymphoma incidence increases in both genders and worldwide with increasing age. According to the Surveillance Epidemiology and End Results (SEER) factsheet, the median age at diagnosis between 2007-2011 in the United States was 66 years of age, with most individuals diagnosed between the ages of 65 and 74 (23.1%). In this same time period, only 1.6% were diagnosed under the age of 20, 3.8% between the ages of 20 and 34, jumping up to 13.2% between the ages of 45 and 54, and 20.7% of individuals diagnosed between the ages of 55 and 64.[12]

In the United States specifically, but also a trend noted worldwide, Caucasian men experience both higher incidence and mortality rates compared to other racial and ethnic groups. In fact, non-Hodgkin cases—with the exception of peripheral T-cell NHL—occur 40-70% more in United States Caucasians compared to individuals of African American decent.[13] Asian/Pacific Islanders, American Indian/Alaska Natives and Hispanics all have lower incidence rates as well when compared to either Caucasians or African Americans.

In the HIV/AIDS population specifically, where 25-30% of deaths are attributable to cancer and non-Hodgkin lymphoma is an AIDS-defining illness, the risk of NHL is increased 60-200 fold compared to the risk observed in the general population.[14] In the United States, the incidence rate of NHL occurring in individuals who have transitioned to AIDS is estimated to be 109 cases per 100,000 person-years, with a standardized incidence ratio of 7.3.[15] Furthermore, the relative risk when comparing incidence before and after the transition from HIV to AIDS is striking. The incidence of NHL in individuals who prior to diagnosis had not yet transitioned to AIDS is estimated to be 82 cases per 100,000 person-years, while in persons who progressed to AIDS prior to cancer diagnosis the incidence is estimated to be 349 cases per 100,000 person-years—indicating the even greater increase in AIDS-NHL risk among severely immunocompromised individuals.[15] This translates into a relative risk (AIDS vs. HIV+) of 4.3, and standardized incidence ratios of 5.6 and 20.8 between those without a prior AIDS diagnosis and those with an AIDS diagnosis prior to NHL detection, respectively.[15]

Trends in NHL: From Prior to the AIDS Epidemic to the HAART era

Between the 1950's and 70's, U.S. NHL mortality rates were associated with years of education and family income as little else was understood regarding NHL etiology.[14] During the 1970's, both the age-adjusted incidence and mortality rates of NHL increased by ~4%, and continued to increase until they stabilized in about 1990.[14] In the pre-HIV era, exposures including pesticides and both identified and unidentified infections were suggested to contribute to the increase.[14] However, given that these exposures would not have been common in the general population, other risk factors including sun exposure, diet, genetic, and lifestyle factors may have contributed. After the HIV/AIDS epidemic began, the geographic distribution of NHL remained consistent with the geographic spread of HIV/AIDS over time.[14] According to SEER

results in white males, incidence increased gradually from 1973 to 1983, then peaked rapidly in 1995.[16] Incidence that year was highest in regions with higher AIDS prevalence, including San Francisco and Seattle, compared to regions with lower AIDS prevalence (such as Utah and Iowa).[14, 16] However, despite these parallel distributions, researchers have sought to separate NHL trends between HIV-infected and -uninfected populations.[14] Over the course of approximately the last decade, overall NHL rates in the United States (U.S.) have stabilized to a degree in both populations.[17] In the U.S. during the period of 1992-2009, overall NHL rates increased 0.3 percent per year. However, among uninfected individuals from 1992-2003, NHL rates increased by 1.4% per year, and then flattened, remaining stable.[17] This indicates that NHL rates have stabilized in HIV-uninfected individuals in the U.S. since the early 2000's, and that researchers are able to begin separating out NHL trends between HIV-infected and -uninfected individuals.[17]

Among HIV-infected individuals, NHL previously accounted for three percent of all AIDS-defining conditions. However, this proportion began declining around 1993, when the AIDS definition was revised to include low $CD4^+$ T-cell count and the number of reportable AIDS cases increased.[16] The introduction of HAART in 1996 also changed the epidemiology of HIV/AIDS.[18] Four studies reported reduced annual NHL incidence by about 50% in the HAART era compared to the pre-HAART era.[16, 18] A study of SEER data reported decreased NHL incidence for AIDS-related subtypes (immunoblastic, Burkitt, and PCNS lymphoma), and an international meta-analysis reported similar decreases except for Burkitt lymphoma.[17] Several reports state conflicting trends, such as increasing incidence in the Multi-Ethnic AIDS Cohort from 1996-7, and no change at all in the Swiss HIV cohort.[16] However, short follow-up and/or few cases may have limited these studies. Reports are more certain that AIDS-NHL prognosis

has improved in the HAART era, especially due to higher $CD4^+$ T-cell counts imparted by anti-retroviral treatment allowing patients to tolerate chemotherapy, as well as due to improved cancer therapies.[19, 20] According to a 2002 MACS study, HAART was associated with an 84% reduction in mortality after NHL diagnosis, and a 2010 study re-emphasized a 77% decrease in NHL incidence among MACS participants in the post-HAART era.[21] However, when compared to HIV-uninfected MACS participants in the HAART era, the incidence of NHL was still substantially increased among HIV-infected individuals (IRR of 11.18), however this increase in risk was diminished compared to the risk previously observed in HIV-infected participants prior to wide-spread HAART treatment.[21] As NHL risk is still increased in individuals infected with HIV-1 compared to the general population despite these advances in treatment options, the need for better preventive and control measures, along with continued elucidation of NHL etiological factors, is clear.

1.1.3. Risk Factors for Non-Hodgkin Lymphoma

Overall, the greatest risk factor for NHL is long-standing immunodeficiency, as reflected in the substantial increase in NHL risk observed among people with HIV/AIDS or those who have experienced an organ transplant and immune suppressive therapy.[22] While immunodeficiency is the main risk factor for this malignancy, there are several additional risk factors identified in both the HIV-infected and -uninfected populations. For example, as with most cancers, increasing age is a risk factor for NHL development, with a specific increase in individuals who are over 50 years of age in the HIV-uninfected population.[14] This risk factor is also of concern to those with HIV, as due to expanded treatment options, these individuals are living longer and thus extending their period at risk for NHL development. Thus, as individuals infected with HIV continue to age, their risk for developing AIDS-NHL will also increase.[23] In

both HIV-infected and –uninfected populations, males experience a higher risk of NHL compared to females. Specifically, men who have sex with men are at the greatest risk due to increased exposure to HIV and other infectious agents through physical contact and high-risk lifestyle behaviors (i.e. injection drug use).[24] This is a point of public health relevance; as the MSM population continues to grow in size, and increase in age given improved treatment options, the population and period at risk of developing NHL will also increase.

Among HIV-infected and -uninfected populations, European decent increases tumor susceptibility compared to other racial/ethnic groups. The risk attributable to tobacco and alcohol use is minimal in HIV/AIDS- and overall-NHL.[8, 25] In HIV-uninfected populations, more recent tobacco use and longer duration of smoking have been suspected to marginally increase NHL risk.[26, 27] In terms of alcohol, a slight decrease in risk associated with recent consumption has been suggested in HIV-uninfected populations, however this risk differs across gender.[28] Additional risk factors that may be associated with both overall- and AIDS-NHL susceptibility include but are not limited to: obesity, environmental, agricultural and occupational exposures, and other infectious agents. For example, obesity has been found to increase risk of NHL, while high intake of cruciferous vegetables and fibers has been shown to lower risk.[29] As for environmental exposures, there is a possible link between UV radiation and NHL: UV radiation increases T-suppressor cells and decreases T-helper and NK cells, indicating that UV radiation may increase NHL susceptibility.[14] Occupational exposures such as pesticides, insecticides and dust have been found to increase risk, especially in those who work in the farming, construction, forestry, agricultural or leather industries.[30] Furthermore, benzene, gasoline and arsenic also impart an increase in risk for non-Hodgkin malignancies.[31] While, the predominant infectious agent that contributes to NHL development is EBV, several other infectious agents, including

human T-cell leukemia virus type 1, hepatitis C, Helicobacter pylori and Campylobacter jejuni, contribute to general- and AIDS-NHL susceptibility.[8] Familial predisposition has also been suggested to increase the risk of NHL development in siblings of patients, while familial monoclonal B-cell disease is thought to act as a risk factor for children and young adults who are immunodeficient.[8,32]

Genetic Predisposition to NHL in the General and HIV-AIDS Populations

Genetic predisposition to NHL exists among both immunocompetent and immunocompromised individuals, and variants in several genes such as *HLA*, *TNF*-α, and *IL10* have been shown as associated with NHL susceptibility.[33, 34] The major histocompatibility complex, HLA, presents infectious agent peptides to T-cells to elicit an immune response. This complex consists of 128 genes located at 6p21 which have large allele variation across individuals and populations, including two HLA alleles that delay AIDS onset (B*27, B*57) and four that accelerate the disease progression from HIV-1 infection (A,B,C,B*35-Px).[34] Furthermore, immune competence is a risk factor for both non-Hodgkin's lymphoma and AIDS, and studies have reported associations between HLA variants and NHL risk.[35] Additionally, significant associations between HLA and NHL subtypes have been identified, in particular for follicular and DLBC lymphomas.[36] Given the established function of the HLA complex in immune response and lymphomagenesis, understanding the impact of inherited variants within this gene region may help inform potential biomarkers of prevention.

The tumor necrosis factor-alpha gene (*TNF*), located near HLA at chromosome 6p21, activates the NF-κB inflammation pathway. A pooled study from the International Lymphoma Epidemiology (InterLymph) Consortium identified that *TNF* −308G→A was associated with increased risk of diffuse large B-cell lymphoma. Furthermore, *TNF* −308G→A was in linkage

disequilibrium with lymphotoxin-alpha (*LTA*) 252A→G, which was also associated with increased risk of DLBCL. The haplotype including both *TNF* −308G→A and *LTA* 252A→G was associated with increased risk of DLBCL for all individual studies in the analysis overall.[37] Further, two SNPs on *LTA* and *TNF* comprise the G-A haplotype in the MHC region, which is associated with NHL risk in HIV-1 uninfected individuals.[38] To further investigate this association in those infected with HIV, SNPs in *LTA, TNF,* and 6 other proximal genes were typed in 140 MACS participants, with the G-A haplotype and a four-SNP haplotype in this neighboring gene cluster found to be associated with AIDS-NHL.[38] *Wang et al.,* echoed this result in a study which identified NHL cases from the SEER registry, in which the haplotype including *TNF* G-308A, *LTA* A252G, *LTA* C-91A, *LTA* A252G, *TNF* C-857T, *TNF* G-308A, and *TNF* G-238A, increased NHL risk.[39] A MACS study also found a nominally significant positive association between elevated TNF-α serum levels and NHL risk at three time points preceding AIDS-NHL diagnosis in HIV-positive patients.[40] These last points are of note as similar associations for both AIDS-related and non-AIDS-related NHL may suggest an important HLA-specific disease mechanism common to both. Further these data emphasize the critical importance of understanding how inherited variants work together to influence NHL susceptibility.

In contrast to *TNF*, interleukin 10 *(IL10)* is an anti-inflammatory cytokine which downregulates the expression of MHC class I and II.[41] Elevated IL-10 serum levels were significantly associated with NHL risk at three time points before diagnosis in the MACS.[33, 40] Additionally, a variant was identified within the promoter of *IL10* in this study that was overrepresented among AIDS-NHL cases. The InterLymph pooled study reported that the homozygous *TNF*−308G→A variant allele combined with at least one *IL10* −3575T→A SNP

allele was associated with a doubled risk for DLBLC.[37] Other authors have suggested that a series of gene expression abnormalities in the inflammation pathway could affect lymphomagenesis. Overall, a complicated genetic network is active and at work in general- and AIDS-NHL susceptibility. Given the established function of the HLA complex in immune response, and the critical implication of immune function in NHL development, understanding the impact of inherited variants within these highly crucial genes and region on NHL may help identify markers for prevention and control.

1.1.4. Pathogenesis Overview

Non-Hodgkin lymphomas are highly heterogeneous in nature, with distinct genetic lesions associated with each subtype.[42] Proto-oncogene activation results in the majority of NHL neoplasia that occurs. In non-Hodgkin lymphomas, these proto-oncogenes are mostly activated through chromosomal translocations, although point mutations and amplifications are involved. Most often, these chromosomal translocations result in the generation of a fusion protein, in which genes located at the breaksite of the chromosomes are fused together resulting in a hybrid protein that is deviant from the wild type.[43] Chromosomal translocation also causes homotopic deregulation in which the proto-oncogene is relocated to a region that enhances function resulting in constitutive oncogenic expression. Heterotopic expression can also occur in which the proto-oncogene is solely activated in malignant tissue.[44] As brief examples, three main proto-oncogenes undergo chromosomal translocations and are linked to NHL neoplasia, *BCL-1/CCND1*, *BCL-6*, and *c-MYC*.[43] *BCL-1* is a transcription factor and cell cycle regulator, and has a chromosomal abnormality at t(11;14)(q13;q32). The mechanism of this lesion leads to transcriptional deregulation and is a distinct genetic profile for Mantel Cell lymphoma. Lesions in the *BCL-6* gene have become known as the genetic signature for DLBCL, especially those

involving the chromosomal translocation at 3q27—an alteration detected in about 15% of DLBCL—which results in this transcription factor undergoing transcriptional deregulation. Lastly, abnormalities in the *c-MYC* gene are characteristic of Burkitt lymphoma and also a genetic signature for specific subtypes of AIDS-related NHL. The most common structural abnormalities in *c-MYC* occur at t(8;14)(q24;q32), t(2;8)(p11;q24), and t(8;22)(q24;q11), again resulting in transcriptional deregulation of this proto-oncogene.[43]

In addition, inactivation of the common tumor suppressor gene *p53,* and deletion of the long arm of chromosome 6 which harbors other crucial tumor suppressors, contributes to the genetic pathogenesis of NHL. Furthermore, other chromosomal anomalies, point mutations and infectious agents that compromise the immune system, including Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus/ human herpesvirus 8 (KSHV), participate in this oncogenesis.[3]

While HIV is one infectious agent that contributes to NHL through immune suppression, EBV and KSHV increase NHL susceptibility through their innate oncogenic potential. EBV infection is detected in about 40-50% of all AIDS-NHLs and nearly 100% of all tumors that present in the primary central nervous system (Table 1.1).[6] KSHV infection is associated with the development of primary effusion lymphoma on top of Kaposi sarcoma. These oncogenic viruses inhibit tumor suppressor genes, activate oncogenes and influence biologic mechanisms including cell cycle regulation and apoptosis.

1.1.5. Pathogenesis Specific to the HIV/AIDS Population

The greatest risk factor for non-Hodgkin Lymphoma development in the HIV/AIDS population is long-standing immunodeficiency and loss of immune function caused by the infection.[6, 45, 46] Two general pathways are accepted for the pathogenesis of AIDS-NHL.[47, 48]

First, HIV infection causes the depletion of helper T-cells which normally regulate EBV-infected B-cells. T-cell depletion allows uncontrolled EBV viral replication and cellular proliferation of these EBV-infected B-cells which results in the transformation of B-cells—an established step in lymphomagenesis. This pathway underlies NHLs such as central nervous system and primary effusion lymphomas, which are predominately linked to either EBV or KHSV infection and tend to occur among individuals with extremely low $CD4^+$ T-cell counts (<50 cell/mm^3). Second, HIV infection may cause chronic B-cell activation, in which an environment rich in the expression of AID (activation induced cytidine deaminase) and other inflammatory cytokines leads to an accumulation of molecular lesions which in turn contribute to lymphomagenesis.[47-50] Diffuse large B-cell and Burkitt lymphoma may be driven via this mechanism, as these NHLs are associated with EBV to a smaller degree, occur among people with moderate $CD4^+$ T-cell counts (>200 cell/mm^3), and have tumors that exhibit high proportions of oncogenic hypermutations and translocations that are indicative of hyper B-cell activation (Table 1.1).

In patients with Burkitt lymphoma, EBV infection is present in 30-50% of cases, a modest $CD10^+$ B-cell population is observed, and *c-MYC* is activated via translocation in 100% of cases. Further, a critical abnormality associated with Burkitt lymphoma is the deletion of *p53* in nearly 60% of individuals.[3] Diffuse large B-cell lymphoma can be divided into centroblastic and immunoblastic lymphomas—immunoblastic being those that occur extranodally, like PCNSL (15-30% of AIDS-NHL). EBV infection is present in over 40% of individuals with centroblastic DLBCL and nearly 100% of immunoblastic DLBCLs. Further, EBV infection is regarded to drive immunoblastic tumorigenesis as LMP1, an antigen encoded by EBV, is expressed in those with immunoblastic disease.[7] Last, immunoblastic DLBCLs express the non-germinal center B-cell activated phenotype (lacking *BCL6*/*CD10* and expressing

CD138^{+}*/MUM1*). In regards to centroblastic lymphomas, although ~40-60% EBV positive, the LMP1 antigen is not expressed. However, *BCL-6* proto-oncogene rearrangement is exhibited in 20% of these cases.[5] Further, centroblastic DLBCLs are characterized by CD20 expression, do not experience *BCL6* rearrangement and express the germinal center B-cell phenotypes (*BCL6*/*CD10* and lack of *CD138*$^{+}$*/MUM1*).[7] Primary effusion lymphoma is a rare group of KSHV/HHV8^{+}-related lymphoma, with ~80% of cases exhibiting EBV positivity. It is important to note that LMP1 is not expressed in this malignancy.[5] The most frequently expressed antigens of PEL include, *CD30*, *CD45*, *MUM1*, and *CD138*, with aberrant expression also associated with CD3/CD7. Last, all PEL cases exhibit HHV8/KSHV-associated latent nuclear antigen. In total, these mechanisms work specifically by taking advantage of compromised immune systems to promote specific genetic lesions that result in lymphomagenesis.

1.2. microRNA

1.2.1. Overview

Since the discovery of miRNA involvement in cancer in 2001, the important role of microRNAs (miRNAs) in the regulation of cell cycle control, immune response and other biologic mechanisms has become clearer. Variant miRNAs have been found to increase the susceptibility to and progression of various complex diseases, including cancer.[51, 52]

miRNAs are small non-coding RNA sequences, no greater than approximately 24 nucleotides in length, that regulate gene expression on the post-transcriptional level. There have been over 700 miRNAs identified in humans and they are thought to regulate at least 30% of all protein-coding genes. As they are non-coding RNAs, although transcribed via normal processes,

they are not translated. Instead they work at the post-transcriptional/pre-translational level to modify and regulate protein-coding messengerRNAs (mRNA).[53-56]

The first indication that miRNAs affect cancer development arose in 2001 from the Calin research group at MD Anderson Cancer Center. This research group identified two miRNA coding regions, *miR-15* and *miR-16*, located on a portion of chromosome 13 that was deleted in the majority of chronic lymphocytic leukemia patients (in their initial study, 65% of samples).[57] Furthermore, Calin's group confirmed that most miRNA coding regions (~50%) are located in genomic regions either associated with cancer or associated with fragile regions that promote DNA instability, such as intronic and intergenic regions.[58] In addition, unique miRNA profiles or signatures are being established in various cancers. For example, in breast cancer, differential miRNA expression in tumor tissues is used to classify the tissue into tumor subtypes: Luminal A, Luminal B, Basal-like and HER_{2+}.[52] Future use of these differential miRNA expression patterns has the potential to clarify and promote diagnostics, prognostics and cancer therapy.[59]

1.2.2. miRNA Biogenesis, Target Regulation, and SNP-Interaction

miRNAs undergo a series of complex interactions during their processing and maturation, all of which are highly sequence dependent.[60] It is important to keep in mind that germline variation, such as SNPs, has the potential to influence miRNA processing at each of these steps.[61-63] First, miRNAs are transcribed by RNA PolII as a long, primary transcript known as a pri-miRNA in the nucleus, which then folds into a stem-loop structure. SNPs located within the pri-miRNA sequence may influence both formation of the secondary stem-loop structure resulting in less efficient miRNA processing from the pri- to pre-miRNA, and subsequent maturation.[62] Further, DGCR8, an RNA binding protein, interacts with Drosha (another RNA enzyme) in order to cleave pri-miRNA and facilitate the transition from pri-miRNA to pre-

miRNA. Interestingly, one SNP located within this RNA binding protein, *DGCR8* rs417309, was found to interrupt this initial miRNA processing step by influencing the target site binding capacity of miRNA-106b and miR-579 to *DGCR8*, as reflected by higher luciferase activity in plasmids containing the SNP.[64] This example highlights the potential for variation with miRNA processing genes to change miRNA interaction and activity.

Next, this ~70 nucleotide stem-loop structure is bound by *XPO5* and *RAN* and transported into the cytoplasm. Once in cytoplasm, *DICER* binds the pre-miRNA and cleaves both strands to produce a 22 nucleotide duplex.[65] This duplex is then unwound, producing one mature miRNA strand that is ready for interaction with the family of *Ago* proteins, and a secondary strand that is most often times degraded.[66] The mature miRNA interaction with the RNA Induced Silencing Complex (RISC) and associated proteins (including Gemin3 and Gemin4) is a highly critical step. It is through the miRNA::RISC complex that miRNAs interact with and regulate target genes. SNPs within the RISC complex or within pre-miRNAs loaded into the complex have been identified to interrupt target interaction, resulting in deregulated mechanisms.[67-69] Last, the miRNA leads the complex to the specific target location within the 3' UTR (untranslated region) of the target gene. The 5' end of the miRNA then interacts with the 3' UTR of the target to either cleave mRNA or inhibit translation.[69]

In order to understand our study rationale, it is important understand the complexity behind this miRNA::target gene interaction. Target selection is based on the amount of pairing between the miRNA seed region (nucleotides 2-7 beginning from the 5' end of the miRNA UTR) and the target mRNA. The degree to which a gene is silenced depends on this interaction. If there is high complementation between the base pairs of the miRNA and the target messengerRNA, the RISC complex will attach and the miRNA will bind to the target mRNA at

the 3' UTR. The RISC complex then cleaves the mRNA at the designated region causing RNA interference which results in a portion of that mRNA being silenced. If high complementation is lacking, as is the case when a SNP is located within either the miRNA seed region or mRNA target region, the RISC complex binds to, but does not cleave, the target mRNA.[69] Instead the complex remains bound, blocking the open reading frame and inhibiting gene expression.[70] The fact that miRNAs can act on target genes that exhibit perfect and skewed complementation critically means that miRNAs have the ability to impart wide-spread influence and regulation on various steps critical to tumorigenesis.[65] Further, this indicates that miRNAs are still able to influence target gene regulation, to a degree, despite the presence of a SNP. However, it is the degree to which this interaction is impaired in relation to disease development that is of most concern.[64] Take into consideration the *miR-196a* coding region which contains a SNP within the precursor strand (rs11614913). *miR-196a* consists of two mature miRNA sequences, with this SNP located within what becomes the 3' passenger strand of the primary and mature sequence.[71] Given this location, altered target gene regulation and less efficient miRNA maturation from the precursor miRNA has been identified.[65] In fact, this SNP alters mature miR-196a levels *in vitro*, in which the variant allele of the precursor sequence imparts decreased levels of mature miR-196a compared to the levels produced by the wild-type precursor. Further, genome-wide expression microarray data demonstrated pre-miR-196a -T to result in fewer than half the amount of altered transcripts compared to pre-miR-196a -C, indicating decreased regulatory capacity due to this SNP.[71] Other SNPs within miRNAs themselves, or within miRNA promoter regions, have also been identified to influence regulatory capacity. [62, 68, 72]

In addition to the requirement of high Watson-Crick base pairing between the miRNA and target mRNA, miRNA activity can also be influenced by sequence variation near the site of

interaction. For example, an adenosine located within the target at nucleotide 1 in combination with complementation of the miRNA at position eight increases miRNA target recognition.[73] Further, several SNPs within miRNA target sites have been identified to increase cancer risk through altering proper interaction and processing. For example, *IL1A* rs3783553 interrupts a miR-122 binding site and has been associated with nearly a 40% decrease in hepatocellular carcinoma risk.[74] *KRAS* rs61764370 interrupts a *let*-7 miRNA family binding site and is associated with an increase in lung cancer risk.[75] These examples highlight the importance of high sequence complementation between the miRNA and the 3' UTR of the target gene. These examples also indicate that SNPs within either the miRNA, target mRNA, or processing machinery have the ability to interrupt this highly sequence-dependent target interaction.

In summary, SNPs located in miRNA binding sites are thought to interrupt miRNA target interaction, thus deregulating function and subsequent gene expression.[76, 77] Further, SNPs within miRNA binding sites result in the RISC complex not efficiently silencing mRNAs or initiating translation repression, thus resulting in variant gene sequences thought to be involved in tumorigenesis. Additionally, research indicates that SNPs in miRNA seed regions, transcripts and miRNA processing machinery also influence target interaction, miRNA maturation and activity to affect tumorigenesis.[65, 78]

1.2.3. miRNA and Cancer

The regulation of miRNA plays a role in tumorigenesis, and miRNAs have the potential to act either as tumor suppressor genes or as oncogenes in both normal and malignant tissues.[79-81] We will discuss the potential for both options briefly. miRNAs found as down-regulated in cancerous tissues have been coined tumor suppressor genes. If a miRNA is down-regulated, their ability to function properly is interrupted resulting in uncontrolled cell proliferation. For

example, a recent study investigated miRNA expression in colorectal cancer tissue samples and compared them to the expression found in normal tissues of the same patient. Two miRNAs were found to have significantly decreased expression patterns in the cancerous tissue when compared to the non-cancerous tissue: *miR-143* and *miR-145*.[82] Their decreased expression would imply that normal expression helps maintain cell function, thus these miRNAs may act as tumor suppressor genes, although this is just one possibility. Another study investigated 20 prostate tumors via a qPCR analysis and concluded that *miR-15a* and *miR-16* were under-expressed compared to normal tissue in 85% of their samples, suggesting that these miRNAs might also act as tumor suppressors in prostate tumors.[83] Also, among several others, *miR-101* acts as a tumor suppressor in prostate cancer, as reflected by the fact that *miR-101-1* on chromosome 1 and *miR-101-2* on chromosome 9 are deleted in nearly 40% of clinically localized prostate tumor tissues.[84] In addition, decreased expression of the *Let-7* family of miRNAs found in non-small cell lung cancer contributes to tumorigenesis through deactivating this miRNA family's ability to repress cell proliferation.[85] Another study by *Nasser et al.*, investigated the tumor suppressor activities of *miR-1* in lung cancer and came to several interesting conclusions.[86] These researchers found *miR-1* to be down-regulated in primary human lung cancers, determined through under-expression of this miRNA in 87.5% of their samples. Furthermore, *miR-1* function was restored in the cancerous lung tissue with treatment of either the tumor suppressor C/EBP, or a histone deacetylase inhibitor. Lastly, *Nasser et al*., investigated the anti-tumorigenic properties of *miR-1* and found that expression of *miR-1* in specific non-expressing cells reversed the tumorigenic properties of those cells, including tumor formation, growth and replication. Future studies should focus on identifying the anti-tumorigenic properties of miRNA.

miRNAs are also involved in the *p53* pathway. Initially members of the *miR-34* family were classified as tumor suppressor genes due to their ability to induce apoptosis and promote cell cycle checkpoints in neuroblastoma cells.[87] Furthermore, this family has been found to be down-regulated in several cancers. During a genome-wide screen performed by the Tarasov research group in 2007, the *miR-34* family was found to be up-regulated by increased *p53* activity. This was accentuated by the fact that the increased expression promoted by *p53* resulted in *miR-34*'s participation in apoptosis and G1-cell cycle arrest. Additionally, the first non-coding exon region of *miR-34* is located next to an evolutionarily conserved *p53* binding site, reiterating this miRNA family's association with the *p53* network.[88] Future research is bound to identify additional miRNA families whose work as tumor suppressors partake in the *p53* mechanism.

The majority of miRNAs that are over-expressed in cancerous tissues are considered to have oncogenic function. One example, the *miR-17-92* gene family, is over-expressed in B-cell lymphoma samples as well as in colorectal and lung cancer tissues. The mechanisms underlying this gene family's functions are not fully elucidated, however, there is a suggested mechanism.[89] Two miRNAs in the *miR-17-92* cluster—*miR-20a* and *miR17-5p*—target the E2F1 transcription factor that promotes cell cycle progression and disrupts the *c-Myc* feedback-loop resulting in uncontrolled cellular proliferation.[90] Furthermore, this same miRNA cluster is thought to evade apoptosis through deregulating the *BIM* pro-apoptotic gene.[91] And specifically, once *miR-20a* and *miR17* are inhibited, they allow the induction of apoptosis in lung cancer cells.[92] Moreover, a colorectal cancer study identified twenty-one different miRNAs as over-expressed in cancerous tissues, of which six were confirmed to be over-expressed in a statistically significant manner when compared to normal tissue.[82] Additionally, *miR-21* is up-regulated in prostate cancer tumors, suggesting an oncogenic function of this aberrant miRNA.[93, 94] miRNAs are also thought

to regulate *c-MYC* at the established 8q24 locus risk site in prostate cancers, although this research is continually evolving.[95, 96] Further research identifying miRNAs that are up-regulated in human cancers can only promote our understanding of miRNA mechanisms in tumorigenesis.

1.2.4. microRNA and Immune Function and B-cell development

Emerging research indicates that microRNAs also are involved in the formation of both the adaptive and innate immune systems.[97] Furthermore, research indicates that immune function directly influences miRNA transcription. These observations make the involvement of miRNA in normal immune function a new and exciting field of research.[98]

There seem to be immune system responses that affect the processing of miRNAs along each portion of their biogenesis. For example, *miR-155* is a miRNA with an active role in B-cell differentiation, germinal center response and proper maturation of Ig class-switched plasma cells; hence critical for normal immune function and development.[99] The transcription of *miR-155* has been found to be up-regulated in response to various immune responses, particularly the inflammatory stimuli of toll-like receptors (TLRs).[100]

miRNAs have also been found to act as regulators of haematopoietic stem cells, and cells of the innate immune system such as granulocytes, monocytes, macrophages and dendritic cells.[98] For example, the *miR-17-92* cluster down-regulates the production of runt-related transcription factor 1, which in turn regulates the colony-stimulating factor receptor during the maturation of monocytes.[101] miRNAs are also substantially involved in B-cell development and maturation. Research indicates that *miR-150* and the *miR-17-92* cluster are required to support early development and survival during the progenitor to precursor B-cell transition.[102] *miR-150* acts to repress *MYB* and *miR-17-92* acts to repress *BIM* and the BCL-2-interacting mediator of death expression. In addition, other miRNAs may also be associated with this developmental

transition. Furthermore, the development of mature B cells is regulated by *miR-155*, through a mechanism in which *miR-155* affects the encoding of *AID* and PU.1. Both *AID* and PU.1 are also found to be involved in lymphomagenesis.[98] Through *miR-155*'s influence on these mechanisms, antibody production and class-switching are promoted, again indicating the importance of miRNAs in B-cell development.[103]

miRNAs also act in T-cell development, where their production by the miRNA processing gene *DICER* is critical for proper T-cell development. Like in B-cell development, the *miR-17-92* cluster targets *BIM* and additionally *PTEN* to influence T-cell selection and survival during their differentiation in the thymus. Furthermore, *miR-181a* is also found to promote T-cell selection and survival through affecting various protein phosphatases along the path to maturation.[104] Once in the periphery, *miR-155* and other miRNAs act on T-cell maturation to promote maturation into T helper 1 cells, while *miR-326* acts to promote cells towards T helper 17 cells, active in inflammatory disease.[98] Thus, miRNAs contribute to the development, maturation and fate of both B- and T-cells to promote adaptive immune function.

1.2.5. microRNA and Non-Hodgkin Lymphoma

While much remains to be understood about the role of microRNAs in lymphomas, deregulation of miRNAs including, miR-21, miR-142-3p/-5p, miR-143, miR-145, miR-150, miR-195, miR-497, and the *miR-17-92* family have been described.[89, 105] For example, increased serum levels of miR-21 are associated with relapse-free survival in DLBCL, and downregulation of miR-142-3p/-5p, miR-143, miR-145 and miR-150 has been observed in Burkitts lymphoma, EBV-transformed cell lines, and mantle cell lymphoma—with implications in apoptosis and cell growth.[105] Further, miR-142-3p, miR-142-5p, and miR-101 are downregulated by the *BCL6* proto-oncogene in DLBCL. The *miR-17-92* cluster is over-expressed in several human

malignancies, including AIDS-NHL.[89] Initially, *miR-17-92* was found to be transcriptionally activated by *c-MYC*, suggesting a pertinent relationship between the miRNA cluster and this oncogene in specific lymphomas. For example, the *miR-17-92* cluster was found to be over-expressed in Burkitt's lymphoma in patients with non-HIV-related malignancies, where the microRNA is activated by a mechanism reliant on *c-MYC*.[106] Furthermore, *miR-17-92* has been found to suppress apoptosis in Burkitt's lymphomas, with *miR-17-92*, *miR-19a* and *miR-19b* having a forefront position in promoting the oncogenic potential of the cluster through regulating *PTEN* expression. The *miR-17-92* cluster has also been identified as repeatedly over-expressed in diffuse large B-cell lymphoma, in which this cluster regulates *PTEN* and *BIM*, both well-known tumor suppressors.[107]

miR-155 is also a critical miRNA related to lymphogenesis. This miRNA is related to viral immune response, is substantially involved in immune and inflammatory response, and has been identified in relation to both HIV/AIDS and cancer progression.[108] *miR-155* acts by partaking in B-cell differentiation, and through working to suppress *BIM*. In addition, this miRNA expression is induced by *BIC*, the B-cell integration cluster.[109, 110] Several studies have confirmed *miR-155* to be over-expressed in DLBCL. In addition, one recent study indicated that *miR-155* targets the inositol phosphatase SHIP1, and through that regulation over-expression of *miR-155*, leads to decreased levels of SHIP1 in DLBCL. It is thought that this reaction is due to tumor necrosis factor α causing an autocrine stimulation that enhances *miR-155* expression.[99]

Another miRNA, *miR-142*, is located within a group of *Ig* and non-*Ig* genes that are direct targets of *AID*—whose wide-known deregulation is directly linked to the genomic instability of B-cell lymphoma initiation. As *miR-142* is located in this specific location, it is translocated in B-cell malignancies.[111] Furthermore, there is potential for other miRNAs, such as

miR-26, *miR-34a*, *miR-153* and *miR-203*, to contribute to mechanisms related to aberrant lymphogenesis and hence non-Hodgkin lymphoma development.[112-115] Further research into the extensive roles that these miRNAs play in non-Hodgkin lymphoma susceptibility and development has the potential to improve mechanistic understanding, and promote better preventive and control measures.

1.2.6. microRNAs and miRNA-related SNPs as Potential Biomarkers of Cancer Development

MicroRNAs have much potential as biomarkers for cancer. They may serve as diagnostic markers for cancer development, or as prognostic indicators, distinguishing indolent from aggressive disease.[116] Oncogenesis and tumor progression involve many pathways including cell growth, apoptosis, differentiation, DNA repair, inflammation, motility, and metabolism; microRNAs, in theory, can affect all of these biological pathways.[56]

SNPs in microRNAs have been shown to have a profound impact on cancer.[117, 118] They have the ability to inhibit the expression or activity of a tumor suppressor, or can enhance expression or activity of an oncogene. For example, miRNA *Let-7* functions as a tumor suppressor by regulating the *KRAS* oncogene, among others, and is frequently downregulated in cancer.[85] One SNP, rs61764370, in the 3' UTR of *KRAS* is associated with increased risk for ovarian, oral, non-small cell lung cancer, triple negative breast cancer, and melanoma, and moreover is associated with resistance to chemotherapy in ovarian and colorectal cancers.[69] A SNP in the 3'UTR of *TGFBR1*, which has a major role in regulating cell growth, was shown to increase the risk of familial breast cancer.[119] Examples of alterations in DNA repair miRNAs include polymorphisms in the 3' UTRs of *PARP1* and *RAD51*, which have consequences in bladder cancer, as well as RAD52 (rs7963551), which is associated with reduced breast cancer

risk.[120-122] As such, a specific miRNA-related SNP such as this may be associated with a HER-2 positive or triple negative breast cancer, and thus may help pinpoint the best treatment approach.

Such differences highlight the diverse nature of miRNA-related SNPs and suggest their involved activity as both tumor suppressors and tumor promoters. Thus, miRNA-related polymorphisms have significance in cancer development, disease prognosis, and therapeutics that will only increase with our own increasing knowledge of the field.

1.3. Gaps in the Literature

miRNA expression profiles have been associated with AIDS- NHL risk. However, germline SNPs in miRNAs and their processing machinery have yet to be investigated in relation to AIDS-NHL. Additionally, SNPs within miRNA target sites of genes implicated in HIV progression and immune function have not been fully investigated in relation to AIDS-NHL. Our study will fill these gaps through the use of the highly detailed MACS cohort, and to our knowledge is the first study of its kind investigating germline SNPs related to miRNA pathways and AIDS-related NHL

Table 1.1 Molecular Features of AIDS-NHLs						
Subtype	**% of all NHL in general population**	**% of all NHL in HIV**	**Oncovirus associations (EBV, KSHV)**	**% EBV positive**	**CD4 T cell count at Dx**	**Characteristic molecular lesion of the tumor**
Burkitt lymphoma	1.4% [17]	8.4% [17]	EBV[123]	30-50%[123]	125 cells/ μL [124]	*C-MYC* translocation, point mutations in *MYC* regulatory regions and *p53* mutations common. [5]
Diffuse large B-cell lymphoma	Overall DLBC: 23.1%[2]	Overall DLBC: 45.5% [117]	EBV[123]	PCNSL: ~100%[123] Centro DLBCL 40-60% [123]	94 cells/μL [124]	Immunoblastic Characterized by CD20 expression and EBV-encoded LMP1, with a small sector expressing CD30. Non germinal center b-cell activated phenotype- lacking CD10 and BCL6 and expresses CD138 and MUM1. Centroblastic: Characterized by CD20 expression, with CD30 expressed in some tumors. Also, expression of germinal center B-cell phenotypes BCL6 and CD10 and lack of CD138 and MUM1.
PEL	<1% [125]	4% [125]	KSHV, EBV[123]	100%/50-80%[123]	98 cells/μL [126]	LNA1 positive (HHV8/KSHV-associated latent nuclear antigen), CD45 [5]
Plasmablastic of oral cavity	<1%[2]	2.6% [5]	EBV[123]	75%[5]	120 cells/μL [127]	EBER positive

CHAPTER 2: RESEARCH OBJECTIVES AND METHODOLOGY

2.1. Research Objectives

The objective of this proposed study was to assess whether SNPs located within miRNA binding sites, miRNA coding regions, or miRNA processing genes—collectively referred to as miRNA-related SNPs—were associated with susceptibility to non-Hodgkin lymphoma in an HIV/AIDS population through a candidate gene approach and a case-control study nested within the Multicenter AIDS Cohort Study. Further, we sought to assess the potential association between SNPs in miRNA processing genes and serum levels of four miRNAs (miR-21, miR-122, miR-222, and miR-223) associated with AIDS-NHL in a previous study.

2.2. Specific Aims and Hypotheses

Objective 1: To identify miRNA-related SNPs through integrative bioinformatic techniques, and to estimate any association between these SNPs and AIDS-NHL susceptibility using a nested case-control study within the Multicenter AIDS Cohort Study.

Hypothesis 1

We hypothesize that microRNA-related SNPs (SNPs within miRNA coding regions and miRNA target sites) will be associated with AIDS-NHL risk. This hypothesis is based on biologic evidence that demonstrates SNPs within miRNA coding regions, or miRNA target sites, affect miRNA::Target gene binding affinity, which in turn influences target gene regulation and expression.[118, 128-131] In addition, we hypothesize that SNPs within miRNA processing genes may be associated with non-Hodgkin lymphoma risk. This hypothesis is based on biologic observation that SNPs within miRNA processing genes may affect miRNA formation and

maturation, which in turn may alter downstream miRNA expression levels and subsequent target gene regulation.[65]

Specific Aim 1

i) To identify SNPs in genetic sequences coding for miRNAs, miRNA target sites, and miRNA processing machinery that are associated with lymphomagenesis through the use of bioinformatic tools including prediction and pathway analysis.

ii) To evaluate potential associations between SNPs in miRNA, miRNA target sites, or in miRNA processing machinery, and AIDS-NHL susceptibility.

Objective 2: To investigate the relationship between SNPs in miRNA-processing genes and relative expression levels of four serum miRNAs related to AIDS-NHL susceptibility.

Hypothesis 2

We hypothesize that SNPs within genes that process and regulate miRNA maturation will be associated with differential miRNA serum levels. The basis for this hypothesis is that SNPs within miRNA processing machinery interrupt miRNA maturation resulting in aberrant miRNA expression levels which may further increase or decrease cancer susceptibility. Additionally, these aberrant miRNA expression levels have the potential to act as markers of patient prognosis, treatment responsiveness and survival—reemphasizing their biologic significance.[131-134] As example, one SNP, *DGCR8* rs417309, influenced the target site binding capacity of miRNA-106b and miR-579 to *DGCR8*, reflected by higher luciferase activity in plasmids containing the SNP.[64] This SNP, located within the 3' UTR of this miRNA processing gene, was also associated with breast cancer risk. Further, *XPO5* rs11077—a SNP investigated in the present study—is suggested to decrease both overall survival and disease-free survival among individuals with Hodgkin lymphoma, as well as influence treatment-related pulmonary toxicity.[133] Given this

scientific observation, we hypothesize that genetic variation within miRNA biogenesis genes has the ability to alter miRNA maturation such that downstream expression levels are impacted. We further accept that these aberrant miRNA expression levels may influence AIDS-NHL development further downstream.

Specific Aim 2

To assess whether genetic variation in miRNA-processing genes is associated with mean $\log_e$ miRNA serum expression levels among a subgroup of MACS participants using linear regression.

2.3. Study Design and Methods

2.3.1. Study Overview

To test our hypotheses, we designed a case-control study nested within the Multicenter AIDS Cohort Study (MACS). The MACS is the longest running prospective study investigating both the natural and treated progression of HIV/AIDS, with 7,087 men who have sex with men followed for up to twenty-nine years, and recruited over four requirement periods (April 1984-March1985, n=4,954 recruited; April 1987-September 1991, n=668 recruited; October 2001-August 2003, n=1,350 recruited; and 2010+, n=115 recruited thus far).[135] Four study sites were used to collect these data (Baltimore, Chicago, Los Angeles, and Pittsburgh), and for this study, data were utilized from HIV positive men enrolled in either the first or second MACS cohorts (1984-1985 and 1987-1991, respectively), as these were the only two recruitment periods with eligible AIDS-NHL cases. Individuals participated in a baseline visit at study enrollment and subsequent semi-annual visits for the entirety of their follow-up time. At these visits, each

participant responded to detailed questionnaires that captured demographic data and information such as medical history and lifestyle and behavioral risk factors (including history of substance abuse, consumption of alcohol, use of tobacco, and sexual behaviors). Additionally, data regarding current medications, including anti-retroviral use and adherence, were documented. AIDS-defining outcomes were also documented. Biologically relevant specimens were provided by participants for risk and prognostic factors such as $CD4^+$ T-cell count, HIV RNA viral loads, and HBV/HCV serology. Biological specimens are stored at the CAMACS Central Repository. There is over an 86% follow-up rate for HIV-positive participants in the MACS.[136, 137] The study was approved by the Institutional Review Boards (IRBs) associated with each MACS site, including the University of California at Los Angeles, USA.

2.4. Study Population

2.4.1. Case Definition and Reference Date

Cases were defined as MACS participants enrolled in either the first (1984-1985) or second (1987-1991) recruitment periods who were diagnosed with a pathologically-confirmed (ICD-O-3 coded) non-Hodgkin lymphoma as of July 2010 and had DNA samples available for analysis. Pathological confirmation was obtained through the combination of state cancer registries and pathology reports, or, in some cases, at the time of death via autopsy (30 cases out of 180, or 16.7%, had identical NHL diagnoses and death dates). A total of 225 cases of NHL have been diagnosed in the MACS. Of these, 183 NHL cases fulfilled our inclusion criteria and had DNA cells available for extraction and subsequent analysis. For cases, reference date was defined as the NHL diagnosis date.

2.4.2. Control Definition, Reference Date and Matching Criteria

To be eligible as a control, participants had to be HIV-positive, enrolled in either the first or second MACS recruitment periods, and, cancer-free at the time of case-NHL diagnosis.

Cases were matched to controls in up to a 3 controls:1 ratio on the following variables: recruitment period (April 1984-March1985, n=4,954 recruited; April 1987-September 1991, n=668 recruited), duration of HIV positive and cancer-free follow-up time (controls had to be followed at least as long as the cases), race (white or non-white), seroprevalence at time of recruitment or seroconversion during follow-up, and $CD4^+$ T-cell count (categorical: 0-49 mm^3, 50-99 mm^3, 100-199 mm^3, 200-349 mm^3, 350-499 mm^3, and 500 mm^3 and above). The time point for $CD4^+$ T-cell count in cases was the last measurement before NHL diagnosis, while in controls the time point for $CD4^+$ T-cell count was within one year of case duration of follow-up. A total of 533 HIV-positive controls were selected. For controls, reference date was defined as the date of first HIV-positive study visit plus the time period in days between first HIV-positive study visit date and the NHL diagnosis date of the corresponding case.

2.4.3. Data Collection

We obtained peripheral blood mononuclear cells from cases and controls. Whole genomic DNA was extracted using the QIAamp DNA blood mini kit, purified, and then amplified using the REPLI-g Mini Kit in accordance with Qiagen, Inc protocol. This amplified DNA was then genotyping on the Fluidigm platform.[138] In addition, we requested numerous covariates of interest from both the baseline and follow-up questionnaire data from the central data coordinating center.

2.5. miRNA-related SNP Selection (Specific Aim 1)

An overview of our miRNA-related SNP selection is displayed in Figure 2.1. The process can be simplified as follows, and is described in detail beginning in section 2.5.1. (1) First, we will describe our selection starting from miRNA coding regions of interest and ending in SNP selection. Initially, we selected candidate miRNAs from literature and laboratory measurements. (a) SNPs with MAFs greater than 5% in Caucasians and documented within those miRNA coding regions identified from the literature and laboratory measurements were automatically added to our list (n=5). (b) Additionally, we used TargetScan5.2 to predict the target genes of miRNAs of interest from the initial list. (i) If a predicted target gene overlapped with a candidate miRNA from the initial list, then the target gene binding sequence and location span were documented. (ii) Using the UCSC Genome Browser "Blat", the predicted binding location on the gene was translated to chromosomal location. Next, using dbSNP, the gene region was sorted by 3'UTR SNPs. Through using the identified chromosomal binding location, we identified and documented SNPs located within or near that position. Last, SNPs had to have a MAF>5% in the Caucasian population in order to be included in our final SNP list (total n=12).

(2) Second, we will describe our selection starting from a candidate gene list, turning that into a target gene list from which we selected additional SNPs. Initially, a list of candidate genes was compiled using scientific literature and KEEG pathway analysis to include genes associated with general tumorigenesis and lymphomagenesis. a) We then used TargetScan5.2 to predict miRNAs that targeted these candidate genes of interest to create our target gene list. (i) If a predicted miRNA overlapped with a candidate gene from the list, then the target gene binding sequence and location span were documented. (ii) Using the UCSC Genome Browser "Blat", the predicted binding location on the target gene was translated to chromosomal location. Next,

using dbSNP, the gene region was sorted by 3'UTR SNPs. By using the identified chromosomal binding location, we identified and documented SNPs located within or near that position. Last, SNPs had to have a MAF>5% in the Caucasian population in order to be included in our final SNP list (total n=12).

(3) In order to select our miRNA biogenesis SNPs, first we conducted a literature review of the greater than 25 genes associated with miRNA biogenesis. At study onset, seven of those genes were highlighted in the literature with potential miRNA-related SNPs (*Drosha, XPO5, RAN, DICER, AGO, GEMIN3, GEMIN4*). We selected eight functional SNPs highlighted in the literature from these genes that also had a European decent MAF greater than 5%. (n=8).

The final SNP list included 25 SNPs within miRNA coding regions (n=5), miRNA processing genes (n=8), and within or near predicted miRNA target sites (n=12).

2.5.1. Detailed miRNA Selection (Step 1)

To initiate our miRNA selection we conducted an extensive literature review aimed to identify miRNAs putatively linked to AIDS-NHL susceptibility. These aspects included miRNAs related to immune function, inflammatory response, cell cycle progression and regulation, B-cell differentiation and those established to be differentially expressed across NHL subtypes. From this investigation, we established our initial list of pertinent miRNAs.

<u>miRNA identified from the literature:</u>

miR-155, miR-17-92 cluster (includes: miR-17-5p, miR-17-3p, miR-18, miR-19a, miR-19b-1, miR-20, miR-92-1), miR-15a-497 cluster (Includes: miR-15a, miR-15b, miR-16, miR- 497), miR-29a, miR-29b, miR- 29c, miR-34a, 34c-5p, miR-106b-25 cluster (Includes: miR-106b, miR-93,miR- 25), miR-142-3p, miR-146a, 146b-5p, miR-150, miR-153, miR-181a, 181b, 181c, 181d, and miR-203.

Additionally, our miRNA selection was informed by research conducted in the laboratory of Dr. Martínez-Maza by Dr. Dharma Thapa. Briefly, Dr. Thapa profiled the expression of miRNAs in normal B-cells derived from tonsils to identify a miRNA signature associated with normal B-cell differentiation and maturation. This miRNA signature was then used to compare miRNA expression profiles in tumors across subtypes of non-Hodgkin lymphoma.[89] Our list of miRNAs was expanded to include miRNAs implicated in AIDS-NHL subtypes, in addition to those listed above:

Let7-a, let7-b, Let7-i, Let7-c, Let7-d, Let7-g, miR-30a, miR-30b, miR-30c, miR-30d, miR-30e, miR-21, miR-23a, miR-23b, miR-24, miR-26a, miR-26b, miR-1297, miR-101, miR-129, miR-185, miR-221, miR-222, miR-494, and miR-342.

2.5.2. Detailed Target Gene Selection (Step 2)

A list of pertinent target genes was identified through an extensive literature review at study onset. We sought to identify genes related to immune function, inflammatory response, cell cycle progression and regulation, B-cell differentiation and those established to be differentially expressed across NHL subtypes. Furthermore, we sought to identify genes related to these mechanisms and also known to be regulated by, or regulators of, miRNAs of interest.

To expand our target gene list, we used the TargetScan miRNA target prediction algorithm, a cornerstone in miRNA target prediction from the Whitehead Institute for Biomedical Research, MIT.[139-143] Several studies have confirmed that a multitude of TargetScan predicted miRNA::mRNA interactions are accurate through the use of microarray and proteomic approaches.[144] Furthermore, these studies have shown that the TargetScan algorithm is one of the top-predicting programs of correct targets.[145-148] Using the TargetScan algorithm we identified the binding sites for each of our chosen 55 miRNAs of interest. Briefly, the TargetScan

algorithm predicts which functional genes are targeted by a specific miRNA for post-transcriptional regulation. TargetScan is comprised of independent features experimentally and computationally found to influence miRNA targeting. The TargetScan algorithm takes into account the following items to predict miRNA target sites: 1) the amount of nucleotide pairing within the 3' UTR, and specifically the amount within the highly efficient region of 7mer-m8 sites, 2) the amount of local AU content within the 7mer-m8 sites (the higher the content the more efficient the pairing), and 3) the distance of the 7mer-m8 site to the 3' UTR ends. Each of these influences is individually scored and then combined using a regression model to produce a total score called the "Context Score". Multiple in vivo and in vitro studies were used to model the correspondence between the context score and miRNA::mRNA interaction efficiency. When a researcher runs the algorithm for a particular miRNA on their publicly available website (www.targetscan.org) the target genes are ranked by context score so that researchers are able to identify target genes with the most efficient predicted binding site. The more negative a context score, the higher the likelihood and efficiency of the miRNA::target gene interaction due to the modeling structure. For each of our miRNAs, we selected the TargetScan algorithm's top-ranked genes as reflected in their context score to jumpstart our investigation, corresponding to approximately the top 30% of hits for each gene as suggested in recent literature.[146, 148] This list of predicted target genes produced from the algorithm for our 55 miRNAs of interest consisted of 855 genes.

Function and Pathway Analysis

Analyzing the basic function of these 855 genes became critical in order to create a list of target genes, 1) related to functions and pathways of interest in our study and 2) that could be further investigated using the pathway analysis program KEGG.[149-152] Thus, we produced a list

of target genes that were known to be involved in cell cycle regulatory mechanisms, cell proliferation, apoptosis, B-cell development, tumor suppression, have oncogenic potential, genes related to miRNA processing machinery, and/or involved in pathways known to affect tumorigenesis or NHL development. This narrowed our initial predicted genes from 855 to 157 genes. Pathway analysis of these 157 genes was our next step.

Literature suggests that miRNAs tend to work through entire pathways to exert full regulatory potential.[65, 88, 145, 153, 154] As such, it was important for us to investigate the pathways of each of the remaining 157 genes to identify other closely related genes that could be targets of our miRNAs of interest. To do this, we used the publically available KEGG (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/).[151, 152] In this program, we identified all of the pathways of interest in which our 157 target genes occurred. This included 34 different pathways ranging from cell cycle regulation, to mRNA surveillance, to various signaling pathways. Once the pathways for our 157 target genes were identified, a total of 258 closely related genes in their pathways were also noted. We identified these genes through literally investigating the pathways from start to finish and focusing on genes linking the process-initiating genes to the product (ex. apoptosis or cell proliferation). The majority of these 258 genes were seen to occur multiple times over the pathways of interest.

As the focus of our study was to identify genes that target our miRNAs of interest, we then had to determine whether any of these additional 258 pathway-related genes were putative targets of our 55 miRNAs. We turned back to the TargetScan algorithm and ran all of the new genes in the algorithm to identify which were predicted targets of our miRNAs of interest. In this iteration, we considered genes to be eligible for our study if they were 1) targeted by a miRNA of interest, and 2) if their Context Score ranked them within the top half of predicted target sites.

After this analysis 150 genes from the pathway analysis were considered eligible to be included in our target genes of interest.

In conclusion, our final candidate gene list for SNP investigation included 55 miRNAs of interest, 297 of their predicted targets, and 7 genes directly related to miRNA processing machinery (n=359). Through uniting the features of miRNA target prediction and pathway analysis, our list of candidate genes encompassed a wide-range of genes related to specific aspects of NHL susceptibility and development.

2.5.3. SNP Selection

Our next step was to identify whether or not SNPs of potential interest were located within the target regions of our candidate gene list. *A priori* we decided to include SNPs for genotyping with a minor allele frequency (MAF) greater than 5% in the Caucasian population according to dbSNP, as 83.5% of cases and 90.6% of controls are Caucasian in this study. We set the minor allele frequency threshold to 5% so as to help ensure that the minor allele cell counts analyzed were not under-powered. In other words, setting a 5% MAF threshold helped ensure that minor allele cell counts would be large enough to make meaningful statements about the rare allele.

First, SNPs meeting this criterion and established in the literature as within miRNA processing genes or within miRNA coding regions, were automatically added to our SNP list (Step 1a). Next, we took each miRNA coding region or target gene of interest from our candidate list and called that coding region or gene in TargetScan5.2 (list collected from Steps 1b and 2). TargetScan5.2 then informed us which miRNAs targeted a specific gene, or which genes targeted a specific miRNA. If a miRNA was also on our candidate miRNA list, or a gene also on our candidate gene list, then the binding sequence and binding location on the target gene were

recorded. Next, through using the UCSC Genome Browser "BLAT", and entering the recorded target sequence of interest, the chromosomal position of the target location within our gene of interest was provided (Step: miRNA-SNP and Candidate gene 3' region SNP selection). In dbSNP, we sort by gene region to compile a list of all 3' UTR SNPs, and then mapped the predicted target location over the SNP chromosome position. We thus were able to identify SNPs within or near a miRNA binding site of interest. As SNPs located near miRNA binding sites have also been shown to affect miRNA binding, and given our interest in a thorough investigation of the 3'UTR region, we compiled a list of 164 SNPs, spanning 115 genes, predicted to be within at least 200 base pairs of a miRNA of interest binding site. We prioritized this list by distance to miRNA binding site and gene pertinence to NHL development, selecting 25 SNPs for genotyping on the Fluidigm platform. Table 2.1 displays these SNPs.

2.6. Genotyping

In total, 25 SNPs were selected to be genotyped from amplified DNA on the Fluidigm platform across 12 plates containing 716 MACS samples.[138] One SNP related to miRNA maturation and processing (*DROSHA* rs10719) continually failed the platform and was excluded from further genotyping at onset. This study thus assessed data from 24 successfully genotyped SNPs.

Genotyping was performed with a customized Fluidigm Dynamic 96.96 Array™ Assay at the UCLA Genetics Laboratory. These assays were based on allele-specific PCR SNP detection chemistry with the reliability of Dynamic Array™ integrated fluidic circuits (IFCs). The SNPtype Assay we used employed tagged, allele specific PCR primers and a common reverse primer. A universal probe set was used in every reaction, producing uniform

fluorescence. Additionally, Fluidigm provided locus-specific primer sequences that allowed one to confirm target locations.

2.7. Quality Control Measures

2.7.1. Post-genotyping SNP and Sample Inclusion and Exclusion Criteria

SNPs were excluded if (1) genotyping SNP call rate <95%, (2) Hardy-Weinberg equilibrium (HWE) p-value < Bonferroni-adjusted p-value 0.002, (3) <95% Coriell sample concordance for SNPs, and (4) pairwise r^2 value for linkage disequilibrium higher than our accepted threshold of 0.80. Samples were excluded if they did not meet the following criteria: (5) sample coverage <90%, and (6) <95% replicate sample concordance on the duplicate plate. These criteria were established as follows: (1) SNP call rate reflects the proportion of SNPs successfully called (genotyped) out of the total population genotyped for that SNP and is used as a screening tool for genomic data which reflects genotyping platform efficiency. We set our SNP call rate threshold to greater than 95% to ensure inclusion of valid data in our analysis; (2) Hardy-Weinberg equilibrium tests were conducted among controls to evaluate potential genotyping errors identifiable through serious deviations from equilibrium, given the HWE underlying assumption that the control group is representative of the general population. Any deviation of the control allele from Hardy Weinberg Equilibrium may be indicative of a potential genotyping error; (3) 12 Coriell samples of established race/ethnicity and gender with genotypes sequenced and replicated in HapMap were included in our study and genotyped (one per plate) via the Fluidigm platform. The genotyped sequences were then compared to the replicated Coriell sample sequence established in HapMap. As any deviation between the two sets indicates an error in genotyping, we set our threshold to 95%; (4) In order to investigate correlation

between SNPs located within the same gene post-genotyping, we set a pairwise r^2 threshold to 0.80, meaning that if two SNPs were found to be correlated at an r^2 value greater than 80%, only one would be chosen for further analysis as we are confident in the ability for one SNP to act as a proxy for the estimate of the second, paired SNP; (5) Sample coverage investigates the number of SNPs successfully genotyped out of the total number of genotyped SNPs per sample (per participant) as a quality control measure reflecting the overall quality of each DNA sample. We set our sample coverage threshold to greater than 90%—a standard quality threshold for sample inclusion; and (6) Greater than a 95% replicate plate concordance level was required in order to add further confidence to the validity of the chosen genotyping platform and sample quality.

2.7.2. Initial Measures

Initially, all blank and water samples (one blank and one water sample per plate) were confirmed as having been called appropriately as "no calls" using the Fluidigm SNP Genotyping Analysis software (100% of blank and water samples called accurately). We also compared the expected base calls, minor allele and genotype frequencies of our data to those found in NCBI's dbSNP database and HapMap to check that our genotypes fit the expected distributions (Table 2.1).

In order to calculate the study minor allele frequency, we investigated the minor homozygous allele and heterozygous allele calls. Using the following equation for MAF, the corresponding values for each SNP was determined: MAF= [2(qq)+1(pq)]/[2(# of samples genotyped for a particular SNP)], where qq=minor allele call and pq=heterozygous allele call.

Although, zero SNPs were excluded due to low post-genotyping minor allele frequency, two SNPs were found to have 'flipped' minor alleles. We suspect this is due to the fact that their documented dbSNP MAFs hovered around 0.50 (rs2075993: HapMap MAF A=0.47, major to

minor GG/AG/AA; Our data MAF G=0.4937 and rs7813: HapMap MAF T=0.49; major to minor CC/CT/TT; Our data MAF C=0.3945).

2.7.3. Coriell Sample Concordance

Next, Coriell sample concordance was investigated. We plated one Coriell sample per plate as a quality control measure (12 Coriell samples total), and confirmed that 100% of the genotype calls for the Coriell samples were concordant with the genotypes for those same samples reported in HapMap.

2.7.4. Call Rate Determination

In order to calculate call rate we used the following equation: 1-(# of samples that failed to be called for a particular SNP/716), where 716 was the total sample size for our project that was genotyped. All call rates were above 95% (lowest was 98.5%) thus no SNPs were excluded from further analysis based solely on call rate.

2.7.5. Sample Coverage

In order to investigate sample coverage, we calculated how many SNPs were called successfully out of the successfully genotyped 25 SNPs for each of our 716 samples. Sample coverage investigates the number of SNPs successfully genotyped out of the total number of genotyped SNPs per sample (per participant) as a quality control measure reflecting the overall quality of each DNA sample. We set our sample coverage threshold to greater than 90% —a standard quality threshold for sample inclusion. We identified four samples with sample coverage under 90% and these were excluded from further analysis. One of these samples would have been excluded from conditional logistic regression analysis anyway due to having zero matched controls.

2.7.6. Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium tests were conducted among controls to identify any potential genotyping errors as identifiable through serious deviations from equilibrium. Given the issue of multiple comparisons, we adjusted our p-value for comparison using the following equation: (standard p-value of 0.05/# of SNPs genotyped, in our case 25). Our adjusted p-value then became: 0.002. Using the chi-squared test for HWE we determined one SNP out of HWE and was excluded from further analysis: *TAB3* rs3816757 (HWE p-value: 0.000000).

2.7.7. Linkage Disequilibrium

Zero SNPs in this study were identified as being in high linkage disequilibrium at a pairwise r^2 threshold of 0.80.

2.7.8. Duplicate Sample Concordance

We re-genotyped 5% of our samples as randomly selected by a random number generator for quality control purposes (n=53/1066).

On our replicate plate, 21 out of 24 SNPs were duplicated with 100% concordance. *WWOX* rs12828 had ~8% discordance, higher than our accepted threshold of 5% and thus this SNP was dropped from further analysis. Of the 53 samples re-genotyped, one sample was previously excluded due to low sample coverage. Thus, 98.1% of our samples were re-genotyped with 100% accuracy (51/52). Two SNPs were replicated incorrectly out of 25 SNPs replicated in one sample (8.0% discordance), which was higher than our accepted threshold of 5%, and thus this sample was removed.

2.7.9. QC Conclusions and Summary of Data for Analysis

SNPs to exclude: *TAB3* rs3816757 (HWE p-value: 0.000000), and *WWOX* rs12828 due to 8% replicate discordance. Thus we completed statistical analysis on 22 SNPs (out of 25 originally genotyped).

We excluded four samples that were unable to be genotyped in >10% of SNPs (one of which was previously notated as an unmatched case in the original dataset and excluded), and another sample that exhibited poor replicate concordance. Thus, we completed statistical analysis on 709 MACS participants including 180 cases and 529 controls, with 180 matched risk sets available for conditional logistic regression.

2.8. Statistical Analysis

2.8.1. SNP Frequencies

Table 3.1 displays the frequencies of SNPs investigated in our study. Initially, we examined the attributes of the dataset, and investigated the SNP frequencies in both cases and controls. To assess deviations from HWE, the SNP frequencies for the controls were analyzed to confirm which genotypes were representative of the 'general population' to assure interpretation of our results.[155] Fourteen SNPs out of 25 had values missing for either the cases or controls (Table 3.1).

Initial SNP-cancer associations were examined under additive, dominant and recessive genetic models. To investigate patient demographic characteristics we used the Pearson Chi-squared or Fisher's Exact test dependent on sample size, with Student's t-test used to evaluate

continuous variables (Table 3.2). All analyses were conducted using SAS v9.1.3 software (SAS Institute, Cary NC).

2.8.2. Description of Study Variables

Exposure Variables:

The exposure variables of interest were the identified miRNA-related SNPs.

Outcome of Interest:

The outcome of interest in this study was AIDS-related non-Hodgkin lymphoma. This outcome was further stratified by site: systemic or primary central nervous system (CNS AIDS-NHL) lymphomas.

Risk Factors for NHL:

Important risk factors for AIDS-NHL include lower $CD4^+$ T-cell counts and higher HIV viral loads.[5, 47, 156] Given that we conducted a genetic association study, it was important for us to consider potential confounders that were strongly associated with NHL risk or HIV progression as it was unlikely that the potential variables affected the inherited miRNA-related SNPs. However, risk factors such as $CD4^+$ T-cell count and HIV viral load are so strongly associated with both HIV-1 progression and AIDS-NHL development, that it was necessary to take factors such as those into consideration as covariates that may have muddle our observations.[5, 157, 158]

2.8.3. Longitudinal Covariate Data and Variable Creation

Prior to the start of our statistical analysis, we investigated the longitudinal dataset in order to create variables for the following potential covariates: history of methamphetamine use, history of tobacco use, cumulative pack years among smokers, history of alcohol consumption, intensity of alcohol consumption and history of cannabis use. We investigated the literature to

help inform our decision of up until which time period before reference date we should investigate each potential confounder.

In order to investigate history of tobacco use, exposure history was taken up to one year prior to diagnosis for cases and one year prior to reference date for controls. Literature suggests that NHL risk is associated with smoking to a greater extent among those with the longest duration and pack-years [hazards ratio for NHL development among individuals with less than 20 years as a smoker: 1.09 (0.67-1.78) compared to the hazards ratio for NHL development among individuals with more than 20 years as a smoker: 1.76 (1.25-2.47)].[159] Additionally, time since quitting seems to influence NHL risk, however this is inconsistent. Thus, we sought to use data as close to the diagnosis date as possible, without having the disease effect smoking status (for example, if the measure was taken at diagnosis). Additionally, a one-year prior to diagnosis time-window had been previously used to control for tobacco use in InterLymph publications.[27, 159-162]

In order to investigate the history of alcohol consumption, exposure history was taken up to one year prior to diagnosis for cases and one year prior to reference date in controls. Literature suggests that alcohol drinking decreases NHL risk marginally in men, and that these estimates are lower for individuals who currently consume alcohol [OR: 0.73 (0.64-0.84)], compared to individuals who formerly consumed alcohol [OR: 0.95 (0.80-1.14)].[163] Additionally, this time-window was previously used to control for alcohol use in InterLymph publications.[28, 163-165]

In order to investigate history of methamphetamine use, exposure history was taken up to one year prior to diagnosis for cases and one year prior to reference date in controls as informed by a recent publication by *Chao et al.*, in which recent methamphetamine use prior to diagnosis was found to infer the greatest AIDS-NHL risk: 4.73 (1.41-15.8).[166]

In order to investigate history of cannabis use, exposure history was taken up to one year prior to diagnosis for cases and one year prior to reference date for controls. A few papers have looked exclusively at marijuana use and NHL development. One paper from 1999 looked at cannabis use one year prior to interview, which loosely translated into about one year prior to diagnosis as the majority of cases were recruited within one month of diagnosis. This paper emphasized that decreased risk of NHL was associated with increased and more recent cannabis use. As such, capturing information one year prior to diagnosis reflects recent use.[167]

2.8.4. Covariate Selection

A priori, given biologic rationale based on our extensive knowledge of AIDS-NHL etiology, the following variables were established as necessary for model inclusion: age at reference date, AIDS diagnosis prior to reference date, $CD4^+$ T-cell count at reference date, HIV viral load at set point prior to HAART, history of anti-retroviral therapy, and race. In order to select other important covariates, such as alcohol consumption and tobacco use which traditionally are not as strikingly associated with lymphoma risk, covariate selection was appropriate. Both backward and forward selection methods can result in confounded or imprecise estimates if used without caution.[168] Backward selection was chosen here as a forward selection approach would not have maximized the benefit of our *a priori* selected variables.

In our backwards variable selection under a logistic regression model, set at an alpha level of 0.10 with *a priori* and matching variables forced into the model, the following covariates were investigated, with the resultant p-value displayed: smoking status one year prior to reference date (categorical: ever, former, never; p-value: 0.91), during follow-up, cumulative pack-years up to one year prior to reference date (continuous; p-value: 0.25), alcohol consumption up to one year prior to reference date (categorical: current, former and never; p-

value: 0.27), intensity of alcohol consumption at visit one year prior to reference date (categorical: 1-3 drinks per week, 4-13 drinks per week, 13 or more drinks per week; p-value: 0.56), recreational cannabis use up to one year prior to reference date (categorical: current, former and never; p-value: 0.99), recreational amphetamine use (RAU) one year prior to reference date (categorical: current, former and never; p-value: 0.17), and HCV infection status prior to reference date (categorical: ever, never; p-value: 0.09). In order to deal with issues of co-linearity, we investigated backwards selection using combinations of the above variables such that the impact of highly correlated variables (ex. smoking and pack-years) was investigated separately. After backwards variable selection, only HCV infection (ever vs. never) remained evident as a potential confounding factor at an alpha level of 0.10 (p-value= 0.09). We thus adjusted for HCV in our analyses.

HCV status was measured at the baseline visit, retrospectively using archived serum samples, as well as prospectively at every bi-annual visit after 2001. A participant was considered to be HCV negative if all HCV antibody tests prior to reference date were negative. Among those who tested positive for HCV antibodies, a participant was considered to have cleared their HCV infection if at least two anti-HCV tests were negative, thus being labeled as "ever" HCV infected, while not considered chronically infected. Recently, Dr. Po-yin Chang determined that history of HCV infection prior to reference date was marginally associated with AIDS-NHL risk in this same study population (OR_{adj}: 1.75, 95% CI: 0.97-3.13), which increased our confidence in adjustment for this covariate.[169]

2.8.5. Statistical Analyses and Model Decisions

We estimated adjusted odds ratios (OR_{adjs}) and 95% confidence intervals (95% CI) for each SNP*AIDS-NHL association using conditional logistic regression which was appropriate given our matched case-control study design. This analysis maintained matched pairs and study design integrity. We estimated the association between our SNPs of interest and 1) overall NHL, 2) systemic NHL, 3) central nervous system (CNS) AIDS-related NHL, under a main effects model, adjusting for age at reference date (continuous), HIV viral load at set point (continuous, before HAART), AIDS diagnosis prior to reference date (reference= no), anti-retroviral therapy prior to reference date (reference= no), $CD4^+$ T-cell count at last measurement within 18 months prior to reference date (continuous), race (reference=non-Hispanic white), and history of HCV infection (reference=no; Tables in Chapter 3)**.**

2.8.6. Sensitivity Analyses

In addition to conditional logistic regression which was appropriate given the matched case-control study design, we conducted unconditional logistic regression as a sensitivity analysis. Under this modeling scheme, we adjusted for all matching and adjustment factors in order to estimate each SNP*AIDS-NHL association. This highlighted any potential difference in estimates imparted by risk-sets dropped from the conditional logistic regression analysis due to missing covariate data (Table 3.8- no differences to report). For example, under a conditional logistic regression model such as ours where the cases are matched to controls in up to a 3:1 ratio, if there is missing covariate data for the case, the entire set (all four participants) are dropped from the analysis. However, under an unconditional logistic regression model where there are no matched sets, only the participant with missing covariate data would be excluded from the analysis. The unconditional and conditional logistic regression models would show

prominent differences if there were large amounts of missing covariate data, especially among the cases. However, as established by Breslow and Day, if the study was designed with matched data (as was the case in the current study), then the conditional likelihood should be used as there is a beneficial increase in estimate precision compared to the use of unconditional methods. This is especially true as unconditional methods may produce biased risk estimates if there are a large number of strata with scarce data.[170] As this study design was a matched case-control study, where the potential for small strata and sparse data was of concern, conditional logistic regression was the optimal modeling scheme and became the focus of this dissertation. However to be prudent, we considered our analysis under an unconditional logistic regression model in a sensitivity analysis.

Additionally, we ran sub-group analyses 1) among Caucasians only, to examine SNP associations in a homogeneous population, and 2) among individuals without a primary Kaposi Sarcoma (KS) diagnosis prior to reference date (Tables in Chapter 3). Forty-three controls and twenty AIDS-NHL cases experienced a KS event prior to reference date. We excluded these individuals as i) in the case of controls: due to experiencing a previous oncogenic event, they were unrepresentative of the underlying HIV-infected source population from which the controls were sampled, ii) in the case of cases: these individuals may have experienced different genetic or behavioral risk factors that contributed to the initial primary KS diagnosis, and iii) overall: these individuals may be more susceptible to a second primary cancer.[171, 172] Last, we investigated the multiplicative interaction and joint effects between a SNP located within the *microRNA-196a* coding region (*microRNA-196a* rs11614913) and a second SNP located within a *microRNA-196a* binding site (*HIF1A* rs2057482) and overall, systemic and CNS AIDS-NHL.

2.8.7. Missing Data Issues

One of the most informative covariates in our analysis, HIV viral load at set point, was missing on one hundred and twenty-eight participants, including all individuals enrolled during the second MACS recruitment period (missing in total n=128; missing in cohort 1 n=56; missing in cohort 2 n=72). HIV viral load at set point is a time point in which there is a natural equilibrium between viral replication and viral clearance, and has been found as a good predictor of time to AIDS diagnosis. For example, a lower HIV viral load at set point translates into a longer time until the transition to AIDS. Further, Dr. Po-yin Chang recently established HIV viral load at set point as a reasonable predictor of AIDS-NHL risk in this same study population nested within the MACS, compared to viral load measurements taken at other time points. This again emphasizes the importance of a viral load measurement taken at this time point (OR_{adj}: 3.77, 95% CI: 1.48-9.65). The time point in which natural equilibrium between viral replication and viral clearance occurs is usually achieved several months after initial infection in individuals who are unexposed to anti-retroviral treatment. For MACS sero-converters in the first cohort, the estimated viral load at set point was the average viral load 12 to 24.5 months after sero-conversion. For individuals who had already undergone sero-conversion prior to entry (HIV-prevalent participants) in the first cohort, their viral set point was approximated from visit 3 or 4 to avoid potential measurement errors. However, for HIV prevalent individuals enrolled in the second cohort (100% of participants enrolled in our study from cohort two were HIV-prevalent), the viral load at set point measurement was missing as they had probably already sero-converted several years prior to study enrollment (3-5 years prior). In order to address this missing data issue, we used overall median imputation to estimate the median HIV RNA viral load at set point

(median control value= 4.316 vc/ml) in controls, and applied that median value to the missing data for the 128 participants.

Another critical covariate of interest, $CD4^+$ T-cell count at last measurement within 18 months of reference date, was missing on 56 participants. This $CD4^+$ T-cell count data had 17.2% missingness among cases (31 AIDS-NHL cases) and 4.7% missingness present among controls (25 controls; overall 7.9% missing, n=56). In order to account for this missing data we also applied the overall median value in controls (median control value= 113 cells/mm^3) to those with missing values.

Missing data is a common occurrence in molecular epidemiologic research, and there are several ways to handle missingness such as that found within our study.[173] As this biomarker data was missing systematically on individuals enrolled in the second recruitment period (i.e., the missingness of the data was related to the patient characteristic of being recruited in the second enrollment period), we cannot conclude that these data are missing completely at random. This is especially true as differences between the cohorts are established on key covariates including race.[137] Therefore, one viable way to handle missing data is to apply the overall median value in controls to participants with missing covariate data.[174] We used median over mean imputation as the normality of these two variables were slightly skewed. One advantage of an overall median imputation approach is the ability to incorporate an investigator's knowledge of the variable to help inform if the median value applied is in fact an appropriate and biologically plausible value.[173, 174] Additionally, this method is easily computable without advanced statistical software. One disadvantage of this approach includes that the single value used for imputation does not take into consideration a model for the missingness, nor does it take into consideration sampling variability.[173, 174] However, as long as the investigator is aware of this potential overstatement of

precision, and presents the data such that the reader is made aware of this caveat, appropriate interpretation of the data is attainable.

As a sensitivity analysis, Monte Carlo Markov Chain (MCMC) multiple imputation for missing data was explored. The following variables of interest with missing values were imputed: $CD4^+$ T-cell count at last measurement within 18 months of reference date (n=56 missing), HIV viral load at set point (n=128 missing), history of tobacco use (n=1 missing), history of alcohol consumption (n=14 missing), history of methamphetamine use (n=26 missing), and history of cannabis use (n=14 missing). The following variables were used to help inform the imputation: race, sero-conversion status, AIDS diagnosis prior to reference date, use of anti-retroviral therapy prior to reference date, duration of HIV follow-up time, age at reference date, history of HCV infection, and $CD4^+$ T-cell count at matching. As a modest proportion of individuals had missing covariates, it was appropriate to choose a small number of imputations; after running various combinations (n=10, 30, 50, 100), we decided on 30 imputations.[168] The multiply imputed parameter estimates for the mean value for HIV viral load at set point was 4.36 vc/ml (95% CI: 4.31-4.42), and the mean value of $CD4^+$ T-cell count at last measurement within 18 months of reference date was 185.75 cells/mm^3 (95% CI: 170.6-200.90). No reportable differences in odds ratio estimation were noted in our sensitivity analysis using multiply imputed data compared to median imputed data, as demonstrated in Chapter 3, Table 3.9.

2.8.8. Interpretation of Results Given Missing Data

In this study, interpretation of the complete case data was not representative of the entire study population as each individual in the second recruitment period was excluded in this analysis due to missing biomarker data on HIV viral load at set point (18.1% missingness; total n=128; cohort 2 n=72). It is established that complete case approaches only produce unbiased

risk estimates in situations where the data are missing completely at random and thus are representative of the larger, underlying study population.[174, 175] We believed estimation using the complete case data may produce biased risk estimates as i) this biomarker data was missing systematically on individuals enrolled in the second recruitment period (i.e., the missingness of the data was related to the patient characteristic of being recruited in the second enrollment period); and ii) we were unable to rule out that the source of missingness was related to unobserved variables at work. Further, individuals enrolled in cohort one are documented to differ on several key covariates, including racial composition, level of education attained, and $CD4^+$ T-cell counts, compared to those enrolled in cohort two.[137] In order to assess our suspicion that the complete case approach excluded individuals (those enrolled in cohort 2) who differed from included participants (those enrolled in cohort 1), we compared the distribution patterns of the following pertinent variables between cohorts: age at reference date, HIV duration in years prior to reference date, AIDS diagnosis prior to reference date, history of ART use prior to reference date, history of HAART use prior to reference date, $CD4^+$ T-cell count at date of matching, race, history of HCV infection and sero-conversion status (Table 3.3). We calculated chi-squared p-values comparing cohort 1 to cohort 2 across these key covariates. HIV sero-conversion status at cohort entry ($p=0.004$), duration of follow-up in years prior to reference date ($p=0.0001$), age at reference date ($p=0.0001$), ethnicity ($p=0.004$), history of ART use ($p=0.02$), and $CD4^+$ T-cell count at date of matching ($p=0.006$) all differed significantly at an alpha level of 0.05 between individuals in cohort 1 and cohort 2. After taking these differences between those included and excluded in the complete case analysis into consideration, we were more inclined to focus our result interpretation to data utilizing the entire study population through the overall median imputation approach.

Further, we compared results from three highlighted SNPs and overall AIDS-NHL across a complete case analysis, a median imputed analysis, a multiply imputed sensitivity analysis, a model not adjusted for HIV viral load at set point, and a model only adjusted for matching factors, to investigate result robustness across analytic schemes (Table 3.9).

2.8.9. Semi-Bayesian Approach for Multiple Comparisons

Molecular and cancer epidemiological studies often times explore the associations between multiple biomarkers and various outcomes, making the issues of multiple comparison and false discovery of analytic importance. In 1992, Greenland generalized an application of a semi-Bayes approach to the analysis of correlated multiple associations using data from an occupational cancer-mortality study.[176] In this paper, data was described using conventional modeling approaches, parametric empirical-Bayes approaches, and what he described as a "semi-Bayes" approach. In this approach, he did not augment the original data with a prior estimated from the data, but rather with prior data emphasizing the investigator's *a priori* thoughts regarding the potential range an odds ratio may realistically lie within. Over the years, this semi-Bayesian approach was expanded to include setting priors of null associations to pull the observed associations towards the null, with normal coefficient priors of null associations [$\beta \sim N(0, 0.5)$], emphasizing investigator uncertainty.[177-179] Further, although the semi-Bayes (SB) approach had been proposed to provide corrected estimates for the issue of multiple comparisons, this approach had not been well-employed in cancer or molecular epidemiological studies due to a lack of detailed procedures using statistical software until recently.[180] However, ongoing work between our T32 Molecular Cancer Epidemiology group, in particular Dr. Po-yin Chang, and Dr. Greenland expanded this methodology to provide a protocol with SAS code for the SB approach to correct for multiple comparisons in cancer molecular epidemiological

studies.[181, 182] Following this methodology, and fitting it appropriately for our matched case-control study, we applied the semi-Bayes approach to any observed statistically significant association with normal coefficient priors of null associations [$\beta \sim N(0, 0.5)$], as described in detail below, to correct for multiple comparisons (code provided in the appendix).

The prior specification used for augmentation in the current study was based on previously described, detailed methodology.[177-179] As such, we assigned a normal prior for coefficient β with prior mean (m_{prior}) and variance (v_{prior}): $\beta \sim N(m_{prior}, v_{prior})$. The median, mean and mode of coefficient priors all were equal to the m_{prior}. The antilog of β (e^{β}) has a log-normal distribution which corresponded to the median of the hazards ratio in cox regression, or odds ratio in logistic regression. After exponentiation, coefficient priors demonstrate the investigator's probability percent confidence (P%), or P/(1-P) bets, that an association is focused on e^{β} with upper (OR_{upper}) and lower (OR_{lower}) limits. If, for example, an investigator is 95% certain that an OR_{prior} lies between 0.25 and 4.00 with equal chance of being above or below the center, then the $OR_{prior} = (OR_{upper}*OR_{lower})^{1/2} = 1$, $m_{prior} = \ln(OR_{prior}) = 0$, and $v_{prior} = [\ln(OR_{upper}/OR_{lower})/(2*1.96)]^2 = [\ln(4/0.25)/(2*1.96)]^2 = 0.50$, the prior variance traditionally applied in this methodology and used in our study.

As cancer is considered a rare disease, the odds ratio approximates the incidence rate ratio (IRR), and as such we are able to make an extension using pseudo-populations.[177, 178, 183] As investigators analyzing a pseudo-population, we observe A_1 exposed (X = 1) cases and A_0 unexposed (X = 0) cases arising from N_1 exposed and N_0 unexposed participants, respectively ($RR_{prior} = (A_1/N_1)/(A_0/N_0) \approx \exp(m_{prior})$; $v_{prior} \approx 1/A_1+1/A_0$).[178] To approximate a log-normal prior for a RR, A_1 and A_0 are set to equal A, while for mathematical simplicity both N_1 and N_0 are set as large values. For example, if $N_1 = 10^5$ and N_0 is N_1*RR, then $RR_{prior} = (A_1/N_1)/(A_0/N_0)$

= $N_0/N_1 \approx \exp(m_{prior})$, and $v_{prior} \approx 2/A$.[177, 178] As mentioned, in order to obtain a null prior (RR_{prior} = 1.00) with 95% limits of 0.25 and 4.00 and a v_{prior} of 0.50, an investigator would observe four exposed and four non-exposed cases arising respectively from 10^5 exposed and 10^5 unexposed participants in a pseudo-population. For the purposes of cancer molecular epidemiology, these case numbers are reasonable. However, the exact probability of the prior RR being in the interval of 0.25 to 4.00, is actually slightly less than 95%.[178] To improve the 95% coverage, we could i) re-scale the exposure effect X by a scaler S, even if X = 1/S may be meaningless; ii) set RR_{prior} = $N_0/N_1 = \exp(m_{prior}/S) = \exp(m_{prior})^{1/S}$; and/or iii) calculate A from $A= (2*S^2)/v_{prior}$, even A is clinically impractical.[178] Thus, in our matched study, established with a prior variance of 0.50, each prior data record required at minimum two pairs (four data-lines) of records, where the first pair described the concordant risk-set with one line for exposed cases (Y = 1, X = 1/S = 1/10, A = 400), and one for non-exposed non-cases (Y = 0, X = 0, A = 400), and the second pair described the discordant risk-set, where one line represented exposed non-cases (Y = 0, X = 1/10, A = 400), and the other represented non-exposed cases (Y = 1, X = 0, A = 400). All other covariates were set to zero, as described in detail below.

We thus applied the semi-Bayes approach to any observed statistically significant association with normal coefficient priors of a null association [$\beta \sim N(0, 0.5)$], in order to also investigate and correct for false-positive findings in our main effects analyses.[176, 184] Through the incorporation of a normal prior of null-association (prior odds ratio, OR_{prior} = 1.00; 95% prior limits: 0.25, 4.00), we proposed to pull observed associations toward the null, to emphasize our uncertainty.[178] Table 3.7 displays the results of the original significant estimates alongside the SB corrected posterior estimates.

The semi-Bayesian approach for main effect estimation can be described in the following steps: 1) creation of dummy variables; 2) re-scaling and re-centering of pertinent covariates for meaningful interpretation; 3) addition of prior indicators to the original dataset; 4) creation of the cox regression model; 5) quality control check of the semi-Bayesian cox model against the original estimate established under a conditional logistic regression model; 6) building of the prior dataset and quality control check; and 7) data augmentation using the SET statement and posterior hazards ratio estimation (all code provided in the appendix). We will use the association between *GEMIN3* rs197412 and overall AIDS-NHL for illustrative purposes.

Initially, we recoded the following categorical covariates of interest as dummy variables: AIDS diagnosis prior to reference date, HCV infection prior to reference date, history of ART treatment prior to reference date, and race. We then dummy coded the SNPs of interest, which were to be investigated given their statistical significance. For example, a polymorphism with three genotypic levels (homozygous major allele carriers, heterozygotes and homozygous minor allele carriers) under the nominal model would be recoded such that two dummy variables became representative of the initial three-level variable; under the dominant or recessive models, one dummy variable represented the initial two-level variable.

Next, in order for us to have contextually meaningful interpretations under SB correction, regressor re-scaling and/or re-centering was critical. As example, either a one-year change in age or being of zero years of age at cancer diagnosis gives little clinical or practical insight. Thus to increase the meaningfulness of their interpretation, the following variables were re-centered and/or re-scaled: age at reference date was re-scaled and re-centered to the age of HIV-positive controls in the study (age_r0824-40/10), and $CD4^+$ T-cell at matching and at reference date were re-scaled ($CD4^+$ T-cell count/100).

We then indicated weights to our original and prior datasets in order to ensure and incorporate dataset augmentation. For example, A=1 was used as a weight for each "actual" data record, and thus, in our original dataset A was set to 1, whereas in our prior "hypothetical" dataset, A=400. The variable "time" was necessary to add to the original dataset as this variable was matched between the groups in Cox regression, and as such was set to the same value (we used zero) in the original dataset, with all prior records set to time=1. Once our original dataset was updated with these items, we built a Cox regression model using our original dataset, with recoded, re-scaled and re-centered variables. With the risk-set indicator under the STRATA statement, this Cox regression model mimicked the original conditional logistic regression model. Under this modeling scheme, we modeled the hazard of AIDS-NHL diagnosis, while stratifying by the matching indicator variable "r0824_set", using our original dataset with new dummy variables. To ensure quality control, using the original data with new variables, the hazards ratios produced by the Cox regression model were compared to those estimates previously produced under the conditional logistic regression model, as the hazards ratios along with 95% confidence intervals should directly replicate the original estimates, which they did.

Once the built Cox model replicated the conditional logistic results, the prior dataset was created. The covariates included in the prior dataset were the same as those used in the original dataset, just dummy coded if categorical: age at reference date, HIV viral load at set point, AIDS diagnosis prior to reference date, history of ART treatment prior to reference date, $CD4^+$ T-cell count at visit closest to reference date, history of HCV infection prior to reference date, and race. R0824_case represented our case-control indicator.

To create our prior dataset, we used the "r0824_set" variable to maintain matching between our cases and controls. For each genotypic model (nominal, log-additive, dominant, and

recessive) a different prior dataset was created as displayed in the appendix. As example, we will describe our prior dataset creation in detail for the nominal model of *GEMIN3* rs197412, under which the heterozygote carriers are compared to the homozygote major allele carriers, and homozygote minor alleles carriers are compared to the homozygous major alleles carriers.

In our prior dataset, for the nominal genetic model, we added the matched risk-set indicator "r0824_set". In the original dataset, these matched sets were identified by r0824_sets 1 through 185. Thus to differentiate between the original and prior datasets, the prior dataset was labeled with the first risk-set starting at 201, followed by 202, and so forth. Additionally, under the nominal model, we needed one discordant and concordant pair for the heterozygote carriers who were compared to the homozygote major allele carriers, and one discordant and concordant pair for the homozygote minor allele carriers who were compared to the homozygous major alleles carriers. Thus, for the first pair data-line in the prior dataset, r0824_set was labeled as 201 for the first concordant pair (exposed case and unexposed control) estimating heterozygote carriers who were compared to the homozygote major allele carriers, and 202 for the second, discordant pair (unexposed case and exposed control) estimating heterozygote carriers who were compared to the homozygote major allele carriers. The third and fourth paired data-lines, were labeled 203 for the concordant pair estimating homozygote minor allele carriers compared to the homozygous major allele carriers, and 204 for the discordant pair estimating homozygote minor allele carriers compared to the homozygous major allele carriers. For each prior of two data-lines per risk-set, one prior data-line was set as an exposed case (outcome Y = 1) with Y=1, X= 1/S = 1/10, $A = S^2*(2/v_{prior}) = 10^2*(2/0.50) = 400$, and all other covariates being zero, and one prior data-line was set as an unexposed control (outcome Y = 0) with Y=0, X= 0, $A = S^2*(2/v_{prior}) = 10^2*(2/0.50) = 400$, and all other covariates being zero. Next, one prior data-line was set for an

unexposed case with Y= 1, X = 0, A = 400, and all other covariates being zero, and one prior data-line was set for an exposed control with Y= 0, X = 1/10, A = 400, and all other covariates being zero.

Once the prior was created, a quality control check was implemented under the cox regression model using the prior dataset. After running the model based on the prior dataset, the prior dataset had a hazards ratio equal to 1.00, and a 95% CI ranging from 0.25-4.00, with standard error of 0.707, reflecting our normal coefficient priors of null associations [$\beta \sim N(0, 0.5)$]. Once the prior dataset passed this quality control check, data augmentation was next. Through augmenting our original dataset with our prior dataset of null association, we obtained posterior SB estimates and 95% posterior limits.[184] The SB posterior estimates tended to be closer to the null than the maximum likelihood estimates observed under conditional logistic regression, with narrower 95% posterior limits. To give a comprehensive overview of our data and results, we presented both suggested (alpha=0.10) and statistically significant results (alpha=0.05) in the results and discussion sections. However, we placed highlighted emphasis on significant results which remained evident after the SB approach as we were more confident in the report of these results due to the fact that they were pulled toward no association (*GEMIN3* rs197412, *miR-196a* rs11614913, and *HIF1A* rs2057482).

2.9. Investigation of the relationship between miRNA-related SNPs and mean $\log_e$ serum levels of four miRNAs established as biomarkers of HIV progression and AIDS-NHL susceptibility (Specific Aim 2)

2.9.1. Study Background

The goal of this aim was to assess whether genetic variation in miRNA-related genes was associated with miRNA serum levels that were measured among a subgroup of seventy-seven MACS participants.[185] In brief, Dr. Dharma Thapa and the laboratory of Dr. Otoniel Martínez-Maza, investigated the hypothesis that serum miRNAs were dysregulated in HIV-infection and that miRNA serum levels have the ability to act as biomarkers of AIDS-NHL susceptibility. In order to assess this hypothesis, serum microRNA levels were estimated and compared in 43 HIV negative men, 45 HIV positive men who did not develop AIDS-NHL, and 62 HIV positive men prior to AIDS-NHL diagnosis from the Multicenter AIDS Cohort Study (n=150). In order to estimate miRNA serum levels among AIDS-NHL individuals, serum samples were taken at a median time of 8.8 months prior to diagnosis. After screening 175 miRNAs for differential expression across 22 samples (AIDS-NHL=7; 8=HIV- and 7=HIV+ controls) four miRNAs—miR-21, miR-122, miR-222, and miR-223—were suggested as deferentially expressed and validated by qPCR across all 150 samples. Higher levels of miR-21 and miR-122, and a decreased level of miR-223, were observed among HIV-1 infected men compared to HIV uninfected men. Furthermore, compared to the HIV-uninfected group, HIV+ individuals who later developed CNS AIDS-NHL or DLBCL experienced higher miR-222 serum levels.

Of the 150 participants investigated in this miRNA serum analysis, 77 overlapped with participants in our genetic association study. Thus, we used this opportunity to investigate whether genetic variation in miRNA-related genes was associated with mature miRNA expression as reflected though these serum levels.

2.9.2. Analytic Plan

We sought to investigate the association between SNPs within miRNA processing genes and miR-21, miR-122, miR-222, and miR-223 serum levels. Initially, univariate analysis was conducted in order to investigate the normality of all four miRNA serum levels. Histogram plots revealed that a $\log_e$ transformation seemed to improve the normality of these serum level estimates (Figure 2.2). Given the continuous nature of these miRNA serum levels, a linear regression approach was used.

In order to inform our covariate selection, we conducted backwards variable selection under a linear regression model, while forcing AIDS-NHL status and race into our model, using an alpha threshold of 0.10. We investigated the following potential covariates, and display the resultant p-value: history of ART use prior to serum sample date (p-value=0.67), AIDS diagnosis prior to serum sample date (p-value=0.57), duration of HIV infection at serum sample date (p-value=0.49), age at serum sample date (p-value=0.96), and HIV viral load at serum sample date (p-value=0.12). Only CD4+ T-cell count at serum date remained evident after this selection (p-value=0.004).

We estimated adjusted mean ratios (MR_{adj}) and 95% confidence intervals (95% CI) for each SNP*serum level association of interest, adjusting for AIDS-NHL case status, race, and CD4+ T-cell count at date of serum sample (Tables 3.10 and 3.11). In this modeling scheme, we estimated differences in mean natural log transformed serum levels across each genotype of interest comparing homozygote minor allele carriers to homozygote major allele carriers, homozygote minor allele carriers combined with heterozygotes to homozygote major allele carriers, and under the log-additive model. Further, these miRNA levels were reported as back-transformed geometric means to illustrate differential mean serum levels across genotype

(Figures 3.4-3.8). Additionally, we stratified this analysis by AIDS-NHL status in order to assess any potential heterogeneity between cases and controls.

Figure 2.1 The Process from miRNA and Candidate Gene Selection, to Target Gene Determination and SNP Identification

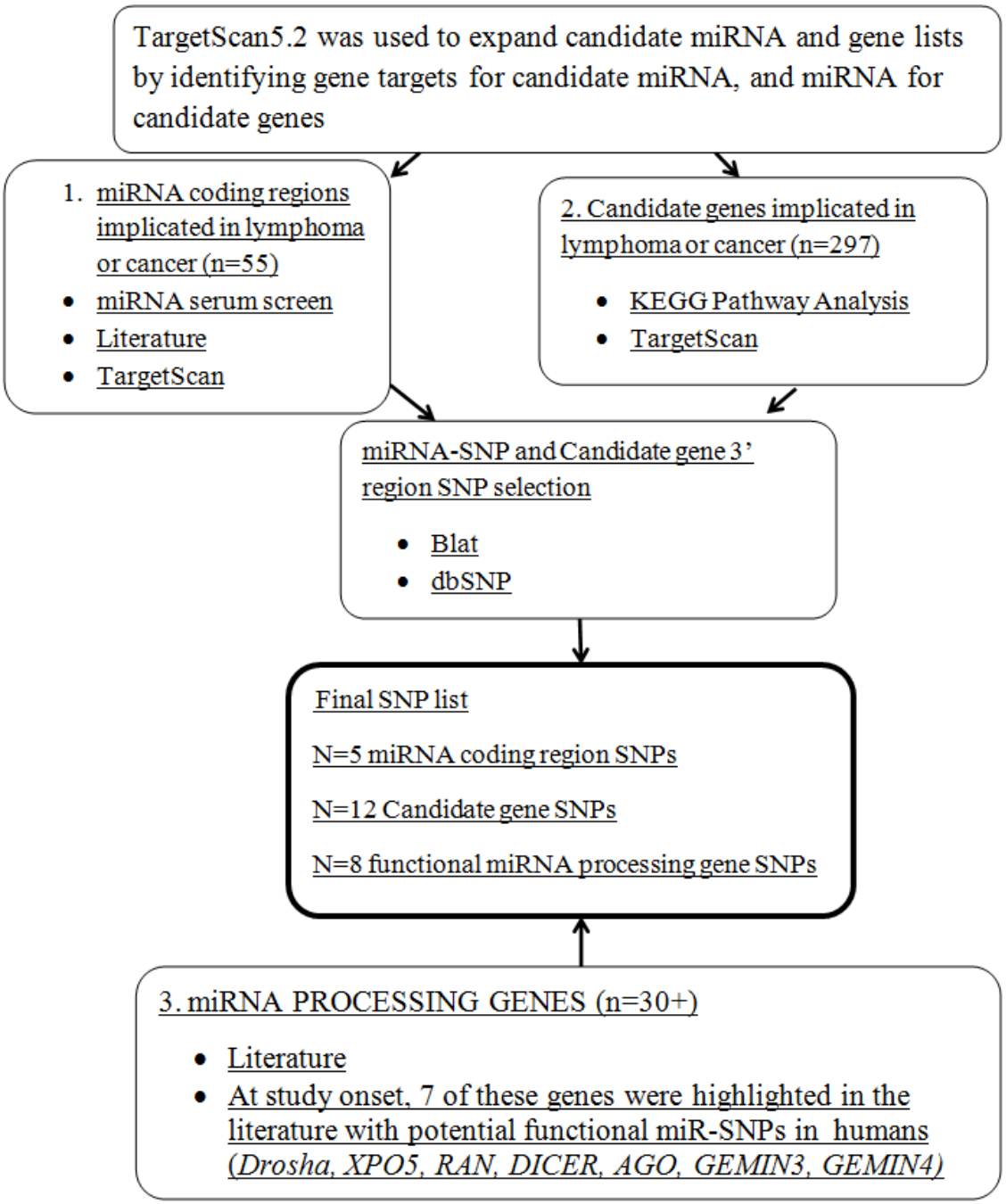
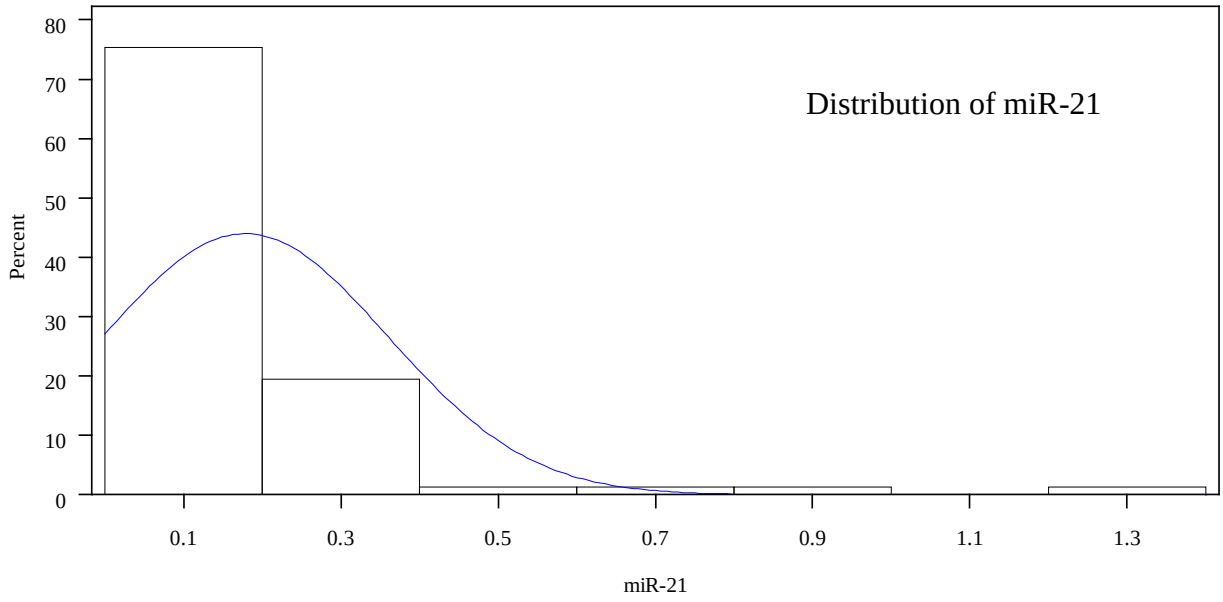


Table 2.1. SNPs genotyped with post-genotyping study MAF and Genotype Distributions

Gene	SNP	Allele and Amino Acid Change (if applicable)	Location and function (if applicable) of SNP	dbSNP MAF Global	Study MAF Overall	dbSNP Genotype Frequency			Study Genotype Frequency		
						Homozygote major allele	Heterozygote	Homozygote minor allele	Homozygote major allele	Heterozygote	Homozygote minor allele
miRNA coding regions											
miR-196a	rs11614913	C→T	Downstream of HOXC8 and HOXC9/ Noncoding RNA	0.38	0.38	0.34	0.44	0.22	0.38	0.49	0.13
miR-26a1	rs7372209	C→T	Within the CTDSPL gene/ Noncoding RNA	0.25	0.27	0.56	0.37	0.07	0.52	0.41	0.07
miR-27a	rs895819	T→C	Downstream of miR-181c/d and NANOS3/ Noncoding RNA	0.36	0.33	0.50	0.41	0.09	0.44	0.47	0.09
miR-300	rs12894467	C→T	Exon Chromosome 14q 32.31/ Noncoding RNA	0.43	0.41	0.31	0.50	0.19	0.34	0.49	0.17
pre-miR-146a	rs2910164	G→C	Upstream of miR-3142 and PTTG1 genes/ Noncoding RNA	0.38	0.22	0.59	0.35	0.06	0.60	0.36	0.04
miRNA processing and maturation genes											
Ago2	rs4961280	C→A	nearGene-5	0.14	0.19	0.77	0.22	0.02	0.66	0.31	0.03
Dicer1	rs3742330	A→G	3' UTR	0.15	0.09	0.85	0.15	0.00	0.82	0.17	0.01
Drosha	rs10719	C→T	3' UTR	0.47	------	0.62	0.35	0.03	------	------	------
Gemin3	rs197412	T→C Ile ⇒Thr	chr1: 112308953/ Missense	0.48	0.42	0.30	0.57	0.13	0.34	0.48	0.18
Gemin4	rs2740348	G→C Gln ⇒Glu	Chr17: 648186/ Missense	0.12	0.17	0.70	0.22	0.08	0.71	0.26	0.04
Gemin4	rs7813	C→T* Arg ⇒Cys	Chr17: 744946/ Missense	0.31	0.39	0.29	0.45	0.27	0.17	0.45	0.38
Ran	rs14035	C→T	nearGene-5	0.30	0.31	0.50	0.40	0.10	0.49	0.41	0.10
XPO5	rs11077	A→C	3' UTR	0.38	0.42	0.34	0.46	0.20	0.33	0.49	0.18
Candidate genes with SNPs near or within a predicted miRNA binding site											
CDK6	rs42031	A→T	3' UTR	0.15	0.20	------	------	------	0.64	0.32	0.04
CXCL12	rs1804429	T→G	3' UTR	0.05	0.03	0.89	0.12	0.00	0.93	0.07	0.00
E2F2	rs2075993	G→A*	3' UTR	0.42	0.49	0.25	0.56	0.20	0.25	0.48	0.26
HIF1A	rs2057482	C→T	3' UTR	0.21	0.16	0.82	0.16	0.02	0.70	0.28	0.03
IL15	rs10519613	C→A	3' UTR	0.20	0.10	0.81	0.19	0.01	0.80	0.19	0.01
IL6R	rs4072391	C→T	3' UTR	0.22	0.20	0.63	0.32	0.05	0.64	0.32	0.04
KRAS	rs9266	C→T	3' UTR	0.45	0.47	0.29	0.47	0.24	0.29	0.49	0.22
RCHY1	rs2126852	A→G	3' UTR	0.26	0.28	0.49	0.38	0.13	0.53	0.39	0.08
TAB3	rs3816757	C→G	3' UTR	0.21	0.22	0.72	0.17	0.12	0.78	0.00	0.22
TP53INP1	rs7760	T→G	3' UTR	0.17	0.13	0.85	0.15	0.00	0.76	0.22	0.02
TP53INP1	rs896849	T→C	3' UTR	0.20	0.17	0.81	0.17	0.02	0.70	0.27	0.03
WWOX	rs12828	A→G	3' UTR	0.44	0.42	0.41	0.44	0.15	0.35	0.47	0.18

*Allele changes presented represent those found on HAPMAP/dbSNP. The minor alleles of two SNPs, *E2F2* rs2075993 and *Gemin4* rs7813, as described in HAPMAP/dbSNP, were found to be the major alleles in our post-genotyped samples. We suspect this is due to the fact that the documented HAPMAP/dbSNP MAFs hovered around 0.50 (rs2075993: HAPMAP MAF A=0.47, major to minor GG/AG/AA; Our data MAF G=0.4937 and rs7813: HAPMAP MAF T=0.49; major to minor CC/CT/TT; Our data MAF C=0.3945).

Figure 2.2 Comparison of miRNA serum levels to natural log transformed miRNA serum levels to investigate normality



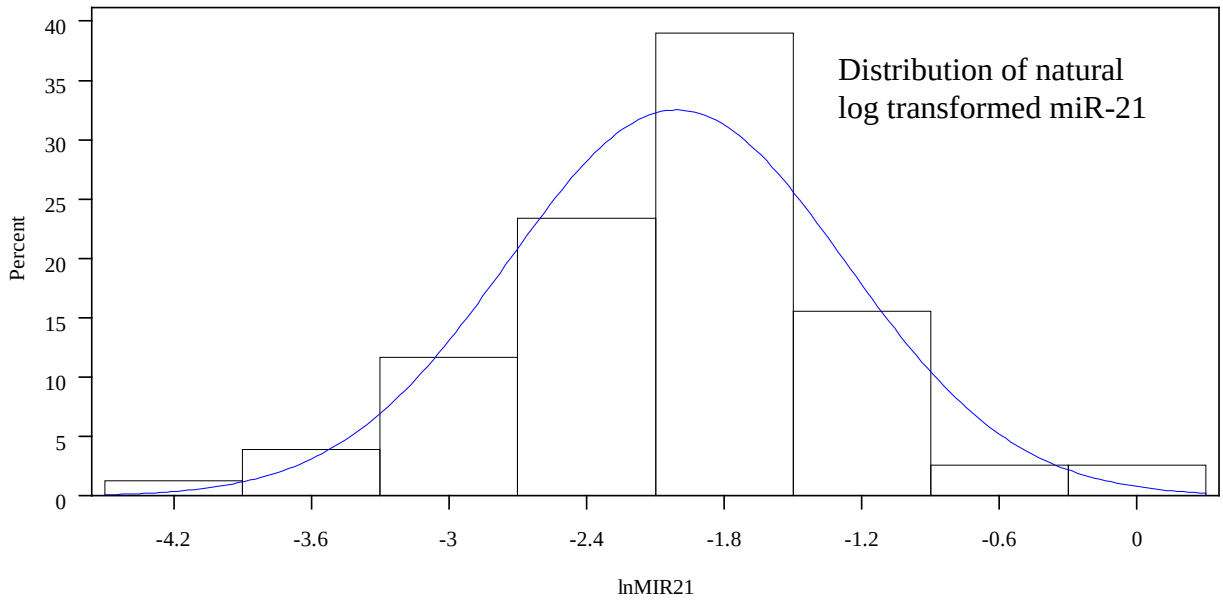
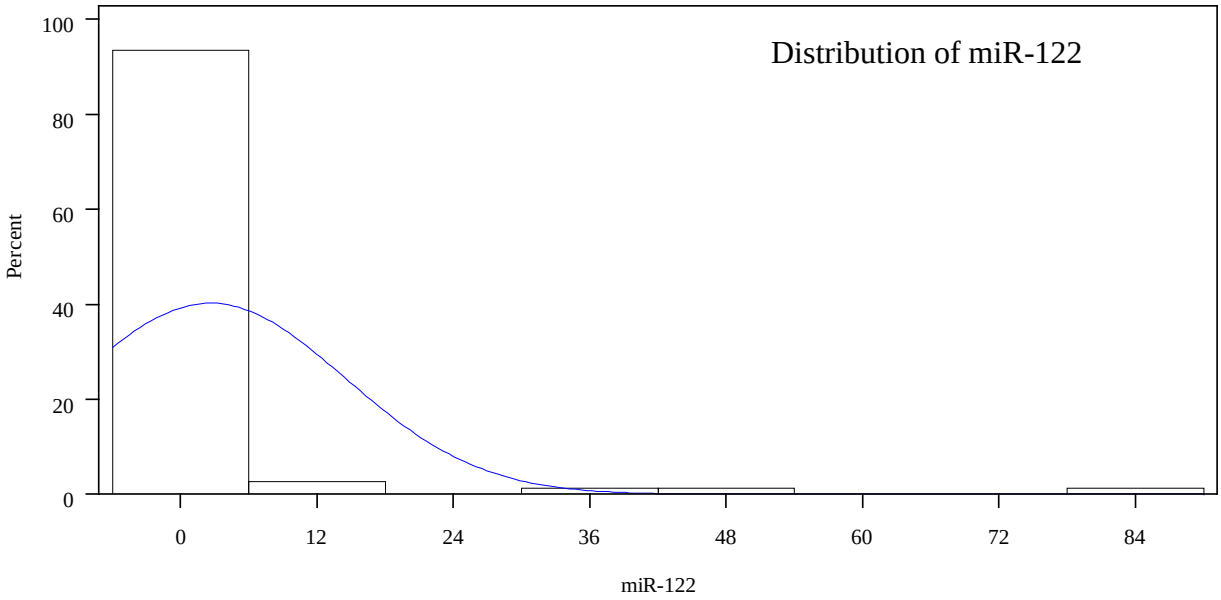


Figure 2.2 Continued. Comparison of miRNA serum levels to natural log transformed miRNA



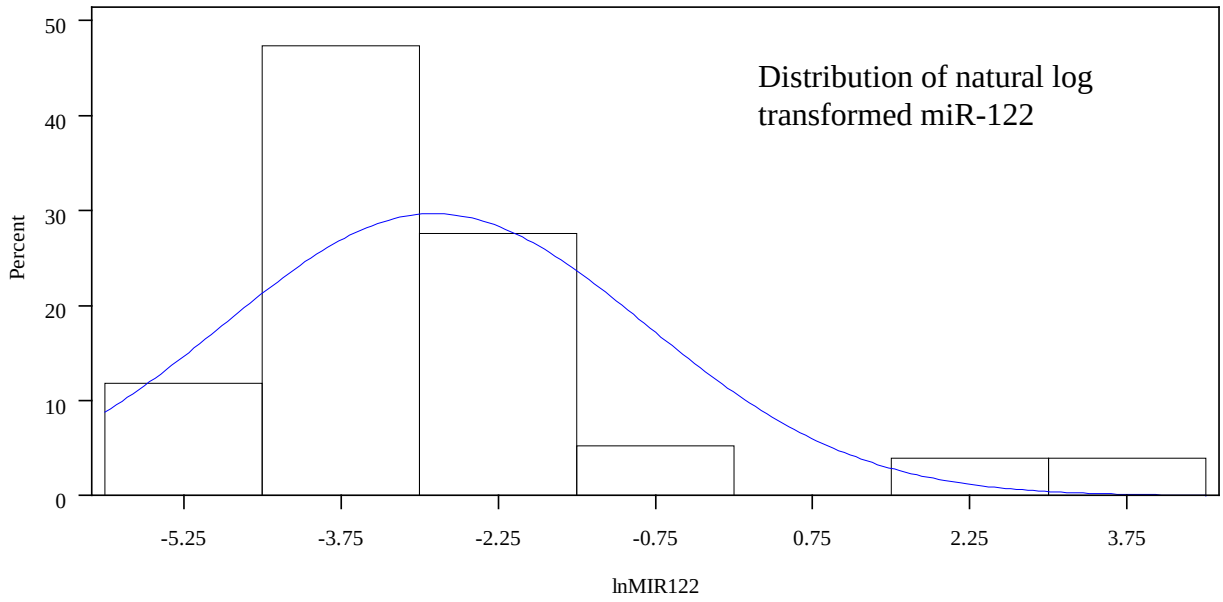
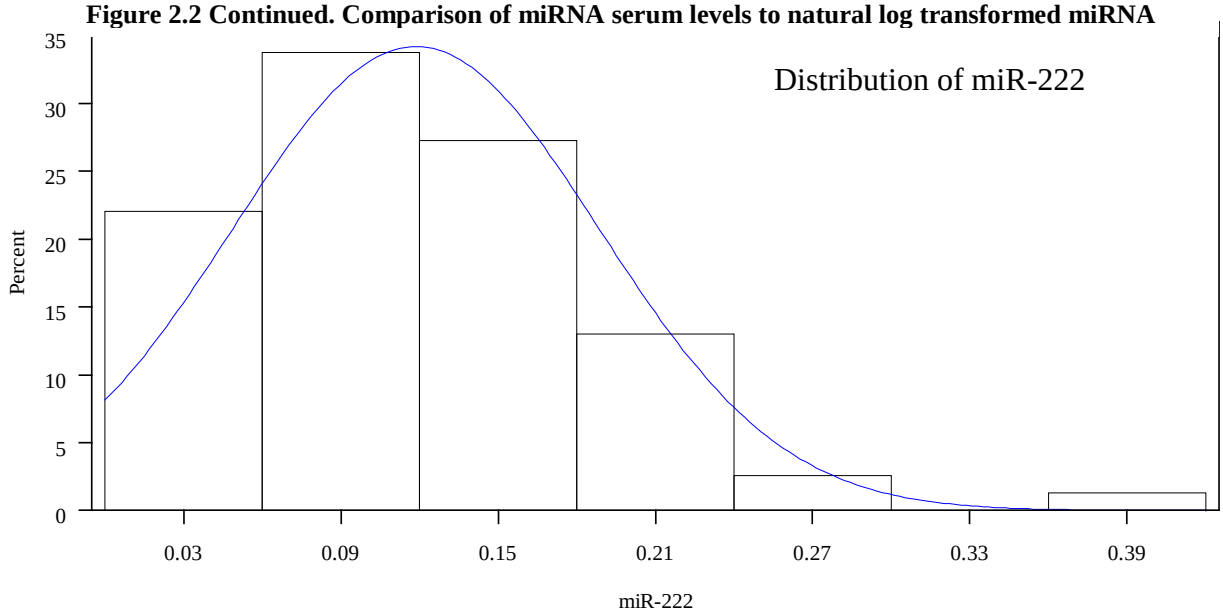


Figure 2.2 Continued. Comparison of miRNA serum levels to natural log transformed miRNA



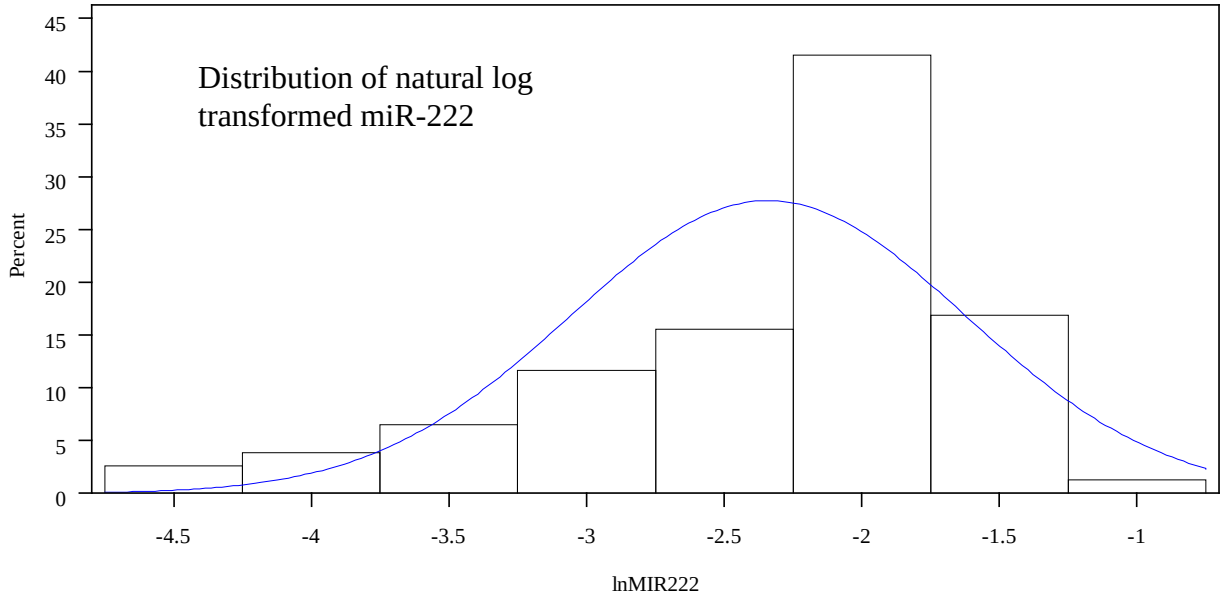
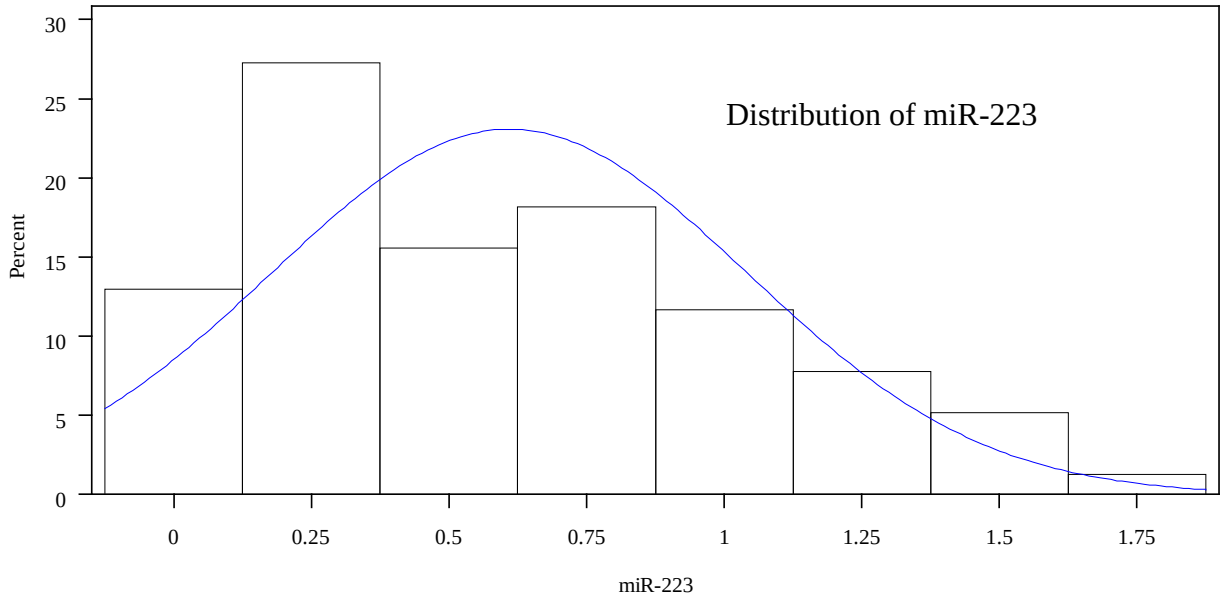
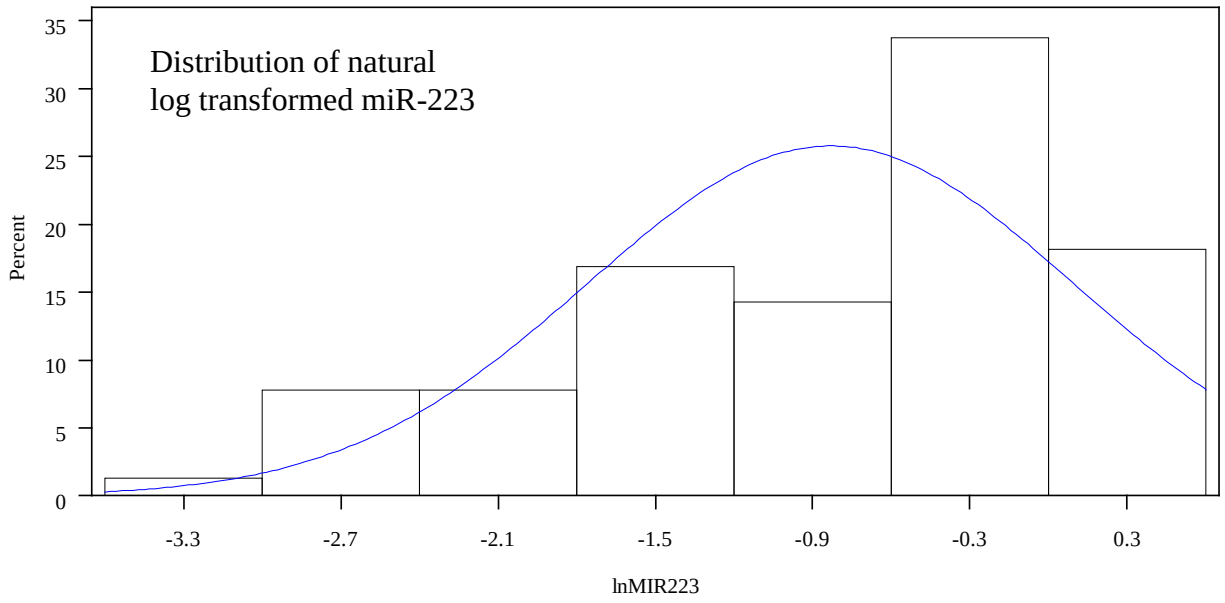


Figure 2.2 Continued. Comparison of miRNA serum levels to natural log transformed miRNA





CHAPTER 3: RESULTS

3.1. Genotyped SNPs

Table 3.1 displays the genotype distribution of the 22 SNPs successfully genotyped in this study. The distribution of allele frequencies for *miR-196a* rs11614913 differed between cases and controls (p-value=0.05), while the distribution of allele frequencies for *XPO5* rs11077 was suggested to differ between AIDS-NHL cases and HIV^+ controls (p-value=0.07). Through combining bioinformatic tools to inform our SNP selection, we identified the location of 11 SNPs previously undocumented as within or near a putative miRNA target site. Nine of these eleven SNPs were successfully genotyped on our platform, with putative binding locations noted in Table 3.1.

3.2. AIDS-NHL Case and HIV Positive Control Demographics

In our study population of 180 AIDS-NHL cases and 529 HIV^+ controls, AIDS-NHL cases were older than HIV^+ controls (Student's T-test, p-value =0.03; 180 matched sets are displayed in Table 3.2; Cohort comparison Table 3.3). The majority of both AIDS-NHL cases (48.3%) and HIV^+ controls (52.6%) were between the ages of 35-44 at reference date. Further, the majority of both AIDS-NHL cases and HIV^+ controls were Caucasian in this study (83.5% and 90.6%, respectively), with a larger proportion of AIDS-NHL cases reporting themselves as White-Hispanic (9.9%) and Black non-Hispanic (6.6%) than HIV^+ controls (4.7% and 4.0%, respectively in controls).

Individuals were matched on seroprevalence at time of recruitment or seroconversion during follow-up, and years of known infection time prior to reference date, so as to ensure cases and controls had the same HIV^+ follow-up time within the study. This is reflected in regards to the similarity of their proportional distribution. The similarity in distribution between $CD4^+$ T-cell counts at date of matching and pre-HAART $CD4^+$ T-cell counts slope among both AIDS-NHL cases and HIV^+ controls also reflects the effectiveness of matching in this study.

As the majority of these individuals entered into the MACS prior to the advent and widespread use of HAART, only six-percent of AIDS-NHL cases and under nine-percent of HIV^+ controls were exposed to HAART treatment prior to reference date. On the other hand, over sixty- and seventy- percent of AIDS-NHL cases and HIV^+ controls, respectively, were exposed to ART treatment, with a significantly larger proportion of AIDS-NHL cases compared to HIV^+ controls reporting no ART use prior to reference date (38.5% vs. 29.1%, Chi-squared p-value=0.02). Further, higher mean HIV viral loads, both at set-point prior to HAART and closest measurement one year prior to reference date, were evident in AIDS-NHL cases compared to HIV^+ controls. In addition to ART use, HAART use, and HIV viral load, another main cofactor for AIDS-NHL susceptibility is the transition to AIDS. This was reflected in our study as over fifty percent of AIDS-NHL cases were diagnosed as having transitioning to AIDS prior to reference date in comparison to only forty-one percent of HIV^+ controls (Chi-square p-value= 0.01).

Among those who had transitioned to AIDS prior to reference date, the largest proportion of diagnoses classified as AIDS-defining events were attributable to Pneumoncystis pneumonia (PCP; 35.8% in AIDS-NHL cases and 38.7% of diagnoses in HIV^+ controls) and Kaposi's Sarcoma (KS; 21.1% in AIDS-NHL cases and 19.8% of diagnoses in HIV^+ controls).

AIDS-NHL cases and HIV^+ controls did not differ significantly among covariates such as history of tobacco and alcohol use, or substance abuse.

Last, among AIDS-NHL cases, the majority (n=123, 68.1%) of NHL cases were of systemic origin compared to those of the primary central nervous system (CNS; n=57, 31.9%). Among systemic NHL cases, nearly thirty-percent were of the diffuse large b-cell lymphoma subtype compared to only seventeen-percent of those diagnosed with Burkitt's lymphoma. EBV positivity was tested in 47.7% of cases (86/180), 58 tumors which were of systemic origin and 28 tumors which were CNS-site specific. Among those with systemic tumors, 56.9% tested positive for EBV infection (33/58), while 89.3% of CNS tumors were EBV positive (25/28).

3.3. microRNA-related SNPs and AIDS-NHL Susceptibility

3.3.1. SNPs and Overall, Systemic and CNS AIDS-NHL Susceptibility

Table 3.4 displays the main effect associations between miRNA-related SNPs of interest and overall, systemic and CNS AIDS-NHLs, while figures 3.1-3.3 illustrate an overview of log-additive p-values for each analysis. For the associations between miRNA-related SNPs and overall AIDS-NHL, i) complete case and ii) analyses imputed for $CD4^+$ T-cell count at reference date and HIV viral load at set point, are displayed. These analyses were adjusted for age at reference date, HIV viral load at set point prior to HAART, AIDS diagnosis prior to reference date, history of ART use, $CD4^+$ T-cell count at reference date, race, and history of HCV infection. Cases and controls were matched on recruitment period duration of HIV positive and cancer-free follow-up time, race, $CD4^+$ T-cell count at matching, and seroprevalence at time of recruitment or seroconversion during follow-up. The following discussed results focus on the

analyses imputed for $CD4^+$ T-cell count at reference date and HIV viral load at set point, as discussed in Chapter 2. Further, to give a comprehensive overview of our results, we present those that were suggested to be statistically significant (alpha=0.10) and those statistically significant at alpha=0.05, while noting and adding highlighted emphasis to those that remained statistically significant, or in other words "evident", after the SB approach (posterior SB estimates for significant results available for viewing in Table 3.7).

The variant allele (T) of *microRNA-196a2* rs11614913 was suggested as inversely associated with AIDS-NHL overall (OR_{adj} = 0.80 per variant allele; 95% CI: 0.61-1.05). A positive association was suggested between *microRNA-27* rs895819 and AIDS-NHL overall (OR_{adj} = 1.31 per variant allele; 95% CI: 0.97-1.76). A positive association was observed between *GEMIN3* rs197412 and AIDS-NHL overall which remained evident after the SB approach (OR_{adj} = 1.35 per variant allele; 95% CI: 1.03-1.78). The variant allele (G) of *RCHY1* rs2126852 was suggested as inversely associated with AIDS-NHL overall (OR_{adj} = 0.81 per variant allele; 95% CI: 0.60-1.09).

In site specific analysis, the variant allele (T) of *microRNA-196a2* rs11614913 was inversely associated with CNS AIDS-NHL and remained evident after the SB approach (TT/CT vs. CC: OR_{adj} = 0.43; 95% CI: 0.22-0.87). Unobserved in AIDS-NHL overall, an association was suggested between *microRNA-300* rs12894467 and systemic AIDS-NHL (TT vs. CT/CC: OR_{adj} = 1.67; 95% CI: 0.95-2.96). A positive association between *GEMIN3* rs197412 and systemic AIDS-NHL was also suggested (OR_{adj} = 1.33 per variant allele; 95% CI: 0.96-1.84). Unobserved in AIDS-NHL overall, a positive association was suggested between *GEMIN4* rs7813 and systemic AIDS-NHL (OR_{adj} = 1.27 per variant allele; 95% CI: 0.92-1.76). Although unobserved in AIDS-NHL overall, *HIF1A* rs2057482 was associated with systemic and CNS AIDS-NHL in

a differential manner: *HIF1A* rs2057482 was positively associated with systemic AIDS-NHL, whereas among those with CNS AIDS-NHL, an inverse association was observed (systemic AIDS-NHL OR_{adj} = 1.83 per variant allele; 95% CI: 1.16-2.90; CNS AIDS-NHL OR_{adj} = 0.38 per variant allele; 95% CI: 0.17-0.87). These associations remained evident after SB analysis. The variant allele (G) of *RCHY1* rs2126852 was suggested as inversely associated with systemic AIDS-NHL (OR_{adj} = 0.74 per variant allele; 95% CI: 0.52-1.06). Results investigating the association between these miRNA-related SNPs and overall AIDS-NHL under unconditional logistic regression are presented in Table 3.8. There are no major discrepancies between the overall results from conditional logistic regression and unconditional logistic regression.

3.3.2. Association of microRNA-196a rs11614913 and HIF1A rs2057482 on Overall, Systemic and CNS AIDS-NHL

We investigated the multiplicative interaction and joint effects between a SNP located within the *microRNA-196a* coding region (*microRNA-196a* rs11614913) and a second SNP located within a *microRNA-196a* binding site (*HIF1A* rs2057482) and overall, systemic and CNS AIDS-NHL, under a dominant model (Table 3.5). We did not observe multiplicative interaction between these two SNPs and AIDS-NHL, as reflected in the ratio of odds ratios (overall AIDS-NHL: ROR_{adj}= 0.75; 95% CI: 0.34-1.65; systemic AIDS-NHL: ROR_{adj}= 0.48; 95% CI: 0.18-1.27; CNS AIDS-NHL: ROR_{adj}= 1.18; 95% CI: 0.22-6.42). Among those with systemic AIDS-NHL, a joint effect between the variant genotypes of these two SNPs was suggested in comparison to individuals wild-type for both SNPs, however this association was not as pronounced compared to those with the *HIF1A* rs2057482 variant and *miR-196a* wild-type genotypes (Joint effect: OR_{adj11} = 2.14; 95% CI: 1.04-4.39; *HIF1A* rs2057482: OR_{adj10} =3.28; 95% CI: 1.42-7.57). Among those with CNS AIDS-NHL, the joint effect of the two variant SNPs resulted in a more

prominent inverse association compared to either SNP singularly, although limited sample size is of prominent concern (4 cases compared to 39 controls; Joint effect: OR_{adj11} = 0.10; 95% CI: 0.02-0.45; *HIF1A* rs2057482: OR_{adj10} =0.22; 95% CI: 0.06-0.81; *miR-196a* rs11614913: OR_{adj01} =0.38; 95% CI: 0.17-0.88).

3.3.3. Subgroup analyses: SNPs and Overall AIDS-NHL among Caucasians and those without a Primary Kaposi's Sarcoma Diagnosis

Table 3.6 displays the main effect associations between selected miRNA-related SNPs (those associated or suggested as associated with AIDS-NHL in the main effects analyses; Table 3.4) and overall AIDS-NHL among Caucasians (n=631) and those without a primary Kaposi's Sarcoma Diagnosis prior to reference date (n=646).

The inverse association suggested among all study participants between the variant allele (T) of *miR-196a* rs11614913 and overall AIDS-NHL was mimicked among Caucasian-only participants (OR_{adj} = 0.76 per variant allele; 95% CI: 0.57-1.02). The suggested association observed among all study participants between *miR-27a* rs895819 and overall AIDS-NHL was mimicked among Caucasian-only participants (OR_{adj} = 1.28 per variant allele; 95% CI: 0.92-1.79). The suggested association between *microRNA-300* rs12894467 and AIDS-NHL remained among Caucasian-only participants (OR_{adj} = 1.25 per variant allele; 95% CI: 0.92-1.70). The association observed between *GEMIN3* rs197412 and overall AIDS-NHL was suggested among Caucasians (OR_{adj} = 1.34 per variant allele; 95% CI: 1.00-1.81). A suggested inverse association between *RCHY1* rs2126852 and overall-AIDS NHL persisted among Caucasians (OR_{adj} = 0.79 per variant allele; 95% CI: 0.57-1.09).

The inverse association suggested among all study participants between the variant allele (T) of *miR-196a* rs11614913 and overall AIDS-NHL was mimicked among those without a primary KS diagnosis (OR_{adj} = 0.77 per variant allele; 95% CI: 0.57-1.03). The association observed between *GEMIN3* rs197412 and overall AIDS-NHL became even stronger among those without a primary KS diagnosis and remained evident after SB analysis (OR_{adj} = 1.48 per variant allele; 95% CI: 1.09-1.99).

3.3.4. Investigation of Study Result Robustness

Table 3.9 displays associations from three SNPs highlighted of interest in this study and overall AIDS-NHL (*GEMIN3* rs197412, *miR-196a* rs11614913 and *HIFIA* rs2057482) under four modeling schemes to investigate study result robustness. This table displays results from the complete case analyses, median imputed analyses, multiply imputed sensitivity analyses, a model not adjusted for HIV viral load at set point, and a model adjusted for matching factors only. Let us take for example, the associations between *GEMIN3* rs197412 and AIDS-NHL overall under a log-additive model. In the complete case analysis, where covariates with missing data included HIV viral load at set point and $CD4^+$ T-cell count at reference date, a marginal association was suggested between *GEMIN3* rs197412 and AIDS-NHL (OR_{adj} = 1.27 per variant allele; 95% CI: 0.90-1.79). In the median imputed analysis, towards which we focused our discussion, where the control value of 4.316 vc/ml was applied to missing HIV viral load at set point data and the control value of 113 cells/mm^3 was applied to missing $CD4^+$ T-cell count at reference date data, an association was observed between *GEMIN3* rs197412 and AIDS-NHL (OR_{adj} = 1.35 per variant allele; 95% CI: 1.03-1.78). In the multiply imputed sensitivity analysis for HIV viral load at set point and $CD4^+$ T-cell count at reference date, an association was observed between *GEMIN3* rs197412 and AIDS-NHL that remained very similar to that found under the median

imputed analysis (OR_{adj} = 1.36 per variant allele; 95% CI: 1.04-1.79). In the analysis unadjusted for HIV viral load at set point, a marginal association was suggested between *GEMIN3* rs197412 and AIDS-NHL (OR_{adj} = 1.29 per variant allele; 95% CI: 0.98-1.69). And last, in the analysis adjusting only for matching factors, a marginal association was also suggested between *GEMIN3* rs197412 and AIDS-NHL (OR_{adj} = 1.26 per variant allele; 95% CI: 0.98-1.62). Thus, a pattern suggesting an association between *GEMIN3* rs197412 and AIDS-NHL susceptibility is detectable across these results. A similar pattern suggesting an inverse association between *miR-196a* rs11614913 and overall AIDS-NHL susceptibility also remained robust across different analytic strategies, as did a pattern of no association between *HIFIA* rs2057482 and overall AIDS-NHL (Table 3.9).

3.4. Associations between SNPs in miRNA Processing Genes and miRNA-21, miRNA-122, miRNA-222, and miRNA-223 serum levels (Specific Aim 2)

Table 3.10 displays the main effect associations between SNPs in miRNA processing genes and miRNA-21, miRNA-122, miRNA-222, and miRNA-223 serum levels, adjusted for AIDS-NHL status, race and CD4+ count at date of serum sample, among a subgroup of 77 MACS participants. Table 3.11 displays these main effect associations stratified by AIDS-NHL status.

The variant genotype (CC) and allele (C) of *GEMIN3* rs197412 were suggested as associated with higher mean $\log_e$ transformed miRNA-21 serum levels than those with the referent genotype (MR_{adj} = 1.24 per variant allele; 95% CI: 0.97-1.59; CC/CT vs. TT: MR_{adj} = 1.36; 95% CI: 0.95-1.94). This suggested difference in serum level by genotype is reflected in Figure 3.4 which displays mean miR-21 serum levels by the *GEMIN3* rs197412 genotype.

Individuals with the homozygous minor allele genotype exhibited miR-21 relative expression levels of 17.9% compared to 10.8% relative miR-21 expression in individuals with the homozygous major allele genotype. When stratified by cases and controls, this weak association became less prominent, and it is important to note the limited sample size for investigation, especially in controls. No heterogeneity was reflected between AIDS-NHL cases and controls, as reflected in the largely overlapping confidence intervals.

The variant genotype (CC) and allele (C) of *GEMIN3* rs197412 were suggested as associated with higher mean $\log_e$ transformed miRNA-222 serum levels than those with the referent genotype (MR_{adj} = 1.21 per variant allele; 95% CI: 0.98-1.49; CC/CT vs. TT: MR_{adj} = 1.35; 95% CI: 0.99-1.82). This suggested difference in serum level by genotype is reflected in Figure 3.5 which displays mean miR-222 serum levels by the *GEMIN3* rs197412 genotype. Individuals with the homozygous minor allele genotype exhibited miR-222 relative expression levels of 13.2% compared to 7.5% relative miR-222 expression in individuals with the homozygous major allele genotype. When stratified by cases and controls, this association became less prominent but remained suggested among cases and controls (AIDS-NHL cases: MR_{adj} = 1.13 per variant allele; 95% CI: 0.90-1.42; controls: MR_{adj} = 1.61 per variant allele; 95% CI: 0.89-2.89). No heterogeneity was reflected between AIDS-NHL cases and controls.

The variant allele (C) of *GEMIN3* rs197412 was associated with higher mean $\log_e$ transformed miRNA-223 serum levels than those with the referent genotype (MR_{adj} = 1.31 per variant allele; 95% CI: 0.96-1.78). This suggested difference in serum level by genotype is reflected in Figure 3.6 which displays mean miR-223 serum levels by the *GEMIN3* rs197412 genotype. Individuals with the homozygous minor allele genotype exhibited miR-223 relative expression levels of 64.9% compared to 33.2% relative miR-223 expression in individuals with

the homozygous major allele genotype. When stratified by cases and controls, this weak association remained among cases (MR_{adj} = 1.38 per variant allele; 95% CI: 0.98-1.94), however it is important to note the limited sample size for investigation, especially in controls. No heterogeneity was reflected between AIDS-NHL cases and controls, as reflected in the largely overlapping confidence intervals.

An inverse yet marginal association was suggested between the CT genotype of *RAN* rs14035 and mean log_e transformed miRNA-21 serum levels (MR_{adj} = 0.81 per variant allele; 95% CI: 0.62-1.06; CT/TT vs. CC: MR_{adj} = 0.73; 95% CI: 0.53-1.02). This suggested difference in serum level by genotype is reflected in Figure 3.7 which displays mean miR-21 serum levels by the *RAN* rs14035 genotype. Individuals with the homozygous minor allele genotype exhibited miR-21 relative expression levels of 12.3% compared to 15.8% relative miR-21 expression in individuals with the homozygous major allele genotype. When stratified by cases and controls, this weak association remained among cases (CT vs. CC: MR_{adj} = 0.74; 95% CI: 0.50-1.08), however it is important to note the limited sample size for investigation, especially in controls. No heterogeneity was reflected between AIDS-NHL cases and controls, as reflected in the largely overlapping confidence intervals.

An inverse yet marginal association was suggested between the CT genotype of *RAN* rs14035 and mean log_e transformed miRNA-223 serum levels (CT vs. CC: MR_{adj} = 0.69; 95% CI: 0.45-1.06), which remained suggested among cases when stratified by case status (CT vs. CC: MR_{adj} = 0.63; 95% CI: 0.39-1.02), however it is important to note the limited sample size for investigation, especially in controls. No heterogeneity was reflected between AIDS-NHL cases and controls.

An association was marginally suggested between the variant C allele of *XPO5* rs11077 and mean $\log_e$ transformed miRNA-223 serum levels (MR_{adj} = 1.24 per variant allele; 95% CI: 0.91-1.68; AC/CC vs. AA: MR_{adj} = 1.46; 95% CI: 0.92-2.31), which remained suggested among cases when stratified by case and control status (cases MR_{adj} = 1.30 per variant allele 95% CI: 0.91-1.84), however limited sample size is of concern regarding control OR estimation. No heterogeneity was reflected between AIDS-NHL cases and controls. This weakly suggested difference in serum level by genotype is reflected in Figure 3.8, which displays mean miR-223 serum levels by the *XPO5* rs11077 genotype. Individuals with the homozygous minor allele genotype exhibited miR-223 relative expression levels of 47.6% compared to 35.5% relative miR-223 expression in individuals with the homozygous major allele genotype.

Table 3.1. Distribution of SNPs among AIDS-NHL Cases and HIV$^+$ Controls

Genotype	AIDS-NHL Cases	HIV$^+$ Controls	Chi Sq P-Value
AGO2			
rs4961280			
CC	119 (65.4)	348 (65.8)	0.71
CA	56 (30.8)	167 (31.6)	
AA	7 (3.9)	14 (2.7)	
Missing/No Call	0 (0.0)	0 (0.0)	
CDK6 rs42031			
*Within 30 base pairs of a putative miR-26 binding site			
AA	117 (64.3)	335 (63.3)	0.80
AT	56 (30.8)	174 (32.9)	
TT	8 (4.4)	19 (3.6)	
Missing/No Call	1 (0.55)	1 (0.19)	
DICER1			
rs3742330			
AA	152 (83.5)	430 (81.3)	0.82
AG	29 (15.9)	96 (18.2)	
GG	1 (0.55)	3 (0.57)	
Missing/No Call	0 (0.0)	0 (0.0)	
E2F2 rs2075993			
*In 30 base pairs of a putative Let-7 binding site			
AA	45 (24.7)	134 (25.3)	0.71
AG	93 (51.1)	252 (47.6)	
GG	44 (24.2)	142 (26.8)	
Missing/No Call	0 (0.0)	1 (0.19)	
GEMIN3			
rs197412			
TT	54 (29.7)	189 (35.7)	0.14
TC	88 (48.4)	254 (48.0)	
CC	40 (22.0)	86 (16.3)	
Missing/No Call	0 (0.0)	0 (0.0)	
GEMIN4			
rs2740348			
GG	124 (68.1)	376 (71.1)	0.68
GC	50 (27.5)	134 (25.3)	
CC	8 (4.4)	18 (3.4)	
Missing/No Call	0 (0.0)	1 (0.19)	
GEMIN4			
rs7813			
TT	68 (37.4)	200 (37.8)	0.62
TC	79 (43.4)	242 (45.8)	
CC	35 (19.2)	85 (16.1)	
Missing/No Call	0 (0.0)	2 (0.38)	
HIF1A			
rs2057482			
CC	125 (68.7)	369 (69.9)	0.74
CT	51 (28.0)	147 (27.8)	
TT	6 (3.3)	12 (2.3)	
Missing/No Call	0 (0.0)	1 (0.19)	

Table 3.1 Cont. Distribution of SNPs among AIDS-NHL Cases and HIV$^+$ Controls

Genotype	AIDS-NHL Cases	HIV$^+$ Controls	Chi Sq P-Value
IL15 rs10519613			
*Within a putative miR-203 binding site			
CC	137 (75.3)	433 (81.9)	0.18
CA	42 (23.1)	91 (17.2)	
AA	2 (1.1)	5 (0.95)	
Missing/No Call	1 (0.55)	0 (0.0)	
IL6R rs4072391			
*In 30 base pairs of a putative miR-34 Family/miR-23 binding site			
CC	117 (64.3)	337 (63.7)	0.98
CT	58 (31.9)	169 (32.0)	
TT	7 (3.9)	22 (4.2)	
Missing/No Call	0 (0.0)	1 (0.19)	
KRAS rs9266			
*In 15 base pairs of a putative miR-181 binding site			
CC	53 (29.1)	151 (28.5)	0.84
CT	85 (46.7)	261 (49.3)	
TT	43 (23.6)	117 (22.1)	
Missing/No Call	1 (0.55)	0 (0.0)	
miR-196a rs11614913			
CC	72 (39.6)	196 (37.1)	0.40
CT	90 (49.5)	257 (48.6)	
TT	19 (10.4)	76 (14.3)	
Missing/No Call	2 (1.10)	0 (0.0)	
miR-26a1 rs7372209			
CC	95 (52.2)	276 (52.2)	0.55
CT	78 (42.9)	215 (40.6)	
TT	9 (5.0)	38 (7.2)	
Missing/No Call	0 (0.0)	0 (0.0)	
miR-27a rs895819			
TT	68 (37.4)	242 (45.8)	*0.08*
TC	92 (50.6)	242 (45.8)	
CC	22 (12.1)	43 (8.1)	
Missing/No Call	0 (0.0)	2 (0.38)	
MIR-300 rs12894467			
CC	58 (31.9)	183 (34.7)	0.20
CT	86 (47.3)	265 (50.2)	
TT	38 (20.9)	80 (15.2)	
Missing/No Call	0 (0.0)	1 (0.19)	
pre-miR-146a rs2910164			
GG	111 (61.0)	314 (59.4)	0.80
GC	62 (34.1)	193 (36.5)	
CC	9 (5.0)	22 (4.2)	
Missing/No Call	0 (0.0)	0 (0.0)	

Table 3.1 Continued. Distribution of SNPs among AIDS-NHL Cases and HIV$^+$ Controls

Genotype	AIDS-NHL Cases	HIV$^+$ Controls	Chi Sq P-Value
RAN			
rs14035			
CC	87 (47.8)	257 (48.6)	0.59
CT	73 (40.1)	222 (42.0)	
TT	22 (12.1)	50 (9.5)	
Missing/No Call	0 (0.0)	0 (0.0)	
RCHY1 rs2126852			
*In 2 base pairs of a putative miR-153 binding site			
AA	104 (57.1)	272 (51.4)	0.29
AG	67 (36.8)	210 (39.7)	
GG	11 (6.0)	47 (8.9)	
Missing/No Call	0 (0.0)	0 (0.0)	
TP53INP1 rs7760			
*In 30 base pairs of a putative miR-24/miR-153 binding site			
TT	135 (74.2)	407 (77.0)	0.75
TG	43 (23.6)	111 (20.9)	
GG	4 (2.2)	11 (2.1)	
Missing/No Call	0 (0.0)	0 (0.0)	
TP53INP1 rs896849			
*Within a putative miR-155 binding site			
TT	122 (67.0)	375 (70.9)	0.12
TC	50 (27.5)	140 (26.5)	
CC	10 (5.5)	13 (2.5)	
Missing/No Call	0 (0.0)	1 (0.19)	
XPO5			
rs11077			
AA	62 (34.1)	175 (33.1)	0.49
AC	82 (45.1)	265 (50.1)	
CC	36 (19.8)	89 (16.8)	
Missing/No Call	2 (1.1)	0 (0.0)	
CXCL12 rs1804429			
*Within a putative miR-23a/b binding site			
TT	164 (90.1)	498 (94.1)	0.10
TG	18 (9.9)	30 (5.7)	
GG	0 (0.0)	1 (0.19)	
Missing/No Call	0 (0.0)	0 (0.0)	

Table 3.2 Demographic Characteristics of AIDS-NHL Cases and HIV$^+$ Controls in the MACS

	All AIDS-NHL Cases	HIV$^+$ Controls	*p**
Total, N	180	529	
Age at reference date, mean years (range; SD)	41.8 (24.9-61.3; 7.73)	40.4 (24.1-70.3; 7.53)	**0.03**
Age at reference date in years, n (%)			*0.10*
Less than 25 years of age	1 (0.55)	5 (0.95)	
25-29	12 (6.6)	28 (5.3)	
30-34	21 (11.5)	92 (17.4)	
35-39	44 (24.7)	139 (26.3)	
40-44	43 (23.6)	139 (26.3)	
45-49	27 (15.4)	72 (13.6)	
50-54	21 (11.5)	30 (5.7)	
55-59	8 (4.4)	17 (3.2)	
60-64	3 (1.7)	3 (0.57)	
65+	0 (0.0)	4 (0.76)	
Ethnicity, n (%)*			**0.02**
White, non-Hispanic	152 (84.4)	479 (90.6)	
Non-White	28 (15.6)	50 (9.5)	
Ethnicity, n (%)			**0.02**
White, non-Hispanic	152 (84.4)	479 (90.6)	
Black, non-Hispanic	10 (6.6)	21 (4.0)	
White, Hispanic	18 (11.8)	25 (4.7)	
Other	0 (0.0)	4 (0.76)	
MACS cohort, n (%)*			0.71
1984-1985 Recruits	161 (89.0)	476 (90.0)	0.15
1984 Recruits	145 (90.1)	408 (85.7)	
1985 Recruits	16 (9.9)	68 (14.3)	
1987-1991 Recruits	18 (11.0)	53 (10.0)	0.84
1987 Recruits	11 (61.1)	36 (67.9)	
1988 Recruits	2 (11.1)	8 (15.1)	
1989 Recruits	2 (11.1)	4 (7.5)	
1990 Recruits	1 (5.5)	3 (5.6)	
1991 Recruits	2 (11.1)	2 (3.7)	
Reference Year, n (%)			0.99
1984-1989	47 (26.4)	144 (27.2)	
1990-1995	106 (58.2)	302 (57.1)	
1996-2000	21 (11.5)	62 (11.7)	
2001-2006	6 (3.9)	21 (4.0)	
HIV status at cohort entry, n (%)*			0.23
Prevalent	163 (89.6)	478 (90.4)	
Converter	17 (10.4)	51 (9.6)	
*During follow-up, years of known HIV infection prior to reference date, mean (range; SD)****	7.2 (0.06-21.4; 3.8)	7.3 (0.09-21.4; 3.8)	0.87

Cases and controls were matched on these factors

Table 3.2 Cont. Demographic Characteristics of NHL Cases and Controls in the MACS

	All AIDS-NHL Cases	HIV$^+$ Controls	
HAART prior to reference date, n (%)			0.26
Never	170 (94.0)	483 (91.3)	
Yes	10 (6.0)	46 (8.7)	
Years from first HAART date to reference date, mean (range; SD)	3.9 (0.07-9.7; 2.9)	2.9 (0.08-9.5; 2.8)	0.29
ART prior to reference date, n (%)			**0.02**
Never	69 (38.5)	154 (29.1)	
Yes	111 (61.5)	375 (71.0)	
Years from first ART date to reference date, mean (range; SD)	3.4 (0.07-15.7; 2.6)	3.2 (0.04-17.2; 2.9)	0.52
CD4+ T-cell count slope pre-HAART, mean (range; SD)	-72.8 (-283.4-178.3; 73.16)	-67.9 (-370.0-812.2; 69.86)	0.43
Missing, n (%)	24 (13.2)	4 (0.8)	
CD4+ T-cell count at date of matching (mean)	178.9 (2.0-923.0; 214.6)	184.1 (3.0-1361.0; 224.7)	0.78
Missing, n (%)	0.0 (0.0)	0.0 (0.0)	
CD4+ T-cell count at date of matching, n (cells/ml;%)*			0.99
0-49	73 (40.7)	212 (40.1)	
50-99	25 (13.7)	74 (14.0)	
100-199	28 (15.4)	81 (15.3)	
200-299	14 (7.7)	42 (7.9)	
300-399	14 (7.7)	35 (6.6)	
400-499	7 (3.9)	28 (5.3)	
500+	19 (11.0)	57 (10.8)	
NHL subtypes, n (%)			
Primary Central Nervous System Lymphoma	57 (31.9)	-	
Systemic NHL	123 (68.1)	-	
Diffuse Large B-Cell	36 (29.0)	-	
Diffuse Large B-Cell, Immunoblastic	28 (22.6)	-	
NHL/Lymphoma, Not Specified	32 (25.8)	-	
Burkitt's Lymphoma	21 (16.9)	-	
Others	6 (5.6)	-	
Tumor tested for EBV infection, n (%)			
Tested	86 (47.7)	-	
Not tested	87 (48.3)	-	
Missing	7 (3.9)	-	
Tumor tested positive for EBV infection, n (%)	58 (67.4)	-	
EBV tumor positivity by site, n (%)			
Proportion of tumors tested EBV positive at a systemic site out of 58 systemic tumors tested, n (%)	33 (56.9)	-	
Proportion of tumors tested EBV positive at a CNS site out of 28 CNS tumors tested, n (%)	25 (89.3)	-	
HCV infection prior to reference date, n (%)			**0.09**
Never	156 (86.7)	482 (91.1)	
Ever	24 (13.3)	47 (8.9)	
Among Ever, those with Chronic HCV infection, n (% of total study population)	17 (9.4)	33 (6.2)	0.20

Cases and controls were matched on these factors

Table 3.2 Cont. Demographic Characteristics of NHL Cases and Controls in the MACS

	All AIDS-NHL Cases	HIV$^+$ Controls	
AIDS diagnosis prior to reference date, n (%)			**0.01**
No	86 (47.8)	312 (59.0)	
Yes	94 (52.2)	217 (41.0)	
First AIDS diagnosis outcome among those with AIDS diagnosis prior to reference date, n (%)			0.30
Candida esophagitis	11 (11.6)	18 (8.3)	
Cytomegalovirus infection (CMV)	7 (7.4)	8 (3.7)	
Cytomegalovirus infection retinitis	4 (4.2)	6 (2.8)	
Dementia	2 (2.1)	12 (5.5)	
Kaposi's sarcoma (KS)	20 (21.1)	43 (19.8)	
Mycobacterium Avium Intracellular (MAI)	4 (4.2)	6 (2.8)	
Pneumocystis pneumonia (PCP)	34 (35.8)	84 (38.7)	
Wasting Syndrome	3 (3.2)	19 (8.8)	
Others	10 (10.5)	21 (9.7)	
Total	95 (100.0)	217 (100.0)	
HIV Viral Load at closest point 1 year prior to reference date (vc/ml), n (%)			**0.003**
Less than 3 (vc/ml)	3 (2.2)	28 (5.3)	
3-5 (vc/ml)	63 (34.6)	260 (49.1)	
More than 5 (vc/ml)	83 (46.2)	171 (32.3)	
Missing, n (%)	31 (17.0)	70 (13.2)	
Mean HIV Viral Load at closest point 1 year prior to reference date (vc/mm; continuous; range; SD)	4.99 (1.60-6.64; 0.76)	4.64 (1.60-6.68;0.89)	0.001
HIV RNA Levels at Set Point Pre-HAART (vc/ml), n (%)			*0.10*
Less than 3 (vc/ml)	2 (1.1)	21 (4.0)	
3-5 (vc/ml)	108 (59.3)	362 (68.4)	
More than 5 (vc/ml)	26 (14.31)	62 (11.7)	
Missing, n (%)	44 (25.3)	84 (15.9)	
Mean HIV RNA Levels at Set Point (Before HAART; vc/mm, range; SD)	4.51 (2.60-5.98; 0.58)	4.29 (2.48-5.83; 0.66)	**0.004**
Tobacco use one year prior to reference date, n (%)			0.54
Never	63 (35.2)	175 (33.1)	
Former	71 (39.6)	192 (36.3)	
Current	46 (25.3)	161 (30.4)	
Missing, n (%)	0 (0.0)	1 (0.2)	
During follow-up, cumulative pack-years among ever smokers up to one year prior to reference date, mean (range; SD)	21.4 (0.12-85.5; 18.8)	22.5 (0.03-83.3; 17.8)	0.61
Missing, n (%)	17 (14.4)	46 (13.0)	

***Cases and controls were matched on these factors**

Table 3.2 Cont. Demographic Characteristics of NHL Cases and Controls in the MACS

	All AIDS-NHL Cases	HIV^+ Controls	
Alcohol consumption one year prior to reference date, n (%)			0.70
Never	5 (2.8)	25 (4.7)	
Former	28 (15.4)	84 (15.9)	
Current	143 (79.7)	410 (77.5)	
Missing, n (%)	4 (2.2)	10 (1.9)	
Intensity of alcohol consumption at visit one year prior to reference date, n (%)			0.30
Non-drinker since last visit	32 (18.1)	108 (20.4)	
Light (1-3 drinks per week)	89 (49.5)	217 (41.0)	
Moderate (4-13 drinks per week)	46 (25.3)	152 (28.7)	
Heavy (13 or more drinks per week)	9 (5.0)	42 (7.9)	
Missing, n (%)	4 (2.2)	10 (1.9)	
Recreational cannabis use one year prior reference date, n (%)			0.49
Never	6 (3.3)	11 (2.1)	
Former	95 (52.8)	256 (48.4)	
Current	75 (41.8)	252 (47.6)	
Missing, n (%)	4 (2.2)	10 (1.9)	
Recreational amphetamine use (RAU) one year prior to reference date, n (%)			0.36
Never	9 (5.2)	34 (6.7)	
Former	152 (88.4)	429 (84.0)	
Current	11 (6.4)	48 (9.4)	
Missing, n (%)	8 (4.4)	18 (3.4)	

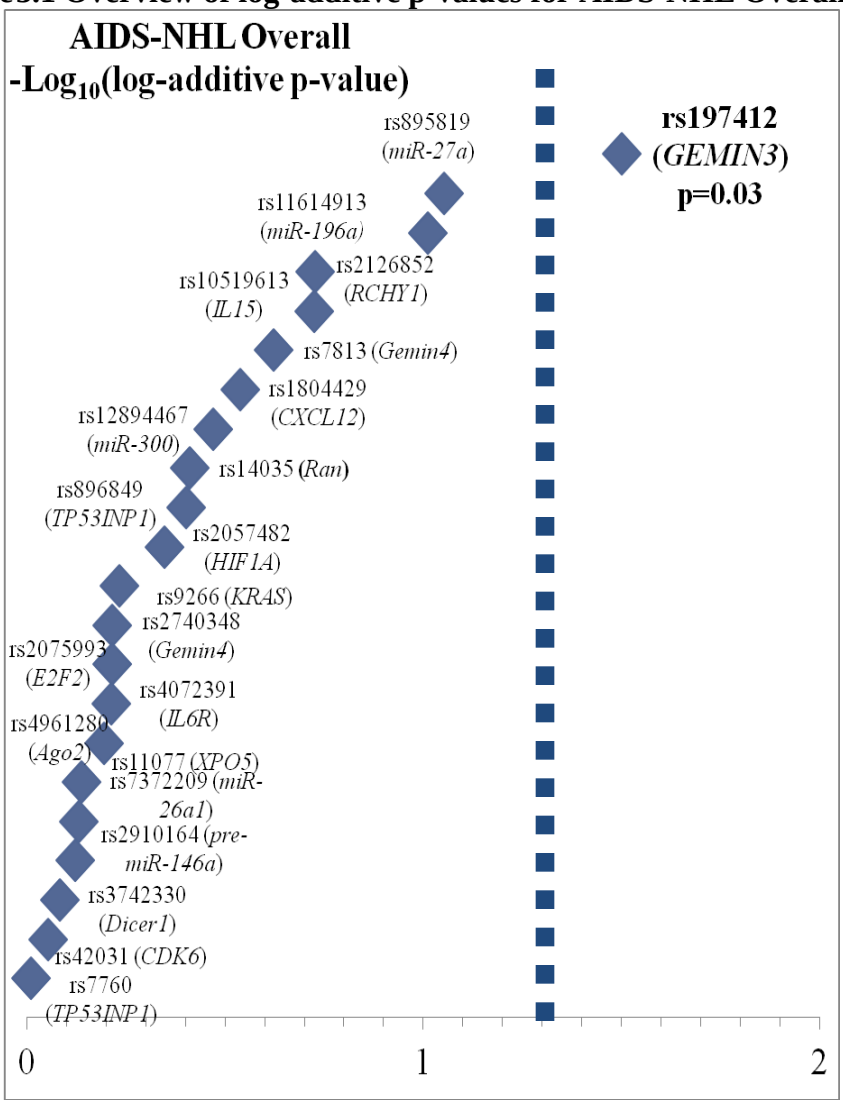
Cases and controls were matched on these factors

Table 3.3. Comparison of Demographic Characteristics of Cohort 1 versus Cohort 2 to Inform Result Interpretation

	Cohort 1 AIDS-NHL Cases	Cohort 1 HIV+ Controls	Cohort 2 AIDS-NHL Cases	Cohort 2 HIV+ Controls	*p-value Cohort 1 v. Cohort 2*
Total, N	161	476	19	53	
HIV status at cohort entry, n (%)*					**0.004**
Prevalent	144 (89.4)	425 (89.3)	19 (100.0)	53 (100.0)	
Converter	17 (10.6)	51 (10.7)	0 (0.0)	0 (0.0)	
*During follow-up, years of known HIV infection prior to reference date, mean (SD)**	7.5 (3.7)	7.5 (3.6)	5.1 (3.3)	5.2 (3.3)	**<0.0001**
Age at reference date in years, n (%)					**<0.0001**
Less than 25 years of age	1 (0.62)	4 (0.84)	0 (0.0)	1 (1.9)	
25-29	12 (7.5)	20 (4.2)	0 (0.0)	8 (15.1)	
30-34	13 (8.1)	79 (16.6)	8 (42.1)	13 (24.5)	
35-39	40 (24.8)	127 (26.7)	4 (21.1)	12 (22.6)	
40-44	39 (24.2)	132 (27.7)	4 (21.1)	7 (13.2)	
45-49	26 (16.2)	69 (14.5)	1 (5.3)	3 (5.7)	
50-54	19 (11.8)	26 (5.5)	2 (10.5)	4 (7.6)	
55-59	8 (5.0)	16 (3.4)	0 (0.0)	1 (1.8)	
60-64	3 (1.9)	0 (0.0)	0 (0.0)	3 (5.7)	
65+	0 (0.0)	3 (0.63)	0 (0.0)	1 (1.9)	
AIDS diagnosis prior to reference date, n					0.72
No	77 (47.8)	282 (59.2)	9 (47.4)	30 (56.6)	
Yes	84 (52.2)	194 (40.8)	10 (52.6)	23 (43.4)	
Ethnicity, n (%)					**<0.0001**
White, non-Hispanic	142 (88.2)	446 (93.7)	10 (52.6)	33 (62.3)	
Black, non-Hispanic	3 (1.9)	6 (1.3)	7 (36.8)	15 (28.3)	
White, Hispanic	16 (9.9)	24 (5.0)	2 (10.5)	1 (1.9)	
Other	0 (0.0)	0 (0.0)	0 (0.0)	4 (7.6)	
HAART prior to reference date, n (%)					0.44
Never	152 (94.4)	433 (91.0)	18 (94.7)	50 (94.3)	
Yes	9 (5.6)	43 (9.0)	1 (5.3)	3 (5.7)	
ART prior to reference date, n (%)					**0.02**
Never	65 (40.4)	144 (30.3)	4 (21.1)	10 (18.9)	
Yes	96 (59.6)	332 (69.8)	15 (79.0)	43 (81.1)	
CD4+ T-cell count at date of matching, n (cells/ml; %)*					
0-49	65 (25.4)	191 (40.1)	8 (42.1)	21 (39.6)	**0.006**
50-99	23 (14.3)	69 (14.5)	2 (10.5)	5 (9.4)	
100-199	23 (14.3)	66 (13.9)	5 (26.3)	15 (28.3)	
200-299	12 (7.5)	35 (7.4)	2 (10.5)	7 (13.2)	
300-399	13 (8.1)	33 (6.9)	1 (5.3)	2 (3.8)	
400-499	7 (4.4)	28 (5.9)	0 (0.0)	0 (0.0)	
500+	18 (11.2)	54 (11.3)	1 (5.3)	3 (4.2)	
HCV infection prior to reference date, n (%)					0.62
Never	139 (86.3)	433 (91.0)	17 (89.5)	49 (92.5)	
Ever	22 (13.7)	43 (9.0)	2 (10.5)	4 (7.6)	

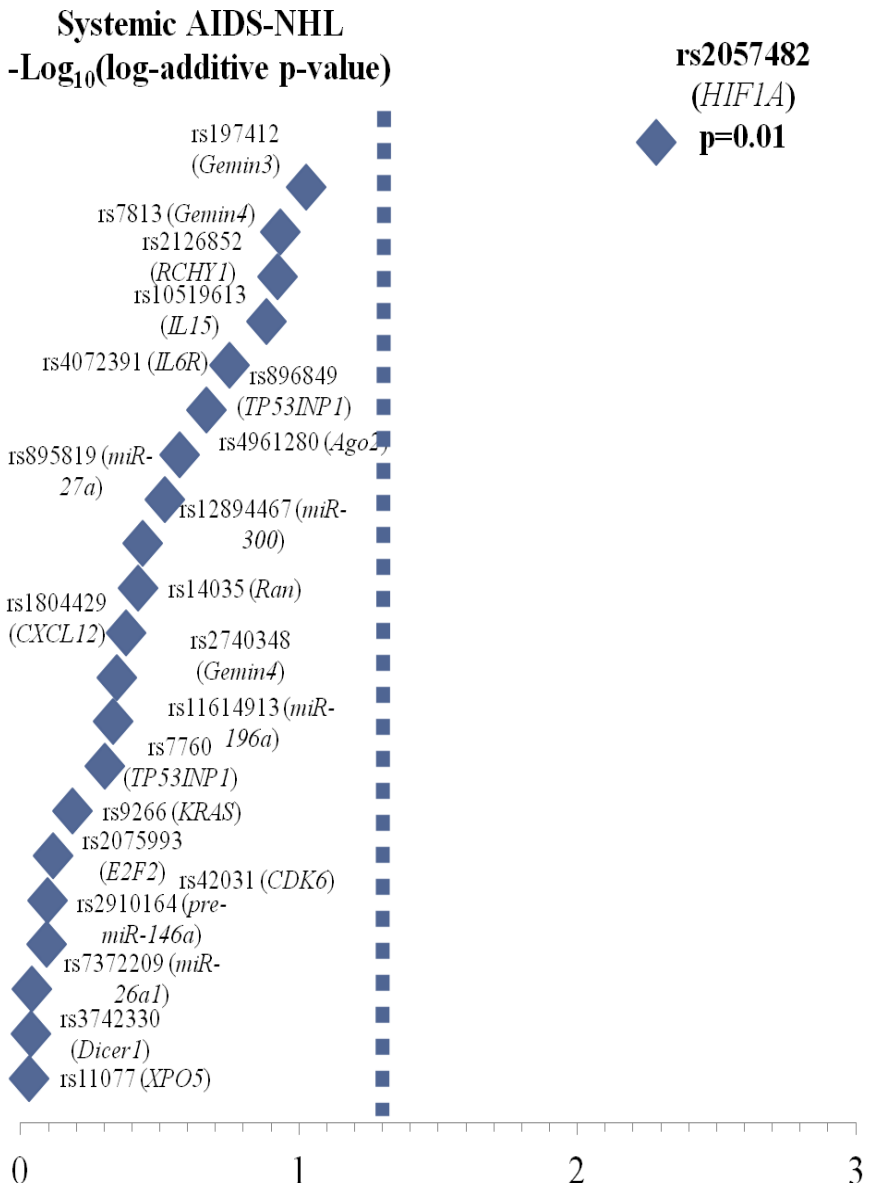
Cases and controls were matched on these factors

Figure 3.1 Overview of log-additive p-values for AIDS-NHL Overall



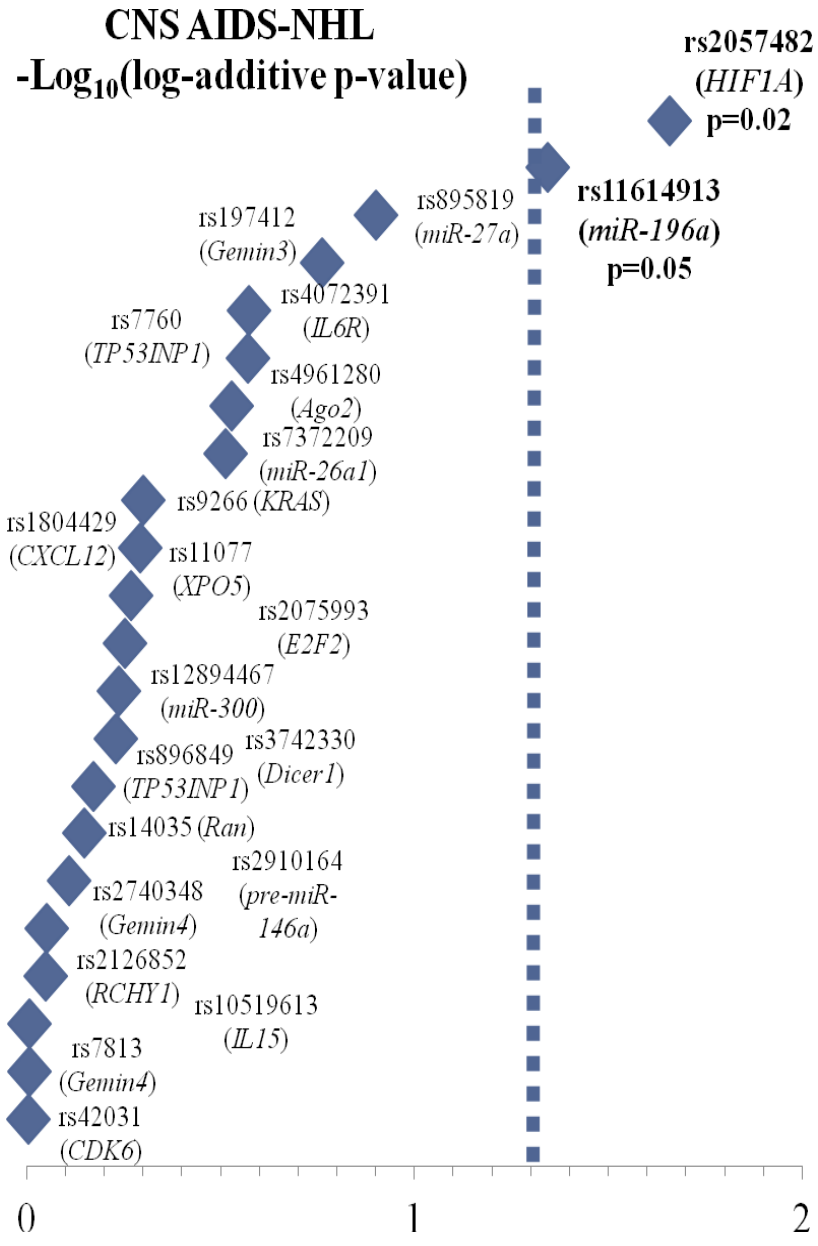
p-value=0.05 threshold $-\log_{10}(0.05)=1.30$

Figure 3.2 Overview of log-additive p-values for Systemic AIDS-NHL



p-value=0.05 threshold $-\log_{10}(0.05)=1.30$

Figure 3.3 Overview of log-additive p-values for CNS AIDS-NHL



p-value=0.05 threshold $-\log_{10}(0.05)=1.30$

Table 3.4. Association of miRNA-related SNPs and Overall, Systemic, or CNS AIDS-NHL

	Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL		Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)
microRNA coding regions	**Complete Case Analysis**						**Analysis imputed for HIV Viral Load and CD4+ T-cell count at reference date**					
miR-196a rs11614913												
CC	43/164	1.00 (Ref)	23/120	1.00 (Ref)	20/44	1.00 (Ref)	72/196	1.00 (Ref)	41/138	1.00 (Ref)	31/58	1.00 (Ref)
CT	63/212	1.00 (0.62-1.59)	46/139	1.33 (0.72-2.43)	17/73	0.51 (0.20-1.29)	88/257	0.88 (0.60-1.29)	67/169	1.18 (0.73-1.90)	21/88	**0.41 (0.19-0.88)**^
TT	11/59	0.59 (0.26-1.33)	9/46	0.87 (0.34-2.21)	2/13	0.17 (0.02-1.25)	19/76	0.57 (0.31-1.06)	14/58	0.62 (0.29-1.33)	5/18	0.53 (0.16-1.78)
Log-Add	117/435	0.85 (0.60-1.19)	78/305	1.02 (0.67-1.56)	39/130	**0.46 (0.23-0.93)**	179/529	0.80 (0.61-1.05)	122/365	0.89 (0.64-1.24)	57/164	**0.57 (0.33-0.99)**
Dominant[1]	74/271	0.91 (0.58-1.44)	55/185	1.23 (0.68-2.22)	19/86	0.42 (0.17-1.01)	107/333	0.81 (0.56-1.17)	81/227	1.04 (0.66-1.65)	26/106	**0.43 (0.22-0.87)**^
Recessive[2]	11/59	0.59 (0.28-1.27)	9/46	0.72 (0.31-1.68)	2/13	0.20 (0.03-1.56)	19/76	0.61 (0.34-1.10)	14/58	0.57 (0.28-1.14)	5/18	0.73 (0.22-2.42)
miR-26a1 rs7372209												
CC	59/226	1.00 (Ref)	36/152	1.00 (Ref)	23/74	1.00 (Ref)	94/276	1.00 (Ref)	61/180	1.00 (Ref)	33/96	1.00 (Ref)
CT	50/180	1.43 (0.88-2.33)	37/131	1.47 (0.83-2.62)	13/49	1.73 (0.65-4.62)	77/215	1.26 (0.86-1.85)	56/155	1.29 (0.82-2.04)	21/60	1.51 (0.71-3.18)
TT	8/29	1.82 (0.69-4.84)	5/22	1.93 (0.59-6.34)	3/7	2.01 (0.26-15.71)	9/38	0.80 (0.36-1.78)	6/30	0.71 (0.27-1.83)	3/8	1.53 (0.28-8.40)
Log-Add	117/435	1.39 (0.94-2.06)	78/305	1.43 (0.89-2.29)	39/130	1.58 (0.69-3.59)	180/529	1.06 (0.79-1.43)	123/365	1.03 (0.73-1.46)	57/164	1.39 (0.74-2.60)
Dominant	58/209	1.46 (0.91-2.36)	42/153	1.51 (0.85-2.65)	16/56	1.75 (0.66-4.61)	86/253	1.19 (0.82-1.73)	62/185	1.19 (0.76-1.84)	24/68	1.51 (0.72-3.15)
Recessive	8/29	1.46 (0.58-3.69)	5/22	1.53 (0.49-4.77)	3/7	1.37 (0.20-9.21)	9/38	0.71 (0.33-1.56)	6/30	0.63 (0.25-1.59)	3/8	1.20 (0.23-6.15)
miR-27a rs895819												
TT	47/206	1.00 (Ref)	35/150	1.00 (Ref)	12/56	1.00 (Ref)	67/242	1.00 (Ref)	50/170	1.00 (Ref)	17/72	1.00 (Ref)
TC	57/191	1.24 (0.75-2.05)	34/127	1.04 (0.57-1.92)	23/64	1.74 (0.65-4.65)	92/242	1.28 (0.85-1.93)	58/162	1.16 (0.71-1.89)	34/80	1.66 (0.74-3.73)
CC	13/36	1.42 (0.60-3.33)	9/27	1.28 (0.43-3.79)	4/9	1.49 (0.35-6.38)	21/43	1.75 (0.90-3.41)	15/32	1.61 (0.71-3.62)	6/11	2.18 (0.64-7.46)
Log-Add	117/433	1.21 (0.84-1.75)	78/304	1.09 (0.68-1.75)	39/129	1.37 (0.73-2.56)	180/527	1.31 (0.97-1.76)	123/364	1.23 (0.85-1.77)	57/163	1.53 (0.89-2.65)
Dominant	70/227	1.27 (0.79-2.06)	43/154	1.07 (0.59-1.94)	27/73	1.67 (0.69-4.04)	113/285	1.35 (0.91-2.00)	73/194	1.21 (0.76-1.95)	40/91	1.77 (0.84-3.75)
Recessive	13/36	1.27 (0.56-2.86)	9/27	1.24 (0.44-3.50)	4/9	1.28 (0.31-5.24)	21/43	1.52 (0.81-2.85)	15/32	1.46 (0.69-3.11)	6/11	1.77 (0.55-5.66)
miR-300 rs12894467												
CC	40/159	1.00 (Ref)	24/100	1.00 (Ref)	16/59	1.00 (Ref)	58/183	1.00 (Ref)	38/114	1.00 (Ref)	20/69	1.00 (Ref)
CT	56/212	1.08 (0.65-1.81)	38/162	0.89 (0.47-1.68)	18/50	1.34 (0.50-3.59)	85/265	1.01 (0.66-1.54)	57/198	0.83 (0.49-1.40)	28/67	1.43 (0.64-3.18)
TT	21/63	1.52 (0.73-3.16)	16/42	1.87 (0.77-4.55)	5/21	1.00 (0.23-4.35)	37/80	1.36 (0.79-2.37)	28/52	1.49 (0.77-2.87)	9/28	1.21 (0.40-3.64)
Log-Add	117/434	1.20 (0.84-1.70)	78/304	1.25 (0.81-1.94)	39/130	1.08 (0.55-2.14)	180/528	1.14 (0.87-1.51)	123/364	1.17 (0.84-1.63)	57/164	1.16 (0.69-1.95)
Dominant	77/275	1.16 (0.71-1.90)	54/204	1.04 (0.57-1.92)	23/71	1.26 (0.49-3.24)	122/345	1.09 (0.73-1.63)	85/250	0.97 (0.59-1.58)	37/95	1.37 (0.64-2.94)
Recessive	21/63	1.44 (0.75-2.79)	16/42	2.02 (0.91-4.47)	5/21	0.84 (0.22-3.28)	37/80	1.36 (0.84-2.20)	28/52	1.67 (0.95-2.96)	9/28	0.98 (0.36-2.66)
pre-miR-146a rs2910164												
GG	71/260	1.00 (Ref)	44/174	1.00 (Ref)	27/86	1.00 (Ref)	110/314	1.00 (Ref)	72/204	1.00 (Ref)	38/110	1.00 (Ref)
GC	43/156	1.25 (0.76-2.05)	32/116	1.44 (0.79-2.62)	11/40	1.06 (0.41-2.76)	62/193	0.98 (0.66-1.44)	45/145	0.98 (0.61-1.57)	17/48	1.06 (0.50-2.21)
CC	3/19	0.40 (0.10-1.64)	2/15	0.45 (0.09-2.20)	1/4	0.38 (0.01-25.09)	8/22	0.90 (0.36-2.22)	6/16	0.93 (0.33-2.62)	2/6	1.43 (0.19-10.83)
Log-Add	117/435	0.97 (0.65-1.45)	78/305	1.05 (0.65-1.67)	39/130	0.97 (0.40-2.33)	180/529	0.96 (0.70-1.32)	123/365	0.97 (0.67-1.41)	57/164	1.10 (0.58-2.09)
Dominant	46/175	1.12 (0.69-1.81)	34/131	1.26 (0.71-2.24)	12/44	1.02 (0.40-2.64)	70/215	0.97 (0.66-1.41)	51/161	0.97 (0.62-1.53)	19/54	1.08 (0.52-2.22)
Recessive	3/19	0.36 (0.09-1.49)	2/15	0.39 (0.08-1.89)	1/4	0.37 (0.01-24.16)	8/22	0.91 (0.37-2.21)	6/16	0.94 (0.34-2.60)	2/6	1.40 (0.19-10.41)

Table 3.4 Continued. Association of miRNA-related SNPs and Overall, Systemic, or CNS AIDS-NHL

	Complete Case Analysis						Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference					
	Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL		Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL	
	Overall	Adjusted*	Systemic	Adjusted*	CNS	Adjusted*	Overall	Adjusted*	Systemic	Adjusted*	CNS	Adjusted*
Functional SNPs in miRNA processing and maturation genes												
AGO2 rs4961280												
CC	69/291	1.00 (Ref)	42/204	1.00 (Ref)	27/87	1.00 (Ref)	117/348	1.00 (Ref)	75/241	1.00 (Ref)	42/107	1.00 (Ref)
CA	42/132	1.37 (0.82-2.27)	32/93	1.50 (0.82-2.71)	10/39	1.01 (0.35-2.89)	56/167	1.01 (0.67-1.53)	43/114	1.28 (0.77-2.11)	13/53	0.55 (0.25-1.23)
AA	6/12	2.24 (0.71-7.09)	4/8	2.44 (0.58-10.28)	2/4	1.44 (0.18-11.62)	7/14	1.48 (0.54-4.04)	5/10	1.50 (0.46-4.93)	2/4	1.27 (0.16-10.06)
Log-Add	117/435	1.42 (0.94-2.14)	78/305	1.52 (0.93-2.49)	39/130	1.10 (0.50-2.43)	180/529	1.08 (0.77-1.53)	123/365	1.26 (0.83-1.90)	57/164	0.70 (0.36-1.37)
Dominant	48/144	1.45 (0.89-2.36)	36/101	1.57 (0.88-2.80)	12/43	1.07 (0.40-2.86)	63/181	1.05 (0.70-1.56)	48/124	1.30 (0.80-2.12)	15/57	0.60 (0.28-1.29)
Recessive	6/12	2.02 (0.65-6.29)	4/8	2.12 (0.51-8.78)	2/4	1.44 (0.18-11.43)	7/14	1.47 (0.55-3.97)	5/10	1.36 (0.42-4.39)	2/4	1.43 (0.19-10.92)
DICER1 rs3742330												
AA	96/353	1.00 (Ref)	62/246	1.00 (Ref)	34/107	1.00 (Ref)	150/430	1.00 (Ref)	102/297	1.00 (Ref)	48/133	1.00 (Ref)
AG	20/80	0.84 (0.45-1.54)	15/58	0.95 (0.46-1.95)	5/22	0.62 (0.18-2.09)	29/96	0.94 (0.57-1.53)	20/66	0.95 (0.52-1.72)	9/30	0.80 (0.32-1.96)
GG	1/2	1.54 (0.12-20.16)	1/1	1.89 (0.11-32.93)	0/1	0.00 (0.00-.)	1/3	1.04 (0.10-10.47)	1/2	1.17 (0.10-13.21)	0/1	0.00 (0.00-.)
Log-Add	117/435	0.89 (0.51-1.56)	78/305	1.02 (0.53-1.95)	39/130	0.61 (0.18-2.03)	180/529	0.95 (0.60-1.50)	123/365	0.97 (0.56-1.67)	57/164	0.78 (0.32-1.90)
Dominant	21/82	0.86 (0.48-1.56)	16/59	0.98 (0.49-1.99)	5/23	0.61 (0.18-2.05)	30/99	0.94 (0.58-1.53)	21/68	0.96 (0.53-1.72)	9/31	0.79 (0.32-1.93)
Recessive	1/2	1.53 (0.12-19.90)	1/1	1.90 (0.11-32.95)	0/1	0.00 (0.00-.)	1/3	1.04 (0.10-10.48)	1/2	1.17 (0.10-13.25)	0/1	0.00 (0.00-.)
GEMIN3 rs197412												
TT	38/163	1.00 (Ref)	28/120	1.00 (Ref)	10/43	1.00 (Ref)	54/189	1.00 (Ref)	40/140	1.00 (Ref)	14/49	1.00 (Ref)
TC	56/212	1.24 (0.73-2.08)	37/145	1.24 (0.67-2.31)	19/67	1.12 (0.39-3.23)	87/254	1.34 (0.88-2.03)	60/169	1.49 (0.91-2.45)	27/85	1.04 (0.45-2.42)
CC	23/60	1.63 (0.81-3.30)	13/40	1.42 (0.59-3.44)	10/20	2.36 (0.60-9.21)	39/86	**1.85 (1.05-3.23)**	23/56	1.66 (0.84-3.32)	16/30	2.24 (0.77-6.46)
Log-Add	117/435	1.27 (0.90-1.79)	78/305	1.20 (0.79-1.84)	39/130	1.50 (0.76-3.00)	180/529	**1.35 (1.03-1.78)**^	123/365	1.33 (0.96-1.84)	57/164	1.46 (0.85-2.51)
Dominant	79/272	1.31 (0.80-2.16)	50/185	1.28 (0.70-2.32)	29/87	1.32 (0.47-3.68)	126/340	1.44 (0.97-2.15)	83/225	1.53 (0.95-2.45)	43/115	1.26 (0.56-2.81)
Recessive	23/60	1.42 (0.77-2.64)	13/40	1.24 (0.57-2.72)	10/20	2.18 (0.70-6.73)	39/86	1.54 (0.94-2.53)	23/56	1.32 (0.71-2.46)	16/30	2.18 (0.89-5.35)
GEMIN4 rs2740348												
GG	80/308	1.00 (Ref)	48/212	1.00 (Ref)	32/96	1.00 (Ref)	123/376	1.00 (Ref)	81/254	1.00 (Ref)	42/122	1.00 (Ref)
GC	32/111	0.98 (0.58-1.66)	26/85	1.21 (0.66-2.23)	6/26	0.56 (0.19-1.67)	49/134	1.12 (0.73-1.71)	36/100	1.07 (0.64-1.77)	13/34	1.15 (0.50-2.62)
CC	5/16	0.90 (0.25-3.16)	4/8	1.77 (0.42-7.55)	1/8	0.10 (0.00-6.41)	8/18	1.04 (0.40-2.70)	6/10	1.56 (0.49-4.98)	2/8	0.54 (0.07-4.02)
Log-Add	117/435	0.97 (0.63-1.49)	78/305	1.26 (0.76-2.08)	39/130	0.50 (0.18-1.35)	180/528	1.07 (0.77-1.50)	123/364	1.14 (0.76-1.70)	57/164	0.95 (0.49-1.87)
Dominant	37/127	0.97 (0.59-1.62)	30/93	1.26 (0.70-2.27)	7/34	0.52 (0.18-1.51)	57/152	1.11 (0.74-1.66)	42/110	1.12 (0.69-1.81)	15/42	1.06 (0.47-2.35)
Recessive	5/16	0.90 (0.26-3.13)	4/8	1.65 (0.39-6.88)	1/8	0.14 (0.00-6.89)	8/18	1.01 (0.39-2.60)	6/10	1.54 (0.48-4.87)	2/8	0.51 (0.07-3.71)
GEMIN4 rs7813												
TT	46/158	1.00 (Ref)	26/106	1.00 (Ref)	20/52	1.00 (Ref)	67/200	1.00 (Ref)	41/133	1.00 (Ref)	26/67	1.00 (Ref)
TC	48/203	0.98 (0.59-1.61)	34/149	1.22 (0.64-2.34)	14/54	0.64 (0.27-1.50)	78/242	1.07 (0.71-1.60)	56/173	1.17 (0.71-1.94)	22/69	0.87 (0.43-1.77)
CC	23/73	1.20 (0.62-2.32)	18/49	1.76 (0.79-3.95)	5/24	0.65 (0.15-2.69)	35/85	1.38 (0.81-2.35)	26/57	1.67 (0.87-3.21)	9/28	1.12 (0.41-3.03)
Log-Add	117/434	1.07 (0.78-1.48)	78/304	1.32 (0.88-1.97)	39/130	0.74 (0.39-1.38)	180/527	1.16 (0.89-1.50)	123/363	1.27 (0.92-1.76)	57/164	1.00 (0.63-1.60)
Dominant	71/276	1.03 (0.65-1.65)	52/198	1.36 (0.74-2.49)	19/78	0.64 (0.28-1.45)	113/327	1.14 (0.78-1.67)	82/230	1.28 (0.80-2.07)	31/97	0.93 (0.48-1.79)
Recessive	23/73	1.22 (0.67-2.21)	18/49	1.57 (0.78-3.15)	5/24	0.81 (0.21-3.20)	35/85	1.33 (0.83-2.15)	26/57	1.51 (0.85-2.69)	9/28	1.19 (0.46-3.08)

Table 3.4 Continued. Association of miRNA-related SNPs and Overall, Systemic, or CNS AIDS-NHL

	Complete Case Analysis						Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference					
	Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL		Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)
Functional SNPs in miRNA processing and maturation genes												
RAN rs14035												
CC	49/219	1.00 (Ref)	31/158	1.00 (Ref)	18/61	1.00 (Ref)	85/257	1.00 (Ref)	60/182	1.00 (Ref)	25/75	1.00 (Ref)
CT	57/175	1.35 (0.83-2.21)	39/123	1.55 (0.86-2.77)	18/52	1.01 (0.39-2.56)	73/222	1.07 (0.72-1.59)	50/154	1.15 (0.72-1.83)	23/68	0.86 (0.40-1.86)
TT	11/41	1.06 (0.48-2.37)	8/24	1.26 (0.47-3.35)	3/17	0.95 (0.22-4.08)	22/50	1.31 (0.72-2.38)	13/29	1.34 (0.62-2.87)	9/21	1.38 (0.50-3.80)
Log-Add	117/435	1.13 (0.81-1.59)	78/305	1.26 (0.83-1.91)	39/130	0.98 (0.53-1.84)	180/529	1.12 (0.85-1.47)	123/365	1.15 (0.82-1.61)	57/164	1.10 (0.67-1.78)
Dominant	68/216	1.28 (0.81-2.04)	47/147	1.49 (0.85-2.61)	21/69	0.99 (0.42-2.31)	95/272	1.12 (0.77-1.62)	63/183	1.18 (0.76-1.84)	32/89	0.99 (0.49-2.00)
Recessive	11/41	0.93 (0.43-2.01)	8/24	1.01 (0.40-2.54)	3/17	0.95 (0.23-3.97)	22/50	1.27 (0.72-2.23)	13/29	1.25 (0.61-2.59)	9/21	1.47 (0.56-3.84)
XPO5 rs11077												
AA	42/142	1.00 (Ref)	28/94	1.00 (Ref)	14/48	1.00 (Ref)	62/175	1.00 (Ref)	44/115	1.00 (Ref)	18/60	1.00 (Ref)
AC	53/216	0.76 (0.44-1.32)	37/158	0.70 (0.35-1.40)	16/58	0.80 (0.29-2.24)	82/265	0.98 (0.64-1.51)	55/189	0.84 (0.50-1.41)	27/76	1.33 (0.58-3.05)
CC	20/77	0.83 (0.41-1.69)	12/53	0.70 (0.28-1.76)	8/24	1.10 (0.30-3.95)	34/89	1.11 (0.64-1.93)	23/61	1.01 (0.52-1.97)	11/28	1.35 (0.48-3.76)
Log-Add	115/435	0.89 (0.62-1.28)	77/305	0.82 (0.52-1.31)	38/130	1.02 (0.54-1.94)	178/529	1.05 (0.80-1.38)	122/365	0.98 (0.70-1.38)	56/164	1.17 (0.71-1.94)
Dominant	73/293	0.78 (0.46-1.32)	49/211	0.70 (0.35-1.38)	24/82	0.87 (0.33-2.30)	116/354	1.01 (0.67-1.53)	78/250	0.88 (0.53-1.44)	38/104	1.34 (0.61-2.95)
Recessive	20/77	1.00 (0.55-1.83)	12/53	0.92 (0.43-1.95)	8/24	1.27 (0.43-3.78)	34/89	1.13 (0.70-1.80)	23/61	1.13 (0.64-2.01)	11/28	1.11 (0.47-2.62)
Candidate genes with SNPs near or within a predicted miRNA binding site												
CDK6 rs42031												
AA	67/277	1.00 (Ref)	45/198	1.00 (Ref)	22/79	1.00 (Ref)	116/335	1.00 (Ref)	80/237	1.00 (Ref)	36/98	1.00 (Ref)
AT	44/140	1.33 (0.81-2.18)	30/95	1.43 (0.78-2.61)	14/45	1.15 (0.48-2.79)	55/174	0.96 (0.64-1.44)	37/116	1.02 (0.62-1.69)	18/58	0.85 (0.41-1.75)
TT	6/17	1.21 (0.41-3.58)	3/11	0.89 (0.20-3.88)	3/6	1.90 (0.34-10.67)	8/19	1.31 (0.52-3.26)	5/11	1.27 (0.40-4.00)	3/8	1.52 (0.31-7.46)
Log-Add	117/434	1.22 (0.82-1.81)	78/304	1.20 (0.73-1.98)	39/130	1.26 (0.65-2.45)	179/528	1.03 (0.74-1.43)	122/364	1.06 (0.71-1.60)	57/164	1.00 (0.56-1.77)
Dominant	50/157	1.31 (0.82-2.11)	33/106	1.36 (0.76-2.45)	17/51	1.25 (0.55-2.87)	63/193	0.99 (0.67-1.47)	42/127	1.05 (0.65-1.69)	21/66	0.92 (0.47-1.81)
Recessive	6/17	1.09 (0.37-3.16)	3/11	0.76 (0.18-3.24)	3/6	1.83 (0.33-10.08)	8/19	1.33 (0.54-3.27)	5/11	1.26 (0.40-3.93)	3/8	1.59 (0.33-7.70)
CXCL12 rs1804429												
TT	109/407	1.00 (Ref)	74/287	1.00 (Ref)	35/120	1.00 (Ref)	162/498	1.00 (Ref)	112/345	1.00 (Ref)	50/153	1.00 (Ref)
TG	8/28	0.81 (0.32-2.04)	4/18	0.58 (0.17-1.98)	4/10	2.01 (0.38-10.58)	18/30	1.71 (0.87-3.35)	11/20	1.47 (0.65-3.34)	7/10	2.50 (0.68-9.11)
GG	0/0	NAC	0/0	NAC	0/0	NAC	0/1	0.00 (0.00-.)	0/0	1.00 (Ref)	0/1	0.00 (0.00-.)
Log-Add	117/435	0.81 (0.32-2.04)	78/305	0.58 (0.17-1.98)	39/130	2.01 (0.38-10.58)	180/529	1.47 (0.77-2.80)	123/365	1.47 (0.65-3.34)	57/164	1.45 (0.48-4.39)
Dominant	8/28	0.81 (0.32-2.04)	4/18	0.58 (0.17-1.98)	4/10	2.01 (0.38-10.58)	18/31	1.60 (0.82-3.12)	11/20	1.47 (0.65-3.34)	7/11	1.93 (0.56-6.67)
Recessive	0/0	NAC	0/0	NAC	0/0	NAC	0/1	0.00 (0.00-.)	0/0	1.00 (Ref)	0/1	0.00 (0.00-.)
E2F2 rs2075993												
AA	27/116	1.00 (Ref)	18/85	1.00 (Ref)	9/31	1.00 (Ref)	43/142	1.00 (Ref)	30/102	1.00 (Ref)	13/40	1.00 (Ref)
AG	60/206	1.70 (0.93-3.10)	41/138	1.75 (0.86-3.56)	19/68	1.73 (0.52-5.81)	92/252	1.39 (0.86-2.27)	63/165	1.48 (0.84-2.63)	29/87	1.38 (0.53-3.63)
GG	30/112	1.57 (0.78-3.15)	19/81	1.47 (0.64-3.39)	11/31	1.74 (0.45-6.74)	45/134	1.17 (0.66-2.07)	30/97	1.11 (0.57-2.15)	15/37	1.45 (0.45-4.68)
Log-Add	117/434	1.23 (0.88-1.73)	78/304	1.20 (0.80-1.81)	39/130	1.28 (0.67-2.46)	180/528	1.07 (0.81-1.41)	123/364	1.05 (0.76-1.44)	57/164	1.19 (0.67-2.11)
Dominant	90/318	1.66 (0.93-2.95)	60/219	1.66 (0.84-3.27)	30/99	1.74 (0.54-5.55)	137/386	1.32 (0.83-2.11)	93/262	1.35 (0.78-2.32)	44/124	1.39 (0.54-3.61)
Recessive	30/112	1.07 (0.62-1.82)	19/81	0.98 (0.51-1.91)	11/31	1.15 (0.43-3.08)	45/134	0.92 (0.59-1.43)	30/97	0.85 (0.50-1.44)	15/37	1.11 (0.48-2.57)

Table 3.4 Continued. Association of miRNA-related SNPs and Overall, Systemic, or CNS AIDS-NHL

	Complete Case Analysis						Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference					
	Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL		Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)
Candidate genes with SNPs near or within a predicted miRNA binding site												
HIF1A rs2057482												
CC	78/312	1.00 (Ref)	47/227	1.00 (Ref)	31/85	1.00 (Ref)	125/369	1.00 (Ref)	79/266	1.00 (Ref)	46/103	1.00 (Ref)
CT	35/117	1.13 (0.66-1.95)	29/76	1.89 (1.00-3.59)	6/41	**0.20 (0.05-0.84)**^	49/147	1.10 (0.71-1.71)	40/92	**1.85 (1.10-3.09)**^	9/55	**0.23 (0.08-0.73)**^
TT	4/5	3.72 (0.74-18.58)	2/1	8.18 (0.58-116.1)	2/4	2.19 (0.18-26.08)	6/12	1.29 (0.43-3.84)	4/6	3.23 (0.66-15.86)	2/6	0.33 (0.05-2.03)
Log-Add	117/434	1.33 (0.83-2.11)	78/304	**2.01 (1.09-3.68)**^	39/130	0.55 (0.22-1.38)	180/528	1.12 (0.78-1.60)	123/364	**1.83 (1.16-2.90)**^	57/164	**0.38 (0.17-0.87)**^
Dominant	39/122	1.25 (0.74-2.11)	31/77	**1.94 (1.03-3.66)**	8/45	0.34 (0.10-1.10)	55/159	1.12 (0.74-1.71)	44/98	**1.91 (1.15-3.16)**^	11/61	**0.25 (0.09-0.72)**^
Recessive	4/5	3.60 (0.73-17.82)	2/1	5.26 (0.41-67.83)	2/4	2.13 (0.20-23.28)	6/12	1.25 (0.42-3.69)	4/6	2.59 (0.54-12.37)	2/6	0.52 (0.09-2.89)
IL15 rs10519613												
CC	92/356	1.00 (Ref)	60/252	1.00 (Ref)	32/104	1.00 (Ref)	135/433	1.00 (Ref)	89/303	1.00 (Ref)	46/130	1.00 (Ref)
CA	23/76	1.22 (0.67-2.24)	17/50	1.33 (0.65-2.74)	6/26	1.15 (0.33-3.96)	42/91	1.39 (0.88-2.21)	32/57	1.57 (0.91-2.70)	10/34	0.99 (0.39-2.50)
AA	1/3	1.72 (0.15-19.78)	1/3	1.84 (0.16-21.25)	0/0	1.00 (Ref)	2/5	1.19 (0.22-6.39)	2/5	1.33 (0.25-7.16)		1.00 (Ref)
Log-Add	116/435	1.24 (0.72-2.14)	78/305	1.34 (0.72-2.50)	38/130	1.15 (0.33-3.96)	179/529	1.32 (0.87-1.98)	123/365	1.43 (0.90-2.28)	56/164	0.99 (0.39-2.50)
Dominant	24/79	1.25 (0.69-2.25)	18/53	1.37 (0.68-2.74)	6/26	1.15 (0.33-3.96)	44/96	1.38 (0.88-2.16)	34/62	1.55 (0.91-2.62)	10/34	0.99 (0.39-2.50)
Recessive	1/3	1.71 (0.15-19.54)	1/3	1.83 (0.16-21.05)	0/0	1.00 (Ref)	2/5	1.12 (0.21-5.97)	2/5	1.21 (0.23-6.44)		1.00 (Ref)
IL6R rs4072391												
CC	76/284	1.00 (Ref)	53/201	1.00 (Ref)	23/83	1.00 (Ref)	117/337	1.00 (Ref)	86/236	1.00 (Ref)	31/101	1.00 (Ref)
CT	35/133	0.99 (0.60-1.63)	24/91	1.10 (0.59-2.06)	11/42	0.87 (0.32-2.34)	57/169	0.99 (0.67-1.45)	36/114	0.91 (0.56-1.47)	21/55	1.32 (0.65-2.72)
TT	6/17	1.16 (0.40-3.42)	1/12	0.34 (0.04-2.87)	5/5	3.13 (0.54-18.26)	6/22	0.70 (0.26-1.87)	1/14	0.17 (0.02-1.39)	5/8	1.97 (0.47-8.14)
Log-Add	117/434	1.03 (0.69-1.53)	78/304	0.91 (0.53-1.54)	39/130	1.29 (0.65-2.56)	180/528	0.93 (0.67-1.28)	123/364	0.76 (0.49-1.16)	57/164	1.36 (0.79-2.36)
Dominant	41/150	1.01 (0.62-1.63)	25/103	1.00 (0.55-1.84)	16/47	1.13 (0.47-2.70)	63/191	0.95 (0.66-1.39)	37/128	0.82 (0.51-1.32)	26/63	1.41 (0.71-2.77)
Recessive	6/17	1.17 (0.40-3.39)	1/12	0.33 (0.04-2.75)	5/5	3.21 (0.56-18.50)	6/22	0.70 (0.26-1.86)	1/14	0.18 (0.02-1.43)	5/8	1.80 (0.44-7.28)
KRAS rs9266												
CC	37/124	1.00 (Ref)	26/93	1.00 (Ref)	11/31	1.00 (Ref)	52/151	1.00 (Ref)	37/114	1.00 (Ref)	15/37	1.00 (Ref)
CT	49/215	0.81 (0.46-1.41)	35/150	0.91 (0.47-1.78)	14/65	0.46 (0.15-1.44)	85/261	0.86 (0.54-1.35)	61/177	1.01 (0.58-1.74)	24/84	0.46 (0.18-1.13)
TT	30/96	0.78 (0.40-1.49)	16/62	0.64 (0.28-1.47)	14/34	0.84 (0.26-2.72)	42/117	0.88 (0.52-1.49)	24/74	0.87 (0.45-1.68)	18/43	0.68 (0.26-1.76)
Log-Add	116/435	0.88 (0.63-1.22)	77/305	0.81 (0.54-1.22)	39/130	0.93 (0.51-1.71)	179/529	0.94 (0.72-1.22)	122/365	0.94 (0.68-1.30)	57/164	0.84 (0.51-1.38)
Dominant	79/311	0.80 (0.47-1.34)	51/212	0.82 (0.44-1.53)	28/99	0.59 (0.22-1.63)	127/378	0.86 (0.56-1.32)	85/251	0.96 (0.58-1.60)	42/127	0.54 (0.24-1.22)
Recessive	30/96	0.89 (0.51-1.55)	16/62	0.68 (0.33-1.39)	14/34	1.33 (0.51-3.53)	42/117	0.97 (0.62-1.51)	24/74	0.87 (0.49-1.53)	18/43	1.12 (0.52-2.42)
RCHY1 rs2126852												
AA	70/218	1.00 (Ref)	51/154	1.00 (Ref)	19/64	1.00 (Ref)	102/272	1.00 (Ref)	75/187	1.00 (Ref)	27/85	1.00 (Ref)
AG	41/179	0.66 (0.40-1.08)	23/128	**0.47 (0.25-0.88)**^	18/51	1.01 (0.41-2.48)	67/210	0.89 (0.60-1.31)	39/147	0.70 (0.43-1.13)	28/63	1.40 (0.68-2.87)
GG	6/38	0.49 (0.19-1.32)	4/23	0.64 (0.19-2.15)	2/15	0.29 (0.04-1.98)	11/47	0.55 (0.26-1.19)	9/31	0.61 (0.26-1.45)	2/16	0.37 (0.07-2.10)
Log-Add	117/435	0.68 (0.46-1.00)	78/305	**0.60 (0.37-0.99)**	39/130	0.74 (0.37-1.46)	180/529	0.81 (0.60-1.09)	123/365	0.74 (0.52-1.06)	57/164	0.96 (0.55-1.68)
Dominant	47/217	0.63 (0.40-1.02)	27/151	**0.49 (0.27-0.89)**^	20/66	0.85 (0.37-1.99)	78/257	0.83 (0.57-1.20)	48/178	0.68 (0.43-1.07)	30/79	1.20 (0.61-2.37)
Recessive	6/38	0.60 (0.23-1.54)	4/23	0.82 (0.25-2.68)	2/15	0.29 (0.05-1.83)	11/47	0.58 (0.28-1.22)	9/31	0.69 (0.29-1.60)	2/16	0.31 (0.06-1.73)

Table 3.4 Continued. Association of miRNA-related SNPs and Overall, Systemic, or CNS AIDS-NHL

	Complete Case Analysis						Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference					
	Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL		Overall AIDS-NHL		Systemic AIDS-NHL		CNS AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Systemic Ca/Co	Adjusted* OR_{adj} (95% CI)	CNS Ca/Co	Adjusted* OR_{adj} (95% CI)
Candidate genes with SNPs near or within a predicted miRNA binding site												
TP53INP1 rs7760												
TT	87/341	1.00 (Ref)	57/246	1.00 (Ref)	30/95	1.00 (Ref)	134/407	1.00 (Ref)	91/285	1.00 (Ref)	43/122	1.00 (Ref)
TG	28/87	1.13 (0.67-1.92)	19/56	1.29 (0.69-2.42)	9/31	0.68 (0.23-1.97)	42/111	1.02 (0.66-1.58)	29/73	1.18 (0.70-1.96)	13/38	0.67 (0.28-1.59)
GG	2/7	1.70 (0.28-10.53)	2/3	3.78 (0.45-31.48)	.0/4	0.00 (0.00-.)	4/11	0.98 (0.27-3.65)	3/7	1.33 (0.28-6.36)	1/4	0.36 (0.03-5.32)
Log-Add	117/435	1.17 (0.74-1.87)	78/305	1.42 (0.82-2.48)	39/130	0.63 (0.23-1.71)	180/529	1.01 (0.70-1.46)	123/365	1.17 (0.76-1.81)	57/164	0.65 (0.31-1.39)
Dominant	30/94	1.16 (0.70-1.94)	21/59	1.38 (0.75-2.53)	9/35	0.63 (0.22-1.80)	46/122	1.01 (0.67-1.54)	32/80	1.19 (0.73-1.95)	14/42	0.64 (0.28-1.49)
Recessive	2/7	1.68 (0.27-10.37)	2/3	3.56 (0.43-29.20)	0/4	0.00 (0.00-.)	4/11	0.98 (0.27-3.64)	3/7	1.30 (0.27-6.19)	1/4	0.42 (0.03-5.70)
TP53INP1 rs896849												
TT	77/318	1.00 (Ref)	52/229	1.00 (Ref)	25/89	1.00 (Ref)	121/375	1.00 (Ref)	84/266	1.00 (Ref)	37/109	1.00 (Ref)
TC	35/107	1.30 (0.78-2.15)	22/72	1.32 (0.72-2.44)	13/35	1.02 (0.39-2.67)	50/140	1.04 (0.69-1.58)	32/91	1.06 (0.64-1.75)	18/49	0.92 (0.43-1.98)
CC	5/9	3.47 (0.82-14.78)	4/3	6.19 (0.96-39.98)	1/6	1.47 (0.08-26.07)	9/13	1.98 (0.74-5.29)	7/7	3.67 (1.05-12.84)	2/6	0.62 (0.09-4.29)
Log-Add	117/434	1.45 (0.94-2.23)	78/304	1.59 (0.95-2.66)	39/130	1.07 (0.46-2.50)	180/528	1.17 (0.83-1.65)	123/364	1.32 (0.88-1.99)	57/164	0.87 (0.46-1.66)
Dominant	40/116	1.40 (0.86-2.28)	26/75	1.49 (0.84-2.66)	14/41	1.05 (0.41-2.68)	59/153	1.11 (0.75-1.66)	39/98	1.20 (0.74-1.93)	20/55	0.89 (0.42-1.88)
Recessive	5/9	3.21 (0.77-13.47)	4/3	5.71 (0.91-35.99)	1/6	1.45 (0.08-25.35)	9/13	1.95 (0.74-5.13)	7/7	3.59 (1.04-12.34)	2/6	0.64 (0.10-4.29)

[1]Dominant: Homozygous minor allele + Heterozygous vs. Homozygous major allele.

[2]Recessive: Homozygous minor allele vs. Heterozygous + Homozygous major allele.

*Adjusted for age at case diagnosis or reference date in controls (continuous), HIV RNA levels before set point (continuous), AIDS diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), race (reference=White), and history of HCV infection (reference=No).

Cases and controls were matched on: cohort (1 or 2), date of infection (continuous), duration of follow-up (continuous), seroconversion status (categorical), race (white or non-white), and CD4+ T-cell count at date of matching (categorical).

NAC= Not able to calculate due to sample size.

^Remained evident after SB approach.

Table 3.5. Association of *microRNA-196a* rs11614913 and *HIF1A* rs2057482 on Overall, Systemic and CNS AIDS-NHL, under a dominant model using data imputed for HIV Viral Load and CD4+ T-cell count at reference date

	Overall AIDS-NHL			Systemic AIDS-NHL				CNS AIDS-NHL			
Overall Ca/Co	Joint Effects		OR_{adj} (95% CI)	Systemic Ca/Co	Joint Effects		OR_{adj} (95% CI)	CNS Ca/Co	Joint Effects		OR_{adj} (95% CI)
	rs2057482	rs11614913			rs2057482	rs11614913			rs2057482	rs11614913	
47/143	0	0	1.00 (Ref)	23/107	0	0	1.00 (Ref)	24/36	0	0	1.00 (Ref)
25/53	1	0	1.37 (0.72-2.61)	18/31	1	0	**3.28 (1.42-7.57)**	7/22	1	0	**0.22 (0.06-0.81)**
77/226	0	1	0.88 (0.57-1.38)	55/159	0	1	1.37 (0.76-2.49)	22/67	0	1	**0.38 (0.17-0.88)**
30/106	1	1	0.90 (0.51-1.60)	26/67	1	1	**2.14 (1.04-4.39)**	4/39	1	1	**0.10 (0.02-0.45)**
30/106	ROR: 0.75 (0.34-1.65)			26/67	ROR: 0.48 (0.18-1.27)			4/39	ROR: 1.18 (0.22-6.42)		

[1]Dominant Model: Homozygous minor allele + Heterozygous vs. Homozygous major allele.

*Adjusted for age at case diagnosis or reference date in controls (continuous), HIV RNA levels before set point (continuous), AIDS diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), race (reference=White), and history of HCV infection (reference=No).

Cases and controls were matched on: cohort (1 or 2), date of infection (continuous), duration of follow-up (continuous), seroconversion status (categorical), race (white or non-white), and CD4+ T-cell count at date of matching (categorical).

Table 3.6. Association of miRNA-related SNPs and Overall AIDS-NHL Among Those without a Primary KS diagnosis, and Among Caucasian-Only Participants

	Those without a Primary KS diagnosis				Caucasian-Only Participants			
	Complete Case Analysis		Imputed Analysis		Complete Case Analysis		Imputed Analysis	
	Non-1°KS Ca/Co	Adjusted* OR_{adj} (95% CI)	Non-1°KS Ca/Co	Adjusted* OR_{adj} (95% CI)	Caucasian Ca/Co	Adjusted** OR_{adj} (95% CI)	Caucasian Ca/Co	Adjusted** OR_{adj} (95% CI)
microRNA coding regions								
miR-196a rs11614913								
CC	40/150	1.00 (Ref)	65/178	1.00 (Ref)	39/149	1.00 (Ref)	59/171	1.00 (Ref)
CT	53/195	0.90 (0.54-1.50)	76/236	0.76 (0.50-1.15)	54/201	0.94 (0.56-1.57)	75/236	0.86 (0.56-1.32)
TT	10/56	0.73 (0.30-1.81)	18/72	0.59 (0.30-1.16)	10/57	0.48 (0.21-1.14)	17/72	0.53 (0.27-1.02)
Log-Add	103/40	0.88 (0.60-1.28)	159/486	0.77 (0.57-1.03)	103/407	0.77 (0.54-1.11)	151/479	0.76 (0.57-1.02)
Dominant[1]	63/25	0.88 (0.54-1.43)	94/308	0.72 (0.48-1.08)	64/258	0.83 (0.51-1.36)	92/308	0.77 (0.51-1.15)
Recessive[2]	10/56	0.78 (0.33-1.83)	18/72	0.69 (0.37-1.30)	10/57	0.50 (0.22-1.12)	17/72	0.57 (0.31-1.06)
miR-26a1 rs7372209								
CC	51/205	1.00 (Ref)	83/251	1.00 (Ref)	51/211	1.00 (Ref)	76/248	1.00 (Ref)
CT	44/167	1.18 (0.69-2.00)	68/197	1.22 (0.81-1.84)	45/170	1.38 (0.83-2.30)	68/198	1.36 (0.90-2.05)
TT	8/29	1.59 (0.58-4.37)	9/38	0.75 (0.33-1.69)	7/26	1.88 (0.63-5.60)	8/33	1.04 (0.43-2.50)
Log-Add	103/401	1.22 (0.80-1.87)	160/486	1.01 (0.74-1.39)	103/407	1.37 (0.90-2.09)	152/479	1.18 (0.86-1.64)
Dominant	52/196	1.21 (0.72-2.04)	77/235	1.14 (0.76-1.70)	52/196	1.42 (0.86-2.34)	76/231	1.32 (0.88-1.96)
Recessive	8/29	1.43 (0.55-3.69)	9/38	0.68 (0.31-1.49)	7/26	1.57 (0.55-4.52)	8/33	0.90 (0.38-2.12)
miR-27a rs895819								
TT	42/192	1.00 (Ref)	62/223	1.00 (Ref)	42/194	1.00 (Ref)	58/226	1.00 (Ref)
TC	49/173	1.12 (0.65-1.92)	78/220	1.12 (0.72-1.74)	51/180	1.20 (0.70-2.07)	77/214	1.34 (0.86-2.10)
CC	12/34	1.40 (0.54-3.64)	20/41	1.71 (0.84-3.49)	10/31	1.09 (0.40-2.93)	17/37	1.55 (0.72-3.33)
Log-Add	103/399	1.16 (0.77-1.73)	160/484	1.24 (0.90-1.71)	103/405	1.11 (0.74-1.67)	152/477	1.28 (0.92-1.79)
Dominant	61/207	1.16 (0.70-1.94)	98/261	1.20 (0.79-1.84)	61/211	1.19 (0.70-2.00)	94/251	1.37 (0.89-2.11)
Recessive	12/34	1.33 (0.53-3.33)	20/41	1.61 (0.82-3.15)	10/31	0.99 (0.38-2.56)	17/37	1.31 (0.64-2.70)
miR-300 rs12894467								
CC	36/144	1.00 (Ref)	53/165	1.00 (Ref)	36/152	1.00 (Ref)	51/171	1.00 (Ref)
CT	48/199	0.95 (0.54-1.66)	73/247	0.85 (0.54-1.35)	48/204	1.05 (0.61-1.80)	72/245	0.96 (0.61-1.50)
TT	19/57	1.53 (0.70-3.34)	34/73	1.35 (0.75-2.45)	19/50	**3.28 (1.41-7.61)**^	29/62	1.80 (0.96-3.35)
Log-Add	103/400	1.17 (0.80-1.70)	160/485	1.12 (0.83-1.50)	103/406	**1.54 (1.04-2.27)**^	152/478	1.25 (0.92-1.70)
Dominant	67/256	1.05 (0.62-1.79)	107/320	0.96 (0.62-1.48)	67/254	1.30 (0.78-2.17)	101/307	1.10 (0.72-1.68)
Recessive	19/57	1.58 (0.78-3.20)	34/73	1.50 (0.89-2.52)	19/50	**3.20 (1.46-7.03)**^	29/62	**1.85 (1.05-3.23)**
Functional SNPs in miRNA processing and maturation genes								
AGO2 rs4961280								
CC	63/270	1.00 (Ref)	108/319	1.00 (Ref)	60/275	1.00 (Ref)	98/314	1.00 (Ref)
CA	35/120	1.33 (0.76-2.31)	46/154	0.87 (0.55-1.37)	38/121	1.30 (0.76-2.21)	48/152	1.06 (0.68-1.64)
AA	5/11	1.85 (0.54-6.38)	6/13	1.31 (0.44-3.87)	5/11	1.84 (0.53-6.46)	6/13	1.47 (0.48-4.48)
Log-Add	103/401	1.34 (0.86-2.10)	160/486	0.96 (0.66-1.42)	103/407	1.32 (0.85-2.05)	152/479	1.11 (0.77-1.61)
Dominant	40/131	1.38 (0.81-2.35)	52/167	0.90 (0.58-1.41)	43/132	1.34 (0.80-2.25)	54/165	1.09 (0.71-1.67)
Recessive	5/11	1.67 (0.49-5.65)	6/13	1.39 (0.48-4.04)	5/11	1.68 (0.49-5.78)	6/13	1.44 (0.48-4.33)

Table 3.6 Continued. Association of miRNA-related SNPs and Overall AIDS-NHL Among Those without a Primary KS diagnosis, and Among Caucasian-Only Participants

	Those without a Primary KS diagnosis				Caucasian-Only Participants			
	Complete Case Analysis Overall AIDS-NHL		Imputed Analysis Overall AIDS-NHL		Complete Case Analysis Overall AIDS-NHL		Imputed Analysis Overall AIDS-NHL	
	Non-1°KS Ca/Co	Adjusted* OR_{adj} (95% CI)	Non-1°KS Ca/Co	Adjusted* OR_{adj} (95% CI)	Caucasian Ca/Co	Adjusted** OR_{adj} (95% CI)	Caucasian Ca/Co	Adjusted** OR_{adj} (95% CI)
Functional SNPs in miRNA processing and maturation genes								
GEMIN3 rs197412								
TT	33/150	1.00 (Ref)	46/174	1.00 (Ref)	33/159	1.00 (Ref)	46/182	1.00 (Ref)
TC	48/194	1.30 (0.74-2.30)	76/232	1.29 (0.81-2.04)	53/196	1.39 (0.80-2.41)	80/233	1.49 (0.95-2.32)
CC	22/57	1.82 (0.87-3.80)	38/80	**2.28 (1.24-4.18)**^	17/52	1.41 (0.64-3.09)	26/64	1.72 (0.92-3.21)
Log-Add	103/401	1.34 (0.93-1.93)	160/486	**1.48 (1.09-1.99)**^	103/407	1.23 (0.84-1.78)	152/479	1.34 (1.00-1.81)
Dominant	70/251	1.42 (0.82-2.44)	114/312	1.47 (0.95-2.28)	70/248	1.39 (0.82-2.37)	106/297	1.53 (1.00-2.35)
Recessive	22/57	1.53 (0.81-2.89)	38/80	**1.94 (1.15-3.30)**^	17/52	1.14 (0.57-2.27)	26/64	1.34 (0.77-2.34)
GEMIN4 rs7813								
TT	43/144	1.00 (Ref)	62/183	1.00 (Ref)	38/140	1.00 (Ref)	53/168	1.00 (Ref)
TC	42/189	0.84 (0.48-1.45)	69/224	0.95 (0.61-1.47)	44/196	1.10 (0.64-1.89)	68/227	1.13 (0.73-1.76)
CC	18/68	1.01 (0.49-2.11)	29/78	1.34 (0.75-2.41)	21/70	1.41 (0.69-2.87)	31/82	1.45 (0.81-2.58)
Log-Add	103/401	0.97 (0.68-1.38)	160/485	1.11 (0.84-1.48)	103/406	1.17 (0.83-1.66)	152/477	1.19 (0.90-1.58)
Dominant	60/257	0.88 (0.53-1.47)	98/302	1.04 (0.69-1.56)	65/266	1.17 (0.71-1.95)	99/309	1.21 (0.80-1.83)
Recessive	18/68	1.11 (0.56-2.19)	29/78	1.39 (0.81-2.36)	21/70	1.33 (0.71-2.52)	31/82	1.34 (0.80-2.25)
Candidate genes with SNPs near or within a predicted miRNA binding site								
E2F2 rs2075993								
AA	24/107	1.00 (Ref)	39/131	1.00 (Ref)	22/102	1.00 (Ref)	31/113	1.00 (Ref)
AG	52/191	1.40 (0.74-2.66)	80/232	1.33 (0.78-2.28)	54/198	1.40 (0.73-2.68)	79/238	1.14 (0.66-1.96)
GG	27/103	1.38 (0.66-2.87)	41/123	1.16 (0.64-2.13)	27/106	1.30 (0.61-2.79)	42/127	1.03 (0.55-1.90)
Log-Add	103/401	1.16 (0.81-1.67)	160/486	1.07 (0.79-1.44)	103/406	1.13 (0.78-1.64)	152/478	1.00 (0.74-1.36)
Dominant	79/294	1.40 (0.76-2.57)	121/355	1.27 (0.77-2.11)	81/304	1.37 (0.73-2.57)	121/365	1.10 (0.65-1.85)
Recessive	27/103	1.08 (0.61-1.92)	41/123	0.95 (0.59-1.53)	27/106	1.01 (0.57-1.80)	42/127	0.93 (0.59-1.49)
HIF1A rs2057482								
CC	67/285	1.00 (Ref)	111/336	1.00 (Ref)	70/292	1.00 (Ref)	107/341	1.00 (Ref)
CT	32/111	1.33 (0.75-2.37)	44/138	1.09 (0.69-1.73)	30/109	1.09 (0.60-1.96)	41/130	1.14 (0.70-1.85)
TT	4/4	3.64 (0.71-18.73)	5/11	1.25 (0.38-4.08)	3/5	3.34 (0.58-19.18)	4/7	1.27 (0.31-5.21)
Log-Add	103/400	1.49 (0.91-2.43)	160/485	1.10 (0.75-1.63)	103/406	1.26 (0.76-2.10)	152/478	1.14 (0.75-1.71)
Dominant	36/115	1.45 (0.83-2.52)	49/149	1.11 (0.71-1.73)	33/114	1.18 (0.67-2.09)	45/137	1.15 (0.72-1.83)
Recessive	4/4	3.36 (0.67-16.96)	5/11	1.20 (0.37-3.89)	3/5	3.26 (0.57-18.51)	4/7	1.24 (0.30-5.04)
RCHY1 rs2126852								
AA	64/201	1.00 (Ref)	91/250	1.00 (Ref)	62/204	1.00 (Ref)	88/246	1.00 (Ref)
AG	33/166	0.65 (0.38-1.11)	58/195	0.90 (0.59-1.38)	36/167	0.64 (0.38-1.10)	55/189	0.89 (0.58-1.36)
GG	6/34	0.58 (0.21-1.56)	11/41	0.68 (0.31-1.50)	5/36	0.49 (0.17-1.40)	9/44	0.51 (0.22-1.17)
Log-Add	103/401	0.71 (0.47-1.07)	160/486	0.86 (0.63-1.18)	103/407	0.67 (0.44-1.02)	152/479	0.79 (0.57-1.09)
Dominant	39/200	0.64 (0.38-1.06)	69/236	0.86 (0.58-1.29)	41/203	0.61 (0.37-1.02)	64/233	0.81 (0.54-1.21)
Recessive	6/34	0.71 (0.27-1.84)	11/41	0.71 (0.33-1.53)	5/36	0.58 (0.21-1.62)	9/44	0.53 (0.24-1.21)

Table 3.6 Continued. Association of miRNA-related SNPs and Overall AIDS-NHL Among Those without a Primary KS diagnosis, and Among Caucasian-Only Participants

	Those without a Primary KS diagnosis				Caucasian-Only Participants			
	Complete Case Analysis Overall AIDS-NHL		Imputed Analysis Overall AIDS-NHL		Complete Case Analysis Overall AIDS-NHL		Imputed Analysis Overall AIDS-NHL	
	Non-1°KS Ca/Co	Adjusted* OR_{adj} (95% CI)	Non-1°KS Ca/Co	Adjusted* OR_{adj} (95% CI)	Caucasian Ca/Co	Adjusted** OR_{adj} (95% CI)	Caucasian Ca/Co	Adjusted** OR_{adj} (95% CI)
Candidate genes with SNPs near or within a predicted miRNA binding site								
TP53INP1 rs896849								
TT	65/292	1.00 (Ref)	105/342	1.00 (Ref)	70/298	1.00 (Ref)	107/347	1.00 (Ref)
TC	34/100	1.54 (0.90-2.62)	48/131	1.13 (0.73-1.76)	29/99	1.02 (0.59-1.78)	39/121	0.91 (0.57-1.46)
CC	4/8	2.84 (0.60-13.40)	7/12	2.10 (0.73-6.06)	4/9	3.73 (0.73-19.08)	6/10	0.99 (0.29-3.31)
Log-Add	103/400	1.58 (1.00-2.50)	160/485	1.24 (0.86-1.80)	103/406	1.23 (0.77-1.98)	152/478	0.94 (0.63-1.40)
Dominant	38/108	1.61 (0.96-2.70)	55/143	1.20 (0.78-1.83)	33/108	1.13 (0.66-1.92)	45/131	0.92 (0.59-1.45)
Recessive	4/8	2.52 (0.55-11.59)	7/12	1.99 (0.70-5.65)	4/9	3.71 (0.73-18.85)	6/10	1.02 (0.31-3.37)

[1]Dominant: Homozygous minor allele + Heterozygous vs. Homozygous major allele.

[2]Recessive: Homozygous minor allele vs. Heterozygous + Homozygous major allele.

*Adjusted for age at case diagnosis or reference date in controls (continuous), HIV RNA levels before set point (continuous), AIDS diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), race (reference=White), and history of HCV infection (reference=No).

**Adjusted for age at case diagnosis or reference date in controls (continuous), HIV RNA levels before set point (continuous), AIDS diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), and history of HCV infection (reference=No).

Cases and controls were matched on: cohort (1 or 2), date of infection (continuous), duration of follow-up (continuous), seroconversion status (categorical), race (white or non-white), and CD4+ T-cell count at date of matching (categorical).

NAC= Not able to calculate due to sample size.

^Remained evident after SB approach.

Table 3.7. Display of original statistically significant estimates alongside semi-Bayes corrected posterior estimates to highlight results that remained evident after SB correction, listed by table number, data source (imputed dataset or complete case data), and NHL site

Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference *GEMIN3* rs197412 and overall AIDS-NHL

Analytic Model	Overall Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
CC vs. TT	38/84	1.85 (1.05-3.23)	1.67 (1.00-2.81)	Table 3.4
Log-Additive	179/520	1.35 (1.03-1.78)^	**1.35 (1.03-1.76)**	Table 3.4

Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference *miR-196a rs11614913* and CNS AIDS-NHL

Analytic Model	CNS Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
CT vs. CC	22/86	0.41 (0.19-0.88)^	**0.52 (0.27-0.99)**	Table 3.4
Log-Additive	57/161	0.57 (0.33-0.99)	0.62 (0.37-1.02)	Table 3.4
Dominant	26/104	0.43 (0.22-0.87)^	**0.51 (0.28-0.95)**	Table 3.4

Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference *HIF1A* rs2057482 and Systemic AIDS-NHL

Analytic Model	Systemic Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
CT vs. CC	40/92	1.85 (1.10-3.09)^	**1.67 (1.04-2.70)**	Table 3.4
Log-Additive	122/358	1.83 (1.16-2.90)^	**1.72 (1.12-2.67)**	Table 3.4
Dominant	44/96	1.91 (1.15-3.16)^	**1.77 (1.10-2.84)**	Table 3.4

Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference *HIF1A* rs2057482 and CNS AIDS-NHL

Analytic Model	CNS Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
CT vs. CC	10/54	0.23 (0.08-0.73)^	**0.42 (0.19-0.94)**	Table 3.4
Log-Additive	57/161	0.38 (0.17-0.87)^	**0.49 (0.25-0.95)**	Table 3.4
Dominant	12/60	0.25 (0.09-0.72)^	**0.40 (0.19-0.88)**	Table 3.4

Complete Cases Analysis
miR-196a rs11614913 and CNS AIDS-NHL

Analytic Model	CNS Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
Log-Additive	39/128	0.46 (0.23-0.93)	0.54 (0.29-1.00)	Table 3.4

Complete Cases Analysis
HIF1A rs2057482 and Systemic AIDS-NHL

Analytic Model	Systemic Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
Log-Add	78/301	2.01 (1.09-3.68)^	**1.80 (1.03-3.12)**	Table 3.4
Dominant	31/77	1.94 (1.03-3.66)	1.73 (0.98-3.08)	Table 3.4

Complete Cases Analysis
RCHY1 rs2126852 and Systemic AIDS-NHL

Analytic Model	Systemic Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
AG vs. AA	23/128	0.47 (0.25-0.88)^	**0.54 (0.31-0.95)**	Table 3.4
Log-Add	78/302	0.60 (0.37-0.99)	0.64 (0.40-1.01)	Table 3.4
Dominant	27/150	0.49 (0.27-0.89)^	**0.55 (0.32-0.94)**	Table 3.4

Table 3.7 Continued. Display of original statistically significant estimates alongside semi-Bayes corrected posterior estimates to highlight results that remained evident after SB correction, listed by table number, data source (imputed dataset or complete case data), and NHL site

Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference
miR-300 rs12894467 and overall AIDS-NHL, among Caucasians

Analytic Model	Overall Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
Recessive	29/62	1.85 (1.05-3.23)	1.24 (0.92-1.67)	Table 3.6

Complete Cases Analysis
miR-300 rs12894467 and overall AIDS-NHL, among Caucasians

Analytic Model	Overall Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
TT vs. CC	19/50	3.28 (1.41-7.61)^	**2.40 (1.18-4.86)**	Table 3.6
Log-Additive	67/250	1.54 (1.04-2.27)^	**1.49 (1.03-2.17)**	Table 3.6
Recessive	19/50	3.20 (1.46-7.03)^	**2.42 (1.23-4.76)**	Table 3.6

Analysis with data imputed for HIV Viral Load and CD4+ T-cell count at reference
GEMIN3 rs197412 and overall AIDS-NHL, among those without a KS diagnosis prior to reference date

Analytic Model	Overall Ca/Co	Original Estimate	SB Corrected Posterior Estimate	Table
CC vs. TT	22/57	2.28 (1.24-4.18)^	**1.98 (1.14-3.42)**	Table 3.6
Log-Additive	103/396	1.48 (1.09-1.99)^	**1.45 (1.08-1.94)**	Table 3.6
Recessive	22/57	1.94 (1.15-3.30)^	**1.79 (1.09-2.93)**	Table 3.6

^Remained evident after SB analysis

Table 3.8. Association of miRNA-related SNPs and Overall AIDS-NHL, using Unconditional Logistic Regression

	Complete Case Analysis AIDS-NHL		Imputed Analysis AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)
microRNA coding regions				
miR-196a rs11614913				
CC	43/164	1.00 (Ref)	72/196	1.00 (Ref)
CT	63/212	1.06 (0.66-1.68)	90/257	0.90 (0.61-1.33)
TT	11/59	0.56 (0.25-1.24)	19/76	0.56 (0.30-1.05)
Log-Add	117/435	0.84 (0.60-1.18)	181/529	0.80 (0.60-1.05)
Dominant[1]	74/271	0.95 (0.61-1.50)	109/333	0.83 (0.57-1.20)
Recessive[2]	11/59	0.54 (0.26-1.13)	19/76	0.60 (0.34-1.07)
miR-26a1 rs7372209				
CC	59/226	1.00 (Ref)	95/276	1.00 (Ref)
CT	50/180	1.17 (0.74-1.83)	78/215	1.13 (0.78-1.64)
TT	8/29	1.15 (0.48-2.78)	9/38	0.71 (0.32-1.58)
Log-Add	117/435	1.12 (0.79-1.58)	182/529	0.98 (0.73-1.31)
Dominant	58/209	1.17 (0.75-1.80)	87/253	1.06 (0.74-1.53)
Recessive	8/29	1.07 (0.46-2.51)	9/38	0.67 (0.31-1.46)
miR-27a rs895819				
TT	47/206	1.00 (Ref)	68/242	1.00 (Ref)
TC	57/191	1.31 (0.82-2.08)	92/242	1.26 (0.85-1.86)
CC	13/36	1.56 (0.74-3.32)	22/43	1.84 (0.99-3.43)
Log-Add	117/433	1.27 (0.91-1.77)	182/527	1.32 (1.00-1.75)
Dominant	70/227	1.35 (0.86-2.10)	114/285	1.35 (0.93-1.96)
Recessive	13/36	1.36 (0.67-2.77)	22/43	1.62 (0.90-2.91)
miR-300 rs12894467				
CC	40/159	1.00 (Ref)	58/183	1.00 (Ref)
CT	56/212	1.01 (0.62-1.63)	86/265	0.97 (0.65-1.46)
TT	21/63	1.13 (0.59-2.17)	38/80	1.24 (0.73-2.08)
Log-Add	117/434	1.05 (0.77-1.44)	182/528	1.09 (0.84-1.41)
Dominant	77/275	1.04 (0.66-1.63)	124/345	1.04 (0.71-1.52)
Recessive	21/63	1.13 (0.63-2.02)	38/80	1.26 (0.79-2.00)
pre-miR-146a rs2910164				
GG	71/260	1.00 (Ref)	111/314	1.00 (Ref)
GC	43/156	1.15 (0.73-1.81)	62/193	0.95 (0.65-1.40)
CC	3/19	0.54 (0.14-2.03)	9/22	0.96 (0.40-2.31)
Log-Add	117/435	0.98 (0.67-1.44)	182/529	0.96 (0.70-1.32)
Dominant	46/175	1.07 (0.69-1.67)	71/215	0.95 (0.66-1.38)
Recessive	3/19	0.51 (0.14-1.91)	9/22	0.98 (0.42-2.32)

Table 3.8 Continued. Association of miRNA-related SNPs and Overall AIDS-NHL, using Unconditional Logistic Regression

	Complete Case Analysis AIDS-NHL		Imputed Analysis AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)
Functional SNPs in miRNA processing and maturation genes				
AGO2 rs4961280				
CC	69/291	1.00 (Ref)	119/348	1.00 (Ref)
CA	42/132	1.34 (0.84-2.13)	56/167	0.98 (0.66-1.46)
AA	6/12	2.32 (0.77-6.93)	7/14	1.55 (0.57-4.23)
Log-Add	117/435	1.41 (0.96-2.06)	182/529	1.07 (0.76-1.50)
Dominant	48/144	1.41 (0.90-2.21)	63/181	1.02 (0.69-1.51)
Recessive	6/12	2.08 (0.71-6.14)	7/14	1.57 (0.58-4.22)
DICER1 rs3742330				
AA	96/353	1.00 (Ref)	152/430	1.00 (Ref)
AG	20/80	0.92 (0.52-1.64)	29/96	0.89 (0.55-1.44)
GG	1/2	1.89 (0.16-23.07)	1/3	1.20 (0.12-12.39)
Log-Add	117/435	0.98 (0.58-1.66)	182/529	0.91 (0.58-1.43)
Dominant	21/82	0.95 (0.54-1.66)	30/99	0.90 (0.56-1.44)
Recessive	1/2	1.90 (0.16-23.16)	1/3	1.22 (0.12-12.57)
GEMIN3 rs197412				
TT	38/163	1.00 (Ref)	54/189	1.00 (Ref)
TC	56/212	1.18 (0.73-1.92)	88/254	1.23 (0.81-1.86)
CC	23/60	1.74 (0.91-3.31)	40/86	1.67 (0.98-2.85)
Log-Add	117/435	1.29 (0.94-1.78)	182/529	1.28 (0.99-1.67)
Dominant	79/272	1.29 (0.82-2.05)	128/340	1.33 (0.90-1.97)
Recessive	23/60	1.58 (0.89-2.80)	40/86	1.47 (0.92-2.35)
GEMIN4 rs2740348				
GG	80/308	1.00 (Ref)	124/376	1.00 (Ref)
GC	32/111	1.18 (0.73-1.93)	50/134	1.13 (0.75-1.69)
CC	5/16	1.08 (0.35-3.32)	8/18	1.00 (0.39-2.61)
Log-Add	117/435	1.12 (0.76-1.65)	182/528	1.07 (0.77-1.49)
Dominant	37/127	1.17 (0.74-1.86)	58/152	1.11 (0.75-1.64)
Recessive	5/16	1.04 (0.34-3.16)	8/18	0.97 (0.38-2.50)
GEMIN4 rs7813				
TT	46/158	1.00 (Ref)	68/200	1.00 (Ref)
TC	48/203	0.95 (0.59-1.55)	79/242	1.08 (0.72-1.61)
CC	23/73	1.29 (0.70-2.36)	35/85	1.37 (0.82-2.31)
Log-Add	117/434	1.11 (0.82-1.50)	182/527	1.16 (0.90-1.49)
Dominant	71/276	1.04 (0.67-1.63)	114/327	1.15 (0.79-1.67)
Recessive	23/73	1.32 (0.76-2.28)	35/85	1.32 (0.83-2.10)

Table 3.8 Continued. Association of miRNA-related SNPs and Overall AIDS-NHL, using Unconditional Logistic Regression

	Complete Case Analysis AIDS-NHL		Imputed Analysis AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)
Functional SNPs in miRNA processing and maturation genes				
RAN rs14035				
CC	49/219	1.00 (Ref)	87/257	1.00 (Ref)
CT	57/175	1.50 (0.95-2.36)	73/222	1.03 (0.71-1.52)
TT	11/41	1.21 (0.56-2.60)	22/50	1.13 (0.62-2.04)
Log-Add	117/435	1.23 (0.89-1.69)	182/529	1.05 (0.81-1.37)
Dominant	68/216	1.44 (0.93-2.22)	95/272	1.05 (0.74-1.51)
Recessive	11/41	1.00 (0.48-2.06)	22/50	1.11 (0.63-1.96)
XPO5 rs11077				
AA	42/142	1.00 (Ref)	62/175	1.00 (Ref)
AC	53/216	0.81 (0.50-1.31)	82/265	0.91 (0.61-1.37)
CC	20/77	0.84 (0.44-1.58)	36/89	1.14 (0.68-1.92)
Log-Add	115/435	0.90 (0.66-1.22)	180/529	1.05 (0.81-1.35)
Dominant	73/293	0.81 (0.52-1.29)	118/354	0.97 (0.66-1.43)
Recessive	20/77	0.95 (0.53-1.68)	36/89	1.21 (0.76-1.91)
Candidate genes with SNPs near or within a predicted miRNA binding site				
CDK6 rs42031				
AA	67/277	1.00 (Ref)	117/335	1.00 (Ref)
AT	44/140	1.28 (0.81-2.02)	56/174	0.93 (0.63-1.37)
TT	6/17	1.75 (0.63-4.86)	8/19	1.48 (0.60-3.62)
Log-Add	117/434	1.30 (0.90-1.87)	181/528	1.03 (0.75-1.42)
Dominant	50/157	1.32 (0.85-2.05)	64/193	0.98 (0.67-1.42)
Recessive	6/17	1.61 (0.59-4.40)	8/19	1.51 (0.62-3.68)
CXCL12 rs1804429				
TT	109/407	1.00 (Ref)	164/498	1.00 (Ref)
TG	8/28	0.91 (0.39-2.14)	18/30	1.79 (0.93-3.44)
GG	0/0	NAC	0/1	NAC
Log-Add	117/435	0.91 (0.39-2.14)	182/529	1.60 (0.85-3.02)
Dominant	8/28	0.91 (0.39-2.14)	18/31	1.71 (0.89-3.29)
Recessive	0/0	NAC	0/1	NAC
E2F2 rs2075993				
AA	27/116	1.00 (Ref)	44/142	1.00 (Ref)
AG	60/206	1.44 (0.83-2.51)	93/252	1.33 (0.84-2.11)
GG	30/112	1.30 (0.70-2.44)	45/134	1.16 (0.68-1.98)
Log-Add	117/434	1.13 (0.83-1.53)	182/528	1.07 (0.83-1.39)
Dominant	90/318	1.39 (0.82-2.35)	138/386	1.27 (0.82-1.97)
Recessive	30/112	1.02 (0.62-1.66)	45/134	0.96 (0.63-1.45)

Table 3.8 Continued. Association of miRNA-related SNPs and Overall AIDS-NHL, using Unconditional Logistic Regression

	Complete Case Analysis AIDS-NHL		Imputed Analysis AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)
Candidate genes with SNPs near or within a predicted miRNA binding site				
HIF1A rs2057482				
CC	78/312	1.00 (Ref)	125/369	1.00 (Ref)
CT	35/117	1.34 (0.83-2.17)	51/147	1.12 (0.75-1.68)
TT	4/5	2.87 (0.66-12.48)	6/12	1.40 (0.46-4.22)
Log-Add	117/434	1.43 (0.94-2.17)	182/528	1.14 (0.81-1.61)
Dominant	39/122	1.41 (0.89-2.25)	57/159	1.14 (0.77-1.69)
Recessive	4/5	2.63 (0.61-11.34)	6/12	1.34 (0.45-4.01)
IL15 rs10519613				
CC	92/356	1.00 (Ref)	137/433	1.00 (Ref)
CA	23/76	1.21 (0.69-2.13)	42/91	1.47 (0.93-2.31)
AA	1/3	1.11 (0.11-11.01)	2/5	1.20 (0.22-6.59)
Log-Add	116/435	1.18 (0.71-1.96)	181/529	1.37 (0.92-2.06)
Dominant	24/79	1.20 (0.69-2.09)	44/96	1.45 (0.93-2.26)
Recessive	1/3	1.08 (0.11-10.69)	2/5	1.12 (0.21-6.14)
IL6R rs4072391				
CC	76/284	1.00 (Ref)	117/337	1.00 (Ref)
CT	35/133	1.00 (0.62-1.60)	58/169	1.04 (0.70-1.53)
TT	6/17	1.21 (0.44-3.34)	7/22	0.78 (0.31-1.99)
Log-Add	117/434	1.04 (0.72-1.51)	182/528	0.97 (0.71-1.33)
Dominant	41/150	1.02 (0.65-1.61)	65/191	1.01 (0.69-1.46)
Recessive	6/17	1.21 (0.44-3.31)	7/22	0.77 (0.31-1.94)
KRAS rs9266				
CC	37/124	1.00 (Ref)	53/151	1.00 (Ref)
CT	49/215	0.76 (0.46-1.26)	85/261	0.89 (0.58-1.36)
TT	30/96	0.85 (0.47-1.54)	43/117	0.92 (0.55-1.52)
Log-Add	116/435	0.91 (0.67-1.23)	181/529	0.95 (0.74-1.23)
Dominant	79/311	0.79 (0.49-1.26)	128/378	0.90 (0.60-1.33)
Recessive	30/96	1.00 (0.60-1.67)	43/117	0.99 (0.64-1.52)
RCHY1 rs2126852				
AA	70/218	1.00 (Ref)	104/272	1.00 (Ref)
AG	41/179	0.70 (0.45-1.11)	67/210	0.87 (0.60-1.27)
GG	6/38	0.47 (0.18-1.20)	11/47	0.58 (0.28-1.21)
Log-Add	117/435	0.69 (0.49-0.99)	182/529	0.81 (0.61-1.08)
Dominant	47/217	0.66 (0.43-1.02)	78/257	0.82 (0.57-1.17)
Recessive	6/38	0.54 (0.22-1.37)	11/47	0.61 (0.30-1.26)

Table 3.8 Continued. Association of miRNA-related SNPs and Overall AIDS-NHL, using Unconditional Logistic Regression

	Complete Case Analysis AIDS-NHL		Imputed Analysis AIDS-NHL	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)
Candidate genes with SNPs near or within a predicted miRNA binding site				
TP53INP1 rs7760				
TT	87/341	1.00 (Ref)	135/407	1.00 (Ref)
TG	28/87	1.09 (0.65-1.83)	43/111	1.00 (0.65-1.54)
GG	2/7	0.94 (0.18-4.97)	4/11	0.92 (0.27-3.12)
Log-Add	117/435	1.06 (0.68-1.66)	182/529	0.98 (0.68-1.43)
Dominant	30/94	1.08 (0.65-1.79)	47/122	0.99 (0.65-1.51)
Recessive	2/7	0.92 (0.18-4.84)	4/11	0.92 (0.27-3.10)
TP53INP1 rs896849				
TT	77/318	1.00 (Ref)	122/375	1.00 (Ref)
TC	35/107	1.28 (0.79-2.06)	50/140	1.01 (0.67-1.52)
CC	5/9	1.64 (0.49-5.42)	10/13	1.82 (0.72-4.58)
Log-Add	117/434	1.28 (0.86-1.90)	182/528	1.14 (0.82-1.58)
Dominant	40/116	1.31 (0.83-2.08)	60/153	1.08 (0.73-1.59)
Recessive	5/9	1.51 (0.46-4.95)	10/13	1.81 (0.73-4.51)

[1]Dominant: Homozygous minor allele + Heterozygous vs. Homozygous major allele.

[2]Recessive: Homozygous minor allele vs. Heterozygous + Homozygous major allele.

*Adjusted for cohort (1 or 2), date of infection (continuous), duration of follow-up (continuous), seroconversion status (categorical), race (white or non-white), CD4+ T-cell count at date of matching (categorical), age at case diagnosis or reference date in controls (continuous), HIV RNA levels before set-point (continuous), AIDS Diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), race (reference=White), and history of HCV infection (reference=No).

NAC= Not able to calculate due to sample size.

Table 3.9. Investigating Model and Highlighted Result Robustness, miR-related SNPs and Overall AIDS-NHL

	Complete Case Analysis		Median Imputed Analysis		Multiply Imputed Analysis		Without adjusting for HIV Viral Load[3]		Matched analysis	
	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted* OR_{adj} (95% CI)	Overall Ca/Co	Adjusted** OR_{adj} (95% CI)	Overall Ca/Co	Adjusted+ OR_{adj} (95% CI)
Candidate genes with SNPs near or within a predicted miRNA binding site										
miR-196a rs11614913										
CC	43/164	1.00 (Ref)	72/196	1.00 (Ref)	72/196	1.00 (Ref)	72/196	1.00 (Ref)	72/196	1.00 (Ref)
CT	63/212	1.00 (0.62-1.59)	88/257	0.88 (0.60-1.29)	88/257	0.88 (0.61-1.29)	88/257	0.93 (0.64-1.35)	88/257	0.97 (0.68-1.38)
TT	11/59	0.59 (0.26-1.33)	19/76	0.57 (0.31-1.06)	19/76	0.56 (0.30-1.04)	19/76	0.61 (0.33-1.12)	19/76	0.69 (0.39-1.24)
Log-Add	**117/435**	**0.85 (0.60-1.19)**	**179/529**	**0.80 (0.61-1.05)**	**179/529**	**0.79 (0.61-1.04)**	**179/529**	**0.83 (0.63-1.08)**	**179/52**	**0.88 (0.68-1.13)**
Dominant	74/271	0.91 (0.58-1.44)	107/333	0.81 (0.56-1.17)	107/333	0.81 (0.56-1.16)	107/333	0.85 (0.60-1.23)	107/33	0.91 (0.65-1.27)
Recessive[2]	11/59	0.59 (0.28-1.27)	19/76	0.61 (0.34-1.10)	19/76	0.60 (0.34-1.07)	19/76	0.63 (0.35-1.13)	19/76	0.70 (0.41-1.22)
Functional SNPs in miRNA processing and maturation genes										
GEMIN3 rs197412										
TT	38/163	1.00 (Ref)	54/189	1.00 (Ref)	54/189	1.00 (Ref)	54/189	1.00 (Ref)	54/189	1.00 (Ref)
TC	56/212	1.24 (0.73-2.08)	87/254	1.34 (0.88-2.03)	87/254	1.31 (0.87-1.98)	87/254	1.24 (0.82-1.86)	87/254	1.19 (0.81-1.75)
CC	23/60	1.63 (0.81-3.30)	39/86	1.85 (1.05-3.23)	39/86	1.88 (1.08-3.28)	39/86	1.69 (0.97-2.93)	39/86	1.62 (0.98-2.68)
Log-Add	**117/435**	**1.27 (0.90-1.79)**	**180/529**	**1.35 (1.03-1.78)**	**180/529**	**1.36 (1.04-1.79)**	**180/529**	**1.29 (0.98-1.69)**	**180/52**	**1.26 (0.98-1.62)**
Dominant	79/272	1.31 (0.80-2.16)	126/340	1.44 (0.97-2.15)	126/340	1.43 (0.96-2.12)	126/340	1.33 (0.90-1.96)	126/34	1.29 (0.89-1.86)
Recessive	23/60	1.42 (0.77-2.64)	39/86	1.54 (0.94-2.53)	39/86	1.60 (0.98-2.61)	39/86	1.48 (0.91-2.41)	39/86	1.46 (0.93-2.28)
Candidate genes with SNPs near or within a predicted miRNA binding site										
HIF1A rs2057482										
CC	78/312	1.00 (Ref)	125/369	1.00 (Ref)	125/369	1.00 (Ref)	125/369	1.00 (Ref)	125/36	1.00 (Ref)
CT	35/117	1.13 (0.66-1.95)	49/147	1.10 (0.71-1.71)	49/147	1.08 (0.70-1.66)	49/147	1.11 (0.72-1.70)	49/147	0.97 (0.65-1.45)
TT	4/5	3.72 (0.74-18.58)	6/12	1.29 (0.43-3.84)	6/12	1.38 (0.45-4.19)	6/12	1.28 (0.42-3.86)	6/12	1.36 (0.48-3.86)
Log-Add	**117/434**	**1.33 (0.83-2.11)**	**180/528**	**1.12 (0.78-1.60)**	**180/528**	**1.11 (0.77-1.59)**	**180/528**	**1.12 (0.78-1.60)**	**180/52**	**1.03 (0.74-1.45)**
Dominant	39/122	1.25 (0.74-2.11)	55/159	1.12 (0.74-1.71)	55/159	1.10 (0.73-1.67)	55/159	1.13 (0.74-1.70)	55/159	1.00 (0.68-1.47)
Recessive	4/5	3.60 (0.73-17.82)	6/12	1.25 (0.42-3.69)	6/12	1.34 (0.44-4.05)	6/12	1.24 (0.41-3.69)	6/12	1.37 (0.49-3.86)

* Adjusted for age at case diagnosis or reference date in controls (continuous), HIV RNA levels before set point (continuous), AIDS diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), race (reference=White), and history of HCV infection (reference=No).

** Adjusted for age at case diagnosis or reference date in controls (continuous), AIDS diagnosis prior to case diagnosis or reference date in controls (reference=No), history of ART Treatment (reference=No), CD4+ T-cell at reference date (continuous), race (reference=White), and history of HCV infection (reference=No).

+ Cases and controls were matched on: cohort (1 or 2), date of infection (continuous), duration of follow-up (continuous), seroconversion status (categorical), race (white or non-white), and CD4+ T-cell count at date of matching (categorical).

[3]We also ran this analysis using CD4_match, so as to not use any imputed data while removing HIV viral load at set-point. The results did not change more than in the hundredths.

Table 3.10. Association of SNPs within miRNA Processing Genes and Pre-Reference Date miRNA serum levels among subgroup of 77 study participants (n=61 AIDS-NHL and 16 controls), Using Linear Regression

		Mean log_e(microRNA-21)	**Mean log_e(microRNA-122)**	**Mean log_e(microRNA-223)**	**Mean log_e(microRNA-222)**
Genotype	N	Mean $Ratios_{asj}$ (95% CI)	Mean $Ratios_{asj}$ (95% CI)	Mean $Ratios_{asj}$ (95% CI)	Mean $Ratios_{asj}$ (95% CI)
AGO2 rs4961280					
CC	54	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
CA	21	0.80 (0.55-1.16)	0.71 (0.25-2.03)	0.78 (0.48-1.24)	0.88 (0.64-1.21)
AA	2	---	---	---	---
Log-Add	77	0.92 (0.67-1.28)	1.21 (0.49-3.01)	0.80 (0.54-1.20)	0.97 (0.73-1.28)
p for trend	77	0.63	0.68	0.27	0.81
CA+AA	23	0.85 (0.59-1.22)	0.92 (0.32-2.59)	0.77 (0.49-1.22)	0.92 (0.67-1.25)
DICER1 rs3742330					
AA	64	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
AG	12	0.81 (0.51-1.29)	0.54 (0.14-2.06)	0.84 (0.47-1.50)	0.98 (0.66-1.45)
GG	1	---	---	---	---
Log-Add	77	0.83 (0.56-1.23)	0.55 (0.18-1.74)	0.83 (0.50-1.36)	0.89 (0.63-1.25)
p for trend	77	0.35	0.31	0.45	0.49
AG+GG	13	0.81 (0.52-1.26)	0.52 (0.14-1.89)	0.82 (0.47-1.43)	0.93 (0.63-1.36)
GEMIN3 rs197412					
TT	26	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
TC	37	1.31 (0.89-1.92)	0.98 (0.33-2.91)	1.29 (0.80-2.09)	1.33 (0.96-1.84)
CC	14	1.51 (0.90-2.53)	1.92 (0.44-8.36)	1.72 (0.90-3.31)	1.40 (0.90-2.17)
Log-Add	77	1.24 (0.97-1.59)	1.29 (0.64-2.60)	1.31 (0.96-1.78)	1.21 (0.98-1.49)
p for trend	77	0.08	0.47	0.09	0.07
TC+CC	51	1.36 (0.95-1.94)	1.17 (0.42-3.24)	1.39 (0.89-2.19)	1.35 (0.99-1.82)
GEMIN4 rs2740348					
GG	49	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
GC	26	1.12 (0.78-1.62)	1.83 (0.67-5.04)	1.23 (0.78-1.94)	1.12 (0.82-1.52)
CC	2	---	---	---	---
Log-Add	77	1.12 (0.82-1.54)	1.52 (0.63-3.69)	1.14 (0.77-1.70)	1.10 (0.84-1.45)
p for trend	77	0.47	0.35	0.51	0.48
GC+CC	28	1.13 (0.79-1.61)	1.75 (0.65-4.70)	1.21 (0.77-1.88)	1.12 (0.83-1.51)
GEMIN4 rs7813					
TT	29	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
TC	35	1.02 (0.71-1.47)	1.21 (0.43-3.38)	1.09 (0.68-1.72)	0.98 (0.72-1.35)
CC	13	0.75 (0.46-1.22)	0.63 (0.15-2.60)	0.75 (0.40-1.38)	0.75 (0.49-1.14)
Log-Add	77	0.89 (0.71-1.13)	0.87 (0.45-1.71)	0.90 (0.67-1.21)	0.89 (0.73-1.09)
p for trend	77	0.34	0.69	0.50	0.24
TC+CC	48	0.94 (0.66-1.32)	1.02 (0.39-2.67)	0.98 (0.64-1.51)	0.91 (0.68-1.22)

Table 3.10 Cont. Association of SNPs within miRNA Processing Genes and Pre-Reference Date miRNA serum levels among subgroup of 77 study participants (n=61 AIDS-NHL and 16 controls), Using Linear Regression

		Mean $\log_e$(microRNA-21)	**Mean $\log_e$(microRNA-122)**	**Mean $\log_e$(microRNA-223)**	**Mean $\log_e$(microRNA-222)**
Genotype	N	Mean Ratios$_{asj}$ (95% CI)	Mean Ratios$_{asj}$ (95% CI)	Mean Ratios$_{asj}$ (95% CI)	Mean Ratios$_{asj}$ (95% CI)
RAN rs14035					
CC	38	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
CT	34	0.72 (0.51-1.01)	0.45 (0.17-1.18)	0.69 (0.45-1.06)	0.83 (0.61-1.11)
TT	5	0.85 (0.43-1.68)	0.82 (0.12-5.60)	1.43 (0.61-3.35)	0.98 (0.54-1.77)
Log-Add	77	0.81 (0.62-1.06)	0.64 (0.30-1.36)	0.91 (0.65-1.28)	0.90 (0.72-1.14)
p for trend	77	0.13	0.24	0.58	0.39
CT+TT	39	0.73 (0.53-1.02)	0.49 (0.19-1.23)	0.76 (0.50-1.15)	0.85 (0.64-1.12)
XPO5 rs11077					
AA	22	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
AC	41	0.83 (0.56-1.23)	0.82 (0.27-2.52)	1.46 (0.90-2.37)	0.97 (0.69-1.35)
CC	14	0.75 (0.45-1.23)	0.52 (0.13-2.19)	1.46 (0.78-2.74)	0.80 (0.52-1.23)
Log-Add	77	0.86 (0.67-1.10)	0.73 (0.36-1.48)	1.24 (0.91-1.68)	0.90 (0.73-1.12)
p for trend	77	0.23	0.38	0.17	0.34
AC+CC	55	0.81 (0.56-1.17)	0.73 (0.25-2.13)	1.46 (0.92-2.31)	0.92 (0.67-1.27)

Adjusted for AIDS-NHL case indicator (reference=No), race (reference=White), and CD4+ T-cell count at date of serum sample (continuous).

---Unable to estimate due to sample size.

Table 3.11. Association of SNPs within miRNA Processing Genes and Pre-Reference Date miRNA serum levels among subgroup of 77 study participants (n=61 AIDS-NHL and 16 controls), Using Linear Regression, stratified by case-control status

	Ca	Co	**Mean $\log_e$(microRNA-21)**		**Mean $\log_e$(microRNA-122)**		**Mean $\log_e$(microRNA-223)**		**Mean $\log_e$(microRNA-222)**	
Genotyp			Mean Ratios$_{asj}$ (95% CI)		Mean Ratios$_{asj}$ (95% CI)		Mean Ratios$_{asj}$ (95% CI)		Mean Ratios$_{asj}$ (95% CI)	
AGO2 rs4961280			AIDS-NHL	Controls	AIDS-NHL	Controls	AIDS-NHL	Controls	AIDS-NHL	Controls
CC	45	9	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
CA	15	6	0.84 (0.54-1.30)	0.87 (0.39-1.94)	0.96 (0.30-3.09)	0.15 (0.01-3.28)	0.72 (0.42-1.25)	1.51 (0.42-5.43)	0.94 (0.66-1.36)	0.88 (0.32-2.42)
AA	1	1	---	---	---	---	---	---	---	---
Log-Add	61	16	0.78 (0.53-1.15)	1.77 (0.89-3.53)	0.86 (0.31-2.39)	2.82 (0.18-43.17)	0.72 (0.42-1.25)	1.51 (0.64-3.60)	0.93 (0.67-1.28)	1.31 (0.64-2.70)
p for trend	61	16	0.21	0.10	0.77	0.43	0.12	0.32	0.64	0.43
CA+AA	16	7	0.80 (0.52-1.22)	1.33 (0.49-3.65)	0.90 (0.29-2.80)	0.67 (0.02-27.17)	0.69 (0.40-1.17)	1.64 (0.52-5.23)	0.93 (0.66-1.32)	1.11 (0.42-2.96)
DICER1 rs3742330										
AA	50	14	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
AG	10	2	0.78 (0.47-1.30)	---	0.43 (0.11-1.67)	---	0.84 (0.44-1.60)	---	0.88 (0.58-1.33)	---
GG	1	0	---	---	---	---	---	---	---	---
Log-Add	61	16	0.81 (0.53-1.24)	1.00 (0.28-3.53)	0.47 (0.15-1.45)	1.28 (0.01-124.10)	0.82 (0.48-1.42)	0.88 (0.20-3.83)	0.81 (0.57-1.15)	1.69 (0.53-5.44)
p for trend	61	16	0.32	1.00	0.19	0.91	0.48	0.85	0.23	0.35
AG+GG	11	2	0.78 (0.48-1.27)	---	0.42 (0.12-1.52)	---	0.82 (0.44-1.52)	---	0.83 (0.56-1.24)	---
GEMIN3 rs197412										
TT	17	9	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
TC	31	6	1.28 (0.81-2.02)	1.21 (0.64-2.30)	1.20 (0.37-3.90)	0.49 (0.04-6.71)	1.49 (0.84-2.64)	0.78 (0.27-2.20)	1.25 (0.86-1.81)	1.42 (0.65-3.11)
CC	13	1	1.23 (0.70-2.17)	---	1.09 (0.25-4.69)	---	1.85 (0.91-3.76)	---	1.24 (0.78-1.98)	---
Log-Add	61	16	1.13 (0.86-1.49)	**1.86 (1.06-3.29)**	1.06 (0.52-2.16)	3.13 (0.29-33.42)	1.38 (0.98-1.94)	1.01 (0.46-2.24)	1.13 (0.90-1.42)	1.61 (0.89-2.89)
p for trend	61	16	0.37	**0.03**	0.86	0.32	0.07	0.97	0.28	0.10
TC+CC	44	7	1.26 (0.83-1.92)	1.59 (0.72-3.52)	1.16 (0.40-3.41)	1.27 (0.06-26.61)	1.60 (0.94-2.70)	0.87 (0.33-2.31)	1.25 (0.88-1.76)	1.60 (0.75-3.40)
GEMIN4 rs2740348										
GG	37	12	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
GC	22	4	1.14 (0.76-1.71)	---	2.27 (0.83-6.24)	---	1.15 (0.69-1.93)	---	1.00 (0.72-1.40)	---
CC	2	0	---	---	---	---	---	---	---	---
Log-Add	61	16	1.14 (0.81-1.61)	1.09 (0.41-2.87)	1.72 (0.72-4.10)	0.70 (0.02-23.49)	1.08 (0.70-1.68)	1.71 (0.58-5.06)	1.02 (0.77-1.36)	1.85 (0.79-4.34)
p for trend	61	16	0.44	0.85	0.22	0.83	0.72	0.31	0.87	0.14
GC+CC	24	4	1.15 (0.78-1.70)	---	2.13 (0.80-5.68)	---	1.13 (0.69-1.87)	---	1.01 (0.73-1.40)	---
GEMIN4 rs7813										
TT	23	6	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
TC	30	5	0.93 (0.62-1.39)	1.58 (0.58-4.35)	1.09 (0.39-3.09)	2.50 (0.06-104.10)	0.96 (0.58-1.60)	1.58 (0.49-5.13)	0.82 (0.60-1.14)	2.14 (0.86-5.34)
CC	8	5	0.76 (0.42-1.37)	0.84 (0.30-2.36)	0.84 (0.17-4.24)	0.32 (0.01-14.65)	0.52 (0.24-1.09)	1.99 (0.60-6.64)	0.64 (0.39-1.03)	1.31 (0.52-3.34)
Log-Add	61	16	0.88 (0.67-1.17)	0.96 (0.57-1.61)	0.97 (0.47-2.01)	0.64 (0.10-4.10)	0.78 (0.55-1.10)	1.42 (0.81-2.52)	0.80 (0.64-1.01)	1.20 (0.74-1.95)
p for trend	61	16	0.38	0.86	0.93	0.61	0.15	0.20	0.06	0.43
TC+CC	38	10	0.89 (0.60-1.30)	1.17 (0.50-2.75)	1.04 (0.39-2.79)	0.93 (0.04-21.17)	0.84 (0.52-1.38)	1.77 (0.69-4.56)	0.78 (0.57-1.06)	1.69 (0.79-3.63)

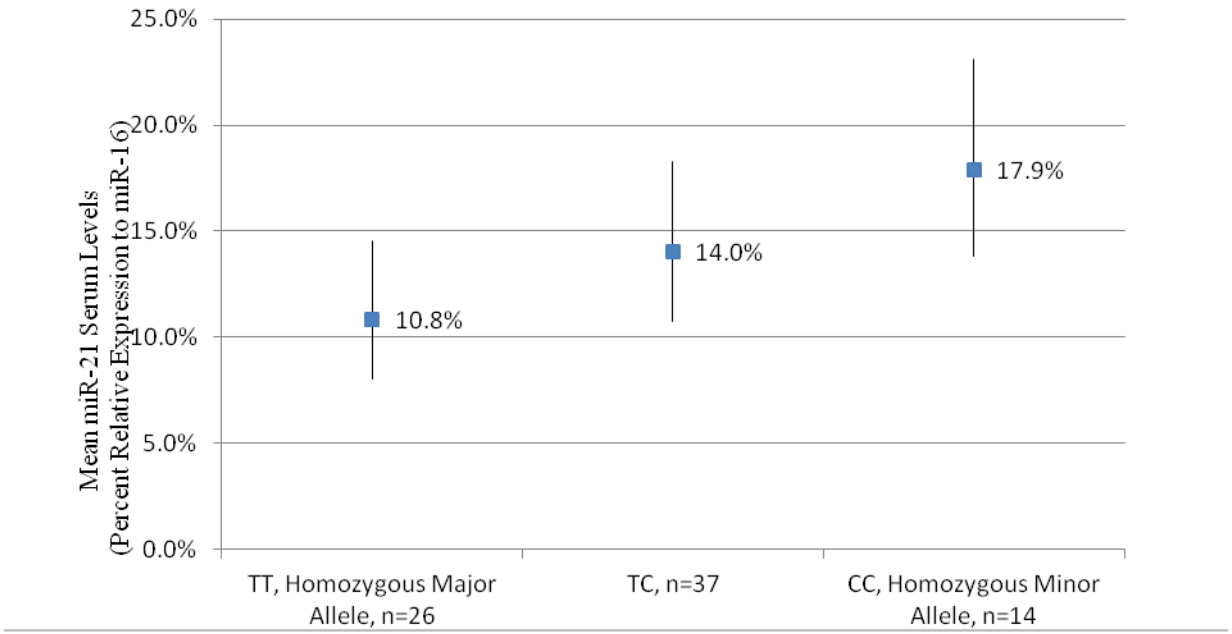
Table 3.11 Cont. Association of SNPs within miRNA Processing Genes and Pre-Reference Date miRNA serum levels among subgroup of 77 study participants (n=61 AIDS-NHL and 16 controls), Using Linear Regression , stratified by case-control status

			Mean $\log_e$(microRNA-21)		**Mean $\log_e$(microRNA-122)**		**Mean $\log_e$(microRNA-223)**		**Mean $\log_e$(microRNA-222)**	
Genotyp	Ca	Co	Mean Ratios$_{asj}$ (95% CI)		Mean Ratios$_{asj}$ (95% CI)		Mean Ratios $_{asj}$ (95% CI)		Mean Ratios$_{asj}$ (95% CI)	
R̂AN rs14035			AIDS-NHL	Controls	AIDS-NHL	Controls	AIDS-NHL	Controls	AIDS-NHL	Controls
CC	29	9	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
CT	29	5	0.74 (0.50-1.08)	0.61 (0.25-1.50)	0.51 (0.19-1.37)	0.28 (0.01-8.81)	0.63 (0.39-1.02)	0.92 (0.30-2.84)	0.82 (0.60-1.12)	0.83 (0.33-2.07)
TT	3	2	---	---	---	---	---	---	---	---
Log-Add	61	16	0.85 (0.62-1.17)	0.68 (0.40-1.18)	0.72 (0.32-1.64)	0.49 (0.06-4.03)	0.83 (0.55-1.25)	1.09 (0.54-2.17)	0.92 (0.71-1.20)	0.83 (0.47-1.44)
p for trend	61	16	0.32	0.16	0.43	0.48	0.36	0.80	0.52	0.47
CT+TT	32	7	0.76 (0.53-1.11)	0.58 (0.27-1.26)	0.56 (0.21-1.46)	0.30 (0.02-5.90)	0.69 (0.43-1.10)	1.02 (0.38-2.74)	0.85 (0.62-1.15)	0.78 (0.35-1.73)
XPO5 rs11077										
AA	19	3	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
AC	32	9	0.73 (0.48-1.10)	1.66 (0.54-5.10)	0.46 (0.16-1.38)	16.81 (0.34-831.25)	1.54 (0.90-2.62)	1.08 (0.27-4.30)	0.82 (0.58-1.17)	2.14 (0.81-5.66)
CC	10	4	0.70 (0.40-1.24)	---	0.29 (0.07-1.24)	---	1.56 (0.76-3.22)	---	0.78 (0.49-1.26)	---
Log-Add	61	16	0.82 (0.62-1.08)	1.02 (0.54-1.93)	0.52 (0.26-1.06)	2.60 (0.28-24.03)	1.30 (0.91-1.84)	1.01 (0.48-2.11)	0.87 (0.70-1.10)	1.00 (0.54-1.83)
p for trend	61	16	0.15	0.94	0.07	0.37	0.15	0.98	0.24	0.99
AC+CC	42	13	0.72 (0.48-1.07)	1.47 (0.51-4.20)	0.41 (0.15-1.17)	13.82 (0.39-495.64)	1.54 (0.93-2.56))	1.06 (0.30-3.75)	0.81 (0.58-1.13)	1.74 (0.65-4.63)

Adjusted for race (reference=White) and CD4+ T-cell count at date of serum sample (continuous).

--- = Not able to estimate due to sample size.

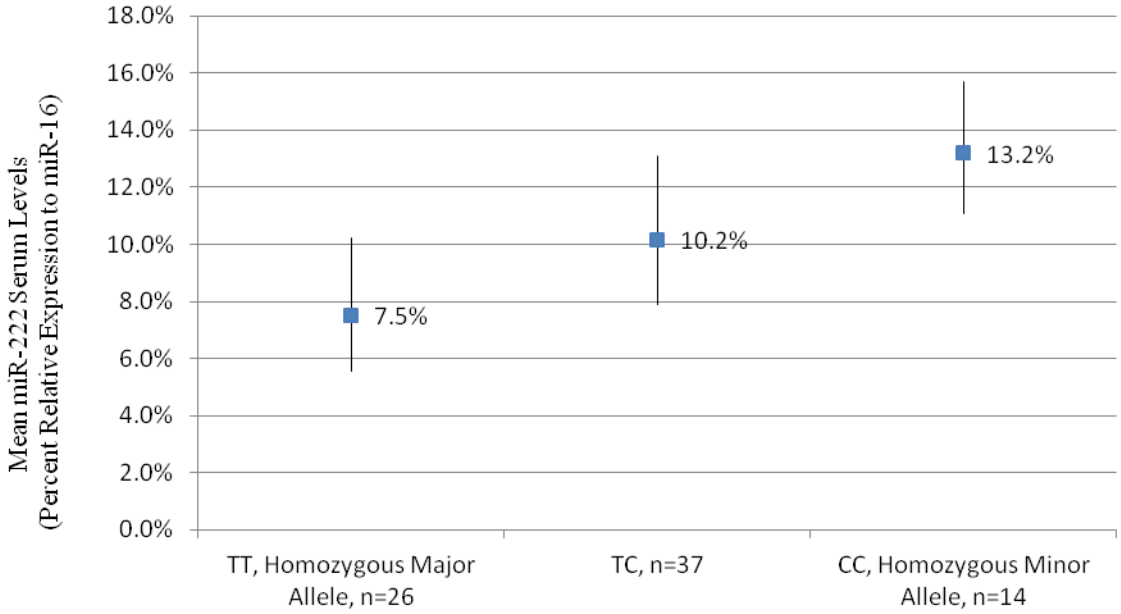
Figure 3.4 Mean miR-21 serum levels by Gemin3 rs197412



Mean ratios estimating the association between *GEMIN3 rs197412* and mean $\log_e$ miRNA-21 levels: $MR_{CC\ vs.\ TT} = 1.51\ (0.90\text{-}2.53)$; $MR_{CC+CT\ vs.\ TT} = 1.36\ (0.95\text{-}1.94)$.

Bars represent the 95% confidence limits of the means.

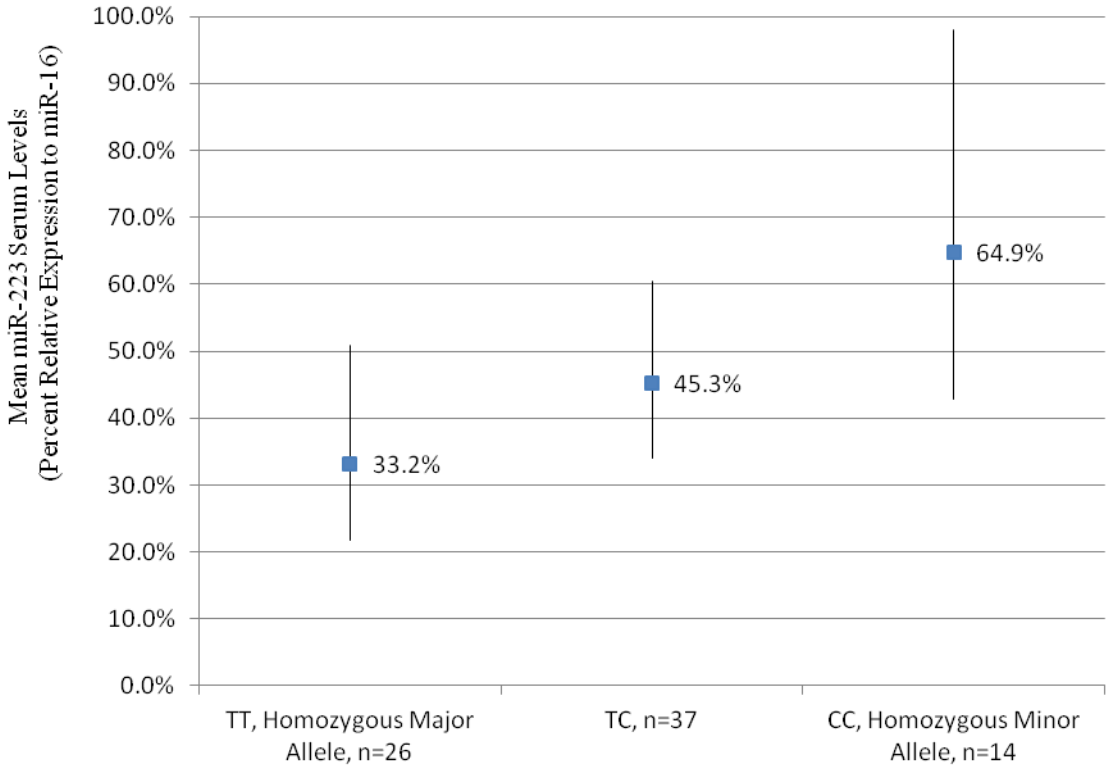
Figure 3.5 Mean miR-222 serum levels by Gemin3 rs197412



Mean ratios estimating the association between *GEMIN3 rs197412* and mean $\log_e$ miRNA-222 levels: $MR_{CC\ vs.\ TT} = 1.40\ (0.90\text{-}2.17)$; $MR_{CC+CT\ vs.\ TT} = 1.35\ (0.99\text{-}1.82)$.

Bars represent the 95% confidence limits of the means.

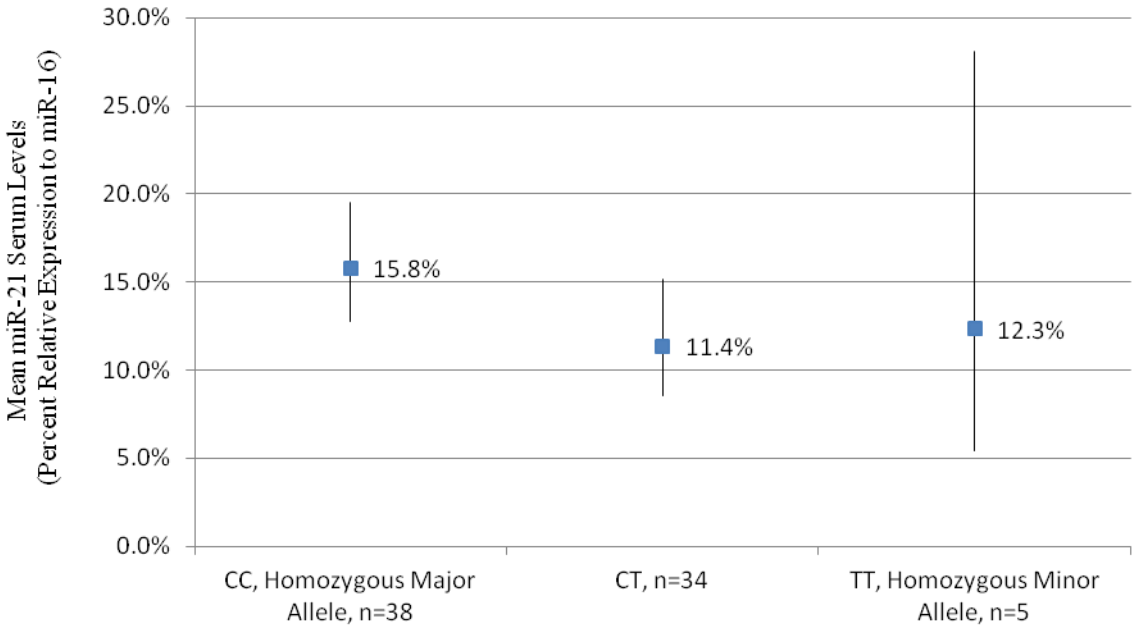
Figure 3.6 Mean miR-223 serum levels by Gemin3 rs197412



Mean ratios estimating the association between *GEMIN3 rs197412* and mean $\log_e$ miRNA-223 levels: $MR_{CC\ vs.\ TT}$= 1.72 (0.90-3.31); $MR_{CC+CT\ vs.\ TT}$= 1.39 (0.89-2.19).

Bars represent the 95% confidence limits of the means.

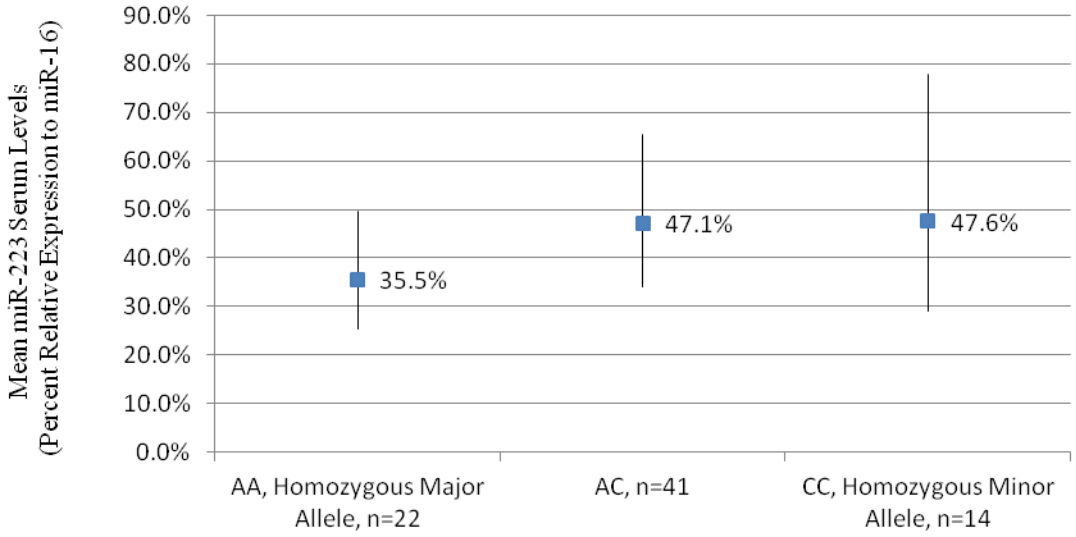
Figure 3.7 Mean miR-21 serum levels by RAN rs14035



Mean ratios estimating the association between *RANrs14035* and mean $\log_e$ miRNA-21 levels:
$MR_{TT\ vs.\ CC}$= 0.85 (0.43-1.68); $MR_{TT+CT\ vs.\ CC}$= 0.73 (0.0.53-1.02)

Bars represent the 95% confidence limits of the means.

Figure 3.8 Mean miR-223 serum levels by XPO5 rs11077



Mean ratios estimating the association between *XPO5* rs11077 and mean $\log_e$ miRNA-223 levels: $MR_{AC\ vs.\ AA} = 1.46\ (0.90\text{-}2.74)$; $MR_{CC+CA\ vs.\ AA} = 1.46\ (0.91\text{-}1.68)$.

Bars represent the 95% confidence limits of the means.

CHAPTER 4: DISCUSSION

4.1 Highlighted Result

GEMIN3 rs197412, a non-synonymous SNP resulting in a residue change of isoleucine to threonine within a RNA helicase gene involved in miRNA processing and maturation, was associated with AIDS-NHL overall (OR_{adj} = 1.35 per variant allele; 95% CI: 1.03-1.78). In addition, the variant genotype (CC) and allele (C) of *GEMIN3* rs197412 remained suggested as associated with higher mean log_e transformed miRNA-222 serum levels compared to those with the referent genotype MR_{adj} = 1.21 per variant allele; 95% CI: 0.98-1.49; CT/CC vs. TT: MR_{adj} = 1.35; 95% CI: 0.99-1.82). Additionally, this SNP was suggested as associated with higher mean log_e transformed miRNA-21 and miR-223 levels. These suggested associations are in support of growing literature demonstrating that germline variation within miRNA-processing genes affect miRNA maturation and subsequent levels of expression—levels which in turn may also contribute to tumorigenesis.[69] As non-synonymous coding SNPs have the ability to impart phenotypic change through the amino acid substitution, our interest regarding this SNP was heightened.[186] To investigate any potential consequence of this amino acid change the bioinformatic algorithm PolyPhen-2 was utilized, with this SNP categorized as "benign", indicating no gain-of- or loss-of-function predicted to be associated with this residue change.[187] However our results are still interesting as although no literature exists regarding this SNP and lymphoma, *GEMIN3* is a binding partner of Epstein-Barr nuclear antigen 2 (*EBNA2*) and Epstein-Barr nuclear antigen 3C (*EBNA3C*)—highlighting potential mechanisms through which *GEMIN3* may impact AIDS-NHL.[188] *GEMIN3* also participates in RNA transport, RNA metabolism and decay, ribosome biogenesis, and RNA translation initiation—all processes influenced by germline variation. These items considered in combination suggest that it is likely

that our observed associations between *GEMIN3* rs197412 and both AIDS-NHL and elevated miRNA serum levels, reflect the hypothesis that SNPs within miRNA processing machinery have the potential impact miRNA maturation and downstream functionality related to tumor susceptibility.

4.2. miRNA-Related SNPs and AIDS-NHL Susceptibility

In this dissertation, we analyzed the association between 22 miRNA-related SNPs and AIDS-NHL susceptibility. Our results indicate the importance of SNPs within miRNA coding regions and processing genes, and suggest that a few miRNA-related SNPs may be related to AIDS-NHL susceptibility.

The variant allele (T) of *microRNA-196a2* rs11614913 was inversely associated with AIDS-NHL overall, and CNS AIDS-NHL (AIDS-NHL Overall OR_{adj} = 0.80 per variant allele; 95% CI: 0.61-1.05; CNS AIDS-NHL TT/CT vs. CC: OR_{adj} = 0.43; 95% CI: 0.22-0.87 evident after SB analysis). A recent meta-analysis investigating *miR-196a* rs11614913 across 21 studies observed a decrease in overall cancer risk associated with the T allele compared to the C allele (T vs. C OR_{adj}: 0.89; 95% CI: 0.85-0.94)—an inverse association echoed across breast, liver, lung, gastric and colorectal cancer sites. No studies to our knowledge have investigated this SNP in relation to AIDS-NHL.[189, 190] *miR-196a* is composed of two mature miRNA sequences processed from the same stem-loop, with *microRNA-196a2* rs11614913 located within the precursor strand of what becomes the 3' passenger strand of the primary and mature sequence.[71] This location has been identified to alter target gene regulation and critically to interfere with the formation of the secondary stem-loop structure which results in less efficient miRNA processing and maturation from the precursor miRNA.[65] In fact, this SNP alters mature miR-196a levels *in*

vitro, in which the variant allele of the precursor sequence resulted in decreased levels of mature miR-196a compared to the levels produced by the wild-type precursor. Further, genome-wide expression microarray data demonstrated pre-miR-196a -T to result in fewer than half of the amount of altered transcripts compared to the pre-miR-196a –C (wild-type), indicating decreased regulatory capacity imparted by this SNP.[71] Given that this SNP decreases target gene regulation, lowers mature miR-196a levels, and is inversely associated with overall cancer development, the tumor suppressive potential of *miR-196a* rs11614913 to impact AIDS-NHL, as seen in our study, is biologically plausible.

This inverse association was most prominent among those with CNS AIDS-NHL, thus perhaps this SNP impacts AIDS-NHL via an EBV-related mechanism. In our study population, 89% of CNS AIDS-NHL tumors were EBV positive, reflecting the fact that nearly all CNS AIDS-related lymphomas are EBV^+.[191] EBV has the potential to induce somatic hypermutations (SHM) upon B-cell infection through inducing AID expression—a DNA mutation-creating enzyme of critical importance in B-cell activation— via the viral LMP-1 protein.[47] AID expression results in an accumulation of AID-related mutations, including SHM, that contribute to lymphoma development.[48] Interestingly, AID-induced SHM decreases p53 activity—activity found to protect B-cells from developing these mutations— and further, AID has been found to directly cause SHM of the *p53* gene.[47, 48] EBV-infection also induces miRNA expression and the EB virus is predicted to encode over twenty viral miRNAs.[47] Further, miRNAs interact heavily with the *p53* pathway, with miRNA expression contributing to *p53* regulation, and p53 expression contributing to miRNA regulation.[88, 192, 193] Thus, one proposed mechanism for our observation is that decreased mature miR-196a levels imparted by *miR-196a* rs11614913 may

promote p53 (wide-type protein) activity which in turn would decrease the accumulation of AID-induced SHM, and thus decrease AIDS-NHL susceptibility.

A weak, positive association was suggested between *microRNA-27a* rs895819 and overall AIDS-NHL in this study (AIDS-NHL Overall OR_{adj} = 1.31 per variant allele; 95% CI: 0.97-1.76). *miRNA-27a* is considered to have oncogenic potential as highlighted by *Tang et al.*, who reported miRNA-27a levels as considerably over-expressed in individuals with invasive breast cancer, with these levels negatively correlated with survival.[55] The variant allele (C) of *microRNA-27a* rs895819, located within the terminal pre-miRNA loop, may influence both mature miRNA function and the minimum free energy available for target gene interaction, as demonstrated in gastric cancer.[194] *Sun et al.* demonstrated increased miR-27a expression among those with the variant genotype (GG) of rs895819, and found that *mir-27a* rs895819 presence impacted target gene binding affinity, as reflected in decreased target gene (*ZBTB10*) mRNA levels among those with the variant genotype.[194] Further, decreased *ZBTB10* expression levels were inversely correlated with *mir-27a* expression in those with gastric cancer, again suggesting that this SNP influenced gene interaction. Although, studies investigating this SNP have experienced conflicting results, including a recent meta-analysis which concluded zero association with overall-cancer risk, this is the first study to investigate and suggest *mir-27a* rs895819 in relation to AIDS-NHL.[195] We believe that the suggested association observed in this study is reasonable given the biologic plausibility for this SNP to impair *mir-27a* expression and target gene interaction; however quantitative measurement of wild-type and variant expression in lymphoma cell lines with a larger sample size may better elucidate an AIDS-NHL-specific association.

We observed a suggested association between *microRNA-300* rs12894467 and AIDS-NHL overall among Caucasian-only participants (AIDS-NHL overall among Caucasians OR_{adj} = 1.25 per variant allele; 95% CI: 0.92-1.70), and among those with systemic AIDS-NHL under a recessive model (systemic AIDS-NHL TT vs. CT+CC: OR_{adj} = 1.67; 95% CI: 0.95-2.96). This miRNA coding region is associated with the stem cell pathway and has tumor suppressive properties through enhancing the self renewal of stem cells and participating in stem-like cell differentiation.[196] To our knowledge, we are the first to investigate this SNP in relation to AIDS-related NHL, with very little published regarding this miRNA coding region. Our study implies that this SNP located within the miR-300 coding region may influence the susceptibility of AIDS-NHL, especially among Caucasian individuals.

We observed a positive association between the variant genotype (CC) of *GEMIN3* rs197412 and overall AIDS-NHL (OR_{adj} = 1.35 per variant allele; 95% CI: 1.03-1.78; evident after SB correction). *GEMIN3* rs197412, a non-synonymous SNP within a RNA helicase gene involved in miRNA processing and maturation is inversely association with premalignant oral lesions.[65, 67, 78] To investigate any potential consequence of this missense mutation, the bioinformatic algorithm PolyPhen-2 was utilized, with this SNP categorized as “benign”, indicating no gain-of- or loss-of-function predicted to be associated with this residue change.[187] However, while no literature exists regarding this SNP and either lymphoma or AIDS-NHL, *GEMIN3* is a binding partner of Epstein-Barr nuclear antigen 2 (*EBNA2*) and Epstein-Barr nuclear antigen 3C (*EBNA3C*) —viral genes involved in EBV transformation of B lymphocytes and essential genes for EBV latency establishment.[188] *EBNA2* interacts with *IL18* on B-cells, and is associated with *IL-18* receptor expression—an expression already heightened due to HIV-1 or EBV infection. Of critical importance is the fact that *GEMIN3* binds directly to *p53*, acting as a

negative regulator of this tumor suppressor—demonstrated through blocking apoptotic function in EBV-transformed B-cell lymphoma cell lines.[188] Further, the presence of *EBNA3C* enhances the protein stability and production of GEMIN3, causing an accumulation of GEMIN3 species that blocks *p53* DNA-binding, again decreasing apoptosis and contributing to uncontrolled cellular proliferation. Last, *GEMIN3* knockdown results in the death of overall B-cell lymphoma cells, and in particular *EBNA3C* positive cells.[188] This point is critical, as it further suggests this gene participates in both EBV- and non-EBV-related AIDS-NHL mechanisms. Thus, one potential mechanism through which *GEMIN3* rs197412 may increase AIDS-NHL susceptibility is through negatively influencing *p53* activity and promoting uncontrolled cellular proliferation. This decreased *p53* activity would also increase AID activity, which additionally influences the cycle of accumulated mutations that impact lymphomagenesis.

Further, *GEMIN3* also participates in RNA transport, RNA metabolism and decay, ribosome biogenesis, and RNA translation initiation—all processes able to be influenced by germline variation. As *GEMIN3* rs197412 is involved miRNA biogenesis, a process highly dependent on sequence complementation, it is also likely that our observation reflects the hypothesis that SNPs within miRNA processing machinery do impact miRNA maturation and downstream functionality related to tumor susceptibility.[69]

The variant allele (G) of *RCHY1* rs2126852—estimated in our bioinformatic analysis to be near a binding site of brain-specific *miR-153*—was weakly and inversely associated with overall and systemic AIDS-NHL in this study (AIDS-NHL overall OR_{adj} = 0.81 per variant allele; 95% CI: 0.60-1.09; AIDS-NHL systemic OR_{adj} = 0.74 per variant allele; 95% CI: 0.52-1.06). *RCHY1*—an E3 ubiquitin ligase—promotes p53 degradation and impacts cell cycle regulation.[197, 198] Further, this gene interacts with *p53* to promote ubiquitination and thus results

in decreased *p53* expression. *Leng, et al*., demonstrated that through binding *p53* directly, *RCHY1* regulates *p53* homeostasis and is able to ubiquinate *p53* independent of other genes, such as *MDM2*.[197] Thus, this SNP may impact the interaction between *RCHY1* and *p53*, contributing to the suggested decrease in AIDS-NHL susceptibility observed in our study. As *p53* functions to protect B-cells from developing SHM mutations, if *RCHY1* rs2126852 does in fact promote *p53* activity, the suggested decrease in AIDS-NHL risk observed in our study is understandable. Additional investigation into the potential for *RCHY1* rs2126852 to impact *p53* expression is warranted. Further, this SNP's predicted location near a *miR-153* binding site may help elucidate any potential for this gene to influence AIDS-NHL. *miR-153* is described to have potential tumor suppressor power, as *miR-153* levels are significantly decreased in tumor cell lines derived from the central nervous system.[113] *Xu et al.,* demonstrated that in glioblastoma cell lines, *miR-153* directly targets *BCL-2* in order to decrease uncontrolled cellular proliferation and promote apoptosis.[113] Thus, functional studies investigating the biologic mechanisms between *RCHY1* rs2126852 and *miR-153* binding is also warranted, especially as our study suggests this germline variant may affect tumor susceptibility. Investigation of this SNP in a larger sample of CNS-AIDS NHL cases may also prove fruitful as only two CNS cases were currently under study with the variant genotype, limiting additional site-specific conclusions.

We observed a suggested association between the variant allele (C) of *GEMIN4* rs7813—a miRNA processing gene—and systemic AIDS-NHL (OR_{adj} = 1.27 per variant allele; 95% CI: 0.92-1.76). *GEMIN4* interacts with the RNA-induced silencing complex (RISC) to process mature miRNAs, and assists RISC in target RNA identification and repression.[199] Thus, there is potential for genetic variation within this gene to impact both miRNA maturation and target gene regulation through disrupting proper interaction with the RISC complex. Additionally, *GEMIN4*

rs7813 is a functional SNP located within the exon that may play a role in protein expression through resulting in an arginine to cysteine amino acid substitution, and is suggested to be associated with prostate, esophageal, ovarian and renal cell carcinomas.[199-201] To our knowledge we are the first to investigate this SNP in relation to AIDS-NHL. We hypothesize this SNP impacts miRNA processing such that the downstream effects influence AIDS-NHL susceptibility.

In our study, *HIF1A* rs2057482 was found to have opposing associations with systemic and CNS AIDS-NHL: *HIF1A* rs2057482 was positively associated with systemic AIDS-NHL, whereas among those with CNS AIDS-NHL, an inverse association was observed. These associations remained evident after SB analysis (systemic AIDS-NHL OR_{adj} = 1.83 per variant allele; 95% CI: 1.16-2.90; CNS AIDS-NHL OR_{adj} = 0.38 per variant allele; 95% CI: 0.17-0.87). To our knowledge, we are the first group to investigate this SNP in relation to AIDS-NHL—a SNP without substantial literature, and which per a recent meta-analysis does not seem to be associated with overall-cancer risk.[202] However, there is biologic plausibility for this SNP to be related to AIDS-NHL susceptibility under our hypotheses. Interestingly, the introduction of *HIF1A* rs2057482 creates novel miRNA biding sites for miR-196a-5p, miR-196b-5p and miR-921, implying that this SNP may be related to AIDS-NHL through a miRNA-related mechanism.[203] *HIF1*—critically involved in oxygen homeostasis—activates the expression of over 60 genes, including *BCL-XL*, and contributes to cell regulation, proliferation and survival.[204] Important for tumorigenesis, *HIF1A*—a catalytic subunit that controls *HIF1* activity through oxygen regulation—increases *p53* stability to promote apoptosis, with this interaction unobserved among carriers of the mutant *p53* gene.[205-207] Further, an accumulation of genetic alterations are induced by *HIF1A* over-expression in cancer cells, including lymphoma, while

low levels of *HIF1A* are found in normal cells, suggesting that *HIF1A* may provide selective advantages for the survival and promotion of cancer cells.[208-210] As, *HIF1A* is over-expressed in lymphoma cells, and this over-expression promotes glycolysis which results in an accumulation of reactive oxygen species (ROS), it is plausible that systemic AIDS-NHL susceptibility may be influenced by SNPs within *HIF1A* though such a mechanism.

In regards to the inverse association between *HIF1A* rs2057482 and CNS AIDS-NHL observed in our study, it is important to note that *HIF1A* protein levels are stabilized in lymphoblastoid cells (whereas levels are low in normal cells), even under conditions mimicking hypoxia.[211] In fact, this stabilization is made stronger through an EBV-related mechanism in which EBNA antigens bind to prolylhydroxylases to inhibit hydroxylation and the eventual degradation of *HIF1A.* This in turn, increases the cell's preference for aerobic glycolysis, resulting in an accumulation of reactive oxygen species (ROS).[212] It is important to note, that the ROS are elevated in lymphoblastoid cells, regardless of whether they have been infected with EBV. Thus, *HIF1A* levels are stabilized in lymphoblastoid due to inhibited hydroxylation, and this results in an accumulation of ROS (which innately promote tumorigenesis), indicating that high levels of *HIF1A* have the potential to promote lymphomagenesis. Thus, suggested potential mechanisms for our observation between *HIF1A* rs2057482 and CNS AIDS-NHL susceptibility include: a) this SNP may inhibit the interaction between EBNA antigens that promote *HIF1A* stability, thus promoting decreased *HIF1A* levels, or b) this SNP may enhance the ability of the cell to undergo aerobic glycolysis, thus resulting in the production of fewer ROS, and in turn decreasing AIDS-NHL susceptibility.

As *microRNA-196a2* binds to *HIF1A* in the presence of the *HIF1A* rs2057482 SNP, we investigated the joint effect of *miR-196a2* rs11614913 and *HIF1A* rs2057482 and overall,

systemic and CNS AIDS-NHLs. A stronger inverse association with CNS AIDS-NHL by the joint effect of these two SNPs was suggested, which was consistent with the direction both these SNPs worked in individually in relation to CNS AIDS-NHL (Joint effect: OR_{adj11} = 0.10 95% CI: 0.02-0.45; *HIF1A* rs2057482: OR_{adj10} =0.22 95% CI: 0.06-0.81; *miR-196a* rs11614913: OR_{adj01} =0.38 95% CI: 0.17-0.88). Despite limited sample size, this observation was consistent with our hypothesis, implying that miRNA-related SNPs, especially SNPs in miRNA coding regions and their target genes, have the potential to affect tumor susceptibility. As interaction on the additive scale was not suggested, the observed joint effect emphasizes a greater decrease in CNS AIDS-NHL among those who carry both SNPs, however biologic interaction may not be present. Validation on a larger scale is needed.

4.3. SNPs within miRNA processing genes and Differential miRNA Serum Levels

In this dissertation, we analyzed the association between 7 SNPs within genes involved in miRNA biogenesis and subsequent miRNA serum levels to investigate the association between genetic variation within miRNA processing machinery and subsequent miRNA dysfunction, as reflected by differential miRNA expression across genotype. Given that differential serum levels were suggested across specific genotypes, our results suggest that germline variation within these genes may be related to downstream miRNA expression; expression that may additionally influence AIDS-NHL susceptibility.

The variant genotype (CC) and allele (C) of *GEMIN3* rs197412 were suggested as associated with higher mean $\log_e$ transformed miRNA-21, miR-222 and miR-223 sera levels compared to wild-type individuals, while no associations with liver-specific miR-122 were detected (miR-21 mean ratio: MR_{adj} = 1.24 per variant allele; 95% CI: 0.97-1.59; miR222 mean

ratio: MR_{adj} = 1.21 per variant allele; 95% CI: 0.98-1.49; miR-223 mean ratio: MR_{adj} = 1.31 per variant allele; 95% CI: 0.96-1.78; miR-122 mean ratio: MR_{adj} = 1.29 per variant allele; 95% CI: 0.64-2.60). These suggested associations are in support of growing literature that germline variation within miRNA-processing genes affect miRNA maturation and subsequent levels of expression—levels which in turn may also contribute to tumorigenesis.[69] It is interesting that this SNP was not specific, but associated with three different miRNAs. This re-emphasizes the overall importance of *GEMIN3* rs197412 in miRNA biogenesis, and supports our hypothesis that genetic variation within miRNA processing genes may affect miRNA maturation to some degree, which is likely due to disrupted interaction during processing in the RISC complex. Our observation suggests that this SNP does interfere with miRNA biogenesis.

While miRNA-21 and miR-223 sera levels may be used to distinguish individuals without HIV-1 infection from HIV+ individuals given differential expression patterns, miR-222 does not make this distinction.[185] Rather, miR-222 is elevated in AIDS-NHL individuals compared to those uninfected with HIV-1, and further miR-222 over-expression is able to distinguish HIV+ individuals from those who go on to develop CNS or DLBC AIDS-NHL.[185] This indicates that miR-222 may be related to AIDS-NHL susceptibility over HIV progression. Thus, this increase in sera associated with the C allele of *GEMIN3* rs197412 suggests the possibility of this miRNA processing gene in HIV-related lymphomagenesis, and the potential for inherent genetic variation within miRNA-related genes to influence miRNA function, as reflected through differential expression downstream.

GEMIN3 rs197412 was the only miRNA-biogenesis SNP out of the seven suggested as associated with multiple miRNA sera. For example, plots of back-transformed geometric means visually suggested a pattern of increased sera levels that accompanied the variant allele of

GEMIN3 rs197412 (i.e. the *GEMIN3* rs197412 variant (CC) genotype exhibited miR-21 relative expression levels of 17.9% compared to 10.8% relative miR-21 expression in wild-type (TT) individuals; the *GEMIN3* rs197412 variant (CC) genotype exhibited miR-222 relative expression levels of 13.2% compared to 7.5% relative miR-222 expression in wild-type (TT) individuals; the *GEMIN3* rs197412 variant (CC) genotype exhibited miR-223 relative expression levels of 64.9% compared to 33.2% relative miR-223 expression in wild-type (TT) individuals).

The CT genotype of *RAN* rs14035 was suggested as associated with lower mean $\log_e$ transformed miRNA-21 serum levels (miR-21 mean ratio: $MR_{adj} = 0.81$ per variant allele; 95% CI: 0.62-1.06; CT+TT vs. CC $MR_{adj} = 0.73$; 95% CI: 0.53-1.02). This SNP is located within the *RAN* 3' UTR, indicating that on top of interfering with miRNA maturation, this SNP also has the potential to influence miRNA binding affinity. However the extent to which this occurs is still unclear.[65,213] An association was also suggested between the variant C allele of *XPO5* rs11077 and mean $\log_e$ transformed miRNA-223 levels compared to wild-type individuals (miR-223 mean ratio: log additive: $MR_{adj} = 1.24$; 95% CI: 0.91-1.68; AC+CC vs. AA: $MR_{adj} = 1.46$; 95% CI: 0.92-2.31). *XPO5* is critically involved in the transport of pre-miRNAs from the nucleus into the cytoplasm, and downregulation of *XPO5* results in decreased production of miRNA levels, reduced miRNA processing, and differential miRNA-target inhibition.[133] Further, our suggested observation is consistent with literature indicating *XPO5* mutations sequester pri-miRNA to result in lower miRNA levels.[69, 213] Thus, although marginal, our observations support evidence that demonstrates miRNA biogenesis-related SNPs may have the potential to influence overall, downstream miRNA expression.

4.4. Strengths and Limitations

A main strength of our study was the selection of AIDS-NHL cases and HIV-infected controls from an established, longitudinal prospective cohort with rich biologic data available for analysis. The ability for us to supplement our genotypic data with sera data enabled an even more comprehensive investigation of the biologic mechanisms at play. Additionally, detailed covariate data collected at multiple time points helped to investigate potential confounders, such as history of alcohol consumption, in a comprehensive manner. Last, the application of the semi-Bayesian approach to correct the issue of multiple comparisons in this study added confidence in our highlighted report of associations that remained evident after the application of this correction which pulled our estimates towards the null (*GEMIN3* rs197412, *miR-196a* rs11614913, and *HIF1A* rs2057482).

Although confident in our covariate selection and confounder control, there is the potential for residual confounding by both known and unknown confounders. Additionally, due to concerns regarding sparse data, we were limited in our ability to stratify by specific AIDS-NHL histological subtypes, or other covariates of potential interest. Furthermore, for the study of SNPs with a lower minor allele frequency, the possibility of observations arising due to chance could not be fully excluded.

Additionally, due to missing data on two of our most critical covariates—HIV viral load at set point pre-HAART and $CD4^+$ T-cell count at reference date—imputation was necessary when investigating the association between our miRNA-related SNPs of interest and AIDS-NHL susceptibility. Further, as HIV viral load at set point was missing on all participants enrolled in the second recruitment period, our result interpretation was limited towards those results based on our median-control value imputed data, which took into consideration participants from both

cohorts. In order to investigate underlying differences between participants in the complete case and imputed analyses, we compared the descriptive characteristics of individuals in both recruitment periods, and determined that individuals enrolled in cohort two differed significantly across several key covariates compared to men enrolled in cohort one. Thus, we believe the results from our imputed analysis are more representative of the entire population compared to the complete case analysis, as the underlying group components differ.

Additionally, we compared the robustness of our main results across different analytic techniques (complete case analyses, median imputed analyses, multiply imputed sensitivity analyses, a model not adjusted for HIV viral load at set point, and a model adjusted for matching factors only). A pattern suggesting an association between *GEMIN3* rs197412 and AIDS-NHL susceptibility was robust across these analytic schemes. A similar pattern suggesting an inverse association between *miR-196a* rs11614913 and overall AIDS-NHL susceptibility also remained robust across different analytic strategies. And, last a pattern of no association between *HIFIA* rs2057482 and overall AIDS-NHL was evident across analyses. The fact that these results remained suggested across analytic strategies increased our confidence in reporting that these polymorphisms may influence AIDS-NHL susceptibility in the setting of HIV/AIDS.

Our genotypic data came from whole genome amplified DNA extracted from B-cell pellets. There are two items to consider in regards to these samples. First, there are disadvantages of using amplified DNA compared to original DNA, as the process of amplifying DNA may introduce bias. Second, the DNA was extracted from B-cells which were immortalized by EBV infection, and thus over long periods of time and cell growth, EBV-driven somatic mutations may have accumulated in our samples.

As original, genomic DNA is extremely valuable and oftentimes is available in limited quantities, methods for the whole genome amplification of DNA have been developed.[214, 215] However, these methods are prone to specific biases, including preferential amplification and allele dropout.[216] Preferential amplification occurs when one allele fails to reach the appropriate threshold of detection. Several factors can account for preferential amplification, including the short allele product being amplified preferentially, differential denaturation due to guanine and cytosine base-richness differences (%GC-bias), nucleotide copy errors, mismatches between the primer and specific allelic template, and low annealing temperatures that may cause an increase or decrease in amplification of simple tandom repeats.[215, 216] Allele dropout is a random amplification failure that occurs when one of the two heterozygous alleles fails to amplify, while the other allele proceeds to do so. In the current study, perhaps the use of amplified DNA as opposed to original DNA contributed to the replicate discordance noted for one excluded SNP (*WWOX* rs12828). These potential errors in DNA amplification are caveats that we as investigators must take into account when using whole genome amplified DNA as opposed to the original artifact. However, as our study allele and genotype frequencies were similar to those previously reported in dbSNP and HapMap, we are confident that SNPs in this study were not affected by these potential errors in whole genome DNA amplification.

In regards to the fact that our samples used DNA that was extracted from B-cells which had been immortalized by EBV, there are a few items to take into consideration. In 2007 *Epeldegui et al.*, demonstrated that EBV-infection of B-cells resulted in AID expression approximately ten days post-infection. Additionally, EBV infection of these B-cells also resulted in expression of polymerase-η within 2-10 days of infection. AID expression is critically involved in Ig gene class switch recombination (CSR) and somatic hypermutation (SHM), while

the DNA error-prone polymerase-η is involved in SHM. Thus as, EBV infection moves B-cells towards a semi-activated state, in which Ig gene class switch recombination is promoted, one concern may be bias in genotyping if CSR has occurred. However, this potential bias would be subject to mutations located in that specific region of recombination. A more serious concern would be polymerase-η- or AID expression-induced somatic hypermutations, such as mutations of *p53* and other genes, including perhaps those in the present study, that may have accumulated over time, especially in the setting of immortalized B-cells where proliferation is promoted.[217] In the 2007 study, polymerase-η- and AID expression in EBV-infected B-cells resulted in increased mutations within the *p53*, *BCL6*, and *beta-globin* genes (the majority of these being C-T or G-A transitions). The reported frequency of these mutations was approximately 24.4 mutations per 10,000 basepairs in *BCL6*, 5.5 mutations per 10,000 basepairs in *p53*, and 4.7 mutations per 10,000 basepairs in *beta-globin*.[217] The fact that mutations were seen to arise in the non-oncogenic *beta-globin* suggests the far reaching potential of polymerase-η- or AID-induced mutations in EBV-immortalized B-cells and is noteworthy. However, given the somewhat low frequency and wide-spread of these mutations (24.4/10,000bp and 5.5/10,000bp, *BCL6* and *p53*, respectively), in conjunction with the fact that other studies using DNA extracted from this same source have not noted substantial deviations in expected genotype frequencies, reasonably accurate genotyping from the extracted DNA was probably achieved.

One item to keep in mind is that residual confounding may still be present when investigating the relationship between miRNA-processing SNPs and serum levels, especially as these mechanisms are not fully elucidated as of yet. There is the opportunity for these serum levels to be altered by HIV disease progression or other unknown confounders. Further, as these sera samples were taken on average at 8.8 months prior to AIDS-NHL diagnosis, it is difficult to

attribute the elevated sera levels to miRNA-related genetic variation exclusively. However, given our adjustment of case-control status and lack of suggested heterogeneity between cases and controls, we believe that our observations are able to be interpreted and add a point of interest to our study. Given i) specific circulating biomarker levels have been detected up to thirteen years prior to AIDS-NHL diagnoses[46, 185]; ii) there is not an established temporal threshold to mark when miRNAs first become elevated prior to disease onset, especially as it is more than likely the accumulation of deregulated miRNA function that contributes to disease development over time; and iii) our study rationale that germline variation influences disease susceptibility in part through altering downstream biologic function, it is reasonable to think that these elevated serum level miRNAs may not be caused by undiagnosed disease, but rather reflect the fact that SNPs in miRNA biogenesis machinery have long-term, wide reaching effects on downstream biologic processes (such as miRNA over-expression) which in turn may additionally increase tumor susceptibility. It is also critical to remember that SNPs are involved in complex biologic networks that run longitudinally. We are only able to see brief cross-sectional slices of the longitudinal implications imparted by genetic variation, and thus are limited in our conclusions.

CHAPTER 5: CONCLUSION AND PUBLIC HEALTH IMPLICATIONS

In this dissertation, we observed that a few miRNA-related SNPs, including *GEMIN3* rs197412, *miR-196a* rs11614913, and *HIF1A* rs2057482, were associated with AIDS-NHL susceptibility. We further highlighted the importance of miRNA processing genes such as *GEMIN3*, and suggest that SNPs within these genes may influence miRNA maturation and downstream functionality as reflected in differential sera levels across genotypes. Specifically, our study found that *GEMIN3* rs197412 was associated with AIDS-NHL susceptibility, and that this SNP was associated with higher mean miR-21, miR-222, and miR-223 serum levels. Given that higher mean miR-222 levels are suggested to result in poorer patient prognosis and survival of AIDS-NHL, it is of importance to explore the function of miRNA-related genes in AIDS-NHL development, and the potential for this SNP, as well as others like it, to act as biomarkers of susceptibility. The fact that *GEMIN3* rs197412 was suggested to be associated with three of our four mean miRNA sera levels further supports our hypothesis that germline variation within miRNA biogenesis genes may have the ability to influence downstream serum miRNA expression. In summary, we observed that *GEMIN3* rs197412 is associated with AIDS-NHL susceptibility, and that this SNP may interrupt miRNA processing in such a manner to impart a functional change in miRNA serum levels.

Further, *miR-196a* rs11614913 and *HIF1A* rs2057482 emerged as potential biomarkers of AIDS-NHL, highlighting the diverse array of biologic mechanisms underlying AIDS-lymphomagenesis. Additionally, on a small scale we were able to investigate the joint effects of these two miRNA-related SNPs (*HIF1A* rs2057482 and *microRNA-196* rs11614913). As *HIF1A*

rs2057482 creates a *microRNA-196a* binding site, it was interesting to observe the joint decrease in CNS AIDS-NHL susceptibility suggested by these SNPs.

As the processes of miRNA biogenesis, regulation and target determination are all critically dependent on sequence complementation, SNPs located within these regions have the ability to alter normal miRNA function through interrupting or impairing sequence interaction. These miRNA-related SNPs contribute to cancer susceptibility, and are on the rise as markers of disease prognosis and treatment responsiveness. Identification of miRNA-related SNPs that increase cancer risk is relevant to the development of screening approaches, including genetic profiling. Further, identification of miRNA-related SNPs associated with poorer survival or treatment response, disease subtypes, or patho-clinical features have implications in prognosis and treatment approaches. AIDS-related NHL is one type of cancer influenced by miRNAs and miRNA-related SNPs. Further, AIDS-related NHL is one cancer in particular that could benefit from further elucidation of underlying miRNA-related mechanisms, especially as despite the widespread use of HAART treatment, individuals with HIV-1 infection still experience a nearly 50 times increase in NHL risk compared to that found in the general population. In this dissertation, we found the potential for miRNA-related SNPs to contribute to the susceptibility of AIDS-NHL, while highlighting the extent of biologic interplay and networking-reach of these highly diverse, biologic mini-machines.

CHAPTER 6: FUTURE RESEARCH

Overall, establishing the functionality of miRNA-related SNPs through quantifying gene expression will help clarify the roles of SNPs pertinent to AIDS-NHL. To begin, experimental validation of miRNA-related SNPs identified in this study as near or within putative miRNA binding sites through the use of a reporter assay system or qPCR is critical. Second, it would be prudent to investigate the functionality and impact on gene expression of SNPs suggested as associated with HIV/AIDS in order to help elucidate the underlying biologic mechanisms. For example, it would be important to assess whether *GEMIN3* rs197412 increases or decreases *GEMIN3* expression through comparing mRNA levels between wild-type and variant construct systems, especially as although this is a non-synonymous SNP, it was categorized as "benign" by the functional effect algorithm PolyPhen-2. Next, it would be prudent to develop a system investigating the impact of this SNP on mature miRNA products to determine if this SNP makes a functional impact on miRNA processing. As *GEMIN3* rs197412 stood apart from other miRNA biogenesis machinery-related functional SNPs investigated in this study, these analyses may help flesh out what made this SNP so unique compared to the others under investigation. Deep-sequencing of this genomic region may also help to elucidate whether *GEMIN3* rs197412 is actually the SNP imparting the effect, or if another SNP in the region is responsible for the observation. Additionally, it would be crucial to further investigate *GEMIN3* rs197412 across other cancers to see if the same patterns arise and to further disentangle the role of *GEMIN3* rs197412 in tumorigenesis. Functional assessment of *HIF1A* rs2057482 would also be relevant, in addition to investigation of this SNP with finer stratification by histological sub-type, to further illuminate the site-specific patterns noted in this study.

One additional avenue of research involves quantifying the effect of SNPs within miRNA coding regions or processing machinery on subsequent serum miRNA expression profiles, as well as on other tumor suppressor/oncogene gene expressions. For example, there is limited evidence for the extent to which functional miRNA SNPs affect *p53* protein levels and related biological mechanisms. Expanding the scope of investigation will help us further understand how miRNA and miRNA-related SNP networks influence tumor development.

In regards to expanding the serum analyses in this study, although natural log transformation improved the normality of the data, investigation under non-parametric analytic methods may also be fruitful.[218-220] One option would be to use a Kruskal-Wallis test for non-parametric data, in which one nominal variable with greater than two levels (such as that of three-level genotypes: *aa ab bb*) and one measurement variable that does not meet the assumption of normality are investigated.[221] The null hypothesis of the Kruskal-Wallis test is not that two populations have identical means, but rather that the samples come from populations where the probability of a random observation from one group being greater than that in the second group is equal to 0.50. Under this modeling scheme, each measurement value is replaced with a rank score. The sum of the ranks is then calculated for each group, followed by calculation of the approximately chi-squared distributed test statistic, H and subsequent p-value.[221] If the results from a Kruskal-Wallis test were to be significant in our study investigating miRNA-related SNPs and miRNA serum associations, the interpretation would be that the mean ranks of serum levels were different significantly across genotype level. One consideration may be limited sample size. If due to limited sample size cell counts (n less than 5) it is appropriate to collapse the genotypic data such that we are only investigating a two-level nominal variable, the Mann-Whitney U test may be appropriate.[222, 223]

APPENDIX

Example semi-Bayesian code for multiple comparison correction in a matched case-control study investigating the association between miRNA-related SNPs and AIDS-NHL

```
/*restructuring of original dataset*/
data  MACSSB;
set  MACSSB;

/*dummy variable recoding for covariates of interest*/
if prev_aids_r0824 ne . then do;
if prev_aids_r0824= 1 then prev_aids_r0824sb1=1;
else prev_aids_r0824sb1=0;
end;
if everhcv ne . then do;
if everhcv= 1 then everhcvsb1=1;
else everhcvsb1=0;
end;
if artuse ne . then do;
if artuse= 1 then artusesb1=1;
else artusesb1=0;
end;
if racenew ne . then do;
if racenew=2 then racenewsb2=1;
else racenewsb2=0;
if racenew=3 then racenewsb3=1;
else racenewsb3=0;
end;

/*dummy variable recoding for polymorphisms*/
if genSNP1974t ne . then do;
if genSNP1974t= 2 then gen1974sb2=1;
else gen1974sb2=0;
if genSNP1974t=3 then gen1974sb3=1;
else gen1974sb3=0;
end;
if genSNP1974dom ne . then do;
if genSNP1974dom= 2 then gen1974domsb2=1;
else gen1974domsb2=0;
end;
if genSNP1974res ne . then do;
if genSNP1974res= 2 then gen1974ressb2=1;
else gen1974ressb2=0;
end;
if genSNP2126t ne . then do;
```

```
if genSNP2126t= 2 then gen2126sb2=1;
else gen2126sb2=0;
if genSNP2126t=3 then gen2126sb3=1;
else gen2126sb3=0;
end;
if genSNP2126dom ne . then do;
if genSNP2126dom= 2 then gen2126domsb2=1;
else gen2126domsb2=0;
end;
if genSNP2126res ne . then do;
if genSNP2126res= 2 then gen2126ressb2=1;
else gen2126ressb2=0;
end;
if genSNP1161t ne . then do;
if genSNP1161t= 2 then gen1161sb2=1;
else gen1161sb2=0;
if genSNP1161t=3 then gen1161sb3=1;
else gen1161sb3=0;
end;
if genSNP1161dom ne . then do;
if genSNP1161dom= 2 then gen1161domsb2=1;
else gen1161domsb2=0;
end;
if genSNP1161res ne . then do;
if genSNP1161res= 2 then gen1161ressb2=1;
else gen1161ressb2=0;
end;
if genSNP20574t ne . then do;
if genSNP20574t= 2 then gen20574sb2=1;
else gen20574sb2=0;
if genSNP20574t=3 then gen20574sb3=1;
else gen20574sb3=0;
end;
if genSNP20574dom ne . then do;
if genSNP20574dom= 2 then gen20574domsb2=1;
else gen20574domsb2=0;
end;
if genSNP20574res ne . then do;
if genSNP20574res= 2 then gen20574ressb2=1;
else gen20574ressb2=0;
end;
if genSNP1289t ne . then do;
if genSNP1289t= 2 then gen1289sb2=1;
else gen1289sb2=0;
if genSNP1289t=3 then gen1289sb3=1;
else gen1289sb3=0;
```

```
end;
if genSNP1289dom ne . then do;
if genSNP1289dom= 2 then gen1289domsb2=1;
else gen1289domsb2=0;
end;
if genSNP1289res ne . then do;
if genSNP1289res= 2 then gen1289ressb2=1;
else gen1289ressb2=0;
end;
run;

/*re-scaling and re-centering of pertinent covariates*/
data  MACSSB;
set  MACSSB;
if age_r0824 ne . then do;
ager0824sb=(age_r0824-40/10);
end;
if cd4_match ne . then do;
cd4matchsb= (cd4_match/100);
end;
if cd4atref ne . then do;
cd4atrefsb= (cd4atref/100);
end;
run;

/*adding prior indicators*/
data MACSSB;
set MACSSB;
A=1;
time = 0;
Run;

/*Quality control check to ensure that we obtain same odds ratios and 95% CI from original data
as done under the conditional logistic model*/
          PROC PHREG DATA=MACSSB;
  MODEL time*r0824_case(0) = gen1974sb2 gen1974sb3 ager0824sb rnasetlg
prev_aids_r0824sb1 artusesb1 cd4atrefsb racenewsb2 racenewsb3 everhcvsb1
                  /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*SNP1974 log additive cancer*/
PROC PHREG DATA=MACSSB;
  MODEL time*r0824_case(0) = genSNP1974t ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb racenewsb2 racenewsb3  everhcvsb1
                  /RL;
```

```
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*SNP1974 dominant cancer*/
PROC PHREG DATA=MACSSB;
  MODEL time*r0824_case(0) = gen1974domsb2 ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb racenewsb2 racenewsb3  everhcvsb1
                  /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*SNP1974 recessive cancer*/
PROC PHREG DATA=MACSSB;
  MODEL time*r0824_case(0) = gen1974ressb2 ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb racenewsb2 racenewsb3  everhcvsb1
                  /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*ready to build prior dataset and begin data augmentation*/
/*MAIN-EFFECT MODEL: coefficient for SNP (GEMIN3 rs197412) to AIDS-NHL risk*/
/*Prior specification gen1974sb2 gen1974sb3*/
DATA PRIORd;
  INPUT  r0824_set time r0824_case gen1974sb2 gen1974sb3 A ager0824sb rnasetlg
prev_aids_r0824sb1 artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3   ;
  DATALINES;
  201  1  1  0.1 0    400  0  0  0  0  0  0  0  0
  201  1  0  0   0    400  0  0  0  0  0  0  0  0
  202  1  1  0   0    400  0  0  0  0  0  0  0  0
  202  1  0  0.1 0    400  0  0  0  0  0  0  0  0
  203  1  1  0   0.1  400  0  0  0  0  0  0  0  0
  203  1  0  0   0    400  0  0  0  0  0  0  0  0
  204  1  1  0   0    400  0  0  0  0  0  0  0  0
  204  1  0  0   0.1  400  0  0  0  0  0  0  0  0
;
RUN;
/*Shrinkage estimation to check prior, should have HR=1.00 95% CI: 0.25-4.00*/
PROC PHREG DATA= PRIORd;
  MODEL time*r0824_case(0) = gen1974sb2 gen1974sb3 ager0824sb rnasetlg
prev_aids_r0824sb1 artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                  /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*Merge original and prior datasets*/
DATA COMBINEDd;
```

```
  SET PRIORd MACSSB;
RUN;
/*Shrinkage estimation to produce posterior estimates*/
PROC PHREG DATA= COMBINEDd;
  MODEL time*r0824_case(0) = gen1974sb2 gen1974sb3 ager0824sb rnasetlg
prev_aids_r0824sb1 artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                    /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;

/*Prior specification genSNP1974t*/
DATA PRIORd;
  INPUT  r0824_set time r0824_case genSNP1974t A ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3   ;
  DATALINES;
  201  1  1  0.1  400  0  0  0  0  0  0 0 0 0
  201  1  0  0    400  0  0  0  0  0  0 0 0 0
  202  1  1  0    400  0  0  0  0  0  0 0 0 0
  202  1  0  0.1  400  0  0  0  0  0  0 0 0 0
;
RUN;
/*Shrinkage estimation to check prior, should have HR=1.00 95% CI: 0.25-4.00*/
PROC PHREG DATA= PRIORd;
  MODEL time*r0824_case(0) = genSNP1974t ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                    /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*Merge original and prior datasets*/
DATA COMBINEDd;
  SET PRIORd MACSSB;
RUN;
/*Shrinkage estimation to produce posterior estimates*/
PROC PHREG DATA= COMBINEDd;
  MODEL time*r0824_case(0) = genSNP1974t ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                    /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;

/*Prior specification gen1974domsb2*/
DATA PRIORd;
```

```
 INPUT  r0824_set time r0824_case gen1974domsb2 A ager0824sb rnasetlg
prev_aids_r0824sb1 artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3   ;
 DATALINES;
 201  1  1  0.1  400   0  0  0  0  0  0  0 0 0
 201  1  0  0    400   0  0  0  0  0  0  0 0 0
 202  1  1  0    400  0  0  0  0  0  0  0 0  0
 202  1  0  0.1  400  0  0  0  0  0  0  0 0  0
;
RUN;
/*Shrinkage estimation to check prior, should have HR=1.00 95% CI: 0.25-4.00*/
PROC PHREG DATA= PRIORd;
 MODEL time*r0824_case(0) = gen1974domsb2 ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                  /RL;
 STRATA r0824_set;
 WEIGHT A;
RUN;
/*Merge original and prior datasets*/
DATA COMBINEDd;
 SET PRIORd MACSSB;
RUN;
/*Shrinkage estimation to produce posterior estimates*/
PROC PHREG DATA= COMBINEDd;
 MODEL time*r0824_case(0) = gen1974domsb2 ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                  /RL;
 STRATA r0824_set;
 WEIGHT A;
RUN;


/*Prior specification gen1974ressb2*/
DATA PRIORd;
 INPUT  r0824_set time r0824_case gen1974ressb2 A ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3   ;
 DATALINES;
 201  1  1  0.1  400  0  0  0  0  0  0 0 0 0
 201  1  0  0    400  0  0  0  0  0  0 0 0 0
 202  1  1  0    400  0  0  0  0  0  0 0 0 0
 202  1  0  0.1  400  0  0  0  0  0  0 0 0 0
;
RUN;
/*Shrinkage estimation to check prior, should have HR=1.00 95% CI: 0.25-4.00*/
PROC PHREG DATA= PRIORd;
 MODEL time*r0824_case(0) = gen1974ressb2 ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
```

```
                /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
/*Merge original and prior datasets*/
DATA COMBINEDd;
  SET PRIORd MACSSB;
RUN;
/*Shrinkage estimation to produce posterior estimates*/
PROC PHREG DATA= COMBINEDd;
  MODEL time*r0824_case(0) = gen1974ressb2 ager0824sb rnasetlg prev_aids_r0824sb1
artusesb1 cd4atrefsb everhcvsb1 racenewsb2 racenewsb3
                /RL;
  STRATA r0824_set;
  WEIGHT A;
RUN;
```

REFERENCES

1. Cheson BD, Leonard JP. Monoclonal Antibody Therapy for B-Cell Non-Hodgkin's Lymphoma. New England Journal of Medicine. 2008; 359(6): 613-26.

2. Morton LM Turner JJ, Cerhan JR, et al. Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). 2007; 110(2): 695-708.

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

4. Raphael M, Borisch B, Jaffe ES. Lymphomas associated with infection by the human immune deficiency virus (HIV). World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. 2001; IARC Press, Lyon: 260-3.

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

6. Rubinstein PG, Aboulafia DM, Zloza A. Malignancies in HIV/AIDS: from epidemiology to therapeutic challenges. Aids. 2014; 28(4): 453-65.

7. Carbone A, Gloghini A. AIDS-related lymphomas: from pathogenesis to pathology. British Journal of Haematology. 2005; 130(5): 662-70.

8. Müller AMS, Ihorst G, Mertelsmann R, Engelhardt M. Epidemiology of non-Hodgkin's lymphoma (NHL): trends, geographic distribution, and etiology. Annals of Hematology. 2005; 84(1): 1-12.

9. American Cancer Society. Cancer Facts & Figures 2014. Atlanta: American Cancer Society; 2014.

10. Ekström-Smedby K. Epidemiology and etiology of non-Hodgkin lymphoma – a review. Acta Oncologica. 2006; 45(3): 258-71.

11. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr, accessed on 07/14/2014

12. Howlader N, Noone A, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2011. National Cancer Institute Bethesda, MD, http://seercancergov/csr/1975_2011/, based on November 2013 SEER data submission. 2014.

13. Groves FD, Linet MS, Travis LB, Devesa SS. Cancer Surveillance Series: Non-Hodgkin's Lymphoma Incidence by Histologic Subtype in the United States From 1978 Through 1995. Journal of the National Cancer Institute. 2000; 92(15): 1240-51.

14. Alexander DD, Mink PJ, Adami H-O, Chang ET, Cole P, Mandel JS, et al. The non-Hodgkin lymphomas: A review of the epidemiologic literature. International Journal of Cancer. 2007; 120(S12): 1-39.

15. Engels EA, Biggar RJ, Hall HI, Cross H, Crutchfield A, Finch JL, et al. Cancer risk in people infected with human immunodeficiency virus in the United States. International Journal of Cancer. 2008; 123(1): 187-94.

16. Mbulaiteye SM, Parkin D, Rabkin CS. Epidemiology of AIDS-related malignancies an international perspective. Hematol Oncol Clin North Am. 2003; 17(3): 673-96.

17. Shiels MS, Engels EA, Linet MS, Clarke CA, Li J, Hall HI, et al. The Epidemic of Non–Hodgkin Lymphoma in the United States: Disentangling the Effect of HIV, 1992–2009. Cancer Epidemiology Biomarkers & Prevention. 2013; 22(6): 1069-78.

18. Vishnu P, Aboulafia DM. AIDS-Related Non-Hodgkin's Lymphoma in the Era of Highly Active Antiretroviral Therapy. Advances in Hematology. vol. 2012; 2012: (article ID 485943).

19. Ezzat H, Filipenko D, Vickars L, Galbraith P, Li C, Murphy K, et al. Improved Survival in HIV-Associated Diffuse Large B-Cell Lymphoma with the Addition of Rituximab to Chemotherapy in Patients Receiving Highly Active Antiretroviral Therapy. HIV Clinical Trials. 2007; 8(3): 132-44.

20. Kaplan LD. HIV-associated lymphoma. Best Practice & Research Clinical Haematology. 2012; 25(1): 101-17.

21. Seaberg EC, Wiley D, Martínez-Maza O, Chmiel JS, Kingsley L, Tang Y, et al. Cancer incidence in the multicenter aids cohort study before and during the HAART era. Cancer. 2010; 116(23): 5507-16.

22. Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. The Lancet. 2007; 370(9581): 59-67.

23. Goedert JJ, Bower M. Impact of highly effective antiretroviral therapy on the risk for Hodgkin lymphoma among people with human immunodeficiency virus infection. Current Opinion in Oncology. 2012; 24(5): 531-6 10.1097/CCO.0b013e3283560697.

24. 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.

25. Schroeder JC, Olshan AF, Baric R, Dent GA, Weinberg CR, Yount B, et al. A case–control study of tobacco use and other non-occupational risk factors for t(14;18) subtypes of non-Hodgkin's lymphoma (United States). Cancer Causes and Control. 2002; 13(2): 159-68.

26. Lu Y, Wang SS, Reynolds P, Chang ET, Ma H, Sullivan-Halley J, et al. Cigarette Smoking, Passive Smoking, and Non-Hodgkin Lymphoma Risk: Evidence From the California Teachers Study. American Journal of Epidemiology. 2011; 174(5): 563-73.

27. Morton L, P Hartge, et al. Cigarette Smoking and Risk of Non-Hodgkin Lymphoma: A Pooled Analysis from the International Lymphoma Epidemiology Consortium (InterLymph). Cigarette Smoking and Risk of Non-Hodgkin Lymphoma: A Pooled Analysis from the International Lymphoma Epidemiology Consortium (InterLymph). 2005; 14(4): 925-33.

28. Chang ET, Clarke CA, Canchola AJ, Lu Y, Wang SS, Ursin G, et al. Alcohol Consumption Over Time and Risk of Lymphoid Malignancies in the California Teachers Study Cohort. American Journal of Epidemiology. 2010; 172(12): 1373-83.

29. Hartge P, Smith MT. Environmental and Behavioral Factors and the Risk of Non-Hodgkin Lymphoma. Cancer Epidemiology Biomarkers & Prevention. 2007; 16(3): 367-8.

30. Fabbro-Peray P, Daures J-P, Rossi J-F. Environmental Risk Factors for non-Hodgkin's Lymphoma: A Population-based Case–Control Study in Languedoc-Roussillon, France. Cancer Causes and Control. 2001; 12(3): 201-12.

31. Richardson DB, Terschüren C, Hoffmann W. Occupational risk factors for non-Hodgkin's lymphoma: A population-based case-control study in Northern Germany. American Journal of Industrial Medicine. 2008; 51(4): 258-68.

32. Hemminki K, Li X, Czene K. Familial risk of cancer: Data for clinical counseling and cancer genetics. International Journal of Cancer. 2004; 108(1): 109-14.

33. Breen EC, Boscardin WJ, Detels R, Jacobson LP, Smith MW, O'Brien SJ, 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. Clinical Immunology. 2003; 109(2): 119-29.

34. O'Brien SJ, Nelson GW. Human genes that limit AIDS. Nat Genet. 2004; 36(6): 565-74.

35. Alcoceba M, Sebastian E, Marin L, Balanzategui A, Sarasquete ME, Chillon MC, et al. HLA specificities are related to development and prognosis of diffuse large B-cell lymphoma. Blood. 2013; 122(8): 1448-54.

36. Wang SS, Abdou AM, Morton LM, Thomas R, Cerhan JR, Gao X, et al. Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. Blood. 2010; 115(23): 4820-3.

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

38. Aissani B, Ogwaro KM, Shrestha S, Tang J, Breen EC, Wong HL, et al. The major histocompatibility complex conserved extended haplotype 8.1 in AIDS-related non-Hodgkin lymphoma. Journal of acquired immune deficiency syndromes (1999). 2009; 52(2): 170-9.

39. Wang SS, Cerhan JR, Hartge P, Davis S, Cozen W, Severson RK, et al. Common Genetic Variants in Proinflammatory and Other Immunoregulatory Genes and Risk for Non-Hodgkin Lymphoma. Cancer Research. 2006; 66(19): 9771-80.

40. Vendrame E, Hussain SK, Breen EC, Magpantay LI, Widney DP, Jacobson LP, et al. Serum levels of cytokines and biomarkers for inflammation and immune activation, and HIV-associated non-Hodgkin B-cell lymphoma risk. Cancer Epidemiol Biomarkers Prev. 2014; 23(2): 343-9.

41. Naicker DD, Wang B, Losina E, Zupkosky J, Bryan S, Reddy S, et al. Association of IL-10-promoter genetic variants with the rate of CD4 T-cell loss, IL-10 plasma levels, and breadth of cytotoxic T-cell lymphocyte response during chronic HIV-1 infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2012; 54(2): 294-302.

42. Lenz G, Wright GW, Emre NCT, Kohlhammer H, Dave SS, Davis RE, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proceedings of the National Academy of Sciences. 2008; 105(36): 13520-5.

43. Gaidano G, Pastore C, Volpe G. Molecular pathogenesis of non-Hodgkin lymphoma: a clinical perspective. Haematologica. 1995; 80(5): 454-72.

44. Ballerini P, Gaidano G, Gong J, Tassi V, Saglio G, Knowles D, et al. Multiple genetic lesions in acquired immunodeficiency syndrome-related non-Hodgkin's lymphoma. Blood. 1993; 81(1): 166-76.

45. Grulich AE, Vajdic CM, Cozen W. Altered Immunity as a Risk Factor for Non-Hodgkin Lymphoma. Cancer Epidemiology Biomarkers & Prevention. 2007; 16(3): 405-8.

46. Hussain SK, Hessol NA, Levine AM, Breen EC, Anastos K, Cohen M, et al. Serum Biomarkers of Immune Activation and Subsequent Risk of Non-Hodgkin B-Cell Lymphoma among HIV-Infected Women. Cancer Epidemiology Biomarkers & Prevention. 2013; 22(11): 2084-93.

47. Epeldegui M, Vendrame E, Martínez-Maza O. HIV-associated immune dysfunction and viral infection: role in the pathogenesis of AIDS-related lymphoma. Immunol Res. 2010; 48(1-3): 72-83.

48. Epeldegui M, Widney DP, Martínez-Maza O. Pathogenesis of AIDS lymphoma: role of oncogenic viruses and B cell activation-associated molecular lesions. Current Opinion in Oncology. 2006; 18(5): 444-8 10.1097/01.cco.0000239882.23839.e5.

49. Engels EA, Pfeiffer RM, Landgren OM, Moore RD. Immunologic and Virologic Predictors of AIDS-Related Non-Hodgkin Lymphoma in the Highly Active Antiretroviral Therapy Era. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2010; 54(1): 78-84.

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

51. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nature Genetics. 2005; 37(7): 766-70.

52. Blenkiron C, Miska EA. miRNAs in cancer: approaches, aetiology, diagnostics and therapy. Human Molecular Genetics. 2007; 16(R1): R106-R13.

53. van Rooij E, and Kauppinen, Sakari. Development of microRNA therapeutics is coming of age. EMBO Mol Med. 2014; 6(7): 851-864.

54. Wang X, Wang H-K, Li Y, Hafner M, Banerjee NS, Tang S, et al. microRNAs are biomarkers of oncogenic human papillomavirus infections. Proceedings of the National Academy of Sciences. 2014; 111(11): 4262-4267.

55. Tang W, Zhu J, Su S, Wu W, Liu Q, Su F, et al. MiR-27 as a Prognostic Marker for Breast Cancer Progression and Patient Survival. PLoS ONE. 2012; 7(12): e51702.

56. Nelson KM, Weiss GJ. MicroRNAs and cancer: past, present, and potential future. Molecular Cancer Therapeutics. 2008; 7(12): 3655-60.

57. Calin GA. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences. 2002; 99(24): 15524-9.

58. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(9): 2999-3004.

59. Shen J, Liu Z, Todd N, Zhang H, Liao J, Yu L, et al. Diagnosis of lung cancer in individuals with solitary pulmonary nodules by plasma microRNA biomarkers. BMC Cancer. 2011; 11(1): 374.

60. Starega-Roslan J, Koscianska E, Kozlowski P, Krzyzosiak W. The role of the precursor structure in the biogenesis of microRNA. Cell Mol Life Sci. 2011; 68(17): 2859-71.

61. Bartel D. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 2004; 116(2): 281-97.

62. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, et al. SNPs in human miRNA genes affect biogenesis and function. RNA. 2009; 15(9): 1640-51.

63. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet. 2010; 11(9): 597 - 610.

64. Jiang Y, Chen J, Wu J, Hu Z, Qin Z, Liu Xa, et al. Evaluation of genetic variants in microRNA biosynthesis genes and risk of breast cancer in Chinese women. International Journal of Cancer. 2013; 133(9): 2216-24.

65. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nature Reviews Cancer. 2010; 10(6): 389-402.

66. Seitz H, Ghildiyal M, Zamore PD. Argonaute loading improves the 5′ precision of both MicroRNAs and their miRNA* strands in flies. Curr Biol. 2008; 18: 147-51.

67. Ke HL, Chen M, Ye Y, Hildebrandt MAT, Wu WJ, Wei H, et al. Genetic variations in micro-RNA biogenesis genes and clinical outcomes in non-muscle-invasive bladder cancer. Carcinogenesis. 2013; 34(5): 1006-11.

68. Gong J, Tong Y, Zhang H, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat. 2012; 33(1): 254 - 63.

69. Salzman DW, Weidhaas JB. SNPing cancer in the bud: MicroRNA and microRNA-target site polymorphisms as diagnostic and prognostic biomarkers in cancer. Pharmacology & Therapeutics. 2013; 137(1): 55-63.

70. Lynam-Lennon N, Maher SG, Reynolds JV. The roles of microRNA in cancer and apoptosis. Biological Reviews. 2009; 84(1): 55-71.

71. Hoffman AE, Zheng T, Yi C, Leaderer D, Weidhaas J, Slack F, et al. microRNA miR-196a-2 and Breast Cancer: A Genetic and Epigenetic Association Study and Functional Analysis. Cancer Research. 2009; 69(14): 5970-7.

72. Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Human Molecular Genetics. 2007; 16(9): 1124-31.

73. Lewis B, Burge C, Bartel D. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 120(1): 15 - 20.

74. Gao Y. An insertion/deletion polymorphism at miRNA-122-binding site in the interleukin-1alpha 3′ untranslated region confers risk for hepatocellular carcinoma. Carcinogenesis. 2009; 30(12): 2064-9.

75. Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I, et al. A SNP in a let-7 microRNA Complementary Site in the KRAS 3' Untranslated Region Increases Non-Small Cell Lung Cancer Risk. Cancer Research. 2008; 68(20): 8535-40.

76. Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W. Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. Carcinogenesis. 2008; 29(7): 1306-11.

77. Christensen BC, Moyer BJ, Avissar M, Ouellet LG, Plaza SL, McClean MD, et al. A let-7 microRNA-binding site polymorphism in the KRAS 3' UTR is associated with reduced survival in oral cancers. Carcinogenesis. 2009; 30(6): 1003-7.

78. Ye Y, Wang KK, Gu J, Yang H, Lin J, Ajani JA, et al. Genetic Variations in MicroRNA-Related Genes Are Novel Susceptibility Loci for Esophageal Cancer Risk. Cancer Prevention Research. 2008; 1(6): 460-9.

79. Gaur A, Jewell DA, Liang Y, Ridzon D, Moore JH, Chen C, et al. Characterization of MicroRNA Expression Levels and Their Biological Correlates in Human Cancer Cell Lines. Cancer Research. 2007; 67(6): 2456-68.

80. Liang Y, Ridzon D, Wong L, Chen C. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics. 2007; 8: 166.

81. Muiños-Gimeno M, Montfort M, Bayés M, Estivill X, Espinosa-Parrilla Y. Design and evaluation of a panel of single-nucleotide polymorphisms in microRNA genomic regions for association studies in human disease. European Journal of Human Genetics. 2009; 18(2): 218-26.

82. Motoyama K. IH, Takatsuno Y., et al. Over- and under-expressed microRNAs in human colorectal cancer. International Journal of Oncology. 2009; 34(4): 1069-75.

83. Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R, Memeo L, et al. The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. Nat Med. 2008; 14(11): 1271-7.

84. Varambally S, Cao Q, Mani R-S, Shankar S, Wang X, Ateeq B, et al. Genomic Loss of microRNA-101 Leads to Overexpression of Histone Methyltransferase EZH2 in Cancer. Science. 2008; 322(5908): 1695-9.

85. Johnson CD, Esquela-Kerscher A, Stefani G, Byrom M, Kelnar K, Ovcharenko D, et al. The let-7 MicroRNA Represses Cell Proliferation Pathways in Human Cells. Cancer Research. 2007; 67(16): 7713-22.

86. Nasser MW, Datta J, Nuovo G, Kutay H, Motiwala T, Majumder S, et al. Down-regulation of Micro-RNA-1 (miR-1) in Lung Cancer: Suppression of tumorigenic property of lung cancer cells and their sensitization to doxorubicin-induced apoptosis by miR-1. Journal of Biological Chemistry. 2008; 283(48): 33394-405.

87. Welch C, Chen Y, Stallings RL. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. Oncogene. 2007; 26(34): 5017-22.

88. He X, He L, Hannon GJ. The Guardian's Little Helper: MicroRNAs in the p53 Tumor Suppressor Network. Cancer Research. 2007; 67(23): 11099-101.

89. Thapa DR, Li X, Jamieson BD, Martínez-Maza O. Overexpression of MicroRNAs from the miR-17-92 Paralog Clusters in AIDS-Related Non-Hodgkin's Lymphomas. PLoS ONE. 2011; 6(6): e20781.

90. Bracken AP CM, Cocito A, Helin K. E2F target genes: unraveling the biology. Trends Biochem Sci. 2004; 29(8): 409-17.

91. Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol. 2008; 9(4): 405-14.

92. Matsubara H, Takeuchi T, Nishikawa E, Yanagisawa K, Hayashita Y, Ebi H, et al. Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92. Oncogene. 2007; 26(41): 6099-105.

93. Li T, Li D, Sha J, Sun P, Huang Y. MicroRNA-21 directly targets MARCKS and promotes apoptosis resistance and invasion in prostate cancer cells. Biochemical and Biophysical Research Communications. 2009; 383(3): 280-5.

94. Volinia S, Calin GA, Liu C-G, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(7): 2257-61.

95. Pomerantz MM, Beckwith CA, Regan MM, Wyman SK, Petrovics G, Chen Y, et al. Evaluation of the 8q24 Prostate Cancer Risk Locus and MYC Expression. Cancer Research. 2009; 69(13): 5568-74.

96. Coppola V, De Maria R, Bonci D. MicroRNAs and prostate cancer. Endocr Relat Cancer. 2010; 17(1): F1-17.

97. Carissimi C, Fulci V, Macino G. MicroRNAs: Novel regulators of immunity. Autoimmunity Reviews. 2009; 8(6): 520-4.

98. O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D. Physiological and pathological roles for microRNAs in the immune system. Nature Reviews Immunology. 2010; 10(2): 111-22.

99. Pedersen IM, Otero D, Kao E, Miletic AV, Hother C, Ralfkiaer E, et al. Onco-miR-155 targets SHIP1 to promote TNFα-dependent growth of B cell lymphomas. EMBO Molecular Medicine. 2009; 1(5): 288-95.

100. O'Connell RM, Taganov KD, Boldin MP, Cheng G, Baltimore D. MicroRNA-155 is induced during the macrophage inflammatory response. Proceedings of the National Academy of Sciences. 2007; 104(5): 1604-9.

101. Fontana L, Pelosi E, Greco P, Racanicchi S, Testa U, Liuzzi F, et al. MicroRNAs 17-5p-20a-106a control monocytopoiesis through AML1 targeting and M-CSF receptor upregulation. Nat Cell Biol. 2007; 9(7): 775-87.

102. Xiao C, Rajewsky K. MicroRNA Control in the Immune System: Basic Principles. Cell. 2009; 136(1): 26-36.

103. Pauley KM, Cha S, Chan EKL. MicroRNA in autoimmunity and autoimmune diseases. Journal of Autoimmunity. 2009; 32(3-4): 189-94.

104. Li Q-J, Chau J, Ebert PJR, Sylvester G, Min H, Liu G, et al. miR-181a Is an Intrinsic Modulator of T Cell Sensitivity and Selection. Cell. 2007; 129(1): 147-61.

105. Caramuta S, Lee L, Özata DM, Akçakaya P, Georgii-Hemming P, Xie H, Amini RM, Lawrie CH, et al. Role of microRNAs and microRNA machinery in the pathogenesis of diffuse large B-cell lymphoma. Blood Cancer. 2013; 3(10): e152.

106. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, et al. Genetic dissection of the miR-17 92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes & Development. 2009; 23(24): 2806-11.

107. Mazan-Mamczarz K, Gartenhaus RB. Role of microRNA deregulation in the pathogenesis of diffuse large B-cell lymphoma (DLBCL). Leukemia Research. 2013; 37(11): 1420-8.

108. Ludwig L. RNA silencing and HIV: A hypothesis for the etiology of the severe combined immunodeficiency induced by the virus. Retrovirology. 2008; 5(1): 79.

109. Rai D. Coordinated expression of microRNA-155 and predicted target genes in diffuse large B-cell lymphoma. Cancer Genetics and Cytogenetics. 2008; 181(1): 8-15.

110. Rai D, Kim SW, McKeller MR, Dahia PLM, Aguiar RCT. Targeting of SMAD5 links microRNA-155 to the TGF- pathway and lymphomagenesis. Proceedings of the National Academy of Sciences. 2010; 107(7): 3111-6.

111. Robbiani DF, Bunting S, Feldhahn N, Bothmer A, Camps J, Deroubaix S, et al. AID Produces DNA Double-Strand Breaks in Non-Ig Genes and Mature B Cell Lymphomas with Reciprocal Chromosome Translocations. Molecular Cell. 2009; 36(4): 631-41.

112. Asslaber D, Pinon JD, Seyfried I, Desch P, Stocher M, Tinhofer I, et al. microRNA-34a expression correlates with MDM2 SNP309 polymorphism and treatment-free survival in chronic lymphocytic leukemia. Blood. 2010; 115(21): 4191-7.

113. Xu J, Liao X, Wong C. Downregulations of B-cell lymphoma 2 and myeloid cell leukemia sequence 1 by microRNA 153 induce apoptosis in a glioblastoma cell line DBTRG-05MG. International Journal of Cancer. 2010; 126(4): 1029-35.

114. Bueno MJ, Pérez de Castro I, Gómez de Cedrón M, Santos J, Calin GA, Cigudosa Juan C, et al. Genetic and Epigenetic Silencing of MicroRNA-203 Enhances ABL1 and BCR-ABL1 Oncogene Expression. Cancer Cell. 2008; 13(6): 496-506.

115. Zhao J-J, Lin J, Lwin T, Yang H, Guo J, Kong W, et al. microRNA expression profile and identification of miR-29 as a prognostic marker and pathogenetic factor by targeting CDK6 in mantle cell lymphoma. Blood. 2010; 115(13): 2630-9.

116. Negrini M, Ferracin M, Sabbioni S, Croce CM. MicroRNAs in human cancer: from research to therapy. Journal of Cell Science. 2007; 120(11): 1833-40.

117. Slaby O, Bienertova-Vasku J, Svoboda M, Vyzula R. Genetic polymorphisms and microRNAs: new direction in molecular epidemiology of solid cancer. Journal of Cellular and Molecular Medicine. 2012; 16(1): 8-21.

118. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, et al. Single-Nucleotide Polymorphisms Inside MicroRNA Target Sites Influence Tumor Susceptibility. Cancer Research. 2010; 70(7): 2789-98.

119. Iuliano R, Vismara MFM, Dattilo V, Trapasso F, Baudi F, Perrotti N. The Role of MicroRNAs in Cancer Susceptibility. BioMed Research International. 2013; vol. 2013(article ID 591931).

120. Ichimi T, Enokida H, Okuno Y, Kunimoto R, Chiyomaru T, Kawamoto K, et al. Identification of novel microRNA targets based on microRNA signatures in bladder cancer. International Journal of Cancer. 2009; 125(2): 345-52.

121. Teo MTW, Landi D, Taylor CF, Elliott F, Vaslin L, Cox DG, et al. The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes. Carcinogenesis. 2012; 33(3): 581-6.

122. Jiang Y, Qin Z, Hu Z, Guan X, Wang Y, He Y, et al. Genetic variation in a hsa-let-7 binding site in RAD52 is associated with breast cancer susceptibility. Carcinogenesis. 2013; 34(3): 689-93.

123. Gopal S, Achenbach CJ, Yanik EL, Dittmer DP, Eron JJ, Engels EA. Moving Forward in HIV-Associated Cancer. Journal of Clinical Oncology. 2014; 32(9): 876-80.

124. Lim ST, Karim R, Nathwani BN, Tulpule A, Espina B, Levine AM. AIDS-related Burkitt's lymphoma versus diffuse large-cell lymphoma in the pre-highly active antiretroviral therapy (HAART) and HAART eras: significant differences in survival with standard chemotherapy. J Clin Oncol. 2005; 23(19): 4430-8.

125. Chen YB, Rahemtullah A, Hochberg E. Primary effusion lymphoma. Oncologist. 2007; 12(5): 569-76.

126. Simonelli C, Spina M, Cinelli R, Talamini R, Tedeschi R, Gloghini A, et al. Clinical features and outcome of primary effusion lymphoma in HIV-infected patients: a single-institution study. J Clin Oncol. 2003; 21(21): 3948-54.

127. Riedel DJ, Gonzalez-Cuyar LF, Zhao XF, Redfield RR, Gilliam BL. Plasmablastic lymphoma of the oral cavity: a rapidly progressive lymphoma associated with HIV infection. The Lancet infectious diseases. 2008; 8(4): 261-7.

128. Preskill C, Weidhaas JB. SNPs in MicroRNA Binding Sites as Prognostic and Predictive Cancer Biomarkers. 2013; 18(4): 327-40.

129. Yang B, Liu C, Diao L, Wang C, Guo Z. A polymorphism at the microRNA binding site in the 3′ untranslated region of C14orf101 is associated with non-Hodgkin lymphoma overall survival. Cancer Genetics. 2014; 207(4): 141-6.

130. Dzikiewicz-Krawczyk A, Macieja A, May E, Januszkiewicz-Lewandowska D, Mosor M, Fichna M, et al. Polymorphisms in microRNA target sites modulate risk of lymphoblastic and myeloid leukemias and affect microRNA binding. Journal of Hematology & Oncology. 2014; 7(1): 43.

131. Li X, Tian X, Zhang B, Chen J. Polymorphisms in MicroRNA-Related Genes Are Associated With Survival of Patients With T-Cell Lymphoma. The Oncologist. 2014; 19(3): 243-9.

132. López-López E, Gutiérrez-Camino Á, Piñán MÁ, Sánchez-Toledo J, Uriz JJ, Ballesteros J, et al. Pharmacogenetics of MicroRNAs and MicroRNAs Biogenesis Machinery in Pediatric Acute Lymphoblastic Leukemia. PLoS ONE. 2014; 9(3): e91261.

133. Navarro A, Muñoz C, Gaya A, Díaz-Beyá M, Gel B, Tejero R, et al. MiR-SNPs as Markers of Toxicity and Clinical Outcome in Hodgkin Lymphoma Patients. PLoS ONE. 2013; 8(5): e64716.

134. Lawrie CH. MicroRNAs and lymphomagenesis: a functional review. British Journal of Haematology. 2013; 160(5): 571-81.

135. Multicenter AIDS Cohort Study. CAMACS MACS Dossier (Accessed March 27th, 2014 http://www.statepi.jhsph.edu/macs/macs.html).

136. Detels R, Jacobson L, Margolick J, Martinez-Maza O, Muñoz A, Phair J, et al. The multicenter AIDS Cohort Study, 1983 to …. Public Health. 2012; 126(3): 196-8.

137. Becker JT, Kingsley LA, Molsberry S, Reynolds S, Aronow A, Levine AJ, et al. Cohort Profile: Recruitment cohorts in the neuropsychological substudy of the Multicenter AIDS Cohort Study. International Journal of Epidemiology. 2014; doi: 10.1093/ije/dyu092.

138. Fluidigm SNP Genotyping Analysis Version 3.1.2 Build 20111017.1807. South San Francisco CA, USA.

139. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Research. 2009; 19(1): 92-105.

140. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. Molecular Cell. 2007; 27(1): 91-105.

141. Lewis BP, Burge CB, Bartel DP. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell. 2005; 120(1): 15-20.

142. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell. 2003; 115: 787-98.

143. Landi D. MicroRNA and Cancer: Methods and Protocols, Methods in Molecular Biology, vol. 676, Chapter 14: Prediction of the Biological Effect of Polymorphisms Within MicroRNA Binding Sites, pgs 197-210., DOI 10.1007/978-1-60761-863-8_14 SSBM, LLC 2011.

144. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, et al. Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. Carcinogenesis. 2008; 29(3): 579-84.

145. Chen and Chen. Wei Wu (ed.) MaCMaP, Methods in Molecular Biology, vol. 676, Chapter 16: Analysis of Targets and Functions Coregulated by MicroRNAs, pgs 225-241., DOI 10.1007/978-1-60761-863-8_14 SSBM, LLC 2011.

146. Wang YP, Li KB. Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. BMC Genomics. 2009; 10(1): 218.

147. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008; 455(7209): 58-63.

148. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 2008; 455(7209): 64-71.

149. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000; 28(1): 27-30.

150. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research. 2010; 38(suppl 1): D355-D60.

151. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Research. 2006; 34(suppl 1): D354-D7.

152. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Research. 2004; 32(suppl 1): D277-D80.

153. Liu Q, Fu H, Sun F, Zhang H, Tie Y, Zhu J, et al. miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. Nucleic Acids Research. 2008; 36(16): 5391-404.

154. Cloonan N, Brown M, Steptoe A, Wani S, Chan W, Forrest A, et al. The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. Genome Biology. 2008; 9(8): R127.

155. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research. 2001; 29(1): 308-11.

156. Clifford GM, Franceschi S. Cancer risk in HIV-infected persons: influence of CD4+ count. Future Oncology. 2009; 5(5): 669-78.

157. 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. The Lancet Oncology. 2009; 10(12): 1152-9.

158. Silverberg MJ, Chao C, Abrams DI. New insights into the role of HIV infection on cancer risk. The Lancet Oncology. 2009; 10(12): 1133-4.

159. Geyer SM, Morton LM, Habermann TM, Allmer C, Davis S, Cozen W, et al. Smoking, alcohol use, obesity, and overall survival from non-Hodgkin lymphoma. Cancer. 2010; 116(12): 2993-3000.

160. Castillo JJ, Dalia S. Cigarette smoking is associated with a small increase in the incidence of non-Hodgkin lymphoma: a meta-analysis of 24 observational studies. Leukemia & Lymphoma. 2012; 53(10): 1911-9.

161. Diver W PA, Thun M, et al. The association between cigarette smoking and non-Hodgkin lymphoid neoplasms in a large US cohort study. Cancer Cases and Control 2012; 23(8): 1231-40.

162. Sergentanis TN, Kanavidis P, Michelakos T, Petridou ET. Cigarette smoking and risk of lymphoma in adults: a comprehensive meta-analysis on Hodgkin and non-Hodgkin disease. European Journal of Cancer Prevention. 2013; 22(2): 131-150.

163. Morton L, Zheng T, Holford T, et al. Alcohol consumption and risk of non-Hodgkin lymphoma: a pooled analysis. The Lancet Oncology 2005; 6(7): 469-76.

164. Gapstur SM, Diver WR, McCullough ML, Teras LR, Thun MJ, Patel AV. Alcohol Intake and the Incidence of Non-Hodgkin Lymphoid Neoplasms in the Cancer Prevention Study II Nutrition Cohort. American Journal of Epidemiology. 2012; 176(1): 60-9.

165. Tramacere I, Pelucchi C, Bonifazi M, Bagnardi V, Rota M, Bellocco R, et al. Alcohol drinking and non-Hodgkin lymphoma risk: a systematic review and a meta-analysis. Annals of Oncology. 2012; 23(11): 2791-8.

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

167. Holly E LA, Bracci P, et al. Case-Control Study of Non-Hodgkin's Lymphoma among Women and Heterosexual Men in the San Francisco Bay Area, California. American Journal of Epidemiology. 1999; 150(4): 375-89.

168. Abdelmonem Afifi VC, Susanne May. Computer-Aided Multivariate Analysis. 4 ed. Boca Raton: Chapman and Hall, CRC; 2004: 281-329.

169. Chang PY, Detels R, Martínez-Maza O, Zhang ZF, Jacobson LP, Margolick JB, et al. Comment on "Characteristics of B-Cell Lymphomas in HIV/HCV-Coinfected Patients During the Combined Antiretroviral Therapy Era: An ANRS CO16 LYMPHOVIR Cohort Study". JAIDS Journal of Acquired Immune Deficiency Syndromes. 2014; Publish Ahead of Print: 10.1097/QAI.0000000000000204.

170. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume I - The Analysis of Case-Control Studies. Lyon, International Agency for Research on Cancer (IARC Scientific Publications No 32). 1980: 252-79.

171. Dong C, Hemminki K. Second primary neoplasms in 633,964 cancer patients in Sweden, 1958–1996. International Journal of Cancer. 2001; 93(2): 155-61.

172. Royle JS, Baade P, Joske D, Fritschi L. Risk of second cancer after lymphohematopoietic neoplasm. International Journal of Cancer. 2011; 129(4): 910-9.

173. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology. 2006; 59(10): 1087-91.

174. Desai M, Kubo J, Esserman D, Terry MB. The Handling of Missing Data in Molecular Epidemiology Studies. Cancer Epidemiology Biomarkers & Prevention. 2011; 20(8): 1571-9.

175. van der Heijden GJMG, T. Donders AR, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. Journal of Clinical Epidemiology. 2006; 59(10): 1102-9.

176. Greenland S. A semi-bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer-mortality study. Statistics in Medicine. 1992; 11(2): 219-30.

177. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. International Journal of Epidemiology. 2006; 35(3): 765-75.

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

179. Sullivan SG, Greenland S. Re: Sullivan SG, Greenland S. Bayesian regression in SAS software. Int J Epidemiol 2013;42:308–17. International Journal of Epidemiology. 2014; 43(3): 974.

180. Greenland S, Robins JM. Empirical-Bayes Adjustments for Multiple Comparisons Are Sometimes Useful. Epidemiology. 1991; 2(4): 244-51.

181. Yang T, Chang P-Y, Park S, Bastani D, Chang S-C, Morgenstern H, et al. Tobacco smoking, NBS1 polymorphisms, and survival in lung and upper aerodigestive tract cancers with semi-Bayes adjustment for hazard ratio variation. Cancer Causes & Control. 2014; 25(1): 11-23.

182. Chang P, Chang S, Zhang ZF, and Greenland S. (Working draft and title) Semi-Bayes (SB) Approach for Correction of Potential False-positive Findings of Main-effects, Stratum-specific Associations, and Product-term Coefficients: Application in Cancer Molecular Epidemiology. 2014.

183. Sullivan SG, Greenland S. Bayesian regression in SAS software. Int J Epidemiol. 2013; 42: 308-17.

184. Greenland S, Christensen R. Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. Statistics in Medicine. 2001; 20(16): 2421-8.

185. Thapa DR, Hussain SK, Tran W-C, D'souza G, Bream JH, Achenback CJ, et al. Serum microRNAs in HIV-infected individuals as pre-diagnosis biomarkers for AIDS-related non-Hodgkin lymphomas (AIDS-NHL). JAIDS Journal of Acquired Immune Deficiency Syndromes. 2014; 66(2): 229-37.

186. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Research. 2002; 30(17): 3894-900.

187. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Chapter 7: Current Protocols in Human Genetics: John Wiley & Sons, Inc.; 2001: 7.20.1-7.20.41.

188. Cai Q, Guo Y, Xiao B, Banerjee S, Saha A, Lu J, et al. Epstein-Barr Virus Nuclear Antigen 3C Stabilizes Gemin3 to Block p53-mediated Apoptosis. PLoS Pathog. 2011; 7(12): e1002418.

189. Wang J, Wang Q, Liu H, Shao N, Tan B, Zhang G, et al. The association of miR-146a rs2910164 and miR-196a2 rs11614913 polymorphisms with cancer risk: a meta-analysis of 32 studies. Mutagenesis. 2012; 27(6): 779-88.

190. Wang F, Ma Y-L, Zhang P, Yang J-J, Chen H-Q, Liu Z-H, et al. A genetic variant in microRNA-196a2 is associated with increased cancer risk: a meta-analysis. Molecular Biology Reports. 2012; 39(1): 269-75.

191. Kristinsson S, Rosenquist R, Landgren O, Goldin L, Björkholm M. Genetics in Lymphomagenesis. In: Wiernik PH, Goldman JM, Dutcher JP, Kyle RA, editors. Neoplastic Diseases of the Blood: Springer New York; 2013. p. 835-47.

192. Liao J-M, Cao B, Zhou X, Lu H. New insights into p53 functions through its target microRNAs. Journal of Molecular Cell Biology. 2014; 6(3): 206-13.

193. Boominathan L. The Tumor Suppressors p53, p63, and p73 Are Regulators of MicroRNA Processing Complex. PLoS ONE. 2010; 5(5): e10615.

194. Sun Q, Gu H, Zeng Y, Xia Y, Wang Y, Jing Y, et al. Hsa-mir-27a genetic variant contributes to gastric cancer susceptibility through affecting miR-27a and target gene expression. Cancer Science. 2010; 101(10): 2241-7.

195. Zhong S, Chen Z, Xu J, Li W, Zhao J. Pre-mir-27a rs895819 polymorphism and cancer risk: a meta-analysis. Molecular Biology Reports. 2013; 40(4): 3181-6.

196. Zhang D, Yang G, Chen X, Li C, Wang L, Liu Y, et al. mir-300 Promotes Self-Renewal and Inhibits the Differentiation of Glioma Stem-Like Cells. J Mol Neurosci. 2014: epub 1-8.

197. Leng RP, Lin Y, Ma W, Wu H, Lemmers B, Chung S, et al. Pirh2, a p53-Induced Ubiquitin-Protein Ligase, Promotes p53 Degradation. Cell. 2003; 112(6): 779-91.

198. Bergiers I, Bridoux L, Nguyen N, Twizere J-C, Rezsöhazy R. The Homeodomain Transcription Factor Hoxa2 Interacts with and Promotes the Proteasomal Degradation of the E3 Ubiquitin Protein Ligase RCHY1. PLoS ONE. 2013; 8(11): e80387.

199. Liu J, Wei M, He Y, et al. Genetic Variants in the MicroRNA Machinery Gene *GEMIN4* Are Associated with Risk of Prostate Cancer: A Case-control Study of the Chinese Han Population. DNA and Cell Biology. 2012; 31(7): 1296-1302.

200. Yang P-W, Huang Y-C, Hsieh C-Y, Hua K-T, Huang Y-T, Chiang T-H, et al. Association of miRNA-related Genetic Polymorphisms and Prognosis in Patients with Esophageal Squamous Cell Carcinoma. Annals of Surgical Oncology. 2014: epub 1-9.

201. Liang D, Meyer L, Chang DW, Lin J, Pu X, Ye Y, et al. Genetic Variants in MicroRNA Biosynthesis Pathways and Binding Sites Modify Ovarian Cancer Risk, Survival, and Treatment Response. Cancer Research. 2010; 70(23): 9765-76.

202. Hu X, Fang Y, Zheng J, He Y, Zan X, Lin S, et al. The association between HIF-1α polymorphism and cancer risk: a systematic review and meta-analysis. Tumor Biology. 2014; 35(2): 903-16.

203. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. 2014; 42(1): D86-91.

204. Hernandez-Luna MA, Rocha-Zavaleta L, Vega MI, Huerta-Yepez S. Hypoxia inducible factor-1α induces chemoresistance phenotype in non-Hodgkin lymphoma cell line via up-regulation of Bcl-xL. Leukemia & Lymphoma. 2013; 54(5): 1048-55.

205. Wincewicz A, Sulkowska M, Koda M, Sulkowski S. Cumulative expression of HIF-1-alpha, Bax, Bcl-xL and P53 in human colorectal cancer. Pathology. 2007; 39(3): 334-8.

206. Greijer A, Van der Wall E. The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. Journal of clinical pathology. 2004; 57(10): 1009-14.

207. Hammond EM, Giaccia AJ. The role of p53 in hypoxia-induced apoptosis. Biochemical and biophysical research communications. 2005; 331(3): 718-25.

208. Semenza GL. HIF-1 and tumor progression: pathophysiology and therapeutics. Trends in molecular medicine. 2002; 8(4): S62-S7.

209. Bertout JA, Patel SA, Simon MC. The impact of O2 availability on human cancer. Nature Reviews Cancer. 2008; 8(12): 967-75.

210. Ryan HE, Poloni M, McNulty W, Elson D, Gassmann M, Arbeit JM, et al. Hypoxia-inducible factor-1α is a positive factor in solid tumor growth. Cancer Res. 2000; 60(15): 4010.

211. Evens AM, Schumacker PT, Helenowski IB, Singh ATK, Dokic D, Keswani A, et al. Hypoxia inducible factor-alpha activation in lymphoma and relationship to the thioredoxin family. British Journal of Haematology. 2008; 141(5): 676-80.

212. Darekar S, Georgiou K, Yurchenko M, Yenamandra SP, Chachami G, Simos G, et al. Epstein-Barr Virus Immortalization of Human B-Cells Leads to Stabilization of Hypoxia-Induced Factor 1 Alpha, Congruent with the Warburg Effect. PLoS ONE. 2012; 7(7): e42072.

213. Melo SA, Moutinho C, Ropero S, Calin GA, Rossi S, Spizzo R, et al. A Genetic Defect in Exportin-5 Traps Precursor MicroRNAs in the Nucleus of Cancer Cells. Cancer Cell. 2010; 18(4): 303-15.

214. Lasken RS, Egholm M. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. Trends in Biotechnology. 2003; 21(12): 531-5.

215. Hughes S, Arneson N, Done S, Squire J. The use of whole genome amplification in the study of human disease. Progress in Biophysics and Molecular Biology. 2005; 88(1): 173-89.

216. Macaulay IC, Voet T. Single Cell Genomics: Advances and Future Perspectives. PLoS Genet. 2014; 10(1): e1004126.

217. Epeldegui M, Hung YP, McQuay A, Ambinder RF, Martínez-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. Molecular Immunology. 2007; 44(5): 934-42.

218. Chen X-Z, Zhang W-K, Yang K, Wang L-L, Liu J, Wang L, et al. Correlation between serum CA724 and gastric cancer: multiple analyses based on Chinese population. Molecular Biology Reports. 2012; 39(9): 9031-9.

219. Harima Y, Ikeda K, Utsunomiya K, Komemushi A, Kanno S, Shiga T, et al. Apolipoprotein C-II Is a Potential Serum Biomarker as a Prognostic Factor of Locally Advanced Cervical Cancer After Chemoradiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2013; 87(5): 1155-61.

220. Khemka VK, Bagchi D, Ghosh A, Sen O, Bir A, Chakrabarti S, et al. Raised Serum Adenosine Deaminase Level in Nonobese Type 2 Diabetes Mellitus. The Scientific World Journal. 2013; 2013: 404320.

221. Chan Y, Walmsley RP. Learning and Understanding the Kruskal-Wallis One-Way Analysis-of-Variance-by-Ranks Test for Differences Among Three or More Independent Groups. Physical Therapy. 1997; 77(12): 1755-61.

222. Fan C, Zhang D. A Note on Power and Sample Size Calculations for the Kruskal–Wallis Test for Ordered Categorical Data. Journal of Biopharmaceutical Statistics. 2012; 22(6): 1162-73.

223. Rosner B, Glynn RJ. Power and Sample Size Estimation for the Wilcoxon Rank Sum Test with Application to Comparisons of C Statistics from Alternative Prediction Models. Biometrics. 2009; 65(1): 188-97.